Statistics for engineers and scientists 5th edition by william navidi pdf

Page i Statistics for Engineers and Scientists Fifth Edition William Navidi Colorado School of Mines Page ii STATISTICS FOR ENGINEERS AND SCIENTISTS, FIFTH EDITION Published by McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121. Copyright © 2020 by McGraw-Hill Education. All rights reserved. Printed in the United States of America. Previous editions © 2015, 2011, and 2008. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of McGraw-Hill Education, including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning. Some ancillaries, including electronic and print components, may not be available to customers outside the United States. This book is printed on acid-free paper. 1 2 3 4 5 6 7 8 9 LCR 21 20 19 ISBN 978-1-259-71760-4 (bound edition) MHID 1-259-71760-7 (bound edition) ISBN 978-1-260-43099-8 (loose-leaf edition) MHID 1-260-43099-5 (loose-leaf edition) Portfolio Manager: Thomas Scaife, Ph.D. Product Developer: Tina Bower Marketing Manager: Shannon O'Donnell Content Project Managers: Jeni McAtee, Samantha Donisi-Hamm Buyer: Susan K. Culbertson Design: Debra Kubiak Content Licensing Specialist: Lorraine Buczek Cover images created by William Navidi using MATLAB 7.10.0, Release 2010A. Compositor: Aptara, Inc "Portions of information contained in this publication/book are printed with permission of Minitab Inc. All such material remains the exclusive property and copyright of Minitab Inc. All rights reserved." All credits appearing on page or at the end of the book are considered to be an extension of the copyright page. Library of Congress Cataloging-in-Publication Data Names: Navidi, William Cyrus, author. | Colorado School of Mines. Title: Statistics for engineers and scientists / William Navidi (Colorado School of Mines). Description: Fifth edition. | New York, NY : McGraw-Hill Education, [2020] | Prepared for the Colorado School of Mines. | Includes bibliographical references and index. Identifiers: LCCN 2018022787| ISBN 9781259717604 (alk. paper) | ISBN 1259717607 (alk. paper) Subjects: LCSH: Mathematical statistics—Simulation methods. | Bootstrap (Statistics) | Linear models (Statistics) | Engineering—Statistical methods. | Science—Statistical methods. Classification: LCC QA276.4 .N38 2020 | DDC 519.5—dc23 LC record available at The Internet addresses listed in the text were accurate at the time of publication. The inclusion of a website does not indicate an endorsement by the authors or McGraw-Hill Education, and McGraw-Hill Education does not guarantee the accuracy of the information presented at these sites. mheducation.com/highered To Catherine, Sarah, and Thomas Page iii Page iv ABOUT THE AUTHOR William Navidi is Professor of Applied Mathematics and Statistics at the Colorado School of Mines. He received his B.A. degree in mathematics from New College, his M.A. in mathematics from Michigan State University, and his Ph.D. in statistics from the University of California at Berkeley. Professor Navidi has authored more than 70 research papers both in statistical theory and in a wide variety of applications including computer networks, epidemiology, molecular biology, chemical engineering, and geophysics. Page v BRIEF CONTENTS Preface xi 1 Sampling and Descriptive Statistics 1 2 Probability 48 3 Propagation of Error 164 4 Commonly Used Distributions 200 5 Confidence Intervals 323 6 Hypothesis Testing 402 7 Correlation and Simple Linear Regression 514 8 Multiple Regression 603 9 Factorial Experiments 669 10 Statistical Quality Control 772 Appendix A: Tables 811 Appendix B: Partial Derivatives 836 Appendix C: Bibliography 838 Answers to Odd-Numbered Exercises 841 Index 912 Page vii CONTENTS Preface xi Chapter 1 Sampling and Descriptive Statistics 1 Introduction 1 1.1 Sampling 3 1.2 Summary Statistics 13 1.3 Graphical Summaries 25 Chapter 2 Probability 48 Introduction 48 2.1 Basic Ideas 48 2.2 Counting Methods 62 2.3 Conditional Probability and Independence 69 2.4 Random Variables 90 2.5 Linear Functions of Random Variables 116 2.6 Jointly Distributed Random Variables 127 Chapter 3 Propagation of Error 164 Introduction 164 3.1 Measurement Error 164 3.2 Linear Combinations of Measurements 170 3.3 Uncertainties for Functions of One Measurement 180 3.4 Uncertainties for Functions of Several Measurements 186 Chapter 4 Commonly Used Distributions 200 Introduction 200 4.1 The Bernoulli Distribution 200 4.2 The Binomial Distribution 203 4.3 The Poisson Distribution 215 4.4 Some Other Discrete Distributions 230 4.5 The Normal Distribution 241 4.6 The Lognormal Distribution 256 4.7 The Exponential Distribution 262 4.8 Some Other Continuous Distributions 272 4.9 Some Principles of Point Estimation 280 4.10 Probability Plots 285 4.11 The Central Limit Theorem 290 4.12 Simulation 303 Chapter 5 Confidence Intervals 323 Introduction 323 5.1 Large-Sample Confidence Intervals for a Population Mean 324 5.2 Confidence Intervals for Proportions 339 5.3 Small-Sample Confidence Intervals for a Population Mean 345 5.4 Confidence Intervals for the Difference Between Two Means 355 5.5 Confidence Intervals for the Difference Between Two Proportions 359 5.6 Small-Sample Confidence Intervals for the Difference Between Two Means 364 5.7 Confidence Intervals with Paired Data 371 5.8 Confidence Intervals for the Variance and Standard Deviation of a Normal Population 376 5.9 Prediction Intervals and Tolerance Intervals 381 Page viii 5.10 Using Simulation to Construct Confidence Intervals 385 Chapter 6 Hypothesis Testing 402 Introduction 402 6.1 Large-Sample Tests for a Population Mean 402 6.2 Drawing Conclusions from the Results of Hypothesis Tests 412 6.3 Tests for a Population Proportion 420 6.4 Small-Sample Tests for a Population Mean 425 6.5 Large-Sample Tests for the Difference Between Two Means 431 6.6 Tests for the Difference Between Two Proportions 437 6.7 Small-Sample Tests for the Difference Between Two Means 443 6.8 Tests with Paired Data 452 6.9 Distribution-Free Tests 458 6.10 Tests with Categorical Data 467 6.11 Tests for Variances of Normal Populations 477 6.12 Fixed-Level Testing 483 6.13 Power 488 6.14 Multiple Tests 497 6.15 Using Simulation to Perform Hypothesis Tests 501 Chapter 7 Correlation and Simple Linear Regression 514 Introduction 514 7.1 Correlation 514 7.2 The Least-Squares Line 532 7.3 Uncertainties in the Least-Squares Coefficients 549 7.4 Checking Assumptions and Transforming Data 570 Chapter 8 Multiple Regression 603 Introduction 603 8.1 The Multiple Regression Model 603 8.2 Confounding and Collinearity 621 8.3 Model Selection 630 Chapter 9 Factorial Experiments 669 Introduction 669 9.1 One-Factor Experiments 669 9.2 Pairwise Comparisons in One-Factor Experiments 694 9.3 Two-Factor Experiments 707 9.4 Randomized Complete Block Designs 732 9.5 2P Factorial Experiments 742 Page ix Chapter 10 Statistical Quality Control 772 Introduction 772 10.1 Basic Ideas 772 10.2 Control Charts for Variables 775 10.3 Control Charts for Attributes 795 10.4 The CUSUM Chart 800 10.5 Process Capability 804 Appendix A: Tables 811 Appendix B: Partial Derivatives 836 Appendix C: Bibliography 838 Answers to Odd-Numbered Exercises 841 Index 912 Page xi PREFACE MOTIVATION The idea for this book grew out of discussions between the statistics faculty and the engineering faculty at the Colorado School of Mines regarding our introductory statistics course for engineers. Our engineering faculty felt that the students needed substantial coverage of propagation of error, as well as more emphasis on model-fitting skills. The statistics faculty believed that students needed to become more aware of some important practical statistical issues such as the checking of model assumptions and the use of simulation. My view is that an introductory statistics text for students in engineering and science should offer all these topics in some depth. In addition, it should be flexible enough to allow for a variety of choices to be made regarding coverage, because there are many different ways to design a successful introductory statistics course. Finally, it should provide examples that present important ideas in realistic settings. Accordingly, the book has the following features: • The book is flexible in its presentation of probability, allowing instructors wide latitude in choosing the depth and extent of their coverage of this topic. • The book contains many examples that feature real, contemporary data sets, both to motivate students and to show connections to industry and scientific research. • The book contains many examples of computer output and exercises suitable for solving with computer software. • The book provides extensive coverage of propagation of error. • The book presents a solid introduction to simulation methods and the bootstrap, including applications to verifying normality assumptions, computing probabilities, estimating bias, computing confidence intervals, and testing hypotheses. • The book provides more extensive coverage of linear model diagnostic procedures than is found in most introductory texts. This includes material on examination of residual plots, transformations of variables, and principles of variable selection in multivariate models. • The book covers the standard introductory topics, including descriptive statistics, probability, confidence intervals, hypothesis tests, linear regression, factorial experiments, and statistical quality control. MATHEMATICAL LEVEL Most of the book will be mathematically accessible to those whose background includes one semester of calculus. The exceptions are multivariate propagation of error, which requires partial derivatives, and joint probability distributions, which require multiple integration. These topics may be skipped on first reading, if desired. Page xii COMPUTER USE Over the past 40 years, the development of fast and cheap computing has revolutionized statistical practice; indeed, this is one of the main reasons that statistical methods have been penetrating ever more deeply into scientific work. Scientists and engineers today must not only be adept with computer software packages, they must also have the skill to draw conclusions from computer output and to state those conclusions in words. Accordingly, the book contains exercises and examples that involve interpreting, as well as generating, computer output, especially in the chapters on linear models and factorial experiments. Many statistical software packages are available for instructors who wish to integrate their use into their courses, and this book can be used effectively with any of these packages. The modern availability of computers and statistical software has produced an important educational benefit as well, by making simulation methods accessible to introductory students. Simulation makes the fundamental principles of statistics come alive. The material on simulation presented here is designed to reinforce some basic statistical ideas, and to introduce students to some of the uses of this powerful tool. CONTENT Chapter 1 covers sampling and descriptive statistics. The reason that statistical methods work is that samples, when properly drawn, are likely to resemble their populations. Therefore Chapter 1 begins by describing some ways to draw valid samples. The second part of the chapter discusses descriptive statistics. Chapter 2 is about probability. There is a wide divergence in preferences of instructors regarding how much and how deeply to cover this subject. Accordingly, I have tried to make this chapter as flexible as possible. The major results are derived from axioms, with proofs given for most of them. This should enable instructors to take a mathematically rigorous approach. On the other hand, I have attempted to illustrate each result with an example or two, in a scientific context where possible, that is designed to present the intuition behind the result. Instructors who prefer a more informal approach may therefore focus on the examples rather than the proofs. Chapter 3 covers propagation of error, which is sometimes called "error analysis" or, by statisticians, "the delta method." The coverage is more extensive than in most texts, but because the topic is so important to many engineers I thought it was worthwhile. The presentation is designed to enable instructors to adjust the amount of coverage to fit the needs of of the course. In particular, Sections 3.2 through 3.4 can be omitted without loss of continuity. Chapter 4 presents many of the probability distribution functions commonly used in practice. Point estimation, probability plots and the Central Limit Theorem are also covered. The final section introduces simulation methods to assess normality assumptions, compute probabilities, and estimate bias. Page xiii Chapters 5 and 6 cover confidence intervals and hypothesis testing, respectively. The P-value approach to hypothesis testing is emphasized, but fixed-level testing and power calculations are also covered. The multiple testing problem is covered in some depth. Simulation methods to compute confidence intervals and to test hypotheses are introduced as well. Chapter 7 covers correlation and simple linear regression. I have worked hard to emphasize that linear models are appropriate only when the relationship between the variables is linear. This point is all the more important since it is often overlooked in practice by engineers and scientists (not to mention statisticians). It is not hard to find in the scientific literature straight-line fits and correlation coefficient summaries for plots that show obvious curvature or for which the slope of the line is determined by a few influential points. Therefore this chapter includes a lengthy section on checking model assumptions and transforming variables. Chapter 8 covers multiple regression. Model selection methods are given particular emphasis, because choosing the variables to include in a model is an essential step in many reallife analyses. The topic of confounding is given careful treatment as well. Chapter 9 discusses some commonly used experimental designs and the methods by which their data are analyzed. One-way and two-way analysis of variance methods, along with randomized complete block designs and 2p factorial designs, are covered fairly extensively. Chapter 10 presents the topic of statistical quality control, discussing control charts, CUSUM charts, and process capability; and concluding with a brief discussion of sixsigma quality. NEW FOR THIS EDITION The fifth edition of this book is intended to extend the strengths of the fourth. Some of the changes are: • A large number of new exercises have been included, many of which involve real data from recently published sources. • Many examples have been updated. • Material on resistance to outliers has been added to Chapter 1. • Chapter 7 now contains material on interpreting the slope of the least-squares line. • The exposition has been improved in a number of places. RECOMMENDED COVERAGE The book contains enough material for a year-long course. For a one-semester course, there are a number of options. In our three-hour course at the Colorado School of Mines, we cover all of the first four chapters, except for joint distributions, the more theoretical aspects of point estimation, and the exponential, gamma, and Weibull distributions. We then cover the material on confidence intervals and hypothesis testing in Chapters 5 and 6, going quickly over the Page xiv two-sample methods and power calculations and omitting distribution-free methods and the chi-square and F tests. We finish by covering as much of the material on correlation and simple linear regression in Chapter 7 as time permits. A course with a somewhat different emphasis can be fashioned by including more material on probability, spending more time on two-sample methods and power, and reducing coverage of propagation of error, simulation, or regression. Many other options are available; for example, one may choose to include material on factorial experiments in place of some of the preceding topics. INSTRUCTOR RESOURCES The following resources are available on the book website www.mhhe.com/navidi. • Solutions Manual • PowerPoint Lecture Notes • Suggested Syllabi ACKNOWLEDGMENTS I am indebted to many people for contributions at every stage of development. I received valuable suggestions from my colleagues Barbara Moskal, Gus Greivel, Ashlyn Munson, and Melissa Laeser at the Colorado School of Mines. Mike Colagrosso developed some excellent applets, and Lesley Strawderman developed PowerPoint slides to supplement the text. I am particularly grateful to Jack Miller of the University of Michigan, who has corrected many errors and made many valuable suggestions for improvement. The staff at McGraw-Hill has been extremely capable and supportive. In particular, I would like to express my thanks to Product Developer Tina Bower, Content Project Manager Jeni McAtee, and Senior Portfolio Manager Thomas Scaife for their patience and guidance in the preparation of this edition. William Navidi Key Features Page xv Content Overview This book allows flexible coverage because there are many ways to design a successful introductory statistics course. • Flexible coverage of probability addresses the needs of different courses. Allowing for a mathematically rigorous approach, the major results are derived from axioms, with proofs given for most of them. On the other hand, each result is illustrated with an example or two to promote intuitive understanding. Instructors who prefer a more informal approach may therefore focus on the examples rather than the proofs and skip the optional sections. • Extensive coverage of propagation of error , sometimes called "error analysis" or "the delta method," is provided in a separate chapter. The coverage is more thorough than in most texts. The format is flexible so that the amount of coverage can be tailored to the needs of the course. • A solid introduction to simulation methods and the bootstrap is presented in the final sections of Chapters 4, 5, and 6. • Extensive coverage of linear model diagnostic procedures in Chapter 7 includes a lengthy section on checking model assumptions and transforming variables. The chapter emphasizes that linear models are appropriate only when the relationship between the variables is linear. This point is all the more important since it is often overlooked in practice by engineers and scientists (not to mention statisticians). Real-World Data Sets With a fresh approach to the subject, the author uses contemporary real-world data sets to motivate students and show a direct connection to industry and research. Computer Output The book contains exercises and examples that involve interpreting, as well as generating, computer output. Students—study more efficiently, retain more and achieve better outcomes. Instructors—focus on what you love—teaching. Page xvi SUCCESSFUL SEMESTERS INCLUDE CONNECT FOR INSTRUCTORS You're in the driver's seat. Want to build your own course? No problem. Prefer to use our turnkey, prebuilt course? Easy. Want to make changes throughout the semester? Sure. And you'll save time with Connect's auto-grading too. They'll thank you for it. Adaptive study resources like SmartBook® help your students be better prepared in less time. You can transform your class time from dull definitions to dynamic debates. Hear from your peers about the benefits of Connect at www.mheducation.com/highered/connect Make it simple, make it affordable. Connect makes it easy with seamless integration using any of the major Learning Management Systems—Blackboard®, Canvas, and D2L, among others—to let you organize your course in one convenient location. Give your students access to digital mate rials at a discount with our inclusive access program. Ask your McGraw-Hill representative for more information. ©Hill Street Studios/Tobin Rogers/Blend Images LLC Solutions for your challenges. A product isn't a solution. Real solutions are affordable, reliable, and come with training and ongoing support when you need it and how you want it. Our Customer Experience Group can also help you troubleshoot tech problems — although Connect's 99% uptime means you might not need to call them. See for yourself at status.mheducation.com Page xvii FOR STUDENTS ©Shutterstock/wavebreakmedia Effective, efficient studying. Connect helps you be more productive with your study time and get better grades using tools like SmartBook, which highlights key concepts and creates a personalized study plan. Connect sets you up for success, so you walk into class with confidence and walk out with better grades. "I really liked this app — it made it easy to study when you don't have your textbook in front of you. " - Jordan Cunningham, Eastern Washington University Study anytime, anywhere. Download the free ReadAnywhere app and access your online eBook when it's convenient, even if you're offline. And since the app automatically syncs with your eBook in Connect, all of your notes are available every time you open it. Find out more at www.mheducation.com/readanywhere No surprises. The Connect Calendar and Reports tools keep you on track with the work you need to get done and your assignment scores. Life gets busy; Connect tools help you keep learning through it all. Learning for everyone. McGraw-Hill works directly with Accessibility Services Departments and faculty to meet the learning needs of all students. Please contact your Accessibility Services office and ask them to email m, or visit www.mheducation.com/accessibility for more information. Page 1 Chapter 1 Sampling and Descriptive Statistics Introduction The collection and analysis of data are fundamental to science and engineering. Scientists discover the principles that govern the physical world, and engineers learn how to design important new products and processes, by analyzing data collected in scientific experiments. A major difficulty with scientific data is that they are subject to random variation, or uncertainty. That is, when scientific measurements are repeated, they come out somewhat differently each time. This poses a problem: How can one draw conclusions from the results of an experiment when those results could have come out differently? To address this question, a knowledge of statistics is essential. Statistics is the field of study concerned with the collection, analysis, and interpretation of uncertain data. The methods of statistics allow scientists and engineers to design valid experiments and to draw reliable conclusions from the data they produce. Although our emphasis in this book is on the applications of statistics to science and engineering, it is worth mentioning that the analysis and interpretation of data are playing an ever-increasing role in all aspects of modern life. For better or worse, huge amounts of data are collected about our opinions and our lifestyles, for purposes ranging from the creation of more effective marketing campaigns to the development of social policies designed to improve our way of life. On almost any given day, newspaper articles are published that purport to explain social or economic trends through the analysis of data. A basic knowledge of statistics is therefore necessary not only to be an effective scientist or engineer, but also to be a wellinformed member of society. The Basic Idea The basic idea behind all statistical methods of data analysis is to make inferences about a population by studying a relatively small sample chosen from it. As an illustration, Page 2 consider a machine that makes steel rods for use in optical storage devices. The specification for the diameter of the rods is 0.45 ± 0.02 cm. During the last hour, the machine has made 1000 rods. The quality engineer wants to know approximately how many of these rods meet the specification. He does not have time to measure all 1000 rods. So he draws a random sample of 50 rods, measures them, and finds that 46 of them (92%) meet the diameter specification. Now, it is unlikely that the sample of 50 rods represents the population of 1000 perfectly. The proportion of good rods in the population is likely to differ somewhat from the sample proportion of 92%. What the engineer needs to know is just how large that difference is likely to be. For example, is it plausible that the population percentage could be as high as 95%? 98%? As low as 90%? 85%? Here are some specific questions that the engineer might need to answer on the basis of these sample data: 1. The engineer needs to compute a rough estimate of the likely size of the difference between the sample proportion and the population proportion. How large is a typical difference for this kind of sample? 2. The quality engineer needs to note in a logbook the percentage of acceptable rods manufactured in the last hour. Having observed that 92% of the sample rods were good, he will indicate the percentage of acceptable rods in the population as an interval of the form 92% ± x%, where x is a number calculated to provide reasonable certainty that the true population percentage is in the interval. How should x be calculated? 3. The engineer wants to be fairly certain that the percentage of good rods is at least 90%; otherwise he will shut down the process for recalibration. How certain can he be that at least 90% of the 1000 rods are good? Much of this book is devoted to addressing questions like these. The first of these questions requires the computation of a standard deviation, which we will discuss in Chapters 2 and 4. The second question requires the construction of a confidence interval, which we will learn about in Chapter 5. The third calls for a hypothesis test, which we will study in Chapter 6. The remaining chapters in the book cover other important topics. For example, the engineer in our example may want to know how the amount of carbon in the steel rods is related to their tensile strength. Issues like this can be addressed with the methods of correlation and regression, which are covered in Chapters 7 and 8. It may also be important to determine how to adjust the manufacturing process with regard to several factors, in order to produce optimal results. This requires the design of factorial experiments, which are discussed in Chapter 9. Finally, the engineer will need to develop a plan for monitoring the quality of the product manufactured by the process. Chapter 10 covers the topic of statistical quality control, in which statistical methods are used to maintain quality in an industrial setting. The topics listed here concern methods of drawing conclusions from data. These methods form the field of inferential statistics. Before we discuss these topics, we must first learn more about methods of collecting data and of summarizing clearly the basic information they contain. These are the topics of sampling and descriptive statistics, and they are covered in the rest of this chapter. Page 3 1-1 Sampling As mentioned, statistical methods are based on the idea of analyzing a sample drawn from a population. For this idea to work, the sample must be chosen in an appropriate way. For example, let us say that we wished to study the heights of students at the Colorado School of Mines by measuring a sample of 100 students. How should we choose the 100 students to measure? Some methods are obviously bad. For example, choosing the students from the rosters of the football and basketball teams would undoubtedly result in a sample that would fail to represent the height distribution of the population of students. You might think that it would be reasonable to use some conveniently obtained sample, for example, all students living in a certain dorm or all students enrolled in engineering statistics. After all, there is no reason to think that the heights of these students would tend to differ from the heights of students in general. Samples like this are not ideal, however, because they can turn out to be misleading in ways that are not anticipated. The best sampling methods involve random sampling. There are many different random sampling methods, the most basic of which is simple random sampling. To understand the nature of a simple random sample, think of a lottery. Imagine that 10,000 lottery tickets have been sold and that 5 winners are to be chosen. What is the fairest way to choose the winners? The fairest way is to put the 10,000 tickets in a drum, mix them thoroughly, and then reach in and one by one draw 5 tickets out. These 5 winning tickets are a simple random sample from the population of 10,000 lottery tickets. Each ticket is equally likely to be one of the 5 tickets drawn. More importantly, each collection of 5 tickets that can be formed from the 10,000 is equally likely to be the group of 5 that is drawn. It is this idea that forms the basis for the definition of a simple random sample. Summary ■ ■ ■ A population is the entire collection of objects or outcomes about which information is sought. A sample is a subset of a population, containing the objects or outcomes that are actually observed. A simple random sample of size n is a sample chosen by a method in which each collection of n population items is equally likely to make up the sample, just as in a lottery. Since a simple random sample is analogous to a lottery, it can often be drawn by the same method now used in many lotteries: with a computer random number generator. Suppose there are N items in the population. One assigns to each item in the population an integer between 1 and N. Then one generates a list of random integers between 1 and N and chooses the corresponding population items to make up the simple random sample. Page 4 Example 1.1 A physical education professor wants to study the physical fitness levels of students at her university. There are 20,000 students enrolled at the university, and she wants to draw a sample of size 100 to take a physical fitness test. She obtains a list of all 20,000 students, numbered from 1 to 20,000. She uses a computer random number generator to generate 100 random integers between 1 and 20,000 and then invites the 100 students corresponding to those numbers to participate in the study. Is this a simple random sample? Solution Yes, this is a simple random sample. Note that it is analogous to a lottery in which each student has a ticket and 100 tickets are drawn. Example 1.2 A quality engineer wants to inspect rolls of wallpaper in order to obtain information on the rate at which flaws in the printing are occurring. She decides to draw a sample of 50 rolls of wallpaper from a day's production. Each hour for 5 hours, she takes the 10 most recently produced rolls and counts the number of flaws on each. Is this a simple random sample? Solution No. Not every subset of 50 rolls of wallpaper is equally likely to make up the sample. To construct a simple random sample, the engineer would need to assign a number to each roll produced during the day and then generate random numbers to determine which rolls make up the sample. In some cases, it is difficult or impossible to draw a sample in a truly random way. In these cases, the best one can do is to sample items by some convenient method. For example, imagine that a construction engineer has just received a shipment of 1000 concrete blocks, each weighing approximately 50 pounds. The blocks have been delivered in a large pile. The engineer wishes to investigate the crushing strength of the blocks by measuring the strengths in a sample of 10 blocks. To draw a simple random sample would require removing blocks from the center and bottom of the pile, which might be quite difficult. For this reason, the engineer might construct a sample simply by taking 10 blocks off the top of the pile. A sample like this is called a sample of convenience. Definition A sample of convenience is a sample that is obtained in some convenient way, and not drawn by a well-defined random method. The big problem with samples of convenience is that they may differ systematically in some way from the population. For this reason samples of convenience should not be used, except in situations where it is not feasible to draw a random sample. When it is necessary to take a Page 5 sample of convenience, it is important to think carefully about all the ways in which the sample might differ systematically from the population. If it is reasonable to believe that no important systematic difference exists, then it may be acceptable to treat the sample of convenience as if it were a simple random sample. With regard to the concrete blocks, if the engineer is confident that the blocks on the top of the pile do not differ systematically in any important way from the rest, then he may treat the sample of convenience as a simple random sample. If, however, it is possible that blocks in different parts of the pile may have been made from different batches of mix or may have different curing times or temperatures, a sample of convenience could give misleading results. Some people think that a simple random sample is guaranteed to reflect its population perfectly. This is not true. Simple random samples always differ from their populations in some ways, and occasionally may be substantially different. Two different samples from the same population will differ from each other as well. This phenomenon is known as sampling variation. Sampling variation is one of the reasons that scientific experiments produce somewhat different results when repeated, even when the conditions appear to be identical. Example 1.3 A quality inspector draws a simple random sample of 40 bolts from a large shipment and measures the length of each. He finds that 34 of them, or 85%, meet a length specification. He concludes that exactly 85% of the bolts in the shipment meet the specification. The inspector's supervisor concludes that the proportion of good bolts is likely to be close to, but not exactly equal to, 85%. Which conclusion is appropriate? Solution Because of sampling variation, simple random samples don't reflect the population perfectly. They are often fairly close, however. It is therefore appropriate to infer that the proportion of good bolts in the lot is likely to be close to the sample proportion, which is 85%. It is not likely that the population proportion is equal to 85%, however. Example 1.4 Continuing Example 1.3, another inspector repeats the study with a different simple random sample of 40 bolts. She finds that 36 of them, or 90%, are good. The first inspector claims that she must have done something wrong, since his results showed that 85%, not 90%, of bolts are good. Is he right? Solution No, he is not right. This is sampling variation at work. Two different samples from the same population will differ from each other and from the population. Since simple random samples don't reflect their populations perfectly, why is it important that sampling be done at random? The benefit of a simple random sample is that there is no systematic mechanism tending to make the sample unrepresentative. The differences Page 6 between the sample and its population are due entirely to random variation. Since the mathematical theory of random variation is well understood, we can use mathematical models to study the relationship between simple random samples and their populations. For a sample not chosen at random, there is generally no theory available to describe the mechanisms that caused the sample to differ from its population. Therefore, nonrandom samples are often difficult to analyze reliably. In Examples 1.1 to 1.4, the populations consisted of actual physical objects—the students at a university, the concrete blocks in a pile, the bolts in a shipment. Such populations are called tangible populations. Tangible populations are always finite. After an item is sampled, the population size decreases by 1. In principle, one could in some cases return the sampled item to the population, with a chance to sample it again, but this is rarely done in practice. Engineering data are often produced by measurements made in the course of a scientific experiment, rather than by sampling from a tangible population. To take a simple example, imagine that an engineer measures the length of a rod five times, being as careful as possible to take the measurements under identical conditions. No matter how carefully the measurements are made, they will differ somewhat from one another, because of variation in the measurement process that cannot be controlled or predicted. It turns out that it is often appropriate to consider data like these to be a simple random sample from a population. The population, in these cases, consists of all the values that might possibly have been observed. Such a population is called a conceptual population, since it does not consist of actual objects. A simple random sample may consist of values obtained from a process under identical experimental conditions. In this case, the sample comes from a population that consists of all the values that might possibly have been observed. Such a population is called a conceptual population. Example 1.5 involves a conceptual population. Example 1.5 A geologist weighs a rock several times on a sensitive scale. Each time, the scale gives a slightly different reading. Under what conditions can these readings be thought of as a simple random sample? What is the population? Solution If the physical characteristics of the scale remain the same for each weighing, so that the measurements are made under identical conditions, then the readings may be considered to be a simple random sample. The population is conceptual. It consists of all the readings that the scale could in principle produce. Note that in Example 1.5, it is the physical characteristics of the measurement process that determine whether the data are a simple random sample. In general, when deciding Page 7 whether a set of data may be considered to be a simple random sample, it is necessary to have some understanding of the process that generated the data. Statistical methods can sometimes help, especially when the sample is large, but knowledge of the mechanism that produced the data is more important. Example 1.6 A new chemical process has been designed that is supposed to produce a higher yield of a certain chemical than does an old process. To study the yield of this process, we run it 50 times and record the 50 yields. Under what conditions might it be reasonable to treat this as a simple random sample? Describe some conditions under which it might not be appropriate to treat this as a simple random sample. Solution To answer this, we must first specify the population. The population is conceptual and consists of the set of all yields that will result from this process as many times as it will ever be run. What we have done is to sample the first 50 yields of the process. If, and only if, we are confident that the first 50 yields are generated under identical conditions, and that they do not differ in any systematic way from the yields of future runs, then we may treat them as a simple random sample. Be cautious, however. There are many conditions under which the 50 yields could fail to be a simple random sample. For example, with chemical processes, it is sometimes the case that runs with higher yields tend to be followed by runs with lower yields, and vice versa. Sometimes yields tend to increase over time, as process engineers learn from experience how to run the process more efficiently. In these cases, the yields are not being generated under identical conditions and would not be a simple random sample. Example 1.6 shows once again that a good knowledge of the nature of the process under consideration is important in deciding whether data may be considered to be a simple random sample. Statistical methods can sometimes be used to show that a given data set is not a simple random sample. For example, sometimes experimental conditions gradually change over time. A simple but effective method to detect this condition is to plot the observations in the order they were taken. A simple random sample should show no obvious pattern or trend. Figure 1.1 (page 8) presents plots of three samples in the order they were taken. The plot in Figure 1.1a shows an oscillatory pattern. The plot in Figure 1.1b shows an increasing trend. Neither of these samples should be treated as a simple random sample. The plot in Figure 1.1c does not appear to show any obvious pattern or trend. It might be appropriate to treat these data as a simple random sample. However, before making that decision, it is still important to think about the process that produced the data, since there may be concerns that don't show up in the plot (see Example 1.7). FIGURE 1.1 Three plots of observed values versus the order in which they were made. (a) The values show a definite pattern over time. This is not a simple random sample. (b) The values show a trend over time. This is not a simple random sample. (c) The values do not show a pattern or trend. It may be appropriate to treat these data as a simple random sample. Sometimes the question as to whether a data set is a simple random sample depends on the population under consideration. This is one case in which a plot can look good, yet the data are not a simple random sample. Example 1.7 provides an illustration. Page 8 Example 1.7 A new chemical process is run 10 times each morning for five consecutive mornings. A plot of yields in the order they are run does not exhibit any obvious pattern or trend. If the new process is put into production, it will be run 10 hours each day, from 7 A.M. until 5 P.M. Is it reasonable to consider the 50 yields to be a simple random sample? What if the process will always be run in the morning? Solution Since the intention is to run the new process in both the morning and the afternoon, the population consists of all the yields that would ever be observed, including both morning and afternoon runs. The sample is drawn only from that portion of the population that consists of morning runs, and thus it is not a simple random sample. There are many things that could go wrong if this is used as a simple random sample. For example, ambient temperatures may differ between morning and afternoon, which could affect yields. If the process will be run only in the morning, then the population consists only of morning runs. Since the sample does not exhibit any obvious pattern or trend, it might well be appropriate to consider it to be a simple random sample. Independence The items in a sample are said to be independent if knowing the values of some of them does not help to predict the values of the others. With a finite, tangible population, the items in a simple random sample are not strictly independent, because as each item is drawn, the population changes. This change can be substantial when the population is small. However, when the population is very large, this change is negligible and the items can be treated as if they were independent. Page 9 To illustrate this idea, imagine that we draw a simple random sample of 2 items from the population For the first draw, the numbers 0 and 1 are equally likely. But the value of the second item is clearly influenced by the first; if the first is 0, the second is more likely to be 1, and vice versa. Thus the sampled items are dependent. Now assume we draw a sample of size 2 from this population: Again on the first draw, the numbers 0 and 1 are equally likely. But unlike the previous example, these two values remain almost equally likely on the second draw as well, no matter what happens on the first draw. With the large population, the sample items are for all practical purposes independent. It is reasonable to wonder how large a population must be in order that the items in a simple random sample may be treated as independent. A rule of thumb is that when sampling from a finite population, the items may be treated as independent so long as the sample contains 5% or less of the population. Interestingly, it is possible to make a population behave as though it were infinitely large, by replacing each item after it is sampled. This method is called sampling with replacement. With this method, the population is exactly the same on every draw and the sampled items are truly independent. With a conceptual population, we require that the sample items be produced under identical experimental conditions. In particular, then, no sample value may influence the conditions under which the others are produced. Therefore, the items in a simple random sample from a conceptual population may be treated as independent. We may think of a conceptual population as being infinite, or equivalently, that the items are sampled with replacement. Summary ■ ■ The items in a sample are independent if knowing the values of some of the items does not help to predict the values of the others. Items in a simple random sample may be treated as independent in many cases encountered in practice. The exception occurs when the population is finite and the sample consists of a substantial fraction (more than 5%) of the population. Other Sampling Methods Page 10 In addition to simple random sampling there are other sampling methods that are useful in various situations. In weighted sampling, some items are given a greater chance of being selected than others, like a lottery in which some people have more tickets than others. In stratified random sampling, the population is divided up into subpopulations, called strata, and a simple random sample is drawn from each stratum. In cluster sampling, items are drawn from the population in groups, or clusters. Cluster sampling is useful when the population is too large and spread out for simple random sampling to be feasible. For example, many U.S. government agencies use cluster sampling to sample the U.S. population to measure sociological factors such as income and unemployment. A good source of information on sampling methods is Scheaffer, et al., (2012). Simple random sampling is not the only valid method of random sampling. But it is the most fundamental, and we will focus most of our attention on this method. From now on, unless otherwise stated, the terms "sample" and "random sample" will be taken to mean "simple random sample." Types of Experiments There are many types of experiments that can be used to generate data. We briefly describe a few of them. In a one-sample experiment, there is only one population of interest, and a single sample is drawn from it. For example, imagine that a process is being designed to produce polyethylene that will be used to line pipes. An experiment in which several specimens of polyethylene are produced by this process, and the tensile strength of each is measured, is a onesample experiment. The measured strengths are considered to be a simple random sample from a conceptual population of all the possible strengths that can be observed for specimens manufactured by this process. One-sample experiments can be used to determine whether a process meets a certain standard, for example, whether it provides sufficient strength for a given application. In a multisample experiment, there are two or more populations of interest, and a sample is drawn from each population. For example, if several competing processes are being considered for the manufacture of polyethylene, and tensile strengths are measured on a sample of specimens from each process, this is a multisample experiment. Each process corresponds to a separate population, and the measurements made on the specimens from a particular process are considered to be a simple random sample from that population. The usual purpose of multisample experiments is to make comparisons among populations. In this example, the purpose might be to determine which process produced the greatest strength or to determine whether there is any difference in the strengths of polyethylene made by the different processes. In many multisample experiments, the populations are distinguished from one another by the varying of one or more factors that may affect the outcome. Such experiments are called factorial experiments. For example, in his M.S. thesis at the Colorado School of Mines, G. Fredrickson measured the Charpy V-notch impact toughness for a large number of welds. Each weld was made with one of two types of base metals and had its toughness measured at one of several temperatures. This was a factorial experiment with two factors: base metal and temperature. The data consisted of several toughness measurements made at each Page 11 combination of base metal and temperature. In a factorial experiment, each combination of the factors for which data are collected defines a population, and a simple random sample is drawn from each population. The purpose of a factorial experiment is to determine how varying the levels of the factors affects the outcome being measured. In his experiment Fredrickson found that for each type of base metal, the toughness remained unaffected by temperature unless the temperature was very low—below −100°C. As the temperature was decreased from −100°C to −200°C, the toughness dropped steadily. Types of Data When a numerical quantity designating how much or how many is assigned to each item in a sample, the resulting set of values is called numerical or quantitative. In some cases, sample items are placed into categories, and category names are assigned to the sample items. Then the data are categorical or qualitative. Example 1.8 provides an illustration. Example 1.8 The article "Wind-Uplift Capacity of Residential Wood Roof-Sheathing Panels Retrofitted with Insulating Foam Adhesive" (P. Datin, D. Prevatt, and W. Pang, Journal of Architectural Engineering, 2011:144–154) presents tests in which air pressure was applied to roof-sheathing panels until failure. The pressure at failure for each panel was recorded, along with the type of sheathing, sheathing thickness, and wood species. The following table presents results of four tests. Sheathing Type Failure Pressure (kPa) 5-ply plywood 2.63 Oriented Strand 1.69 Board Cox plywood 5.26 4-ply plywood 5.03 Thickness (mm) Wood Species 11.9 Douglas Fir 15.1 Spruce-Pine-Fir 12.7 Southern Yellow Pine 15.9 Douglas Fir Which data are numerical and which are categorical? Solution The failure pressure and thickness are numerical. The sheathing type and wood species are categorical. Controlled Experiments and Observational Studies Many scientific experiments are designed to determine the effect of changing one or more factors on the value of a response. For example, suppose that a chemical engineer wants to determine how the concentrations of reagent and catalyst affect the yield of a process. The engineer can run the process several times, changing the concentrations each time, and compare the yields that result. This sort of experiment is called a controlled experiment, because the values of Page 12 the factors, in this case the concentrations of reagent and catalyst, are under the control of the experimenter. When designed and conducted properly, controlled experiments can produce reliable information about cause-and-effect relationships between factors and response. In the yield example just mentioned, a well-done experiment would allow the experimenter to conclude that the differences in yield were caused by differences in the concentrations of reagent and catalyst. There are many situations in which scientists cannot control the levels of the factors. For example, there have been many studies conducted to determine the effect of cigarette smoking on the risk of lung cancer. In these studies, rates of cancer among smokers are compared with rates among non-smokers. The experimenters cannot control who smokes and who doesn't; people cannot be required to smoke just to make a statistician's job easier. This kind of study is called an observational study, because the experimenter simply observes the levels of the factor as they are, without having any control over them. Observational studies are not nearly as good as controlled experiments for obtaining reliable conclusions regarding cause and effect. In the case of smoking and lung cancer, for example, people who choose to smoke may not be representative of the population as a whole, and may be more likely to get cancer for other reasons. For this reason, although it has been known for a long time that smokers have higher rates of lung cancer than non-smokers, it took many years of carefully done observational studies before scientists could be sure that smoking was actually the cause of the higher rate. Exercises for Section 1.1 1. 2. 3. Each of the following processes involves sampling from a population. Define the population, and state whether it is tangible or conceptual. a. A chemical process is run 15 times, and the yield is measured each time. b. A pollster samples 1000 registered voters in a certain state and asks them which candidate they support for governor. c. In a clinical trial to test a new drug that is designed to lower cholesterol, 100 people with high cholesterol levels are recruited to try the new drug. d. Eight concrete specimens are constructed from a new formulation, and the compressive strength of each is measured. e. A quality engineer needs to estimate the percentage of bolts manufactured on a certain day that meet a strength specification. At 3:00 in the afternoon he samples the last 100 bolts to be manufactured. If you wanted to estimate the mean height of all the students at a university, which one of the following sampling strategies would be best? Why? Note that none of the methods are true simple random samples. i. Measure the heights of 50 students found in the gym during basketball intramurals. ii. Measure the heights of all engineering majors. iii. Measure the heights of the students selected by choosing the first name on each page of the campus phone book. True or false: A simple random sample is guaranteed to reflect exactly the population from which it was drawn. a. A simple random sample is free from any systematic tendency to differ from the population from which it was drawn. b. A simple random sample of 100 college students is selected from all students registered at a certain college, and it turns out that 38 of them participate in intramural sports. True or false: a. The proportion of students at this college who participate in intramural sports is likely to be close to 0.38, but not equal to 0.38. a. 4. Page 13 5. 6. 7. 8. 9. A certain process for manufacturing integrated circuits has been in use for a period of time, and it is known that 12% of the circuits it produces are defective. A new process that is supposed to reduce the proportion of defectives is being tested. In a simple random sample of 100 circuits produced by the new process, 12 were defective. a. One of the engineers suggests that the test proves that the new process is no better than the old process, since the proportion of defectives in the sample is the same. Is this conclusion justified? Explain. b. Assume that there had been only 11 defective circuits in the sample of 100. Would this have proven that the new process is better? Explain. c. Which outcome represents stronger evidence that the new process is better: finding 11 defective circuits in the sample, or finding 2 defective circuits in the sample? Refer to Exercise 5. True or false: a. If the proportion of defectives in the sample is less than 12%, it is reasonable to conclude that the new process is better. b. If the proportion of defectives in the sample is only slightly less than 12%, the difference could well be due entirely to sampling variation, and it is not reasonable to conclude that the new process is better. c. If the proportion of defectives in the sample is a lot less than 12%, it is very unlikely that the difference is due entirely to sampling variation, so it is reasonable to conclude that the new process is better. To determine whether a sample should be treated as a simple random sample, which is more important: a good knowledge of statistics, or a good knowledge of the process that produced the data? A medical researcher wants to determine whether exercising can lower blood pressure. At a health fair, he measures the blood pressure of 100 individuals, and interviews them about their exercise habits. He divides the individuals into two categories: those whose typical level of exercise is low, and those whose level of exercise is high. a. Is this a controlled experiment or an observational study? b. The subjects in the low exercise group had considerably higher blood pressure, on the average, than subjects in the high exercise group. The researcher concludes that exercise decreases blood pressure. Is this conclusion well-justified? Explain. A medical researcher wants to determine whether exercising can lower blood pressure. She recruits 100 people with high blood pressure to participate in the study. She assigns a random sample of 50 of them to pursue an exercise program that includes daily swimming and jogging. She assigns the other 50 to refrain from vigorous activity. She measures the blood pressure of each of the 100 individuals both before and after the study. a. Is this a controlled experiment or an observational study? b. On the average, the subjects in the exercise group substantially reduced their blood pressure, while the subjects in the no-exercise group did not experience a reduction. The researcher concludes that exercise decreases blood pressure. Is this conclusion better justified than the conclusion in Exercise 8? Explain. 1-2 Summary Statistics A sample is often a long list of numbers. To help make the important features of a sample stand out, we compute summary statistics. The two most commonly used summary statistics are the sample mean and the sample standard deviation. The mean gives an indication of the center of the data, and the standard deviation gives an indication of how spread out the data are. Page 14 The Sample Mean The sample mean is also called the "arithmetic mean," or, more simply, the "average." It is the sum of the numbers in the sample, divided by how many there are. Definition Let X1, …, Xn be a sample. The sample mean is (1.1) Note that it is customary to use a letter with a bar over it (e.g.,) to denote a sample mean. Note also that the sample mean has the same units as the sample values X1,…, Xn. Example 1.9 A simple random sample of five men is chosen from a large population of men, and their heights are measured. The five heights (in inches) are 65.51, 72.30, 68.31, 67.05, and 70.68. Find the sample mean. Solution We use Equation (1.1). The sample mean is The Standard Deviation Here are two lists of numbers: 28, 29, 30, 31, 32 and 10, 20, 30, 40, 50. Both lists have the same mean of 30. But clearly the lists differ in an important way that is not captured by the mean: the second list is much more spread out than the first. The standard deviation is a quantity that measures the degree of spread in a sample. Let X1, …, Xn be a sample. The basic idea behind the standard deviation is that when the spread is large, the sample values will tend to be far from their mean, but when the spread is small, the values will tend to be close to their mean. So the first step in calculating the standard deviation is to compute the differences (also called deviations) between each sample value and the sample mean. The deviations are , …, . Now some of these deviations are positive and some are negative. Large negative deviations are just as indicative of spread as large positive deviations are. To make all the deviations positive we square them, obtaining the squared deviations . From the squared deviations we can compute a measure of spread called the sample variance. The sample variance is the average of the squared deviations, except that we divide by n − 1 instead of n. It is customary to denote the Page 15 sample variance by s2. Definition Let X1,…, Xn be a sample. The sample variance is the quantity (1.2) An equivalent formula, which can be easier to compute, is (1.3) While the sample variance is an important quantity, it has a serious drawback as a measure of spread. Its units are not the same as the units of the sample values; instead they are the squared units. To obtain a measure of spread whose units are the same as those of the sample values, we simply take the square root of the variance. This quantity is known as the sample standard deviation. It is customary to denote the sample standard deviation by s (the square root of s2). Definition Let X1, …, Xn be a sample. The sample standard deviation is the quantity (1.4) An equivalent formula, which can be easier to compute, is (1.5) The sample standard deviation is the square root of the sample variance. It is natural to wonder why the sum of the squared deviations is divided by n − 1 rather than n. The purpose in computing the sample standard deviation is to estimate the amount of spread in the population from which the sample was drawn. Ideally, therefore, we would compute deviations from the mean of all the items in the population, rather than the deviations from the sample mean. However, the population mean is in general unknown, so the sample mean is used in its place. It is a mathematical fact that the deviations around the sample mean tend to be a bit smaller than the deviations around the population mean and that dividing by n − 1 rather than n provides exactly the right correction. Example 1.10 Find the sample variance and the sample standard deviation for the height data in Example 1.9. Solution We'll first compute the sample variance by using Equation (1.2). The sample mean is (see Example 1.9). The sample variance is therefore Alternatively, we can use Equation (1.3): The sample standard deviation is the square root of the sample variance: What would happen to the sample mean, variance, and standard deviation if the heights in Example 1.9 were measured in centimeters rather than inches? Let's denote the heights in inches by X1, X2, X3, X4, X5, and the heights in centimeters by Y1, Y2, Y3, Y4, Y5. The relationship between Xi and Yi is then given by Yi = 2.54Xi. If you go back to Example 1.9, convert to centimeters, and compute the sample mean, you will find that the sample means in centimeters and in inches are related by the equation . Thus if we multiply each sample item by a constant, the sample mean is multiplied by the same constant. As for the sample variance, you will find that the deviations are related by the equation . It follows that , and that sY = 2.54sX. What if each man in the sample put on 2-inch heels? Then each sample height would increase by 2 inches and the sample mean would increase by 2 inches as well. In general, if a constant is added to each sample item, the sample mean increases (or decreases) by the same constant. The deviations, however, do not change, so the sample variance and standard deviation are unaffected. Summary ■ If X1, …, Xn is a sample and Yi = a + bXi, where a and b are constants, then . ■ If X1, …, Xn is a sample and Yi = a + bXi, where a and b are constants, then and sY = |b|sX. , The Sample Median Page 17 The median, like the mean, is a measure of center. To compute the median of a sample, order the values from smallest to largest. The sample median is the middle number. If the sample size is an even number, it is customary to take the sample median to be the average of the two middle numbers. Definition If n numbers are ordered from smallest to largest: ■ If n is odd, the sample median is the number in position ■ If n is even, the sample median is the average of the numbers in positions and . 1.11 Find the sample median for the height data in Example 1.9. Solution The five heights, arranged in increasing order, are 65.51, 67.05, 68.31, 70.68, 72.30. The sample median is the middle number, which is 68.31. The Trimmed Mean Like the median, the trimmed mean is a measure of center that is designed to be unaffected by outliers. The trimmed mean is computed by arranging the sample values in order, "trimming" an equal number of them from each end, and computing the mean of those remaining. If p% of the data are trimmed from each end, the resulting trimmed mean is called the "p% trimmed mean." There are no hard-and-fast rules on how many values to trim. The most commonly used trimmed means are the 5%, 10%, and 20% trimmed means. Note that the median can be thought of as an extreme form of trimmed mean, obtained by trimming away all but the middle one or two sample values. Since the number of data points trimmed must be a whole number, it is impossible in many cases to trim the exact percentage of data that is called for. If the sample size is denoted by n, and a p% trimmed mean is desired, the number of data points to be trimmed is np/100. If this is not a whole number, the simplest thing to do when computing by hand is to round it to the nearest whole number and trim that amount. Example 1.12 In the article "Evaluation of Low-Temperature properties of HMA Mixtures" (P. Sebaaly, A. Lake, and J. Epps, Journal of Transportation Engineering, 2002: 578–583), the following values of fracture stress (in megapascals) were measured for a sample of 24 mixtures of hot-mixed asphalt (HMA). 30 75 79 80 105 126 138 149 179 179 191 223 232 232 236 240 242 245 247 254 274 384 470 Compute the mean, median, and the 5%, 10%, and 20% trimmed means. Page 18 Solution The mean is found by averaging together all 24 numbers, which produces a value of 195.42. The median is the average of the 12th and 13th numbers, which is (191 + 223)/2 = 207.00. To compute the 5% trimmed mean, we must drop 5% of the data from each end. This comes to (0.05)(24) = 1.2 observations. We round 1.2 to 1, and trim one observation off each end. The 5% trimmed mean is the average of the remaining 22 numbers: To compute the 10% trimmed mean, round off (0.1)(24) = 2.4 to 2. Drop 2 observations from each end, and then average the remaining 20: To compute the 20% trimmed mean, round off (0.2)(24) = 4.8 to 5. Drop 5 observations from each end, and then average the remaining 14: Outliers Sometimes a sample may contain a few points that are much larger or smaller than the rest. Such points are called outliers. See Figure 1.2 for an example. Sometimes outliers result from data entry errors; for example, a misplaced decimal point can result in a value that is an order of magnitude different from the rest. Outliers should always be scrutinized, and any outlier that is found to result from an error should be corrected or deleted. Not all outliers are errors. Sometimes a population may contain a few values that are much different from the rest, and the outliers in the sample reflect this fact. FIGURE 1.2 A data set that contains an outlier. Outliers are a real problem for data analysts. For this reason, when people see outliers in their data, they sometimes try to find a reason, or an excuse, to delete them. An outlier should not be deleted, however, unless there is reasonable certainty that it results from an error. If a population truly contains outliers, but they are deleted from the sample, the sample will not characterize the population correctly. Resistance to Outliers A statistic whose value does not change much when an outlier is added to or removed from a sample is said to be resistant. The median and trimmed mean are both resistant, but the mean and standard deviation are not. Page 19 We illustrate the fact with a simple example. Annual salaries for a sample of six engineers, in $1000s, are 51, 58, 65, 75, 84, and 93. The mean is 71, the standard deviation is 15.96, the median is 70, and the 10% trimmed mean is 70.5. Now we add the salary of the CEO, which is $300,000, to the list. The list is now 51, 58, 65, 75, 84, 93, and 300. Now the mean is 124.71, the standard deviation is 87.77, the median is 75, and the 10% trimmed mean is 75. Clearly the mean and standard deviation have changed considerably, while the median and trimmed mean have changed much less. Because it is resistant, the median is often used as a measure of center for samples that contain outliers. To see why, Figure 1.3 presents a plot of the salary data we have just considered. It is reasonable to think that the median is more representative of the sample than the mean is. FIGURE 1.3 When a sample contains outliers, the median may be more representative of the sample than the mean is. The Mode and the Range The mode and the range are summary statistics that are of limited use but are occasionally seen. The sample mode is the most frequently occurring value in a sample. If several values occur with equal frequency, each one is a mode. The range is the difference between the largest and smallest values in a sample. It is a measure of spread, but it is rarely used, because it depends only on the two extreme values and provides no information about the rest of the sample. Example 1.13 Find the modes and the range for the sample in Example 1.12. Solution There are three modes: 80, 179, and 232. Each of these values appears twice, and no other value appears more than once. The range is 470 − 30 = 440. Quartiles The median divides the sample in half. Quartiles divide it as nearly as possible into quarters. A sample has three quartiles. There are several different ways to compute quartiles, but all of them give approximately the same result. The simplest method when computing by hand is as follows: Let n represent the sample size. Order the sample values from smallest to largest. To find the first quartile, compute the value 0.25(n + 1). If this is an integer, then the sample value in that position is the first quartile. If not, then take the average of the sample values on either side of this value. The third quartile is computed in the same way, except that the value 0.75(n +1) is used. The second quartile uses the value 0.5(n + 1). The second quartile is identical to the median. We note that some computer packages use slightly different methods to Page 20 compute quartiles, so their results may not be quite the same as the ones obtained by the method described here. Example 1.14 Find the first and third quartiles of the asphalt data in Example 1.12. Solution The sample size is n = 24. To find the first quartile, compute (0.25)(25) = 6.25. The first quartile is therefore found by averaging the 6th and 7th data points, when the sample is arranged in increasing order. This yields (105 + 126)/2 = 115.5. To find the third quartile, compute (0.75)(25) = 18.75. We average the 18th and 19th data points to obtain (242 + 245)/2 = 243.5. Percentiles The pth percentile of a sample, for a number p between 0 and 100, divides the sample so that as nearly as possible p% of the sample values are less than the pth percentile, and (100 − p)% are greater. There are many ways to compute percentiles; they all produce similar results. We describe here a method analogous to the method described for computing quartiles. Order the sample values from smallest to largest, and then compute the quantity (p/100)(n + 1), where n is the sample size. If this quantity is an integer, the sample value in this position is the pth percentile. Otherwise average the two sample values on either side. Note that the first quartile is the 25th percentile, the median is the 50th percentile, and the third quartile is the 75th percentile. Some computer packages use slightly different methods to compute percentiles, so their results may differ slightly from the ones obtained by this method. Percentiles are often used to interpret scores on standardized tests. For example, if a student is informed that her score on a college entrance exam is on the 64th percentile, this means that 64% of the students who took the exam got lower scores. Example 1.15 Find the 65th percentile of the asphalt data in Example 1.12. Solution The sample size is n = 24. To find the 65th percentile, compute (0.65)(25) = 16.25. The 65th percentile is therefore found by averaging the 16th and 17th data points, when the sample is arranged in increasing order. This yields (236 + 240)/2 = 238. In practice, the summary statistics we have discussed are often calculated on a computer, using a statistical software package. The summary statistics are sometimes called descriptive statistics because they describe the data. We present an example of the calculation of summary statistics from the software package MINITAB. Then we will show how these statistics can be used to discover some important features of the data. For a Ph.D. thesis that investigated factors affecting diesel vehicle emissions, J. Yanowitz of the Colorado School of Mines obtained data on emissions of particulate matter (PM) for a sample of 138 vehicles driven at low altitude (near sea level) and for a sample of 62 vehicles driven at high altitude (approximately one mile above sea level). All the vehicles were Page 21 manufactured between 1991 and 1996. The samples contained roughly equal proportions of high- and low-mileage vehicles. The data, in units of grams of particulates per gallon of fuel consumed, are presented in Tables 1.1 and 1.2. At high altitude, the barometric pressure is lower, so the effective air/fuel ratio is lower as well. For this reason it was thought that PM emissions might be greater at higher altitude. We would like to compare the samples to determine whether the data support this assumption. It is difficult to do this simply by examining the raw data in the tables. Computing summary statistics makes the job much easier. Figure 1.4 presents summary statistics for both samples, as computed by MINITAB. TABLE 1.1 Particulate matter (PM) emissions (in g/gal) for 138 vehicles driven at low altitude 1.50 1.48 2.98 1.40 3.12 0.25 6.73 5.30 9.30 6.96 7.21 0.87 1.06 7.39 1.37 2.37 0.53 7.82 3.93 6.50 1.67 1.12 1.11 2.66 1.81 2.12 3.36 4.59 3.52 4.62 9.14 4.49 TABLE 1.2 1.25 2.15 11.00 11.44 3.60 4.57 1.63 1.74 2.67 5.27 3.12 4.93 8.23 2.80 1.11 8.81 4.38 3.67 3.34 1.87 4.07 0.87 3.63 1.11 2.24 4.24 2.74 5.66 3.24 7.74 3.62 6.70 0.94 2.14 0.61 3.03 6.55 3.59 5.79 5.41 2.67 5.93 0.64 6.64 1.46 1.23 1.18 3.10 3.59 3.40 2.75 4.51 1.31 4.04 0.97 1.04 3.06 3.33 3.48 4.97 8.92 9.04 2.49 2.48 0.90 1.63 4.08 2.56 2.96 11.23 9.93 7.71 Particulate matter (PM) emissions (in g/gal) for 62 vehicles driven at high altitude 7.59 2.06 8.86 8.67 5.61 6.28 6.07 5.23 5.54 3.46 2.44 3.01 13.63 13.02 23.38 9.24 3.22 4.04 17.11 12.26 19.91 8.50 7.81 7.18 6.95 18.64 7.10 6.04 5.66 4.40 3.57 4.35 3.84 2.37 3.81 5.32 5.84 2.89 4.68 1.85 9.14 8.67 9.52 2.68 10.14 9.20 7.31 2.09 6.32 6.53 6.32 2.01 5.91 5.60 5.61 1.50 6.46 5.29 5.64 2.07 1.11 3.32 1.83 7.56 FIGURE 1.4 MINITAB output presenting descriptive statistics for the PM data in Tables 1.1 and 1.2. In Figure 1.4 (page 22), the quantity labeled "N" is the sample size. Following that is the sample mean. The next quantity (SE Mean) is the standard error of the mean. The standard error of the mean is equal to the standard deviation divided by the square root of the sample size. This is a quantity that is not used much as a descriptive statistic, although it is important for applications such as constructing confidence intervals and hypothesis tests, which we will cover in Chapters 5 and 6. Following the standard error of the mean is the 5% trimmed mean (TrMean), and the standard deviation (StDev). Finally, the second line of the output provides the minimum, median, and maximum, as well as the first and third quartiles (Q1 and Q3). We note that the values of the quartiles produced by the computer package differ slightly from the values that would be computed by the methods we describe. This is not surprising, since there are several ways to compute these values. The differences are not large enough to have any practical importance. The summary statistics tell a lot about the differences in PM emissions between high- and low-altitude vehicles. First, note that the mean is indeed larger for the high-altitude vehicles than for the low-altitude vehicles (6.596 vs. 3.715), which supports the hypothesis that Page 22 emissions tend to be greater at high altitudes. Now note that the maximum value for the high-altitude vehicles (23.38) is much higher than the maximum for the low-altitude vehicles (11.23). This shows that there are one or more high-altitude vehicles whose emissions are much higher than the highest of the low-altitude vehicles. Could the difference in mean emissions be due entirely to these vehicles? To answer this, compare the medians, the first and third quartiles, and the trimmed means. These statistics are not affected much by a few large values, yet all of them are noticeably larger for the high-altitude vehicles. Therefore, we can conclude that the high-altitude vehicles not only contain a few very high emitters, they also have higher emissions than the low-altitude vehicles in general. Finally note that the standard deviation is larger for the high-altitude vehicles, which indicates that the values for the high-altitude vehicles are more spread out than those for the low-altitude vehicles. At least some of this difference in spread must be due to the one or more high-altitude vehicles with very high emissions. Summary Statistics for Categorical Data With categorical data, each sample item is assigned a category rather than a quantitative value. But to work with categorical data, numerical summaries are needed. The two most commonly used ones are the frequencies and the sample proportions (sometimes called relative frequencies). The frequency for a given category is simply the number of sample items that fall into that category. The sample proportion is the frequency divided by the sample size. Example 1.16 A process manufactures crankshaft journal bearings for an internal combustion engine. Bearings whose thicknesses are between 1.486 and 1.490 mm are classified as conforming, which means that they meet the specification. Bearings thicker than this are reground, and bearings thinner than this are scrapped. In a sample of 1000 bearings, 910 were conforming, 53 were reground, and 37 were scrapped. Find the frequencies and sample proportions. Solution The frequencies are 910, 53, and 37. The sample proportions are 910/1000 = 0.910, 53/1000 = 0.053, and 37/1000 = 0.037. Page 23 Sample Statistics and Population Parameters Each of the sample statistics we have discussed has a population counterpart. This is easy to see when the population is finite. For example, for a finite population of numerical values, the population mean is simply the average of all the values in the population; the population median is the middle value, or average of the two middle values; and so on. In fact, any numerical summary used for a sample can be used for a finite population, just by applying the methods of calculation to the population values rather than the sample values. One small exception occurs for the population variance, where we divide by n rather than n − 1. There is a difference in terminology for numerical summaries of populations as opposed to samples. Numerical summaries of a sample are called statistics, while numerical summaries of a population are called parameters. Of course, in practice, the entire population is never observed, so the population parameters cannot be calculated directly. Instead, the sample statistics are used to estimate the values of the population parameters. The methods for computing sample statistics require that the sample be finite. Therefore, when a population contains an infinite number of values, the methods for computing sample statistics cannot be applied to compute population parameters. For infinite populations, parameters such as the mean and variance are computed by procedures that generalize the methods used to compute sample statistics, and which involve infinite sums or integrals. We will describe these procedures in Chapter 2. Summary ■ ■ ■ A numerical summary of a sample is called a statistic. A numerical summary of a population is called a parameter. Statistics are often used to estimate parameters. Exercises for Section 1.2 1. True or false: For any list of numbers, half of them will be below the mean. 2. Is the sample mean always the most frequently occurring value? If so, explain why. If not, give an example. 3. Is the sample mean always equal to one of the values in the sample? If so, explain why. If not, give an example. 4. Is the sample median always equal to one of the values in the sample? If so, explain why. If not, give an example. 5. Find a sample size for which the median will always equal one of the values in the sample. 6. For a list of positive numbers, is it possible for the standard deviation to be greater than the mean? If so, give an example. If not, explain why not. 7. Is it possible for the standard deviation of a list of numbers to equal 0? If so, give an example. If not, explain why not. 8. In a certain company, every worker received a $50-per-week raise. How does this affect the mean salary? The standard deviation of the salaries? 9. In another company, every worker received a 5% raise. How does this affect the mean salary? The standard deviation of the salaries? Page 24 10. A sample of 100 cars driving on a freeway during a morning commute was drawn, and the number of occupants in each car was recorded. The results were as follows: 1. 2. b. c. d. e. f. Find the sample mean number of occupants. Find the sample standard deviation of the number of occupants. Find the sample median number of occupants. Compute the first and third quartiles of the number of occupants. What proportion of cars had more than the mean number of occupants? For what proportion of cars was the number of occupants more than one standard deviation greater than the mean? g. For what proportion of cars was the number of occupants within one standard deviation of the mean? 11. In a sample of 20 men, the mean height was 178 cm. In a sample of 30 women, the mean height was 164 cm. What was the mean height for both groups put together? 12. Each of 16 students measured the circumference of a tennis ball by four different methods, which were: Method A: Estimate the circumference by eye. Method B: Measure the diameter with a ruler, and then compute the circumference. Method C: Measure the circumference with a ruler and string. Method D: Measure the circumference by rolling the ball along a ruler. The results (in cm) are as follows, in increasing order for each method: Method A: 18.0, 18.0, 18.0, 20.0, 22.0, 22.0, 22.5, 23.0, 24.0, 24.0, 25.0, 25.0, 25.0, 25.0, 26.0, 26.4. Method B: 18.8, 18.9, 18.9, 19.6, 20.1, 20.4, 20.4, 20.4, 20.4, 20.5, 21.2, 22.0, 22.0, 22.0, 22.0, 23.6. Method C: 20.2, 20.5, 20.5, 20.7, 20.8, 20.9, 21.0, 21.0, 21.0, 21.0, 21.5, 21.5, 21.5, 21.5, 21.6. Method D: 20.0, 20.0, 20.0, 20.0, 20.2, 20.5, 20.5, 20.7, 20.7, 20.7, 21.0, 21.1, 21.5, 21.6, 22.1, 22.3. a. Compute the mean measurement for each method. b. Compute the median measurement for each method. c. Compute the 20% trimmed mean measurement for each method. d. Compute the first and third quartiles for each method. e. Compute the standard deviation of the measurements for each method. f. For which method is the standard deviation the largest? Why should one expect this method to have the largest standard deviation? g. Other things being equal, is it better for a measurement method to have a smaller standard deviation or a larger standard deviation? Or doesn't it matter? Explain. 13. Refer to Exercise 12. a. If the measurements for one of the methods were converted to inches (1 inch = 2.54 cm), how would this affect the mean? The median? The quartiles? The standard deviation? b. If the students remeasured the ball, using a ruler marked in inches, would the effects on the mean, median, quartiles, and standard deviation be the same as in part (a)? Explain. 14. There are 10 employees in a particular division of a company. Their salaries have a mean of $70,000, a median of $55,000, and a standard deviation of $20,000. The largest number on the list is $100,000. By accident, this number is changed to $1,000,000. a. What is the value of the mean after the change? b. What is the value of the median after the change? c. What is the value of the standard deviation after the change? 15. Quartiles divide a sample into four nearly equal pieces. In general, a sample of size n can be broken into k nearly equal pieces by using the cutpoints (i/k)(n + 1) for i = 1, ..., k − 1. Consider the following ordered sample: 2 18 23 41 44 46 49 61 62 74 76 79 82 89 92 95 a. Tertiles divide a sample into thirds. Find the tertiles of this sample. b. Quintiles divide a sample into fifths. Find the quintiles of this sample. Page 25 16. In each of the following data sets, tell whether the outlier seems certain to be due to an error, or whether it could conceivably be correct. a. The length of a rod is measured five times. The readings in centimeters are 48.5, 47.2, 4.91, 49.5, 46.3. b. The prices of five cars on a dealer's lot are $25,000, $30,000, $42,000, $110,000, $31,000. 1.3 Graphical Summaries Stem-and-Leaf Plots The mean, median, and standard deviation are numerical summaries of a sample or of a population. Graphical summaries are used as well to help visualize a list of numbers. The graphical summary that we will discuss first is the stem-and-leaf plot. A stem-and-leaf plot is a simple way to summarize a data set. As an example, the data in Table 1.3 concern the geyser Old Faithful in Yellowstone National Park. This geyser alternates periods of eruption, which typically last from 1.5 to 4 minutes, with periods of dormancy, which are considerably longer. Table 1.3 presents the durations, in minutes, of 60 dormant periods. The list has been sorted into numerical order. TABLE 1.3 42 55 68 75 80 84 45 55 69 75 80 84 Durations (in minutes) of dormant periods of the geyser Old Faithful 49 56 70 76 80 85 50 56 71 76 81 86 51 57 72 76 82 86 51 58 73 76 82 86 51 60 74 76 83 86 51 60 75 76 83 86 53 60 75 76 83 86 55 60 75 79 83 91 53 66 75 79 84 93 Figure 1.5 presents a stem-and-leaf plot of the geyser data. Each item in the sample is divided into two parts: a stem, consisting of the leftmost one or two digits, and the leaf, which consists of the next digit. In Figure 1.5, the stem consists of the tens digit and the leaf consists of the ones digit. Each line of the stem-and-leaf plot contains all of the sample items with a given stem. The stem-and-leaf plot is a compact way to represent the data. It also gives some indication of its shape. For the geyser data, we can see that there are relatively few durations in the 60–69 minute interval, compared with the 50–59, 70–79, or 80–89 minute intervals. FIGURE 1.5 Stem-and-leaf plot for the geyser data in Table 1.3. When there are a great many sample items with the same stem, it is often necessary to assign more than one row to that stem. As an example, Figure 1.6 presents a Page 26 computergenerated stem-and-leaf plot, produced by MINITAB, for the PM data in Table 1.2 in Section 1.2. The middle column, consisting of 0s, 1s, and 2s, contains the stems, which are the tens digits. To the right of the stems are the leaves, consisting of the ones digits for each of the sample items. Note that the digits to the right of the decimal point have been truncated, so that the leaf will have only one digit. Since many numbers are less than 10, the 0 stem must be assigned several lines, five in this case. Specifically, the first line contains the sample items whose ones digits are either 0 or 1, the next line contains the items whose ones digits are either 2 or 3, and so on. For consistency, all the stems are assigned several lines in the same way, even though there are few enough values for the 1 and 2 stems that they could have fit on fewer lines. FIGURE 1.6 Stem-and-leaf plot of the PM data in Table 1.2 in Section 1.2 as produced by MINITAB. The output in Figure 1.6 contains a cumulative frequency column to the left of the stem-and-leaf plot. The upper part of this column provides a count of the number of items at or above the current line, and the lower part of the column provides a count of the number of items at or below the current line. Next to the line that contains the median is the count of items in that line, shown in parentheses. A good feature of stem-and-leaf plots is that they display all the sample values. One can reconstruct the sample in its entirety from a stem-and-leaf plot, although some digits may be truncated. In addition, the order in which the items were sampled cannot be determined. Dotplots A dotplot is a graph that can be used to give a rough impression of the shape of a sample. It is useful when the sample size is not too large and when the sample contains some repeated values. Figure 1.7 presents a dotplot for the geyser data in Table 1.3. For each value in the sample a vertical column of dots is drawn, with the number of dots in the column equal to the number of times the value appears in the sample. The dotplot gives a good indication of where the sample values are concentrated and where the gaps are. For example, it is immediately apparent Page 27 from Figure 1.7 that the sample contains no dormant periods between 61 and 65 minutes in length. FIGURE 1.7 Dotplot for the geyser data in Table 1.3. Stem-and-leaf plots and dotplots are good methods for informally examining a sample, and they can be drawn fairly quickly with pencil and paper. They are rarely used in formal presentations, however. Graphics more commonly used in formal presentations include the histogram and the boxplot, which we will now discuss. Histograms A histogram is a graphic that gives an idea of the "shape" of a sample, indicating regions where sample points are concentrated and regions where they are sparse. We will construct a histogram for the PM emissions of 62 vehicles driven at high altitude, as presented in Table 1.2 (Section 1.2). The sample values range from a low of 1.11 to a high of 23.38, in units of grams of emissions per gallon of fuel. The first step is to construct a frequency table, shown in Table 1.4. TABLE 1.4 Frequency table for PM emissions of 62 vehicles driven at high altitude Class Interval (g/gal) 1–< 3 3–< 5 5–< 7 7–< 9 9–< 11 11–< 13 13–< 15 15–< 17 17–< 19 19–< 21 Frequency 12 11 18 9 5 1 2 0 2 1 Relative Frequency 0.1935 0.1774 0.2903 0.1452 0.0806 0.0161 0.0323 0.0000 0.0323 0.0161 Density 0.0968 0.0887 0.1452 0.0726 0.0403 0.0081 0.0161 0.0000 0.0161 0.0081 21–< 23 23–< 25 0 1 0.0000 0.0161 0.0000 0.0081 The intervals in the left-hand column are called class intervals. They divide the sample into groups. For most histograms, the class intervals all have the same width. In Table 1.4, all classes have width 2. The notation 1–< 3, 3–< 5, and so on, indicates that a point on the boundary will go into the class on its right. For example, a sample value equal to 3 will go into the class 3–< 5, not 1–< 3. There is no hard-and-fast rule as to how to choose the endpoints of the class intervals. In general, it is good to have more intervals rather than fewer, but it is also good to have large numbers of sample points in the intervals. Striking the proper balance is a matter of judgment and of trial and error. When the number of observations n is large (several hundred or Page 28 more), some have suggested that reasonable starting points for the number of classes may be log2 n or 2n1/3. When the number of observations is smaller, more classes than these are often needed. The column labeled "Frequency" in Table 1.4 presents the numbers of data points that fall into each of the class intervals. The column labeled "Relative Frequency" presents the frequencies divided by the total number of data points, which for these data is 62. The relative frequency of a class interval is the proportion of data points that fall into the interval. Note that since every data point is in exactly one class interval, the relative frequencies must sum to 1. Finally, the column labeled "Density" presents the relative frequency divided by the class width. In this case all classes have width 2, so the densities are found by dividing the relative frequencies by 2. Note that when the classes are of equal width, the frequencies, relative frequencies, and densities are proportional to one another. Figure 1.8 presents a histogram for Table 1.4. The units on the horizontal axis are the units of the data, in this case grams per gallon. Each class interval is represented by a rectangle. When the class intervals are of equal width, the heights of the rectangles may be set equal to the frequencies, the relative frequencies, or the densities. Since these three quantities are proportional, the shape of the histogram will be the same in each case. For the histogram in Figure 1.8, the heights of the rectangles are the relative frequencies. FIGURE 1.8 Histogram for the data in Table 1.4. In this histogram the heights of the rectangles are the relative frequencies. Since the class widths are all the same, the frequencies, relative frequencies, and densities are proportional to one another, so it would have been equally appropriate to set the heights equal to the frequencies or to the densities. Page 29 TABLE 1.5 Frequency table, with unequal class widths, for PM emissions of 62 vehicles driven at high altitude Class Interval (g/gal) 1–< 3 3–< 5 5–< 7 7–< 9 9–< 11 11–< 15 15–< 25 Frequency 12 11 18 9 5 3 4 Relative Frequency 0.1935 0.1774 0.2903 0.1452 0.0806 0.0484 0.0645 Density 0.0968 0.0887 0.1452 0.0726 0.0403 0.0121 0.0065 Unequal Class Widths In some cases, histograms are drawn with class intervals of differing widths. This may be done when it is desired for the histogram to have a smoother appearance, or when the data come in the form of a frequency table in which the classes have unequal widths. Table 1.5 presents the PM data of Table 1.4 with the last seven classes collapsed into two. It is important to note that because the class widths vary in size, the densities are no longer proportional to the relative frequencies. Instead, the densities adjust for the width of the class. Other things being equal, wider classes tend to contain more sample items than the narrower classes, and thus tend to have larger relative frequencies. Dividing the relative frequency by the class width to obtain the density adjusts for this tendency. For this reason, when the classes have unequal widths, the heights of the rectangles must be set equal to the densities. The areas of the rectangles then represent the relative frequencies. Figure 1.9 presents the histogram for Table 1.5. Comparing this histogram to the one in Figure 1.8 shows that the string of small rectangles on the right has been smoothed out. FIGURE 1.9 Histogram for the PM emissions for high-altitude vehicles. The frequency table is presented in Table 1.5. Since the classes have differing widths, the heights of the rectangles must be set equal to the densities. The areas of the rectangles are then equal to the relative frequencies. Compare with the equal-class-width histogram in Figure 1.8. Summary When the class intervals are of unequal widths, the heights of the rectangles must be set equal to the densities. The areas of the rectangles will then be the relative frequencies. Page 30 Example 1.17 Use the histogram in Figure 1.8 to determine the proportion of the vehicles in the sample with emissions between 7 and 11 g/gal. Solution The proportion is the sum of the relative frequencies of the classes spanning the range between 7 and 11. This is found by adding the heights of the rectangles for the two class intervals covered. The result is 0.1452 + 0.0806 = 0.2258. Note that this result can also be obtained from the frequency table. The proportion of data points with values between 7 and 9 is 0.1452, and the proportion between 9 and 11 is 0.0806. The proportion between 7 and 11 is therefore equal to 0.1452 + 0.0806 = 0.2258. Example 1.18 Use the histogram in Figure 1.9 to determine the proportion of the vehicles in the sample with emissions between 9 and 15 g/gal. Solution The proportion is the sum of the relative frequencies of the two classes spanning the range between 9 and 15. Since the heights of the rectangles represent densities, the areas of the rectangles represent relative frequencies. The sum of the areas of the rectangles is (2)(0.0403) + (4)(0.0121) = 0.129. Note that this result can also be obtained from the frequency table. The proportion of data points with values between 9 and 11 is 0.0806, and the proportion between 11 and 15 is 0.0484. The proportion between 9 and 15 is therefore equal to 0.0806 + 0.0484 = 0.129. Summary To construct a histogram: ■ Choose boundary points for the class intervals. ■ Compute the frequency and relative frequency for each class. (Relative frequency is optional if the classes all have the same width.) ■ Compute the density for each class, according to the formula ■ (This step is optional if the classes all have the same width.) Draw a rectangle for each class. If the classes all have the same width, the heights of the rectangles may be set equal to the frequencies, the relative frequencies, or the densities. If the classes do not all have the same width, the heights of the rectangles must be set equal to the densities. Symmetry and Skewness Page 31 A histogram is perfectly symmetric if its right half is a mirror image of its left half. Histograms that are not symmetric are referred to as skewed. In practice, virtually no sample has a perfectly symmetric histogram; all exhibit some degree of skewness. In a skewed histogram, one side, or tail, is longer than the other. A histogram with a long right-hand tail is said to be skewed to the right, or positively skewed. A histogram with a long left-hand tail is said to be skewed to the left, or negatively skewed. While there is a formal mathematical method for measuring the skewness of a histogram, it is rarely used; instead people judge the degree of skewness informally by looking at the histogram. Figure 1.10 presents some histograms for hypothetical samples. In most cases, for a histogram that is skewed to the right (Figure 1.10c), the mean is greater than the median. The reason for this is that the mean is near the center of mass of the histogram, that is, it is near the point where the histogram would balance if supported there. For a histogram skewed to the right, more than half the data will be to the left of the center of mass. Similarly, in most cases the mean is less than the median for a histogram that is skewed to the left (Figure 1.10a). The histogram for the PM data (Figure 1.8) is skewed to the right. The sample mean is 6.596, which is greater than the sample median of 5.75. FIGURE 1.10 (a) A histogram skewed to the left. The mean is less than the median. (b) A nearly symmetric histogram. The mean and median are approximately equal. (c) A histogram skewed to the right. The mean is greater than the median. Unimodal and Bimodal Histograms We have used the term "mode" to refer to the most frequently occurring value in a sample. This term is also used in regard to histograms and other curves to refer to a peak, or local maximum. A histogram is unimodal if it has only one peak, or mode, and bimodal if it has two clearly distinct modes. In principle, a histogram can have more than two modes, but this does not happen often in practice. The histograms in Figure 1.10 are all unimodal. Figure 1.11 (page 32) presents a bimodal histogram for a hypothetical sample. FIGURE 1.11 A bimodal histogram. In some cases, a bimodal histogram indicates that the sample can be divided into Page 32 two subsamples that differ from each other in some scientifically important way. Each sample corresponds to one of the modes. As an example, Table 1.6 presents the durations of 60 dormant periods of the geyser Old Faithful (originally presented in Table 1.3). Along with the durations of the dormant period, in minutes, the duration of the eruption immediately preceding the dormant period is classified either as short (less than 3 minutes) or long (more than 3 minutes). TABLE 1.6 Durations of dormant periods (in minutes) and of the previous eruptions of the geyser Old Faithful Dormant 76 80 84 50 93 55 76 58 74 75 80 56 80 69 57 90 42 Eruption Long Long Long Short Long Short Long Short Long Long Long Short Long Long Long Long Short Dormant 91 51 79 53 82 51 76 82 84 53 86 51 85 45 88 51 80 49 82 75 73 67 68 86 72 75 75 66 84 70 79 60 86 71 67 81 76 83 76 55 Long Short Long Short Long Short Long Long Long Short Long Short Long Short Long Short Long Short Long Long Long Long Long Long Long Long Long Short Long Long Long Short Long Long Long Long Long Long Long Short Figure 1.12a presents a histogram for all 60 durations. Figures 1.12b and 1.12c present histograms for the durations following short and long eruptions, respectively. The histogram for all the durations is clearly bimodal. The histograms for the durations following short or long eruptions are both unimodal, and their modes form the two modes of the histogram for the full sample. FIGURE 1.12 (a) Histogram for all 60 durations in Table 1.6. This histogram is bimodal. (b) Histogram for the durations in Table 1.6 that follow short eruptions. (c) Histogram for the durations in Table 1.6 that follow long eruptions. The histograms for the durations following short eruptions and for those following long eruptions are both unimodal, but the modes are in different places. When the two samples are combined, the histogram is bimodal. Boxplots A boxplot is a graphic that presents the median, the first and third quartiles, and any Page 33 outliers that are present in a sample. Boxplots are easy to understand, but there is a bit of terminology that goes with them. The interquartile range is the difference between the third quartile and the first quartile. Note that since 75% of the data is less than the third quartile, and 25% of the data is less than the first quartile, it follows that 50%, or half, of the data are between the first and third quartiles. The interquartile range is therefore the distance needed to span the middle half of the data. We have defined outliers as points that are unusually large or small. If IQR represents the interquartile range, then for the purpose of drawing boxplots, any point that is more than 1.5 IQR above the third quartile, or more than 1.5 IQR below the first quartile, is considered an outlier. Some texts define a point that is more than 3 IQR from the first or third quartile as an extreme outlier. These definitions of outliers are just conventions for drawing boxplots and need not be used in other situations. Figure 1.13 (page 34) presents a boxplot for some hypothetical data. The plot consists of a box whose bottom side is the first quartile and whose top side is the third quartile. A horizontal line is drawn at the median. The "outliers" are plotted individually and are indicated by crosses in the figure. Extending from the top and bottom of the box are vertical lines called "whiskers." The whiskers end at the most extreme data point that is not an outlier. FIGURE 1.13 Anatomy of a boxplot. Apart from any outliers, a boxplot can be thought of as having four pieces: the two parts of the box separated by the median line, and the two whiskers. Again apart from outliers, each of these four parts represents one-quarter of the data. The boxplot therefore indicates how large an interval is spanned by each quarter of the data, and in this way it can be used to determine the regions in which the sample values are more densely crowded and the regions in which they are more sparse. Page 34 Steps in the Construction of a Boxplot 1. 2. 3. Compute the median and the first and third quartiles of the sample. Indicate these with horizontal lines. Draw vertical lines to complete the box. Find the largest sample value that is no more than 1.5 IQR above the third quartile, and the smallest sample value that is no more than 1.5 IQR below the first quartile. Extend vertical lines (whiskers) from the quartile lines to these points. Points more than 1.5 IQR above the third quartile, or more than 1.5 IQR below the first quartile, are designated as outliers. Plot each outlier individually. Figure 1.14 presents a boxplot for the geyser data presented in Table 1.6. First note that there are no outliers in these data. Comparing the four pieces of the boxplot, we can tell that the sample values are comparatively densely packed between the median and the first quartile, and more sparse between the median and the third quartile. The lower whisker Page 35 is a bit longer than the upper one, indicating that the data has a slightly longer lower tail than an upper tail. Since the distance between the median and the first quartile is greater than the distance between the median and the third quartile, and since the lower quarter of the data produces a longer whisker than the upper quarter, this boxplot suggests that the data are skewed to the left. FIGURE 1.14 Boxplot for the Old Faithful dormant period data presented in Table 1.6. A histogram for these data was presented in Figure 1.12a. The histogram presents a more general impression of the spread of the data. Importantly, the histogram indicates that the data are bimodal, which a boxplot cannot do. Comparative Boxplots A major advantage of boxplots is that several of them may be placed side by side, allowing for easy visual comparison of the features of several samples. Tables 1.1 and 1.2 (in Section 1.2) presented PM emissions data for vehicles driven at high and low altitudes. Figure 1.15 presents a side-by-side comparison of the boxplots for these two samples. FIGURE 1.15 Comparative boxplots for PM emissions data for vehicles driven at high versus low altitudes. The comparative boxplots in Figure 1.15 show that vehicles driven at low altitude tend to have lower emissions. In addition, there are several outliers among the data for high-altitude vehicles whose values are much higher than any of the values for the low-altitude vehicles (there is also one low-altitude value that barely qualifies as an outlier). We conclude that at high altitudes, vehicles have somewhat higher emissions in general, and that a few vehicles have much higher emissions. The box for the high-altitude vehicles is a bit taller, and the lower whisker a bit longer, than that for the low-altitude vehicles. We conclude that apart from the outliers, the spread in values is slightly larger for the high-altitude vehicles and is much larger when the outliers are considered. In Figure 1.4 (in Section 1.2) we compared the values of some numerical descriptive statistics for these two samples, and reached some conclusions similar to the previous Page 36 ones. The visual nature of the comparative boxplots in Figure 1.15 makes comparing the features of samples much easier. We have mentioned that it is important to scrutinize outliers to determine whether they have resulted from errors, in which case they may be deleted. By identifying outliers, boxplots can be useful in this regard. The following example provides an illustration. The article "Virgin Versus Recycled Wafers for Furnace Qualification: Is the Expense Justified?" (V. Czitrom and J. Reece, in Statistical Case Studies for Industrial Process Improvement, ASA and SIAM, 1997:87–104) describes a process for growing a thin silicon dioxide layer onto silicon wafers that are to be used in semiconductor manufacture. Table 1.7 presents thickness measurements, in angstroms (Å), of the oxide layer for 24 wafers. Nine measurements were made on each wafer. The wafers were produced in two separate runs, with 12 wafers in each run. TABLE 1.7 Run 1 Run 2 Oxide layer thicknesses for silicon wafers Wafer 1 90.0 92.2 94.9 92.7 2 91.8 94.5 93.9 77.3 3 90.3 91.1 93.3 93.5 4 92.6 90.3 92.8 91.6 5 91.1 89.8 91.5 91.5 6 76.1 90.2 96.8 84.6 7 92.4 91.7 91.6 91.1 8 91.3 90.1 95.4 89.6 9 96.7 93.7 93.9 87.9 10 92.0 94.6 93.7 94.0 11 94.1 91.5 95.3 92.8 12 91.7 97.4 95.1 96.7 1 93.0 89.9 93.6 89.0 2 91.4 90.6 92.2 91.9 3 91.9 91.8 92.8 96.4 4 90.6 91.3 94.9 88.3 5 93.1 91.8 94.6 88.9 6 90.8 91.5 91.5 91.5 7 88.0 91.8 90.5 90.4 8 88.3 96.0 92.8 93.7 9 94.2 92.2 95.8 92.5 10 101.5 103.1 103.2 103.5 11 92.8 90.8 92.2 91.7 12 92.1 93.4 94.0 94.7 Thicknesses (Å) 91.6 88.2 92.0 89.9 87.2 88.1 92.7 91.7 90.6 93.1 93.3 95.7 88.0 92.4 90.7 95.8 90.4 92.0 89.3 90.1 93.4 92.2 77.5 91.4 93.6 90.9 92.4 87.6 93.8 86.5 87.9 92.2 90.0 97.9 94.0 91.0 90.3 91.5 89.6 91.0 91.4 96.1 102.5 89.0 88.5 90.8 92.1 92.0 98.2 96.0 87.9 92.8 93.3 90.7 94.1 91.1 91.2 92.6 89.1 90.1 91.9 94.5 89.3 95.5 93.6 88.9 92.4 92.5 92.4 90.9 100.3 95.2 88.7 92.9 92.6 91.7 97.9 95.7 90.5 95.2 94.3 91.3 92.7 94.5 89.4 94.5 95.2 94.2 95.1 93.1 89.8 92.4 93.0 88.9 90.9 92.8 92.7 90.9 92.8 90.7 91.3 93.6 92.1 91.6 98.4 92.1 91.8 94.0 89.4 93.2 93.9 90.2 95.3 93.0 92.8 93.6 91.0 102.0 106.7 105.4 87.5 93.8 91.4 91.2 92.3 91.1 The 12 wafers in each run were of several different types and were processed in several different furnace locations. The purpose in collecting the data was to determine whether the thickness of the oxide layer was affected either by the type of wafer or the furnace location. This was therefore a factorial experiment, with wafer type and furnace location as the factors, and oxide layer thickness as the outcome. The experiment was designed so that there was not supposed to be any systematic difference in the thicknesses between one run and another. The first step in the analysis was to construct a boxplot for the data in each run to help Page 37 determine if this condition was in fact met, and whether any of the observations should be deleted. The results are presented in Figure 1.16. FIGURE 1.16 Comparative boxplots for oxide layer thickness data. The boxplots show that there were several outliers in each run. Note that apart from these outliers, there are no striking differences between the samples, and therefore no evidence of any systematic difference between the runs. The next task is to inspect the outliers, to determine which, if any, should be deleted. By examining the data in Table 1.7, it can be seen that the eight largest measurements in run 2 occurred on a single wafer: number 10. It was then determined that this wafer had been contaminated with a film residue, which caused the large thickness measurements. It would therefore be appropriate to delete these measurements. In the actual experiment, the engineers had data from several other runs available, and for technical reasons, decided to delete the entire run, rather than to analyze a run that was missing one wafer. In run 1, the three smallest measurements were found to have been caused by a malfunctioning gauge, and were therefore appropriately deleted. No cause could be determined for the remaining two outliers in run 1, so they were included in the analysis. Multivariate Data Sometimes the items in a population may have several values associated with them. For example, imagine choosing a random sample of days and determining the average temperature and humidity on each day. Each day in the population provides two values, temperature and humidity. The random sample therefore would consist of pairs of numbers. If the precipitation were measured on each day as well, the sample would consist of triplets. In principle, any number of quantities could be measured on each day, producing a sample in which each item is a list of numbers. Data for which each item consists of more than one value is called multivariate data. When each item is a pair of values, the data are said to be bivariate. One of the most useful graphical summaries for numerical bivariate data is the scatterplot. If the data consist Page 38 of ordered pairs (x1, y1), ..., (xn, yn), then a scatterplot is constructed simply by plotting each point on a two-dimensional coordinate system. Scatterplots can also be used to summarize multivariate data when each item consists of more than two values. One simply constructs separate scatterplots for each pair of values. The following example illustrates the usefulness of scatterplots. The article "Advances in Oxygen Equivalence Equations for Predicting the Properties of Titanium Welds" (D. Harwig, W. Ittiwattana, and H. Castner, The Welding Journal, 2001:126s–136s) presents data concerning the chemical composition and strength characteristics of a number of titanium welds. Figure 1.17 presents two scatterplots. Figure 1.17a is a plot of the yield strength [in thousands of pounds per square inch (ksi)] versus carbon content (in percent) for some of these welds. Figure 1.17b is a plot of the yield strength (in ksi) versus nitrogen content (in percent) for the same welds. FIGURE 1.17 (a) A scatterplot showing that there is not much of a relationship between carbon content and yield strength for a certain group of welds. (b) A scatterplot showing that for these same welds, higher nitrogen content is associated with higher yield strength. The plot of yield strength versus nitrogen content (Figure 1.17b) shows some clear structure —the points seem to be following a line from lower left to upper right. In this way, the plot illustrates a relationship between nitrogen content and yield strength: Welds with higher nitrogen content tend to have higher yield strength. This scatterplot might lead investigators to try to predict strength from nitrogen content or to try to increase nitrogen content to increase strength. (The fact that there is a relationship on a scatterplot does not guarantee that these attempts will be successful, as we will discuss in Section 7.1.) In contrast, there does not seem to be much structure to the scatterplot of yield strength versus carbon content, and thus there is no evidence of a relationship between these two quantities. This scatterplot would discourage investigators from trying to predict strength from carbon content. Page 39 Exercises for Section 1.3 1. 2. 3. 4. The weather in Los Angeles is dry most of the time, but it can be quite rainy in the winter. The rainiest month of the year is February. The following table presents the annual rainfall in Los Angeles, in inches, for each February from 1965 to 2006. 0.2 3.7 1.2 13.7 1.5 0.2 1.7 0.6 0.1 8.9 1.9 5.5 0.5 3.1 3.1 8.9 8.0 12.7 4.1 0.3 2.6 1.5 8.0 4.6 0.7 0.7 6.6 4.9 0.1 4.4 3.2 11.0 7.9 0.0 1.3 2.4 0.1 2.8 4.9 3.5 6.1 0.1 a. Construct a stem-and-leaf plot for these data. b. Construct a histogram for these data. c. Construct a dotplot for these data. d. Construct a boxplot for these data. Does the box-plot show any outliers? Forty-five specimens of a certain type of powder were analyzed for sulfur trioxide content. Following are the results, in percent. The list has been sorted into numerical order. 14.1 14.4 14.7 14.8 15.3 15.6 16.1 16.6 17.3 14.2 14.4 14.7 14.9 15.3 15.7 16.2 17.2 17.3 14.3 14.4 14.8 15.0 15.4 15.7 16.4 17.2 17.8 14.3 14.4 14.8 15.0 15.4 15.9 16.4 17.2 21.9 14.3 14.6 14.8 15.2 15.5 15.9 16.5 17.2 22.4 a. Construct a stem-and-leaf plot for these data. b. Construct a histogram for these data. c. Construct a dotplot for these data. d. Construct a boxplot for these data. Does the box-plot show any outliers? Refer to Table 1.2 (in Section 1.2). Construct a stem-and-leaf plot with the ones digit as the stem (for values greater than or equal to 10 the stem will have two digits) and the tenths digit as the leaf. How many stems are there (be sure to include leafless stems)? What are some advantages and disadvantages of this plot, compared to the one in Figure 1.6 (page 26)? Following are measurements of soil concentrations (in mg/kg) of chromium (Cr) and nickel (Ni) at 20 sites in the area of Cleveland, Ohio. These data are taken from the article "Variation in North American Regulatory Guidance for Heavy Metal Surface Soil Contamination at Commercial and Industrial Sites" (A. Jennings and J. Ma, J Environment Eng, 2007:587–609). Cr: 34 1 511 2 574 496 322 424 269 140 244 252 76 108 24 38 18 34 30 191 Ni: 23 22 55 39 283 34 159 37 61 34 163 140 32 23 54 837 5. 64 354 376 471 a. Construct a histogram for each set of concentrations. b. Construct comparative boxplots for the two sets of concentrations. c. Using the boxplots, what differences can be seen between the two sets of concentrations? A certain reaction was run several times using each of two catalysts, A and B. The catalysts were supposed to control the yield of an undesirable side product. Results, in units of percentage yield, for 24 runs of catalyst A and 20 runs of catalyst B are as follows: Catalyst A 4.4 4.9 4.1 3.6 4.3 4.4 3.4 4.6 2.6 2.9 3.9 3.1 2.6 5.2 6.7 2.6 4.8 5.7 3.8 4.7 4.1 4.0 4.5 4.5 2.9 5.8 3.1 6.7 4.3 5.5 2.5 1.6 5.2 3.8 Catalyst B 3.4 6.4 3.7 3.5 6.3 1.1 5.0 3.8 5.9 2.6 a. b. c. 6. Construct a histogram for the yields of each catalyst. Construct comparative boxplots for the yields of the two catalysts. Using the boxplots, what differences can be seen between the results of the yields of the two catalysts? Sketch a histogram for which a. the mean is greater than the median. b. the mean is less than the median. c. the mean is approximately equal to the median. Page 40 7. The figure below is a histogram showing the distribution of serum cholesterol level for a sample of men. Use the histogram to answer the following questions: a. Is the percentage of men with cholesterol levels above 240 mg/dL closest to 30%, 50%, or 70%? b. In which interval are there more men: 240–260 mg/dL or 280–340 mg/dL? 8. The histogram below presents the compressive strengths of a sample of concrete blocks hardened for 28 days. One rectangle from the histogram is missing. What is its height? Refer to Table 1.4 (in Section 1.3). a. Using the class intervals in the table, construct a histogram in which the heights of the rectangles are equal to the frequencies. b. Using the class intervals in the table, construct a histogram in which the heights of the rectangles are equal to the densities. c. Compare the histograms in parts (a) and (b) with the histogram in Figure 1.8, for which the heights are the relative frequencies. Are the shapes of the histograms the same? 10. Refer to Table 1.5 (in Section 1.3). a. Using the class intervals in the table, construct a histogram in which the heights of the rectangles are equal to the relative frequencies. b. Compare the histogram in part (a) with the histogram in Figure 1.9, for which the heights are the densities. Are the shapes of the histograms the same? c. Explain why the heights should not be set equal to the relative frequencies in this case. d. Which classes are visually exaggerated by making the heights equal to the relative frequencies? 11. The following table presents the number of students absent in a middle school in 9. northwestern Montana for each school day in January 2008. Date Jan. 2 Jan. 3 Jan. 4 Jan. 7 Jan. 8 Jan. 9 Jan. 10 Jan. 11 Jan. 14 Jan. 15 Jan. 16 Jan. 17 Jan. 18 Jan. 21 Jan. 22 Jan. 23 Jan. 24 Jan. 25 Jan. 28 Jan. 29 Jan. 30 Jan. 31 a. b. Number Absent 65 67 71 57 51 49 44 41 59 49 42 56 45 77 44 42 45 46 100 59 53 51 Construct a boxplot. There was a snowstorm on January 27. Was the number of absences the next day an outlier? 12. Which of the following statistics cannot be determined from a boxplot? i. The median ii. The mean iii. The first quartile iv. The third quartile v. The interquartile range 13. A sample of 100 resistors has an average resistance of 50 Ω and a standard deviation of 5 Ω. A second sample of 100 resistors has an average resistance of 100 Ω and a standard deviation of 5 Ω. If the two samples are combined, the standard deviation of all 200 resistances will be __________. Page 41 i. ii. less than 5 Ω greater than 5 Ω iii. equal to 5 Ω iv. can't tell from the information given (Hint: Don't do any calculations. Just try to sketch, very roughly, histograms for each sample separately, then for the combined sample.) 14. Following are boxplots comparing the amount of econozole nitrate (in µg/cm2) absorbed into skin for a brand name and a generic antifungal ointment (from the article "Improved Bioequivalence Assessment of Topical Dermatological Drug Products Using Dermatopharmacokinetics." B. N Dri-Stempfer, W. Navidi, R. Guy, and A. Bunge, Pharmaceutical Research, 2009:316–328). True or false: a. The median amount absorbed for the brand name drug is greater than the 25th percentile of the amount absorbed for the generic drug. b. The median amount absorbed for the brand name drug is greater than the median amount absorbed for the generic drug. c. About half the sample values for the brand name drug are between 2 and 3. d. There is a greater proportion of values outside the box for the brand name drug than for the generic drug. e. Both samples are skewed to the right. f. Both samples contain outliers. 15. Following are summary statistics for two data sets, A and B. Minimum 1st Quartile Median 3rd Quartile Maximum A 0.066 1.42 2.60 6.02 10.08 B −2.235 5.27 8.03 9.13 10.51 a. b. Compute the interquartile ranges for both A and B. Do the summary statistics for A provide enough information to construct a boxplot? If so, construct the boxplot. If not, explain why. c. Do the summary statistics for B provide enough information to construct a boxplot? If so, construct the boxplot. If not, explain why. 16. Match each histogram to the boxplot that represents the same data set. Page 42 17. Refer to the asphalt data in Example 1.12 (page 17). a. Construct a boxplot for the asphalt data. b. Which values, if any, are outliers? c. Construct a dotplot for the asphalt data. d. For purposes of constructing boxplots, an outlier is defined to be a point whose distance from the nearest quartile is more than 1.5 IQR. A more general, and less precise, definition is that an outlier is any point that is detached from the bulk of the data. Are there any points in the asphalt data set that are outliers under this more general definition, but not under the boxplot definition? If so, which are they? 18. Refer to Exercise 5. Construct a back-to-back stem-and-leaf plot by constructing a list of stems, then write the yields for Catalyst A to the right and the yields for Catalyst B to the left. Using this plot, what differences can be seen between the results of the yields of the two catalysts? 19. In general a histogram is skewed to the left when the median is greater than the mean and to the right when the median is less than the mean. There are exceptions, however. Consider the following data set: a. b. Compute the mean and median. Based on the mean and median, would you expect the data set to be skewed to the left, skewed to the right, or approximately symmetric? Explain. c. Construct a histogram. Does the histogram have the shape you expected? 20. Match each scatterplot to the statement that best describes it. Page 43 i. The relationship between x and y is approximately linear. ii. The relationship between x and y is nonlinear. iii. There isn't much of any relationship between x and y. iv. The relationship between x and y is approximately linear, except for an outlier. 21. For the following data: x 1.4 y 2.3 a. b. c. 2.4 3.7 4.0 5.7 4.9 9.9 5.7 6.9 6.3 7.8 9.0 9.3 11.0 15.8 15.4 36.9 34.6 53.2 Make a scatterplot of y versus x. Is the relationship between x and y approximately linear, or is it nonlinear? Compute the natural logarithm of each y value. This is known as making a log transformation of y. Make a scatterplot of ln y versus x. Is the relationship between x and ln y approximately linear, or is it nonlinear? In general, it is easier to work with quantities that have an approximate linear relationship than with quantities that have a nonlinear relationship. For these data, do you think it would be easier to work with x and y or with x and ln y? Explain. Supplementary Exercises for Chapter 1 1. 2. 3. 4. 5. A vendor converts the weights on the packages she sends out from pounds to kilograms (1 kg ≈ 2.2 lb). a. How does this affect the mean weight of the packages? b. How does this affect the standard deviation of the weights? Refer to Exercise 1. The vendor begins using heavier packaging, which increases the weight of each package by 50 g. a. How does this affect the mean weight of the packages? b. How does this affect the standard deviation of the weights? The specification for the pull strength of a wire that connects an integrated circuit to its frame is 10 g or more. Units made with aluminum wire have a defect rate of 10%. A redesigned manufacturing process, involving the use of gold wire, is being investigated. The goal is to reduce the rate of defects to 5% or less. Out of the first 100 units manufactured with gold wire, only 4 are defective. True or false: a. Since only 4% of the 100 units were defective, we can conclude that the goal has been reached. b. Although the sample percentage is under 5%, this may represent sampling variation, so the goal may not yet be reached. c. There is no use in testing the new process, because no matter what the result is, it could just be due to sampling variation. d. If we sample a large enough number of units, and if the percentage of defective units is far enough below 5%, then it is reasonable to conclude that the goal has been reached. A coin is tossed twice and comes up heads both times. Someone says, "There's something wrong with this coin. A coin is supposed to come up heads only half the time, not every time." a. Is it reasonable to conclude that something is wrong with the coin? Explain. b. If the coin came up heads 100 times in a row, would it be reasonable to conclude that something is wrong with the coin? Explain. The smallest number on a list is changed from 12.9 to 1.29. a. Is it possible to determine by how much the mean changes? If so, by how much does it change? b. Is it possible to determine by how much the median changes? If so, by how much does it change? What if the list consists of only two numbers? c. Is it possible to determine by how much the standard deviation changes? If so, by how much does it change? Page 44 6. There are 15 numbers on a list, and the smallest number is changed from 12.9 to 1.29. a. Is it possible to determine by how much the mean changes? If so, by how much does it change? b. Is it possible to determine the value of the mean after the change? If so, what is the value? c. Is it possible to determine by how much the median changes? If so, by how much does 7. 8. 9. it change? d. Is it possible to determine by how much the standard deviation changes? If so, by how much does it change? There are 15 numbers on a list, and the mean is 25. The smallest number on the list is changed from 12.9 to 1.29. a. Is it possible to determine by how much the mean changes? If so, by how much does it change? b. Is it possible to determine the value of the mean after the change? If so, what is the value? c. Is it possible to determine by how much the median changes? If so, by how much does it change? d. Is it possible to determine by how much the standard deviation changes? If so, by how much does it change? The article "The Selection of Yeast Strains for the Production of Premium Quality South African Brandy Base Products" (C. Steger and M. Lambrechts, Journal of Industrial Microbiology and Biotechnology, 2000:431–440) presents detailed information on the volatile compound composition of base wines made from each of 16 selected yeast strains. Following are the concentrations of total esters (in mg/L) in each of the wines. 284.34 173.01 229.55 312.95 215.34 188.72 144.39 172.79 139.38 197.81 303.28 256.02 658.38 105.14 295.24 170.41 a. Compute the mean concentration. b. Compute the median concentration. c. Compute the first quartile of the concentrations. d. Compute the third quartile of the concentrations. e. Construct a boxplot for the concentrations. What features does it reveal? Concerning the data represented in the following boxplot, which one of the following statements is true? i. The mean is greater than the median. ii. The mean is less than the median. iii. The mean is approximately equal to the median. 10. True or false: In any boxplot, a. b. c. d. The length of the whiskers is equal to 1.5 IQR, where IQR is the interquartile range. The length of the whiskers may be greater than 1.5 IQR, where IQR is the interquartile range. The length of the whiskers may be less than 1.5 IQR, where IQR is the interquartile range. The values at the ends of the whiskers are always values in the data set used to construct the boxplot. Page 45 11. For each of the following histograms, determine whether the vertical axis has been labeled correctly. 12. In the article "Occurrence and Distribution of Ammonium in Iowa Groundwater" (K. Schilling, Water Environment Research, 2002:177–186), ammonium concentrations (in mg/L) were measured at a total of 349 alluvial wells in the state of Iowa. The mean concentration was 0.27, the median was 0.10, and the standard deviation was 0.40. If a histogram of these 349 measurements were drawn, i. it would be most likely skewed to the right. ii. it would be most likely skewed to the left. iii. it would be approximately symmetric. iv. its shape could not be determined without knowing the relative frequencies. 13. The article "Vehicle-Arrival Characteristics at Urban Uncontrolled Intersections" (V. Rengaraju and V. Rao, Journal of Transportation Engineering, 1995: 317–323) presents data on traffic characteristics at 10 intersections in Madras, India. One characteristic measured was the speeds of the vehicles traveling through the intersections. The accompanying table gives the 15th, 50th, and 85th percentiles of speed (in km/h) for two intersections. Intersection A B 15th 27.5 24.5 Percentile 50th 37.5 26.5 85th 40.0 36.0 a. If a histogram for speeds of vehicles through intersection A were drawn, would it most likely be skewed to the left, skewed to the right, or approximately symmetric? Explain. b. If a histogram for speeds of vehicles through intersection B were drawn, would it most likely be skewed to the left, skewed to the right, or approximately symmetric? Explain. 14. The cumulative frequency and the cumulative relative frequency for a given class interval are the sums of the frequencies and relative frequencies, respectively, over all classes up to and including the given class. For example, if there are five classes, with frequencies 11, 7, 3, 14, and 5, the cumulative frequencies would be 11, 18, 21, 35, and 40, and the Page 46 cumulative relative frequencies would be 0.275, 0.450, 0.525, 0.875, and 1.000. Construct a table presenting frequencies, relative frequencies, cumulative frequencies, and cumulative relative frequencies, for the data in Exercise 2 of Section 1.3, using the class intervals 14–< 15, 15–< 16, ..., 22–< 23. 15. The article "Computing and Using Rural versus Urban Measures in Statistical Applications" (C. Goodall, K. Kafadar, and J. Tukey, The American Statistician, 1998:101–111) discusses methods to measure the degree to which U.S. counties are urban rather than rural. The following frequency table presents population frequencies of U.S. counties. Populations are on the log2 scale; thus the first interval contains counties whose populations are at least 26 = 64 but less than 212.4 = 5404, and so on. log2 Population 6.0–< 12.4 12.4–< 13.1 13.1–< 13.6 13.6–< 14.0 14.0–< 14.4 Number of Counties 305 294 331 286 306 14.4–< 14.8 14.8–< 15.3 15.3–< 16.0 16.0–< 17.0 17.0–< 23.0 273 334 326 290 323 a. b. c. d. Construct a histogram from the frequency table. Estimate the proportion of counties whose populations are greater than 100,000. Is the histogram skewed to the left, skewed to the right, or approximately symmetric? Construct a histogram using the actual populations rather than their logs. Why do you think the article transformed the populations to the log scale? 16. The article "Hydrogeochemical Characteristics of Groundwater in a Mid-Western Coastal Aquifer System" (S. Jeen, J. Kim, et al., Geosciences Journal, 2001:339–348) presents measurements of various properties of shallow groundwater in a certain aquifer system in Korea. Following are measurements of electrical conductivity (in microsiemens per centimeter) for 23 water samples. 2099 528 2030 1350 1018 384 1499 1265 375 424 789 810 522 513 488 200 215 486 257 557 260 461 500 a. Find the mean. b. Find the standard deviation. c. Find the median. d. Construct a dotplot. e. Find the 10% trimmed mean. f. Find the first quartile. g. Find the third quartile. h. Find the interquartile range. i. Construct a boxplot. j. Which of the points, if any, are outliers? k. If a histogram were constructed, would it be skewed to the left, skewed to the right, or approximately symmetric? 17. Water scarcity has traditionally been a major concern in the Canary Islands. Water rights are divided into shares, which are privately owned. The article "The Social Construction of Scarcity. The Case of Water in Tenerife (Canary Islands)" (F. Aguilera-Klink, E. PerezMoriana, and J. Sanchez-Garcia, Ecological Economics, 2000:233–245) discusses the extent to which many of the shares are concentrated among a few owners. The following table presents the number of owners who own various numbers of shares. (There were 15 owners who owned 50 shares or more; these are omitted.) Note that it is possible to own a fractional number of shares; for example, the interval 2–< 3 contains 112 individuals who owned at least 2 but less than 3 shares. Number of Shares Number of Owners 0–< 1 1–< 2 2–< 3 3–< 4 4–< 5 5–< 10 10–< 15 15–< 20 20–< 25 25–< 30 30–< 50 18 165 112 87 43 117 51 32 10 8 8 Page 47 a. b. Construct a histogram for these data. Approximate the median number of shares owned by finding the point for which the areas on either side are equal. c. Approximate the first quartile of the number of shares owned by finding the point for which 25% of the area is to the left. d. Approximate the third quartile of the number of shares owned by finding the point for which 75% of the area is to the left. 18. The Editor's Report in the November 2003 issue of Technometrics provides the following information regarding the length of time taken to review articles that were submitted for publication during the year 2002. A few articles took longer than 9 months to review, these are omitted from the table. Time (months) 0–< 1 1–< 2 2–< 3 3–< 4 4–< 5 5–< 6 6–< 7 7–< 8 8–< 9 Number of Articles 45 17 18 19 12 14 13 22 11 a. Construct a histogram for these data. b. Which class interval contains the median review time? c. Which class interval contains the first quartile of the review times? d. Which class interval contains the third quartile of the review times? 19. The article " The Ball-on-Three-Ball Test for Tensile Strength: Refined Methodology and Results for Three Hohokam Ceramic Types" (M. Beck, American Antiquity, 2002:558–569) discusses the strength of ancient ceramics. Several specimens of each of three types of ceramic were tested. The loads (in kg) required to crack the specimens are as follows: Ceramic Type Loads (kg) Sacaton 15, 30, 51, 20, 17, 19, 20, 32, 17, 15, 23, 19, 15, 18, 16, 22, 29, 15, 13, 15 Gila 27, 18, 28, 25, 55, 21, 18, 34, 23, 30, 20, 30, 31, 25, Plain 28, 26, 17, 19, 16, 24, 19, 9, 31, 19, 27, 20, 43, 15 Casa 20, 16, 20, 36, 27, 35, 66, 15, 18, 24, 21, 30, 20, 24, Grande 23, 21, 13, 21 a. b. c. Construct comparative boxplots for the three samples. How many outliers does each sample contain? Comment on the features of the three samples. Page 48 Chapter 2 Probability Introduction The development of the theory of probability was financed by seventeenth-century gamblers, who hired some of the leading mathematicians of the day to calculate the correct odds for certain games of chance. Later, people realized that scientific processes involve chance as well, and since then the methods of probability have been used to study the physical world. Probability is now an extensive branch of mathematics. Many books are devoted to the subject, and many researchers have dedicated their professional careers to its further development. In this chapter we present an introduction to the ideas of probability that are most important to the study of statistics. 2.1 Basic Ideas To make a systematic study of probability, we need some terminology. An experiment is a process that results in an outcome that cannot be predicted in advance with certainty. Tossing a coin, rolling a die, measuring the diameter of a bolt, weighing the contents of a box of cereal, and measuring the breaking strength of a length of fishing line are all examples of experiments. To discuss an experiment in probabilistic terms, we must specify its possible outcomes: Definition The set of all possible outcomes of an experiment is called the sample space for the experiment. For tossing a coin, we can use the set {Heads, Tails} as the sample space. For rolling a sixsided die, we can use the set {1, 2, 3, 4, 5, 6}. These sample spaces are finite. Some experiments have sample spaces with an infinite number of outcomes. For example, imagine that a punch with diameter 10 mm punches holes in sheet metal. Because of variations in the angle of Page 49 the punch and slight movements in the sheet metal, the diameters of the holes vary between 10.0 and 10.2 mm. For the experiment of punching a hole, then, a reasonable sample space is the interval (10.0, 10.2), or in set notation, {x | 10.0 < x < 10.2}. This set obviously contains an infinite number of outcomes. For many experiments, there are several sample spaces to choose from. For example, assume that a process manufactures steel pins whose lengths vary between 5.20 and 5.25 cm. An obvious choice for the sample space for the length of a pin is the set {x | 5.20 < x < 5.25}. However, if the object were simply to determine whether the pin was too short, too long, or within specification limits, a good choice for the sample space might be {too short, too long, within specifications}. When discussing experiments, we are often interested in a particular subset of outcomes. For example, we might be interested in the probability that a die comes up an even number. The sample space for the experiment is {1, 2, 3, 4, 5, 6}, and coming up even corresponds to the subset {2, 4, 6}. In the hole punch example, we might be interested in the probability that a hole has a diameter less than 10.1 mm. This corresponds to the subset {x | 10.0 < x < 10.1}. There is a special name for a subset of a sample space: Definition A subset of a sample space is called an event. Note that for any sample space, the empty set ∅ is an event, as is the entire sample space. A given event is said to have occurred if the outcome of the experiment is one of the outcomes in the event. For example, if a die comes up 2, the events {2, 4, 6} and {1, 2, 3} have both occurred, along with every other event that contains the outcome "2." Example 2.1 An electrical engineer has on hand two boxes of resistors, with four resistors in each box. The resistors in the first box are labeled 10 Ω (ohms), but in fact their resistances are 9, 10, 11, and 12 Ω. The resistors in the second box are labeled 20 Ω, but in fact their resistances are 18, 19, 20, and 21 Ω. The engineer chooses one resistor from each box and determines the resistance of each. Let A be the event that the first resistor has a resistance greater than 10, let B be the event that the second resistor has a resistance less than 19, and let C be the event that the sum of the resistances is equal to 28. Find a sample space for this experiment, and specify the subsets corresponding to the events A, B, and C. Solution A good sample space for this experiment is the set of ordered pairs in which the first component is the resistance of the first resistor and the second component is the resistance of the second resistor. We will denote this sample space by . Page 50 The events A, B, and C are given by Combining Events We often construct events by combining simpler events. Because events are subsets of sample spaces, it is traditional to use the notation of sets to describe events constructed in this way. We review the necessary notation here. ■ The union of two events A and B, denoted A ∪ B, is the set of outcomes that belong either to A, to B, or to both. In words, A ∪ B means "A or B." Thus the event A ∪ B occurs whenever either A or B (or both) occurs. ■ The intersection of two events A and B, denoted A ∪ B, is the set of outcomes that belong both to A and to B. In words, A ∩ B means "A and B." Thus the event A ∩ B occurs whenever both A and B occur. ■ The complement of an event A, denoted Ac, is the set of outcomes that do not belong to A. In words, Ac means "not A." Thus the event Ac occurs whenever A does not occur. Events can be graphically illustrated with Venn diagrams. Figure 2.1 illustrates the events A ∪ B, A ∩ B, and B ∩ Ac. FIGURE 2.1 Venn diagrams illustrating various events: (a) A ∪ B, (b) A ∩ B, (c) B ∩ Ac. Example 2.2 Refer to Example 2.1. Find B ∪ C and A ∩ Bc. Solution The event B ∪ C contains all the outcomes that belong either to B or to C, or to both. Therefore The event Bc contains those outcomes in the sample space that do not belong to B. It Page 51 follows that the event A ∩ Bc contains the outcomes that belong to A and do not belong to B. Therefore Mutually Exclusive Events There are some events that can never occur together. For example, it is impossible that a coin can come up both heads and tails, and it is impossible that a steel pin can be both too long and too short. Events like this are said to be mutually exclusive. Definition ■ ■ The events A and B aresaidtobe mutually exclusive if they have no outcomes in common. More generally, a collection of events A1,A2,...,An is said to be mutually exclusive if no two of them have any outcomes in common. The Venn diagram in Figure 2.2 illustrates mutually exclusive events. FIGURE 2.2 The events A and B are mutually exclusive. Example 2.3 Refer to Example 2.1. If the experiment is performed, is it possible for events A and B both to occur? How about B and C? A and C? Which pair of events is mutually exclusive? Solution If the outcome is (11, 18) or (12, 18), then events A and B both occur. If the outcome is (10, 18), then both B and C occur. It is impossible for A and C both to occur, because these events are mutually exclusive, having no outcomes in common. Page 52 Probabilities Each event in a sample space has a probability of occurring. Intuitively, the probability is a quantitative measure of how likely the event is to occur. Formally speaking, there are several interpretations of probability; the one we shall adopt is that the probability of an event is the proportion of times the event would occur in the long run, if the experiment were to be repeated over and over again. We often use the letter P to stand for probability. Thus when tossing a coin, the notation "P(heads) = 1/2" means that the probability that the coin lands heads is equal to 1/2. Summary Given any experiment and any event A: ■ The expression P(A) denotes the probability that the event A occurs. ■ P(A) is the proportion of times that event A would occur in the long run, if the experiment were to be repeated over and over again. In many situations, the only way to estimate the probability of an event is to repeat the experiment many times and determine the proportion of times that the event occurs. For example, if it is desired to estimate the probability that a printed circuit board manufactured by a certain process is defective, it is usually necessary to produce a number of boards and test them to determine the proportion that are defective. In some cases, probabilities can be determined through knowledge of the physical nature of an experiment. For example, if it is known that the shape of a die is nearly a perfect cube and that its mass is distributed nearly uniformly, it may be assumed that each of the six faces is equally likely to land upward when the die is rolled. Once the probabilities of some events have been found through scientific knowledge or experience, the probabilities of other events can be computed mathematically. For example, if it has been estimated through experimentation that the probability that a printed circuit board is defective is 0.10, an estimate of the probability that a board is not defective can be calculated to be 0.90. As another example, assume that steel pins manufactured by a certain process can fail to meet a length specification either by being too short or too long. By measuring a large number of pins, it is estimated that the probability that a pin is too short is 0.02, and the probability that a pin is too long is 0.03. It can then be estimated that the probability that a pin fails to meet the specification is 0.05. In practice, scientists and engineers estimate the probabilities of some events on the basis of scientific understanding and experience and then use mathematical rules to compute estimates of the probabilities of other events. In the rest of this section and in Section 2.2, we will explain some of these rules and show how to use them. Axioms of Probability The subject of probability is based on three commonsense rules, known as axioms. They are: Page 53 The Axioms of Probability 1. Let be a sample space. Then . 2. For any event A, 0 ≤ P(A) ≤ 1. 3. If A and B are mutually exclusive events, then P(A ∪ B) = P(A) + P(B). More generally, if A1,A2, ... are mutually exclusive events, then P(A1 ∪ A2 ∪ ···) = P(A1) + P(A2) + ···. With a little thought, it is easy to see that the three axioms do indeed agree with common sense. The first axiom says that the outcome of an experiment is always in the sample space. This is obvious, because by definition the sample space contains all the possible outcomes of the experiment. The second axiom says that the long-run frequency of any event is always between 0 and 100%. For an example illustrating the third axiom, we previously discussed a process that manufactures steel pins, in which the probability that a pin is too short is 0.02 and the probability that a pin is too long is 0.03. The third axiom says that the probability that the pin is either too short or too long is 0.02 + 0.03 = 0.05. We now present two simple rules that are helpful in computing probabilities. These rules are intuitively obvious, and they can also be proved from the axioms. Proofs are provided at the end of the section. For any event A, (2.1) Let ∅ denote the empty set. Then (2.2) Equation (2.1) says that the probability that an event does not occur is equal to 1 minus the probability that it does occur. For example, if there is a 40% chance of rain, there is a 60% chance that it does not rain. Equation (2.2) says that it is impossible for an experiment to have no outcome. Example 2.4 A target on a test firing range consists of a bull's-eye with two concentric rings around it. A projectile is fired at the target. The probability that it hits the bull’s-eye is 0.10, the probability that it hits the inner ring is 0.25, and the probability that it hits the outer ring is 0.45. What is the probability that the projectile hits the target? What is the probability that it misses the target? Solution Hitting the bull’s-eye, hitting the inner ring, and hitting the outer ring are mutually exclusive events, since it is impossible for more than one of these events to occur. Therefore, using Axiom 3, Page 54 We can now compute the probability that the projectile misses the target by using Equation (2.1): Example 2.5 The following table presents probabilities for the number of times that a certain computer system will crash in the course of a week. Let A be the event that there are more than two crashes during the week, and let B be the event that the system crashes at least once. Find a sample space. Then find the subsets of the sample space that correspond to the events A and B. Then find P(A) and P(B). Number of Crashes 0 1 2 3 4 Probability 0.60 0.30 0.05 0.04 0.01 Solution A sample space for the experiment is the set {0, 1, 2, 3, 4}. The events are A = {3, 4} and B = {1, 2, 3, 4}. To find P(A), notice that A is the event that either 3 crashes happen or 4 crashes happen. The events “3 crashes happen” and “4 crashes happen” are mutually exclusive. Therefore, using Axiom 3, we conclude that We will compute P(B) in two ways. First, note that Bc is the event that no crashes happen. Therefore, using Equation (2.1), For a second way to compute P(B), note that B is the event that 1 crash happens or 2 crashes happen or 3 crashes happen or 4 crashes happen. These events are mutually exclusive. Therefore, using Axiom 3, we conclude that In Example 2.5, we computed the probabilities of the events A = {3, 4} and B = {1, 2, 3, 4} by summing the probabilities of the outcomes in each of the events: P(A) = P(3) + P(4) and P(B) = P(1) + P(2) + P(3) + P(4). This method works in general. Since any two outcomes in a sample space are mutually exclusive, the probability of any event that contains a finite number of outcomes can be found by summing the probabilities of the outcomes that make up the event. If A is an event containing outcomes O1,...,On, that is, if A = {O1,...,On}, then (2.3) Sample Spaces with Equally Likely Outcomes For some experiments, a sample space can be constructed in which all the outcomes are equally likely. A simple example is the roll of a fair die, in which the sample space is {1, 2, 3, 4, 5, 6} and each of these outcomes has probability 1/6. Another type of experiment that results in equally likely outcomes is the random selection of an item from a population of items. The items in the population can be thought of as the outcomes in a sample space, and each item is equally likely to be selected. A population from which an item is sampled at random can be thought of as a sample space with equally likely outcomes. If a sample space contains N equally likely outcomes, the probability of each outcome is 1/N. This is so, because the probability of the whole sample space must be 1, and this probability is equally divided among the N outcomes. If A is an event that contains k outcomes, then P(A) can be found by summing the probabilities of the k outcomes, so P(A) = k/N. If S is a sample space containing N equally likely outcomes, and if A is an event containing k outcomes, then (2.4) Page 56 Example 2.6 An extrusion die is used to produce aluminum rods. Specifications are given for the length and the diameter of the rods. For each rod, the length is classified as too short, too long, or OK, and the diameter is classified as too thin, too thick, or OK. In a population of 1000 rods, the number of rods in each class is as follows: Length Too Short OK Too Long Too Thin 10 38 2 Diameter OK 3 900 25 Too Thick 5 4 13 A rod is sampled at random from this population. What is the probability that it is too short? Solution We can think of each of the 1000 rods as an outcome in a sample space. Each of the 1000 outcomes is equally likely. We’ll solve the problem by counting the number of outcomes that correspond to the event. The number of rods that are too short is 10 + 3 + 5 = 18. Since the total number of rods is 1000, The Addition Rule If A and B are mutually exclusive events, then P(A ∪ B) = P(A) + P(B). This rule can be generalized to cover the case where A and B are not mutually exclusive. Example 2.7 illustrates the reasoning. Example 2.7 Refer to Example 2.6. If a rod is sampled at random, what is the probability that it is either too short or too thick? Solution First we’ll solve this problem by counting the number of outcomes that correspond to the event. In the following table the numbers of rods that are too thick are circled, and the numbers of rods that are too short have rectangles around them. Note that there are 5 rods that are both too short and too thick. Page 57 Of the 1000 outcomes, the number that are either too short or too thick is 10 + 3 + 5 + 4 + 13 = 35. Therefore Now we will solve the problem in a way that leads to a more general method. In the sample space, there are 10 + 3 + 5 = 18 rods that are too short and 5 + 4 + 13 = 22 rods that are too thick. But if we try to find the number of rods that are either too short or too thick by adding 18 + 22, we get too large a number (40 instead of 35). The reason is that there are five rods that are both too short and too thick, and these are counted twice. We can still solve the problem by adding 18 and 22, but we must then subtract 5 to correct for the double counting. We restate this reasoning, using probabilities: The method of Example 2.7 holds for any two events in any sample space. In general, to find the probability that either of two events occurs, add the probabilities of the events and then subtract the probability that they both occur. Summary Let A and B be any events. Then (2.5) A proof of this result, based on the axioms, is provided at the end of this section. Note that if A and B are mutually exclusive, then P(A ∩ B) = 0, so Equation (2.5) reduces to Axiom 3 in this case. Page 58 Example 2.8 In a process that manufactures aluminum cans, the probability that a can has a flaw on its side is 0.02, the probability that a can has a flaw on the top is 0.03, and the probability that a can has a flaw on both the side and the top is 0.01. What is the probability that a randomly chosen can has a flaw? What is the probability that it has no flaw? Solution We are given that P(flaw on side) = 0.02, P(flaw on top) = 0.03, and P(flaw on side and flaw on top) = 0.01. Now P(flaw) = P(flaw on side or flaw on top). Using Equation (2.5), To find the probability that a can has no flaw, we compute Venn diagrams can sometimes be useful in computing probabilities by showing how to express an event as the union of disjoint events. Example 2.9 Refer to Example 2.8. What is the probability that a can has a flaw on the top but not on the side? Solution Let A be the event that a can has a flaw on the top and let B be the event that a can has a flaw on the side. The following Venn diagram (Figure 2.3) shows that A is the union of the mutually exclusive events A ∩ B and A ∩ Bc. It follows that P(A) = P(A ∩ B) + P(A ∩ Bc). We know that P(A) = 0.03 and P(A ∩ B) = 0.01. Therefore 0.03 = 0.01 + P(A ∩ Bc), so P(A ∩ Bc) = 0.02. FIGURE 2.3 The events A ∩ B and A ∩ Bc are mutually exclusive, and their union is the event A. Page 59 Proof that P(Ac) = 1 − P(A) Let S be a sample space and let A be an event. Then A and Ac are mutually exclusive, so by Axiom 3, But by Axiom 1, Therefore It follows that P(Ac) = 1 − P(A). Proof that P(∅) = 0 Let be a sample space. Therefore . Proof that P(A ∪ B) = P(A) + P(B) − P(A ∩ B) Let A and B be any two events. The key to the proof is to write A ∪ B as the union of three mutually exclusive events: A ∩ Bc, A ∩ B, and Ac ∩ B. (2.6) The following figure illustrates Equation (2.6). By Axiom 3, (2.7) Now A = (A ∩ Bc) ∪ (A ∩ B), and B = (Ac ∩ B) ∪ (A ∩ B). Therefore (2.8) and (2.9) Summing Equations (2.8) and (2.9) yields (2.10) Comparing Equations (2.10) and (2.7) shows that (2.11) It follows that P(A ∪ B) = P(A) + P(B) − P(A ∩ B). Exercises for Section 2.1 1. 2. 3. 4. Page 60 The probability that a bolt meets a strength specification is 0.87. What is the probability that the bolt does not meet the specification? A die (six faces) has the number 1 painted on three of its faces, the number 2 painted on two of its faces, and the number 3 painted on one face. Assume that each face is equally likely to come up. a. Find a sample space for this experiment. b. Find P(odd number). c. If the die were loaded so that the face with the 3 on it were twice as likely to come up as each of the other five faces, would this change the sample space? Explain. d. If the die were loaded so that the face with the 3 on it were twice as likely to come up as each of the other five faces, would this change the value of P(odd number)? Explain. A section of an exam contains four True-False questions. A completed exam paper is selected at random, and the four answers are recorded. a. List all 16 outcomes in the sample space. b. Assuming the outcomes to be equally likely, find the probability that all the answers are the same. c. Assuming the outcomes to be equally likely, find the probability that exactly one of the four answers is “True.” d. Assuming the outcomes to be equally likely, find the probability that at most one of the four answers is “True.” Three times each day, a quality engineer samples a component from a recently manufactured batch and tests it. Each part is classified as conforming (suitable for its intended use), downgraded (unsuitable for the intended purpose but usable for another purpose), or scrap (not usable). An experiment consists of recording the categories of the three parts tested in a particular day. a. List the 27 outcomes in the sample space. b. Let A be the event that all the parts fall into the same category. List the outcomes in A. c. Let B be the event that there is one part in each category. List the outcomes in B. d. Let C be the event that at least two parts are conforming. List the outcomes in C. e. List the outcomes in A ∩ C. f. List the outcomes in A ∪ B. g. List the outcomes in A ∩ Cc. h. List the outcomes in Ac ∩ C. i. Are events A and C mutually exclusive? Explain. j. Are events B and C mutually exclusive? Explain. Four candidates are to be interviewed for a job. Two of them, numbered 1 and 2, are qualified, and the other two, numbered 3 and 4, are not. The candidates are interviewed at random, and the first qualified candidate interviewed will be hired. The outcomes are the sequences of candidates that are interviewed. So one outcome is 2, and another is 431. a. List all the possible outcomes. b. Let A be the event that only one candidate is interviewed. List the outcomes in A. c. Let B be the event that three candidates are interviewed. List the outcomes in B. d. Let C be the event that candidate 3 is interviewed. List the outcomes in C. e. Let D be the event that candidate 2 is not interviewed. List the outcomes in D. f. Let E be the event that candidate 4 is interviewed. Are A and E mutually exclusive? How about B and E, C and E, D and E? Refer to Exercise 5. Two candidates are randomly selected. a. List the equally likely outcomes. b. What is the probability that both are qualified? c. What is the probability that exactly one is qualified? According to a report by the Agency for Healthcare Research and Quality, the age distribution for people admitted to a hospital for an asthma-related illness was as follows. Age (years) Less than 1 1–17 18–44 45–64 65–84 85 and up 8. Proportion 0.02 0.25 0.19 0.30 0.20 0.04 Page 61 a. What is the probability that an asthma patient is between 18 and 64 years old? b. What is the probability that an asthma patient is less than 85 years old? A company audit showed that of all bills that were sent out, 71% of them were paid on time, 18% were paid up to 30 days late, 9% were paid between 31 and 90 days late, and 2% remained unpaid after 90 days. One bill is selected at random. a. What is the probability that the bill was paid on time? b. What is the probability that the bill was paid late? 9. 10. 11. 12. 13. 14. 15. Among the cast aluminum parts manufactured on a certain day, 80% were flawless, 15% had only minor flaws, and 5% had major flaws. Find the probability that a randomly chosen part a. has a flaw (major or minor). b. has no major flaw. The article “High Cumulative Risk of Lung Cancer Death among Smokers and Nonsmokers” (P. Brennan, et al. American Journal of Epidemiology, 2006:1233–1241) states that the probability is 0.24 that a man who is a heavy smoker will contract lung cancer. True or false: a. In a sample of 100 men who are heavy smokers, exactly 24 of them will contract lung cancer. b. In a sample of 100 men who are heavy smokers, the number who will contract lung cancer is likely to be close to 24, but not exactly equal to 24. c. As more and more heavy-smoking men are sampled, the proportion who contract lung cancer will approach 0.24. Resistors manufactured by a certain process are labeled as having a resistance of 5 Ω. A sample of 100 resistors is drawn, and 87 of them have resistances between 4.9 and 5.1 Ω. True or false: a. The probability that a resistor has a resistance between 4.9 and 5.1 Ω is 0.87. b. The probability that a resistor has a resistance between 4.9 and 5.1 Ω is likely to be close to 0.87, but not exactly equal to 0.87. Let V be the event that a computer contains a virus, and let W be the event that a computer contains a worm. Suppose P(V) = 0.15, P(W) = 0.05, and P(V ∪ W) = 0.17. a. Find the probability that the computer contains both a virus and a worm. b. Find the probability that the computer contains neither a virus nor a worm. c. Find the probability that the computer contains a virus but not a worm. Let S be the event that a randomly selected college student has taken a statistics course, and let C be the event that the same student has taken a chemistry course. Suppose P(S) = 0.4, P(C) = 0.3, and P(S ∩ C) = 0.2. a. Find the probability that a student has taken statistics, chemistry, or both. b. Find the probability that a student has taken neither statistics nor chemistry. c. Find the probability that a student has taken statistics but not chemistry. Six hundred paving stones were examined for cracks, and 15 were found to be cracked. The same 600 stones were then examined for discoloration, and 27 were found to be discolored. A total of 562 stones were neither cracked nor discolored. One of the 600 stones is selected at random. a. Find the probability that it is cracked, discolored, or both. b. Find the probability that it is both cracked and discolored. c. Find the probability that it is cracked but not discolored. All the fourth-graders in a certain elementary school took a standardized test. A total of 85% of the students were found to be proficient in reading, 78% were found to be proficient in mathematics, and 65% were found to be proficient in both reading and mathematics. A student is chosen at random. a. b. c. What is the probability that the student is proficient in mathematics but not in reading? What is the probability that the student is proficient in reading but not in mathematics? What is the probability that the student is proficient in neither reading nor mathematics? Page 62 16. A system contains two components, A and B. The system will function so long as either A or B functions. The probability that A functions is 0.95, the probability that B functions is 0.90, and the probability that both function is 0.88. What is the probability that the system functions? 17. A system contains two components, A and B. The system will function only if both components function. The probability that A functions is 0.98, the probability that B functions is 0.95, and the probability that either A or B functions is 0.99. What is the probability that the system functions? 18. Human blood may contain either or both of two antigens, A and B. Blood that contains only the A antigen is called type A, blood that contains only the B antigen is called type B, blood that contains both antigens is called type AB, and blood that contains neither antigen is called type O. At a certain blood bank, 35% of the blood donors have type A blood, 10% have type B, and 5% have type AB. a. What is the probability that a randomly chosen blood donor is type O? b. A recipient with type A blood may safely receive blood from a donor whose blood does not contain the B antigen. What is the probability that a randomly chosen blood donor may donate to a recipient with type A blood? 19. True or false: If A and B are mutually exclusive, a. P(A ∪ B) = 0 b. P(A ∩ B) = 0 c. P(A ∪ B) = P(A ∩ B) d. P(A ∪ B) = P(A) + P(B) 20. A flywheel is attached to a crankshaft by 12 bolts, numbered 1 through 12. Each bolt is checked to determine whether it is torqued correctly. Let A be the event that all the bolts are torqued correctly, let B be the event that the #3 bolt is not torqued correctly, let C be the event that exactly one bolt is not torqued correctly, and let D be the event that bolts #5 and #8 are torqued correctly. State whether each of the following pairs of events is mutually exclusive. a. A and B b. B and D c. C and D d. B and C 2.2 Counting Methods When computing probabilities, it is sometimes necessary to determine the number of outcomes in a sample space. In this section we will describe several methods for doing this. The basic rule, which we will call the fundamental principle of counting, is presented by means of Example 2.10. Example 2.10 A certain make of automobile is available in any of three colors: red, blue, green, and with either a large or small engine. In how many ways can a buyer choose a car? Solution There are three choices of color and two choices of engine. A complete list of choices is written in the following 3 × 2 table. The total number of choices is (3)(2) = 6. Red Blue Green Large Red, Large Blue, Large Green, Large Small Red, Small Blue, Small Green, Small To generalize Example 2.10, if there are n1 choices of color and n2 choices of Page 63 engine, a complete list of choices can be written in an n1×n2 table, so the total number of choices is If an operation can be performed in n1 ways, and if for each of these ways a second operation can be performed in n2 ways, then the total number of ways to perform the two operations is n1 n2. The fundamental principle of counting states that this reasoning can be extended to any number of operations. The Fundamental Principle of Counting Assume that k operations are to be performed. If there are n1 ways to perform the first operation, and if for each of these ways there are n2 ways to perform the second operation, and if for each choice of ways to perform the first two operations there are n3 ways to perform the third operation, and so on, then the total number of ways to perform the sequence of k operations is n1n2 ⋯ nk. Example 2.11 When ordering a certain type of computer, there are 3 choices of hard drive, 4 choices for the amount of memory, 2 choices of video card, and 3 choices of monitor. In how many ways can a computer be ordered? Solution The total number of ways to order a computer is (3)(4)(2)(3) = 72. Permutations A permutation is an ordering of a collection of objects. For example, there are six permutations of the letters A, B, C: ABC, ACB, BAC, BCA, CAB, and CBA. With only three objects, it is easy to determine the number of permutations just by listing them all. But with a large number of objects this would not be feasible. The fundamental principle of counting can be used to determine the number of permutations of any set of objects. For example, we can determine the number of permutations of a set of three objects as follows. There are 3 choices for the object to place first. After that choice is made, there are 2 choices remaining for the object to place second. Then there is 1 choice left for the object to place last. Therefore, the total number of ways to order three objects is (3)(2)(1) = 6. This reasoning can be generalized. The number of permutations of a collection of n objects is the product of the integers from 1 to n. This product can be written with the symbol n!, read “n factorial.” Page 64 Definition For any positive integer n, n! = n(n − 1)(n − 2) ⋯ (3)(2)(1). Also, we define 0! = 1. The number of permutations of n objects is n!. Example 2.12 Five people stand in line at a movie theater. Into how many different orders can they be arranged? Solution The number of permutations of a collection of five people is 5! = (5)(4)(3)(2)(1) = 120. Sometimes we are interested in counting the number of permutations of subsets of a certain size chosen from a larger set. This is illustrated in Example 2.13. Example 2.13 Five lifeguards are available for duty one Saturday afternoon. There are three lifeguard stations. In how many ways can three lifeguards be chosen and ordered among the stations? Solution We use the fundamental principle of counting. There are 5 ways to choose a lifeguard to occupy the first station, then 4 ways to choose a lifeguard to occupy the second station, and finally 3 ways to choose a lifeguard to occupy the third station. The total number of permutations of three lifeguards chosen from 5 is therefore (5)(4)(3) = 60. The reasoning used to solve Example 2.13 can be generalized. The number of permutations of k objects chosen from a group of n objects is This expression can be simplified by using factorial notation: Page 65 Summary The number of permutations of k objects chosen from a group of n objects is Combinations In some cases, when choosing a set of objects from a larger set, we don’t care about the ordering of the chosen objects; we care only which objects are chosen. For example, we may not care which lifeguards occupy which stations; we might care only which three lifeguards are chosen. Each distinct group of objects that can be selected, without regard to order, is called a combination. We will now show how to determine the number of combinations of k objects chosen from a set of n objects. We will illustrate the reasoning with the result of Example 2.13. In that example, we showed that there are 60 permutations of three objects chosen from five. Denoting the objects A, B, C, D, E, Figure 2.4 presents a list of all 60 permutations. FIGURE 2.4 The 60 permutations of three objects chosen from five. The 60 permutations in Figure 2.4 are arranged in 10 columns of 6 permutations each. Within each column, the three objects are the same, and the column contains the six different permutations of those three objects. Therefore, each column represents a distinct combination of three objects chosen from five, and there are 10 such combinations. Figure 2.4 thus shows that the number of combinations of three objects chosen from five can be found by dividing the number of permutations of three objects chosen from five, which is 5!/(5 − 3)!, by the number of permutations of three objects, which is 3! In summary, the number of combinations of three objects chosen from five is . The number of combinations of k objects chosen from n is often denoted by the symbol . The reasoning used to derive the number of combinations of three objects chosen from five can be generalized to derive an expression for . Page 66 Summary The number of combinations of k objects chosen from a group of n objects is (2.12) Example 2.14 At a certain event, 30 people attend, and 5 will be chosen at random to receive door prizes. The prizes are all the same, so the order in which the people are chosen does not matter. How many different groups of five people can be chosen? Solution Since the order of the five chosen people does not matter, we need to compute the number of combinations of 5 chosen from 30. This is Choosing a combination of k objects from a set of n divides the n objects into two subsets: the k that were chosen and the n − k that were not chosen. Sometimes a set is to be divided up into more than two subsets. For example, assume that in a class of 12 students, a project is assigned in which the students will work in groups. Three groups are to be formed, consisting of five, four, and three students. We can calculate the number of ways in which the groups can be formed as follows. We consider the process of dividing the class into three groups as a sequence of two operations. The first operation is to select a combination of 5 students to make up the group of 5. The second operation is to select a combination of 4 students from the remaining 7 to make up the group of 4. The group of 3 will then automatically consist of the 3 students who are left. The number of ways to perform the first operation is After the first operation has been performed, the number of ways to perform the second operation is The total number of ways to perform the sequence of two operations is therefore Notice that the numerator in the final answer is the factorial of the total group size, while the Page 67 while the denominator is the product of the factorials of the sizes of the groups chosen from it. This holds in general. Summary The number of ways of dividing a group of n objects into groups of k1,...,kr objects, where k1 + ⋯ + kr = n, is (2.13) Example 2.15 Twenty lightbulbs are to be arranged on a string. Six of the bulbs are red, four are blue, seven are green, and three are white. In how many ways can the bulbs be arranged? Solution There are 20 sockets on the string. They must be divided into four groups, with one group for each color of light bulb. The number of ways to divide 20 objects into four groups of the specified sizes is When a sample space consists of equally likely outcomes, the probability of an event can be found by dividing the number of outcomes in the event by the total number of outcomes in the sample space. Counting rules can sometimes be used to compute these numbers, as the following example illustrates: Example 2.16 A box of bolts contains 8 thick bolts, 5 medium bolts, and 3 thin bolts. A box of nuts contains 6 that fit the thick bolts, 4 that fit the medium bolts, and 2 that fit the thin bolts. One bolt and one nut are chosen at random. What is the probability that the nut fits the bolt? Solution The sample space consists of all pairs of nuts and bolts, and each pair is equally likely to be chosen. The event that the nut fits the bolt consists of all matching pairs of nuts and bolts. Therefore There are 6 + 4 + 2 = 12 nuts, and 8 + 5 + 3 = 16 bolts. Therefore The number of matching pairs is found by summing the number of pairs of thick nuts Page 68 and bolts, the number of pairs of medium nuts and bolts, and the number of pairs of thin nuts and bolts. These numbers are Therefore Exercises for Section 2.2 1. 2. 3. 4. 5. 6. 7. 8. DNA molecules consist of chemically linked sequences of the bases adenine, guanine, cytosine, and thymine, denoted A, G, C, and T. A sequence of three bases is called a codon. A base may appear more than once in a codon. a. How many different codons are there? b. The bases A and G are purines, while C and T are pyrimidines. How many codons are there whose first and third bases are purines and whose second base is a pyrimidine? c. How many codons consist of three different bases? A metallurgist is designing an experiment to determine the effect of flux, base metal, and energy input on the hardness of a weld. She wants to study four different fluxes, two different base metals, and three different amounts of energy input. If each run of the experiment involves a choice of one flux, one base metal, and one amount of energy input, how many different runs are possible? The article “Improved Bioequivalence Assessment of Topical Dermatological Drug Products Using Dermatopharmacokinetics” (B. N’Dri-Stempfer, W. Navidi, et al., Pharmaceutical Research, 2009:316–328) describes a study in which a new type of ointment was applied to forearms of volunteers to study the rates of absorption into the skin. Eight locations on the forearm were designated for ointment application. The new ointment was applied to four locations, and a control was applied to the other four. How many different ways were there to choose the four locations to apply the new ointment? A group of 18 people have gotten together to play baseball. They will divide themselves into two teams of 9 players each, with one team wearing green uniforms and the other wearing yellow uniforms. In how many ways can this be done? In horse racing, one can make a trifecta bet by specifying which horse will come in first, which will come in second, and which will come in third, in the correct order. One can make a box trifecta bet by specifying which three horses will come in first, second, and third, without specifying the order. a. In an eight-horse field, how many different ways can one make a trifecta bet? b. In an eight-horse field, how many different ways can one make a box trifecta bet? A college math department consisting of 10 faculty members must choose a department head, an assistant department head, and a faculty senate representative. In how many ways can this be done? A test consists of 15 questions. Ten are true-false questions, and five are multiple-choice questions that have four choices each. A student must select an answer for each question. In how many ways can this be done? License plates in a certain state consist of three letters followed by three digits. a. How many different license plates can be made? b. How many license plates are there that contain neither the letter “Q” nor the digit “9”? c. A license plate is drawn at random. What is the probability that it contains neither the letter “Q” nor the digit “9”? Page 69 A computer password consists of eight characters. a. How many different passwords are possible if each character may be any lowercase letter or digit? b. How many different passwords are possible if each character may be any lowercase letter or digit, and at least one character must be a digit? c. A computer system requires that passwords contain at least one digit. If eight characters are generated at random, and each is equally likely to be any of the 26 letters or 10 digits, what is the probability that a valid password will be generated? 10. Joe, Megan, and Santana are salespeople. Their sales manager has 18 accounts and must assign six accounts to each of them. In how many ways can this be done? 11. One drawer in a dresser contains 8 blue socks and 6 white socks. A second drawer contains 4 blue socks and 2 white socks. One sock is chosen from each drawer. What is the probability that they match? 12. A drawer contains 6 red socks, 4 green socks, and 2 black socks. Two socks are chosen at random. What is the probability that they match? 9. 2.3 Conditional Probability and Independence A sample space contains all the possible outcomes of an experiment. Sometimes we obtain some additional information about an experiment that tells us that the outcome comes from a certain part of the sample space. In this case, the probability of an event is based on the outcomes in that part of the sample space. A probability that is based on a part of a sample space is called a conditional probability. We explore this idea through some examples. In Example 2.6 (in Section 2.1) we discussed a population of 1000 aluminum rods. For each rod, the length is classified as too short, too long, or OK, and the diameter is classified as too thin, too thick, or OK. These 1000 rods form a sample space in which each rod is equally likely to be sampled. The number of rods in each category is presented in Table 2.1. Of the 1000 rods, 928 meet the diameter specification. Therefore, if a rod is sampled, P(diameter OK) = 928/1000 = 0.928. This probability is called the unconditional probability, since it is based on the entire sample space. Now assume that a rod is sampled, and its length is measured and found to meet the specification. What is the probability that the diameter also meets the specification? The key to computing this probability is to realize that knowledge that the length meets the specification reduces the sample space from which the rod is drawn. Table 2.2 (page 70) presents this idea. Once we know that the length specification is met, we know that the rod will be one of the 942 rods in the sample space presented in Table 2.2. TABLE 2.1 Length Too Short OK Too Long Sample space containing 1000 aluminum rods Too Thin 10 38 2 Diameter OK 3 900 25 Too Thick 5 4 13 Page 70 TABLE 2.2 Reduced sample space containing 942 aluminum rods that meet the length specification Length Too Short OK Too Long Too Thin — 38 — Diameter OK — 900 — Too Thick — 4 — Of the 942 rods in this sample space, 900 of them meet the diameter specification. Therefore, if we know that the rod meets the length specification, the probability that the rod meets the diameter specification is 900/942. We say that the conditional probability that the rod meets the diameter specification given that it meets the length specification is equal to 900/942, and we write P(diameter OK | length OK) = 900/942 = 0.955. Note that the conditional probability P(diameter OK | length OK) differs from the unconditional probability P(diameter OK), which was computed from the full sample space (Table 2.1) to be 0.928. Example 2.17 Compute the conditional probability P(diameter OK | length too long). Is this the same as the unconditional probability P(diameter OK)? Solution The conditional probability P(diameter OK | length too long) is computed under the assumption that the rod is too long. This reduces the sample space to the 40 items indicated in boldface in the following table. Length Too Short OK Too Long Too Thin 10 38 2 Diameter OK 3 900 25 Too Thick 5 4 13 Of the 40 outcomes, 25 meet the diameter specification. Therefore The unconditional probability P(diameter OK) is computed on the basis of all 1000 outcomes in the sample space and is equal to 928/1000 = 0.928. In this case, the conditional probability differs from the unconditional probability. Let’s look at the solution to Example 2.17 more closely. We found that In the answer 25/40, the denominator, 40, represents the number of outcomes that satisfy the condition that the rod is too long, while the numerator, 25, represents the number of Page 71 outcomes that satisfy both the condition that the rod is too long and that its diameter is OK. If we divide both the numerator and denominator of this answer by the number of outcomes in the full sample space, which is 1000, we obtain Now 40/1000 represents the probability of satisfying the condition that the rod is too long. That is, The quantity 25/1000 represents the probability of satisfying both the condition that the rod is too long and that the diameter is OK. That is, We can now express the conditional probability as This reasoning can be extended to construct a definition of conditional probability that holds for any sample space: Definition Let A and B be events with P(B) ≠ 0. The conditional probability of A given B is (2.14) Figure 2.5 presents Venn diagrams to illustrate the idea of conditional probability. FIGURE 2.5 (a) The diagram represents the unconditional probability P(A). P(A) is illustrated by considering the event A in proportion to the entire sample space, which is represented by the rectangle. (b) The diagram represents the conditional probability P(A|B). Since the event B is known to occur, the event B now becomes the sample space. For the event A to occur, the outcome must be in the intersection A ∩ B.The conditional probability P(A|B) is therefore illustrated by considering the intersection A ∩ B in proportion to the entire event B. Page 72 Example 2.18 Refer to Example 2.8 (in Section 2.1). What is the probability that a can will have a flaw on the side, given that it has a flaw on top? Solution We are given that P(flaw on top) = 0.03, and P(flaw on side and flaw on top) = 0.01. Using Equation (2.14), Example 2.19 Refer to Example 2.8 (in Section 2.1). What is the probability that a can will have a flaw on the top, given that it has a flaw on the side? Solution We are given that P(flaw on side) = 0.02, and P(flaw on side and flaw on top) = 0.01. Using Equation (2.14), The results of Examples 2.18 and 2.19 show that in most cases, P(A|B) ≠ P(B|A). Independent Events Sometimes the knowledge that one event has occurred does not change the probability that another event occurs. In this case the conditional and unconditional probabilities are the same, and the events are said to be independent. We present an example. Example 2.20 If an aluminum rod is sampled from the sample space presented in Table 2.1, find P(too long) and P(too long | too thin). Are these probabilities different? Page 73 Solution The conditional probability and the unconditional probability are the same. The information that the rod is too thin does not change the probability that the rod is too long. Example 2.20 shows that knowledge that an event occurs sometimes does not change the probability that another event occurs. In these cases, the two events are said to be independent. The event that a rod is too long and the event that a rod is too thin are independent. We now give a more precise definition of the term, both in words and in symbols. Definition Two events A and B are independent if the probability of each event remains the same whether or not the other occurs. In symbols: If P(A) ≠ 0 and P(B) ≠ 0, then A and B are independent if (2.15) If either P(A) = 0 or P(B) = 0, then A and B are independent. If A and B are independent, then the following pairs of events are also independent: A and Bc, Ac and B, and Ac and Bc. The proof of this fact is left as an exercise. The concept of independence can be extended to more than two events: Definition Events A1,A2,...,An are independent if the probability of each remains the same no matter which of the others occur. In symbols: Events A1,A2,...,An are independent if for each Ai, and each collection Aj1,...,Ajm of events with P(Aj1 ∩ ⋯ ∩ Ajm) ≠ 0, (2.16) The Multiplication Rule Page 74 Sometimes we know P(A|B) and we wish to find P(A ∩ B). We can obtain a result that is useful for this purpose by multiplying both sides of Equation (2.14) by P(B). This leads to the multiplication rule. If A and B are two events with P(B) ≠ 0, then (2.17) If A and B are two events with P(A) ≠ 0, then (2.18) If P(A) ≠ 0 and P(B) ≠ 0, then Equations (2.17) and (2.18) both hold. When two events are independent, then P(A|B) = P(A) and P(B|A) = P(B), so the multiplication rule simplifies: If A and B are independent events, then (2.19) This result can be extended to any number of events. If A1,A2,...,An are independent events, then for each collection Aj1,...,Ajm of events (2.20) In particular, (2.21) Example 2.21 A vehicle contains two engines, a main engine and a backup. The engine component fails only if both engines fail. The probability that the main engine fails is 0.05, and the probability that the backup engine fails is 0.10. Assume that the main and backup engines function independently. What is the probability that the engine component fails? Solution The probability that the engine component fails is the probability that both engines fail. Therefore Since the engines function independently, we may use Equation (2.19): Example Page 75 2.22 A system contains two components, A and B. Both components must function for the system to work. The probability that component A fails is 0.08, and the probability that component B fails is 0.05. Assume the two components function independently. What is the probability that the system functions? Solution The probability that the system functions is the probability that both components function. Therefore Since the components function independently, Example 2.23 Tests are performed in which structures consisting of concrete columns welded to steel beams are loaded until failure. Eighty percent of the failures occur in the beam, while 20% occur in the weld. Five structures are tested. What is the probability that all five failures occur in the beam? Solution For i = 1,..., 5, let Ai denote the event that the ith failure is a beam failure. Then Example 2.24 In Example 2.23, what is the probability that at least one failure occurs in the beam? Solution The easiest way to solve the problem is to notice that Now, letting Di denote the event that the ith failure is a weld failure, Page 76 Therefore P(at least one beam failure) = 1 − 0.0003 = 0.9997. Equations (2.19) and (2.20) tell us how to compute probabilities when we know that events are independent, but they are usually not much help when it comes to deciding whether two events really are independent. In most cases, the best way to determine whether events are independent is through an understanding of the process that produces the events. Here are a few illustrations: ■ A die is rolled twice. It is reasonable to believe that the outcome of the second roll is not affected by the outcome of the first roll. Therefore, knowing the outcome of the first roll does not help to predict the outcome of the second roll. The two rolls are independent. ■ A certain chemical reaction is run twice, using different equipment each time. It is reasonable to believe that the yield of one reaction will not affect the yield of the other. In this case the yields are independent. ■ A chemical reaction is run twice in succession, using the same equipment. In this case, it might not be wise to assume that the yields are independent. For example, a low yield on the first run might indicate that there is more residue than usual left behind. This might tend to make the yield on the next run higher. Thus knowing the yield on the first run could help to predict the yield on the second run. ■ The items in a simple random sample may be treated as independent, unless the population is finite and the sample comprises more than about 5% of the population (see the discussion of independence in Section 1.1). The Law of Total Probability The law of total probability is illustrated in Figure 2.6. A sample space contains the events A1,A2,A3, and A4. These events are mutually exclusive, since no two overlap. They are also exhaustive, which means that their union covers the whole sample space. Each outcome in the sample space belongs to one and only one of the events A1,A2,A3,A4. The event B can be any event. In Figure 2.6, each of the events Ai intersects B, forming the events A1 ∩ B, A2 ∩ B, A3 ∩ B, and A4 ∩ B. It is clear from Figure 2.6 that the events A1 ∩ B, A2 ∩ B, A3 ∩ B, and A4 ∩ B are mutually exclusive and that they cover B. Every outcome in B belongs to one and only one of the events A1 ∩ B, A2 ∩ B, A3 ∩ B, A4 ∩ B. It follows that Page 77 FIGURE 2.6 The mutually exclusive and exhaustive events A1,A2,A3,A4 divide the event B into mutually exclusive subsets. which is a union of mutually exclusive events. Therefore, by Axiom 3 on page 53, we have Since P(Ai ∩ B) = P(B|Ai)P(Ai), (2.22) Equation (2.22) is a special case of the law of total probability, restricted to the case where there are four mutually exclusive and exhaustive events. The intuition behind the law of total probability is quite simple. The events A1,A2,A3,A4 break the event B into pieces. The probability of B is found by adding up the probabilities of the pieces. We could redraw Figure 2.6 to have any number of events Ai. This leads to the general case of the law of total probability. Law of Total Probability If A1,...,An are mutually exclusive and exhaustive events, and B is any event, then (2.23) Equivalently, if P(Ai) ≠ 0 for each Ai, (2.24) Example 2.25 Customers who purchase a certain make of car can order an engine in any of three sizes. Of all cars sold, 45% have the smallest engine, 35% have the medium-sized one, and 20% have the largest. Of cars with the smallest engine, 10% fail an emissions test within two years of purchase, while 12% of those with the medium size and 15% of those with the largest engine fail. What is the probability that a randomly chosen car will fail an emissions test within two years? Page 78 Solution Let B denote the event that a car fails an emissions test within two years. Let A1 denote the event that a car has a small engine, A2 the event that a car has a medium-size engine, and A3 the event that a car has a large engine. Then The probability that a car will fail a test, given that it has a small engine, is 0.10. That is, P(B|A1) = 0.10. Similarly, P(B|A2) = 0.12, and P(B|A3) = 0.15. By the law of total probability (Equation 2.24), Sometimes problems like Example 2.25 are solved with the use of tree diagrams. Figure 2.7 presents a tree diagram for Example 2.25. There are three primary branches on the tree, corresponding to the three engine sizes. The probabilities of the engine sizes are listed on their respective branches. At the end of each primary branch are two secondary branches, representing the events of failure and no failure. The conditional probabilities of failure and no Page 79 failure, given engine size, are listed on the secondary branches. By multiplying along each of the branches corresponding to the event B = fail, we obtain the probabilities P(B|Ai)P(Ai). Summing these probabilities yields P(B), as desired. FIGURE 2.7 Tree diagram for the solution to Example 2.25. Bayes’ Rule If A and B are two events, we have seen that in most cases P(A|B) ≠ P(B|A). Bayes’ rule provides a formula that allows us to calculate one of the conditional probabilities if we know the other one. To see how it works, assume that we know P(B|A) and we wish to calculate P(A|B). Start with the definition of conditional probability (Equation 2.14): Now use Equation (2.18) to substitute P(B|A)P(A) for P(A ∩ B): (2.25) Equation (2.25) is essentially Bayes’ rule. When Bayes’ rule is written, the expression P(B) in the denominator is usually replaced with a more complicated expression derived from the law of total probability. Specifically, since the events A and Ac are mutually exclusive and exhaustive, the law of total probability shows that (2.26) Substituting the right-hand side of Equation (2.26) for P(B) in Equation (2.25) yields Bayes’ rule. A more general version of Bayes’ rule can be derived as well, by considering a collection A1,...,An of mutually exclusive and exhaustive events and using the law of total probability to replace P(B) with the expression on the right-hand side of Equation (2.24). Bayes’ Rule Special Case: Let A and B be events with P(A) ≠ 0, P(Ac) ≠ 0, and P(B) ≠ 0. Then (2.27) General Case: Let A1,...,An be mutually exclusive and exhaustive events with P(Ai) ≠ 0 for each Ai. Let B be any event with P(B) ≠ 0. Then (2.28) Example 2.26 shows how Bayes’ rule can be used to discover an important and surprising result in the field of medical testing. Page 80 Example 2.26 The proportion of people in a given community who have a certain disease is 0.005. A test is available to diagnose the disease. If a person has the disease, the probability that the test will produce a positive signal is 0.99. If a person does not have the disease, the probability that the test will produce a positive signal is 0.01. If a person tests positive, what is the probability that the person actually has the disease? Solution Let D represent the event that the person actually has the disease, and let + represent the event that the test gives a positive signal. We wish to find P(D|+). We are given the following probabilities: Using Bayes’ rule (Equation 2.27), In Example 2.26, only about a third of the people who test positive for the disease actually have the disease. Note that the test is fairly accurate; it correctly classifies 99% of both diseased and nondiseased individuals. The reason that a large proportion of those who test positive are actually disease-free is that the disease is rare—only 0.5% of the population has it. Because many diseases are rare, it is the case for many medical tests that most positives are false positives, even when the test is fairly accurate. For this reason, when a test comes out positive, a second test is usually given before a firm diagnosis is made. Example 2.27 Refer to Example 2.25. A record for a failed emissions test is chosen at random. What is the probability that it is for a car with a small engine? Solution Let B denote the event that a car failed an emissions test. Let A1 denote the event that a car has a small engine, A2 the event that a car has a medium-size engine, and A3 the event that a car has a large engine. We wish to find P(A1|B). The following probabilities are given in Example 2.25: By Bayes’ rule, Page 81 Application to Reliability Analysis Reliability analysis is the branch of engineering concerned with estimating the failure rates of systems. While some problems in reliability analysis require advanced mathematical methods, there are many problems that can be solved with the methods we have learned so far. We begin with an example illustrating the computation of the reliability of a system consisting of two components connected in series. Example 2.28 A system contains two components, A and B, connected in series as shown in the following diagram. The system will function only if both components function. The probability that A functions is given by P(A) = 0.98, and the probability that B functions is given by P(B) = 0.95. Assume that A and B function independently. Find the probability that the system functions. Solution Since the system will function only if both components function, it follows that Example 2.29 illustrates the computation of the reliability of a system consisting of two components connected in parallel. Example 2.29 A system contains two components, C and D, connected in parallel as shown in the following diagram. Page 82 The system will function if either C or D functions. The probability that C functions is 0.90, and the probability that D functions is 0.85. Assume C and D function independently. Find the probability that the system functions. Solution Since the system will function so long as either of the two components functions, it follows that The reliability of more complex systems can often be determined by decomposing the system into a series of subsystems, each of which contains components connected either in series or in parallel. Example 2.30 illustrates the method. Example 2.30 The thesis “Dynamic, Single-stage, Multiperiod, Capacitated Production Sequencing Problem with Multiple Parallel Resources” (D. Ott, M.S. thesis, Colorado School of Mines, 1998) describes a production method used in the manufacture of aluminum cans. The following schematic diagram, slightly simplified, depicts the process. The initial input into the process consists of coiled aluminum sheets, approximately 0.25 mm thick. In a process known as “cupping,” these sheets are uncoiled and shaped into can bodies, which are cylinders that are closed on the bottom and open on top. These can bodies are then washed and sent to the printer, which prints the label on the can. In practice there are several printers on a line; the diagram presents a line with three printers. The printer deposits the cans onto pallets, which are wooden structures that hold 7140 cans each. The cans next go to be filled. Some fill lines can accept cans directly from the pallets, but others can accept Page 83 them only from cell bins, which are large containers holding approximately 100,000 cans each. To use these fill lines, the cans must be transported from the pallets to cell bins, in a process called depalletizing. In practice there are several fill lines; the diagram presents a case where there are two fill lines, one of which will accept cans from the pallets, and the other of which will not. In the filling process the cans are filled, and the can top is seamed on. The cans are then packaged and shipped to distributors. It is desired to estimate the probability that the process will function for one day without failing. Assume that the cupping process has probability 0.995 of functioning successfully for one day. Since this component is denoted by “A” in the diagram, we will express this probability as P(A) = 0.995. Assume that the other process components have the following probabilities of functioning successfully during a one-day period: P(B) = 0.99, P(C) = P(D) = P(E) = 0.95, P(F) = 0.90, P(G) = 0.90, P(H) = 0.98. Assume the components function independently. Find the probability that the process functions successfully for one day. Solution We can solve this problem by noting that the entire process can be broken down into subsystems, each of which consists of simple series or parallel component systems. Specifically, subsystem 1 consists of the cupping and washing components, which are connected in series. Subsystem 2 consists of the printers, which are connected in parallel. Subsystem 3 consists of the fill lines, which are connected in parallel, with one of the two lines consisting of two components connected in series. We compute the probabilities of successful functioning for each subsystem, denoting the probabilities p1, p2, and p3. Subsystem 3 functions if F functions, or if both G and H function. Therefore The entire process consists of the three subsystems connected in series. Therefore, Page 84 for the process to function, all three subsystems must function. We conclude that We remark that the assumption that the components function independently is crucial in the solutions of Examples 2.28, 2.29, and 2.30. When this assumption is not met, it can be very difficult to make accurate reliability estimates. If the assumption of independence is used without justification, reliability estimates may be misleading. Exercises for Section 2.3 1. 2. 3. 4. Let A and B be events with P(A) = 0.8and P(A ∩ B) = 0.2. For what value of P(B) will A and B be independent? Let A and B be events with P(A) = 0.5 and P(A∩Bc) = 0.4. For what value of P(B) will A and B be independent? A box contains 15 resistors. Ten of them are labeled 50 Ω and the other five are labeled 100 Ω. a. What is the probability that the first resistor is 100 Ω? b. What is the probability that the second resistor is 100 Ω, given that the first resistor is 50 Ω? c. What is the probability that the second resistor is 100 Ω, given that the first resistor is 100 Ω? Refer to Exercise 3. Resistors are randomly selected from the box, one by one, until a 100 Ω 5. 6. 7. 8. resistor is selected. a. What is the probability that the first two resistors are both 50 Ω? b. What is the probability that a total of two resistors are selected from the box? c. What is the probability that more than three resistors are selected from the box? On graduation day at a large university, one graduate is selected at random. Let A represent the event that the student is an engineering major, and let B represent the event that the student took a calculus course in college. Which probability is greater, P(A|B) or P(B|A)? Explain. The article “Integrating Risk Assessment and Life Cycle Assessment: A Case Study of Insulation” (Y. Nishioka, J. Levy, et al., Risk Analysis, 2002: 1003–1017) estimates that 5.6% of a certain population has asthma, and that an asthmatic has probability 0.027 of suffering an asthma attack on a given day. A person is chosen at random from this population. What is the probability that this person has an asthma attack on that day? Suppose that 90% of bolts and 85% of nails meet specifications. One bolt and one nail are chosen independently. a. What is the probability that both meet specifications? b. What is the probability that neither meets specifications? c. What is the probability that at least one of them meets specifications? A drag racer has two parachutes, a main and a backup, that are designed to bring the vehicle to a stop after the end of a run. Suppose that the main chute deploys with probability 0.99, and that if the main fails to deploy, the backup deploys with probability 0.98. a. What is the probability that one of the two parachutes deploys? Page 85 b. What is the probability that the backup parachute deploys? 9. At a certain car dealership, 20% of customers who bought a new vehicle bought an SUV, and 3% of them bought a black SUV. Given that a customer bought an SUV, what is the probability that it was black? 10. At a certain college, 30% of the students major in engineering, 20% play club sports, and 10% both major in engineering and play club sports. A student is selected at random. a. What is the probability that the student is majoring in engineering? b. What is the probability that the student plays club sports? c. Given that the student is majoring in engineering, what is the probability that the student plays club sports? d. Given that the student plays club sports, what is the probability that the student is majoring in engineering? e. Given that the student is majoring in engineering, what is the probability that the student does not play club sports? f. Given that the student plays club sports, what is the probability that the student is not majoring in engineering? 11. In the process of producing engine valves, the valves are subjected to a first grind. Valves whose thicknesses are within the specification are ready for installation. Those valves whose thicknesses are above the specification are reground, while those whose thicknesses are below the specification are scrapped. Assume that after the first grind, 70% of the valves 12. 13. 14. 15. meet the specification, 20% are reground, and 10% are scrapped. Furthermore, assume that of those valves that are reground, 90% meet the specification, and 10% are scrapped. a. Find the probability that a valve is ground only once. b. Given that a valve is not reground, what is the probability that it is scrapped? c. Find the probability that a valve is scrapped. d. Given that a valve is scrapped, what is the probability that it was ground twice? e. Find the probability that the valve meets the specification (after either the first or second grind). f. Given that a valve meets the specification (after either the first or second grind), what is the probability that it was ground twice? g. Given that a valve meets the specification, what is the probability that it was ground only once? Sarah and Thomas are going bowling. The probability that Sarah scores more than 175 is 0.4, and the probability that Thomas scores more than 175 is 0.2. Their scores are independent. a. Find the probability that both score more than 175. b. Given that Thomas scores more than 175, the probability that Sarah scores higher than Thomas is 0.3. Find the probability that Thomas scores more than 175 and Sarah scores higher than Thomas. A particular automatic sprinkler system has two different types of activation devices for each sprinkler head. One type has a reliability of 0.9; that is, the probability that it will activate the sprinkler when it should is 0.9. The other type, which operates independently of the first type, has a reliability of 0.8. If either device is triggered, the sprinkler will activate. Suppose a fire starts near a sprinkler head. a. What is the probability that the sprinkler head will be activated? b. What is the probability that the sprinkler head will not be activated? c. What is the probability that both activation devices will work properly? d. What is the probability that only the device with reliability 0.9 will work properly? Laura and Philip each fire one shot at a target. Laura has probability 0.5 of hitting the target, and Philip has probability 0.3. The shots are independent. a. Find the probability that the target is hit. b. Find the probability that the target is hit by exactly one shot. c. Given that the target was hit by exactly one shot, find the probability that Laura hit the target. A population of 600 semiconductor wafers contains wafers from three lots. The wafers are categorized by lot and by whether they conform to a thickness specification. The Page 86 following table presents the number of wafers in each category. A wafer is chosen at random from the population. Lot A B Conforming 88 165 Nonconforming 12 35 C 260 40 a. If the wafer is from Lot A, what is the probability that it is conforming? b. If the wafer is conforming, what is the probability that it is from Lot A? c. If the wafer is conforming, what is the probability that it is not from Lot C? d. If the wafer is not from Lot C, what is the probability that it is conforming? 16. Refer to Exercise 15. Let E1 be the event that the wafer comes from Lot A, and let E2 be the event that the wafer is conforming. Are E1 and E2 independent? Explain. 17. A geneticist is studying two genes. Each gene can be either dominant or recessive. A sample of 100 individuals is categorized as follows. Gene 2 Gene 1 Dominant Recessive Dominant 56 14 Recessive 24 6 a. b. c. d. What is the probability that a randomly sampled individual, gene 1 is dominant? What is the probability that a randomly sampled individual, gene 2 is dominant? Given that gene 1 is dominant, what is the probability that gene 2 is dominant? These genes are said to be in linkage equilibrium if the event that gene 1 is dominant is independent of the event that gene 2 is dominant. Are these genes in linkage equilibrium? 18. A car dealer sold 750 automobiles last year. The following table categorizes the cars sold by size and color and presents the number of cars in each category. A car is to be chosen at random from the 750 for which the owner will win a lifetime of free oil changes. Size Small Midsize Large White 102 86 26 Color Black Red 71 33 63 36 32 22 Grey 134 105 40 a. If the car is small, what is the probability that it is black? b. If the car is white, what is the probability that it is midsize? c. If the car is large, what is the probability that it is red? d. If the car is red, what is the probability that it is large? e. If the car is not small, what is the probability that it is not grey? 19. The following table presents the 100 senators of the 115th U.S. Congress on January 3, 2017, classified by political party affiliation and gender. Democrat Male 30 Female 16 Republican Independent 47 2 5 0 A senator is selected at random from this group. Compute the following probabilities. a. The senator is a male Republican. b. The senator is a Democrat or a female. c. The senator is a Republican. d. The senator is not a Republican. e. The senator is a Democrat. f. The senator is an Independent. g. The senator is a Democrat or an Independent. 20. An automobile insurance company divides customers into three categories, good risks, medium risks, and poor risks. Assume that 70% of the customers are good risks, Page 87 20% are medium risks, and 10% are poor risks. Assume that during the course of a year, a good risk customer has probability 0.005 of filing an accident claim, a medium risk customer has probability 0.01, and a poor risk customer has probability 0.025. A customer is chosen at random. a. What is the probability that the customer is a good risk and has filed a claim? b. What is the probability that the customer has filed a claim? c. Given that the customer has filed a claim, what is the probability that the customer is a good risk? 21. Nuclear power plants have redundant components in important systems to reduce the chance of catastrophic failure. Assume that a plant has two gauges to measure the level of coolant in the reactor core and that each gauge has probability 0.01 of failing. Assume that one potential cause of gauge failure is that the electric cables leading from the core to the control room where the gauges are located may burn up in a fire. Someone wishes to estimate the probability that both gauges fail, and makes the following calculation: a. What assumption is being made in this calculation? b. Explain why this assumption is probably not justified in the present case. c. Is the probability of 0.0001 likely to be too high or too low? Explain. 22. Refer to Exercise 21. Is it possible for the probability that both gauges fail to be greater than 0.01? Explain. 23. A lot of 10 components contains 3 that are defective. Two components are drawn at random and tested. Let A be the event that the first component drawn is defective, and let B be the event that the second component drawn is defective. a. Find P(A). b. Find P(B|A). 24. 25. 26. 27. c. Find P(A ∩ B). d. Find P(Ac ∩ B). e. Find P(B). f. Are A and B independent? Explain. A lot of 1000 components contains 300 that are defective. Two components are drawn at random and tested. Let A be the event that the first component drawn is defective, and let B be the event that the second component drawn is defective. a. Find P(A). b. Find P(B|A). c. Find P(A ∩ B). d. Find P(Ac ∩ B). e. Find P(B). f. Find P(A|B). g. Are A and B independent? Is it reasonable to treat A and B as though they were independent? Explain. In a lot of n components, 30% are defective. Two components are drawn at random and tested. Let A be the event that the first component drawn is defective, and let B be the event that the second component drawn is defective. For which lot size n will A and B be more nearly independent: n = 10 or n = 10,000? Explain. A certain delivery service offers both express and standard delivery. Seventy-five percent of parcels are sent by standard delivery, and 25% are sent by express. Of those sent standard, 80% arrive the next day, and of those sent express, 95% arrive the next day. A record of a parcel delivery is chosen at random from the company’s files. a. What is the probability that the parcel was shipped express and arrived the next day? b. What is the probability that it arrived the next day? c. Given that the package arrived the next day, what is the probability that it was sent express? Each day, a weather forecaster predicts whether or not it will rain. For 80% of rainy days, she correctly predicts that it will rain. For 90% of non-rainy days, she correctly predicts that it will not rain. Suppose that 10% of days are rainy and 90% are non-rainy. a. What proportion of the forecasts are correct? b. Another forecaster always predicts that there will be no rain. What proportion of these forecasts are correct? Page 88 28. Items are inspected for flaws by two quality inspectors. If a flaw is present, it will be detected by the first inspector with probability 0.9, and by the second inspector with probability 0.7. Assume the inspectors function independently. a. If an item has a flaw, what is the probability that it will be found by both inspectors? b. If an item has a flaw, what is the probability that it will be found by at least one of the two inspectors? c. Assume that the second inspector examines only those items that have been passed by the first inspector. If an item has a flaw, what is the probability that the second inspector will find it? 29. 30. 31. 32. 33. Refer to Exercise 28. Assume that both inspectors inspect every item and that if an item has no flaw, then neither inspector will detect a flaw. a. Assume that the probability that an item has a flaw is 0.10. If an item is passed by the first inspector, what is the probability that it actually has a flaw? b. Assume that the probability that an item has a flaw is 0.10. If an item is passed by both inspectors, what is the probability that it actually has a flaw? Refer to Example 2.26. a. If a man tests negative, what is the probability that he actually has the disease? b. For many medical tests, it is standard procedure to repeat the test when a positive signal is given. If repeated tests are independent, what is the probability that a man will test positive on two successive tests if he has the disease? c. Assuming repeated tests are independent, what is the probability that a man tests positive on two successive tests if he does not have the disease? Page 89 d. If a man tests positive on two successive tests, what is the probability that he has the disease? 34. A system consists of four components connected as shown in the following diagram: Assume A, B, C, and D function independently. If the probabilities that A, B, C, and D fail are 0.10, 0.05, 0.10, and 0.20, respectively, what is the probability that the system functions? 35. A system consists of four components, connected as shown in the diagram. Suppose that the components function independently, and that the probabilities of failure are 0.05 for A, 0.03 for B, 0.07 for C, and 0.14 for D. Find the probability that the system functions. 36. A system contains two components, A and B, connected in series, as shown in the diagram. Assume A and B function independently. For the system to function, both components must function. a. If the probability that A fails is 0.05, and the probability that B fails is 0.03, find the probability that the system functions. b. If both A and B have probability p of failing, what must the value of p be so that the probability that the system functions is 0.90? c. If three components are connected in series, and each has probability p of failing, what must the value of p be so that the probability that the system functions is 0.90? 37. A system contains two components, C and D, connected in parallel as shown in the diagram. Assume C and D function independently. For the system to function, either C or D must function. a. If the probability that C fails is 0.08 and the probability that D fails is 0.12, find the probability that the system functions. b. If both C and D have probability p of failing, what must the value of p be so that the probability that the system functions is 0.99? c. If three components are connected in parallel, function independently, and each has probability p of failing, what must the value of p be so that the probability that the system functions is 0.99? d. If components function independently, and each component has probability 0.5 of failing, what is the minimum number of components that must be connected in parallel so that the probability that the system functions is at least 0.99? 38. If A and B are independent events, prove that the following pairs of events are independent: Ac and B, A and Bc, and Ac and Bc. 39. A fair die is rolled twice. Let A be the event that the number on the first die is odd, let B be the event that the number on the second die is odd, and let C be the event that the sum of the two rolls is equal to 7. a. Show that A and B are independent, A and C are independent, and B and C are independent. This property is known as pairwise independence. b. Show that A, B, and C are not independent. Conclude that it is possible for a set of events to be pairwise independent but not be independent. (In this context, independence is sometimes referred to as mutual independence.) Page 90 2.4 Random Variables In many situations, it is desirable to assign a numerical value to each outcome of an experiment. Such an assignment is called a random variable. To make the idea clear, we present an example. Suppose that an electrical engineer has on hand six resistors. Three of them are labeled 10 Ω and the other three are labeled 20 Ω. The engineer wants to connect a 10 Ω resistor and a 20 Ω resistor in series, to create a resistance of 30 Ω. Now suppose that in fact the three resistors labeled 10 Ω have actual resistances of 9, 10, and 11 Ω, and that the three resistors labeled 20 Ω have actual resistances of 19, 20, and 21 Ω. The process of selecting one resistor of each type is an experiment whose sample space consists of nine equally likely outcomes. The sample space is presented in the following table. Outcome (9, 19) (9, 20) (9, 21) (10, 19) (10, 20) (10, 21) (11, 19) (11, 20) (11, 21) Probability 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 Now what is important to the engineer in this experiment is the sum of the two resistances, rather than their individual values. Therefore we assign to each outcome a number equal to the sum of the two resistances selected. This assignment, represented by the letter X, is presented in the following table. Outcome (9, 19) (9, 20) (9, 21) (10, 19) (10, 20) (10, 21) (11, 19) (11, 20) (11, 21) X 28 29 30 29 30 31 30 31 32 Probability 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 The function X, which assigns a numerical value to each outcome in the sample space, is a random variable. A random variable assigns a numerical value to each outcome in a sample space. It is customary to denote random variables with uppercase letters. The letters X, Y, and Z Page 91 are most often used. We can compute probabilities for random variables in an obvious way. In the example just presented, the event X = 29 corresponds to the event {(9, 20), (10, 19)} of the sample space. Therefore P(X = 29) = P({(9, 20), (10, 19)}) = 2/9. Example 2.31 List the possible values of the random variable X, and find the probability of each of them. Solution The possible values are 28, 29, 30, 31, and 32. To find the probability of one of these values, we add the probabilities of the outcomes in the sample space that correspond to the value. The results are given in the following table: x 28 29 30 31 32 P(X = x) 1/9 2/9 3/9 2/9 1/9 The table of probabilities in Example 2.31 contains all the information needed to compute any probability regarding the random variable X. Note that the outcomes of the sample space are not presented in the table. When the probabilities pertaining to a random variable are known, we usually do not think about the sample space; we just focus on the probabilities. There are two important types of random variables, discrete and continuous. A discrete random variable is one whose possible values form a discrete set; in other words, the values can be ordered, and there are gaps between adjacent values. The random variable X, just described, is discrete. In contrast, the possible values of a continuous random variable always contain an interval, that is, all the points between some two numbers. We will provide precise definitions of these types of random variables later in this section. We present some more examples of random variables. Example 2.32 Computer chips often contain surface imperfections. For a certain type of computer chip, 9% contain no imperfections, 22% contain 1 imperfection, 26% contain 2 imperfections, 20% contain 3 imperfections, 12% contain 4 imperfections, and the remaining 11% contain 5 imperfections. Let Y represent the number of imperfections in a randomly chosen chip. What are the possible values for Y? Is Y discrete or continuous? Find P(Y = y) for each possible value y. Page 92 Solution The possible values for Y are the integers 0, 1, 2, 3, 4, and 5. The random variable Y is discrete, because it takes on only integer values. Nine percent of the outcomes in the sample space are assigned the value 0. Therefore P(Y = 0) = 0.09. Similarly P(Y = 1) = 0.22, P(Y = 2) = 0.26, P(Y = 3) = 0.20, P(Y = 4) = 0.12, and P(Y = 5) = 0.11. Example 2.33 A certain type of magnetic disk must function in an environment where it is exposed to corrosive gases. It is known that 10% of all such disks have lifetimes less than or equal to 100hours, 50% have lifetimes greaterthan100hours but less than or equal to 500 hours, and 40% have lifetimes greater than 500 hours. Let Z represent the number of hours in the lifetime of a randomly chosen disk. Is Z continuous or discrete? Find P(Z ≤ 500). Can we compute all the probabilities for Z? Explain. Solution The lifetime of a component is not limited to a list of discretely spaced values; Z is continuous. Of all the components, 60% have lifetimes less than or equal to 500 hours. Therefore P(Z ≤ 500) = 0.60. We do not have enough information to compute all the probabilities for Z. We can compute some of them, for example, P(Z ≤ 100) = 0.10, P(100 < Z ≤ 500) = 0.50, and P(Z > 500) = 0.40. But we do not know, for example, the proportion of components that have lifetimes between 100 and 200 hours, or between 200 and 300 hours, so we cannot find the probability that the random variable Z falls into either of these intervals. To compute all the probabilities for Z, we would need to be able to compute the probability for every possible interval, for example, P(200 < Z ≤ 300), P(200 < Z ≤ 200.1), and so on. We will see how this can be done later in this section, when we discuss continuous random variables. Random Variables and Populations It is often useful to think of a value of a random variable as having been sampled from a population. For example, consider the random variable Y described in Example 2.32. Observing a value for this random variable is like sampling a value from a population consisting of the integers 0, 1, 2, 3, 4, and 5 in the following proportions: 0s, 9%; 1s, 22%; 2s, 26%; 3s, 20%; 4s, 12%; and 5s, 11%. For a continuous random variable, it is appropriate to imagine an infinite population containing all the possible values of the random variable. For example, for the random variable Z in Example 2.33 we would imagine a population containing all the positive numbers, with 10% of the population values less than or equal to 100, 50% greater than 100 but less than or equal to 500, and 40% greater than 500. The proportion of population values in any interval would be equal to the probability that the variable Z is in that interval. Methods for working with random variables differ somewhat between discrete and continuous random variables. We begin with the discrete case. Page 93 Discrete Random Variables We begin by reviewing the definition of a discrete random variable. Definition A random variable is discrete if its possible values form a discrete set. This means that if the possible values are arranged in order, there is a gap between each value and the next one. The set of possible values may be infinite; for example, the set of all integers and the set of all positive integers are both discrete sets. It is common for the possible values of a discrete random variable to be a set of integers. For any discrete random variable, if we specify the list of its possible values along with the probability that the random variable takes on each of these values, then we have completely described the population from which the random variable is sampled. We illustrate with an example. The number of flaws in a 1-inch length of copper wire manufactured by a certain process varies from wire to wire. Overall, 48% of the wires produced have no flaws, 39% have one flaw, 12% have two flaws, and 1% have three flaws. Let X be the number of flaws in a randomly selected piece of wire. Then P(X = 0) = 0.48, P(X = 1) = 0.39, P(X = 2) = 0.12, and P(X = 3) = 0.01 The list of possible values 0, 1, 2, 3, along with the probabilities for each, provide a complete description of the population from which X is drawn. This description has a name—the probability mass function. Definition The probability mass function of a discrete random variable X is the function p(x) = P(X = x). The probability mass function is sometimes called the probability distribution. Thus for the random variable X representing the number of flaws in a length of wire, p(0) = 0.48, p(1) = 0.39, p(2) = 0.12, p(3) = 0.01, and p(x) = 0 for any value of x other than 0, 1, 2, or 3. Note that if the values of the probability mass function are added over all the possible values of X, the sum is equal to 1. This is true for any probability mass function. The reason is that summing the values of a probability mass function over all the possible values of the corresponding random variable produces the probability that the random variable is equal to one of its possible values, and this probability is always equal to 1. The probability mass function can be represented by a graph in which a vertical line is drawn at each of the possible values of the random variable. The heights of the lines are equal to the probabilities of the corresponding values. The physical interpretation of this graph is that each line represents a mass equal to its height. Figure 2.8 (page 94) presents a graph of the probability mass function of the random variable X. Page 94 FIGURE 2.8 Probability mass function of X, the number of flaws in a randomly chosen piece of wire. The Cumulative Distribution Function of a Discrete Random Variable The probability mass function specifies the probability that a random variable is equal to a given value. A function called the cumulative distribution function specifies the probability that a random variable is less than or equal to a given value. The cumulative distribution function of the random variable X is the function F(x) = P(X ≤ x). Example 2.34 Let F(x) denote the cumulative distribution function of the random variable X that represents the number of flaws in a randomly chosen wire. Find F(2). Find F(1.5). Solution Since F(2) = P(X ≤ 2), we need to find P(X ≤ 2). We do this by summing the probabilities for the values of X that are less than or equal to 2, namely, 0, 1, and 2. Thus Now F(1.5) = P(X ≥ 1.5). Therefore, to compute F(1.5) we must sum the probabilities for the values of X that are less than or equal to 1.5, which are 0 and 1. Thus In general, for any discrete random variable X, the cumulative distribution function F(x) can be computed by Page 95 F(x) can be computed by summing the probabilities of all the possible values of X that are less than or equal to x. Note that F(x) is defined for any number x, not just for the possible values of X. Summary Let X be a discrete random variable. Then ■ The probability mass function of X is the function p(x) = P(X = x). ■ The cumulative distribution function of X is the function F(x) = P(X ≤ x). ■ ■ Example 2.35 Plot the cumulative distribution function F(x) of the random variable X that represents the number of flaws in a randomly chosen wire. Solution First we compute F(x) for each of the possible values of X, which are 0, 1, 2, and 3. For any value x, we compute F(x) by summing the probabilities of all the possible values of X that are less than or equal to x. For example, if 1 ≤ x < 2, the possible values of X that are less than or equal to x are 0 and 1, so F(x) = P(X = 0) + P(X = 1) = 0.87. Therefore Page 96 A plot of F(x) is presented in the following figure. For a discrete random variable, the graph of F(x) consists of a series of horizontal lines (called "steps") with jumps at each of the possible values of X. Note that the size of the jump at any point x is equal to the value of the probability mass function p(x) = P(X = x). Mean and Variance for Discrete Random Variables The population mean of a discrete random variable can be thought of as the mean of a hypothetical sample that follows the probability distribution perfectly. To make this idea concrete, assume that the number of flaws in a wire, X, has the probability mass function given previously, with P(X = 0) = 0.48, P(X = 1) = 0.39, P(X = 2) = 0.12, and P(X = 3) = 0.01. Now imagine that we have a sample of 100 wires, and that the sample follows this distribution perfectly, so that exactly 48 of the wires have 0 flaws, 39 have 1 flaw, 12 have 2 flaws, and 1 has 3 flaws. The sample mean is the total number of flaws divided by 100: This can be rewritten as This shows that the mean of a perfect sample can be obtained by multiplying each possible value of X by its probability, and summing the products. This is the definition of the population mean of a discrete random variable. The population mean of a random variable X may also be called the expectation, or expected value, of X, and can be denoted by μX, by E(X), or simply by μ. Sometimes we will drop the word "population," and simply refer to the population mean as the mean. Page 97 Definition Let X be a discrete random variable with probability mass function p(x) = P(X = x). The mean of X is given by (2.29) where the sum is over all possible values of X. The mean of X is sometimes called the expectation, or expected value, of X and may also be denoted by E(X) or by μ. Example 2.36 A certain industrial process is brought down for recalibration whenever the quality of the items produced falls below specifications. Let X represent the number of times the process is recalibrated during a week, and assume that X has the following probability mass function. x p(x) 0 0.35 1 0.25 2 0.20 3 0.15 4 0.05 Find the mean of X. Solution Using Equation (2.29), we compute The population mean has an important physical interpretation. It is the horizontal component of the center of mass of the probability mass function, that is, it is the point on the horizontal axis at which the graph of the probability mass function would balance if supported there. Figure 2.9 illustrates this property for the probability mass function described in Example 2.36, where the population mean is μ = 1.30. FIGURE 2.9 The graph of a probability mass function will balance if supported at the population mean. We will describe the population variance by making Page 98 an analogy with the sample variance. The sample variance was discussed in Section 1.2. Recall that for a sample X1,...,Xn, the sample variance is given by . The sample variance is thus essentially the average of the squared differences between the sample points and the sample mean (except that we divide by n − 1 instead of n). By analogy, the population variance of a discrete random variable is a weighted average of the squared differences (x − μX)2 where x ranges through all the possible values of the random variable X. This weighted average is computed by multiplying each squared difference (x − μX)2 by the probability P(X = x) and summing the results. The population variance of a random variable X can be denoted by σ2X, by V(X), or simply by σ2. The population variance is given by the formula By performing some algebra, an alternate formula can be obtained. A derivation of the alternate formula is given at the end of this section. We also define the population standard deviation to be the square root of the population variance. We denote the population standard deviation of a random variable X by σX or simply by σ. As with the mean, we will sometimes drop the word "population," and simply refer to the population variance and population standard deviation as the variance and standard deviation, respectively. Summary Let X be a discrete random variable with probability mass function p(x) = P(X = x). Then ■ The variance of X is given by (2.30) ■ An alternate formula for the variance is given by (2.31) ■ ■ The variance of X may also be denoted by V(X) or by σ2. The standard deviation is the square root of the variance: . Page 99 Example 2.37 Find the variance and standard deviation for the random variable X described in Example 2.36, representing the number of times a process is recalibrated. Solution In Example 2.36 we computed the mean of X to be μX = 1.30. We compute the variance by using Equation (2.30): The standard deviation is . Example 2.38 Use the alternate formula, Equation (2.31), to compute the variance of X, the number of times a process is recalibrated. Solution In Example 2.36 the mean was computed to be μX = 1.30. The variance is therefore Example 2.39 A resistor in a certain circuit is specified to have a resistance in the range 99 Ω–101 Ω. An engineer obtains two resistors. The probability that both of them meet the specification is 0.36, the probability that exactly one of them meets the specification is 0.48, and the probability that neither of them meets the specification is 0.16. Let X represent the number of resistors that meet the specification. Find the probability mass function, and the mean, variance, and standard deviation of X. Solution The probability mass function is P(X = 0) = 0.16, P(X = 1) = 0.48, P(X = 2) = 0.36, and P(X = x) = 0 for x ≠ 0, 1, or 2. The mean is The variance is The standard deviation is Page 100 . To develop a physical interpretation for the population variance, imagine that each line in the graph of the probability mass function has mass proportional to its length, and that a solid rod is inserted vertically into the graph through the center of mass (population mean). Now imagine grasping the rod and twirling the graph around it. The more spread out the graph, the more difficult it would be to twirl. The physical quantity that measures the difficulty in twirling is the moment of inertia. For each line in the graph, the moment of inertia is equal to the mass of the line multiplied by the squared distance from the line to the center of mass, multiplied by the length of the line. The moment of inertia for the entire graph is the sum of the moments of the lines, which is the population variance. The Probability Histogram When the possible values of a discrete random variable are evenly spaced, the probability mass function can be represented by a histogram, with rectangles centered at the possible values of the random variable. The area of a rectangle centered at a value x is equal to P(X = x). Such a histogram is called a probability histogram, because the areas represent probabilities. Table 2.3 presents the probability mass function for a random variable X that represents the number of defects in a printed circuit board. Figure 2.10 presents the probability histogram for this random variable. The probability that a value of a random variable falls into a given interval is given by an area under the probability histogram. Example 2.40 illustrates the idea. TABLE 2.3 x 0 1 2 3 p(x) 0.45 0.35 0.15 0.05 FIGURE 2.10 Probability histogram for X, the number of defects in a printed circuit board. Page 101 Example 2.40 Find the probability that a randomly chosen printed circuit board has more than one defect. Indicate this probability as an area under the probability histogram. Solution We wish to find P(X > 1). Since no board has more than three defects, the proportion of wires that have more than one defect can be found by adding the proportion that have two defects to the proportion that have three defects. In symbols, P(X > 1) = P(X = 2) + P(X = 3) The probability mass function specifies that P(X = 2) = 0.15 and P(X = 3) = 0.05. Therefore P(X > 1) = 0.15 + 0.05 = 0.20. This probability is given by the area under the probability histogram corresponding to those rectangles centered at values greater than 1 (see Figure 2.11). There are two such rectangles; their areas are P(X = 2) = 0.15 and P(X = 3) = 0.05. This is another way to show that P(X > 1) = 0.15 + 0.05 = 0.20. FIGURE 2.11 Probability histogram for X, the number of defects in a printed circuit board. The area corresponding to values of X greater than 1 is shaded. This area is equal to P(X > 1). In Chapter 4 we will see that probabilities for discrete random variables can sometimes be approximated by computing the area under a curve. Representing the discrete probabilities with a probability histogram will make it easier to understand how this is done. Continuous Random Variables Figure 1.9 (in Section 1.3) presents a histogram for the emissions, in grams of particulates per gallon of fuel consumed, of a sample of 62 vehicles. Note that emissions is a continuous variable, because its possible values are not restricted to some discretely spaced set. The class intervals are chosen so that each interval contains a reasonably large number of vehicles. If the sample were larger, we could make the intervals narrower. In particular, if we had information on the entire population, containing millions of vehicles, we could make the intervals Page 102 extremely narrow. The histogram would then look quite smooth and could be approximated with a curve, which might look like Figure 2.12. FIGURE 2.12 The histogram for a large continuous population could be drawn with extremely narrow rectangles and might look like this curve. If a vehicle were chosen at random from this population to have its emissions measured, the emissions level X would be a random variable. The probability that X falls between any two values a and b is equal to the area under the histogram between a and b. Because the histogram in this case is represented by a curve, the probability would be found by computing an integral. The random variable X described here is an example of a continuous random variable. A continuous random variable is defined to be a random variable whose probabilities are represented by areas under a curve. This curve is called the probability density function. Because the probability density function is a curve, the computations of probabilities involve integrals, rather than the sums that are used in the discrete case. Definition A random variable is continuous if its probabilities are given by areas under a curve. The curve is called a probability density function for the random variable. The probability density function is sometimes called the probability distribution. Computing Probabilities with the Probability Density Function Let X be a continuous random variable. Let the function f(x) be the probability density function of X. Let a and b be any two numbers, with a < b. The proportion of the population whose values of X lie between a and b is given by , the area under the probability density function between a and b. This is the probability that the random variable X takes on a value between a and b. Note that the area under the curve does not depend on whether the endpoints a and b are included in the interval. Therefore probabilities involving X do not depend on whether endpoints are included. Page 103 Summary Let X be a continuous random variable with probability density function f(x). Let a and b be any two numbers, with a < b. Then (2.32) In addition, (2.33) (2.34) If f(x) is the probability density function of a random variable X, then the area under the entire curve from −∞ to ∞ is the probability that the value of X is between −∞ and ∞. This probability must be equal to 1, because the value of X is always between −∞ and ∞. Therefore the area under the entire curve f(x) is equal to 1. Summary Let X be a continuous random variable with probability density function f(x). Then (2.35) Example 2.41 A hole is drilled in a sheet-metal component, and then a shaft is inserted through the hole. The shaft clearance is equal to the difference between the radius of the hole and the radius of the shaft. Let the random variable X denote the clearance, in millimeters. The probability density function of X is Components with clearances larger than 0.8 mm must be scrapped. What proportion of components are scrapped? Solution Figure 2.13 (page 104) presents the probability density function of X. Note that the density f(x) is 0 for x ≤ 0 and for x ≥ 1. This indicates that the clearances are always between 0 and 1 mm. The proportion of components that must be scrapped is P(X > 0.8), which is equal to the area under the probability density function to the right of 0.8. Page 104 FIGURE 2.13 Graph of the probability density function of X, the clearance of a shaft. The area shaded is equal to P(X > 0.8). This area is given by The Cumulative Distribution Function of a Continuous Random Variable The cumulative distribution function of a continuous random variable X is F(x) = P(X ≤ x), just as it is for a discrete random variable. For a discrete random variable, F(x) can be found by summing values of the probability mass function. For a continuous random variable, the value of F(x) is obtained by integrating the probability density function. Since F(x) = P(X ≤ x), it follows from Equation (2.33) that , where f(t) is the probability density function. Definition Let X be a continuous random variable with probability density function f(x). The cumulative distribution function of X is the function (2.36) Page 105 Example 2.42 Refer to Example 2.41. Find the cumulative distribution function F(x) and plot it. Solution The probability density function of X is given by f(t) = 0 if t ≤ 0, f(t) = 1.25(1 − t4) if 0 < t < 1, and f(t) = 0 if t ≥ 1. The cumulative distribution function is given by . Since f(t) is defined separately on three different intervals, the computation of the cumulative distribution function involves three separate cases. If x ≤ 0: If 0 < x < 1: If x ≥ 1: Therefore Page 106 A plot of F(x) is presented here. Note that the cumulative distribution function F(x) in Example 2.42 is a continuous function —there are no jumps in its graph. This is characteristic of continuous random variables. The cumulative distribution function of a continuous random variable will always be continuous, while the cumulative distribution function of a noncontinuous random variable will never be continuous. Example 2.43 Refer to Example 2.41. Use the cumulative distribution function to find the probability that the shaft clearance is less than 0.5 mm. Solution Let X denote the shaft clearance. We need to find P(X ≤ 0.5). This is equivalent to finding F(0.5), where F(x) is the cumulative distribution function. Using the results of Example 2.42, F(0.5) = 1.25(0.5 − 0.55/5) = 0.617. Mean and Variance for Continuous Random Variables The population mean and variance of a continuous random variable are defined in the same way as those of a discrete random variable, except that the probability density function is used instead of the probability mass function. Specifically, if X is a continuous random variable, its population mean is defined to be the center of mass of its probability density function, and its population variance is the moment of inertia around a vertical axis through the population mean. The formulas are analogous to Equations (2.29) through (2.31), with the sums replaced by integrals. As was the case with discrete random variables, we will sometimes drop the word "population" and refer to the population mean, population variance, and population standard deviation more simply as the mean, variance, and standard deviation, respectively. Page 107 Definition Let X be a continuous random variable with probability density function f(x). Then the mean of X is given by (2.37) The mean of X is sometimes called the expectation, or expected value, of X and may also be denoted by E(X) or by μ. Definition Let X be a continuous random variable with probability density function f(x). Then ■ The variance of X is given by (2.38) ■ An alternate formula for the variance is given by (2.39) ■ ■ The variance of X may also be denoted by V(X) or by σ2. The standard deviation is the square root of the variance: . 2.44 Refer to Example 2.41. Find the mean clearance and the variance of the clearance. Solution Using Equation (2.37), the mean clearance is given by Having computed μX = 0.4167, we can now compute . It is easiest to use the Page 108 alternate formula, Equation (2.39): The Population Median and Percentiles In Section 1.2, we defined the median of a sample to be the middle number, or the average of the two middle numbers, when the sample values are arranged from smallest to largest. Intuitively, the sample median is the point that divides the sample in half. The population median is defined analogously. In terms of the probability density function, the median is the point at which half the area under the curve is to the left, and half the area is to the right. Thus if X is a continuous random variable with probability density function f(x), the median of X is the point xm that solves the equation . The median is a special case of a percentile. Let 0 < p < 100. The pth percentile of a population is the value xp such that p% of the population values are less than or equal to xp. Thus if X is a continuous random variable with probability density function f , the pth percentile of X is the point xp that solves the equation p/100. Note that the median is the 50th percentile. Figure 2.14 illustrates the median and the 90th percentile for a hypothetical population. FIGURE 2.14 (a) Half of the population values are less than the median. (b) Ninety percent of the population values are less than the 90th percentile x90. Page 109 Definition Let X be a continuous random variable with probability mass function f(x) and cumulative distribution function F(x). ■ The median of X is the point xm that solves the equation ■ If p is any number between 0 and 100, the pth percentile is the point xp that solves the equation ■ The median is the 50th percentile. We note that it is possible to construct continuous random variables for which there is an interval of points that satisfy the definition of the median or a percentile. Such random variables are seldom found in practice. Example 2.45 A certain radioactive mass emits alpha particles from time to time. The time between emissions, in seconds, is random, with probability density function Find the median time between emissions. Find the 60th percentile of the times. Solution The median is the solution to . We therefore must solve Half of the times between emissions are less than 6.931 s, and half are greater. The 60th percentile x60 is the solution to . We proceed as before, substituting x60 for xm, and 0.6 for 0.5. We obtain Page 110 Sixty percent of the times between emissions are less than 9.163 s, and 40% are greater. Chebyshev's Inequality The mean of a random variable is a measure of the center of its distribution, and the standard deviation is a measure of the spread. Chebyshev's inequality relates the mean and the standard deviation by providing a bound on the probability that a random variable takes on a value that differs from its mean by more than a given multiple of its standard deviation. Specifically, the probability that a random variable differs from its mean by k standard deviations or more is never greater than 1/k2. Chebyshev's Inequality Let X be a random variable with mean μX and standard deviation σX. Then Example 2.46 The length of a rivet manufactured by a certain process has mean μX = 50 mm and standard deviation σX = 0.45 mm. What is the largest possible value for the probability that the length of the rivet is outside the interval 49.1–50.9 mm? Solution Let X denote the length of a randomly sampled rivet. We must find P(X ≤ 49.1 or X ≥ 50.9). Now Applying Chebyshev's inequality with k = 2, we conclude that Chebyshev's inequality is valid for any random variable and does not require knowledge of the distribution. Because it is so general, the bound given by Chebyshev's inequality is in most cases much greater than the actual probability. Example 2.47 illustrates this. Example 2.47 The length of a rivet in Example 2.46, it can be verified that μX = 50 and σX = 0.45. Compute the probability that the length of the rivet is outside the interval 49.1–50.9 mm. How close is this probability to the Chebyshev bound of 1/4? Page 111 Solution The actual probability is much smaller than the Chebyshev bound of 1/4. Because the Chebyshev bound is generally much greater than the actual probability, it should only be used when the distribution of the random variable is unknown. When the distribution is known, then the probability mass function or probability density function should be used to compute probabilities. Derivation of the Alternate Formula for the Variance We derive Equation (2.31). We begin with Equation (2.30). Multiplying out (x − μX)2, we obtain Distributing the term P(X = x) over the terms in the parentheses yields Summing the terms separately, (2.40) Now and . Substituting into Equation (2.40) yields We conclude that To derive the alternate formula (2.39) for the variance of a continuous random variable from Equation (2.38), the same steps may be used, replacing ∑x with , and P(X = x) with f(x) dx. Page 112 Exercises for Section 2.4 1. 2. Determine whether each of the following random variables is discrete or continuous. a. The number of heads in 100 tosses of a coin. b. The length of a rod randomly chosen from a day's production. c. The final exam score of a randomly chosen student from last semester's engineering statistics class. d. The age of a randomly chosen Colorado School of Mines student. e. The age that a randomly chosen Colorado School of Mines student will be on his or her next birthday. Computer chips often contain surface imperfections. For a certain type of computer chip, the probability mass function of the number of defects X is presented in the following table. x p(x) a. b. c. d. e. f. 0 1 2 3 4 0.4 0.3 0.15 0.10 0.05 Find P(X ≤ 2). Find P(X > 1). Find μX. Find σ2X. 3. A chemical supply company ships a certain solvent in 10-gallon drums. Let X represent the number of drums ordered by a randomly chosen customer. Assume X has the following probability mass function: x p(x) 1 0.4 2 0.2 3 0.2 4 0.1 5 0.1 a. Find the mean number of drums ordered. b. Find the variance of the number of drums ordered. c. Find the standard deviation of the number of drums ordered. d. Let Y be the number of gallons ordered. Find the probability mass function of Y. e. Find the mean number of gallons ordered. f. Find the variance of the number of gallons ordered. g. Find the standard deviation of the number of gallons ordered. 4. x p1(x) 0 0.2 1 0.2 2 0.3 3 0.1 4 0.1 p2(x) 0.1 0.3 0.1 0.2 0.3 p3(x) 0.1 0.2 0.4 0.2 0.1 3 0.1 4 0.1 A survey of cars on a certain stretch of highway during morning commute hours showed that 70% had only one occupant, 15% had 2, 10% had 3, 3% had 4, and 2% had 5. Let X represent the number of occupants in a randomly chosen car. a. Find the probability mass function of X. b. Find P(X ≤ 2). c. Find P(X > 3). d. Find μX. e. 6. 2 0.2 a. Find the mean number of drums ordered. b. Find the variance of the number of drums ordered. c. Find the standard deviation of the number of drums ordered. d. Let Y be the number of gallons ordered. Find the probability mass function of Y. e. Find the mean number of gallons ordered. f. Find the variance of the number of gallons ordered. g. Find the standard deviation of the number of gallons ordered. 4. Let X represent the number of tires with low air pressure on a randomly chosen car. a. Which of the three functions below is a possible probability mass function of X? Explain. b. 5. 1 0.4 Find σX. The element titanium has five stable occurring isotopes, differing from each other in the number of neutrons an atom contains. If X is the number of neutrons in a randomly chosen titanium atom, the probability mass function of X is given as follows: x p(x) a. Find μX. 24 0.0825 25 0.0744 26 0.7372 27 0.0541 28 0.0518 b. 7. 8. Find σX. A computer sends a packet of information along a channel and waits for a return signal acknowledging that the packet has been received. If no acknowledgment is Page 113 received within a certain time, the packet is re-sent. Let X represent the number of times the packet is sent. Assume that the probability mass function of X is given by where c is a constant. a. Find the value of the constant c so that p(x) is a probability mass function. b. Find P(X = 2). c. Find the mean number of times the packet is sent. d. Find the variance of the number of times the packet is sent. e. Find the standard deviation of the number of times the packet is sent. After manufacture, computer disks are tested for errors. Let X be the number of errors detected on a randomly chosen disk. The following table presents values of the cumulative distribution function F(x) of X. x 0 1 2 3 4 F(x) 0.41 0.72 0.83 0.95 1.00 a. What is the probability that two or fewer errors are detected? b. What is the probability that more than three errors are detected? c. What is the probability that exactly one error is detected? d. What is the probability that no errors are detected? e. What is the most probable number of errors to be detected? 9. On 100 different days, a traffic engineer counts the number of cars that pass through a certain intersection between 5 P.M. and 5:05 P.M. The results are presented in the following table. Number of Cars 0 1 2 3 4 5 Number of Days 36 28 15 10 7 4 Proportion of Days 0.36 0.28 0.15 0.10 0.07 0.04 Let X be the number of cars passing through the intersection between 5 P.M. and 5:05 P.M. on a randomly chosen day. Someone suggests that for any positive integer x, the probability mass function of X is p1(x) = (0.2)(0.8)x. Using this function, compute P(X = x) for values of x from 0 through 5 inclusive. b. Someone else suggests that for any positive integer x, the probability mass function is p2(x) = (0.4)(0.6)x. Using this function, compute P(X = x) for values of x from 0 through 5 inclusive. c. Compare the results of parts (a) and (b) to the data in the table. Which probability mass function appears to be the better model? Explain. d. Someone says that neither of the functions is a good model since neither one agrees with the data exactly. Is this right? Explain. 10. Candidates for a job are interviewed one by one until a qualified candidate is found. Thirty percent of the candidates are qualified. a. What is the probability that the first candidate is qualified? b. What is the probability that the first qualified candidate will be the second one interviewed? c. Let X represent the number of candidates interviewed up to and including the first qualified candidate. Find P(X = 4). d. Find the probability mass function of X. 11. Refer to Exercise 10. The company has two positions open. They will interview candidates until two qualified candidates are found. Let Y represent the number of candidates interviewed up to and including the second qualified candidate. a. What is the smallest possible value for Y? b. What is the probability that Y takes on that value? c. Let X represent the number of applicants who are interviewed up to and including the first qualified candidate. Find P(Y = 4 | X = 1). Page 114 a. d. Find P(Y = 4 | X = 2). e. Find P(Y = 4). 12. Three components are randomly sampled, one at a time, from a large lot. As each component is selected, it is tested. If it passes the test, a success (S) occurs; if it fails the test, a failure (F) occurs. Assume that 80% of the components in the lot will succeed in passing the test. Let X represent the number of successes among the three sampled components. a. What are the possible values for X? b. Find P(X = 3). c. The event that the first component fails and the next two succeed is denoted by FSS. Find P(FSS). d. Find P(SFS) and P(SSF). e. Use the results of parts (c) and (d) to find P(X = 2). f. Find P(X = 1). g. Find P(X = 0). h. Find μX. i. Find . 13. Resistors labeled 100 Ω have true resistances that are between 80 Ω and 120 Ω. Let X be the resistance of a randomly chosen resistor. The probability density function of X is given by a. What proportion of resistors have resistances less than 90 Ω? b. Find the mean resistance. c. Find the standard deviation of the resistances. d. Find the cumulative distribution function of the resistances. 14. Elongation (in percent) of steel plates treated with aluminum are random with probability density function a. What proportion of steel plates have elongations greater than 25%? b. Find the mean elongation. c. Find the variance of the elongations. d. Find the standard deviation of the elongations. e. Find the cumulative distribution function of the elongations. f. A particular plate elongates 28%. What proportion of plates elongate more than this? 15. The lifetime of a transistor in a certain application is random with probability density function a. Find the mean lifetime. b. Find the standard deviation of the lifetimes. c. Find the cumulative distribution function of the lifetime. d. Find the probability that the lifetime will be less than 12 months. 16. A process that manufactures piston rings produces rings whose diameters (in centimeters) vary according to the probability density function a. Find the mean diameter of rings manufactured by this process. b. Find the standard deviation of the diameters of rings manufactured by this process. (Hint: Equation 2.38 may be easier to use than Equation 2.39.) c. Find the cumulative distribution function of piston ring diameters. d. What proportion of piston rings have diameters less than 9.75 cm? e. What proportion of piston rings have diameters between 9.75 and 10.25 cm? 17. Refer to Exercise 16. A competing process produces rings whose diameters (in centimeters) vary according to the probability density function Specifications call for the diameter to be 10.0±0.1cm. Which process is better, this Page 115 one or the one in Exercise 16? Explain. 18. The lifetime, in years, of a certain type of pump is a random variable with probability density function a. What is the probability that a pump lasts more than two years? b. What is the probability that a pump lasts between two and four years? c. Find the mean lifetime. d. Find the variance of the lifetimes. e. Find the cumulative distribution function of the lifetime. f. Find the median lifetime. g. Find the 60th percentile of the lifetimes. 19. The level of impurity (in percent) in the product of a certain chemical process is a random variable with probability density function a. What is the probability that the impurity level is greater than 3%? b. What is the probability that the impurity level is between 2% and 3%? c. Find the mean impurity level. d. Find the variance of the impurity levels. e. Find the cumulative distribution function of the impurity level. 20. The main bearing clearance (in mm) in a certain type of engine is a random variable with probability density function a. b. c. d. e. What is the probability that the clearance is less than 0.02 mm? Find the mean clearance. Find the standard deviation of the clearances. Find the cumulative distribution function of the clearance. Find the median clearance. The specification for the clearance is 0.015 to 0.063 mm. What is the probability that the specification is met? 21. The error in the length of a part (absolute value of the difference between the actual length and the target length), in mm, is a random variable with probability density function a. b. c. d. e. What is the probability that the error is less than 0.2 mm? Find the mean error. Find the variance of the error. Find the cumulative distribution function of the error. The specification for the error is 0 to 0.8 mm. What is the probability that the specification is met? 22. The concentration of a reactant is a random variable with probability density function a. What is the probability that the concentration is greater than 0.5? b. Find the mean concentration. c. Find the probability that the concentration is within ±0.1 of the mean. d. Find the standard deviation σ of the concentrations. e. Find the probability that the concentration is within ±σ of the mean. f. Find the cumulative distribution function of the concentration. 23. The thickness of a washer (in mm) is a random variable with probability density function Page 116 a. What is the probability that the thickness is less than 2.5 m? b. What is the probability that the thickness is between 2.5 and 3.5 m? c. Find the mean thickness. d. Find the standard deviation σ of the thicknesses. e. Find the probability that the thickness is within ±σ of the mean. f. Find the cumulative distribution function of the thickness. 24. Particles are a major component of air pollution in many areas. It is of interest to study the sizes of contaminating particles. Let X represent the diameter, in micrometers, of a randomly chosen particle. Assume that in a certain area, the probability density function of X is inversely proportional to the volume of the particle; that is, assume that where c is a constant. a. Find the value of c so that f(x) is a probability density function. b. Find the mean particle diameter. c. Find the cumulative distribution function of the particle diameter. d. Find the median particle diameter. e. The term PM10 refers to particles 10 μm or less in diameter. What proportion of the contaminating particles are PM10? f. The term PM25 refers to particles 2.5 μm or less in diameter. What proportion of the contaminating particles are PM25? g. What proportion of the PM10 particles are PM25? 25. The repair time (in hours) for a certain machine is a random variable with probability density function a. What is the probability that the repair time is less than 2 hours? b. What is the probability that the repair time is between 1.5 and 3 hours? c. Find the mean repair time. d. Find the cumulative distribution function of the repair times. 26. The diameter of a rivet (in mm) is a random variable with probability density function a. b. c. d. e. What is the probability that the diameter is less than 12.5 mm? Find the mean diameter. Find the standard deviation of the diameters. Find the cumulative distribution function of the diameter. The specification for the diameter is 12.3 to 12.7 mm. What is the probability that the specification is met? 2.5 Linear Functions of Random Variables In practice we often construct new random variables by performing arithmetic operations on other random variables. For example, we might add a constant to a random variable, multiply a random variable by a constant, or add two or more random variables together. In this section, we describe how to compute means and variances of random variables constructed in these Page 117 ways, and we present some practical examples. The presentation in this section is intuitive. A more rigorous presentation is provided in Section 2.6. For those desiring such a presentation, Section 2.6 may be covered in addition to, or in place of, this section. Adding a Constant When a constant is added to a random variable, the mean is increased by the value of the constant, but the variance and standard deviation are unchanged. For example, assume that steel rods produced by a certain machine have a mean length of 5.0 in. and a variance of σ2 = 0.003 in2. Each rod is attached to a base that is exactly 1.0in. long. The mean length of the assembly will be 5.0 + 1.0 = 6.0 in. Since each length is increased by the same amount, the spread in the lengths does not change, so the variance remains the same. To put this in statistical terms, let X be the length of a randomly chosen rod, and let Y = X + 1 be the length of the assembly. Then μY = μX+1 = μX + 1, and . In general, when a constant is added to a random variable, the mean is shifted by that constant, and the variance is unchanged. Summary If X is a random variable and b is a constant, then (2.41) (2.42) Multiplying by a Constant Often we need to multiply a random variable by a constant. This might be done, for example, to convert to a more convenient set of units. We continue the example of steel rod production to show how multiplication by a constant affects the mean, variance, and standard deviation of a random variable. If we measure the lengths of the rods described earlier in centimeters rather than inches, the mean length will be (2.54 cm/in.)(5.0 in.) = 12.7 cm. In statistical terms, let the random variable X be the length in inches of a randomly chosen rod, and let Y = 2.54X be the length in centimeters. Then μY = 2.54μX. In general, when a random variable is multiplied by a constant, its mean is multiplied by the same constant. Summary If X is a random variable and a is a constant, then (2.43) When the length X of a rod is measured in inches, the variance units of in2.If Y = 2.54X is the length in centimeters, then Therefore we obtain by multiplying must have Page 118 must have units of cm2. by 2.542, which is the conversion factor from in2 to cm2. In general, when a random variable is multiplied by a constant, its variance is multiplied by the square of the constant. Summary If X is a random variable and a is a constant, then (2.44) (2.45) If a random variable is multiplied by a constant and then added to another constant, the effect on the mean and variance can be determined by combining Equations (2.41) and (2.43) and Equations (2.42) and (2.44). The results are presented in the following summary. Summary If X is a random variable, and a and b are constants, then (2.46) (2.47) (2.48) Note that Equations (2.46) through (2.48) are analogous to results for the sample mean and standard deviation presented in Section 1.2. Example 2.48 The molarity of a solute in solution is defined to be the number of moles of solute per liter of solution (1 mole = 6.02 × 1023 molecules). If the molarity of a stock solution of concentrated sulfuric acid (H2SO4) is X, and if one part of the solution is mixed with N parts water, the molarity Y of the dilute solution is given by Y = X/(N + 1). Assume that the stock solution is manufactured by a process that produces a molarity with mean 18 and standard deviation 0.1. If 100 mL of stock solution is added to 300 mL of water, find the mean and standard deviation of the molarity of the dilute solution. Page 119 Solution The molarity of the dilute solution is Y = 0.25X. The mean and standard deviation of X are μX = 18 and σX = 0.1, respectively. Therefore (using Equation 2.43) Also, (using Equation 2.45) Means of Linear Combinations of Random Variables Consider the case of adding two random variables. For example, assume that there are two machines that fabricate a certain metal part. The mean daily production of machine A is 100 parts, and the mean daily production of machine B is 150 parts. Clearly the mean daily production from the two machines together is 250 parts. Putting this in mathematical notation, let X be the number of parts produced on a given day by machine A, and let Y be the number of parts produced on the same day by machine B. The total number of parts is X + Y, and we have that μX+Y = μX + μY. This idea extends to any number of random variables. If X1, X2, ..., Xn are random variables, then the mean of the sum X1 + X2 + ... + Xn is given by (2.49) The sum X1 + X2 + ... + Xn is a special case of a linear combination: If X1, ..., Xn are random variables and c1, ..., cn are constants, then the random variable is called a linear combination of X1, ..., Xn. To find the mean of a linear combination of random variables, we can combine Page 120 Equations (2.42) and (2.49): If X and Y are random variables, and a and b are constants, then (2.50) More generally, if X1, X2, ..., Xn are random variables and c1, c2, ..., cn are constants, then the mean of the linear combination c1X1 + c2X2 + ... + cnXn is given by (2.51) Independent Random Variables The notion of independence for random variables is very much like the notion of independence for events. Two random variables are independent if knowledge concerning one of them does not affect the probabilities of the other. When two events are independent, the probability that both occur is found by multiplying the probabilities for each event (see Equations 2.19 and 2.20 in Section 2.3). There are analogous formulas for independent random variables. The notation for these formulas is as follows. Let X be a random variable and let S be a set of numbers. The notation "X ∈ S" means that the value of the random variable X is in the set S. Definition If X and Y are independent random variables, and S and T are sets of numbers, then (2.52) More generally, if X1, ..., Xn are independent random variables, and S1, ..., Sn are sets, then (2.53) Example 2.49 Rectangular plastic covers for a compact disc (CD) tray have specifications regarding length and width. Let X be the length and Y be the width, each measured to the nearest millimeter, of a randomly sampled cover. The probability mass function of X is given by P(X = 129) = 0.2, P(X = 130) = 0.7, and P(X = 131) = 0.1. The probability mass function of Y is given by P(Y = 15) = 0.6 and P(Y = 16) = 0.4. The area of a cover is given by A = XY. Assume X and Y are independent. Find the probability that the area is 1935 mm2. Page 121 Solution The area will be equal to 1935 if X = 129 and Y = 15. Therefore since X and Y are independent Equations (2.52) and (2.53) tell how to compute probabilities for independent random variables, but they are not usually much help in determining whether random variables actually are independent. In general, the best way to determine whether random variables are independent is through an understanding of the process that generated them. Variances of Linear Combinations of Independent Random Variables We have seen that the mean of a sum of random variables is always equal to the sum of the means (Equation 2.49). In general, the formula for the variance of a sum of random variables is a little more complicated than this. But when random variables are independent, the result is simple: the variance of the sum is the sum of the variances. If X1, X2, ..., Xn are independent random variables, then the variance of the sum X1 + X2 + ... + Xn is given by (2.54) To find the variance of a linear combination of random variables, we can combine Equations (2.54) and (2.44): If X1, X2, ..., Xn are independent random variables and c1, c2, ..., cn are constants, then the variance of the linear combination c1X1 + c2X2 + ... + cnXn is given by (2.55) Two frequently encountered linear combinations are the sum and the difference of two random variables. Interestingly enough, when the random variables are independent, the variance of the sum is the same as the variance of the difference. If X and Y are independent random variables with variances and , then the variance of the sum X + Y is (2.56) The variance of the difference X − Y is (2.57) The fact that the variance of the difference is the sum of the variances may seem Page 122 counterintuitive. However, it follows from Equation (2.55) by setting c1 = 1 and c2 = −1. Example 2.50 A piston is placed inside a cylinder. The clearance is the distance between the edge of the piston and the wall of the cylinder and is equal to one-half the difference between the cylinder diameter and the piston diameter. Assume the piston diameter has a mean of 80.85 cm with a standard deviation of 0.02 cm. Assume the cylinder diameter has a mean of 80.95 cm with a standard deviation of 0.03 cm. Find the mean clearance. Assuming that the piston and cylinder are chosen independently, find the standard deviation of the clearance. Solution Let X1 represent the diameter of the cylinder and let X2 the diameter of the piston. The clearance is given by C = 0.5X1 − 0.5X2. Using Equation (2.51), the mean clearance is Since X1 and X2 are independent, we can use Equation (2.55) to find the standard deviation σC: Independence and Simple Random Samples When a simple random sample of numerical values is drawn from a population, each item in the sample can be thought of as a random variable. The items in a simple random sample may be treated as independent, except when the sample is a large proportion (more than 5%) of a finite population (see the discussion of independence in Section 1.1). From here on, unless explicitly stated to the contrary, we will assume this exception has not occurred, so that the values in a simple random sample may be treated as independent random variables. Summary If X1, X2, ...,Xn is a simple random sample, then X1, X2, ..., Xn may be treated as independent random variables, all with the same distribution. When X1, ..., Xn are independent random variables, all with the same distribution, it is Page 123 sometimes said that X1, ..., Xn are independent and identically distributed (i.i.d.). The Mean and Variance of a Sample Mean The most frequently encountered linear combination is the sample mean. Specifically, if X1, ..., Xn is a simple random sample from a population with mean μ and variance σ2, then the sample mean is the linear combination From this fact we can compute the mean and variance of . (using Equation 2.51) As discussed previously, the items in a simple random sample may be treated as independent random variables. Therefore (using Equation 2.55) Summary If X1, ..., Xn is a simple random sample from a population with mean μ and variance σ2, then the sample mean is a random variable with (2.58) (2.59) The standard deviation of is (2.60) Page 124 Example 2.51 A process that fills plastic bottles with a beverage has a mean fill volume of 2.013 L and a standard deviation of 0.005 L. A case contains 24 bottles. Assuming that the bottles in a case are a simple random sample of bottles filled by this method, find the mean and standard deviation of the average volume per bottle in a case. Solution Let V1, ..., V24 represent the volumes in 24 bottles in a case. This is a simple random sample from a population with mean μ = 2.013 and standard deviation σ = 0.005. The average volume is . Using Equation (2.58), Using Equation (2.60), Exercises for Section 2.5 1. If X and Y are independent random variables with means μX = 9.5 and μY = 6.8, and standard deviations σX = 0.4and σY = 0.1, find the means and standard deviations of the following: a. 3X 2. b. Y − X c. X + 4Y The bottom of a cylindrical container has an area of 10 cm2. The container is filled to a height whose mean is 5cm, and whose standard deviation is 0.1 cm. Let V denote the volume of fluid in the container. a. Find μV. b. 3. 4. The lifetime of a certain transistor in a certain application has mean 900 hours and standard deviation 30 hours. Find the mean and standard deviation of the length of time that four bulbs will last. Two batteries, with voltages V1 and V2, are connected in series. The total voltage V is given by V = V1 + V2. Assume that V1 has mean 12 V and standard deviation 1 V, and that V2 has mean 6 V and standard deviation 0.5 V. a. Find μV. b. 5. 6. 7. Assuming V1 and V2 to be independent, find σV. A laminated item is composed of five layers. The layers are a simple random sample from a population whose thickness has mean 1.2 mm and standard deviation 0.04 mm. a. Find the mean thickness of an item. b. Find the standard deviation of the thickness of an item. Two independent measurements are made of the lifetime of a charmed strange meson. Each measurement has a standard deviation of 7×10−15 seconds. The lifetime of the meson is estimated by averaging the two measurements. What is the standard deviation of this estimate? The molarity of a solute in solution is defined to be the number of moles of solute per liter of solution (1 mole = 6.02 × 1023 molecules). If X is the molarity of a solution of magnesium chloride (MgCl2), and Y is the molarity of a solution of ferric chloride (FeCl3), the molarity of chloride ion (Cl−) in a solution made of equal parts of the solutions of MgCl2 and FeCl3 is given by M = X + 1.5Y. Assume that X has mean 0.125 and standard deviation 0.05, and that Y has mean 0.350 and standard deviation 0.10. a. Find μM. b. 8. Find σV. Assuming X and Y to be independent, find σM. A machine that fills bottles with a beverage has a fill volume whose mean is 20.01 ounces, with a standard deviation of 0.02 ounces. A case consists of 24 bottles randomly sampled from the output of the machine. a. Find the mean of the total volume of the beverage in the case. b. Find the standard deviation of the total volume of the beverage in the case. Page 125 c. d. e. Find the mean of the average volume per bottle of the beverage in the case. Find the standard deviation of the volume per bottle of the beverage in the case. How many bottles must be included in a case for the standard deviation of the average volume per bottle to be 0.0025 ounces? The four sides of a picture frame consist of two pieces selected from a population whose mean length is 30 cm with standard deviation 0.1 cm, and two pieces selected from a population whose mean length is 45 cm with standard deviation 0.3 cm. a. Find the mean perimeter. b. Assuming the four pieces are chosen independently, find the standard deviation of the perimeter. 10. A gas station earns $2.60 in revenue for each gallon of regular gas it sells, $2.75 for each gallon of midgrade gas, and $2.90 for each gallon of premium gas. Let X1, X2, and X3 denote the numbers of gallons of regular, midgrade, and premium gasoline sold in a day. Assume that X1, X2, and X3 have means μ1 = 1500, μ2 = 500, and μ3 = 300, and standard deviations σ1 = 180, σ2 = 90, and σ3 = 40, respectively. 9. a. b. Find the mean daily revenue. Assuming X1, X2, and X3 to be independent, find the standard deviation of the daily revenue. 11. A certain commercial jet plane uses a mean of 0.15 gallons of fuel per passenger-mile, with a standard deviation of 0.01 gallons. a. Find the mean number of gallons the plane uses to fly 8000 miles if it carries 210 passengers. b. Assume the amounts of fuel used are independent for each passenger-mile traveled. Find the standard deviation of the number of gallons of fuel the plane uses to fly 8000 miles while carrying 210 passengers. c. The plane used X gallons of fuel to carry 210 passengers 8000 miles. The fuel efficiency is estimated as X/(210 × 8000). Find the mean of this estimate. d. Assuming the amounts of fuel used are independent for each passenger-mile, find the standard deviation of the estimate in part (c). 12. Woodworkers Alex and Brandon make precision pieces for toys. Each toy requires four 8inch long sticks. Alex cuts four sticks one at time, and their lengths are independent. Alex's sticks have mean length 8 inches with a standard deviation 0.15 inches. Brandon clamps four sticks together and cuts them all at once, so that they all have the same length. Brandon's sticks also have a mean length of 8 inches with a standard deviation of 0.15 inches. a. Alex's four sticks are laid end-to-end. What is the mean of the total length? What is the standard deviation of the total length? b. Brandon's four sticks are laid end-to-end. What is the mean of the total length? What is the standard deviation of the total length? c. Four sticks, two from Alex and two from Brandon, are laid end-to-end. What is the mean of the total length? What is the standard deviation of the total length? 13. In the article "An Investigation of the Ca−CO3−CaF2−K2SiO3−SiO2−Fe Flux System Using the Submerged Arc Welding Process on HSLA-100 and AISI-1018 Steels" (G. Fredrickson, M.S. thesis, Colorado School of Mines, 1992), the carbon equivalent P of a weld metal is defined to be a linear combination of the weight percentages of carbon (C), manganese (Mn), copper (Cu), chromium (Cr), silicon (Si), nickel (Ni), molybdenum (Mo), vanadium (V), and boron (B). The carbon equivalent is given by Means and standard deviations of the weight percents of these chemicals were estimated from measurements on 45 weld metals produced on HSLA-100 steel base metal. Assume the means and standard deviations (SD) are as given in the following table. C Mn Cu Cr Si Ni Mo V B Mean 0.0695 1.0477 0.8649 0.7356 0.2171 2.8146 0.5913 0.0079 0.0006 SD 0.0018 0.0269 0.0225 0.0113 0.0185 0.0284 0.0031 0.0006 0.0002 a. Find the mean carbon equivalent of weld metals produced from HSLA-100 steel base metal. Page 126 b. Assuming the weight percents to be independent, find the standard deviation of the carbon equivalent of weld metals produced from HSLA-100 steel base metal. 14. The oxygen equivalence number of a weld is a number that can be used to predict properties such as hardness, strength, and ductility. The article "Advances in Oxygen Equivalence Equations for Predicting the Properties of Titanium Welds" (D. Harwig, W. Ittiwattana, and H. Castner, The Welding Journal, 2001:126s–136s) presents several equations for computing the oxygen equivalence number of a weld. One equation, designed to predict the hardness of a weld, is X = O + 2N + (2/3)C,where X is the oxygen equivalence, and O, H, and C are the amounts of oxygen, nitrogen, and carbon, respectively, in weight percent, in the weld. Suppose that for welds of a certain type, μO = 0.1668, μN = 0.0255, μC = 0.0247, σO = 0.0340, σN = 0.0194, and σC = 0.0131. a. Find μX. b. Suppose the weight percents of O, H, and C are independent. Find σX. 15. Measurements are made on the length and width (in cm) of a rectangular component. Because of measurement error, the measurements are random variables. Let X denote the length measurement and let Y denote the width measurement. Assume that the probability density function of X is and that the probability density function of Y is Assume that the measurements X and Y are independent. a. Find P(X < 9.98). b. Find P(Y > 5.01). c. Find P(X < 9.98 and Y > 5.01). d. Find μx. e. Find μY. 16. The thickness X of a wooden shim (in mm) has probability density function a. Find μx. b. Find c. Let Y denote the thickness of a shim in inches (1 mm = 0.0394 inches). Find μY and . . d. If three shims are selected independently and stacked one atop another, find the mean and variance of the total thickness. 17. The article "Abyssal Peridotites > 3800 Ma from Southern West Greenland: Field Relationships, Petrography, Geochronology, Whole-Rock and Mineral Chemistry of Dunite and Harzburgite Inclusions in the Itsaq Gneiss Complex" (C. Friend, V. Bennett, and A. Nutman, Contrib Mineral Petrol, 2002:71–92) describes the chemical compositions of certain minerals in the early Archaean mantle. For a certain type of olivine assembly, the silicon dioxide (SiO2) content (in weight percent) in a randomly chosen rock has mean 40.25 and standard deviation 0.36. a. Find the mean and standard deviation of the sample mean SiO2 content in a random sample of 10 rocks. b. How many rocks must be sampled so that the standard deviation of the sample mean SiO2 content is 0.05? 18. The number of bytes downloaded per second on an information channel has mean 105 and standard deviation 104. Among the factors influencing the rate is congestion, which produces alternating periods of faster and slower transmission. Let X represent the number of bytes downloaded in a randomly chosen five-second period. a. Is it reasonable to assume that μx = 5 × 105? Explain. b. Is it reasonable to assume that ? Explain. Page 127 2.6 Jointly Distributed Random Variables In this section, results concerning several random variables are presented in a more rigorous fashion than in Section 2.5. For those desiring such a presentation, this section may be covered in addition to, or in place of, Section 2.5. We have said that observing a value of a random variable is like sampling a value from a population. In some cases, the items in the population may each have several random variables associated with them. For example, imagine choosing a student at random from a list of all the students registered at a university and measuring that student's height and weight. Each individual in the population of students corresponds to two random variables, height and weight. If we also determined the student's age, each individual would correspond to three random variables. In principle, any number of random variables may be associated with each item in a population. When two or more random variables are associated with each item in a population, the random variables are said to be jointly distributed. If all the random variables are discrete, they are said to be jointly discrete. If all the random variables are continuous, they are said to be jointly continuous. We will study these two cases separately. Jointly Discrete Random Variables Example 2.49 (in Section 2.5) discussed the lengths and widths of rectangular plastic covers for a CD tray that is installed in a personal computer. Measurements are rounded to the nearest millimeter. Let X denote the measured length and Y the measured width. The possible values of X are 129, 130, and 131, and the possible values for Y are 15 and 16. Both X and Y are discrete, so X and Y are jointly discrete. There are six possible values for the ordered pair (X, Y): (129, 15), (129, 16), (130, 15), (130, 16), (131, 15), and (131, 16). Assume that the probabilities of each of these ordered pairs are as given in the following table. x 129 129 130 130 131 131 y 15 16 15 16 15 16 P(X = x and Y = y) 0.12 0.08 0.42 0.28 0.06 0.04 The joint probability mass function is the function p(x,y) = P(X = x and Y = y). So, for example, we have p(129, 15) = 0.12, and p(130, 16) = 0.28. Sometimes we are given a joint probability mass function of two random variables, but we are interested in only one of them. For example, we might be interested in the probability mass function of X, the length of the CD cover, but not interested in the width Y. We can obtain the probability mass function of either one of the variables X or Y separately by summing the appropriate values of the joint probability mass function. Examples 2.52 and 2.53 illustrate the method. Page 128 Example 2.52 Find the probability that a CD cover has a length of 129 mm. Solution It is clear from the previous table that 12% of the CD covers in the population have a length of 129 and a width of 15, and 8% have a length of 129 and a width of 16. Therefore 20% of the items in the population have a length of 129. The probability that a CD cover has a length of 129 mm is 0.20. In symbols, we have Example 2.53 Find the probability that a CD cover has a width of 16 mm. Solution We need to find P(Y = 16). We can find this quantity by summing the probabilities of all pairs (x,y) for which y = 16. We obtain Examples 2.52 and 2.53 show that we can find the probability mass function of X (or Y) by summing the joint probability mass function over all values of Y (or X). Table 2.4 presents the joint probability mass function of X and Y. The probability mass function of X appears in the rightmost column and is obtained by summing along the rows. The probability mass function of Y appears in the bottom row and is obtained by summing down the columns. Note that the probability mass functions of X and of Y appear in the margins of the table. For this reason they are often referred to as marginal probability mass functions. TABLE 2.4 Joint and marginal probability mass functions for the length X and width Y of a CD cover y x 129 130 131 Py(y) 15 0.12 0.42 0.06 0.60 16 0.08 0.28 0.04 0.40 Px(x) 0.20 0.70 0.10 Finally, if we sum the joint probability density function over all possible values of Page 129 x and y, we obtain the probability that X and Y take values somewhere within their possible ranges, and this probability is equal to 1. Summary If X and Y are jointly discrete random variables: ■ The joint probability mass function of X and Y is the function ■ ■ The marginal probability mass functions of X and of Y can be obtained from the joint probability mass function as follows: where the sums are taken over all the possible values of Y and of X, respectively. The joint probability mass function has the property that where the sum is taken over all the possible values of X and Y. Jointly Continuous Random Variables We have seen that if X is a continuous random variable, its probabilities are found by integrating its probability density function. We say that the random variables X and Y are jointly continuous if their probabilities are found by integrating a function of two variables, called the joint probability density function of X and Y. To find the probability that X and Y take values in any region, we integrate the joint probability density function over that region. Example 2.54 shows how. Example 2.54 Assume that for a certain type of washer, both the thickness and the hole diameter vary from item to item. Let X denote the thickness in millimeters and let Y denote the hole diameter in millimeters, for a randomly chosen washer. Assume that the joint probability density function of X and Y is given by Find the probability that a randomly chosen washer has a thickness between 1.0 and 1.5 mm, and a hole diameter between 4.5 and 5 mm. Page 130 Solution We need to find P(1 ≤ X ≤ 1.5 and 4.5 ≤ Y ≤ 5). The large rectangle in the figure indicates the region where the joint density is positive. The shaded rectangle indicates the region where 1 ≤ x ≤ 1.5 and 4.5 ≤ y ≤ 5, over which the joint density is to be integrated. We integrate the joint probability density function over the indicated region: Note that if a joint probability density function is integrated over the entire plane, that is, if the limits are −∞ to ∞ for both x and y, we obtain the probability that both X and Y take values between −∞ and ∞, which is equal to 1. Summary If X and Y are jointly continuous random variables, with joint probability density function f(x,y), and a < b, c < d, then The joint probability density function has the following properties: f(x, y) ≥ 0 for all x and y We have seen that if X and Y are jointly discrete, the probability mass function of Page 131 either variable may be found by summing the joint probability mass function over all the values of the other variable. When computed this way, the probability mass function is called the marginal probability mass function. By analogy, if X and Y are jointly continuous, the probability density function of either variable may be found by integrating the joint probability density function with respect to the other variable. When computed this way, the probability density function is called the marginal probability density function. Example 2.55 illustrates the idea. Example 2.55 Refer to Example 2.54. Find the marginal probability density function of the thickness X of a washer. Find the marginal probability density function of the hole diameter Y of a washer. Solution Denote the marginal probability density function of X by fX(x), and the marginal probability density function of Y by fY(y). Then and Summary If X and Y are jointly continuous with joint probability density function f(x,y), then the marginal probability density functions of X and of Y are given, respectively, by Page 132 Example 2.56 The article "Performance Comparison of Two Location Based Routing Protocols for Ad Hoc Networks" (T. Camp, J. Boleng, et al., Proceedings of the Twenty-first Annual Joint Conference of IEEE Computer and Communications Societies 2002: 1678–1687) describes a model for the movement of a mobile computer. Assume that a mobile computer moves within the region A bounded by the x axis, the line x = 1, and the line y = x in such a way that if (X,Y) denotes the position of the computer at a given time, the joint density of X and Y is given by Find P(X > 0.5 and Y < 0.5). Solution The region A is the triangle shown in Figure 2.15, with the region X > 0.5 and Y < 0.5 shaded in. To find P(X > 0.5 and Y < 0.5), we integrate the joint density over the shaded region. FIGURE 2.15 The triangle represents the region where the joint density of X and Y is positive. By integrating the joint density over the shaded square, we find the probability that the point (X,Y) lies in the shaded square. Page 133 Example 2.57 Refer to Example 2.56. Find the marginal densities of X and of Y. Solution To compute fX(x), the marginal density of X, we fix x and integrate the joint density along the vertical line through x, as shown in Figure 2.16. The integration is with respect to y, and the limits of integration are y = 0 to y = x. FIGURE 2.16 The marginal density fX(x) is computed by integrating the joint density along the vertical line through x. To compute fY(y), the marginal density of Y, we fix y and integrate the joint density along the horizontal line through y, as shown in Figure 2.17 (page 134). The integration is with respect to x, and the limits of integration are x = y to x = 1. Page 134 FIGURE 2.17 The marginal density fY(y) is computed by integrating the joint density along the horizontal line through y. More than Two Random Variables The ideas of joint probability mass functions and joint probability density functions extend easily to more than two random variables. We present the definitions here. Definition ■ If the random variables X1, ..., Xn are jointly discrete, the joint probability mass function is ■ If the random variables X1, ..., Xn are jointly continuous, they have a joint probability density function f(x1, ..., xn), where for any constants a1 ≤ b1, ..., an ≤ bn. Means of Functions of Random Variables Sometimes we are given a random variable X and we need to work with a function of X. If X is a random variable, and h(X) is a function of X, then h(X) is a random variable as well. If we wish to compute the mean of h(X), it can be done by using the probability mass function or probability density function of X. It is not necessary to know the probability mass function or probability density function of h(X). Page 135 ■ Let X be a random variable, and let h(X) be a function of X. Then If X is discrete with probability mass function p(x), the mean of h(X) is given by (2.61) where the sum is taken over all the possible values of X. ■ If X is continuous with probability density function f(x), the mean of h(X) is given by (2.62) Note that if we substitute h(X) = (X − μX)2 in either Equation (2.61) or (2.62), the right-hand side of the equation becomes an expression for the variance of X. It follows that We can obtain another expression for the variance of X by substituting h(X) = X2 and subtracting . from both sides of the equation. We conclude that Example . 2.58 An internal combustion engine contains several cylinders bored into the engine block. Let X represent the bore diameter of a cylinder, in millimeters. Assume that the probability density function of X is Let A = πX2/4 represent the area of the bore. Find the mean of A. Solution The mean area is 5096 mm2. If h(X) = aX + b is a linear function of X, then the mean μaX+b and the variance can be expressed in terms of μX and can . These results were presented in Equations (2.46) through (2.48) in Section 2.5; we repeat them here. Page 136 If X is a random variable, and a and b are constants, then (2.63) (2.64) (2.65) Proofs of these results are presented at the end of this section. If X and Y are jointly distributed random variables, and h(X,Y) is a function of X and Y, then the mean of h(X,Y) can be computed from the joint probability mass function or joint probability density function of X and Y. ■ If X and Y are jointly distributed random variables, and h(X,Y) is a function of X and Y, then If X and Y are jointly discrete with joint probability mass function p(x,y), (2.66) ■ where the sum is taken over all the possible values of X and Y. If X and Y are jointly continuous with joint probability density function f(x,y), Example 2.59 The displacement of a piston in an internal combustion engine is defined to be the volume that the top of the piston moves through from the top to the bottom of its stroke. Let X represent the diameter of the cylinder bore, in millimeters, and let Y represent the length of the piston stroke in millimeters. The displacement is given by D = πX2Y/4. Assume X and Y are jointly distributed with joint probability density function Find the mean of D. Solution Page 137 The mean displacement is 331,998 mm3, or approximately 332 mL. Conditional Distributions If X and Y are jointly distributed random variables, then knowing the value of X may change probabilities regarding the random variable Y. For example, let Y represent the height in inches and Y represent the weight in pounds of a randomly chosen college student. Let's say that we are interested in the probability P(Y ≥ 200). If we know the joint density of X and Y, we can determine this probability by computing the marginal density of Y. Now let's say that we learn that the student's height is X = 78. Clearly, this knowledge changes the probability that Y ≥ 200. To compute this new probability, the idea of a conditional distribution is needed. We will first discuss the case where X and Y are jointly discrete. Let x be any value for which P(X = x) > 0. Then the conditional probability that Y = y given X = x is P(Y = y | X = x). We will express this conditional probability in terms of the joint and marginal probability mass functions. Let p(x,y) denote the joint probability mass function of X and Y, and let pX(x) denote the marginal probability mass function of X. Then the conditional probability is The conditional probability mass function of Y given X = x is the conditional probability P(Y = y | X = x), considered as a function of y and x. Definition Let X and Y be jointly discrete random variables, with joint probability mass function p(x,y). Let pX(x) denote the marginal probability mass function of X and let x be any number for which pX(x) > 0. The conditional probability mass function of Y given X = x is (2.68) Note that for any particular values of x and y, the valueof pY|X(y | x) is just the conditional probability P(Y = y | X = x). Page 138 Example 2.60 Table 2.4 presents the joint probability mass function of the length X and width Y of a CD cover. Compute the conditional probability mass function pY|X(y | 130). Solution The possible values for Y are y = 15 and y = 16. From Table 2.4, P(Y = 15 and X = 130) = 0.42, and P(X = 130) = 0.70. Therefore, The value of pY|X(16 | 130) can be computed with a similar calculation. Alternatively, note that pY|X(16 | 130) = 1 − pY|X(15 | 130), since y = 15 and y = 16 are the only two possible values for Y. Therefore pY|X(16 | 130) = 0.4. The conditional probability mass function of Y given X = 130 is therefore pY|X(15 | 130) = 0.60, pY|X(16 | 130) = 0.4, and pY|X(y | 130) = 0 for any value of y other than 15 or 16. The analog to the conditional probability mass function for jointly continuous random variables is the conditional probability density function. The definition of the conditional probability density function is just like that of the conditional probability mass function, with mass functions replaced by density functions. Definition Let X and Y be jointly continuous random variables, with joint probability density function f(x,y). Let fX(x) denote the marginal density function of X and let x be any number for which fX(x) > 0. The conditional probability density function of Y given X = x is (2.69) Example 2.61 (Continuing Example 2.54.) The joint probability density function of the thickness X and hole diameter Y (both in millimeters) of a randomly chosen washer is f(x,y) = (1/6)(x + y)for 1 ≤ x ≤ 2 and 4 ≤ y ≤ 5. Find the conditional probability density function of Y given X = 1.2. Find the probability that the hole diameter is less than or equal to 4.8 mm given that the thickness is 1.2 mm. Page 139 Solution In Example 2.55 we computed the marginal probability density functions The conditional probability density function of Y given X = 1.2 is The probability that the hole diameter is less than or equal to 4.8 mm given that the thickness is 1.2 mm is P(Y ≤ 4.8 | X = 1.2). This is found by integrating fY|X(y|1.2) over the region y ≤ 4.8: Conditional Expectation Expectation is another term for mean. A conditional expectation is an expectation, or mean, calculated using a conditional probability mass function or conditional probability density function. The conditional expectation of Y given X = x is denoted E(Y | X = x) or μY|X=x. We illustrate with Examples 2.62 through 2.64. Example 2.62 Table 2.4 presents the joint probability mass function of the length X and width Y of a CD cover. Compute the conditional expectation E(Y | X = 130). Solution We computed the conditional probability mass function pY|X(y | 130) in Example 2.60. The conditional expectation E(Y | X = 130) is calculated using the definition of the mean Page 140 of a discrete random variable and the conditional probability mass function. Specifically, Example 2.63 Refer to Example 2.61. Find the conditional expectation of Y given that X = 1.2. Solution Since X and Y are jointly continuous, we use the definition of the mean of a continuous random variable to compute the conditional expectation. Example 2.64 Refer to Example 2.61. Find the value μY (which can be called the unconditional mean of Y). Does it differ from E(Y | X = 1.2)? Solution The value μY is calculated using the marginal probability mass function of Y. Thus the conditional expectation in this case differs slightly from the unconditional expectation. Independent Random Variables The notion of independence for random variables is very much like the notion of independence for events. Two random variables are independent if knowledge regarding one of them does not affect the probabilities of the other. We present here a definition of independence of random variables in terms of their joint probability mass or joint probability density function. A different but logically equivalent definition was presented in Section 2.5. Page 141 Definition Two random variables X and Y are independent, provided that ■ If X and Y are jointly discrete, the joint probability mass function is equal to the product of the marginals: ■ If X and Y are jointly continuous, the joint probability density function is equal to the product of the marginals: Random variables X1, ..., Xn are independent, provided that ■ If X1, ..., Xn are jointly discrete, the joint probability mass function is equal to the product of the marginals: ■ If X1, ..., Xn are jointly continuous, the joint probability density function is equal to the product of the marginals: Intuitively, when two random variables are independent, knowledge of the value of one of them does not affect the probability distribution of the other. In other words, the conditional distribution of Y given X is the same as the marginal distribution of Y. ■ If X and Y are independent random variables, then If X and Y are jointly discrete, and x is a value for which pX(x) > 0, then ■ If X and Y are jointly continuous, and x is a value for which fX(x) > 0, then Example 2.65 The joint probability mass function of the length X and thickness Y of a CD tray cover is given in Table 2.4. Are X and Y independent? Solution We must check to see if P(X = x and Y = y) = P(X = x)P(Y = y) for every value of x and y. We begin by checking x = 129, y = 15: Continuing in this way, we can verify that P(X = x and Y = y) = P(X = x)P(Y = y) Page 142 for every value of x and y. Therefore X and Y are independent. Example 2.66 (Continuing Example 2.54.) The joint probability density function of the thickness X and hole diameter Y of a randomly chosen washer is f(x,y) = (1/6)(x + y) for 1 ≤ x ≤ 2 and 4 ≤ y ≤ 5. Are X and Y independent? Solution In Example 2.55 we computed the marginal probability functions Clearly f(x,y) ≠ fX(x)fY(y). Therefore X and Y are not independent. Covariance When two random variables are not independent, it is useful to have a measure of the strength of the relationship between them. The population covariance is a measure of a certain type of relationship known as a linear relationship. We will usually drop the term "population," and refer simply to the covariance. Definition Let X and Y be random variables with means μX and μY. The covariance of X and Y is (2.70) An alternate formula is (2.71) A proof of the equivalence of these two formulas is presented at the end of the section. It is important to note that the units of Cov(X,Y) are the units of X multiplied by the units of Y. How does the covariance measure the strength of the linear relationship between X and Y? The covariance is the mean of the product of the deviations (X − μX)(Y − μY). If a Cartesian coordinate system is constructed with the origin at (μX, μY), this product will be positive in the first and third quadrants and negative in the second and fourth quadrants (see Figure 2.18). It follows that if Cov(X,Y) is strongly positive, then values of (X,Y) in the first and third quadrants will be observed much more often than values in the second and fourth quadrants. In a random sample of points, therefore, larger values of X would tend to be paired with larger values of Y, while smaller values of X would tend to be paired with smaller values of Y (see Figure 2.18a). Similarly, if Cov(X,Y) is strongly negative, the points in a random sample would be more likely to lie in the second and fourth quadrants, so larger values of X would tend to be paired Page 143 with smaller values of Y (see Figure 2.18b). Finally, if Cov(X,Y) is near 0, there would be little tendency for larger values of X to be paired with either larger or smaller values of Y (see Figure 2.18c). FIGURE 2.18 (a) A random sample of points from a population with positive covariance. (b) A random sample of points from a population with negative covariance. (c) A random sample of points from a population with covariance near 0. Example 2.67 Continuing Example 2.56, a mobile computer is moving in the region A bounded by the x axis, the line x = 1, and the line y = x (see Figure 2.15). If (X,Y) denotes the position of the computer at a given time, the joint density of X and Y is given by Find Cov(X,Y). Page 144 Solution We will use the formula Cov(X,Y) = μXY − μXμY (Equation 2.71). First we compute μXY: Now the joint density is positive on the triangle shown. To compute the integral over this region, we fix a value of x, as shown. We compute the inner integral by integrating with respect to y along the vertical line through x. The limits of integration along this line are y = 0 to y = x. Then we compute the outer integral by integrating with respect to x over all possible values of x, so the limits of integration on the outer integral are x = 0 to x = 1. Therefore We will use the marginal densities computed in Example 2.57. Page 145 Now Example 2.68 Quality-control checks on wood paneling involve counting the number of surface flaws on each panel. On a given 2 × 8 ft panel, let X be the number of flaws that are found upon application of the final coat of finishing material, and let Y be the number of defects found due to inclusions of foreign particles in the finish. The joint probability mass function p(x,y) of X and Y is presented in the following table. The marginal probability mass functions are presented as well, in the margins of the table. Find the covariance of X and Y. y x 0 1 2 PY(y) 0 0.05 0.05 0.25 0.35 1 0.10 0.15 0.10 0.35 2 0.20 0.05 0.05 0.30 Px(x) 0.35 0.25 0.40 Solution We will use the formula Cov(X,Y) = μXY − μXμY (Equation 2.71). First we compute μXY: We use the marginals to compute μX and μY: Page 146 It follows that Cov(X,Y) = 0.65 − (1.05)(0.95) = −0.3475. Correlation If X and Y are jointly distributed random variables, Cov(X,Y) measures the strength of the linear relationship between them. As mentioned previously, the covariance has units, which are the units of X multiplied by the units of Y. This is a serious drawback in practice, because one cannot use the covariance to determine which of two pairs of random variables is more strongly related, since the two covariances will have different units. What is needed is a measure of the strength of a linear relationship that is unitless. The population correlation is such a measure. We will usually drop the term "population," and refer simply to the correlation. The correlation is denoted ρX,Y or by ρX,Y. The correlation is a scaled version of the covariance. Specifically, to compute the correlation between X and Y, one first computes the covariance, and then gets rid of the units by dividing by the product of the standard deviations of X and Y. It can be proved that the correlation is always between −1 and 1 (see Exercise 29). Summary Let X and Y be jointly distributed random variables with standard deviations σX and σY. The correlation between X and Y is denoted ρX,Y and is given by (2.72) For any two random variables X and Y: Example 2.69 Refer to Example 2.67. Find ρX,Y. Solution In Example 2.67, we computed Cov(X,Y) = 0.01778, μX = 4/5, and μY = 8/15. We now must compute σX and σY. To do this we use the marginal densities of X and of Y, which were computed in Example 2.57. These are fX(x) = 4x3 for 0 < x < 1, and fY(y) = 4y − 4y3 for 0 < y < 1. Then Page 147 It follows that Example . 2.70 Refer to Example 2.68. Find ρX,Y. Solution In Example 2.68, we computed Cov(X,Y) = −0.3475, μX = 1.05, and μY = 0.95. We now compute σX and σY. To do this we use the marginal densities of X and of Y, which were presented in the table in Example 2.68. It follows that As an important special case, note that it follows immediately from the definition (Equation 2.70 or 2.71) that the covariance of any random variable and itself is the variance of that random variable. Also, it follows from the definition of correlation (Equation 2.72) that the correlation between any random variable and itself is 1. Page 148 Summary For any random variable X, Cov(X,X) = and ρX,X = 1. Covariance, Correlation, and Independence When Cov(X,Y) = ρX,Y = 0, there is no linear relationship between X and Y. In this case we say that X and Y are uncorrelated. Note that if Cov(X,Y) = 0, then it is always the case that ρX,Y = 0, and vice versa. If X and Y are independent random variables, then X and Y are always uncorrelated, since there is no relationship, linear or otherwise, between them. It is mathematically possible to construct random variables that are uncorrelated but not independent. This phenomenon is rarely seen in practice, however. Summary ■ If Cov(X,Y) = ρX,Y = 0, then X and Y are said to be uncorrelated. ■ ■ If X and Y are independent, then X and Y are uncorrelated. It is mathematically possible for X and Y to be uncorrelated without being independent. This rarely occurs in practice. A proof of the fact that independent random variables are uncorrelated is presented at the end of this section. An example of random variables that are uncorrelated but not independent is presented in Exercise 22. Linear Combinations of Random Variables We discussed linear combinations of random variables in Section 2.5. We review the results here and include additional results on the variance of a linear combination of dependent random variables. If X1, ..., Xn are random variables and c1, ..., cn are constants, then the random variable is called a linear combination of X1, ..., Xn. If X1, ..., Xn are random variables and c1, ..., cn are constants, then (2.73) (2.74) Proofs of these results for the case n = 2 are presented at the end of this section. Page 149 Equation (2.74) is the most general result regarding the variance of a linear combination of random variables. As a special case, note that if X1, ..., Xn are independent, then all the covariances are equal to 0, so the result simplifies: If X1, ..., Xn are independent random variables and c1, ..., cn are constants, then (2.75) In particular, (2.76) Finally, we present some special cases of Equations (2.74) and (2.76) in which there are only two random variables: If X and Y are random variables, then (2.77) (2.78) If X and Y are independent random variables, then (2.79) (2.80) Note that the variance of the difference X − Y of two independent random variables is the sum of the variances. Example 2.71 (Continuing Example 2.56.) Assume that the mobile computer moves from a random position (X,Y) vertically to the point (X, 0), and then along the x axis to the origin. Find the mean and variance of the total distance traveled. Solution The distance traveled is the sum X + Y. The means of X and of Y were computed in Example 2.67. They are μX = 4/5 = 0.800, and μY = 8/15 = 0.533. We use Equation (2.77). In Example 2.67 we computed Cov(X,Y) = 0.01778. In Example 2.69 we computed Therefore Page 150 The Mean and Variance of a Sample Mean We review the procedures for computing the mean and variance of a sample mean, which were presented in Section 2.5. When a simple random sample is drawn from a population, each item in the sample can be thought of as a random variable. Unless the sample is a large proportion (more than 5%) of the population, the items in the sample may be treated as independent (see the discussion of independence in Section 1.1). From here on, unless stated to the contrary, we will assume that the values in a simple random sample may be treated as independent random variables. If X1, ..., Xn is a simple random sample, then X1, ..., Xn may be treated as independent random variables, all with the same distribution. The most frequently encountered linear combination is the sample mean. Specifically, if X1, ..., Xn is a simple random sample from a population with mean μ and variance σ2, then the sample mean is the linear combination Formulas for the mean and variance of may be derived from Equations (2.73) and (2.75), respectively, by setting c1 = c2 = ··· = cn = 1/n. If X1, ..., Xn is a simple random sample from a population with mean μ and variance σ2, then (2.81) (2.82) The standard deviation of is (2.83) Example 2.72 The article "Water Price Influence on Apartment Complex Water Use" (D. Agthe and R. Billings, Journal of Water Resources Planning and Management, 2002:366–369) discusses the volume of water used in apartments in 308 complexes in Tucson, Arizona. Page 151 The volume used per apartment during the summer had mean 20.4 m3 and standard deviation 11.1m3. Find the mean and standard deviation for the sample mean water use in a sample of 100 apartments. How many apartments must be sampled so that the sample mean water use will have a standard deviation equal to 0.5 m3? Solution Let X1, ..., X100 be the amounts of water used in a sample of 100 apartments. Then X1, ..., X100 come from a population with μ = 20.4 and standard deviation σ = 11.1. We conclude that the sample mean has mean , and standard deviation . Let n be the sample size required so that . Then . Solving for n, we obtain n = 493. Application to Portfolio Management Example (2.73) and its variants play an important role in the field of finance. Assume that an investor has a fixed number of dollars to invest. She may choose from a variety of investments, for example, stocks, bonds, and real estate. After one year she will sell her investment; let X denote her profit (or loss). The value of X cannot be predicted with certainty, so economists treat it as a random variable. The mean μX indicates the amount that the investment can be expected to earn on the average. The standard deviation σX reflects the volatility, or risk, of the investment. If σX is very small, then it is nearly certain that the investment will earn close to its mean return μX, so the risk is low. If σX is large, the return can vary over a wide range, so the risk is high. In general, if two investments have the same mean return, the one with the smaller standard deviation is preferable, since it earns the same return on the average with lower risk. Example 2.73 An investor has $200 to invest. She will invest $100 in each of two investments. Let X and Y denote the returns on the two investments. Assume that μX = μY = $5, σX = σY = $2, and ρX,Y = 0.5. Find the mean and standard deviation of the total return on the two investments. Solution The total return is X + Y. The mean is Using Equation (2.77), the standard deviation is . Now Cov(X,Y) = ρX,YσXσY = (0.5)(2)(2) = 2. Therefore. It is instructive to compare the result of Example 2.73 with Page 152 occur if the entire $200 were invested in a single investment. Example 2.74 Assume that the investor in Example 2.73 invests the entire $200 in one of the two investments. Find the mean and standard deviation of the return. Solution Assume the investor invests in the investment whose return is X. Since $200 is invested, rather than $100, the return will be 2X. The mean return is The standard deviation is Comparing the results of Examples 2.73 and 2.74 shows that the mean returns of the two investment strategies are the same, but the standard deviation (i.e., risk) is lower when the capital is divided between two investments. This is the principle of diversification. When two investments are available whose returns have the same mean and the same risk, it is always advantageous to divide one's capital between them, rather than to invest in only one of them. Proof that μaX+b = aμX + b We assume that X is a continuous random variable with probability density function f(x). Then the proof in the case that X is a discrete random variable is similar, with integrals replaced by sums. Proof that μaX+bY = aμX + bμY Let X and Y be jointly continuous with joint density f(x,y) and marginal densities fX(x) and fY(y). Let a and b be constants. Then Page 153 The proof in the case that X and Y are jointly discrete is similar, with integrals replaced by sums. Proof that We will use the notation E(X) interchangeably with μX, E(Y) interchangeably with μY, and so forth. Let Y = aX + b. Then Proof that We will use the notation E(X) interchangeably with μX, E(Y) interchangeably with μY, and so forth. Proof of the equivalence of Equations (2.70) and (2.71) We will use the notation E(X) interchangeably with μX, E(Y) interchangeably with μY, and so forth. We will show that Now Page 154 Proof that if X and Y are independent, then X and Y are uncorrelated Let X and Y be independent random variables. We will show that μXY = μXμY, from which it will follow that Cov(X,Y) = μXY = 0. We will assume that X and Y are jointly continuous with joint density f(x,y) and marginal densities fX(x) and fY(y). The key to the proof is the fact that since X and Y are independent, f(x,y) = fX(x)fY(y). The proof in the case that X and Y are jointly discrete is similar, with the integrals replaced by sums. Exercises for Section 2.6 1. In a certain community, levels of air pollution may exceed federal standards for ozone or for particulate matter on some days. For a particular summer season, let X be the number of days on which the ozone standard is exceeded and let Y be the number of days on which the particulate matter standard is exceeded. Assume that the joint probability mass function of X and Y is given in the following table: y 0 1 2 0 0.10 0.17 0.06 y 1 0.13 0.23 0.14 2 0.05 0.08 0.04 a. Find P(X = 1and Y = 1). b. Find P(X ≥ 1and Y < 2). c. Find P(X < 1). d. Find P(Y ≥ 1). e. Find the probability that the standard for ozone is exceeded at least once. f. Find the probability that neither standard is ever exceeded. 2. Refer to Exercise 1. a. Find the marginal probability mass function pX(x). b. Find the marginal probability mass function pY(y). c. Find μX. d. Find μY. Page 155 e. Find σX. f. Find σY. g. Find Cov(X, Y). h. Find ρX,Y. i. Are X and Y independent? Explain. 3. Refer to Exercise 1. a. Find the conditional probability mass function pY|X(y|0). b. Find the conditional probability mass function pX|Y(x|1). c. Find the conditional expectation E(Y | X = 0). d. Find the conditional expectation E(X | Y = 1). 4. In a piston assembly, the specifications for the clearance between piston rings and the cylinder wall are very tight. In a lot of assemblies, let X be the number with too little clearance and let Y be the number with too much clearance. The joint probability mass function of X and Y is given in the table below: y x 0 1 2 3 0 0.15 0.12 0.11 0.10 1 0.12 0.07 0.05 0.03 a. b. c. d. e. Find P(X = 0 and Y = 0). f. Find the marginal probability mass function of X. Find the marginal probability mass function of Y. Are X and Y independent? Explain. Find μX and μY. Find σX and σY. 2 0.11 0.05 0.04 0.02 3 0.10 0.04 0.02 0.01 5. Refer to Exercise 4. a. Find pY|X(y | 2). b. Find pX|Y(x | 2). c. Find the conditional expectation E(Y | X = 2). d. Find the conditional expectation E(X | Y = 2). 6. Refer to Example 4. The total number of assemblies that fail to meet specifications is X + Y. a. Find μX+Y. b. Find σX+Y. c. Find P(X + Y = 3). 7. Refer to Exercise 4. Assume that the cost of repairing an assembly whose clearance is too little is $2, and the cost of repairing an assembly whose clearance is too much is $3. a. Express the total cost of repairs in terms of X and Y. b. Find the mean of the total cost of repairs. c. Find the standard deviation of the total cost of repairs. 8. The number of customers in line at a supermarket express checkout counter is a random variable whose probability mass function is given in the following table. x 0 1 2 3 4 5 p(x) 0.10 0.250.30 0.20 0.10 0.05 For each customer, the number of items to be purchased is a random variable with probability mass function y 1 2 3 4 5 6 p(y) 0.05 0.150.25 0.30 0.15 0.10 9. Let X denote the number of customers in line, and let Y denote the total number of items purchased by all the customers in line. Assume the number of items purchased by one customer is independent of the number of items purchased by any other customer. a. Find P(X = 2 and Y = 2). b. Find P(X = 2 and Y = 6). c. Find P(Y = 2). 10. The General Social Survey asked a sample of adults how many siblings (brothers and sisters) they had (X) and also how many children they had (Y). We show results for those who had no more than 4 children and no more than 4 siblings. Assume that the joint probability mass function is given in the following contingency table. Page 156 x 0 1 2 3 4 a. 0 0.03 0.09 0.09 0.04 1 0.01 0.05 0.05 0.04 0.03 y 2 0.08 0.07 0.07 0.04 0.04 The marginal probability mass function of X. 3 0.01 0.03 0.04 0.01 0.03 4 0.01 0.02 0.02 0.02 0.02 b. c. d. Are X and Y independent? Explain. Find μX and μY. Find σx and σY. f. Find Cov(X, Y). g. Find ρX,Y. 11. Refer to Exercise 9. a. Find μX+Y. b. Find σX+Y. c. Find P(X + Y = 5). 11. Refer to Exercise 9. a. Find the conditional probability mass function pY|X(y | 4) b. Find the conditional probability mass function pY|X(x | 3) c. Find the conditional expectation E(Y | X = 4). d. Find the conditional expectation E(X | Y = 3). 12. Automobile engines and transmissions are produced on assembly lines, and are inspected for defects after they come off their assembly lines. Those with defects are repaired. Let X represent the number of engines, and Y the number of transmissions that require repairs in a one-hour time interval. The joint probability mass function of X and Y is as follows: y x 0 1 2 3 0 0.13 0.12 0.02 0.01 1 0.10 0.16 0.06 0.02 2 0.07 0.08 0.08 0.02 3 0.04 0.04 0.02 13. Refer to Exercise 12. Let Z = X + Y represent the total number of repairs needed. a. Find μZ. b. Find σZ. c. Find P(Z = 2). 14. Refer to Exercise 12. Assume that the cost of repairing an engine is $50, and the cost of a transmission repair is $100. Let T represent the total cost of repairs during a one-hour time interval. a. Find μT. b. Find σT. c. Find P(T = 250). 15. Refer to Exercise 12. a. Find the conditional probability mass function pY|X(y | 3). b. Find the conditional probability mass function pX|Y(x | 1). c. Find the conditional expectation E(Y | X = 3). d. Find the conditional expectation E(X | Y = 1). 16. Refer to Example 2.54. a. Find the marginal probability density function fY(y). c. Are X and Y independent? Explain. 17. Refer to Example 2.54. a. Find Cov(X,Y). b. Find ρX,Y. 18. A production facility contains two machines that are used to rework items that are initially defective. Let X be the number of hours that the first machine is in use, and let Y be the number of hours that the second machine is in use, on a randomly chosen day. Page 157 Assume that X and Y have joint probability density function given by a. What is the probability that both machines are in operation for more than half an hour? Find the marginal probability density functions fX(x) and fY(y). c. Are X and Y independent? Explain. 19. Refer to Exercise 18. a. Find Cov(X, Y). b. Find ρX,Y. c. Find the conditional probability density function fY|X(y | 0.5). d. Find the conditional expectation E(Y | X = 0.5). 20. The lifetimes, in months, of two components in a system, denoted X and Y, have joint probability density function a. What is the probability that both components last longer than one month? Find the marginal probability density functions fX(x) and fY(y). c. Are X and Y independent? Explain. 21. The lifetime of a certain component, in years, has probability density function Two such components, whose lifetimes are independent, are available. As soon as the first component fails, it is replaced with the second component. Let X denote the lifetime of the first component, and let Y denote the lifetime of the second component. a. Find the joint probability density function of X and Y. b. Find P(X ≤ 1 and Y > 1). c. Find μX. d. Find μX+Y. Find P(X + Y ≤ 2). (Hint: Sketch the region of the plane where x + y ≤ 2, and then integrate the joint probability density function over that region.) 22. Here are two random variables that are uncorrelated but not independent. Let X and Y have the following joint probability mass function: x −1 0 1 y 0 1 p(x,y) 1/3 1/3 1/3 a. Use the definition of independence on page 141 to show that X and Y are not independent (in fact Y = |X|, so Y is actually a function of X). b. Show that X and Y are uncorrelated. 23. An investor has $100 to invest, and two investments between which to divide it. If she invests the entire amount in the first investment, her return will be X, while if she invests the entire amount in the second investment, her return will be Y. Both X and Y have mean $6 and standard deviation (risk) $3. The correlation between X and Y is 0.3. a. Express the return in terms of X and Y if she invests $30 in the first investment and $70 in the second. b. Find the mean return and the risk if she invests $30 in the first investment and $70 in the second. c. Express the return in terms of X and Y if she invests $K in the first investment and $(100 − K) in the second. d. Find the value of K that minimizes the risk in part (c). e. Prove that the value of K that minimizes the risk in part (c) is the same for any . correlation ρX,Y with ρX,Y ≠ 1. 24. The height H and radius R (in cm) of a cylindrical can are random with joint probability density function The volume of a can is V = πR2H. Find μV. 25. Let R denote the resistance of a resistor that is selected at random from a population of resistors that are labeled 100 Ω. The true population mean resistance is μR = 100 Page 158 Ω, and the population standard deviation is σR = 2 Ω. The resistance is measured twice with an ohmmeter. Let M1 and M2 denote the measured values. Then M1 = R + E1 and M2 = R + E2,where E1 and E2 are the errors in the measurements. Suppose that E1 and E2 are random with and . Further suppose that E1, E2, and R are independent. a. Find and . b. Show that . c. Show that . d. Use the results of (b) and (c) to show that . e. Find . 26. If X is a random variable, prove that Cov(X,X) = . 27. Let X and Y be random variables, and a and b be constants. a. Prove that Cov(aX, bY) = ab Cov(X,Y). b. Prove that if a > 0and b > 0, then ρaX,bY = ρX,Y. Conclude that the correlation coefficient is unaffected by changes in units. 28. Let X, Y, and Z be jointly distributed random variables. Prove that Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z). (Hint: Use Equation 2.71.) 29. Let X and Y be jointly distributed random variables. This exercise leads you through a proof of the fact that −1 ≤ ρX,Y ≤ 1. a. Express the quantity V(X − (σX/σY)Y) in terms of σX, σY, and Cov(X, Y). b. Use the fact that V(X − (σX/σY)Y) ≥ 0 and Cov(X, Y) = ρX,YσXσY to show that ρX,Y ≤ 1. c. Repeat parts (a) and (b) using V(X + (σX/σY)Y) to show that ρX,Y ≥ −1. 30. The oxygen equivalence number of a weld is a number that can be used to predict properties such as hardness, strength, and ductility. The article "Advances in Oxygen Equivalence Equations for Predicting the Properties of Titanium Welds" (D. Harwig, W. Ittiwattana, and H. Castner, The Welding Journal, 2001:126s–136s) presents several equations for computing the oxygen equivalence number of a weld. An equation designed to predict the strength of a weld is X = 1.12C + 2.69N + O − 0.21 Fe, where X is the oxygen equivalence, and C, N, O, and Fe are the amounts of carbon, nitrogen, oxygen, and iron, respectively, in weight percent, in the weld. Suppose that for welds of a certain type, μC = 0.0247, μN = 0.0255, μO = 0.1668, μFe = 0.0597, σC = 0.0131, σN = 0.0194, σO = 0.0340, and σFe = 0.0413. Furthermore assume that correlations are given by ρC,N = −0.44, ρC,O = 0.58, ρC,Fe = 0.39, ρN,O = −0.32, ρN,Fe = 0.09, and ρO,Fe = −0.35. a. Find μX. b. Find Cov(C, N), Cov(C, O), Cov(C, Fe), Cov(N, O), Cov(N, Fe), and Cov(O, Fe). Find σX. 31. Refer to Exercise 30. An equation to predict the ductility of a titanium weld is Y = 7.84C + 11.44N + O − 1.58Fe, where Y is the oxygen equivalence used to predict ductility, and C, N, O, and Fe are the amounts of carbon, nitrogen, oxygen, and iron, respectively, in weight percent, in the weld. Using the means, standard deviations, and correlations presented in Exercise 30, find μY and σY. 32. Let X and Y be jointly continuous with joint probability density function f(x, y) and marginal densities fX(x) and fY(y). Suppose that f(x, y) = g(x)h(y) where g(x) is a function of x alone, h(y) is a function of y alone, and c is a nonzero constant. a. Show that there exists a positive constant c such that fX(x) = cg(x) and fY(y) = (1/c)h(y). b. Use part (a) to show that X and Y are independent. 33. Let a, b, c, d be any numbers with a < b and c < d. Let k be a constant, and let X and Y be jointly continuous random variables with joint probability density function In other words, f(x,y) is constant on the rectangle a < x < b and c < y < d, and zero outside of the rectangle. a. Show that . b. Show that the marginal density of X is fX(x) = 1/(b − a) for a < x < b. c. Show that the marginal density of Y is fY(y) = 1/(d − c) for c < y < d. d. Use parts (a), (b), and (c) to show that X and Y are independent. Page 159 Supplementary Exercises for Chapter 2 1. A system consists of four components connected as shown. 2. 3. 4. 5. 6. 7. Assume A, B, C, and D function independently. If the probabilities that A, B, C, and D fail are 0.1, 0.2, 0.05, and 0.3, respectively, what is the probability that the system functions? A fair coin is tossed until a head appears. What is the probability that more than three tosses are necessary? Silicon wafers are used in the manufacture of integrated circuits. Of the wafers manufactured by a certain process, 10% have resistances below specification and 5% have resistances above specification. a. What is the probability that the resistance of a randomly chosen wafer does not meet the specification? b. If a randomly chosen wafer has a resistance that does not meet the specification, what is the probability that it is too low? Two production lines are used to pack sugar into 5 kg bags. Line 1 produces twice as many bags as does line 2. One percent of the bags from line 1 are defective in that they fail to meet a purity specification, while 3% of the bags from line 2 are defective. A bag is randomly chosen for inspection. a. What is the probability that it came from line 1? b. What is the probability that it is defective? c. If the bag is defective, what is the probability that it came from line 1? d. If the bag is not defective, what is the probability that it came from line 1? Four bricks are sampled from a large load and the crushing strength of each is measured to determine whether it meets a specification. If any of the four fail to meet the specification, the load will be returned. If in fact 10% of the bricks in the load fail to meet the specification, what is the probability that the load will be returned? In a certain type of automobile engine, the cylinder head is fastened to the block by 10 bolts, each of which should be torqued to 60 N · m. Assume that the torques of the bolts are independent. a. If each bolt is torqued correctly with probability 0.99, what is the probability that all the bolts on a cylinder head are torqued correctly? b. The goal is for 95% of the engines to have all their bolts torqued correctly. What must be the probability that a bolt is torqued correctly in order to reach this goal? An electronic message consists of a string of bits (0s and 1s). The message must pass through two relays before being received. At each relay the probability is 0.1 that the bit will be reversed before being relayed (i.e., a 1 will be changed to a 0, or a 0 to a 1). Find the probability that the value of a bit received at its final destination is the same as the value of the bit that was sent. The reading given by a thermometer calibrated in ice water (actual temperature 0°C) is a random variable with probability density function where k is a constant. a. Find the value of k. b. What is the probability that the thermometer reads above 0°C? c. What is the probability that the reading is within 0.25°C of the actual temperature? d. What is the mean reading? e. What is the median reading? f. What is the standard deviation? Two dice are rolled. Given that two different numbers have come up, what is the probability that one of the dice comes up 6? Page 160 10. In a lot of 10 components, 2 are sampled at random for inspection. Assume that in fact exactly 2 of the 10 components in the lot are defective. Let X be the number of sampled components that are defective. a. Find P(X = 0). b. Find P(X = 1). c. Find P(X = 2). d. Find the probability mass function of X. e. Find the mean of X. 11. There are two fuses in an electrical device. Let X denote the lifetime of the first fuse, and let Y denote the lifetime of the second fuse (both in years). Assume the joint probability density function of X and Y is a. Find P(X ≤ 2and Y ≤ 3). b. Find the probability that both fuses last at least 3 years. c. Find the marginal probability density function of X. e. Are X and Y independent? Explain. 12. Let A and B be events with P(A) = 0.3and P(A ∪ B) = 0.7. a. For what value of P(B) will A and B be mutually exclusive? b. For what value of P(B) will A and B be independent? 13. A snowboard manufacturer has three plants, one in the eastern United States, one in the western United States, and one in Canada. Production records show that the U.S. plants each produced 10,000 snowboards last month, while the Canadian plant produced 8000 boards. Of all the boards manufactured in Canada last month, 4% had a defect that caused the boards to delaminate prematurely. Records kept at the U.S. plants show that 3% of the boards manufactured in the eastern United States and 6% of the boards manufactured in the western United States had this defect as well. a. What proportion of the boards manufactured last month were defective? b. What is the probability that a snowboard is defective and was manufactured in Canada? c. Given that a snowboard is defective, what is the probability that it was manufactured in the United States? 14. The article "Traps in Mineral Valuations—Proceed With Care" (W. Lonergan, Journal of the Australasian Institute of Mining and Metallurgy, 2001:18–22) models the value (in millions of dollars) of a mineral deposit yet to be mined as a random variable X with probability mass function p(x) given by p(10) = 0.40, p(60) = 0.50, p(80) = 0.10, and p(x) = 0 for values of x other than 10, 60, or 80. a. Is this article treating the value of a mineral deposit as a discrete or a continuous random variable? b. Compute μX. c. Compute σX. d. The project will be profitable if the value is more than $50 million. What is the probability that the project is profitable? 15. Six new graduates are hired by an engineering firm. Each is assigned at random to one of six cubicles arranged in a row in the back of the room that houses the engineering staff. Two of the graduates are Bill and Cathy. What is the probability that they are assigned adjacent cubicles? 16. A closet contains four pairs of shoes. If four shoes are chosen at random, what is the probability that the chosen shoes do not contain a pair? 17. Let X and Y be independent random variables with μX = 2, σX = 1, μY = 2, and σY = 3. Find the means and variances of the following quantities. a. X + Y b. X − Y c. 3X + 2Y d. 5Y − 2X 19. A steel manufacturer is testing a new additive for manufacturing an alloy of steel. The joint probability mass function of tensile strength (in thousands of pounds/in2) and additive concentration is Concentration of Additive 0.02 0.04 0.06 0.08 Tensile Strength 100 150 200 0.05 0.06 0.11 0.01 0.08 0.10 0.04 0.08 0.17 0.04 0.14 0.12 What are the marginal probability mass functions for X (additive concentration) and Y (tensile strength)? b. Are X and Y independent? Explain. c. Given that a specimen has an additive concentration of 0.04, what is the probability that its strength is 150 or more? d. Given that a specimen has an additive concentration of 0.08, what is the probability that its tensile strength is greater than 125? e. A certain application calls for the tensile strength to be 175 or more. What additive concentration should be used to make the probability of meeting this specification the greatest? 20. Refer to Exercise 19. a. Find μX. b. Find μY. c. Find σX. d. Find σY. e. Find Cov(X,Y). f. Find ρX,Y. 21. Refer to Exercise 19. a. Compute the conditional mass function pY|X(y | 0.06). b. Compute the conditional mass function pX|Y(x | 100). c. Compute the conditional expectation E(Y | X = 0.06). d. Compute the conditional expectation E(X | Y = 100). 22. A certain plant runs three shifts per day. Of all the items produced by the plant, 50% of them are produced on the first shift, 30% on the second shift, and 20% on the third shift. Of all the items produced on the first shift, 1% are defective, while 2% of the items produced on the second shift and 3% of the items produced on the third shift are defective. a. An item is sampled at random from the day's production, and it turns out to be defective. What is the probability that it was manufactured during the first shift? b. An item is sampled at random from the day's production, and it turns out not to be defective. What is the probability that it was manufactured during the third shift? 23. The article "Uncertainty and Climate Change" (G. Heal and B. Kriström, Environmental and Resource Economics, 2002:3–39) considers three scenarios, labeled A, B, and C, for the impact of global warming on income. For each scenario, a probability mass function for the loss of income is specified. These are presented in the following table. Loss (%) 0 2 5 10 15 20 25 a. Scenario A 0.65 0 0.2 0 0.1 0 0.05 Probability Scenario B 0.65 0 0.24 0 0.1 0.01 0 Scenario C 0.65 0.24 0 0.1 0 0.01 0 Compute the mean and standard deviation of the loss under scenario A. b. Compute the mean and standard deviation of the loss under scenario B. Page 162 c. Compute the mean and standard deviation of the loss under scenario C. d. Under each scenario, compute the probability that the loss is less than 10%. 24. Refer to Exercise 23. Assume that the probabilities that each of the three scenarios occurs are P(A) = 0.20, P(B) = 0.30, and P(C) = 0.50. a. Find the probability that scenario A occurs and that the loss is 5%. b. Find the probability that the loss is 5%. c. Find the probability that scenario A occurs given that the loss is 5%. 25. A box contains 75 lightbulbs, 60 watt lightbulbs, and three-hundred-watt lightbulbs. Two bulbs are selected at random from the box. Let X represent the number of 75 W bulbs selected, and let Y represent the number of 60 W bulbs selected. a. Find the joint probability mass function of X and Y. b. Find μX. c. Find μY. d. Find σX. e. Find σY. f. Find Cov(X,Y). g. Find ρX,Y. 26. A stock solution of hydrochloric acid (HCl) supplied by a certain vendor contains small amounts of several impurities, including copper and nickel. Let X denote the amount of copper and let Y denote the amount of nickel, in parts per ten million, in a randomly selected bottle of solution. Assume that the joint probability density function of X and Y is given by a. Find the value of the constant c so that f(x,y) is a joint density function. b. Compute the marginal density function fX(x). c. Compute the conditional density function fY|X(y | x). d. Compute the conditional expectation E(Y | X = 0.4). e. Are X and Y independent? Explain. 27. Refer to Exercise 26. a. Find μX. b. c. d. Find σX. f. Find Cov(X, Y). g. Find ρX,Y. 28. A fair coin is tossed five times. Which sequence is more likely, HTTHH or HHHHH? Or are they equally likely? Explain. 29. A penny and a nickel are tossed. The penny has probability 0.4 of coming up heads, and the nickel has probability 0.6 of coming up heads. Let X = 1 if the penny comes up heads, and let X = 0 if the penny comes up tails. Let Y = 1 if the nickel comes up heads, and let Y = 0 if the nickel comes up tails. a. Find the probability mass function of X. b. Find the probability mass function of Y. c. Is it reasonable to assume that X and Y are independent? Why? d. Find the joint probability mass function of X and Y. 30. Two fair dice are rolled. Let X represent the number on the first die, and let Y represent the number on the second die. Find μXY. 31. A box contains three cards, labeled 1, 2, and 3. Two cards are chosen at random, with the first card being replaced before the second card is drawn. Let X represent the number on the first card, and let Y represent the number on the second card. a. Find the joint probability mass function of X and Y. b. Find the marginal probability mass functions pX(x) and pY(y). c. Find μX and μY. d. Find μXY. e. Find Cov(X,Y). 32. Refer to Exercise 31. Assume the first card is not replaced before the second card is drawn. a. Find the joint probability mass function of X and Y. b. Find the marginal probability mass functions pX(x) and pY(y). c. Find μX and μY. Page 163 d. Find μXY. e. Find Cov(X,Y). 33. This exercise will lead you through a proof of Chebyshev's inequality. Let X be a continuous random variable with probability density function f(x). Suppose that P(X < 0) = 0, so f(x) = 0 for x ≤ 0. a. Show that . b. Let k > 0 be a constant. Show that μX ≥ c. Use part (b) to show that P(X ≥ k) ≤ μX/k.This is called Markov's inequality. It is true for discrete as well as for continuous random variables. Let Y be any random variable with mean μY and variance . Let X = (Y − μY)2. Show that . Let k > 0 be a constant. Show that Use part (e) along with Markov's inequality to prove Chebyshev's inequality: P(|Y − μY| ≥ kσY) ≤ 1/k2. d. e. f. 34. A circle is drawn with radius R, where μR = 10 and πR2. Find μA. . The area of the circle is A = 35. Blood is taken from each of n individuals to be tested for a certain disease. Rather than test each sample separately, a pooled method is used in an attempt to reduce the number of tests needed. Part of each blood sample is taken, and these parts are combined to form a pooled sample. The pooled sample is then tested. If the result is negative, then none of the n individuals has the disease, and no further tests are needed. If the pooled sample tests positive, then each individual is tested to see which of them have the disease. a. Let X represent the number of tests that are carried out. What are the possible values of X? b. Assume that n = 4 individuals are to be tested, and the probability that each has the disease, independent of the others, is p = 0.1. Find μX. c. d. Repeat part (b) with n = 6and p = 0.2. Express μX as a function of n and p. e. The pooled method is more economical than performing individual tests if μX < n. Suppose n = 10. For what values of p is the pooled method more economical than performing n individual tests? Page 164 Chapter 3 Propagation of Error Introduction Measurement is fundamental to scientific work. Scientists and engineers often perform calculations with measured quantities, for example, computing the density of an object by dividing a measurement of its mass by a measurement of its volume, or computing the area of a rectangle by multiplying measurements of its length and width. Any measuring procedure contains error. As a result, measured values generally differ somewhat from the true values that are being measured. When a calculation is performed with measurements, the errors in the measurements produce an error in the calculated value. We say that the error is propagated from the measurements to the calculated value. If we have some knowledge concerning the sizes of the errors in measurements such as the length and width of a rectangle, there are methods for obtaining knowledge concerning the likely size of the error in a calculated quantity such as the area. The subject of propagation of error concerns these methods and is the topic of this chapter. 3.1 Measurement Error A geologist is weighing a rock on a scale. She weighs the rock five times and obtains the following measurements (in grams): 251.3 252.5 250.8 251.1 250.4 The measurements are all different, and it is likely that none of them is equal to the true mass of the rock. The difference between a measured value and the true value is called the error in the measured value. Any measuring procedure contains many sources of error. For example, assume that the rock measurements were read from a dial. If the scale was not calibrated properly, this will pull each measurement away from the true value by some fixed amount. Thus Page 165 imperfect calibration contributes errors of equal magnitude to each measurement. Interpolation between graduation marks on the dial is another source of error. The magnitude of the error due to interpolation is likely to vary from measurement to measurement and is likely to be positive for some measurements and negative for others. It may be reasonable to assume that interpolation errors average out to zero in the long run. In general, we can think of the error in a measurement as being composed of two parts, the systematic error, or bias, and the random error. The bias is the part of the error that is the same for every measurement. The random error, on the other hand, varies from measurement to measurement, and averages out to zero in the long run. Some sources of error contribute both to bias and to random error. For example, consider parallax error. Parallax is the difference in the apparent position of the dial indicator when observed from different angles. The magnitude of the parallax error in any particular measurement depends on the position of the observer relative to the dial. Since the position will vary somewhat from reading to reading, parallax contributes to random error. If the observer tends to lean somewhat to one side rather than another, parallax will contribute to bias as well. Any measurement can be considered to be the sum of the true value plus contributions from each of the two components of error: (3.1) Since part of the error is random, it is appropriate to use a statistical model to study measurement error. We model each measured value as a random variable, drawn from a population of possible measurements. The mean μ of the population represents that part of the measurement that is the same for every measurement. Therefore, μ is the sum of the true value and the bias. The standard deviation σ of the population is the standard deviation of the random error. It represents the variation that is due to the fact that each measurement has a different value for its random error. Intuitively, σ represents the size of a typical random error. We are interested in two aspects of the measuring process. First is its accuracy. The accuracy is determined by the bias, which is the difference between the mean measurement μ and the true value being measured. The smaller the bias, the more accurate the measuring process. If the mean μ is equal to the true value, so that the bias is 0, the measuring process is said to be unbiased. The other aspect of the measuring process that is of interest is the precision. Precision refers to the degree to which repeated measurements of the same quantity tend to agree with each other. If repeated measurements come out nearly the same every time, the precision is high. If they are widely spread out, the precision is low. The precision is therefore determined by the standard deviation σ of the measurement process. The smaller the value of σ, the more precise the measuring process. Engineers and scientists often refer to σ as the random uncertainty or statistical uncertainty in the measuring process. We will refer to σ simply as the uncertainty. Page 166 Summary ■ ■ A measured value is a random variable with mean μ and standard deviation σ. The bias in the measuring process is the difference between the mean measurement and the true value: ■ ■ ■ The uncertainty in the measuring process is the standard deviation σ. The smaller the bias, the more accurate the measuring process. The smaller the uncertainty, the more precise the measuring process. When reporting a measured value, it is important to report an estimate of the bias and uncertainty along with it, in order to describe the accuracy and precision of the measurement. It is in general easier to estimate the uncertainty than the bias. Figures 3.1 and 3.2 illustrate the reason for this. Figure 3.1 illustrates a hypothetical experiment involving repeated measurements, under differing conditions regarding bias and uncertainty. The sets of measurements in Figure 3.1a and b are fairly close together, indicating that the uncertainty is small. The sets of measurements in Figure 3.1a and c are centered near the true value, indicating that the bias is small. FIGURE 3.1 (a) Both bias and uncertainty are small. (b) Bias is large; uncertainty is small. (c) Bias is small; uncertainty is large. (d) Both bias and uncertainty are large. FIGURE 3.2 We can estimate the uncertainty from the set of repeated measurements, but without knowing the true value, we cannot estimate the bias. In real life, of course, we do not know the true value being measured. Thus the plot of measurements shown in Figure 3.1 would look like Figure 3.2. We can still determine that the sets of measurements in Figure 3.2a and b have smaller uncertainty. But without additional information about the true value, we cannot estimate the bias. We conclude from Figures 3.1 and 3.2 that uncertainty can be estimated from repeated measurements, but in order to estimate the bias, we must have additional information about the true value. We might obtain this additional information, for example, by repeatedly measuring a standard quantity whose true value is known and estimating the bias to be the difference between the average of the measurements and the known true value. Another way to estimate the bias would be to compare the average of a large number of measurements to a measurement made with a more elaborate process for which the bias is known to be negligible. Estimating Page 167 the bias is essentially the process of calibration, for which information external to the measuring device is needed. Example 3.1 A laboratory sample of gas is known to have a carbon monoxide (CO) concentration of 50 parts per million (ppm). A spectrophotometer is used to take five independent measurements of this concentration. The five measurements, in ppm, are 51, 47, 53, 53, and 48. Estimate the bias and the uncertainty in a spectrophotometer measurement. Solution The five measurements are regarded as a random sample from the population of possible measurements. The bias is equal to the mean of this population minus the true value of 50. The uncertainty is the standard deviation of the population. We do not know the mean and standard deviation of the population, but we can approximate them with the mean and standard deviation of the sample. The mean of the five measurements is 50.4. Therefore we estimate the bias to be 50.4 − 50 = 0.4 ppm. The standard deviation of the five measurements is 2.8 ppm. Therefore we estimate the uncertainty in each measurement to be 2.8 ppm. Example 3.2 A different spectrophotometer is now used to measure the CO concentration in another gas sample. The true concentration in this sample is unknown. Five measurements are made (in ppm). They are 62, 63, 61, 62, and 59. Estimate the uncertainty in a measurement from this spectrophotometer. Can we estimate the bias? Solution The uncertainty in a single measurement is estimated with the sample standard deviation, which is 1.5 ppm. The sample mean is 61.4 ppm, but to estimate the bias, we would have to subtract the true concentration from this. Since we do not know the true concentration, we cannot estimate the bias. In practice, estimates of uncertainty are sometimes very crude. In Examples 3.1 and 3.2 we suggested estimating the uncertainty σ with the sample standard deviation of five measurements. Estimates based on small samples like this are sometimes widely off the mark. When Page 168 possible, it is best to base estimates of uncertainty on large samples. However, an estimate from a small sample is better than none at all. Summary Let X1, ..., Xn be independent measurements, all made by the same process on the same quantity. ■ The sample standard deviation s can be used to estimate the uncertainty. ■ Estimates of uncertainty are often crude, especially when based on small samples. ■ If the true value is known, the sample mean can be used to estimate the bias: Bias true value. ■ If the true value is unknown, the bias cannot be estimated from repeated measurements. An important example of bias estimation is the calibration of scales in supermarkets and other commercial establishments to ensure that they do not systematically over-or underweigh goods sold to customers. This calibration procedure follows a chain of comparisons with external standards, beginning at the county level and ending up near Paris, France, where the world's ultimate standard for weight (technically mass) is located. This is the International Prototype Kilogram, a platinum–iridium cylinder whose mass is by definition exactly 1 kg. A replica of The Kilogram, located at the National Institute of Standards and Technology in Washington, serves as the standard for measures in the United States. Use of this replica, rather than The Kilogram, introduces a bias into every measure of weight in the United States. By comparing the U.S. replica to The Kilogram, this bias has been estimated to be −1.9 × 10−8 kg. That is, the U.S. replica appears to be lighter than The Kilogram by about 19 parts in a billion. For this reason, all weight measurements made at the National Institute of Standards and Technology are adjusted upward by 19 parts in a billion to compensate. Note that this adjustment factor could not have been estimated by repeated weighing of the replica; comparison with an external standard was required. From here on we will assume, unless otherwise stated, that bias has been reduced to a negligible level. We will describe measurements in the form (3.2) where σ is the uncertainty in the process that produced the measured value. Expression (3.2) has the form a ± b, where a and b are numbers. It is important to realize that expressions containing the symbol ± can have many meanings. The meaning here is that a is a measured value and b is the uncertainty in a. Some people use a ± b to indicate that b is the maximum value for the error, or that b is a multiple of the uncertainty, typically two or three times the uncertainty. Yet another meaning will be presented in Chapter 5, where we will use the notation a ± b to denote a confidence interval, which is an interval computed in a way so as to be likely to contain the true value. Whenever you encounter the symbol ±, you should be Page 169 sure to understand the context in which it is used. Example 3.3 The spectrophotometer in Example 3.1 has been recalibrated, so we may assume that the bias is negligible. The spectrophotometer is now used to measure the CO concentration in another gas sample. The measurement is 55.1 ppm. How should this measurement be expressed? Solution From the repeated measurements in Example 3.1, the uncertainty in a measurement from this instrument was estimated to be 2.8 ppm. Therefore we report the CO concentration in this gas sample as 55.1 ± 2.8 ppm. Exercises for Section 3.1 1. 2. 3. 4. 5. The boiling point of water is measured four times. The results are 110.01°C, 110.02°C, 109.99°C, and 110.01°C. Which of the following statements best describes this measuring process? i. Accurate but not precise ii. Precise but not accurate iii. Neither accurate nor precise iv. Both accurate and precise Two thermometers are calibrated by measuring the freezing point of glacial acetic acid, which is 16.6°C. Equal numbers of measurements are taken with each thermometer. The result from the first thermometer is 16.4 ± 0.2°C and the result from the second thermometer is 16.8 ± 0.1°C. a. Is it possible to tell which thermometer is more accurate? If so, say which one. If not, explain why. b. Is it possible to tell which thermometer is more precise? If so, say which one. If not, explain why. The weight of an object is given as 67.2 ± 0.3 g. True or false: a. The weight was measured to be 67.2 g. b. The true weight of the object is 67.2 g. c. The bias in the measurement is 0.3 g. d. The uncertainty in the measurement is 0.3 g. For some measuring processes, the uncertainty is approximately proportional to the value of the measurement. For example, a certain scale is said to have an uncertainty of ±2%. An object is weighed on this scale. a. Given that the reading is 100 g, express the uncertainty in this measurement in grams. b. Given that the reading is 50 g, express the uncertainty in this measurement in grams. A person stands on a bathroom scale. The reading is 150 lb. After the person gets off the 6. 7. scale, the reading is 2 lb. a. Is it possible to estimate the uncertainty in this measurement? If so, estimate it. If not, explain why not. b. Is it possible to estimate the bias in this measurement? If so, estimate it. If not, explain why not. A person gets on and off a bathroom scale four times. The four readings (in pounds) are 148, 151, 150, and 152. Each time after the person gets off the scale, the reading is 2 lb. a. Is it possible to estimate the uncertainty in these measurements? If so, estimate it. If not, explain why not. b. Is it possible to estimate the bias in these measurements? If so, estimate it. If not, explain why not. In a hypothetical scenario, the National Institute of Standards and Technology has received a new replica of The Kilogram. It is weighed five times. The measurements are as follows, in units of micro-grams above 1 kg: 114.3, 82.6, 136.4, 126.8, 100.7. a. Is it possible to estimate the uncertainty in these measurements? If so, estimate it. If not, explain why not. Page 170 b. Is it possible to estimate the bias in these measurements? If so, estimate it. If not, explain why not. 8. The Kilogram is now weighed five times on a different scale. The measurements are as follows, in units of micrograms above 1 kg: 25.6, 26.8, 26.2, 26.8, 25.4. a. Is it possible to estimate the uncertainty in these measurements? If so, estimate it. If not, explain why not. b. Is it possible to estimate the bias in these measurements? If so, estimate it. If not, explain why not. 9. A new and unknown weight is weighed on the same scale that was used in Exercise 8, and the measurement is 127 μg above 1 kg. Using the information in Exercise 8, is it possible to come up with a more accurate measurement? If so, what is it? If not, explain why not. 10. The article "Calibration of an FTIR Spectrometer" (P. Pankratz, Statistical Case Studies for Industrial and Process Improvement, SIAM-ASA, 1997: 19–38) describes the use of a spectrometer to make five measurements of the carbon content (in ppm) of a certain silicon wafer whose true carbon content was known to be 1.1447 ppm. The measurements were 1.0730, 1.0825, 1.0711, 1.0870, and 1.0979. a. Is it possible to estimate the uncertainty in these measurements? If so, estimate it. If not, explain why not. b. Is it possible to estimate the bias in these measurements? If so, estimate it. If not, explain why not. 11. The length of a rod was measured eight times. The measurements in centimeters, in the order they were taken, were 21.20, 21.22, 21.25, 21.26, 21.28, 21.30, 21.32, 21.35. a. Do these measurements appear to be a random sample from a population of possible measurements? Why or why not? b. Is it possible to estimate the uncertainty in these measurements? Explain. 3.2 Linear Combinations of Measurements Often we add constants to measurements, multiply measurements by constants, or add two or more measurements together. This section describes how uncertainties are affected by these arithmetic operations. Since measurements are random variables, and uncertainties are the standard deviations of these random variables, the results used to compute standard deviations of linear combinations of random variables can be applied to compute uncertainties in linear combinations of measurements. Results for independent random variables were presented in Section 2.5; more general results were presented in Section 2.6. In this section we apply these results to independent measurements. We discuss dependent measurements as well, at the end of the section. We begin by stating the basic results used to compute uncertainties in linear combinations of independent measurements, and then follow with some examples. If X is a measurement and c is a constant, then (3.3) If X1, ..., Xn are independent measurements and c1, ..., cn are constants, then (3.4) Page 171 Example 3.4 The radius of a circle is measured to be 3.0 ± 0.1 cm. Estimate the circumference and find the uncertainty in the estimate. Solution Let R denote the radius of the circle. The measured value of R is 3.0 cm, and the uncertainty is the standard deviation of this measurement, which is σR = 0.1 cm. The circumference is given by C = 2πR. The uncertainty in C is σC, the standard deviation of C. Since 2π is a constant, we have (using Equation 3.3) The circumference is 18.85 ± 0.63 cm. Example 3.5 An item is formed by placing two components end to end. The lengths of the components are measured independently, by a process that yields a random measurement with uncertainty 0.1 cm. The length of the item is estimated by adding the two measured lengths. Assume that the measurements are 4.10 cm and 3.70 cm. Estimate the length of the item and find the uncertainty in the estimate. Solution Let X be the measured length of the first component, and let Y be the measured length of the second component. The estimated length is 7.80 cm. The uncertainty is (using Equation 3.4 with c1 = c2 = 1) The estimated length is 7.80 ± 0.14 cm. Example 3.6 A surveyor is measuring the perimeter of a rectangular lot. He measures two adjacent sides to be 50.11 ± 0.05 m and 75.21 ± 0.08 m. These measurements are independent. Estimate the perimeter of the lot and find the uncertainty in the estimate. Solution Let X = 50.11 and Y = 75.21 be the two measurements. The perimeter is estimated by P = 2X + 2Y = 250.64 m, and the uncertainty in P is (using Equation 3.4) Page 172 The perimeter is 250.64 ± 0.19 m. Example 3.7 In Example 3.6, the surveyor's assistant suggests computing the uncertainty in P by a different method. He reasons that since P = X + X + Y + Y, then This disagrees with the value of 0.19 m calculated in Example 3.6. What went wrong? Solution What went wrong is that the four terms in the sum for P are not all independent. Specifically, X + X is not the sum of independent quantities; neither is Y + Y. In order to use Equation (3.4) to compute the uncertainty in P, we must express P as the sum of independent quantities, that is, P = 2X + 2Y, as in Example 3.6. Repeated Measurements One of the best ways to reduce uncertainty is to take several independent measurements and average them. The measurements in this case are a simple random sample from a population, and their average is the sample mean. Methods for computing the mean and standard deviation of a sample mean were presented in Sections 2.5 and 2.6. These methods can be applied to compute the mean and uncertainty in the average of independent repeated measurements. If X1, ..., Xn are n independent measurements, each with mean μ and uncertainty σ, then the sample mean is a measurement with mean (3.5) and with uncertainty (3.6) With a little thought, we can see how important these results are for applications. What these results say is that if we perform many independent measurements of the same quantity, then the average of these measurements has the same mean as each individual measurement, but the standard deviation is reduced by a factor equal to the square root of the sample size. Page 173 In other words, the average of several repeated measurements has the same accuracy as, and is more precise than, any single measurement. Example 3.8 The length of a component is to be measured by a process whose uncertainty is 0.05 cm. If 25 independent measurements are made and the average of these is used to estimate the length, what will the uncertainty be? How much more precise is the average of 25 measurements than a single measurement? Solution The uncertainty in the average of 25 measurements is . The uncertainty in a single measurement is 0.05 cm. The uncertainty in the average of 25 independent measurements is therefore less than that of a single measurement by a factor of 5, which is the square root of the number of measurements that are averaged. Thus the average of 25 independent measurements is five times more precise than a single measurement. Example 3.9 The mass of a rock is measured five times on a scale whose uncertainty is unknown. The five measurements (in grams) are 21.10, 21.05, 20.98, 21.12, and 21.05. Estimate the mass of the rock and find the uncertainty in the estimate. Solution Let represent the average of the five measurements, and let s represent the sample standard deviation. We compute and s = 0.0543 g. Using Equation (3.6), we would estimate the length of the component to be . We do not know σ, which is the uncertainty, or standard deviation, of the measurement process. However, we can approximate σ with s, the sample standard deviation of the five measurements. We therefore estimate the mass of the rock to be , or 21.06 ± 0.02 g. Example 3.10 In Example 3.6 two adjacent sides of a rectangular lot were measured to be X = 50.11 ± 0.05 m and Y = 75.21 ± 0.08 m. Assume that the budget for this project is sufficient to allow 14 more measurements to be made. Each side has already been measured once. One engineer suggests allocating the new measurements equally to each side, so that each will be measured eight times. A second engineer suggests using all 14 measurements on the longer side, since that side is measured with greater uncertainty. Estimate the uncertainty in filled with water, and then measuring the volume of water displaced. The density D is given by D = m/(V1 − V0), where m is the mass of the rock, V0 is the initial volume of water, and V1 is the volume of water plus rock. Assume the mass of the rock is 750 g, with negligible uncertainty, and that V0 = 500.0 ± 0.1 mL and V1 = 813.2 ± 0.1 mL. Estimate the density of the rock, and find the uncertainty in the estimate. The conversion of ammonium cyanide to urea is a second-order reaction. This means that the concentration C of ammonium cyanide at time t is given by 1/C = kt + 1/C0, where C0 is the initial concentration and k is the rate constant. Assume the initial concentration is known to be 0.1 mol/L exactly. Assume that time can be measured with negligible uncertainty. a. After 45 minutes, the concentration of ammonium cyanide is measured to be 0.0811 ± 0.0005 mol/L. Estimate the rate constant k, and find the uncertainty in the estimate. b. Use the result in part (a) to estimate the time when the concentration of ammonium cyanide will be 0.0750 mol/L, and find the uncertainty in this estimate. Convert the following absolute uncertainties to relative uncertainties. a. 20.9 ± 0.4 b. 15.1 ± 0.8 c. 388 ± 23 d. 2.465 ± 0.009 Convert the following relative uncertainties to absolute uncertainties. a. 48.41 ± 0.3% b. 991.7 ± 0.6% c. 0.011 ± 9% d. 7.86 ± 1% The acceleration g due to gravity is estimated by dropping an object and measuring the time it takes to travel a certain distance. Assume the distance s is known to be exactly 2.2 m. The time is measured to be t = 0.67 ± 0.02 s. Estimate g, and find the relative uncertainty in the estimate. (Note that g = 2s/t2.) 14. Refer to Exercise 4. Assume that T = 298.4 ± 0.2K. Estimate V, and find the relative uncertainty in the estimate. 15. Refer to Exercise 5. a. Assume g = 9.80 m/s2 exactly, and that L = 0.855 ± 0.005 m. Estimate T, and find the relative uncertainty in the estimate. b. Assume L = 0.855 m exactly, and that T = 1.856 ± 0.005 s. Estimate g, and find the relative uncertainty in the estimate. 16. Refer to Exercise 6. Assume that c = 448 J/kg°C and ΔQ = 1210 J are known with negligible uncertainty. Assume the mass is m = 0.75 ± 0.01 kg. Estimate ΔT, and find the relative uncertainty in the estimate. 17. Refer to Exercise 7. Estimate F, and find the relative uncertainty in the estimate, assuming that g = 9.80 m/s2 exactly and that a. d = 0.20 m and l = 35.0 m, both with negligible uncertainty, and h = 4.51 ± 0.03 m. Page 186 b. h = 4.51 m and l = 35.0 m, both with negligible uncertainty, and d = 0.20 ± 0.008 m. c. d = 0.20 m and h = 4.51 m, both with negligible uncertainty, and l = 35.00 ± 0.4 m. 18. Refer to Exercise 8. Assume the critical angle is measured to be 0.90 ± 0.01 rad. Estimate the refractive index and find the relative uncertainty in the estimate. 19. Refer to Exercise 9. Assume that the mass of the rock is 288.2 g with negligible uncertainty, the initial volume of water in the cylinder is 400 ± 0.1 mL, and the volume of water plus rock is 516 ± 0.2 mL. Estimate the density of the rock, and find the relative uncertainty in the estimate. 20. In a chemical reaction run at a certain temperature, the concentration C of a certain reactant at time t is given by 1/C = kt+1 /C0, where C0 is the initial concentration and k is the rate constant. Assume the initial concentration is known to be 0.04 mol/L exactly. Assume that time is measured with negligible uncertainty. a. After 30 s, the concentration C is measured to be 0.0038 ± 2.0 × 10−4 mol/L. Estimate the rate constant k, and find the relative uncertainty in the estimate. b. After 50 s, the concentration C is measured to be 0.0024 ± 2.0 × 10−4 mol/L. Estimate the rate constant k and find the relative uncertainty in the estimate. c. Denote the estimates of the rate constant k in parts (a) and (b) by and , respectively. The geometric mean uncertainty in this estimate. is used as an estimate of k. Find the relative 3.4 Uncertainties for Measurements Functions of Several Often we need to estimate a quantity as a function of several measurements. For example, we might measure the mass m and volume V of a rock and compute the density as D = m/V. In Example 3.15 we saw how to estimate the uncertainty in D when one of the quantities, in this case V, was measured with uncertainty while m was treated as a known constant. However, in practice we might need to estimate the uncertainty in D when both m and V are measured with uncertainty. In this section we will learn how to estimate the uncertainty in a quantity that is a function of several independent uncertain measurements. The basic formula is given here. If X1, X2, ..., Xn are independent measurements whose uncertainties are small, and if U = U(X1, X2, ..., Xn) is a function of X1, X2, ..., Xn, then (3.12) In practice, we evaluate the partial derivatives at the point (X1, X2, ..., Xn). Equation (3.12) is the multivariate propagation of error formula. It is important to note that it is valid only when the measurements X1, X2, ..., Xn are independent. A derivation of the formula is given at the end of the section. As in the case of one measurement, the uncertainties computed by the propagation of error formula are often only rough approximations. Page 187 Nonlinear functions of measurements are generally biased (see the discussion concerning functions of one measurement in Section 3.3). However, so long as the measurements X1, ..., Xn are unbiased, and the uncertainties are all small, the bias in U will usually be small enough to ignore. Exceptions to this rule, which are fairly unusual, can occur when some of the second- or higher-order partial derivatives of U with respect to the Xi are quite large. Of course, if one or more of X1, ..., Xn are substantially biased, then U may be substantially biased as well. These ideas are explored further in Exercise 23 in the Supplementary Exercises at the end of this chapter. We now present some examples that illustrate the use of multivariate propagation of error. Example 3.18 Assume the mass of a rock is measured to be m = 674.0 ± 1.0 g, and the volume of the rock is measured to be V = 261.0 ± 0.1 mL. Estimate the density of the rock and find the uncertainty in the estimate. Solution Substituting m = 674.0 g and V = 261.0 mL, the estimate of the density D is 674.0/261.0 = 2.582 g/mL. Since D = m/V, the partial derivatives of D are The uncertainty in D is therefore The density of the rock is 2.582 ± 0.004 g/mL. One of the great benefits of the multivariate propagation of error formula is that it enables one to determine which measurements are most responsible for the uncertainty in the final result. Example 3.19 illustrates this. Example 3.19 The density of the rock in Example 3.18 is to be estimated again with different equipment, in order to improve the precision. Which would improve the precision of the density estimate more: decreasing the uncertainty in the mass estimate to 0.5 g, or decreasing the uncertainty in the volume estimate to 0.05 mL? Page 188 Solution From Example 3.18, . We have a choice between having σm = 0.5 and σV = 0.1, or having σm = 1.0 and σV = 0.05. The first choice results in σD = 0.002 g/mL, while the second choice results in σD = 0.004 g/mL. It is better to reduce σm to 0.5 g. Example 3.20 Two resistors with resistances R1 and R2 are connected in parallel. The combined resistance R is given by R = (R1R2)/(R1 + R2). If R1 is measured to be 100 ± 10 Ω, and R2 is measured to be 20 ± 1 Ω, estimate R and find the uncertainty in the estimate. Solution The estimate of R is (100)(20)/(100 + 20) = 16.67 Ω. To compute σR, we first compute the partial derivatives of R: Now , and . Therefore The combined resistance is 16.67 ± 0.75 Ω. Example 3.21 In Example 3.20, the 100 ± 10 Ω resistor can be replaced with a more expensive 100 ± 1 Ω resistor. How much would this reduce the uncertainty in the combined resistance? Is it worthwhile to make the replacement? Solution Using the method of Example 3.20, the uncertainty in the combined resistance R with the new resistor would be There is not much reduction from the uncertainty of 0.75 Ω using the old resistor. Almost all the uncertainty in the combined resistance is due to the uncertainty in the 20 Ω resistor. The uncertainty in the 100 Ω resistor can be neglected for most practical purposes. There is little benefit in replacing this resistor. Page 189 Note that in Example 3.20, one component (the 100 Ω resistor) had larger uncertainty, both in absolute terms and relative to the measured value, than the other. Even so, Example 3.21 showed that the uncertainty in the combined resistance was only slightly affected by the uncertainty in this component. The lesson is that one cannot predict the impact that the uncertainties in individual measurements will have on the uncertainty in the final calculation from the magnitudes of those uncertainties alone. One must use the propagation of error formula. Uncertainties for Functions of Dependent Measurements If X1, X2, ..., Xn are not independent, the uncertainty in a function U = U(X1, X2, ..., Xn) can be estimated if the covariance of each pair (Xi, Xj) is known. (Covariance is discussed in Section 2.6.) In many situations, the covariances are not known. In these cases, a conservative estimate of the uncertainty in U may be computed. We present this result here. If X1, X2, ..., Xn are measurements whose uncertainties are small, and if U = U(X1, X2, ..., Xn) is a function of (X1, X2, ..., Xn), then a conservative estimate of σU is given by (3.13) In practice, we evaluate the partial derivatives at the point (X1, X2, ..., Xn). The inequality (3.13) is valid in almost all practical situations; in principle it can fail if some of the second partial derivatives of U are quite large. Example 3.22 Refer to Example 3.20. Find a conservative estimate for the uncertainty in the total resistance R if R1 and R2 are not known to be independent. Solution , ∂R/∂R1 = 0.0278, and ∂R/∂R2 = 0.694. Therefore We have The uncertainty in the total resistance is conservatively estimated by 0.97 Ω. In Example 3.20, we computed the uncertainty to be 0.75 Ω when R1 and R2 are independent. Page 190 Relative Uncertainties for Functions of Several Measurements In Section 3.3, we presented methods for calculating relative uncertainties for functions of one variable. The methods for calculating relative uncertainties for functions of several variables are similar. There are two methods for approximating the relative uncertainty σU/U in a function U = U(X1, X2, ..., Xn): 1. Compute σU using Equation (3.12), and then divide by U. 2. Compute ln U, and then use Equation (3.12) to find σln U, which is equal to σU/U. Both of the methods work in every instance, so one may use whichever is easiest for a given problem. This choice is usually dictated by whether it is easier to compute partial derivatives of U or of ln U. Example 3.23 Two perpendicular sides of a rectangle are measured to be X = 2.0 ± 0.1 cm and Y = 3.2 ± 0.2 cm. Find the relative uncertainty in the area A = XY. Solution This is easily computed by finding the absolute uncertainty in ln A = ln X + ln Y. We begin by computing the partial derivatives of ln A: We are given that σX = 0.1 and σY = 0.2. The relative uncertainty in A is The relative uncertainty in A is 0.080, or 8%. The area of the rectangle is 6.4 cm2 ± 8%. Example 3.24 An Atwood machine consists of two masses X and Y (X > Y) attached to the ends of a light string that passes over a light frictionless pulley. When the masses are released, the larger mass X accelerates down with acceleration Page 191 Suppose that X and Y are measured as X = 100 ± 1 g, and Y = 50 ± 1 g. Assume that g, the acceleration due to gravity, is known with negligible uncertainty. Find the relative uncertainty in the acceleration a. Solution The relative uncertainty in a is equal to the absolute uncertainty in ln a =lng + ln(X − Y) − ln(X + Y). We treat g as a constant, since its uncertainty is negligible. The partial derivatives are The uncertainties in X and Y are σX = σY = 1. The relative uncertainty in a is The relative uncertainty in a is 0.030, or 3%. Note that this value does not depend on g. When a function involves a product or quotient of measurements, the relative uncertainty in the function can be computed directly from the relative uncertainties in the measurements. Example 3.25 illustrates the method. Example 3.25 The height h of a cylinder is measured with a relative uncertainty of 2%, and the radius r is measured with a relative uncertainty of 1%. Find the relative uncertainty in the volume V. Solution The volume is given by V = πr2h. Taking logs, we have Since ln V is a linear combination, and ln π is constant, we can use Equation (3.4) to obtain But σln V = σV/V, σln h = σh/h, and σln r = σr/r are the relative uncertainties in V, h, and r, respectively. Substituting 0.02 for σln h and 0.01 for σln r, we obtain The relative uncertainty in V is 2.8%. We can generalize the result of Example 3.25. If X1, ..., Xn are measurements whose relative uncertainties are Page 192 , and , where m1, ..., mn are any exponents, then the relative uncertainty in U is (3.14) Derivation of the Multivariate Propagation of Error Formula We derive the propagation of error formula for a nonlinear function U of a random variable X by approximating it with a multivariate linear function (i.e., linearizing the problem), and then using the methods of Section 3.2. To find a linear approximation to U, we use a firstorder multivariate Taylor series approximation. Let U = U(X1, X2, ..., Xn) be a function whose partial derivatives all exist. Let (μ1, μ2, ..., μn) be any point. Then if X1, X2, ..., Xn are close to μ1, μ2, ..., μn, respectively, the linearization of U is (3.15) Each partial derivative is evaluated at the point (μ1, μ2, ..., μn). If X1, X2, ..., Xn are independent measurements, the linear approximation leads to a method for approximating the uncertainty in U, given the uncertainties in X1, X2, ..., Xn. The derivation is similar to the one-variable case presented at the end of Section 3.3. Let μ1, μ2, ..., μn be the means of X1, X2, ..., Xn, respectively. Then for any reasonably precise measurements, X1, X2, ..., Xn will be close enough to μ1, μ2, ..., μn for the linearization to be valid. We can rewrite Equation (3.15) as (3.16) The quantities ∂U/∂X1, ∂U/∂X2, ..., ∂U/∂Xn are all constant, since they are evaluated at the point (μ1, μ2, ..., μn). Therefore the quantity is constant as well. It follows from Equation (2.42) (in Section 2.5) and Equation (3.4) (in Section 3.2) that Exercises for Section 3.4 Page 193 1. Find the uncertainty in U, assuming that X = 10.0 ± 0.5, Y = 5.0 ± 0.1, and a. U = XY2 b. U = X2 + Y2 c. U = (X + Y2)/2 2. The volume of a cone is given by V = πr2h/3, where r is the radius of the base and h is the height. Assume the height is measured to be h = 6.00 ± 0.01 cm and the radius is r = 5.00 ± 0.02 cm. a. Estimate the volume of the cone, and find the uncertainty in the estimate. b. Which would provide a greater reduction in the uncertainty in V: reducing the uncertainty in h to 0.005 cm or reducing the uncertainty in r to 0.01 cm? From a fixed point on the ground, the distance to a certain tree is measured to be s = 55.2 ± 0.1 m and the angle from the point to the top of the tree is measured to be θ = 0.50 ± 0.02 radians. The height of the tree is given by h = s tan θ. a. Estimate h, and find the uncertainty in the estimate. b. Which would provide a greater reduction in the uncertainty in h: reducing the uncertainty in s to 0.05 m or reducing the uncertainty in θ to 0.01 radians? Refer to Exercise 10 in Section 3.2. Assume that X = 30.0 ± 0.1 Pa, h = 10.0 ± 0.2 mm, and μ = 1.49 Pa · s with negligible uncertainty. a. Estimate V and find the uncertainty in the estimate. b. Which would provide a greater reduction in the uncertainty in V: reducing the uncertainty in τ to 0.01 Pa or reducing the uncertainty in h to 0.1 mm? When air enters a compressor at pressure P1 and leaves at pressure P2, the intermediate pressure is given by . Assume that P1 = 10.1 ± 0.3 MPa and P2 = 20.1 ± 0.4 MPa. a. Estimate P3, and find the uncertainty in the estimate. 3. 4. 5. b. 6. Which would provide a greater reduction in the uncertainty in P3: reducing the uncertainty in P1 to 0.2 MPa or reducing the uncertainty in P2 to 0.2 MPa? One way to measure the water content of a soil is to weigh the soil both before and after drying it in an oven. The water content is W = (M1 − M2)/M1, where M1 is the mass before drying and M2 is the mass after drying. Assume that M1 = 1.32 ± 0.01 kg and M2 = 1.04 ± 0.01 kg. a. Estimate W, and find the uncertainty in the estimate. b. Which would provide a greater reduction in the uncertainty in W: reducing the uncertainty in M1 to 0.005 kg or reducing the uncertainty in M2 to 0.005 kg? 7. 8. 9. The lens equation says that if an object is placed at a distance p from a lens, and an image is formed at a distance q from the lens, then the focal length f satisfies the equation 1/f = 1/p + 1/q. Assume that p = 2.3 ± 0.2 cm and q = 3.1 ± 0.2 cm. a. Estimate f, and find the uncertainty in the estimate. b. Which would provide a greater reduction in the uncertainty in f: reducing the uncertainty in p to 0.1 cm or reducing the uncertainty in q to 0.1 cm? The pressure P, temperature T, and volume V of one mole of an ideal gas are related by the equation PV = 8.31 T, when P is measured in kilopascals, T is measured in kelvins, and V is measured in liters. a. Assume that P = 242.52 ± 0.03 kPa and V = 10.103 ± 0.002 L. Estimate T, and find the uncertainty in the estimate. b. Assume that P = 242.52 ± 0.03 kPa and T = 290.11 ± 0.02 K. Estimate V, and find the uncertainty in the estimate. c. Assume that V = 10.103 ± 0.002 L and T = 290.11 ± 0.02 K. Estimate P, and find the uncertainty in the estimate. The Beer-Lambert law relates the absorbance A of a solution to the concentration C of a species in solution by A = MLC, where L is the path length and M is the molar absorption coefficient. Assume that C = 1.25 ± 0.03 mol/cm3, L = 1.2 ± 0.1 cm, and A = 1.30 ± 0.05. a. Estimate M and find the uncertainty in the estimate. Page 194 Which would provide a greater reduction in the uncertainty in M: reducing the uncertainty in C to 0.01 mol/cm3, reducing the uncertainty in L to 0.05 cm, or reducing the uncertainty in A to 0.01? 10. In the article "Temperature-Dependent Optical Constants of Water Ice in the Near Infrared: New Results and Critical Review of the Available Measurements" (B. Rajaram, D. Glandorf, et al., Applied Optics, 2001:4449–4462), the imaginary index of refraction of water ice is presented for various frequencies and temperatures. At a frequency of 372.1 cm −1 and a temperature of 166 K, the index is estimated to be 0.00116. At the same frequency and at a temperature of 196 K, the index is estimated to be 0.00129. The uncertainty is reported to be 10−4 for each of these two estimated indices. The ratio of the indices is estimated to be 0.00116/0.00129 = 0.899. Find the uncertainty in this ratio. 11. Refer to Exercise 12 in Section 3.2. Assume that τ0 = 50 ± 1 MPa, w = 1.2 ± 0.1 mm, and k = 0.29 ± 0.05 mm−1. a. Estimate τ, and find the uncertainty in the estimate. b. Which would provide the greatest reduction in the uncertainty in τ: reducing the uncertainty in τ0 to 0.1 MPa, reducing the uncertainty in w to 0.01 mm, or reducing the uncertainty in k to 0.025 mm−1? c. A new, somewhat more expensive process would allow both τ0 and w to be measured b. 12. 13. 14. 15. with negligible uncertainty. Is it worthwhile to implement the process? Explain. According to Snell's law, the angle of refraction θ2 of a light ray traveling in a medium of index of refraction n is related to the angle of incidence θ1 of a ray traveling in a vacuum through the equation sin θ1 = n sin θ2. Assume that θ1 = 0.3672 ± 0.005 radians and θ2 = 0.2943 ± 0.004 radians. Estimate n, and find the uncertainty in the estimate. Archaeologists studying meat storage methods employed by the Nunamiut in northern Alaska have developed a Meat Drying Index. Following is a slightly simplified version of the index given in the article "A Zooarchaeological Signature for Meat Storage: Rethinking the Drying Utility Index" (T. Friesen, American Antiquity, 2001:315–331). Let m represent the weight of meat, b the weight of bone, and g the gross weight of some part of a caribou. The Meat Drying Index y is given by y = mb/g. Assume that for a particular caribou rib, the following measurements are made (in grams): g = 3867.4 ± 0.3, b = 1037.0 ± 0.2, m = 2650.4 ± 0.1. a. Estimate y, and find the uncertainty in the estimate. b. Which would provide the greatest reduction in the uncertainty in y: reducing the uncertainty in g to 0.1 g, reducing the uncertainty in b to 0.1 g, or reducing the uncertainty in m to 0? The resistance R (in ohms) of a cylindrical conductor is given by R = kl/d2, where l is the length, d is the diameter, and k is a constant of proportionality. Assume that l = 14.0 ± 0.1 cm and d = 4.4 ± 0.1 cm. a. Estimate R, and find the uncertainty in the estimate. Your answer will be in terms of the proportionality constant k. b. Which would provide the greater reduction in the uncertainty in R: reducing the uncertainty in l to 0.05 cm or reducing the uncertainty in d to 0.05 cm? A cylindrical wire of radius R elongates when subjected to a tensile force F. Let L0 represent the initial length of the wire and let L1 represent the final length. Young's modulus for the material is given by Assume that F = 800 ± 1 N, R = 0.75 ± 0.1 mm, L0 = 25.0 ± 0.1 mm, and L1 = 30.0 ± 0.1 mm. a. Estimate Y, and find the uncertainty in the estimate. b. Of the uncertainties in F, R, L0, and L1, only one has a non-negligible effect on the uncertainty in Y. Which one is it? 16. According to Newton's law of cooling, the time t needed for an object at an initial temperature T0 to cool to a temperature T in an environment with ambient temperature Ta is given by where k is a constant. Assume that for a certain type of container, k = 0.025 min−1. Let t be the number of minutes needed to cool the container to a temperature of 50°F. Assume that T0 = 70.1 ± 0.2°F and Ta = 35.7 ± 0.1°F. Estimate t, and find the uncertainty in the Page 195 estimate. 17. Refer to Exercise 16. In an experiment to determine the value of k, the temperature T at time t = 10 min is measured to be T = 54.1 ± 0.2°F. Assume that T0 = 70.1 ± 0.2°F and Ta = 35.7 ± 0.1°F. Estimate k, and find the uncertainty in the estimate. 18. The vertical displacement v of a cracked slurry infiltrated fiber concrete member at maximum shear stress is given by v = a + bw, where w is the crack width, and a and b are estimated from data to be a = 2.5 ± 0.1 mm and b = 0.05 ± 0.01. Assume that w = 1.2 ± 0.1 mm. a. Estimate v, and find the uncertainty in the estimate. b. Of the uncertainties in w, a, and b, only one has a non-negligible effect on the uncertainty in v. Which one is it? 19. The shape of a bacterium can be approximated by a cylinder of radius r and height h capped on each end by a hemisphere. The volume and surface area of the bacterium are given by It is known that the rate R at which a chemical is absorbed into the bacterium is R = c(S/V), where c is a constant of proportionality. Assume that for a certain bacterium, r = 0.9 ± 0.1 μm and h = 1.7 ± 0.1 μm. a. Are the computed values of S and V independent? Explain. b. Assuming the measurements of r and h to be independent, estimate R and find the uncertainty in the estimate. Your answer will be in terms of c. 20. Estimate U, and find the relative uncertainty in the estimate, assuming that X = 5.0 ± 0.2, Y = 10.0 ± 0.5, and a. b. c. U = X2 + Y2 21. Refer to Exercise 10 in Section 3.2. Assume that τ = 35.2 ± 0.1 Pa, h = 12.0 ± 0.3 mm, and μ = 1.49 Pa · s with negligible uncertainty. Estimate V, and find the relative uncertainty in the estimate. 22. Refer to Exercise 5. Assume that P1 = 15.3 ± 0.2 MPa and P2 = 25.8 ± 0.1 MPa. Estimate P3, and find the relative uncertainty in the estimate. 23. Refer to Exercise 7. Assume that p = 4.3 ± 0.1 cm and q = 2.1 ± 0.2 cm. Estimate f, and find the relative uncertainty in the estimate. 24. Refer to Exercise 8. a. Assume that P = 224.51 ± 0.04 kPa and V = 11.237 ± 0.002 L. Estimate T, and find the relative uncertainty in the estimate. b. Assume that P = 224.51 ± 0.04 kPa and T = 289.33 ± 0.02 K. Estimate V, and find the relative uncertainty in the estimate. c. Assume that V = 11.203 ± 0.002 L and T = 289.33 ± 0.02 K. Estimate P, and find the relative uncertainty in the estimate. 25. Refer to Exercise 12. Estimate n, and find the relative uncertainty in the estimate, from the following measurements: θ1 = 0.216 ± 0.003 radians and θ2 = 0.456 ± 0.005 radians. 26. Refer to Exercise 14. Assume that l = 10.0 cm ± 0.5% and d = 10.4 cm ± 0.5%. a. Estimate R, and find the relative uncertainty in the estimate. Does the relative uncertainty depend on k? b. Assume that either l or d can be remeasured with relative uncertainty 0.2%. Which should be remeasured to provide the greater improvement in the relative uncertainty of the resistance? 27. Refer to Exercise 15. Assume that F = 750 ± 1 N, R = 0.65 ± 0.09 mm, L0 = 23.7 ± 0.2 mm, and L1 = 27.7 ± 0.2 mm. Estimate Y, and find the relative uncertainty in the estimate. 28. Refer to Exercise 16. Assume that T0 = 73.1 ± 0.1°F, Ta = 37.5 ± 0.2°F, k = 0.032 min−1 with negligible uncertainty, and T = 50°F exactly. Estimate t, and find the relative uncertainty in the estimate. 29. Refer to Exercise 19. Assume that for a certain bacterium, r = 0.8 ± 0.1 μm and h = 1.9 ± 0.1 μm. a. Estimate S, and find the relative uncertainty in the estimate. Page 196 b. Estimate V, and find the relative uncertainty in the estimate. c. Estimate R, and find the relative uncertainty in the estimate. d. Does the relative uncertainty in R depend on c? 30. Refer to Exercise 5. Assume that the relative uncertainty in P1 is 5% and the relative uncertainty in P2 is 2%. Find the relative uncertainty in P3. 31. Refer to Exercise 14. Assume that the relative uncertainty in l is 3% and that the relative uncertainty in d is 2%. Find the relative uncertainty in R. Supplementary Exercises for Chapter 3 1. 2. Assume that X, Y, and Z are independent measurements with X = 25 ± 1, Y = 5.0 ± 0.3, and Z = 3.5 ± 0.2. Find the uncertainties in each of the following quantities: a. XY + Z b. Z/(X + Y) c. d. Assume that X, Y, and Z are independent measurements, and that the relative uncertainty in X is 5%, the relative uncertainty in Y is 10%, and the relative uncertainty in Z is 15%. Find the relative uncertainty in each of the following quantities: a. XYZ b. c. 3. 4. 5. 6. An item is to be constructed by laying three components in a row. The length of each component will be measured. a. If the uncertainty in measuring the length of each component is 1.2 mm, what is the uncertainty in the combined length of the three components? b. If it is desired to estimate the length of the item with an uncertainty of 0.5 mm, what must be the uncertainty in the measurement of each individual component? Assume the uncertainties in the three measurements are the same. For some genetic mutations, it is thought that the frequency of the mutant gene in men increases linearly with age. If m1 is the frequency at age t1, and m2 is the frequency at age t2, then the yearly rate of increase is estimated by r = (m2 − m1)/(t2 − t1). In a polymerase chain reaction assay, the frequency in 20-year-old men was estimated to be 17.7 ± 1.7 per μg DNA, and the frequency in 40-year-old men was estimated to be 35.9 ± 5.8 per μg DNA. Assume that age is measured with negligible uncertainty. a. Estimate the yearly rate of increase, and find the uncertainty in the estimate. b. Find the relative uncertainty in the estimated rate of increase. The Darcy–Weisbach equation states that the power-generating capacity in a hydroelectric system that is lost due to head loss is given by P = ηγQH, where η is the efficiency of the turbine, γ is the specific gravity of water, Q is the flow rate, and H is the head loss. Assume that η = 0.85 ± 0.02, H = 3.71 ± 0.10 m, Q = 60 ± 1m3/s, and γ = 9800 N/m3 with negligible uncertainty. a. Estimate the power loss (the units will be in watts), and find the uncertainty in the estimate. b. Find the relative uncertainty in the estimated power loss. c. Which would provide the greatest reduction in the uncertainty in P: reducing the uncertainty in η to 0.01, reducing the uncertainty in H to 0.05, or reducing the uncertainty in Q to 0.5? Let A and B represent two variants (alleles) of the DNA at a certain locus on the genome. Let p represent the proportion of alleles in a population that are of type A, and let q represent the proportion of alleles that are of type B. The Hardy–Weinberg equilibrium principle states that the proportion PAB of organisms that are of type AB is equal to pq. In a population survey of a particular species, the proportion of alleles of type A is estimated to be 0.360 ± 0.048 and the proportion of alleles of type B is independently estimated to be 0.250 ± 0.043. a. Estimate the proportion of organisms that are of type AB, and find the uncertainty in the estimate. Page 197 b. c. 7. Find the relative uncertainty in the estimated proportion. Which would provide a greater reduction in the uncertainty in the proportion: reducing the uncertainty in the type A proportion to 0.02 or reducing the uncertainty in the type B proportion to 0.02? The heating capacity of a calorimeter is known to be 4 kJ/°C, with negligible uncertainty. The number of dietary calories (kilocalories) per gram of a substance is given by C = 8. 9. 10. 11. 12. 13. cH(ΔT)/m, where C is the number of dietary calories, H is the heating capacity of the calorimeter, ΔT is the increase in temperature in °C caused by burning the substance in the calorimeter, m is the mass of the substance in grams, and c = 0.2390 cal/kJ is the conversion factor from kilojoules to dietary calories. An amount of mayonnaise with mass 0.40 ± 0.01 g is burned in a calorimeter. The temperature increase is 2.75 ± 0.02°C. a. Estimate the number of dietary calories per gram of mayonnaise, and find the uncertainty in the estimate. b. Find the relative uncertainty in the estimated number of dietary calories.

are conducted, and ■ The trials are independent ■ Each trial has the same success probability p ■ X is the number of successes in the n trials then X has the binomial distribution with parameters n and p, denoted X ∼ Bin(n, p). Example 4.5 A fair coin is tossed 10 times. Let X be the number of heads that appear. What is the distribution of X? Solution There are 10 independent Bernoulli trials, each with success probability p = 0.5. The random variable X is equal to the number of successes in the 10 trials. Therefore X ∼ Bin(10, 0.5). Recall from the discussion of independence in Section 1.1 that when drawing a sample from a finite, tangible population, the sample items may be treated as independent if the population is very large compared to the size of the sample. Otherwise the sample items are not independent. In some cases, the purpose of drawing a sample may be to classify each sample item into one of two categories. For example, we may sample a number of items from a lot and classify each one as defective or nondefective. In these cases, each sampled item is a Bernoulli trial, with one category counted as a success and the other counted as a failure. When the population of items is large compared to the number sampled, these Bernoulli trials are nearly independent, and the number of successes among them has, for all practical purposes, a binomial distribution. When the population size is not large compared to the sample, however, the Bernoulli trials are not independent, and the number of successes among them does not have a binomial distribution. A good rule of thumb is that if the sample size is 5% or less of the population, the binomial distribution may be used to model the number of successes. Page 205 Example 4.6 A lot contains several thousand components, 10% of which are defective. Seven components are sampled from the lot. Let X represent the number of defective components in the sample. What is the distribution of X? Solution Since the sample size is small compared to the population (i.e., less than 5%), the number of successes in the sample approximately follows a binomial distribution. Therefore we model X with the Bin(7, 0.1) distribution. Probability Mass Function of a Binomial Random Variable We now derive the probability mass function of a binomial random variable by considering an example. A biased coin has probability 0.6 of coming up heads. The coin is tossed three times. Let X be the number of heads. Then X ∼ Bin(3, 0.6). We will compute P(X = 2). There are three arrangements of two heads in three tosses of a coin, HHT, HTH, and THH. We first compute the probability of HHT. The event HHT is a sequence of independent events: H on the first toss, H on the second toss, T on the third toss. We know the probabilities of each of these events separately: P(H on the first toss) = 0.6, P(H on the second toss) = 0.6, P(T on the third toss) = 0.4 Since the events are independent, the probability that they all occur is equal to the product of their probabilities (Equation 2.20 in Section 2.3). Thus Similarly, P(HTH) = (0.6)(0.4)(0.6) = (0.6)2(0.4)1 and P(THH) = (0.4)(0.6)(0.6) = (0.6)2(0.4)1. It is easy to see that all the different arrangements of two heads and one tail have the same probability. Now Examining this result, we see that the number 3 represents the number of arrangements of two successes (heads) and one failure (tails), 0.6 is the success probability p, the exponent 2 is the number of successes, 0.4 is the failure probability 1 − p, and the exponent 1 is the number of failures. We can now generalize this result to produce a formula for the probability of x successes in n independent Bernoulli trials with success probability p, in terms of x, n, and p. In other words, we can compute P(X = x) where X ∼ Bin(n, p). We can see that Page 206 (4.3) All we need to do now is to provide an expression for the number of arrangements of x successes in n trials. To describe this number, we need factorial notation. For any positive integer n, the quantity n! (read "n factorial") is the number We also define 0! = 1. The number of arrangements of x successes in n trials is n!/[x!(n − x)!]. (A derivation of this result is presented in Section 2.2.) We can now define the probability mass function for a binomial random variable. If X ∼ Bin(n, p), the probability mass function of X is (4.4) Figure 4.2 presents probability histograms for the Bin(10, 0.4) and Bin(20, 0.1) probability mass functions. FIGURE 4.2 (a) The Bin(10, 0.4) probability histogram. (b) The Bin(20, 0.1) probability histogram. Example 4.7 Find the probability mass function of the random variable X if X ∼ Bin(10, 0.4). Find P(X = 5). Page 207 Solution We use Equation (4.4) with n = 10 and p = 0.4. The probability mass function is Example 4.8 A fair die is rolled eight times. Find the probability that no more than 2 sixes come up. Solution Each roll of the die is a Bernoulli trial with success probability 1/6. Let X denote the number of sixes in 8 rolls. Then X ∼ Bin(8, 1/6). We need to find P(X ≤ 2). Using the probability mass function, Table A.1 (in Appendix A) presents binomial probabilities of the form P(X ≤ x) for n ≤ 20 and selected values of p. Examples 4.9 and 4.10 illustrate the use of this table. Example 4.9 According to a 2016 report by J.D. Power and Associates, approximately 20% of registered automobiles have been recalled for a manufacturer's defect but have not been repaired. Twelve automobiles are sampled at random. What is the probability that fewer than four of them have been recalled but not repaired? Solution Let X represent the number of automobiles in the sample that have been recalled but not repaired. Then X ∼ Bin(12, 0.2). The probability that fewer than four have been recalled but not repaired is P(X ≤ 3). We consult Table A.1 with n = 12, p = 0.2, and x = 3. We find that P(X ≤ 3) = 0.795. Page 208 Sometimes the best way to compute the probability of an event is to compute the probability that the event does not occur, and then subtract from 1. Example 4.10 provides an illustration. Example 4.10 Refer to Example 4.9. What is the probability that more than 1 of the 12 cars has been recalled but not repaired? Solution Let X represent the number of cars that have been recalled but not repaired. We wish to compute P(X > 1). Table A.1 presents probabilities of the form P(X ≤ x). Therefore we note that P(X > 1) = 1 − P(X ≤ 1). Consulting the table with n = 12, p = 0.2, x = 1, we find that P(X ≤ 1) = 0.275. Therefore P(X > 1) = 1 − 0.275 = 0.725. A Binomial Random Variable Is a Sum of Bernoulli Random Variables Assume n independent Bernoulli trials are conducted, each with success probability p. Let Y1, ..., Yn be defined as follows: Yi = 1 if the ith trial results in success, and Yi = 0 otherwise. Then each of the random variables Yi has the Bernoulli(p) distribution. Now let X represent the number of successes among the n trials. Then X ∼ Bin(n, p). Since each Yi is either 0 or 1, the sum Y1 + ··· + Yn is equal to the number of the Yi that have the value 1, which is the number of successes among the n trials. Therefore X = Y1 + ··· + Yn. This shows that a binomial random variable can be expressed as a sum of Bernoulli random variables. Put another way, sampling a single value from a Bin(n, p) population is equivalent to drawing a sample of size n from a Bernoulli(p) population, and then summing the sample values. The Mean and Variance of a Binomial Random Variable With a little thought, it is easy to see how to compute the mean of a binomial random variable. For example, if a fair coin is tossed 10 times, we expect on the average to see five heads. The number 5 comes from multiplying the success probability (0.5) by the number of trials (10). This method works in general. If we perform n Bernoulli trials, each with success probability p, the mean number of successes is np. Therefore, if X ∼ Bin(n, p), then μX = np. We can verify this intuition by noting that X is the sum of n Bernoulli variables, each with mean p. The mean of X is therefore the sum of the means of the Bernoulli random variables that compose it, which is equal to np. We can compute σ2X by again noting that X is the sum of independent Bernoulli random variables and recalling that the variance of a Bernoulli random variable is p(1 − p). The variance of X is therefore the sum of the variances of the Bernoulli random variables that compose it, which is equal to np(1 − p). Page 209 Summary If X ∼ Bin(n, p), then the mean and variance of X are given by (4.5) (4.6) Using the binomial probability mass function (Equation 4.4), we could in principle compute the mean and variance of a binomial random variable by using the definitions of mean and variance for a discrete random variable (Equations 2.29 and 2.30 in Section 2.4). These expressions involve sums that are tedious to evaluate. It is much easier to think of a binomial random variable as a sum of independent Bernoulli random variables. Using a Sample Proportion to Estimate a Success Probability In many cases we do not know the success probability p associated with a certain Bernoulli trial, and we wish to estimate its value. A natural way to do this is to conduct n independent trials and count the number X of successes. To estimate the success probability p we compute the sample proportion . This notation follows a pattern that is important to know. The success probability, which is unknown, is denoted by p. The sample proportion, which is known, is denoted . The "hat" (ˆ) indicates that is used to estimate the unknown value p. Example 4.11 A quality engineer is testing the calibration of a machine that packs ice cream into containers. In a sample of 20 containers, 3 are underfilled. Estimate the probability p that the machine underfills a container. Solution The sample proportion of underfilled containers is probability that the machine underfills a container is 0.15 as well. We estimate that the Uncertainty in the Sample Proportion It is important to realize that the sample proportion is just an estimate of the success probability p, and in general, is not equal to p. If another sample were taken, the value of would probably come out differently. In other words, there is uncertainty in . For to be useful, we must compute its bias and its uncertainty. We now do this. Let n denote the sample size, and let X denote the number of successes, where X ∼ Bin(n, p). Page 210 The bias is the difference . Since , it follows from Equation (2.43) (in Section 2.5) that Since , is unbiased; in other words, its bias is 0. The uncertainty is the standard deviation . From Equation (4.6), the standard deviation of X is . Since , it follows from Equation (2.45) (in Section 2.5) that In practice, when computing the uncertainty in we approximate it with , we don't know the success probability p, so . Summary If X ∼ Bin(n, p), then the sample proportion is used to estimate the success probability p. ■ is unbiased. ■ The uncertainty in is (4.7) In practice, when computing Example , we substitute for p, since p is unknown. Example 4.12 The safety commissioner in a large city wants to estimate the proportion of buildings in the city that are in violation of fire codes. A random sample of 40 buildings is chosen for inspection, and 4 of them are found to have fire code violations. Estimate the proportion of buildings in the city that have fire code violations, and find the uncertainty in the estimate. Solution Let p denote the proportion of buildings in the city that have fire code violations. The sample size (number of trials) is n = 40. The number of buildings with violations (successes) is X = 4. We estimate p with the sample proportion : Page 211 Using Equation (4.7), the uncertainty in Substituting is for p and 40 for n, we obtain In Example 4.12, it turned out that the uncertainty in the sample proportion was rather large. We can reduce the uncertainty by increasing the sample size. Example 4.13 shows how to compute the approximate sample size needed to reduce the uncertainty to a specified amount. Example 4.13 In Example 4.12, approximately how many additional buildings must be inspected so that the uncertainty in the sample proportion of buildings in violation will be only 0.02? Solution We need to find the value of n so that . Approximating p with , we obtain Solving for n yields n = 225. We have already sampled 40 buildings, so we need to sample 185 more. The following example requires knowledge of propagation of error, which is covered in Section 3.3. Example 4.14 In a sample of 100 newly manufactured automobile tires, 7 are found to have minor flaws in the tread. If four newly manufactured tires are selected at random and installed on a car, estimate the probability that none of the four tires have a flaw, and find the uncertainty in this estimate. Solution Let p represent the probability that a tire has no flaw. The probability that all four tires have no flaw is p4. We use propagation of error (Section 3.3) to estimate the uncertainty in p4. We begin by computing the sample proportion and finding its uncertainty. The sample proportion is . We substitute n = 100 and . The uncertainty in is given by . for p to obtain Page 212 We estimate p4 with . We use Equation (3.10) to compute the uncertainty in : Exercises for Section 4.2 1. Let X ∼ Bin(7, 0.3). Find a. P(X = 1) b. P(X = 2) c. P(X < 1) d. P(X > 4) e. μX f. Let X ∼ Bin(9, 0.4). Find a. P(X > 6) b. P(X ≥ 2) c. P(2 ≤ X < 5) d. P(2 < X ≤ 5) e. P(X = 0) f. P(X = 7) g. μx h. 3. 4. 5. Find the following probabilities: a. P(X = 2) when X ∼ Bin(4, 0.6) b. P(X > 2) when X ∼ Bin(8, 0.2) c. P(X ≤ 2) when X ∼ Bin(5, 0.4) d. P(3 ≤ X ≤ 5) when X ∼ Bin(6, 0.7) At a certain airport, 75% of the flights arrive on time. A sample of 10 flights is studied. a. Find the probability that all 10 of the flights were on time. b. Find the probability that exactly eight of the flights were on time. c. Find the probability that eight or more of the flights were on time. Of all the registered automobiles in a certain state, 10% violate the state emissions standard. Twelve automobiles are selected at random to undergo an emissions test. a. Find the probability that exactly three of them violate the standard. b. Find the probability that fewer than three of them violate the standard. c. Find the probability that none of them violate the standard. 6. 7. A fair die is rolled 8 times. a. What is the probability that the die comes up 6 exactly twice? b. What is the probability that the die comes up an odd number exactly five times? c. Find the mean number of times a 6 comes up. d. Find the mean number of times an odd number comes up. e. Find the standard deviation of the number of times a 6 comes up. f. Find the standard deviation of the number of times an odd number comes up. Of all the weld failures in a certain assembly, 85% of them occur in the weld metal itself, and the remaining 15% occur in the base metal. A sample of 20 weld failures is examined. a. What is the probability that exactly five of them are base metal failures? b. What is the probability that fewer than four of them are base metal failures? c. What is the probability that none of them are base metal failures? Page 213 d. Find the mean number of base metal failures. e. Find the standard deviation of the number of base metal failures. 8. A general contracting firm experiences cost overruns on 20% of its contracts. In a company audit, 20 contracts are sampled at random. a. What is the probability that exactly four of them experience cost overruns? b. What is the probability that fewer than three of them experience cost overruns? c. What is the probability that none of them experience cost overruns? d. Find the mean number that experience cost overruns. e. Find the standard deviation of the number that experience cost overruns. 9. Several million lottery tickets are sold, and 60% of the tickets are held by women. Five winning tickets will be drawn at random. a. What is the probability that three or fewer of the winners will be women? b. What is the probability that three of the winners will be of one gender and two of the winners will be of the other gender? 10. A marketing manager samples 150 people and finds that 87 of them have made a purchase on the internet within the past month. a. Estimate the proportion of people who have made a purchase on the internet within the past month, and find the uncertainty in the estimate. b. Estimate the number of people who must be sampled to reduce the uncertainty to 0.03. 11. A quality engineer samples 100 steel rods made on mill A and 150 rods made on mill B. Of the rods from mill A, 88 meet specifications, and of the rods from mill B, 135 meet specifications. a. Estimate the proportion of rods from mill A that meet specifications, and find the uncertainty in the estimate. b. Estimate the proportion of rods from mill B that meet specifications, and find the uncertainty in the estimate. c. Estimate the difference between the proportions, and find the uncertainty in the estimate. 12. Of the items manufactured by a certain process, 20% are defective. Of the defective items, 60% can be repaired. a. Find the probability that a randomly chosen item is defective and cannot be repaired. b. Find the probability that exactly 2 of 20 randomly chosen items are defective and cannot be repaired. 13. Of the bolts manufactured for a certain application, 90% meet the length specification and can be used immediately, 6% are too long and can be used after being cut, and 4% are too short and must be scrapped. a. Find the probability that a randomly selected bolt can be used (either immediately or after being cut). b. Find the probability that fewer than 9 out of a sample of 10 bolts can be used (either immediately or after being cut). 14. Gears produced by a grinding process are categorized either as conforming (suitable for their intended purpose), downgraded (unsuitable for the intended purpose but usable for another purpose), or scrap (not usable). Suppose that 80% of the gears produced are conforming, 15% are downgraded, and 5% are scrap. Ten gears are selected at random. a. What is the probability that one or more is scrap? b. What is the probability that eight or more are not scrap? c. What is the probability that more than two are either downgraded or scrap? d. What is the probability that exactly nine are either conforming or downgraded? 15. A commuter must pass through three traffic lights on her way to work. For each light, the probability that it is green when she arrives is 0.6. The lights are independent. a. What is the probability that all three lights are green? b. The commuter goes to work five days per week. Let X be the number of times out of the five days in a given week that all three lights are green. Assume the days are independent of one another. What is the distribution of X? c. Find P(X = 3). Page 214 16. A distributor receives a large shipment of components. The distributor would like to accept the shipment if 10% or fewer of the components are defective and to return it if more than 10% of the components are defective. She decides to sample 10 components, and to return the shipment if more than 1 of the 10 is defective. a. If the proportion of defectives in the batch is in fact 10%, what is the probability that she will return the shipment? b. If the proportion of defectives in the batch is 20%, what is the probability that she will return the shipment? c. If the proportion of defectives in the batch is 2%, what is the probability that she will return the shipment? d. The distributor decides that she will accept the shipment only if none of the sampled items are defective. What is the minimum number of items she should sample if she wants to have a probability no greater than 0.01 of accepting the shipment if 20% of the components in the shipment are defective? 17. A k out of n system is one in which there is a group of n components, and the system will function if at least k of the components function. Assume the components function independently of one another. a. In a 3 out of 5 system, each component has probability 0.9 of functioning. What is the probability that the system will function? b. In a 3 out of n system, in which each component has probability 0.9 of functioning, what is the smallest value of n needed so that the probability that the system functions is at least 0.90? 18. Refer to Exercise 17 for the definition of a k out of n system. For a certain 4 out of 6 system, assume that on a rainy day each component has probability 0.7 of functioning, and that on a nonrainy day each component has probability 0.9 of functioning. a. What is the probability that the system functions on a rainy day? b. What is the probability that the system functions on a nonrainy day? c. Assume that the probability of rain tomorrow is 0.20. What is the probability that the system will function tomorrow? 19. A certain large shipment comes with a guarantee that it contains no more than 15% defective items. If the proportion of defective items in the shipment is greater than 15%, the shipment may be returned. You draw a random sample of 10 items. Let X be the number of defective items in the sample. a. If in fact 15% of the items in the shipment are defective (so that the shipment is good, but just barely), what is P(X ≥ 7)? b. Based on the answer to part (a), if 15% of the items in the shipment are defective, would 7 defectives in a sample of size 10 be an unusually large number? c. If you found that 7 of the 10 sample items were defective, would this be convincing evidence that the shipment should be returned? Explain. d. If in fact 15% of the items in the shipment are defective, what is P(X ≥ 2)? e. Based on the answer to part (d), if 15% of the items in the shipment are defective, would 2 defectives in a sample of size 10 be an unusually large number? f. If you found that 2 of the 10 sample items were defective, would this be convincing evidence that the shipment should be returned? Explain. 20. An insurance company offers a discount to homeowners who install smoke detectors in their homes. A company representative claims that 80% or more of policyholders have smoke detectors. You draw a random sample of eight policyholders. Let X be the number of policyholders in the sample who have smoke detectors. a. b. c. d. e. f. If exactly 80% of the policyholders have smoke detectors (so the representative's claim is true, but just barely), what is P(X ≤ 1)? Based on the answer to part (a), if exactly 80% of the policyholders have smoke detectors, would one policyholder with a smoke detector in a sample of size 8 be an unusually small number? If you found that one of the eight sample policyholders had a smoke detector, would this be convincing evidence that the claim is false? Explain. If exactly 80% of the policyholders have smoke detectors, what is P(X ≤ 6)? Based on the answer to part (d), if 80% of the policyholders have smoke detectors, would six policyholders with smoke detectors in a sample of size 8 be an unusually small number? Page 215 f. If you found that six of the eight sample policyholders had smoke detectors, would this be convincing evidence that the claim is false? Explain. 21. A message consists of a string of bits (0s and 1s). Due to noise in the communications channel, each bit has probability 0.3 of being reversed (i.e., a 1 will be changed to a 0 or a 0 to a 1). To improve the accuracy of the communication, each bit is sent five times, so, for example, 0 is sent as 00000. The receiver assigns the value 0 if three or more of the bits are decoded as 0, and 1 if three or more of the bits are decoded as 1. Assume that errors occur independently. a. A 0 is sent (as 00000). What is the probability that the receiver assigns the correct value of 0? b. Assume that each bit is sent n times, where n is an odd number, and that the receiver assigns the value decoded in the majority of the bits. What is the minimum value of n necessary so that the probability that the correct value is assigned is at least 0.90? 22. Let X ∼ Bin(n, p), and let Y = n − X. Show that Y ∼ Bin(n, 1 − p). 23. Porcelain figurines are sold for $10 if flawless, and for $3 if there are minor cosmetic flaws. Of the figurines made by a certain company, 90% are flawless and 10% have minor cosmetic flaws. In a sample of 100 figurines that are sold, let Y be the revenue earned by selling them and let X be the number of them that are flawless. a. Express Y as a function of X. b. Find μY. c. Find σY. 24. One design for a system requires the installation of two identical components. The system will work if at least one of the components works. An alternative design requires four of these components, and the system will work if at least two of the four components work. If the probability that a component works is 0.9, and if the components function independently, which design has the greater probability of functioning? 25. (Requires material from Section 3.3.) Refer to Example 4.14. Estimate the probability that exactly one of the four tires has a flaw, and find the uncertainty in the estimate. 26. If p is a success probability, the quantity p/(1 − p) is called the odds. Odds are commonly estimated in medical research. The article "A Study of Twelve Southern California Communities with Differing Levels and Types of Air Pollution" (J. Peters, E. Avol, et al., The American Journal of Respiratory and Critical Care Medicine, 1999:760–767) reports an assessment of respiratory health of southern California children. Assume that 88 boys in a sample of 612 reported being diagnosed with bronchitis during the last 12 months. a. Estimate the proportion p of boys who have been diagnosed with bronchitis, and find the uncertainty in the estimate. b. (Requires material from Section 3.3.) Estimate the odds, and find the uncertainty in the estimate. 4.3 The Poisson Distribution The Poisson distribution arises frequently in scientific work. One way to think of the Poisson distribution is as an approximation to the binomial distribution when n is large and p is small. We illustrate with an example. A mass contains 10,000 atoms of a radioactive substance. The probability that a given atom will decay in a one-minute time period is 0.0002. Let X represent the number of atoms that decay in one minute. Now each atom can be thought of as a Bernoulli trial, where success occurs if the atom decays. Thus X is the number of successes in 10,000 independent Bernoulli trials, each with success probability 0.0002, so the distribution of X is Bin(10,000, 0.0002). The mean of X is μX = (10,000)(0.0002) = 2. Page 216 Another mass contains 5000 atoms, and each of these atoms has probability 0.0004 of decaying in a one-minute time interval. Let Y represent the number of atoms that decay in one minute from this mass. By the reasoning in the previous paragraph, Y ∼ Bin(5000, 0.0004) and μY = (5000)(0.0004) = 2. In each of these cases, the number of trials n and the success probability p are different, but the mean number of successes, which is equal to the product np, is the same. Now assume that we wanted to compute the probability that exactly three atoms decay in one minute for each of these masses. Using the binomial probability mass function, we would compute as follows: It turns out that these probabilities are very nearly equal to each other. Although it is not obvious from the formula for the binomial probability mass function, it is the case that when n is large and p is small the mass function depends almost entirely on the mean np, and very little on the specific values of n and p. We can therefore approximate the binomial mass function with a quantity that depends on the product np only. Specifically, if n is large and p is small, and we let λ = np, it can be shown by advanced methods that for all x, (4.8) We are led to define a new probability mass function, called the Poisson probability mass function. The Poisson probability mass function is defined by (4.9) If X is a random variable whose probability mass function is given by Equation (4.9), then X is said to have the Poisson distribution with parameter λ. The notation is X ∼ Poisson(λ). Example 4.15 If X ∼ Poisson(3), compute P(X = 2), P(X = 10), P(X = 0), P(X = −1), and P(X = 0.5). Solution Using the probability mass function (4.9), with λ = 3, we obtain P(X = −1) = 0 because −1 is not in the range of possible values for X. P(X = 0.5) = 0 because 0.5 is not a non-negative integer Example 4.16 If X ∼ Poisson(4), compute P(X ≤ 2) and P(X > 1). Solution To find P(X > 1), we might try to start by writing This leads to an infinite sum that is difficult to compute. Instead, we write For the example discussed at the beginning of this section, we would compute the binomial probabilities using the Poisson mass function to approximate either P(X = x) or P(Y = x) by substituting λ = 2 into Equation (4.9). Table 4.1 (page 218) shows that the approximation is excellent. Page 218 TABLE 4.1 An example of the Poisson approximation to the binomial probability mass function* x 0 1 2 3 4 5 6 7 8 9 P(X = x), X ∼ Bin(10,000, 0.0002) 0.135308215 0.270670565 0.270697637 0.180465092 0.090223521 0.036082189 0.012023787 0.003433993 0.000858069 0.000190568 P(Y = x), Y ∼ Bin(5000, 0.0004) 0.135281146 0.270670559 0.270724715 0.180483143 0.090223516 0.036074965 0.012017770 0.003430901 0.000856867 0.000190186 Poisson Approximation, Poisson(2) 0.135335283 0.270670566 0.270670566 0.180447044 0.090223522 0.036089409 0.012029803 0.003437087 0.000859272 0.000190949 *When n is large and p is small, the Bin(n, p) probability mass function is well approximated by the Poisson(λ) probability mass function (Equation 4.9), with λ = np. Here X ∼ Bin(10,000, 0.0002) and Y ∼ Bin(5000, 0.0004), so λ = np = 2, and the Poisson approximation is Poisson(2). Summary If X ∼ Poisson(λ), then ■ X is a discrete random variable whose possible values are the non-negative integers. ■ The parameter λ is a positive constant. ■ The probability mass function of X is ■ The Poisson probability mass function is very close to the binomial probability mass function when n is large, p is small, and λ = np. The Mean and Variance of a Poisson Random Variable To compute the mean and variance of a Poisson random variable, we can use the probability mass function along with the definitions given by Equations (2.29) and (2.30) (in Section 2.4). Rigorous derivations of the mean and variance using this method are given at the end of the section. We present an intuitive approach here. If X ∼ Poisson(λ), we can think of X as a binomial random variable with large n, small p, and np = λ. Since the mean of a binomial random variable is np, it follows that the mean of a Poisson random variable is λ. The variance of a binomial random variable is np(1 − p). Since p is very small, we replace 1 − p with 1, and conclude that the variance of a Poisson random variable is np = λ. Note that the variance of a Poisson random variable is equal to its mean. Page 219 Summary If X ∼ Poisson(λ), then the mean and variance of X are given by (4.10) (4.11) Figure 4.3 presents probability histograms for the Poisson(1) and Poisson(10) probability mass functions. FIGURE 4.3 (a) The Poisson(1) probability histogram. (b) The Poisson(10) probability histogram. One of the earliest industrial uses of the Poisson distribution involved an application to the brewing of beer. A crucial step in the brewing process is the addition of yeast culture to prepare mash for fermentation. The living yeast cells are kept suspended in a liquid medium. Because the cells are alive, their concentration in the medium changes over time. Therefore, just before the yeast is added, it is necessary to estimate the concentration of yeast cells per unit volume of suspension, so as to be sure to add the right amount. Up until the early part of the twentieth century, this posed a problem for brewers. They estimated the concentration in the obvious way, by withdrawing a small volume of the suspension and counting the yeast cells in it under a microscope. Of course, the estimates determined this way were subject to uncertainty, but no one knew how to compute the uncertainty. Thus no one knew by how much the concentration in the sample was likely to differ from the actual concentration. William Sealy Gosset, a young man in his mid-twenties who was employed by the Guinness Brewing Company of Dublin, Ireland, discovered in 1904 that the number of yeast cells in a sampled volume of suspension follows a Poisson distribution. He was then able to develop methods to compute the needed uncertainty. Gosset's discovery not only enabled Guinness to produce a more consistent product, it showed that the Poisson distribution could have important applications in many situations. Gosset wanted to publish his result, but his managers at Guinness considered his discovery to be proprietary information, and required him to use the pseudonym "Student." In Example 4.17, we will follow a train of thought that leads to Student's result. Before we get to it though, we will mention that shortly after publishing this result, Student made another discovery that solved one of the most important outstanding problems in statistics, and which has profoundly influenced work in virtually all fields of science ever since. We will discuss this result in Section 5.3. Example 4.17 Particles (e.g., yeast cells) are suspended in a liquid medium at a concentration of 10 particles per mL. A large volume of the suspension is thoroughly agitated, and then 1 mL is withdrawn. What is the probability that exactly eight particles are withdrawn? Solution So long as the volume withdrawn is a small fraction of the total, the solution to this problem does not depend on the total volume of the suspension, but only on the concentration of particles in it. Let V be the total volume of the suspension, in mL. Then the total number of particles in the suspension is 10V. Think of each of the 10V particles as a Bernoulli trial. A particle "succeeds" if it is withdrawn. Now 1 mL out of the total of V mL is to be withdrawn. Therefore, the amount to be withdrawn is 1/V of the total, so it follows that each particle has probability 1/V of being withdrawn. Let X denote the number of particles withdrawn. Then X represents the number of successes in 10V Bernoulli trials, each with probability 1/V of success. Therefore X ∼ Bin(10V, 1/V). Since V is large, 10V is large and 1/V is small. Thus to a very close approximation, X ∼ Poisson(10). We compute P(X = 8) with the Poisson probability mass function: P(X = 8) = e−10108/8! = 0.1126. In Example 4.17, λ had the value 10 because the mean number of particles in 1 mL of suspension (the volume withdrawn) was 10. Example 4.18 Particles are suspended in a liquid medium at a concentration of 6 particles per mL. A large volume of the suspension is thoroughly agitated, and then 3 mL are withdrawn. What is the probability that exactly 15 particles are withdrawn? Solution Let X represent the number of particles withdrawn. The mean number of particles in a 3 mL volume is 18. Therefore X ∼ Poisson(18). The probability that exactly 15 particles are withdrawn is Page 221 Note that for the solutions to Examples 4.17 and 4.18 to be correct, it is important that the amount of suspension withdrawn not be too large a fraction of the total. For example, if the total volume in Example 4.18 was 3 mL, so that the entire amount was withdrawn, it would be certain that all 18 particles would be withdrawn, so the probability of withdrawing 15 particles would be zero. Example 4.19 Grandma bakes chocolate chip cookies in batches of 100. She puts 300 chips into the dough. When the cookies are done, she gives you one. What is the probability that your cookie contains no chocolate chips? Solution This is another instance of particles in a suspension. Let X represent the number of chips in your cookie. The mean number of chips is 3 per cookie, so X ∼ Poisson(3). It follows that P(X = 0) = e−330/0! = 0.0498. Example 4.20 Grandma's grandchildren have been complaining that Grandma is too stingy with the chocolate chips. Grandma agrees to add enough chips to the dough so that only 1% of the cookies will contain no chips. How many chips must she include in a batch of 100 cookies to achieve this? Solution Let n be the number of chips to include in a batch of 100 cookies, and let X be the number of chips in your cookie. The mean number of chips is 0.01 n per cookie, so X ∼ Poisson(0.01n). We must find the value of n for which P(X = 0) = 0.01. Using the Poisson(0.01n) probability mass function, Setting e−0.01n = 0.01, we obtain n ≈ 461. Examples 4.17 through 4.20 show that for particles distributed uniformly at random throughout a medium, the number of particles that happen to fall in a small portion of the medium follows a Poisson distribution. In these examples, the particles were actual particles and the medium was spatial in nature. There are many cases, however, when the "particles" represent events and the medium is time. We saw such an example at the beginning of this section, where the number of radioactive decay events in a fixed time interval turned out to follow a Poisson distribution. Example 4.21 presents another. Example 4.21 Assume that the number of cars that pass through a certain intersection during a fixed time interval follows a Poisson distribution. Assume that the mean rate is five cars per minute. Find the probability that exactly 17 cars will pass through the intersection in the next three minutes. Page 222 Solution Let X represent the number of cars that pass through the intersection in the next three minutes. The mean number of cars is (5)(3) = 15, so X ∼ Poisson(15). Using the Poisson(15) probability mass function, Example 4.22 In Example 4.21, let X be the number of cars passing through the intersection in t minutes. Find the probability mass function of X, in terms of t. Solution The mean number of cars in t minutes is 5t, so X ∼ Poisson(5t). The probability mass function of X is Using the Poisson Distribution to Estimate a Rate Often experiments are done to estimate a rate λ that represents the mean number of events that occur in one unit of time or space. In these experiments, the number of events X that occur in t units of time or space is counted, and the rate λ is estimated with the quantity . (Note that since the quantity X/t is used to estimate λ, it is denoted .) If the numbers of events in disjoint intervals of time or space are independent, and if events cannot occur simultaneously, then X follows a Poisson distribution. A process that produces such events is called a Poisson process. Since the mean number of events that occur in t units of time or space is equal to λt, X ∼ Poisson(λt). Summary Let λ denote the mean number of events that occur in one unit of time or space. Let X denote the number of events that are observed to occur in t units of time or space. Then if X ∼ Poisson(λt), λ is estimated with . Example 4.23 A suspension contains particles at an unknown concentration of λ per mL. The suspension is thoroughly agitated, and then 4 mL are withdrawn and 17 particles are counted. Estimate λ. Solution Let X = 17 represent the number of particles counted, and let t = 4 mL be the volume of suspension withdrawn. Then particles per mL. Page 223 Uncertainty in the Estimated Rate It is important to realize that the estimated rate or concentration is just an estimate of the true rate or concentration λ. In general is just an estimate of the true does not equal λ. If the experiment were repeated, the value of would probably come out differently. In other words, there is uncertainty in . For to be useful, we must compute its bias and its uncertainty. The calculations are similar to those for the sample proportion that were presented in Section 4.2. Let X be the number of events counted in t units of time or space, and assume that X ∼ Poisson(λt). The bias is the difference . Since , it follows from Equation (2.43) (in Section 2.5) that Since , is unbiased. The uncertainty is the standard deviation , it follows from Equation . Since X ∼ Poisson(λt), it follows from Equation (4.11) (2.45) (in Section 2.5) that that . Since . Therefore In practice, the value of λ is unknown, so we approximate it with . Summary If X ∼ Poisson(λt), we estimate the rate λ with ■ ■ . is unbiased. The uncertainty in is (4.12) In practice, we substitute for λ in Equation (4.12), since λ is unknown. Example 4.24 A 5 mL sample of a suspension is withdrawn, and 47 particles are counted. Estimate the mean number of particles per mL, and find the uncertainty in the estimate. Page 224 Solution The number of particles counted is X = 47. The volume withdrawn is t = 5 mL. The estimated mean number of particles per mL is The uncertainty in the estimate is Example 4.25 A certain mass of a radioactive substance emits alpha particles at a mean rate of λ particles per second. A physicist counts 1594 emissions in 100 seconds. Estimate λ, and find the uncertainty in the estimate. Solution The estimate of λ is emissions per second. The uncertainty is Example 4.26 In Example 4.25, for how many seconds should emissions be counted to reduce the uncertainty to 0.3 emissions per second? Solution We want to find the time t for which . From Example 4.25, . Substituting this value for λ, we obtain Solving for t yields t = 177 seconds. The following example requires knowledge of propagation of error, which is covered in Section 3.3. Page 225 Example 4.27 The number of flaws on a sheet of aluminum manufactured by a certain process follows a Poisson distribution. In a sample of 100 m2 of aluminum, 200 flaws are counted. Estimate the probability that a given square meter of aluminum has no flaws, and find the uncertainty in the estimate. Solution Let λ represent the mean number of flaws per square meter. We will begin by computing and its uncertainty. We have observed X = 200 flaws in t = 100 m2 of aluminum. Therefore . The uncertainty in is What we need to estimate is the probability that a square meter of aluminum contains no flaws. We first express this probability as a function of λ. To do this, let Y represent the number of flaws in a 1 m2 sheet of aluminum. Then Y ∼ Poisson(λ). We want to estimate P(Y = 0). Using the Poisson probability mass function, this probability is given by The probability that a square meter contains no flaws is therefore estimated with . To find the uncertainty in this estimate, we use the propagation of error method (Equation 3.10). In the case of particles in a suspension, or radioactive decay events, enough is known about the underlying physical principles governing these processes that we were able to argue from first principles to show that the distribution of the number of events is Poisson. There are many other cases where empirical evidence suggests that the Poisson distribution may be appropriate, but the laws governing the process are not sufficiently well understood to make a rigorous derivation possible. Examples include the number of hits on a website, the number of traffic accidents at an intersection, and the number of trees in a section of forest. Page 226 Derivation of the Mean and Variance of a Poisson Random Variable Let X ∼ Poisson(λ). We will show that μX = λ and . Using the definition of population mean for a discrete random variable (Equation 2.29 in Section 2.4): Now the sum is the sum of the Poisson(λ) probability mass function over all its possible values. Therefore , so We use Equation (2.31) (in Section 2.4) to show that , so . (4.13) Substituting x(x − 1) + x for x2 and λ for μX in Equation (4.13), we obtain (4.14) Now x(x − 1) = 0 if x = 0 or 1, and we may therefore begin the sum on the right-hand side of Equation (4.14) at x = 2, and substitute λ for . We obtain Page 227 Exercises for Section 4.3 1. Let X ∼ Poisson(4). Find a. P(X = 1) b. P(X = 0) c. P(X < 2) d. P(X > 1) e. μX f. σX 2. The number of pits in a corroded steel coupon follows a Poisson distribution with a mean of 6 pits per cm2. Let X represent the number of pits in a 1 cm2 area. Find a. P(X = 8) b. P(X = 2) c. P(X < 3) d. P(X > 1) e. μX f. σX 3. The number of large packages delivered by a courier service follows a Poisson distribution with a rate of 5 per day. Let X be the number of large packages delivered on a given day. Find a. P(X = 6) b. P(X ≤ 2) c. P(5 < X < 8) d. μX e. σX 4. Geologists estimate the time since the most recent cooling of a mineral by counting the number of uranium fission tracks on the surface of the mineral. A certain mineral specimen is of such an age that there should be an average of 6 tracks per cm2 of surface area. Assume the number of tracks in an area follows a Poisson distribution. Let X represent the number of tracks counted in 1 cm2 of surface area. Find a. P(X = 7) b. P(X ≥ 3) c. P(2 < X < 7) d. μX e. σX 5. A data center contains 1000 computer servers. Each server has probability 0.003 of failing on a given day. a. What is the probability that exactly two servers fail? b. What is the probability that fewer than 998 servers function? c. What is the mean number of servers that fail? d. What is the standard deviation of the number of servers that fail? One out of every 5000 individuals in a population carries a certain defective gene. A random sample of 1000 individuals is studied. a. What is the probability that exactly one of the sample individuals carries the gene? b. What is the probability that none of the sample individuals carries the gene? c. What is the probability that more than two of the sample individuals carry the gene? d. What is the mean of the number of sample individuals that carry the gene? e. What is the standard deviation of the number of sample individuals that carry the gene? Page 228 7. 8. 9. 6. The number of hits on a certain website follows a Poisson distribution with a mean rate of 4 per minute. a. What is the probability that 5 hits are received in a given minute? b. What is the probability that 9 hits are received in 1.5 minutes? c. What is the probability that fewer than 3 hits are received in a period of 30 seconds? The number of cars arriving at a given intersection follows a Poisson distribution with a mean rate of 4 per second. a. What is the probability that 3 cars arrive in a given second? b. What is the probability that 8 cars arrive in three seconds? c. What is the probability that more than 3 cars arrive in a period of two seconds? A random variable X has a binomial distribution, and a random variable Y has a Poisson distribution. Both X and Y have means equal to 3. Is it possible to determine which random variable has the larger variance? Choose one of the following answers: i. Yes, X has the larger variance. ii. Yes, Y has the larger variance. iii. No, we need to know the number of trials, n, for X. iv. No, we need to know the success probability, p, for X. v. No, we need to know the value of λ for Y. 10. 11. 12. 13. 14. 15. A chemist wishes to estimate the concentration of particles in a certain suspension. She withdraws 3 mL of the suspension and counts 48 particles. Estimate the concentration in particles per mL, and find the uncertainty in the estimate. A microbiologist wants to estimate the concentration of a certain type of bacterium in a wastewater sample. She puts a 0.5 mL sample of the wastewater on a microscope slide and counts 39 bacteria. Estimate the concentration of bacteria, per mL, in this wastewater, and find the uncertainty in the estimate. Two-dimensional Poisson process. The number of plants of a certain species in a certain forest has a Poisson distribution with mean 10 plants per acre. The number of plants in T acres therefore has a Poisson distribution with mean 10T. a. What is the probability that there will be exactly 18 plants in a two-acre region? b. What is the probability that there will be exactly 12 plants in a circle with radius 100 ft? (1 acre = 43,560 ft2.) c. The number of plants of a different type follows a Poisson distribution with mean λ plants per acre, where λ is unknown. A total of 5 plants are counted in a 0.1 acre area. Estimate λ, and find the uncertainty in the estimate. The number of defective components produced by a certain process in one day has a Poisson distribution with mean 20. Each defective component has probability 0.60 of being repairable. a. Find the probability that exactly 15 defective components are produced. b. Given that exactly 15 defective components are produced, find the probability that exactly 10 of them are repairable. c. Let N be the number of defective components produced, and let X be the number of them that are repairable. Given the value of N, what is the distribution of X? d. Find the probability that exactly 15 defective components are produced, with exactly 10 of them being repairable. The probability that a certain radioactive mass emits no particles in a one-minute time period is 0.1353. What is the mean number of particles emitted per minute? The number of flaws in a certain type of lumber follows a Poisson distribution with a rate of 0.45 per linear meter. Page 229 a. What is the probability that a board 3 meters in length has no flaws? b. How long must a board be so that the probability it has no flaw is 0.5? 16. Grandma is trying out a new recipe for raisin bread. Each batch of bread dough makes three loaves, and each loaf contains 20 slices of bread. a. If she puts 100 raisins into a batch of dough, what is the probability that a randomly chosen slice of bread contains no raisins? b. If she puts 200 raisins into a batch of dough, what is the probability that a randomly chosen slice of bread contains 5 raisins? c. How many raisins must she put in so that the probability that a randomly chosen slice will have no raisins is 0.01? 17. Mom and Grandma are each baking chocolate chip cookies. Each gives you two cookies. One of Mom's cookies has 14 chips in it and the other has 11. Grandma's cookies have 6 and 8 chips. a. Estimate the mean number of chips in one of Mom's cookies. b. Estimate the mean number of chips in one of Grandma's cookies. c. Find the uncertainty in the estimate for Mom's cookies. d. Find the uncertainty in the estimate for Grandma's cookies. e. Estimate how many more chips there are on the average in one of Mom's cookies than in one of Grandma's. Find the uncertainty in this estimate. 18. You have received a radioactive mass that is claimed to have a mean decay rate of at least 1 particle per second. If the mean decay rate is less than 1 per second, you may return the product for a refund. Let X be the number of decay events counted in 10 seconds. a. If the mean decay rate is exactly 1 per second (so that the claim is true, but just barely), what is P(X ≤ 1)? b. Based on the answer to part (a), if the mean decay rate is 1 particle per second, would one event in 10 seconds be an unusually small number? c. If you counted one decay event in 10 seconds, would this be convincing evidence that the product should be returned? Explain. d. If the mean decay rate is exactly 1 per second, what is P(X ≤ 8)? e. Based on the answer to part (d), if the mean decay rate is 1 particle per second, would eight events in 10 seconds be an unusually small number? f. If you counted eight decay events in 10 seconds, would this be convincing evidence that the product should be returned? Explain. 19. Someone claims that a certain suspension contains at least seven particles per mL. You sample 1 mL of solution. Let X be the number of particles in the sample. a. If the mean number of particles is exactly seven per mL (so that the claim is true, but just barely), what is P(X ≤ 1)? b. Based on the answer to part (a), if the suspension contains seven particles per mL, would one particle in a 1 mL sample be an unusually small number? c. If you counted one particle in the sample, would this be convincing evidence that the claim is false? Explain. d. If the mean number of particles is exactly 7 per mL, what is P(X ≤ 6)? e. Based on the answer to part (d), if the suspension contains seven particles per mL, would six particles in a 1 mL sample be an unusually small number? f. If you counted six particles in the sample, would this be convincing evidence that the claim is false? Explain. 20. A physicist wants to estimate the rate of emissions of alpha particles from a certain source. He makes two counts. First, he measures the background rate by counting the number of particles in 100 seconds in the absence of the source. He counts 36 background emissions. Then, with the source present, he counts 324 emissions in 100 seconds. This represents the sum of source emissions plus background emissions. a. Estimate the background rate, in emissions per second, and find the uncertainty in the estimate. b. Estimate the sum of the source plus background rate, in emissions per second, and find the uncertainty in the estimate. c. Estimate the rate of source emissions in particles per second, and find the uncertainty in the estimate. d. Page 230 Which will provide the smaller uncertainty in estimating the rate of emissions from the source: (1) counting the background only for 150 seconds and the background plus the source for 150 seconds, or (2) counting the background for 100 seconds and the source plus the background for 200 seconds? Compute the uncertainty in each case. e. Is it possible to reduce the uncertainty to 0.03 particles per second if the background rate is measured for only 100 seconds? If so, for how long must the source plus background be measured? If not, explain why not. 21. (Requires material from Section 3.3.) Refer to Example 4.27. Estimate the probability that a 1 m2 sheet of aluminum has exactly one flaw, and find the uncertainty in this estimate. 4.4 Some Other Discrete Distributions In this section we discuss several discrete distributions that are useful in various situations. The Hypergeometric Distribution When a finite population contains two types of items, which may be called successes and failures, and a simple random sample is drawn from the population, each item sampled constitutes a Bernoulli trial. As each item is selected, the proportion of successes in the remaining population decreases or increases, depending on whether the sampled item was a success or a failure. For this reason the trials are not independent, so the number of successes in the sample does not follow a binomial distribution. Instead, the distribution that properly describes the number of successes in this situation is called the hypergeometric distribution. As an example, assume that a lot of 20 items contains 6 that are defective, and that 5 items are sampled from this lot at random. Let X be the number of defective items in the sample. We will compute P(X = 2). To do this, we first count the total number of different groups of 5 items that can be sampled from the population of 20. (We refer to each group of 5 items as a combination.) The number of combinations of 5 items is the number of different samples that can be drawn, and each of them is equally likely. Then we determine how many of these combinations of 5 items contain exactly 2 defectives. The probability that a combination of 5 items contains exactly 2 defectives is the quotient of the number of combinations of 5 items that contain 2 defectives divided by the total number of combinations of 5 items. In general, the number of combinations of k items that can be chosen from a group of n items is denoted and is equal to (see Example 2.12 in Section 2.2 for a derivation) The number of combinations of 5 items that can be chosen from 20 is therefore Page 231 To determine the number of combinations of 5 that contain exactly 2 defectives, we describe the construction of such a combination as a sequence of two operations. First, select 2 items from the 6 defective ones; second, select 3 items from the 14 nondefective ones. The number of combinations of 2 items chosen from 6 and the number of combinations of 3 items chosen from 14 is The total number of combinations of 5 items that can be made up of 2 defectives and 3 nondefectives is therefore the product (this is an application of the fundamental principle of counting; see Section 2.2 for a more detailed discussion). We conclude that To compute P(X = 2) in the preceding example, it was necessary to know the number of items in the population (20), the number of defective items in the population (6), and the number of items sampled (5). The probability mass function of the random variable X is determined by these three parameters. Specifically, X is said to have the hypergeometric distribution with parameters 20, 6, and 5, which we denote X ∼ H(20, 6, 5). We now generalize this idea. Summary Assume a finite population contains N items, of which R are classified as successes and N − R are classified as failures. Assume that n items are sampled from this population, and let X represent the number of successes in the sample. Then X has the hypergeometric distribution with parameters N, R, and n, which can be denoted X ∼ H(N, R, n). The probability mass function of X is (4.15) Page 232 Example 4.28 Of 50 buildings in an industrial park, 12 have electrical code violations. If 10 buildings are selected at random for inspection, what is the probability that exactly 3 of the 10 have code violations? Solution Let X represent the number of sampled buildings that have code violations. Then X ∼ H(50, 12, 10). We must find P(X = 3). Using Equation (4.15), Mean and Variance of the Hypergeometric Distribution The mean and variance of the hypergeometric distribution are presented in the following box. Their derivations are omitted. If X ∼ H(N, R, n), then (4.16) (4.17) Example 4.29 Refer to Example 4.28. Find the mean and variance of X. Solution X ∼ H(50, 12, 10), so Page 233 Comparison with the Binomial Distribution A population of size N contains R successes and N − R failures. Imagine that a sample of n items is drawn from this population with replacement, that is, with each sampled item being returned to the population after it is sampled. Then the sampled items result from a sequence of n independent Bernoulli trials, and the number of successes X in the sample has a binomial distribution with n trials and success probability p = R/N. In practice, samples are seldom drawn with replacement, because there is no need to sample the same item twice. Instead, sampling is done without replacement, where each item is removed from the population after it is sampled. The sampled items then result from dependent Bernoulli trials, because the population changes as each item is sampled. For this reason the distribution of the number of successes, X, is H(N, R, n) rather than Bin(n, R/N). When the sample size n is small compared to the population size N (i.e., no more than 5%), the difference between sampling with and without replacement is slight, and the binomial distribution Bin(n, R/N) is a good approximation to the hypergeometric distribution H(N, R, n). Note that the mean of H(N, R, n) is nR/N, the same as that of Bin(n, R/N). This indicates that whether the sampling is done with or without replacement, the proportion of successes in the sample is the same on the average as the proportion of successes in the population. The variance of Bin(n, R/N) is n(R/N)(1 − R/N), and the variance of H(N, R, n) is obtained by multiplying this by the factor (N − n)/(N − 1). Note that when n is small relative to N, this factor is close to 1. The Geometric Distribution Assume that a sequence of independent Bernoulli trials is conducted, each with the same success probability p. Let X represent the number of trials up to and including the first success. Then X is a discrete random variable, which is said to have the geometric distribution with parameter p. We write X ∼ Geom(p). Example 4.30 A large industrial firm allows a discount on any invoice that is paid within 30 days. Twenty percent of invoices receive the discount. Invoices are audited one by one. Let X be the number of invoices audited up to and including the first one that qualifies for a discount. What is the distribution of X? Solution Each invoice is a Bernoulli trial, with success occurring if the invoice qualifies for a discount. The success probability is therefore p = 0.2. The number of trials up to and including the first success has a geometric distribution with parameter p = 0.2. Therefore X ∼ Geom(0.2). Page 234 Example 4.31 Refer to Example 4.30. Find P(X = 3). Solution The event X = 3 occurs when the first two trials result in failure and the third trial results in success. It follows that The result of Example 4.31 can be generalized to produce the probability mass function of a geometric random variable. If X ∼ Geom(p), then the probability mass function of X is The Mean and Variance of a Geometric Distribution The mean and variance of the geometric distribution are given in the following box. Their derivations require the manipulation of infinite series and are omitted. If X ∼ Geom(p), then (4.18) (4.19) Example 4.32 Refer to Example 4.30. Let X denote the number of invoices up to and including the first one that qualifies for a discount. Find the mean and variance of X. Solution Since X ∼ Geom(0.2), μX = 1/0.2 = 5, and . The Negative Binomial Distribution The negative binomial distribution is an extension of the geometric distribution. Let r be a positive integer. Assume that independent Bernoulli trials, each with success probability p, are conducted, and let X denote the number of trials up to and including the rth success. Page 235 Then X has the negative binomial distribution with parameters r and p. We write X ∼ NB(r, p). Example 4.33 (Continuing Example 4.30.) In a large industrial firm, 20% of all invoices receive a discount. Invoices are audited one by one. Let X denote the number of invoices up to and including the third one that qualifies for a discount. What is the distribution of X? Find P(X = 8). Solution Since X represents the number of trials up to and including the third success, and since the success probability is p = 0.2, X ∼ NB(3, 0.2). We will compute P(X = 8), and the method of computation will lead to a derivation of the probability mass function of a negative binomial random variable. Since X ∼ NB(3, 0.2), the event X = 8 means that the third success occurred on the eighth trial. Another way to say this is that there were exactly two successes in the first 7 trials, and the eighth trial was a success. Since all the trials are independent, it follows that P(X = 8) = P(exactly 2 successes in first 7 trials)P(success on eighth trial) Now the number of successes in the first 7 trials has the Bin(7, 0.2) distribution, so The probability that the eighth trial (or any other trial) results in success is 0.2. Therefore We generalize the result of Example 4.33 to produce the probability mass function of a negative binomial random variable. If X ∼ NB(r, p), then the probability mass function of X is Note that the smallest possible value for X is r, since it takes at least r trials to Page 236 produce r successes. Note also that when r = 1, the negative binomial distribution is the same as the geometric distribution. In symbols, NB(1, p) = Geom(p). A Negative Binomial Random Variable Is a Sum of Geometric Random Variables Assume that a sequence of 8 independent Bernoulli trials is conducted, each with success probability p, comes out as follows: If X is the number of trials up to and including the third success, then X ∼ NB(3, p), and for this sequence of trials, X = 8. Denote the number of trials up to and including the first success by Y1. For this sequence, Y1 = 3, but in general, Y1 ∼ Geom(p). Now count the number of trials, starting with the first trial after the first success, up to and including the second success. Denote this number of trials by Y2. For this sequence Y2 = 2, but in general, Y2 ∼ Geom(p). Finally, count the number of trials, beginning from the first trial after the second success, up to and including the third success. Denote this number of trials by Y3. For this sequence Y3 = 3, but again in general, Y3 ∼ Geom(p). It is clear that X = Y1 + Y2 + Y3. Furthermore, since the trials are independent, Y1, Y2, and Y3 are independent. This shows that if X ∼ NB(3, p), then X is the sum of three independent Geom(p) random variables. This result can be generalized to any positive integer r. Summary If X ∼ NB(r, p), then where Y1, ..., Yr are independent random variables, each with the Geom(p) distribution. The Mean and Variance of the Negative Binomial Distribution If X ∼ NB(r, p), then X = Y1 + ··· + Yr, where Y1, ..., Yr are independent random variables, each with the Geom(p) distribution. It follows that the mean of X is the sum of the means of the Ys, and the variance of X is the sum of the variances. Each Yi has mean 1/p and variance (1 − p)/p2. Therefore μX = r/p and σ2X = r(1 − p)/p2. Summary If X ∼ NB(r, p), then (4.20) (4.21) Page 237 Example 4.34 Refer to Example 4.33. Find the mean and variance of X, where X represents the number of invoices up to and including the third one that qualifies for a discount. Solution Since X ∼ NB(3, 0.2), it follows that The Multinomial Distribution A Bernoulli trial is a process that results in one of two possible outcomes. A generalization of the Bernoulli trial is the multinomial trial, which is a process that can result in any of k outcomes, where k ≥ 2. For example, the rolling of a die is a multinomial trial, with the six possible outcomes 1, 2, 3, 4, 5, 6. Each outcome of a multinomial trial has a probability of occurrence. We denote the probabilities of the k outcomes by p1, ..., pk. For example, in the roll of a fair die, p1 = p2 = ··· = p6 = 1/6. Now assume that n independent multinomial trials are conducted, each with the same k possible outcomes and with the same probabilities p1, ..., pk. Number the outcomes 1, 2, ..., k. For each outcome i, let Xi denote the number of trials that result in that outcome. Then X1, ..., Xk are discrete random variables. The collection X1, ..., Xk is said to have the multinomial distribution with parameters n, p1, ..., pk. We write X1, ..., Xk ∼ MN(n, p1, ..., pk). Note that it is the whole collection X1, ..., Xk that has the multinomial distribution, rather than any single Xi. Example 4.35 The items produced on an assembly line are inspected, and each is classified as either conforming (acceptable), downgraded, or rejected. Overall, 70% of the items are conforming, 20% are downgraded, and 10% are rejected. Assume that four items are chosen independently and at random. Let X1, X2, X3 denote the numbers among the 4 that are conforming, downgraded, and rejected, respectively. What is the distribution of X1, X2, X3? Solution Each item is a multinomial trial with three possible outcomes, conforming, downgraded, and rejected. The probabilities associated with the outcomes are p1 = 0.7, p2 = 0.2, and p2 = 0.1. The random variables X1, X2, X3 refer to the numbers of each outcome in 4 independent trials. Therefore X1, X2, X3 ∼ MN(4, 0.7, 0.2, 0.1). To show how to compute probabilities concerning multinomial random variables, we will compute P(X1 = 2, X2 = 1, and X3 = 1), where X1, X2, X3 are defined in Example 4.35. Page 238 This will lead to a derivation of the multinomial probability mass function. We begin by noting that there are 12 arrangements of two conforming (C), one downgraded (D), and one rejected (R) among four trials. They are listed here. Each of these 12 arrangements is equally probable. We compute the probability of CCDR. The event CCDR is a sequence of four outcomes: C on the first trial, C on the second trial, D on the third trial, and R on the fourth trial. Since the trials are independent, the probability of the sequence of outcomes is equal to the product of their individual probabilities. Since each of the 12 arrangements has the same probability, In this calculation, the number of arrangements was small enough to count by listing them all. To compute probabilities like this in general, we need a formula. The formula is given in the following box. A derivation is presented in Section 2.2. Assume n independent trials are performed, each of which results in one of k possible outcomes. Let x1, ..., xk be the numbers of trials resulting in outcomes 1, 2, ..., k, respectively. The number of arrangements of the outcomes among the n trials is We can now specify the multinomial probability mass function. If X1, ..., Xk ∼ MN(n, p1, ..., pk), then the probability mass function of X1, ..., Xk is (4.22) Note that the multinomial distribution differs from the other distributions that we have studied in that it concerns several random variables simultaneously. We can express this fact by saying that p(x1, ..., xk) is the joint probability mass function of X1, ..., Xk. Joint probability mass functions are discussed more fully in Section 2.6. Example Page 239 4.36 Alkaptonuria is a genetic disease that results in the lack of an enzyme necessary to break down homogentisic acid. Some people are carriers of alkaptonuria, which means that they do not have the disease themselves, but they can potentially transmit it to their offspring. According to the laws of genetic inheritance, an offspring both of whose parents are carriers of alkaptonuria has probability 0.25 of being unaffected, 0.5 of being a carrier, and 0.25 of having the disease. In a sample of 10 offspring of carriers of alkaptonuria, what is the probability that 3 are unaffected, 5 are carriers, and 2 have the disease? Solution Let X1, X2, X3 denote the numbers among the 10 offspring who are unaffected, carriers, and diseased, respectively. Then X1, X2, X3 ∼ MN(10, 0.25, 0.50, 0.25). It follows from Equation (4.22) that Sometimes we want to focus on only one of the possible outcomes of a multinomial trial. In this situation, we can consider the outcome of interest a "success," and any other outcome a "failure." In this way it can be seen that the number of occurrences of any particular outcome has a binomial distribution. If X1, ..., Xk ∼ MN(n, p1, ..., pk), then for each i Example 4.37 Refer to Example 4.36. Find the probability that exactly 4 of 10 offspring are unaffected. Solution Let X represent the number of unaffected offspring in a sample of 10. Then X ∼ Bin(10, 0.25), so Page 240 Exercises for Section 4.4 1. Twelve items are delivered to a warehouse. Ten of the items are to be worked on. What is the probability that three of them need new brakes? There are 30 restaurants in a certain town. Assume that four of them have health code violations. A health inspector chooses 10 restaurants at random to visit. a. What is the probability that two of the restaurants with health code violations will be visited? b. What is the probability that none of the restaurants that are visited will have health code violations? 3. 4. The probability that a computer running a certain operating system crashes on any given day is 0.1. Find the probability that the computer crashes for the first time on the twelfth day after the operating system is installed. A traffic light at a certain intersection is green 50% of the time, yellow 10% of the time, and red 40% of the time. A car approaches this intersection once each day. Let X represent the number of days that pass up to and including the first time the car encounters a red light. Assume that each day represents an independent trial. a. Find P(X = 3). b. Find P(X ≤ 3). c. Find μX. d. Find σ2X. Refer to Exercise 4. Let Y denote the number of days up to and including the third day on which a red light is encountered. a. Find P(Y = 7). b. Find μY. c. Find σ2Y. Refer to Exercise 4. What is the probability that in a sequence of 10 days, four green lights, one yellow light, and five red lights are encountered? 7. If X ∼ Geom(p), what is the most probable value of X? i. 0. ii. 1/p. iii. p. iv. 1. v. (1 − p)/p2. 8. A process that fills packages is stopped whenever a package is detected whose weight falls outside the specification. Assume that each package has probability 0.01 of falling outside the specification and that the weights of the packages are independent. a. Find the mean number of packages that will be filled before the process is stopped. b. Find the variance of the number of packages that will be filled before the process is stopped. c. Assume that the process will not be stopped until four packages whose weight falls outside the specification are detected. Find the mean and variance of the number of packages that will be filled before the process is stopped. 9. A system is tested for faults once per hour. If there is no fault, none will be detected. If there is a fault, the probability is 0.8 that it will be detected. The tests are independent of one another. a. If there is a fault, what is the probability that it will be detected in 3 hours or less? b. Given that a fault has gone undetected for 2 hours, what is the probability that it will be detected in the next hour? c. What is the mean number of tests that must be conducted in order to detect a fault? 10. A computer program has a bug that causes it to fail once in every thousand runs, on average. In an effort to find the bug, independent runs of the program will be made until the program has failed five times. a. What is the mean number of runs required? b. What is the standard deviation of the number of runs required? 11. In a lot of 15 truss rods, 12 meet a tensile strength specification. Four rods are chosen at random to be tested. Let X be the number of tested rods that meet the specification. a. Find P(X = 3). b. Find μX. c. Find σX. Page 241 12. A lot of parts contains 500 items, 100 of which are defective. Suppose that 20 items are selected at random. Let X be the number of selected items that are defective. a. Express the quantity P(X = 5) using factorials. b. Use the binomial approximation to compute an approximation to P(X = 5). 13. Ten items are to be sampled from a lot of 60. If more than one is defective, the lot will be rejected. Find the probability that the lot will be rejected in each of the following cases. a. The number of defective items in the lot is 5. b. The number of defective items in the lot is 10. c. The number of defective items in the lot is 20. 14. Among smartphone users, 40% use a case but no screen protector, 10% use a screen protector but no case, 45% use both a case and a screen protector, and 5% use neither a case nor a screen protector. Twenty smartphone users are sampled at random. Let X1, X2, X3, X4 be the respective numbers of users in the four given categories. a. Find P(X1 = 7, X2 = 3, X3 = 8, and X4 = 2). b. Find P(X1 = 7). 15. At a certain fast-food restaurant, 25% of drink orders are for a small drink, 35% for a medium, and 40% for a large. A random sample of 20 orders is selected for audit. a. What is the probability that the numbers of orders for small, medium, and large drinks are 5, 7, and 8, respectively? b. What is the probability that more than 10 orders are for large drinks? 16. A thermocouple placed in a certain medium produces readings within 0.1°C of the true temperature 70% of the time, readings more than 0.1°C above the true temperature 10% of the time, and readings more than 0.1°C below the true temperature 20% of the time. a. In a series of 10 independent readings, what is the probability that 5 are within 0.1°C of the true temperature, 2 are more than 0.1°C above, and 3 are more than 0.1°C below? b. What is the probability that more than 8 of the readings are within 0.1°C of the true temperature? 17. Let X ∼ Geom(p), let n be a non-negative integer, and let Y ∼ Bin(n, p). Show that P(X = n) = (1/n)P(Y = 1). 18. Use the result of Exercise 17 and Table A.1 to find P(X = 10) where X ∼ Geom(0.3). 4.5 The Normal Distribution The normal distribution (also called the Gaussian distribution) is by far the most commonly used distribution in statistics. This distribution provides a good model for many, although not all, continuous populations. Part of the reason for this is the Central Limit Theorem, which we shall discuss in Section 4.11. The normal distribution is continuous rather than discrete. The mean of a normal random variable may have any value, and the variance may have any positive value. The probability density function of a normal random variable with mean μ and variance σ2 is given by (4.23) At the end of the section, we derive the fact that μ and σ2 are the mean and variance. If X is a random variable whose probability density function is normal with mean μ and variance σ2, we write X ∼ N(μ, σ2). Page 242 Summary If X ∼ N(μ, σ2), then the mean and variance of X are given by Figure 4.4 presents a plot of the normal probability density function with mean μ and standard deviation σ. The normal probability density function is sometimes called the normal curve. Note that the normal curve is symmetric around μ, so that μ is the median as well as the mean. It is also the case that for any normal population ■ About 68% of the population is in the interval μ ± σ. ■ About 95% of the population is in the interval μ ± 2σ. ■ About 99.7% of the population is in the interval μ ± 3σ. The proportion of a normal population that is within a given number of standard deviations of the mean is the same for any normal population. For this reason, when dealing with normal populations, we often convert from the units in which the population items were originally measured to standard units. Standard units tell how many standard deviations an observation is from the population mean. FIGURE 4.4 Probability density function of a normal random variable with mean μ and variance σ2. Example 4.38 Assume that the heights in a population of women follow the normal curve with mean μ = 64 inches and standard deviation σ = 3 inches. The heights of two randomly chosen women are 67 inches and 62 inches. Convert these heights to standard units. Solution A height of 67 inches is 3 inches more than the mean of 64, and 3 inches is equal to one standard deviation. So 67 inches is one standard deviation above the mean and is thus equivalent to one standard unit. A height of 62 inches is 0.67 standard deviations below the mean, so 62 inches is equivalent to −0.67 standard units. Page 243 In general, we convert to standard units by subtracting the mean and dividing by the standard deviation. Thus, if x is an item sampled from a normal population with mean μ and variance σ2, the standard unit equivalent of x is the number z, where (4.24) The number z is sometimes called the "z-score" of x. The z-score is an item sampled from a normal population with mean 0 and standard deviation 1. This normal population is called the standard normal population. Example 4.39 Ball bearings manufactured for a certain application have diameters (in mm) that are normally distributed with mean 5 and standard deviation 0.08. A particular ball bearing has a diameter of 5.06 mm. Find the z-score. Solution The quantity 5.06 is an observation from a normal distribution with mean μ = 5 and standard deviation σ = 0.08. Example 4.40 Refer to Example 4.39. The diameter of a certain ball bearing has a z-score of −1.5. Find the diameter in the original units of mm. Solution The population mean is μ = 5 and the population standard deviation is σ = 0.08. We use Equation (4.24), substituting −1.5 for z and solving for x. We obtain Solving for x yields x = 4.88. The ball bearing is 4.88 mm in diameter. The proportion of a normal population that lies within a given interval is equal to the area under the normal probability density above that interval. This would suggest that we compute these proportions by integrating the normal probability density given in Equation (4.23). Interestingly enough, areas under this curve cannot be found by the method, taught in elementary calculus, of finding the antiderivative of the function and plugging in the limits of integration. This is because the antiderivative of this function is an infinite series and cannot be written down exactly. Instead, areas under this curve must be approximated numerically. Areas under the standard normal curve (mean 0, variance 1) have been extensively tabulated. A typical such table, called a standard normal table, or z table, is given as Table A.2 (in Appendix A). To find areas under a normal curve with a different mean and variance, we convert to standard units and use the z table. Table A.2 provides areas in the left-hand Page 244 tail of the curve for values of z. Other areas can be calculated by subtraction or by using the fact that the total area under the curve is equal to 1. We now present several examples to illustrate the use of the z table. Example 4.41 Find the area under the normal curve to the left of z = 0.47. Solution From the z table, the area is 0.6808. See Figure 4.5. FIGURE 4.5 Solution to Example 4.41. Example 4.42 Find the area under the normal curve to the right of z = 1.38. Solution From the z table, the area to the left of z = 1.38 is 0.9162. Therefore the area to the right is 1 − 0.9162 = 0.0838. See Figure 4.6. FIGURE 4.6 Solution to Example 4.42. Example 4.43 Find the area under the normal curve between z = 0.71 and z = 1.28. Solution From the z table, the area to the left of z = 1.28 is 0.8997. The area to the left of z = 0.71 is 0.7611. The area between z = 0.71 and z = 1.28 is therefore 0.8997 − 0.7611 = 0.1386. See Figure 4.7. Page 245 FIGURE 4.7 Solution to Example 4.43. Example 4.44 What z-score corresponds to the 75th percentile of a normal curve? The 25th percentile? The median? Solution To answer this question, we use the z table in reverse. We need to find the z-score for which 75% of the area of the curve is to the left. From the body of the table, the closest area to 75% is 0.7486, corresponding to a z-score of 0.67. Therefore the 75th percentile is approximately 0.67. By the symmetry of the curve, the 25th percentile is z = −0.67 (this can also be looked up in the table directly). See Figure 4.8. The median is z = 0. FIGURE 4.8 Solution to Example 4.44. Example 4.45 Lifetimes of batteries in a certain application are normally distributed with mean 50 hours and standard deviation 5 hours. Find the probability that a randomly chosen battery lasts between 42 and 52 hours. Solution Let X represent the lifetime of a randomly chosen battery. Then X ∼ N(50, 52). Figure 4.9 (page 246) presents the probability density function of the N(50, 52) population. The shaded area represents P(42 < X < 52), the probability that a randomly chosen battery has a lifetime between 42 and 52 hours. To compute the area, we will use the z table. First we need to convert the quantities 42 and 52 to standard units. We have From the z table, the area to the left of z = −1.60 is 0.0548, and the area to the left of z = 0.40 is 0.6554. The probability that a battery has a lifetime between 42 and 52 hours is 0.6554 − 0.0548 = 0.6006. Page 246 FIGURE 4.9 Solution to Example 4.45. Example 4.46 Refer to Example 4.45. Find the 40th percentile of battery lifetimes. Solution From the z table, the closest area to 0.4000 is 0.4013, corresponding to a z-score of −0.25. The population of lifetimes has mean 50 and standard deviation 5. The 40th percentile is the point 0.25 standard deviations below the mean. We find this value by converting the z-score to a raw score, using Equation (4.24): Solving for x yields x = 48.75. The 40th percentile of battery lifetimes is 48.75 hours. See Figure 4.10. FIGURE 4.10 Solution to Example 4.46. Example 4.47 A process manufactures ball bearings whose diameters are normally distributed with mean 2.505 cm and standard deviation 0.008 cm. Specifications call for the diameter to be in the interval 2.5 ± 0.01 cm. What proportion of the ball bearings will meet the specification? Solution Let X represent the diameter of a randomly chosen ball bearing. Then X ∼ N(2.505, 0.0082). Figure 4.11 presents the probability density function of the N(2.505, 0.0082) population. The shaded area represents P(2.49 < X < 2.51), which is the proportion of ball bearings that meet the specification. Page 247 We compute the z-scores of 2.49 and 2.51: The area to the left of z = −1.88 is 0.0301. The area to the left of z = 0.63 is 0.7357. The area between z = 0.63 and z = −1.88 is 0.7357 − 0.0301 = 0.7056. Approximately 70.56% of the diameters will meet the specification. FIGURE 4.11 Solution to Example 4.47. Example 4.48 Refer to Example 4.47. The process can be recalibrated so that the mean will equal 2.5 cm, the center of the specification interval. The standard deviation of the process remains 0.008 cm. What proportion of the diameters will meet the specifications? Solution The method of solution is the same as in Example 4.47. The mean is now 2.500 rather than 2.505. The calculations are as follows: The area to the left of z = −1.25 is 0.1056. The area to the left of z = 1.25 is 0.8944. The area between z = 1.25 and z = −1.25 is 0.8944 − 0.1056 = 0.7888. See Figure 4.12. Recalibrating will increase the proportion of diameters that meet the specification to 78.88%. FIGURE 4.12 Solution to Example 4.48. Example Page 248 4.49 Refer to Examples 4.47 and 4.48. Assume that the process has been recalibrated so that the mean is now 2.5 cm. To what value must the standard deviation be lowered so that 95% of the diameters will meet the specification? Solution The specification interval is 2.49–2.51 cm. We must find a value for σ so that this interval spans the middle 95% of the population of ball bearing diameters. See Figure 4.13. The zscore that has 2.5% of the area to its left is z = −1.96. The z-score that has 2.5% of the area to its right is z = 1.96 (this follows from the symmetry of the curve). It follows that the lower specification limit, 2.49, has a z-score of −1.96, while the upper limit of 2.51 has a z-score of 1.96. Either of these facts may be used to find σ. From Equation (4.24), Solving for σ yields σ = 0.0051 cm. FIGURE 4.13 Solution to Example 4.49. If σ = 0.0051, then approximately 95% of the diameters will fall between 2.49 and 2.51. Estimating the Parameters of a Normal Distribution The parameters μ and σ2 of a normal distribution represent its mean and variance, respectively. Therefore, if X1, …, Xn is a random sample from a N(μ, σ2) distribution, μ is estimated with the sample mean and σ2 is estimated with the sample variance s2. As with any sample mean, the uncertainty in is , which we replace with if σ is unknown. In addition, is unbiased for μ. Linear Functions of Normal Random Variables If a normal random variable is multiplied by a nonzero constant, or has a constant added to it, the resulting random variable is also normal, with a mean and variance that are determined by the original mean and variance and the constants. Specifically, Page 249 Summary Let X ∼ N(μ, σ2), and let a ≠ 0 and b be constants. Then (4.25) Example 4.50 A chemist measures the temperature of a solution in °C. The measurement is denoted C, and is normally distributed with mean 40°C and standard deviation 1°C. The measurement is converted to °F by the equation F = 1.8C + 32. What is the distribution of F? Solution Since C is normally distributed, so is F. Now μC = 40, so μF = 1.8(40) + 32 = 104, and , so . Therefore F ∼ N(104, 3.24). Linear Combinations of Independent Normal Random Variables One of the remarkable features of the normal distribution is that linear combinations of independent normal random variables are themselves normal random variables. To be specific, suppose that X1, X2, …, Xn are independent normal random variables. Note that the means and variances of these random variables can differ from one another. Let c1, c2, …, cn be constants. Then the linear combination c1X1 + c2X2 + ⋯ + cnXn is a normally distributed random variable. The mean and variance of the linear combination are c1μ1 + c2μ2 + ⋯ + cnμn and , respectively (see Equations 2.51 and 2.55 in Section 2.5). Summary Let X1, X2, …, Xn be independent and normally distributed with means μ1, μ2, …, μn and variances . Let c1, c2, …, cn be constants, and c1X1 + c2X2 + ⋯ + cnXn be a linear combination. Then (4.26) Example 4.51 In the article "Advances in Oxygen Equivalent Equations for Predicting the Properties of Titanium Welds" (D. Harwig, W. Ittiwattana, and H. Castner, The Welding Journal, 2001:126s–136s), the authors propose an oxygen equivalence equation to predict the strength, ductility, and hardness of welds made from nearly pure titanium. The equation is E = 2C + 3.5N + O, where E is the oxygen equivalence, and C, N, and O are the proportions Page 250 by weight, in parts per million, of carbon, nitrogen, and oxygen, respectively (a constant term involving iron content has been omitted). Assume that for a particular grade of commercially pure titanium, the quantities C, N, and O are approximately independent and normally distributed with means μC = 150, μN = 200, μO = 1500, and standard deviations σC = 30, σN = 60, σO = 100. Find the distribution of E. Find P(E > 3000). Solution Since E is a linear combination of independent normal random variables, its distribution is normal. We must now find the mean and variance of E. Using Equation (4.26), we compute We conclude that E ∼ N(2500, 57,700). To find P(E > 3000), we compute the z-score: . The area to the right of z = 2.08 under the standard normal curve is 0.0188. So P(E > 3000) = 0.0188. If X1, …, Xn is a random sample from any population with mean μ and variance σ2, then the sample mean has mean and variance . If the population is normal, then is normal as well, because it is a linear combination of X1, …, Xn with coefficients c1 = ⋯ = cn = 1/n. Summary Let X1, …, Xn be independent and normally distributed with mean μ and variance σ2. Then (4.27) Other important linear combinations are the sum and difference of two random variables. If X and Y are independent normal random variables, the sum X + Y and the difference X − Y are linear combinations. The distributions of X + Y and X − Y can be determined by using Equation (4.26) with c1 = 1, c2 = 1 for X + Y and c1 = 1, c2 = −1 for X − Y. Page 251 Summary Let X and Y be independent, with and . Then (4.28) (4.29) How Can I Tell Whether My Data Come from a Normal Population? In practice, we often have a sample from some population, and we must use the sample to decide whether the population distribution is approximately normal. If the sample is reasonably large, the sample histogram may give a good indication. Large samples from normal populations have histograms that look something like the normal density function—peaked in the center, and decreasing more or less symmetrically on either side. Probability plots, which will be discussed in Section 4.10, provide another good way of determining whether a reasonably large sample comes from a population that is approximately normal. For small samples, it can be difficult to tell whether the normal distribution is appropriate. One important fact is this: Samples from normal populations rarely contain outliers. Therefore the normal distribution should generally not be used for data sets that contain outliers. This is especially true when the sample size is small. Unfortunately, for small data sets that do not contain outliers, it is difficult to determine whether the population is approximately normal. In general, some knowledge of the process that generated the data is needed. Derivation of the Mean and Variance for a Normal Random Variable Let X ∼ N(μ, σ2). We show that μX = μ and . Using the definition of the population mean of a continuous random variable (Equation 2.37 in Section 2.4), Make the substitution z = (x − μ)/σ. Then x = σz + μ, and dx = σdz. We obtain Direct computation shows that Page 252 Also, since it is the integral of the N(0, 1) probability density function over all its possible values. Therefore To show that , we use Equation (2.38) (in Section 2.4): Make the substitution z = (x − μ)/σ. Recall that μX = μ. Then Integrating by parts twice shows that Therefore . Exercises for Section 4.5 1. 2. 3. 4. 5. Find the area under the normal curve a. To the right of z = −0.85. b. Between z = 0.40 and z = 1.30. c. Between z = −0.30 and z = 0.90. d. Outside z = −1.50 to z = −0.45. Find the area under the normal curve a. To the left of z = 0.56. b. Between z = −2.93 and z = −2.06. c. Between z = −1.08 and z = 0.70. d. Outside z = 0.96 to z = 1.62. Let Z ∼ N(0, 1). Find a constant c for which a. P(Z ≤ c) = 0.1587 b. P(c ≤ Z ≤ 0) = 0.4772 c. P(−c ≤ Z ≤ c) = 0.8664 d. P(0 ≤ Z ≤ c) = 0.2967 e. P(|Z| ≤ c) = 0.1470 If X ∼ N(2, 9), compute a. P(X ≥ 2) b. P(1 ≤ X < 7) c. P(−2.5 ≤ X ≤ −1) d. P(−3 ≤ X − 2 < 3) A process manufactures ball bearings with diameters that are normally distributed with mean 25.1 mm and standard deviation 0.08 mm. a. What proportion of the diameters are less than 25.0 mm? Page 253 b. c. 6. Find the 10th percentile of the diameters. A particular ball bearing has a diameter of 25.2 mm. What percentile is its diameter on? d. To meet a certain specification, a ball bearing must have a diameter between 25.0 and 25.3 millimeters. What proportion of the ball bearings meet the specification? Depths of pits on a corroded steel surface are normally distributed with mean 822 μm and standard deviation 29 μm. a. Find the 10th percentile of pit depths. b. A certain pit is 780 μm deep. What percentile is it on? c. What proportion of pits have depths between 800 and 830 μm? 7. In a recent study, the Centers for Disease Control reported that diastolic blood pressures (in mmHg) of adult women in the U.S. are approximately normally distributed with mean 80.5 and standard deviation 9.9. a. What proportion of women have blood pressures lower than 70? b. What is the 80th percentile of blood pressures? c. A woman has a blood pressure of 84. What percentile is her blood pressure on? d. A diastolic blood pressure greater than 90 is classified as hypertension (high blood pressure). What proportion of women have hypertension? 8. Weights of female cats of a certain breed are normally distributed with mean 4.1 kg and standard deviation 0.6 kg. a. What proportion of female cats have weights between 3.7 and 4.4 kg? b. A certain female cat has a weight that is 0.5 standard deviations above the mean. What proportion of female cats are heavier than this one? c. How heavy is a female cat whose weight is on the 80th percentile? d. A female cat is chosen at random. What is the probability that she weighs more than 4.5 kg? e. Six female cats are chosen at random. What is the probability that exactly one of them weighs more than 4.5 kg? 9. The lifetime of a lightbulb in a certain application is normally distributed with mean μ = 1400 hours and standard deviation σ = 200 hours. a. What is the probability that a lightbulb will last more than 1800 hours? b. Find the 10th percentile of the lifetimes. c. A particular lightbulb lasts 1645 hours. What percentile is its lifetime on? d. What is the probability that the lifetime of a light-bulb is between 1350 and 1550 hours? e. Eight lightbulbs are chosen at random. What is the probability that exactly two of them have lifetimes between 1350 and 1550 hours? 10. In a certain university, math SAT scores for the entering freshman class averaged 650 and had a standard deviation of 100. The maximum possible score is 800. Is it possible that the scores of these freshmen are normally distributed? Explain. 11. Penicillin is produced by the Penicillium fungus, which is grown in a broth whose sugar content must be carefully controlled. The optimum sugar concentration is 4.9 mg/mL. If the concentration exceeds 6.0 mg/mL, the fungus dies and the process must be shut down for the day. a. If sugar concentration in batches of broth is normally distributed with mean 4.9 mg/mL and standard deviation 0.6 mg/mL, on what proportion of days will the process shut down? b. The supplier offers to sell broth with a sugar content that is normally distributed with mean 5.2 mg/mL and standard deviation 0.4 mg/mL. Will this broth result in fewer days of production lost? Explain. 12. Speeds of automobiles on a certain stretch of freeway at 11:00 PM are normally distributed with mean 65 mph. Twenty percent of the cars are traveling at speeds between 55 and 65 mph. What percentage of the cars are going faster than 75 mph? 13. A cylindrical hole is drilled in a block, and a cylindrical piston is placed in the hole. The clearance is equal to one-half the difference between the diameters of the hole and the piston. The diameter of the hole is normally distributed with mean 15 cm and standard deviation 0.025 cm, and the diameter of the piston is normally distributed with mean 14.88 cm and standard deviation 0.015 cm. The diameters of hole and piston are independent. a. Find the mean clearance. b. Find the standard deviation of the clearance. c. What is the probability that the clearance is less than 0.05 cm? d. Find the 25th percentile of the clearance. e. Specifications call for the clearance to be between 0.05 and 0.09 cm. What is the probability that the clearance meets the specification? f. It is possible to adjust the mean hole diameter. To what value should it be adjusted so as to maximize the probability that the clearance will be between 0.05 and 0.09 cm? Shafts manufactured for use in optical storage devices have diameters that are normally distributed with mean μ = 0.652 cm and standard deviation σ = 0.003 cm. The specification for the shaft diameter is 0.650 ± 0.005 cm. a. What proportion of the shafts manufactured by this process meet the specifications? b. The process mean can be adjusted through calibration. If the mean is set to 0.650 cm, what proportion of the shafts will meet specifications? c. If the mean is set to 0.650 cm, what must the standard deviation be so that 99% of the shafts will meet specifications? The fill volume of cans filled by a certain machine is normally distributed with mean 12.05 oz and standard deviation 0.03 oz. a. What proportion of cans contain less than 12 oz? b. The process mean can be adjusted by calibration. To what value should the mean be set so that 99% of the cans will contain 12 oz or more? c. If the process mean remains at 12.05 oz, what must the standard deviation be so that 99% of the cans will contain 12 oz or more? The amount of paint required to paint a surface with an area of 50 m2 is normally distributed with mean 6 L and standard deviation 0.3 L. a. If 6.2 L of paint are available, what is the probability that the entire surface can be painted? b. How much paint is needed so that the probability is 0.9 that the entire surface can be painted? c. What must the standard deviation be so that the probability is 0.9 that 6.2 L of paint will be sufficient to paint the entire surface? A fiber-spinning process currently produces a fiber whose strength is normally distributed with a mean of 75 N/m2. The minimum acceptable strength is 65 N/m2. a. Ten percent of the fiber produced by the current method fails to meet the minimum specification. What is the standard deviation of fiber strengths in the current process? b. If the mean remains at 75 N/m2, what must the standard deviation be so that only 1% of the fiber will fail to meet the specification? c. If the standard deviation is 5 N/m2, to what value must the mean be set so that only 1% of the fiber will fail to meet the specification? The area covered by 1 L of a certain stain is normally distributed with mean 10 m2 and standard deviation 0.2 m2. a. What is the probability that 1 L of stain will be enough to cover 10.3 m2? b. What is the probability that 2 L of stain will be enough to cover 19.9 m2? Scores on an exam were normally distributed. Ten percent of the scores were below 64 and 80% were below 81. Find the mean and standard deviation of the scores. Specifications for an aircraft bolt require that the ultimate tensile strength be at least 18 kN. It is known that 10% of the bolts have strengths less than 18.3 kN and that 5% of the bolts have strengths greater than 19.76 kN. It is also known that the strengths of these bolts are normally distributed. a. Find the mean and standard deviation of the strengths. b. What proportion of the bolts meet the strength specification? Let X ∼ N(μ, σ2), and let Z = (X − μ)/σ. Use Equation (4.25) to show that Z ∼ N(0, 1). The quality-assurance program for a certain adhesive formulation process involves measuring how well the adhesive sticks a piece of plastic to a glass surface. When the process is functioning correctly, the adhesive strength X is normally distributed with a mean of 200 N and a standard deviation of 10 N. Each hour, you make one measurement of the adhesive strength. You are supposed to inform your supervisor if your Page 255 measurement indicates that the process has strayed from its target distribution. a. Find P(X ≤ 160), under the assumption that the process is functioning correctly. b. Based on your answer to part (a), if the process is functioning correctly, would a strength of 160 N be unusually small? Explain. c. If you observed an adhesive strength of 160 N, would this be convincing evidence that the process was no longer functioning correctly? Explain. d. Find P(X ≥ 203), under the assumption that the process is functioning correctly. e. Based on your answer to part (d), if the process is functioning correctly, would a strength of 203 N be unusually large? Explain. f. If you observed an adhesive strength of 203 N, would this be convincing evidence that the process was no longer functioning correctly? Explain. g. Find P(X ≤ 195), under the assumption that the process is functioning correctly. h. Based on your answer to part (g), if the process is functioning correctly, would a strength of 195 N be unusually small? Explain. i. If you observed an adhesive strength of 195 N, would this be convincing evidence that the process was no longer functioning correctly? Explain. Two resistors, with resistances R1 and R2, are connected in series. R1 is normally distributed with mean 100 Ω and standard deviation 5 Ω, and R2 is normally distributed with mean 120 Ω and standard deviation 10 Ω. a. What is the probability that R2 > R1? b. What is the probability that R2 exceeds R1 by more than 30 Ω? 24. The molarity of a solute in solution is defined to be the number of moles of solute per liter of solution (1 mole = 6.02 × 1023 molecules). If X is the molarity of a solution of sodium chloride (NaCl), and Y is the molarity of a solution of sodium carbonate (Na2CO3), the molarity of sodium ion (Na+)inasolution made of equal parts NaCl and Na2CO3 is given by M = 0.5X + Y. Assume X and Y are independent and normally distributed, and that X has mean 0.450 and standard deviation 0.050, and Y has mean 0.250 and standard deviation 0.025. a. What is the distribution of M? b. Find P(M > 0.5). 25. A binary message m,where m is equal either to 0 or to 1, is sent over an information channel. Because of noise in the channel, the message received is X, where X = m + E, and E is a random variable representing the channel noise. Assume that if X ≤ 0.5 then the receiver concludes that m = 0 and that if X > 0.5 then the receiver concludes that m = 1. Assume that E ∼ N(0, 0.25). a. If the true message is m = 0, what is the probability of an error, that is, what is the probability that the receiver concludes that m = 1? b. Let σ2 denote the variance of E. What must be the value of σ2 so that the probability of error when m = 0 is 0.01? 26. Refer to Exercise 25. Assume that if m = 0, the value s = −1.5 is sent, and if m = 1, the value s = 1.5 is sent. The value received is X, where X = s + E, and E ∼ N(0, 0.64). If X ≤ 0.5, then the receiver concludes that m = 0, and if X > 0.5, then the receiver concludes that m = 1. a. If the true message is m = 0, what is the probability of an error, that is, what is the probability that the receiver concludes that m = 1? b. If the true message is m = 1, what is the probability of an error, that is, what is the probability that the receiver concludes that m = 0? c. A string consisting of 60 1s and 40 0s will be sent. A bit is chosen at random from this string. What is the probability that it will be received correctly? d. Refer to part (c). A bit is chosen at random from the received string. Given that this bit is 1, what is the probability that the bit sent was 0? e. Refer to part (c). A bit is chosen at random from the received string. Given that this bit is 0, what is the probability that the bit sent was 1? 27. A company receives a large shipment of bolts. The bolts will be used in an application that requires a torque of 100 J. Before the shipment is accepted, a quality engineer will sample 12 bolts and measure the torque needed to break each of them. The shipment will be accepted if the engineer concludes that fewer than 1% of the bolts in the shipment have a breaking torque of less than 100 J. Page 256 a. If the 12 values are 107, 109, 111, 113, 113, 114, 114, 115, 117, 119, 122, 124, compute the sample mean and sample standard deviation. b. Assume the 12 values are sampled from a normal population, and assume the the sample mean and standard deviation calculated in part (a) are actually the population mean and standard deviation. Compute the proportion of bolts whose breaking torque is less than 100 J. Will the shipment be accepted? c. What if the 12 values had been 108, 110, 112, 114, 114, 115, 115, 116, 118, 120, 123, 140? Use the method outlined in parts (a) and (b) to determine whether the shipment would have been accepted. d. Compare the sets of 12 values in parts (a) and (c). In which sample are the bolts stronger? e. Is the method valid for both samples? Why or why not? 28. Chebyshev's inequality (Section 2.4) states that for any random variable X with mean μ and variance σ2, and for any positive number k, P(|X − μ| ≥ kσ) ≤ 1/k2. Let X ∼ N(μ, σ2). Compute P(|X − μ| ≥ kσ) for the values k = 1, 2, and 3. Are the actual probabilities close to the Chebyshev bound of 1/k2, or are they much smaller? 4.6 The Lognormal Distribution For data that are highly skewed or that contain outliers, the normal distribution is generally not appropriate. The lognormal distribution, which is related to the normal distribution, is often a good choice for these data sets. The lognormal distribution is derived from the normal distribution as follows: If X is a normal random variable with mean μ and variance σ2, then the random variable Y = eX is said to have the lognormal distribution with parameters μ and σ2. Note that if Y has the lognormal distribution with parameters μ and σ2, then the random variable X = ln Y has the normal distribution with mean μ and variance σ2. Summary ■ ■ If X ∼ N(μ, σ2), then the random variable Y = eX has the lognormal distribution with parameters μ and σ2. If Y has the lognormal distribution with parameters μ and σ2, then the random variable X = ln Y has the N(μ, σ2) distribution. The probability density function of a lognormal random variable with parameters μ and σ is (4.30) Figure 4.14 presents a graph of the lognormal density function with parameters μ = 0 and σ = 1. Note that the density function is highly skewed. This is the reason that the lognormal distribution is often used to model processes that tend to produce occasional large values, or outliers. Page 257 FIGURE 4.14 The probability density function of the lognormal distribution with parameters μ = 0 and σ = 1. It can be shown by advanced methods that if Y is a lognormal random variable with parameters μ and σ2, then the mean E(Y) and variance V(Y) are given by (4.31) Note that if Y has the lognormal distribution, the parameters μ and σ2 do not refer to the mean and variance of Y. They refer instead to the mean and variance of the normal random variable ln Y. In Equation (4.31) we used the notation E(Y) instead of μY, and V(Y) instead of , in order to avoid confusion with μ and σ. Example 4.52 Lifetimes of a certain component are lognormally distributed with parameters μ = 1 day and σ = 0.5 days. Find the mean lifetime of these components. Find the standard deviation of the lifetimes. Solution Let Y represent the lifetime of a randomly chosen component. The mean of Y is found by Equation (4.31) to be days. The variance is . The standard deviation is therefore days. To compute probabilities involving lognormal random variables, take logs and use the z table (Table A.2). Examples 4.53 and 4.54 illustrate the method. Example Page 258 4.53 Refer to Example 4.52. Find the probability that a component lasts longer than four days. Solution Let Y represent the lifetime of a randomly chosen component. We need to find P(Y > 4). We cannot use the z table for Y, because Y is not sampled from a normal population. However, ln Y is sampled from a normal population; specifically, ln Y ∼ N(1, 0.52). We express P(Y > 4) as a probability involving ln Y: The z-score of 1.386 is From the z table, we find that P(ln Y > 1.386) = 0.2206. (See Figure 4.15.) We conclude that approximately 22% of the components will last longer than four days. FIGURE 4.15 Solution to Example 4.53. Example 4.54 Refer to Example 4.52. Find the median lifetime. Find the 80th percentile of the lifetimes. Solution Let Y represent the lifetime of a randomly chosen component. Let m denote the median lifetime. Then P(Y ≤ m) = 0.5. Taking logs, we have P(ln Y ≤ ln m) = 0.5. This means that ln m is the median of ln Y. Now ln Y ∼ N(1, 0.52). Therefore ln m = 1, so m = e1 = 2.718. To find the 80th percentile, p80,set P(Y ≤ p80) = 0.80. Then P(ln Y ≤ ln p80) = 0.80. This means that ln p80 is the 80th percentile of ln Y. Now ln Y ∼ N(1, 0.52). From the z table, the z-score of the 80th percentile is 0.84. Therefore ln p80 = 1 + (0.84)(0.5) = 1.42, so p80 = e1.42 = 4.14. Page 259 Estimating the Parameters of a Lognormal Distribution If Y is a random variable whose distribution is lognormal with parameters μ and σ2, then μ and σ2 are the mean and variance, respectively, of ln Y. Therefore if Y1, …, Yn is a random sample from a lognormal population, we first transform to the log scale, defining X1 = ln Y1, …, Xn = ln Yn. Now X1, …, Xn is a random sample from N(μ, σ2). We estimate μ with and σ2 with the sample variance . As with any sample mean, the uncertainty in is , and if σ is unknown, we estimate it with the sample standard deviation sX. Example 4.55 The diameters (in mm) of seeds of a certain plant are lognormally distributed. A random sample of five seeds had diameters 1.52, 2.22, 2.64, 2.00, and 1.69. Estimate the parameters μ and σ. Solution To estimate μ and σ, we take logs of the five sample values, to obtain 0.419, 0.798, 0.971, 0.693, and 0.525. The sample mean is 0.681, and the sample standard deviation is 0.218. We therefore estimate . How Can I Tell Whether My Data Come from a Lognormal Population? As stated previously, samples from normal populations rarely contain outliers. In contrast, samples from lognormal populations often contain outliers in the right-hand tail. That is, the samples often contain a few values that are much larger than the rest of the data. This of course reflects the long right-hand tail in the lognormal density function (Figure 4.14). For samples with outliers on the right, we transform the data, by taking the natural logarithm (or any logarithm) of each value. We then try to determine whether these logs come from a normal population, by plotting them on a histogram, or on a probability plot. Probability plots will be discussed in Section 4.10. Note that the lognormal density has only one long tail, on the right. For this reason, samples from lognormal populations will have outliers on the right, but not on the left. The lognormal distribution should therefore not be used for samples with a few unusually small observations. In addition, lognormal populations contain only positive values, so the lognormal distribution should not be used for samples that contain zeros or negative values. Finally, it is important to note that the log transformation will not always produce a sample that is close to normal. One has to plot a histogram or probability plot (see Section 4.10) to check. Figure 4.16 (page 260) presents two histograms. The first shows the monthly production for 255 gas wells, in units of thousand cubic feet. This histogram clearly has a long right-hand tail, so we conclude that the data do not come from a normal population. The second shows the natural logs of the monthly productions. This histogram is closer to the normal curve, although some departure from normality can be detected. Page 260 FIGURE 4.16 (a) A histogram showing monthly production for 255 gas wells. There is a long right-hand tail. (b) A histogram showing the natural logs of the monthly productions. The distribution of the logged data is much closer to normal. Exercises for Section 4.6 1. The lifetime (in days) of a certain electronic component that operates in a high-temperature environment is lognormally distributed with μ = 1.2 and σ = 0.4. a. Find the mean lifetime. b. Find the probability that a component lasts between three and six days. c. Find the median lifetime. d. Find the 90th percentile of the lifetimes. 2. 3. 4. 5. The article "Assessment of Dermopharmacokinetic Approach in the Bioequivalence Determination of Topical Tretinoin Gel Products" (L. Pershing, J. Nelson, et al., J Am Acad Dermatol 2003:740–751) reports that the amount of a certain antifungal ointment that is absorbed into the skin can be modeled with a lognormal distribution. Assume that the amount (in ng/cm2) of active ingredient in the skin two hours after application is lognormally distributed with μ = 2.2 and σ = 2.1. a. Find the mean amount absorbed. b. Find the median amount absorbed. c. Find the probability that the amount absorbed is more than 100 ng/cm2. d. Find the probability that the amount absorbed is less than 50 ng/cm2. e. Find the 80th percentile of the amount absorbed. f. Find the standard deviation of the amount absorbed. The body mass index (BMI) of a person is defined to be the person's body mass divided by the square of the person's height. The article "Influences of Parameter Uncertainties within the ICRP 66 Respiratory Tract Model: Particle Deposition" (W. Bolch, E. Farfan, et al., Health Physics, 2001:378–394) states that body mass index (in kg/m2) in men aged 25–34 is lognormally distributed with parameters μ = 3.215 and σ = 0.157. a. Find the mean BMI for men aged 25–34. b. Find the standard deviation of BMI for men aged 25–34. c. Find the median BMI for men aged 25–34. d. What proportion of men aged 25–34 have a BMI less than 22? e. Find the 75th percentile of BMI for men aged 25–34. The article "Stochastic Estimates of Exposure and Cancer Risk from Carbon Tetrachloride Released to the Air from the Rocky Flats Plant" (A. Rood, P. McGavran, et al., Page 261 Risk Analysis, 2001:675–695) models the increase in the risk of cancer due to exposure to carbon tetrachloride as lognormal with μ = −15.65 and σ = 0.79. a. Find the mean risk. b. Find the median risk. c. Find the standard deviation of the risk. d. Find the 5th percentile. e. Find the 95th percentile. The period T of a simple pendulum is given by T = where L is the length of the pendulum and g is the acceleration due to gravity. Assume that g = 9.80m/s2 exactly, and that L, in meters, is lognormal with parameters μL = 0.7 and . a. Show that T is lognormally distributed, and compute the parameters μT and ln T = 0.5 ln L + ln 2π − 0.5 ln g.) b. Find P(T > 3). c. Find P(1.5 < T < 3). . (Hint: 6. 7. 8. 9. d. Find the mean of T. e. Find the median of T. f. Find the standard deviation of T. g. Find the 15th percentile of T. h. Find the 85th percentile of T. The volume of a cylinder is given by V = πr2h, where r is the radius of the cylinder and h is the height. Assume the radius, in cm, is lognormal with parameters μr = 1.6 and , the height, in cm, is log-normal with parameters μh = 1.9 and , and that the radius and height are independent. a. Show that V is lognormally distributed, and compute the parameters μV and ln V = ln π+2 ln r + ln h.) b. Find P(V < 600). c. Find P(500 < V < 800). d. Find the mean of V. e. Find the median of V. f. Find the standard deviation of V. g. Find the 5th percentile of V. . (Hint: h. Find the 95th percentile of V. Refer to Exercise 5. Suppose 10 pendulums are constructed. Find the probability that 4 or more have periods greater than 3 seconds. Refer to Exercise 6. Suppose 8 cylinders are constructed. Find the probability that fewer than 5 of them have volumes between 500 and 800 cm3. The article "Withdrawal Strength of Threaded Nails" (D. Rammer, S. Winistorfer, and D. Bender, Journal of Structural Engineering 2001:442–449) describes an experiment comparing the ultimate withdrawal strengths (in N/mm) for several types of nails. For an annularly threaded nail with shank diameter 3.76 mm driven into spruce-pine-fir lumber, the ultimate withdrawal strength was modeled as lognormal with μ = 3.82 and σ = 0.219. For a helically threaded nail under the same conditions, the strength was modeled as lognormal with μ = 3.47 and σ = 0.272. a. What is the mean withdrawal strength for annularly threaded nails? b. What is the mean withdrawal strength for helically threaded nails? c. For which type of nail is it more probable that the withdrawal strength will be greater than 50 N/mm? d. What is the probability that a helically threaded nail will have a greater withdrawal strength than the median for annularly threaded nails? e. An experiment is performed in which withdrawal strengths are measured for several nails of both types. One nail is recorded as having a withdrawal strength of 20 N/mm, but its type is not given. Do you think it was an annularly threaded nail or a helically threaded nail? Why? How sure are you? 10. Choose the best answer, and explain. If X is a random variable with a lognormal distribution, then _______ i. the mean of X is always greater than the median. ii. the mean of X is always less than the median. iii. the mean may be greater than, less than, or equal to the median, depending on the value of σ. 11. The prices of stocks or other financial instruments are often modeled with a lognormal distribution. An investor is considering purchasing stock in one of two companies, A or B. The price of a share of stock today is $1 for both companies. For company A, the value of the stock one year from now is modeled as lognormal with parameters μ = 0.05 and σ = 0.1. For company B, the value of the stock one year from now is modeled as Page 262 lognormal with parameters μ = 0.02 and σ = 0.2. a. Find the mean of the price of one share of company A one year from now. b. Find the probability that the price of one share of company A one year from now will be greater than $1.20. c. Find the mean of the price of one share of company B one year from now. d. Find the probability that the price of one share of company B one year from now will be greater than $1.20. 12. A manufacturer claims that the tensile strength of a certain composite (in MPa) has the lognormal distribution with μ = 5 and σ = 0.5. Let X be the strength of a randomly sampled specimen of this composite. a. If the claim is true, what is P(X < 20)? b. Based on the answer to part (a), if the claim is true, would a strength of 20 MPa be unusually small? c. If you observed a tensile strength of 20 MPa, would this be convincing evidence that the claim is false? Explain. d. If the claim is true, what is P(X < 130)? e. Based on the answer to part (d), if the claim is true, would a strength of 130 MPa be unusually small? f. If you observed a tensile strength of 130 MPa, would this be convincing evidence that the claim is false? Explain. 13. Let X1, …, Xn be independent lognormal random variables and let a1, …, an be constants. Show that the product is lognormal. (Hint: ln P = a1 ln X1 + ⋯ + an ln Xn.) 4.7 The Exponential Distribution The exponential distribution is a continuous distribution that is sometimes used to model the time that elapses before an event occurs. Such a time is often called a waiting time. The exponential distribution is sometimes used to model the lifetime of a component. In addition, there is a close connection between the exponential distribution and the Poisson distribution. The probability density function of the exponential distribution involves a parameter, which is a positive constant λ whose value determines the density function's location and shape. Definition The probability density function of the exponential distribution with parameter λ > 0 is (4.32) Figure 4.17 presents the probability density function of the exponential distribution for various values of λ. If X is a random variable whose distribution is exponential with parameter λ, we write X ∼ Exp(λ). FIGURE 4.17 Plots of the exponential probability density function for various values of X. The cumulative distribution function of the exponential distribution is easy to compute. For x ≤ 0, F(x) = P(X ≤ x) = 0. For x > 0, the cumulative distribution function is Page 263 Summary If X ∼ Exp(λ), the cumulative distribution function of X is (4.33) The mean and variance of an exponential random variable can be computed by using integration by parts. Derivations are provided at the end of the section. If X ∼ Exp(λ), then (4.34) (4.35) Example If X ∼ Exp(2), find μX, 4.56 , and P(X ≤ 1). Solution We compute μX and 0.5, from Equations (4.34) and (4.35), substituting λ = 2. We obtain μX = . Using Equation (4.33), we find that Page 264 Example 4.57 Refer to Example 4.56. Find the median of X. Find the 30th percentile of X. Solution Let m denote the median of X. Then P(X ≤ m) = 0.5. Using Equation (4.33), we find that 1 − e −2m = 0.5. Solving for m, we obtain m = 0.3466. Let p30 denote the 30th percentile. Then P(X ≤ p30) = 0.30. Using Equation (4.33), we find that . Solving for p30, we obtain p30 = 0.1783. The Exponential Distribution and the Poisson Process We mentioned that the exponential distribution is sometimes used to model the waiting time to an event. It turns out that the exponential distribution is the correct model for waiting times whenever the events follow a Poisson process. Recall from Section 4.3 that events follow a Poisson process with rate parameter λ when the numbers of events in disjoint intervals are independent, and the number X of events that occur in any time interval of length t has a Poisson distribution with mean λt, in other words, when X ∼ Poisson(λt). The connection between the exponential distribution and the Poisson process is as follows: If events follow a Poisson process with rate parameter λ, and if T represents the waiting time from any starting point until the next event, then T ∼ Exp(λ). A proof of this fact is given at the end of the section. Example 4.58 A radioactive mass emits particles according to a Poisson process at a mean rate of 15 particles per minute. At some point, a clock is started. What is the probability that more than 5 seconds will elapse before the next emission? What is the mean waiting time until the next particle is emitted? Solution We will measure time in seconds. Let T denote the time in seconds that elapses before the next particle is emitted. The mean rate of emissions is 0.25 per second, so the rate parameter is λ = 0.25, and T ∼ Exp(0.25). The probability that more than 5 seconds will elapse before the next emission is equal to The mean waiting time is seconds. Page 265 Memoryless Property The exponential distribution has a property known as the memoryless property, which we illustrate with Examples 4.59 and 4.60. Example 4.59 The lifetime of a particular integrated circuit has an exponential distribution with mean 2 years. Find the probability that the circuit lasts longer than three years. Solution Let T represent the lifetime of the circuit. Since μT = 2, λ = 0.5. We need to find P(T > 3). Example 4.60 Refer to Example 4.59. Assume the circuit is now four years old and is still functioning. Find the probability that it functions for more than three additional years. Compare this probability with the probability that a new circuit functions for more than three years, which was calculated in Example 4.59. Solution We are given that the lifetime of the circuit will be more than four years, and we must compute the probability that the lifetime will be more than 4 + 3 = 7 years. The probability is given by If T > 7, then T > 4 as well. Therefore P(T > 7 and T > 4) = (T > 7). It follows that The probability that a 4-year-old circuit lasts 3 additional years is exactly the same as the probability that a new circuit lasts 3 years. Examples 4.59 and 4.60 illustrate the memoryless property. The probability that we must wait an additional t units, given that we have already waited s units, is the same as the Page 266 probability that we must wait t units from the start. The exponential distribution does not "remember" how long we have been waiting. In particular, if the lifetime of a component follows the exponential distribution, then the probability that a component that is s time units old will last an additional t time units is the same as the probability that a new component will last t time units. In other words, a component whose lifetime follows an exponential distribution does not show any effects of age or wear. The calculations in Examples 4.59 and 4.60 could be repeated for any values s and t in place of 4 and 3, and for any value of λ in place of 0.5. We now state the memoryless property in its general form: Memoryless Property If T ∼ Exp(λ), and t and s are positive numbers, then Example 4.61 The number of hits on a website follows a Poisson process with a rate of 3 per minute. What is the probability that more than a minute goes by without a hit? If 2 minutes have gone by without a hit, what is the probability that a hit will occur in the next minute? Solution Let T denote the waiting time in minutes until the next hit. Then T ∼ Exp(3). The probability that one minute elapses with no hits is P(T > 1) = e−3(1) = 0.0498. Because of the memoryless property, the probability that one additional minute elapses without a hit, given that two minutes have gone by without a hit, is also equal to 0.0498. The probability that a hit does occur in the next minute is therefore equal to 1 − 0.0498 = 0.9502. Using the Exponential Distribution to Estimate a Rate If X ∼ Exp(λ), then μX = 1/λ, so λ = 1/μX. It follows that if X1, …, Xn is a random sample from Exp(λ), it is reasonable to estimate λ with We will discuss the bias in unbiased for μ. However, . . As with any sample mean , so is , because 1/μ is not a linear function of μ. Therefore is biased for λ = 1/μ. Using advanced methods, it can be shown that , so the bias is approximately λ/n. Thus for a sufficiently large sample size n the bias is negligible, but it can be substantial when the sample size is small. The uncertainty in can be estimated with A derivation is given at the end of the section. Page 267 Summary If X1, …, Xn is a random sample from Exp(λ), then the parameter λ is estimated with (4.36) This estimator is biased. The bias is approximately equal to λ/n. The uncertainty in is estimated with (4.37) This uncertainty estimate is reasonably good when the sample size is more than 20. Correcting the Bias Since quantity , it follows that] has smaller bias for estimating λ than does . In other words, the . This is known as a bias- corrected estimate. Example 4.62 A random sample of size 5 is taken from an Exp(λ) distribution. The values are 7.71, 1.32, 7.46, 6.53, and 0.44. Find a bias-corrected estimate of λ. Solution The sample mean is . The sample size is n = 5. The bias-corrected estimate of λ is 5/[6(4.6920)] = 0.178. Derivation of the Mean and Variance of an Exponential Random Variable To derive Equation (4.34), we begin with Equation (2.37) (in Section 2.4): Substituting the exponential probability density function (4.32) for f (x), we obtain Integrating by parts, setting u = x and dv = λe−λx yields (4.38) Page 268 We evaluate the first quantity on the right-hand side of Equation (4.38): Therefore To derive Equation (4.35), we begin with Equation (2.39) (in Section 2.4): Substituting the exponential probability density function (4.32) for f(x), and 1/λ for μX, we obtain (4.39) We evaluate the integral , using integration by parts. Setting u = x2 and dv = λe−λx we obtain (4.40) We evaluate the first quantity on the right-hand side of Equation (4.40). Therefore Page 269 Substituting into (4.39) yields Derivation of the Relationship Between the Exponential Distribution and the Poisson Process Let T represent the waiting time until the next event in a Poisson process with rate parameter A. We show that T ∼ Exp(λ) by showing that the cumulative distribution function of T is F(t) = 1 − e−λt, which is the cumulative distribution function of Exp(λ). First, if t ≤ 0, then F(t) = P(T ≤ t) = 0. Now let t > 0. We begin by computing P(T > t). The key is to realize that T > t if and only if no events happen in the next t time units. Let X represent the number of events that happen in the next t time units. Now T > t if and only if X = 0, so P(T > t) = P(X = 0). Since X ∼ Poisson(λt), Therefore P(T > t) = e−λt. The cumulative distribution function of T is F(t) = 0 for t − 0, and for t > 0 Since F(t) is the cumulative distribution function of Exp(λ), it follows that T ∼ Exp(λ). Derivation of the Uncertainty in the Estimated Rate We can estimate the uncertainty in 3.10 in Section 3.3): by using the propagation of error method (Equation For this expression to be useful, we need to know . Now the standard deviation of an Exp(λ) distribution is σ = 1/λ (this follows from Equation 4.35; note that the standard deviation is the same as the mean). Therefore . We can Page 270 replace λ with its estimate to obtain Now we can estimate the uncertainty : This propagation of error estimate is fairly good when the sample size is at least 20 or so. For smaller sample sizes it underestimates the uncertainty. Exercises for Section 4.7 1. Let T ∼ Exp(0.45). Find a. μT b. c. 2. P(T > 3) d. The median of T The time between requests to a web server is exponentially distributed with mean 0.5 seconds. a. What is the value of the parameter λ? b. What is the median time between requests? c. What is the standard deviation? 3. 4. 5. d. What is the 80th percentile? A catalyst researcher states that the diameters, in microns, of the pores in a new product she has made have the exponential distribution with parameter λ = 0.25. a. What is the mean pore diameter? b. What is the standard deviation of the pore diameters? c. What proportion of the pores are less than 3 microns in diameter? d. What proportion of the pores are greater than 11 microns in diameter? e. What is the median pore diameter? f. What is the third quartile of the pore diameters? g. What is the 99th percentile of the pore diameters? The distance between flaws on a long cable is exponentially distributed with mean 12 m. a. What is the value of the parameter λ? b. Find the median distance. c. Find the standard deviation of the distances. d. Find the 65th percentile of the distances. Refer to Exercise 3. Suppose that the diameters of 10 pores are independent. a. 6. Find the probability that more than seven of them have diameters less than 3 microns. b. Find the probability that exactly one of them has a diameter greater than 11 microns. Refer to Exercise 2. a. Find the probability that there will be exactly 5 requests in a 2-second time interval. b. Find the probability that there will be more than 1 request in a 1.5-second time interval. c. Find the probability that there will be no requests in a 1-second time interval. d. Find the probability that the time between requests is greater than 1 second. Page 271 e. Find the probability that the time between requests is between 1 and 2 seconds. f. 7. 8. If there have been no requests for the past two seconds, what is the probability that more than one additional second will elapse before the next request? Refer to Exercise 4. a. Find the probability that there are exactly 5 flaws in a 50 m length of cable. b. Find the probability that there are more than two flaws in a 20 m length of cable. c. Find the probability that there are no flaws in a 15 m length of cable. d. Find the probability that the distance between two flaws is greater than 15 m. e. Find the probability that the distance between two flaws is between 8 and 20 m. Someone claims that the waiting time, in minutes, between hits at a certain website has the exponential distribution with parameter λ = 1. a. Let X be the waiting time until the next hit. If the claim is true, what is P(X ≥ 5)? b. Based on the answer to part (a), if the claim is true, is five minutes an unusually long time to wait? c. If you waited five minutes until the next hit occurred, would you still believe the claim? Explain. 9. A certain type of component can be purchased new or used. Fifty percent of all new components last more than five years, but only 30% of used components last more than five years. Is it possible that the lifetimes of new components are exponentially distributed? Explain. 10. A radioactive mass emits particles according to a Poisson process at a mean rate of 2 per second. Let T be the waiting time, in seconds, between emissions. a. What is the mean waiting time? b. What is the median waiting time? c. Find P(T > 2). d. Find P(T < 0.1). e. Find P(0.3 < T < 1.5). f. If 3 seconds have elapsed with no emission, what is the probability that there will be an emission within the next second? 11. The number of traffic accidents at a certain intersection is thought to be well modeled by a Poisson process with a mean of 3 accidents per year. a. Find the mean waiting time between accidents. b. Find the standard deviation of the waiting times between accidents. c. Find the probability that more than one year elapses between accidents. d. Find the probability that less than one month elapses between accidents. e. If no accidents have occurred within the last six months, what is the probability that an accident will occur within the next year? 12. The distance between consecutive flaws on a roll of sheet aluminum is exponentially distributed with mean distance 3 m. Let X be the distance, in meters, between flaws. a. What is the mean number of flaws per meter? b. What is the probability that a 5 m length of aluminum contains exactly two flaws? 13. A radioactive mass emits particles according to a Poisson process. The probability that no particles will be emitted in a two-second period is 0.5. a. What is the probability that no particles are emitted in a four-second period? b. What is the probability that no particles are emitted in a one-second period? c. What is the mean number of particles emitted per second? 14. The lifetime of a transistor is exponentially distributed. The probability that the lifetime is greater than five years is 0.8. a. What is the probability that the lifetime is greater than 15 years? b. What is the probability that the lifetime is less than 2.5 years? c. What is the mean lifetime? 15. A light fixture contains five lightbulbs. The lifetime of each bulb is exponentially distributed with mean 200 hours. Whenever a bulb burns out, it is replaced. Let T be the time of the first bulb replacement. Let Xi, i = 1, …, 5, be the lifetimes of the five bulbs. Assume the lifetimes of the bulbs are independent. a. Find P(X1 > 100). Page 272 b. Find P(X1 > 100 and X2 > 100 and ⋯ and X5 > 100). c. Explain why the event T > 100 is the same as {X1 > 100 and X2 > 100 and ⋯ and X5 > 100}. d. Find P(T ≤ 100). e. Let t be any positive number. Find P(T ≤ t), which is the cumulative distribution function of T. f. Does T have an exponential distribution? g. Find the mean of T. h. If there were n lightbulbs, and the lifetime of each was exponentially distributed with parameter λ, what would be the distribution of T? 4.8 Some Other Continuous Distributions The Uniform Distribution The continuous uniform distribution, which we will sometimes refer to just as the uniform distribution, is the simplest of the continuous distributions. It often plays an important role in computer simulation studies. The uniform distribution has two parameters, a and b,with a < b. Definition The probability density function of the continuous uniform distribution with parameters a and b is (4.41) If X is a random variable with probability density function f (x), we say that X is uniformly distributed on the interval (a, b). Since the probability density function is constant on the interval (a, b), we can think of the probability as being distributed "uniformly" on the interval. If X is a random variable whose distribution is uniform on the interval (a, b), we write X ∼ U(a, b). The mean and variance of a uniform random variable can easily be computed from the definitions (Equations 2.37, 2.38, and 2.39). The derivations are left as an exercise. Let X ∼ U(a, b). Then (4.42) (4.43) Example 4.63 When a motorist stops at a red light at a certain intersection, the waiting time for the light to turn green, in seconds, is uniformly distributed on the interval (0, 30). Find the probability that the waiting time is between 10 and 15 seconds. Page 273 Solution Let X represent the waiting time. Then X ∼ U(0, 30), and we must compute P(10 < X < 15). We will do this by integrating the probability density function over the interval between 10 and 15. From Equation (4.41), the probability density function is Therefore Because the probability density function for a uniform random variable is constant over the range of possible values, probabilities for uniform random variables generally involve areas of rectangles, which can be computed without integrating. Figure 4.18 shows how to compute P(10 < X < 15) without an integral. The probability P(10 < X < 15) is equal to the area of the shaded region under the probability density function. This region is a rectangle with height 1/30 and width equal to 15 − 10 = 5. The probability is therefore 5/30, or 1/6. FIGURE 4.18 Because the probability density function for a uniform random variable is constant over the range of possible values, probabilities for uniform random variables generally involve areas of rectangles, which can be computed without integrating. Here P(10 < X < 15) is the area of a rectangle of dimensions 1/30 × 5, so the probability is 5/30, or 1/6. Example 4.64 Refer to Example 4.63. Find the mean and variance of the waiting time. Solution The waiting time X is uniformly distributed on the interval (0, 30). Using Equation (4.42), μX = (0 + 30)/2 = 15. Using Equation (4.43), . The Gamma Distribution The gamma distribution is a continuous distribution, one of whose purposes is to extend the usefulness of the exponential distribution in modeling waiting times. It involves a certain integral known as the gamma function. We define the gamma function and state some of its properties. Page 274 Definition For r > 0, the gamma function is defined by (4.44) The gamma function has the following properties: 1. If r is an integer, then Γ(r) = (r − 1)!. 2. For any r, Γ(r + 1) = rΓ(r). 3. . The gamma function is used to define the probability density function of the gamma distribution. The gamma probability density function has two parameters, r and λ, both of which are positive constants. Definition The probability density function of the gamma distribution with parameters r > 0 and λ > 0 is (4.45) If X is a random variable whose probability density function is gamma with parameters r and A, we write X ∼ Γ (r, λ). Note that when r = 1, the gamma distribution is the same as the exponential. In symbols, Γ (1, λ) = Exp(λ). Figure 4.19 presents plots of the gamma probability density function for various values of r and λ. FIGURE 4.19 The gamma probability density function for various values of r and λ. When the parameter r is an integer, the gamma distribution is a direct extension of Page 275 the exponential distribution. To be specific, recall that if events follow a Poisson process with rate parameter λ, the waiting time until an event occurs is distributed as Exp(λ). If r is any positive integer, then the waiting time until r events have occurred is distributed as Γ(r, λ). We can say this another way. Let X1 be the waiting time until the first event, and for i > 1let Xi be the waiting time between events i λ 1 and i. The waiting time until the rth event is the sum of the independent random variables X1 + + Xr, each of which is distributed as Exp(λ). Summary If X1, …, Xr are independent random variables, each distributed as Exp(λ), then the sum X1 + ⋯ + Xr is distributed as Γ(r, λ). Since the mean and variance of an exponential random variable are given by 1/λ and 1/λ2, respectively, we can use the fact that a gamma random variable is the sum of independent exponential random variables to compute the mean and variance of a gamma random variable in the case when r is an integer. The results are presented in the following box, and in fact, they are valid for any values of r and λ. If X ∼ Γ (r, λ), then (4.46) (4.47) Example 4.65 Assume that arrival times at a drive-through window follow a Poisson process with mean rate λ = 0.2 arrivals per minute. Let T be the waiting time until the third arrival. Find the mean and variance of T. Find P(T ≤ 20). Solution The random variable T is distributed Γ (3, 0.2). Using Equations (4.46) and (4.47) we compute μT = 3/0.2 = 15 and . To compute P(T ≤ 20) we reason as follows: T 20 means that the third event occurs within 20 minutes. This is the same as saying that the number of events that occur within 20 minutes is greater than or equal to 3. Now let X denote the number of events that occur within 20 minutes. What we have said is that P(T ≤ 20) = P(X ≥ 3). Now the mean of X is (20)(0.2) = 4, and X has a Poisson distribution, so X ∼ Poisson(4). It follows that Page 276 The method used in Example 4.65 to find P(T ≤ 20) can be used to find the cumulative distribution function F(x) = P(T < x) when T ∼ Γ(r, λ) and r is a positive integer. If T ∼ Γ(r, λ), and r is a positive integer, the cumulative distribution function of T is given by (4.48) A gamma distribution for which the parameter r is a positive integer is sometimes called an Erlang distribution. If r = k/2 where k is a positive integer, the Γ(r, 1/2) distribution is called a chi-square distribution with k degrees of freedom. The chi-square distribution is widely used in statistical inference. We will discuss some of its uses in Sections 5.8, 6.10, and 6.11. The Weibull Distribution The Weibull distribution is a continuous distribution that is used in a variety of situations. A common application of the Weibull distribution is to model the lifetimes of components such as bearings, ceramics, capacitors, and dielectrics. The Weibull probability density function has two parameters, both positive constants, that determine its location and shape. We denote these parameters α and β. The probability density function of the Weibull distribution is (4.49) If X is a random variable whose probability density function is Weibull with parameters α and β, we write X ∼ Weibull(α, β). Note that when α = 1, the Weibull distribution is the same as the exponential distribution with parameter λ = β. In symbols, Weibull(1, β) = Exp(β). Figure 4.20 presents plots of the Weibull(α, β) probability density function for several choices of the parameters α and β. By varying the values of α and β, a wide variety of curves can be generated. Because of this, the Weibull distribution can be made to fit a wide variety of data sets. This is the main reason for the usefulness of the Weibull distribution. FIGURE 4.20 The Weibull probability density function for various choices of α and β. The Weibull cumulative distribution function can be computed by integrating the probability density function: (4.50) This integral is not as hard as it looks. Just substitute u = (βt)α, and dt. The mean and variance of the Weibull distribution are expressed in terms of the gamma function. Page 277 If X ∼ Weibull(α, β), then (4.51) (4.52) In the special case that 1/α is an integer, then If the quantity 1/α is an integer, then 1 + 1/α and 1 + 2/α are integers, so Property 1 of the gamma function can be used to compute μX and exactly. If the quantity 1/α is of the form n/2, where n is an integer, then in principle μX and can be computed exactly through repeated applications of Properties 2 and 3 of the gamma function. For other values of α, μX and must be approximated. Many computer packages can do this. Example 4.66 In the article "Snapshot: A Plot Showing Program through a Device Development Laboratory" (D. Lambert, J. Landwehr, and M. Shyu, Statistical Case Studies for Industrial Process Improvement, ASA-SIAM 1997), the authors suggest using a Weibull distribution to model the duration of a bake step in the manufacture of a semiconductor. Let T represent the duration in hours of the bake step for a randomly chosen lot. If T ∼ Weibull(0.3, 0.1), what is the probability that the bake step takes longer than four hours? What is the probability that it takes between two and seven hours? Page 278 Solution We use the cumulative distribution function, Equation (4.50). Substituting 0.3 for a and 0.1 for β, we have Therefore The probability that the step takes between two and seven hours is Exercises for Section 4.8 1. The distance advanced in a day by a tunnel boring machine, in meters, is uniformly distributed on the interval (30, 50). a. Find the mean distance. b. Find the standard deviation of the distances. c. Find the probability that the distance is between 35 and 42 meters. d. 2. Suppose the distances advanced on different days are independent. What is the probability the machine advances more than 45 meters on exactly four out of 10 days? Resistors are labeled 100 Ω. In fact, the actual resistances are uniformly distributed on the interval (95, 103). a. Find the mean resistance. b. Find the standard deviation of the resistances. c. Find the probability that the resistance is between 98 and 102 Ω. d. 3. Suppose that resistances of different resistors are independent. What is the probability that three out of six resistors have resistances greater than 100 Ω? Let T ∼ Γ(4, 0.5). a. Find μT. b. Find σT. c. Find P(T ≤ 1). 4. d. Find P(T ≥ 4). Let T ∼ Γ(r, λ). If μT = 8 and 5. Let T ∼ Γ(r, λ). If μT = 8 and r = 16, find λ and σ2. 6. The lifetime, in years, of a type of small electric motor operating under adverse conditions is exponentially distributed with λ = 3.6. Whenever a motor fails, it is replaced with another of the same type. Find the probability that fewer than six motors fail within one year. , find r and λ. Page 279 7. 8. Let T ∼ Weibull(0.5, 3). a. Find μT. b. Find σT. c. Find P(T < 1). d. Find P(T > 5). e. Find P(2 < T < 4). If T is a continuous random variable that is always positive (such as a waiting time), with probability density function f(t) and cumulative distribution function F(t), then the hazard function is defined to be the function The hazard function is the rate of failure per unit time, expressed as a proportion of the items that have not failed. 9. a. If T ∼ Weibull(α, β), find h(t). b. For what values of α is the hazard rate increasing with time? For what values of α is it decreasing? c. If T has an exponential distribution, show that the hazard function is constant. In the article "Parameter Estimation with Only One Complete Failure Observation" (W. Pang, P. Leung, et al., International Journal of Reliability, Quality, and Safety Engineering, 2001:109–122), the lifetime, in hours, of a certain type of bearing is modeled with the Weibull distribution with parameters α = 2.25 and μ = 4.474 × 10−4. a. Find the probability that a bearing lasts more than 1000 hours. b. Find the probability that a bearing lasts less than 2000 hours. c. Find the median lifetime of a bearing. d. The hazard function is defined in Exercise 8. What is the hazard at t = 2000 hours? 10. The lifetime of a certain battery is modeled with the Weibull distribution with α = 2 and β = 0.1. a. What proportion of batteries will last longer than 10 hours? b. What proportion of batteries will last less than 5 hours? c. What proportion of batteries will last longer than 20 hours? d. The hazard function is defined in Exercise 8. What is the hazard at t = 10 hours? 11. The lifetime of a cooling fan, in hours, that is used in a computer system has the Weibull distribution with α = 1.5 and β = 0.0001. a. What is the probability that a fan lasts more than 10,000 hours? b. What is the probability that a fan lasts less than 5000 hours? c. What is the probability that a fan lasts between 3000 and 9000 hours? 12. Someone suggests that the lifetime T (in days) of a certain component can be modeled with the Weibull distribution with parameters α = 3 and β = 0.01. a. If this model is correct, what is P(T ≤ 1)? b. Based on the answer to part (a), if the model is correct, would one day be an unusually short lifetime? Explain. c. If you observed a component that lasted one day, would you find this model to be plausible? Explain. d. If this model is correct, what is P(T ≤ 90)? e. Based on the answer to part (d), if the model is correct, would 90 days be an unusually short lifetime? An unusually long lifetime? Explain. f. If you observed a component that lasted 90 days, would you find this model to be plausible? Explain. 13. A system consists of two components connected in series. The system will fail when either of the two components fails. Let T be the time at which the system fails. Let X1 and X2 be the lifetimes of the two components. Assume that X1 and X2 are independent and that each has the Weibull distribution with α = 2 and β = 0.2. a. Find P(X1 > 5). b. Find P(X1 > 5 and X2 > 5). c. Explain why the event T > 5 is the same as the event {X1 > 5 and X2 > 5}. d. Find P(T ≤ 5). e. Let t be any positive number. Find P(T ≤ t), which is the cumulative distribution function of T. f. Does T have a Weibull distribution? If so, what are its parameters? Page 280 14. Let X ∼ U(a, b). Use the definition of the mean of a continuous random variable (Equation 2.37) to show that μX = (a + b)/2. 15. Let X ∼ U(a, b). Use the definition of the variance of a continuous random variable (Equation 2.38 or 2.39) to show that . 16. Let X ∼ U(a, b). a. Show that if x ≤ a then P(X ≤ x) = 0. b. Show that if x > b then P(X ≤ x) = 1. c. Show that if a ≤ x ≤ b then P(X ≤ x) = (x − a)/(b − a). Conclude that the cumulative distribution function of X is 17. Let U ∼ U(0, 1). Let a and b be constants with a < b, and let X = (b − a)U + a. a. Find the cumulative distribution function of U (use the result of Exercise 16). b. Show that P(X ≤ x) = P(U ≤ (x − a)/(b − a)). c. Use the result of part (b) to show that X ∼ U(a, b). 4.9 Some Principles of Point Estimation When data are collected, it is often with the purpose of estimating some numerical characteristic of the population from which they came. For example, if X1, ..., Xn is a random sample from a population, the sample mean is often used to estimate the population mean μ, and the sample variance s2 is often used to estimate the population variance σ2. As another example, if X ∼ Bin(n, p), the sample proportion is often used to estimate the unknown population proportion p (see Section 4.2). In general, a quantity calculated from data is called a statistic, and a statistic that is used to estimate an unknown constant, or parameter, is called a point estimator or point estimate. The term point estimate is used when a particular numerical value is specified for the data. For example, if X ∼ Bin(10, p), and we observe X = 3, then the number is a point estimate of the unknown parameter p. On the other hand, if no particular value is specified for X, the random quantity is often called a point estimator of p. Often, point estimators and point estimates are simply called estimators and estimates, respectively. In this section, we will briefly address two questions: 1. Given a point estimator, how do we determine how good it is? 2. What methods can be used to construct good point estimators? In what follows, we will use the letter θ to denote an unknown parameter, and to denote an estimator of θ. Measuring the Goodness of an Estimator Ideally, an estimator should be both accurate and precise. The accuracy of an estimator is measured by its bias, and the precision is measured by its standard deviation, or uncertainty. The quantity most often used to evaluate the overall goodness of an estimator is the mean squared error (abbreviated MSE), which combines both bias and uncertainty. The bias of the estimator is , the difference between the mean of the estimator and the true value. The Page 281 uncertainty is the standard deviation , and is sometimes referred to as the standard error of the estimator. The MSE is found by adding the variance to the square of the bias. Definition Let θ be a parameter, and an estimator of θ. The mean squared error (MSE) of is (4.53) An equivalent expression for the MSE is (4.54) Equation (4.53) says that the MSE is equal to the square of the bias, plus the variance. To interpret Equation (4.54), note that the quantity is the difference between the estimated value and the true value, and is called the error. So Equation (4.54) says that the MSE is the mean of the squared error, and indeed it is this property that gives the MSE its name. Equations (4.53) and (4.54) yield identical results, so either may be used in any situation to compute the MSE. In practice Equation (4.53) is often somewhat easier to use. Example 4.67 Let X ∼ Bin(n, p) where p is unknown. Find the MSE of . Solution We compute the bias and variance of and use Equation (4.53). As shown in Section 4.2, the bias of is 0, and the variance is p(1 − p)/n. Therefore the MSE is 0 + p(1 − p)/n, or p(1 − p)/n. In Example 4.67 the estimator was unbiased, so the MSE was equal to the variance of the estimator. In some cases an unbiased estimator can be altered in a way that adds a small bias, but reduces the variance by a greater amount. Following is an example involving the sample variance. Example 4.68 Let X1, ..., Xn be a random sample from a N(μ, σ2) population. The sample variance is . It can be shown that s2 has mean . Consider the estimator and variance , in which the sum of the squared deviations is divided by n rather than n − 1. Compute the bias, variance, and mean squared error of both s2 and . Show that has smaller mean squared error than s2. Page 282 Solution Since , s2 is unbiased for σ2, so the mean squared error is equal to the variance: . To compute the bias and variance of , note that It follows that Therefore The variance is given by The mean squared error of is therefore To show that has smaller mean squared error than s2, we subtract: We now turn to the question of how to find good estimators. We focus on the most widely used method, which is the method of maximum likelihood. The Method of Maximum Likelihood The idea behind the method of maximum likelihood is to estimate a parameter with the value that makes the observed data most likely. To illustrate the method, let X ∼ Bin(20, p) where p is unknown. Suppose we observe the value X = 7. The probability mass function is Notice that we have included the symbol for the parameter p in the notation for the probability mass function; that is, we have written f(7; p) rather than f(7). The reason for this is that we will treat the probability mass function as a function of p, with the data value 7 being Page 283 constant. When a probability mass function or probability density function is considered to be a function of parameters, it is called a likelihood function. The maximum likelihood estimate (MLE) is the value which, when substituted for p, maximizes the likelihood function. We will now discuss the mechanics of computing the maximum of the likelihood function f(7; p). In principle, we could maximize this function by taking the derivative with respect to p and setting it equal to 0. It is easier, however, to maximize ln f(7; p) instead. Note that the quantity that maximizes the logarithm of a function is always the same quantity that maximizes the function itself. Now We take the derivative with respect to p and set it equal to 0: Solving, we find that the maximizing value is 7/20. Therefore the maximum likelihood estimate is . It is easy to see that whatever value is observed for X, the maximizing value would be X/20. We say, therefore, that the maximum likelihood estimator is . Note that we are using the word “estimate” when a specific value, like 7, is given for X, and the word “estimator” when the data value is unspecified, and represented by the random variable X. In principle, the likelihood function can be a probability density function or a probability mass function. It can also be a joint probability density or mass function, and is often the joint density or mass function of independent random variables. (Joint distributions are covered in Section 2.6.) The likelihood function can also be a function of several parameters. We now present the definition of a maximum likelihood estimator. Definition Let X1, ..., Xn have joint probability density or probability mass function , where θ1, ..., θk are parameters, and x1, ..., xn are the values observed for X1, ..., Xn. The values that maximize f are the maximum likelihood estimates of θ1, ..., θk. If the random variables X1, ..., Xn are substituted for x1, ..., xn, then are called maximum likelihood estimators. The abbreviation MLE is often used for both maximum likelihood estimate and maximum likelihood estimator. Example 4.69 Let X1, ..., Xn be a random sample from an Exp(λ) population, where λ is unknown. Find the MLE of λ. Page 284 Solution The likelihood function is the joint probability density function of X1, ..., Xn, considered as a function of the parameter λ. Since X1, ..., Xn are independent, the joint probability density function is the product of the marginals, each of which is an Exp(λ) probability density function. Therefore The MLE is the value of λ that maximizes the likelihood function. Multiplying out the product yields As is often the case, it is easier to maximize the logarithm of the likelihood: Taking the derivative with respect to λ and setting it equal to 0 yields Solving for λ yields the MLE: Desirable Properties of Maximum Likelihood Estimators Maximum likelihood is the most commonly used method of estimation. The main reason for this is that in most cases that arise in practice, MLEs have two very desirable properties. 1. In most cases, as the sample size n increases, the bias of the MLE converges to 0. 2. In most cases, as the sample size n increases, the variance of the MLE converges to a theoretical minimum. Together, these two properties imply that when the sample size is sufficiently large, the bias of the MLE will be negligible, and the variance will be nearly as small as is theoretically possible. Exercises for Section 4.9 1. Choose the best answer to fill in the blank. If an estimator is unbiased, then ______ i. the estimator is equal to the true value. ii. the estimator is usually close to the true value. iii. the mean of the estimator is equal to the true value. iv. the mean of the estimator is usually close to the true value. 2. Choose the best answer to fill in the blank. The variance of an estimator measures ______ i. how close the estimator is to the true value. Page 285 ii. how close repeated values of the estimator are to each other. iii. how close the mean of the estimator is to the true value. iv. how close repeated values of the mean of the estimator are to each other. 3. Let X1 and X2 be independent, each with unknown mean μ and known variance σ2 = 1. a. Let . Find the bias, variance, and mean squared error of . b. Let . Find the bias, variance, and mean squared error of . c. Let . Find the bias, variance, and mean squared error of . d. For what values of μ does have smaller mean squared error than ? e. For what values of μ does have smaller mean squared error than ? 4. Let X1, ..., Xn be a simple random sample from a N(μ, σ2) population. For any constant k > 0, define . Consider as an estimator of σ2. a. Compute the bias of in terms of k. [Hint: The sample variance s2 is unbiased, and .] b. Compute the variance of in terms of k. [Hint: , and .] c. Compute the mean squared error of in terms of k. d. For what value of k is the mean squared error of minimized? 5. Let X ∼ Geom(p). Find the MLE of p. 6. Let X1, ..., Xn be a random sample from a population with the Poisson(λ) distribution. Find the MLE of λ. 7. Maximum likelihood estimates possess the property of functional invariance, which means that if is the MLE of θ, and h(θ) is any function of θ, then is the MLE of h(θ). a. Let X ∼ Bin(n, p) where n is known and p is unknown. Find the MLE of the odds ratio p/(1 − p). b. Use the result of Exercise 5 to find the MLE of the odds ratio p/(1 − p) if X ∼ Geom(p). c. If X ∼ Poisson(λ), then P(X = 0) = e−λ. Use the result of Exercise 6 to find the MLE of P(X = 0) if X1, ..., Xn is a random sample from a population with the Poisson(λ) distribution. 8. Let X1, ..., Xn be a random sample from a N(μ, 1) population. Find the MLE of μ. 9. Let X1, ..., Xn be a random sample from a N(0, σ2) population. Find the MLE of σ. 10. Let X1, ..., Xn be a random sample from a N(μ, σ2) population. Find the MLEs of μ and of σ. (Hint: The likelihood function is a function of two parameters, μ and σ. Compute partial derivatives with respect to μ and σ and set them equal to 0 to find the values and that maximize the likelihood function.) 4.10 Probability Plots Scientists and engineers frequently work with data that can be thought of as a random sample from some population. In many such cases, it is important to determine a probability distribution that approximately describes that population. In some cases, knowledge of the process that generated the data can guide the decision. More often, though, the only way to determine an appropriate distribution is to examine the sample to find a probability distribution that fits. Probability plots provide a good way to do this. Given a random sample X1, ..., Xn, a probability plot can determine whether the sample might plausibly have come from some specified population. We will present the idea behind probability plots with a simple example. A random sample of size 5 is drawn, and we want to determine whether the population Page 286 from which it came might have been normal. The sample, arranged in increasing order, is Denote the values, in increasing order, by X1, ..., Xn (n = 5 in this case). The first thing to do is to assign increasing, evenly spaced values between 0 and 1 to the Xi. There are several acceptable ways to do this; perhaps the simplest is to assign the value (i − 0.5)/n to Xi. The following table shows the assignment for the given sample. i 1 2 3 4 5 Xi 3.01 3.35 4.79 5.96 7.89 (i − 0.5)/5 0.1 0.3 0.5 0.7 0.9 The value (i − 0.5)/n is chosen to reflect the position of Xi in the ordered sample. There are i − 1 values less than Xi, and i values less than or equal to Xi. The quantity (i − 0.5)/n is a compromise between the proportions (i − 1)/n and i/n. The goal is to determine whether the sample might have come from a normal population. The most plausible normal distribution is the one whose mean and standard deviation are the same as the sample mean and standard deviation. The sample mean is , and the sample standard deviation is s = 2.00. We will therefore determine whether this sample might have come from a N(5, 22) distribution. Figure 4.21 is a plot of the five points (Xi, (i − 0.5)/5). The curve is the cumulative distribution function (cdf) F(x) of the N(5, 22) distribution. Recall that F(x) = P(X ≤ x) where X ∼ N(5, 22). The curve is the cdf of N(5, 22). If the sample points X1, ..., X5 came from this distribution, they are likely to lie close to the curve. FIGURE 4.21 We have drawn horizontal lines through the sample points. We denote the x Page 287 values of the points on the cdf that are crossed by the line, in increasing order, by Q1, ..., Q5. Now the horizontal line through (X1, 0.1) intersects the cdf at the point (Q1, 0.1). This means that the proportion of values in the N(5, 22) population that are less than or equal to Q1 is 0.1. Another way to say this is that Q1 is the 10th percentile of the N(5, 22) distribution. If the sample X1, ..., X5 truly came from a N(5, 22) distribution, then it is reasonable to believe that the lowest value in the sample, X1, would be fairly close to the 10th percentile of the population, Q1. Intuitively, the reason for this is that we would expect that the lowest of five points would be likely to come from the lowest fifth, or 20%, of the population, and the 10th percentile is in the middle of that lowest 20%. Applying similar reasoning to the remaining points, we would expect each Qi to be close to its corresponding Xi. The probability plot consists of the points (Xi, Qi). Since the distribution that generated the Qi was a normal distribution, this is called a normal probability plot. If X1, ..., Xn do in fact come from the distribution that generated the Qi, the points should lie close to a straight line. To construct the plot, we must compute the Qi. These are the 100(i − 0.5)/n percentiles of the distribution that is suspected of generating the sample. In this example the Qi are the 10th, 30th, 50th, 70th, and 90th percentiles of the N(5, 22) distribution. We could approximate these values by looking up the z-scores corresponding to these percentiles, and then converting to raw scores. In practice, the Qi are invariably calculated by a computer software package. The following table presents the Xi and the Qi for this example. i 1 2 3 4 5 Xi 3.01 3.35 4.79 5.96 7.89 Qi 1.44 3.95 5.00 6.05 7.56 Figure 4.22 (page288) presents a normal probability plot for the sample X1, ..., X5. A straight line is superimposed on the plot, to make it easier to judge whether the points lie close to a straight line or not. Two versions of the plot are presented; they are identical except for the scaling on the vertical axis. In the plot on the left, the values on the vertical axis represent the Qi. In the plot on the right, the values on the vertical axis represent the percentiles (as decimals, so 0.1 is the 10th percentile) of the Qi. For example, the 10th percentile of N(5, 22) is 2.44, so the value 0.1 on the right-hand plot corresponds to the value 2.44 on the left-hand plot. The 50th percentile, or median, is 5, so the value 0.5 on the right-hand plot corresponds to the value 5 on the left-hand plot. Computer packages often scale the vertical axis like the plot on the right. In Figure 4.22, the sample points are close to the line, so it is quite plausible that the sample came from a normal distribution. FIGURE 4.22 Normal probability plots for the sample X1, ..., X5. The plots are identical, except for the scaling on the vertical axis. The sample points lie approximately on a straight line, so it is plausible that they came from a normal population. We remark that the points Q1, ..., Qn are called quantiles of the distribution from which they are generated. Sometimes the sample points X1, ..., Xn are called empirical quantiles. For this reason the probability plot is sometimes called a quantile-quantile plot, or QQ plot. In this example, we used a sample of only five points to make the calculations clear. In practice, probability plots work better with larger samples. A good rule of thumb is to Page 288 require at least 30 points before relying on a probability plot. Probability plots can still be used for smaller samples, but they will detect only fairly large departures from normality. Figure 4.23 shows two normal probability plots. The plot in Figure 4.23a is of the monthly productions of 255 gas wells. These data do not lie close to a straight line, and thus do not come from a population that is close to normal. The plot in Figure 4.23b is of the natural logs of the monthly productions. These data lie much closer to a straight line, although some departure from normality can be detected. (Figure 4.16 in Section 4.6 presents histograms of these data.) FIGURE 4.23 Two normal probability plots. (a) A plot of the monthly productions of 255 gas wells. These data do not lie close to a straight line, and thus do not come from a population that is close to normal. (b) A plot of the natural logs of the monthly productions. These data lie much closer to a straight line, although some departure from normality can be detected. See Figure 4.16 for histograms of these data. Page 289 Interpreting Probability Plots It's best not to use hard-and-fast rules when interpreting a probability plot. Judge the straightness of the plot by eye. When deciding whether the points on a probability plot lie close to a straight line or not, do not pay too much attention to the points at the very ends (high or low) of the sample, unless they are quite far from the line. It is common for a few points at either end to stray from the line somewhat. However, a point that is very far from the line when most other points are close is an outlier, and deserves attention. Exercises for Section 4.10 1. 2. 3. Each of three samples has been plotted on a normal probability plot. For each, say whether the sample appears to have come from an approximately normal population. 4. As part of a quality-control study aimed at improving a production line, the weights (in ounces) of 50 bars of soap are measured. The results are as follows, sorted from smallest to largest. 11.612.6 12.7 12.8 13.1 13.3 13.6 13.713.814.1 14.314.3 14.6 14.8 15.1 15.2 15.6 15.615.715.8 15.815.9 15.9 16.1 16.2 16.2 16.3 16.416.516.5 16.516.6 17.0 17.1 17.3 17.3 17.4 17.417.617.7 17.718.1 18.3 18.3 18.3 18.5 18.5 18.819.2 20.3 Construct a normal probability plot for these data. Do these data appear to come from an approximately normal distribution? 5. Below are the durations (in minutes) of 40 eruptions of the geyser Old Faithful in Yellowstone National Park. 4.1 1.8 3.2 1.9 4.6 2.0 4.5 3.9 4.3 2.3 3.8 1.9 4.6 1.8 4.7 1.8 4.6 1.9 3.5 4.0 3.7 3.7 4.3 3.6 3.8 3.8 3.8 2.5 4.5 4.1 3.7 3.8 3.4 4.0 2.3 4.4 4.1 4.3 3.3 2.0 Construct a normal probability plot for these data. Do the data appear to come from an approximately normal distribution? 6. Below are the durations (in minutes) of 40 time intervals between eruptions of the geyser Old Faithful in Yellowstone National Park. 91 51 79 53 82 51 76 82 84 53 86 51 85 45 88 51 80 49 82 75 73 67 68 86 72 75 75 66 84 70 79 60 86 71 67 81 76 83 76 55 Construct a normal probability plot for these data. Do they appear to come from an approximately normal distribution? 7. Construct a normal probability plot for the PM data in Table 1.2 (page 21). Do the PM data appear to come from a normal population? 8. Construct a normal probability plot for the logs of the PM data in Table 1.2. Do the logs of the PM data appear to come from a normal population? 9. Can the plot in Exercise 6 be used to determine whether the data are approximately normal? Explain. 10. In the article “Assessment of Dermatopharmacokinetic Approach in the Bioequivalence Determination of Topical Tretinoin Gel Products” (L. Pershing, J. Nelson, et al., Journal of The American Academy of Dermatology, 2003:740–751), measurements of the concentration of an anti-fungal gel, in ng per square centimeter of skin, were made one hour after application for 49 individuals. Following are the results. The authors claim that these data are well-modeled by a lognormal distribution. Construct an appropriate probability plot, and use it to determine whether the data support this claim. 132.44 76.73 258.46177.46 73.0113.0622.35 141.6 107.54 75.95 70.37 88.76 104.00 19.07174.30 82.87 68.73 41.47120.44 136.52 82.46 67.04 96.92 93.26 72.92138.15 82.43245.41104.68 82.51332.49 147.12129.82 54.83 65.82 75.24 135.52132.21 85.63135.79 65.98349.71 77.84 89.19102.94 166.11168.76 155.20 44.35202.51 Figure 4.23 (page 288) shows that nonnormal data can sometimes be made approximately normal by applying an appropriate function (in this case the natural logarithm). This is known as transforming the data. Exercises 9 and 10 present examples of other transformations. Consider the following data set: 15082 1512 273398 3375 1 3017 366 5545 7 9958 32 4434 233 171310 25 0 90 6643 4218 933820 11335 23 a. Construct a normal probability plot for these data. Do they appear to come from an approximately normal distribution? b. Transform the data by taking the square root of each value in the data set. Show that the square roots appear to come from an approximately normal distribution. 10. Consider the following data set: 0.64 0.29 −0.38 0.490.81 −0.83 10.04 0.79 0.16 0.20 −0.78 0.190.54 1.23 0.74 0.35 0.350.57 −0.94 0.55 0.31 0.67 0.43 0.54 3.390.84 −2.32 0.48 −1.03 −1.17 −2.15 −0.270.340.81 −2.62 0.36 −0.55 a. Construct a normal probability plot for these data. Do they appear to come from an approximately normal distribution? b. Transform the data by taking the reciprocal of each value in the data set. Show that the reciprocals appear to come from an approximately normal distribution. 4.11 The Central Limit Theorem The Central Limit Theorem is by far the most important result in statistics. Many commonly used statistical methods rely on this theorem for their validity. The Central Limit Theorem says that if we draw a large enough sample from a population, then the distribution of the sample mean is approximately normal, no matter what population the sample was drawn from. This allows us to compute probabilities for sample means using the z table, even though the Page 291 population from which the sample was drawn is not normal. We now explain this more fully. Let X1, ..., Xn be a simple random sample from a population with mean μ and variance σ2. Let be the sample mean. Now imagine drawing many such samples and computing their sample means. If one could draw every possible sample of size n from the original population, and compute the sample mean for each one, the resulting collection would be the population of sample means. What do we know about this population? We know from Equations 2.58 and 2.59 in Section 2.5 that its mean is and its variance is . What about the probability density function for this population? One might think that the shape of this probability density function would depend on the shape of the population from which the sample was drawn. The surprising thing is that if the sample size is sufficiently large, this is not so. If the sample size is large enough, the distribution of the sample mean is approximately normal, no matter what the distribution of the population from which the sample was drawn. The Central Limit Theorem Let X1, ..., Xn be a simple random sample from a population with mean μ and variance σ2. Let be the sample mean. Let Sn = X1 + ··· + Xn be the sum of the sample observations. Then if n is sufficiently large, (4.55) and (4.56) Note that the statement of the Central Limit Theorem specifies that and , which hold for any sample mean. The sum of the sample items is equal to the mean multiplied by the sample size, that is, . It follows that and (see Equations 2.43 and 2.44 in Section 2.5). The Central Limit Theorem says that and Sn are approximately normally distributed, if the sample size n is large enough. The natural question to ask is: How large is large enough? The answer depends on the shape of the underlying population. If the sample is drawn from a nearly symmetric distribution, the normal approximation can be good even for a fairly small value of n. However, if the population is heavily skewed, a fairly large n may be necessary. Empirical evidence suggests that for most populations, a sample size of 30 or more is large enough for the normal approximation to be adequate (see Figure 4.24 on page 292). For most populations, if the sample size is greater than 30, the Central Limit Theorem approximation is good. Page 292 FIGURE 4.24 The leftmost plot in each row is the distribution (probability density function or probability mass function) of a random variable. The two plots to its right are the distributions of the sample mean (solid line) for samples of sizes 5 and 30, respectively, with the normal curve (dashed line) superimposed. Top row: Since the original distribution is nearly symmetric, the normal approximation is good even for a sample size as small as 5. Middle row: The original distribution is somewhat skewed. Even so, the normal approximation is reasonably close even for a sample of size 5, and very good for a sample of size 30. Bottom row: The original distribution is highly skewed. The normal approximation is not good for a sample size of 5, but is reasonably good for a sample of size 30. Note that two of the original distributions are continuous, and one is discrete. The Central Limit Theorem holds for both continuous and discrete distributions. Page 293 Example 4.70 Let X denote the number of flaws in a 1 in. length of copper wire. The probability mass function of X is presented in the following table. x P(X = x) 0 0.48 1 0.39 2 0.12 3 0.01 One hundred wires are sampled from this population. What is the probability that the average number of flaws per wire in this sample is less than 0.5? Solution The population mean number of flaws is μ = 0.66, and the population variance is σ2 = 0.5244. (These quantities can be calculated using Equations 2.29 and 2.30 in Section 2.4.) Let X1, ..., X100 denote the number of flaws in the 100 wires sampled from this population. We need to find . Now the sample size is n = 100, which is a large sample. It follows from the Central Limit Theorem (expression 4.55) that . The z-score of 0.5 is therefore From the z table, the area to the left of −2.21 is 0.0136. Therefore , so only 1.36% of samples of size 100 will have fewer than 0.5 flaws per wire. See Figure 4.25. FIGURE 4.25 Solution to Example 4.70. Note that in Example 4.70 we needed to know only the mean and variance of the population, not the probability mass function. Example 4.71 At a large university, the mean age of the students is 22.3 years, and the standard deviation is 4 years. A random sample of 64 students is drawn. What is the probability that the average age of these students is greater than 23 years? Page 294 Solution Let X1, ..., X64 be the ages of the 64 students in the sample. We wish to find . Now the population from which the sample was drawn has mean μ = 22.3 and variance σ2 = 16. The sample size is n = 64. It follows from the Central Limit Theorem (expression 4.55) that . The z-score for 23 is From the z table, the area to the right of 1.40 is 0.0808. Therefore . See Figure 4.26. FIGURE 4.26 Solution to Example 4.71. In Section 4.5 we showed that linear combinations of independent normal random variables are normal (Equations 4.26, 4.28, and 4.29). This fact can be combined with the Central Limit Theorem to find probabilities regarding linear combinations of large sample sums and sample means. Example 4.72 illustrates the method. Example 4.72 The manufacture of a certain part requires two different machine operations. The time on machine 1 has mean 0.4 hours and standard deviation 0.1 hours. The time on machine 2 has mean 0.45 hours and standard deviation 0.15 hours. The times needed on the machines are independent. Suppose that 65 parts are manufactured. What is the distribution of the total time on machine 1? On machine 2? What is the probability that the total time used by both machines together is between 50 and 55 hours? Solution Let X1, ..., X65 represent the times of the 65 parts on machine 1. The population from which this sample was drawn has mean μX = 0.4 and standard deviation σX = 0.1. Let SX = X1 + ··· + X65 be the total time on machine 1. It follows from the Central Limit Theorem (expression 4.56) that Now let Y1, ..., Y65 represent the times of the 65 parts on machine 2. Then μY = 0.45 and σY = 0.15. Let SY = Y1 + ··· + Y65 be the total time on machine 2. Proceeding by the same method used to compute the distribution of the total time on machine 1, we see that Page 295 Finally, let T = SX + SY represent the total time on both machines. Since SX ∼ N(26, 0.65), SY ∼ N(29.25, 1.4625), and SX and SY are independent, it follows that μT = 26 + 29.25 = 55.25, , and T ∼ N(55.25, 2.1125) To find P(50 < T < 55) we compute the z-scores of 50 and of 55. The area to the left of z = −3.61 is 0.0002. The area to the left of z = −0.17 is 0.4325. The area between z = −3.61 and z = −0.17 is 0.4325 − 0.0002 = 0.4323. The probability that the total time used by both machines is between 50 and 55 hours is 0.4323. Normal Approximation to the Binomial Recall from Section 4.2 that if X ∼ Bin(n, p), then , where Y1, ..., Yn is a sample from a Bernoulli(p) population. Therefore X is the sum of the sample observations. The sample proportion is which is also the sample mean . The Bernoulli(p) population has mean μ = p and variance σ2 = p(1 − p). It follows from the Central Limit Theorem that if the number of trials n is large, then X ∼ N(np, np(1 − p)), and . Again the question arises, how large a sample is large enough? In the binomial case, the accuracy of the normal approximation depends on the mean number of successes np and on the mean number of failures n(1 − p). The larger the values of np and n(1 − p), the better the approximation. A common rule of thumb is to use the normal approximation whenever np > 5 and n(1 − p) > 5. A better and more conservative rule is to use the normal approximation whenever np > 10 and n(1 − p) > 10. Summary If X ∼ Bin(n, p), and if np > 10 and n(1 − p) > 10, then (4.57) (4.58) To illustrate the accuracy of the normal approximation to the binomial, Figure 4.27 (page 296) presents the Bin(100, 0.2) probability histogram with the N(20, 16) probability density function superimposed. While a slight degree of skewness can be detected in the binomial distribution, the normal approximation is quite close. Page 296 FIGURE 4.27 The Bin(100, 0.2) probability histogram, with the N(20, 16) probability density function superimposed. The Continuity Correction The binomial distribution is discrete, while the normal distribution is continuous. The continuity correction is an adjustment, made when approximating a discrete distribution with a continuous one, that can improve the accuracy of the approximation. To see how it works, imagine that a fair coin is tossed 100 times. Let X represent the number of heads. Then X ∼ Bin(100, 0.5). Imagine that we wish to compute the probability that X is between 45 and 55. This probability will differ depending on whether the endpoints, 45 and 55, are included or excluded. Figure 4.28 illustrates the case where the endpoints are included, that is, where we wish to compute P(45 ≤ X ≤ 55). The exact probability is given by the total area of the rectangles of the binomial Page 297 probability histogram corresponding to the integers 45 to 55 inclusive. The approximating normal curve is superimposed. To get the best approximation, we should compute the area under the normal curve between 44.5 and 55.5. In contrast, Figure 4.29 illustrates the case where we wish to compute P(45 < X < 55). Here the endpoints are excluded. The exact probability is given by the total area of the rectangles of the binomial probability histogram corresponding to the integers 46 to 54. The best normal approximation is found by computing the area under the normal curve between 45.5 and 54.5. FIGURE 4.28 To compute P(45 ≤ X ≤ 55), the areas of the rectangles corresponding to 45 and to 55 should be included. To approximate this probability with the normal curve, compute the area under the curve between 44.5 and 55.5. FIGURE 4.29 To compute P(45 < X < 55), the areas of the rectangles corresponding to 45 and to 55 should be excluded. To approximate this probability with the normal curve, compute the area under the curve between 45.5 and 54.5. In summary, to apply the continuity correction, determine precisely which rectangles of the discrete probability histogram you wish to include, and then compute the area under the normal curve corresponding to those rectangles. Example 4.73 If a fair coin is tossed 100 times, use the normal curve to approximate the probability that the number of heads is between 45 and 55 inclusive. Solution This situation is illustrated in Figure 4.28. Let X be the number of heads obtained. Then X ∼ Bin(100, 0.5). Substituting n = 100 and p = 0.5 into Equation (4.57), we obtain the normal approximation X ∼ N(50, 25). Since the endpoints 45 and 55 are to be included, we should compute the area under the normal curve between 44.5 and 55.5. The z-scores for 44.5 and 55.5 are From the z table we find that the probability is 0.7286. See Figure 4.30 (page 298). Page 298 FIGURE 4.30 Solution to Example 4.73. Example 4.74 If a fair coin is tossed 100 times, use the normal curve to approximate the probability that the number of heads is between 45 and 55 exclusive. Solution This situation is illustrated in Figure 4.29. Let X be the number of heads obtained. As in Example 4.73, X ∼ Bin(100, 0.5), and the normal approximation is X ∼ N(50, 25). Since the endpoints 45 and 55 are to be excluded, we should compute the area under the normal curve between 45.5 and 54.5. The z-scores for 45.5 and 54.5 are From the z table we find that the probability is 0.6318. See Figure 4.31. FIGURE 4.31 Solution to Example 4.74. Example 4.75 In a large university, 25% of the students are over 21 years of age. In a sample of 400 students, what is the probability that more than 110 of them are over 21? Solution Let X represent the number of students who are over 21. Then X ∼ Bin(400, 0.25). Substituting n = 400 and p = 0.25 into Equation (4.57), we obtain the normal approximation X ∼ N(100, 75). Since we want to find the probability that the number of students is more than Page 299 than 110, the value 110 is excluded. We therefore find P(X > 110.5). We compute the z-score for 110.5, which is Using the z table, we find that P(X > 110.5) = 0.1131. See Figure 4.32. FIGURE 4.32 Solution to Example 4.75. Accuracy of the Continuity Correction The continuity correction improves the accuracy of the normal approximation to the binomial distribution in most cases. For binomial distributions with large n and small p, however, when computing a probability that corresponds to an area in the tail of the distribution, use of the continuity correction can in some cases reduce the accuracy of the normal approximation somewhat. This results from the fact that the normal approximation is not perfect; it fails to account for a small degree of skewness in these distributions. In summary, use of the continuity correction makes the normal approximation to the binomial distribution better in most cases, but not all. Normal Approximation to the Poisson Recall that if X ∼ Poisson(λ), then X is approximately binomial with n large and np = λ. Recall also that μX = λ and . It follows that if λ is sufficiently large, i.e., λ > 10, then X is approximately binomial, with np > 10. It follows from the Central Limit Theorem that X is also approximately normal, with mean and variance both equal to λ. Thus we can use the normal distribution to approximate the Poisson. Summary If X ∼ Poisson(λ), where λ > 10, then Page 300 (4.59) Continuity Correction for the Poisson Distribution Since the Poisson distribution is discrete, the continuity correction can in principle be applied when using the normal approximation. For areas that include the central part of the curve, the continuity correction generally improves the normal approximation, but for areas in the tails the continuity correction sometimes makes the approximation worse. We will not use the continuity correction for the Poisson distribution. Example 4.76 The number of hits on a website follows a Poisson distribution, with a mean of 27 hits per hour. Find the probability that there will be 90 or more hits in three hours. Solution Let X denote the number of hits on the website in three hours. The mean number of hits in three hours is 81, so X ∼ Poisson(81). Using the normal approximation, X ∼ N(81, 81). We wish to find P(X ≥ 90). We compute the z-score of 90, which is Using the z table, we find that P(X ≥ 90) = 0.1587. See Figure 4.33. FIGURE 4.33 Solution to Example 4.76. Exercises for Section 4.11 1. Bottles filled by a certain machine are supposed to contain 12 oz of liquid. In fact the fill volume is random with mean 12.01 oz and standard deviation 0.2 oz. a. What is the probability that the mean volume of a random sample of 144 bottles is less than 12 oz? b. If the population mean fill volume is increased to 12.03 oz, what is the probability that the mean volume of a sample of size 144 will be less than 12 oz? 2. A 500-page book contains 250 sheets of paper. The thickness of the paper used to manufacture the book has mean 0.08 mm and standard deviation 0.01 mm. a. What is the probability that a randomly chosen book is more than 20.2 mm thick (not including the covers)? b. What is the 10th percentile of book thicknesses? c. Someone wants to know the probability that a randomly chosen page is more than 0.1 mm thick. Is enough information given to compute this probability? If so, compute the probability. If not, explain why not. 3. A commuter encounters four traffic lights each day on the way to work. Let X represent the number of these that are red lights. The probability mass function of X is as Page 301 follows. x P(X = x) 0 0.1 1 0.3 2 0.3 3 0.2 4 0.1 What is the probability that in a period of 100 days, the average number of red lights encountered is more than 2 per day? Among all monthly bills from a certain credit card company, the mean amount billed was $485 and the standard deviation was $300. In addition, for 15% of the bills, the amount billed was greater than $1000. A sample of 900 bills is drawn. a. What is the probability that the average amount billed on the sample bills is greater than $500? b. What is the probability that more than 150 of the sampled bills are for amounts greater than $1000? Bags checked for a certain airline flight have a mean weight of 15 kg with a standard deviation of 5 kg. A random sample of 60 bags is drawn. a. What is the probability that the sample mean weight is less than 14 kg? b. Find the 70th percentile of the sample mean weights. c. How many bags must be sampled so that the probability is 0.01 that the sample mean weight is less than 14 kg? 6. The amount of warpage in a type of wafer used in the manufacture of integrated circuits has mean 1.3 mm and standard deviation 0.1 mm. A random sample of 200 wafers is drawn. a. What is the probability that the sample mean warpage exceeds 1.305 mm? b. Find the 25th percentile of the sample mean. c. 7. How many wafers must be sampled so that the probability is 0.05 that the sample mean exceeds 1.305? The time spent by a customer at a checkout counter has mean 4 minutes and standard deviation 2 minutes. a. What is the probability that the total time taken by a random sample of 50 customers is less than 180 minutes? b. Find the 30th percentile of the total time taken by 50 customers. 7. The time taken by an automobile mechanic to complete an oil change is random with mean 29.5 minutes and standard deviation 3 minutes. a. What is the probability that 50 oil changes take more than 1500 minutes? b. What is the probability that a mechanic can complete 80 or more oil changes in 40 hours? c. The mechanic wants to reduce the mean time per oil change so that the probability is 0.99 that 80 or more oil changes can be completed in 40 hours. What does the mean time need to be? Assume the standard deviation remains 3 minutes. 9. The temperature of a solution will be estimated by taking n independent readings and averaging them. Each reading is unbiased, with a standard deviation of σ = 0.5°C. How many readings must be taken so that the probability is 0.90 that the average is within ±0.1 °C of the actual temperature? 10. Among the adults in a large city, 30% have a college degree. A simple random sample of 100 adults is chosen. What is the probability that more than 35 of them have a college degree? 11. In a process that manufactures bearings, 90% of the bearings meet a thickness specification. A shipment contains 500 bearings. A shipment is acceptable if at least 440 of the 500 bearings meet the specification. Assume that each shipment contains a random sample of bearings. a. What is the probability that a given shipment is acceptable? b. What is the probability that more than 285 out of 300 shipments are acceptable? c. What proportion of bearings must meet the specification in order that 99% of the shipments are acceptable? 12. Concrete blocks are produced in lots of 2000. Each block has probability 0.85 of meeting a strength specification. The blocks are independent. a. What is the probability that, in a given lot, fewer than 1690 blocks meet the specification? b. Find the 70th percentile of the number of blocks that meet the specification. c. In a group of six lots, what is the probability that fewer than 1690 blocks meet the specification in three or more of them? Page 302 13. Radioactive mass A emits particles at a mean rate of 20 per minute, and radioactive mass B emits particles at a mean rate of 25 per minute. a. What is the probability that fewer than 200 particles are emitted by both masses together in a five-minute time period? b. What is the probability that mass B emits more particles than mass A in a two-minute time period? 14. The concentration of particles in a suspension is 30 per mL. a. What is the probability that a 2 mL sample will contain more than 50 particles? b. Ten 2 mL samples are drawn. What is the probability that at least 9 of them contain more than 50 particles? c. One hundred 2 mL samples are drawn. What is the probability that at least 90 of them contain more than 50 particles? 15. The concentration of particles in a suspension is 50 per mL. A 5 mL volume of the suspension is withdrawn. a. What is the probability that the number of particles withdrawn will be between 235 and 265? b. What is the probability that the average number of particles per mL in the withdrawn sample is between 48 and 52? c. If a 10 mL sample is withdrawn, what is the probability that the average number per mL of particles in the withdrawn sample is between 48 and 52? d. How large a sample must be withdrawn so that the average number of particles per mL in the sample is between 48 and 52 with probability 95%? 16. A battery manufacturer claims that the lifetime of a certain type of battery has a population mean of 40 hours and a standard deviation of 5 hours. Let represent the mean lifetime of the batteries in a simple random sample of size 100. a. If the claim is true, what is ? b. Based on the answer to part (a), if the claim is true, is a sample mean lifetime of 36.7 hours unusually short? c. If the sample mean lifetime of the 100 batteries were 36.7 hours, would you find the manufacturer's claim to be plausible? Explain. d. If the claim is true, what is ? e. Based on the answer to part (d), if the claim is true, is a sample mean lifetime of 39.8 hours unusually short? f. If the sample mean lifetime of the 100 batteries were 39.8 hours, would you find the manufacturer's claim to be plausible? Explain. 17. A new process has been designed to make ceramic tiles. The goal is to have no more than 5% of the tiles be nonconforming due to surface defects. A random sample of 1000 tiles is inspected. Let X be the number of nonconforming tiles in the sample. a. If 5% of the tiles produced are nonconforming, what is P(X ≥ 75)? b. Based on the answer to part (a), if 5% of the tiles are nonconforming, is 75 nonconforming tiles out of 1000 an unusually large number? c. If 75 of the sample tiles were nonconforming, would it be plausible that the goal had been reached? Explain. d. If 5% of the tiles produced are nonconforming, what is P(X ≥ 53)? e. Based on the answer to part (d), if 5% of the tiles are nonconforming, is 53 nonconforming tiles out of 1000 an unusually large number? f. If 53 of the sample tiles were nonconforming, would it be plausible that the goal had been reached? Explain. 18. The manufacture of a certain part requires two different machine operations. The time on machine 1 has mean 0.5 hours and standard deviation 0.4 hours. The time on machine 2 has mean 0.6 hours and standard deviation 0.5 hours. The times needed on the machines are independent. Suppose that 100 parts are manufactured. a. What is the probability that the total time used by machine 1 is greater than 55 hours? b. What is the probability that the total time used by machine 2 is less than 55 hours? c. What is the probability that the total time used by both machines together is greater than 115 hours? Page 303 d. What is the probability that the total time used by machine 1 is greater than the total time used by machine 2? 19. Seventy percent of rivets from vendor A meet a certain strength specification, and 80% of rivets from vendor B meet the same specification. If 500 rivets are purchased from each vendor, what is the probability that more than 775 of the rivets meet the specifications? 20. Radiocarbon dating: Carbon-14 is a radioactive isotope of carbon that decays by emitting a beta particle. In the earth's atmosphere, approximately one carbon atom in 1012 is carbon14. Living organisms exchange carbon with the atmosphere, so this same ratio holds for living tissue. After an organism dies, it stops exchanging carbon with its environment, and its carbon-14 ratio decreases exponentially with time. The rate at which beta particles are emitted from a given mass of carbon is proportional to the carbon-14 ratio, so this rate decreases exponentially with time as well. By measuring the rate of beta emissions in a sample of tissue, the time since the death of the organism can be estimated. Specifically, it is known that t years after death, the number of beta particle emissions occurring in any given time interval from 1 g of carbon follows a Poisson distribution with rate events per minute. The number of years t since the death of an organism can therefore be expressed in terms of λ: d. An archaeologist finds a small piece of charcoal from an ancient campsite. The charcoal contains 1 g of carbon. a. Unknown to the archaeologist, the charcoal is 11,000 years old. What is the true value of the emission rate λ? b. The archaeologist plans to count the number X of emissions in a 25 minute interval. Find the mean and standard deviation of X. c. The archaeologist then plans to estimate λ with . What is the mean and standard deviation of ? d. What value for would result in an age estimate of 10,000 years? e. What value for would result in an age estimate of 12,000 years? f. What is the probability that the age estimate is correct to within ±1000 years? . What is the mean and 4.12 Simulation When fraternal (nonidentical) twins are born, they may be both boys, both girls, or one of each. Assume that each twin is equally likely to be a boy or a girl, and assume that the sexes of the twins are determined independently. What is the probability that both twins are boys? This probability is easy to compute, using the multiplication rule for independent events. The answer is (0.5)(0.5) = 0.25. But let's say that you did not know the multiplication rule. Is there another way that you could estimate this probability? You could do a scientific experiment, or study. You could obtain records of twin births from hospitals, and count the number in which both were boys. If you obtained a large enough number of records, the proportion in which both twins were boys would likely be close to 0.25, and you would have a good estimate of the probability. Here is an easier way. There are two equally likely outcomes for the birth of a twin: boy and girl. There are also two equally likely outcomes for the toss of a coin: heads and tails. Therefore the number of heads in the toss of two coins has the same distribution as the number of boys in a twin birth (both are binomial with n = 2 trials and success probability p = 0.5). Rather than go to the trouble of monitoring actual births, you could toss two coins a large number of times. The proportion of tosses in which both coins landed heads could be used to estimate the Page 304 proportion of births in which both twins are boys. Estimating the probability that twins are both boys by estimating the probability that two coins both land heads is an example of a simulation experiment. If the sides of the coin are labeled “0” and “1,” then the toss of a coin is an example of a random number generator. A random number generator is a procedure that produces a value that has the same statistical properties as a random quantity sampled from some specified distribution. The random number generated by the toss of a coin comes from a Bernoulli distribution with success probability p = 0.5. Nowadays, computers can generate thousands of random numbers in a fraction of a second, and virtually every statistical software package contains routines that will generate random samples from a wide variety of distributions. When a scientific experiment is too costly, or physically difficult or impossible to perform, and when the probability distribution of the data that would be generated by the experiment is approximately known, computer-generated random numbers from the appropriate distribution can be used in place of actual experimental data. Such computer-generated numbers are called simulated or synthetic data. Summary Simulation refers to the process of generating random numbers and treating them as if they were data generated by an actual scientific experiment. The data so generated are called simulated or synthetic data. Simulation methods have many uses, including estimating probabilities, estimating means and variances, verifying an assumption of normality, and estimating bias. We describe some of these methods in the rest of this section. Using Simulation to Estimate a Probability Simulation is often used to estimate probabilities that are difficult to calculate directly. Here is an example. An electrical engineer will connect two resistors, labeled 100 Ω and 25 Ω, in parallel. The actual resistances may differ from the labeled values. Denote the actual resistances of the resistors that are chosen by X and Y. The total resistance R of the assembly is given by R = XY/(X + Y). Assume that X ∼ N(100, 102) and Y ∼ N(25, 2.52) and that the resistors are chosen independently. Assume that the specification for the resistance of the assembly is 19 < R < 21. What is the probability that the assembly will meet the specification? In other words, what is P(19 < R < 21)? We will estimate this probability with a simulation. The idea is to generate simulated data whose distribution is as close as possible to the distribution of data that would be generated in an actual experiment. In an actual experiment we would take a sample of N resistors labeled 100 Ω, whose actual resistances were X1, ..., XN, and then independently take an equal size sample of resistors labeled 25 Ω, whose actual resistances were Y1, ..., YN. We would then construct N assemblies with resistances . Page 305 The values R1, ..., RN would be a random sample from the population of all possible values of the total resistance. The proportion of the sample values R1, ..., RN that fell between 19 and 21 would be an estimate of P(19 < R < 21). In an actual experiment, X1, ..., XN would be a random sample from N(100, 102) and Y1, ..., YN would be a random sample from N(25, 2.52). Therefore, in the simulated experiment, we will generate a random sample from N(100, 102) and, independently, a random sample from N(25, 2.52). We will then compute simulated total resistances . We use the notation , and to indicate that these are simulated values from a random number generator rather than actual data from a real experiment. Since the sample comes from the same distribution as would an actual sample X1, ..., XN, and since comes from the same distribution as would an actual sample Y1, ..., YN, it follows that the sample comes from the same distribution as would an actual sample of total resistances R1, ..., RN. Therefore we can treat as if it were in fact a sample of actual resistances, even though it is really a sample of random numbers generated by a computer. The results from a simulation with sample size N = 100 are given in Table 4.2 (page 306). This is a smaller sample than one would use in practice. In practice, samples of 1000, 10,000, or more are commonly used. Samples this size pose no problem for modern computers and their software, and the larger the sample, the more precise the results. TABLE 4.2 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 Simulated data for resistances in a parallel circuit X* 102.63 96.83 96.46 88.39 113.07 117.66 108.04 101.13 105.43 103.41 104.89 94.91 92.91 95.86 100.06 90.34 116.74 90.45 97.58 101.19 101.77 100.53 98.00 108.10 91.07 115.94 100.65 Y* 24.30 21.42 26.34 22.10 29.17 27.09 18.20 28.30 23.51 24.64 22.59 27.86 27.06 24.82 23.65 23.75 24.38 25.30 23.05 23.60 31.42 24.93 27.57 27.51 23.38 24.93 28.36 Page 306 R* 19.65 17.54 20.69 17.68 23.19 22.02 15.58 22.11 19.22 19.90 18.58 21.54 20.96 19.71 19.13 18.81 20.17 19.77 18.65 19.14 24.01 19.98 21.52 21.93 18.60 20.52 22.13 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 105.97 92.69 97.48 110.45 89.92 103.78 95.53 113.32 90.82 102.88 93.59 89.19 95.04 109.72 107.35 89.59 101.72 101.24 109.67 100.74 98.44 101.05 88.13 113.94 97.42 109.05 100.65 105.64 78.82 112.31 100.14 88.78 23.59 23.88 25.43 20.70 27.23 25.67 25.55 22.54 23.79 25.99 23.04 27.05 23.76 30.25 27.01 18.55 24.65 25.92 26.61 26.18 23.63 28.81 28.43 29.45 23.78 23.04 26.63 21.57 23.25 22.77 24.95 25.87 19.30 18.99 20.17 17.44 20.90 20.58 20.16 18.80 18.85 20.75 18.49 20.76 19.01 23.71 21.58 15.37 19.84 20.64 21.41 20.78 19.06 22.42 21.49 23.40 19.11 19.02 21.06 17.91 17.95 18.93 19.97 20.03 To make the calculations more transparent, we arrange the 100 values of Ri* found in Table 4.2 in increasing order: 15.3715.48 15.58 16.66 16.94 17.1817.44 17.5417.68 17.69 17.9117.95 18.01 18.06 18.21 18.3118.49 18.5818.60 18.65 18.7118.80 18.81 18.85 18.91 18.9218.93 18.9918.99 19.01 19.0219.03 19.0619.11 19.1319.1419.20 19.2219.24 19.30 19.4719.52 19.5619.58 19.6019.6019.65 19.7119.77 19.81 19.8419.90 19.9119.95 19.9719.9820.03 20.1420.16 20.17 20.1720.49 20.5220.54 20.5520.5520.58 20.6020.60 20.64 20.6920.75 20.7620.78 20.8120.9020.96 21.0621.13 21.24 21.4121.49 21.5221.54 21.5821.7921.84 21.8721.93 21.93 22.0222.06 22.1122.13 22.3622.4223.19 23.4023.71 24.01 To estimate P(19 < R < 21) we determine that 48 values out of the sample of 100 are in this range. We therefore estimate P(19 < R < 21) = 0.48. We note that with a larger sample we might use computer software to make this count. Note the importance of the assumption that the resistance X of the first resistor and the resistance Y of the second resistor were independent. Because of this assumption, we could simulate the experiment by generating independent samples X* and Y*. If X and Y had been dependent, we would have had to generate X* and Y* to have the same joint distribution as X and Y. (Joint distributions are discussed in Section 2.6.) Fortunately, many real problems involve independent samples. We now present another example of a probability estimated with a simulation. Page 307 Example 4.77 An engineer has to choose between two types of cooling fans to install in a computer. The lifetimes, in months, of fans of type A are exponentially distributed with mean 50 months, and the lifetimes of fans of type B are exponentially distributed with mean 30 months. Since type A fans are more expensive, the engineer decides that she will choose type A fans if the probability that a type A fan will last more than twice as long as a type B fan is greater than 0.5. Estimate this probability. Solution Let A represent the lifetime, in months, of a randomly chosen type A fan, and let B represent the lifetime, in months, of a randomly chosen type B fan. We need to compute P(A > 2B). We perform a simulation experiment, using samples of size 1000. We generated a random sample from an exponential distribution with mean 50 (λ = 0.02) and a random sample from an exponential distribution with mean 30 (λ = 0.033). We then count the number of times that . Table 4.3 presents the first 10 values, and the last value. The column labeled “A* > 2B*” contains a “1” if and a “0” if . TABLE 4.3 1 2 3 4 5 6 7 8 9 10 ⋮ 1000 Simulated data for Example 4.77 A* 25.554 66.711 61.189 9.153 98.794 14.577 65.126 13.205 20.535 62.278 ⋮ 19.705 Among the first 10 pairs B* 12.083 11.384 15.191 119.150 45.258 139.149 9.877 12.106 21.613 13.289 ⋮ 12.873 , there are 6 for which A* > 2B* 1 1 1 0 1 0 1 0 0 1 ⋮ 0 . Therefore, if we were to base our results on the first 10 values, we would estimate P(A > 2B) = 6/10 = 0.6. Of course, 10 simulated pairs are not nearly enough to compute a reliable estimate. Among the 1000 simulated pairs, there were 460 for which . We therefore estimate P(A > 2B) = 0.460. The engineer chooses type B. We note that this probability can be computed exactly with a multiple integral. The exact probability is 5/11 = 0.4545. The simulation approximation is quite good. A properly simulated sample from a given probability distribution is in fact a simple random sample from that distribution. Therefore the mean and variance of the simulated sample can be used to estimate the mean and variance of the distribution, and a probability plot may be used to determine whether the probability distribution is well approximated by a standard density function, such as the normal curve. We now present some examples. Page 308 Estimating Means and Variances Example 4.78 shows how simulated values can be used to estimate a population mean and standard deviation. Example 4.78 Use the simulated values in Table 4.2 to estimate the mean μR and standard deviation σR of the total resistance R. Solution We may treat the values as if they were a random sample of actual resistances. Therefore we estimate μR with the sample mean deviation and σR with the sample standard . The sample mean and standard deviation of and are , respectively. These are the estimates of μR and σr, respectively. Using Simulation to Determine Whether a Population Is Approximately Normal One of the most frequently arising issues in data analysis is whether a population is approximately normally distributed. When a simulated sample is available from a population, this issue can be addressed. Example 4.79 Construct a histogram of the simulated values of R* presented in Table 4.2. Construct a normal probability plot to determine whether the density of the total resistance R is approximately normal. Solution The histogram and probability plot are shown in the following figure. The histogram is approximately symmetric and has one mode. This is consistent with normality. Page 309 The normal probability plot suggests a slight departure from normality, especially in the tails. It is fair to say that the distribution appears to be approximately normal. In practice, a sample size of 1000 or more would provide a more precise histogram. A sample of 100 is adequate for the probability plot, although it is no more trouble to generate a larger sample. Example 4.80 shows how simulation can be used to determine whether a sample size is large enough for the Central Limit Theorem to hold, if the distribution from which the sample is drawn is known. Example 4.80 The article “Dermal Absorption from Pesticide Residues” (M. Reddy and A. Bunge, The Practical Applicability of Toxicokinetic Models in the Risk Assessment of Chemicals, 2002:55–79) models the amount of pesticide absorbed into the system as a lognormal random variable whose mean is proportional to the dose. Assume that for a certain dose, the amount absorbed is lognormally distributed with parameters μ = 1 and σ = 0.5. An experiment will be performed in which this dose will be applied in each of five independent trials, and the amount absorbed will be determined each time. Will the average amount absorbed be approximately normally distributed? Solution Let X1, ..., X5 be a random sample from a lognormal distribution with parameters μ = 1 and σ = 0.5. The question asked is whether the sample mean is approximately normally distributed. We will answer this question by generating 1000 simulated random samples of size 5 from this lognormal distribution, computing the sample mean of each of them, and then constructing a normal probability plot for the 1000 sample means. Table 4.4 presents the first three and last three of the samples. The first five columns in each row of Table 4.4 constitute a simple random sample from a lognormal distribution with parameters μ = 1 and σ = 0.5. The sixth column is the sample mean . The 1000 entries in the sixth column are therefore a random sample of sample means. By constructing a normal probability plot of these values, we can determine whether the sample mean is approximately normally distributed. TABLE 4.4 1 2 3 ⋮ 998 999 1000 2.3220 3.3379 2.9338 ⋮ 4.7993 3.7929 3.7680 Simulated data for Example 4.80 1.5087 2.8557 3.0364 ⋮ 3.7609 2.9527 4.5899 1.2144 1.0023 3.1488 ⋮ 1.5751 6.3663 2.8609 2.5092 3.8088 2.0380 ⋮ 3.6382 1.8057 2.1659 3.3408 2.3320 4.7030 ⋮ 2.0254 10.4450 5.0658 2.1790 2.6673 3.1720 ⋮ 3.1598 5.0725 3.6901 Following is a histogram and a normal probability plot of the 1000 values of . Page 310 The histogram shows that the distribution is skewed to the right. The probability plot confirms that the distribution is far from normal. Using Simulation in Reliability Analysis A system is made up of components, each of which has a lifetime that is random. The lifetime of the system is therefore also random. Reliability engineers often know, at least approximately, the probability distributions of the lifetimes of the components and wish to determine the probability distribution of the system. In practice, it can be very difficult to calculate the distribution of the system lifetime directly from the distributions of the component lifetimes. However, if the lifetimes of the components are independent, it can often be done easily with simulation. Following is an example. Example 4.81 A system consists of components A and B connected in parallel as shown in the following schematic illustration. The lifetime in months of component A is distributed Exp(1), and the lifetime in months of component B is distributed Exp(0.5). The system will function until both A and B fail. Estimate the mean lifetime of the system, the probability that the system functions for less than 1 month, and the 10th percentile of the system lifetimes. Solution We generate a sample of simulated lifetimes of component A from an Exp(1) distribution. Then we generate a sample of simulated lifetimes of component B from an Exp(0.5) distribution. Note that the mean lifetime for component A is 1 Page 311 month and the mean lifetime for component B is 1/0.5 = 2 months. The lifetime of the ith simulated system is . Table 4.5 presents results for the first 10 samples and for the last sample. TABLE 4.5 1 2 3 4 Simulated data for Example 4.81 A* 0.0245 0.3623 0.8858 0.1106 B* 0.5747 0.3998 1.7028 14.2252 L* 0.5747 0.3998 1.7028 14.2252 5 6 7 8 9 10 ⋮ 1000 0.1903 2.2259 0.8881 3.3471 2.5475 0.3614 ⋮ 0.3619 0.4665 1.4138 0.9120 3.2134 1.3240 0.8383 ⋮ 1.8799 The sample mean of the first 10 values of 0.4665 2.2259 0.9120 3.3471 2.5475 0.8383 ⋮ 1.8799 is 2.724. Five of them are less than 1. The 10th percentile of these 10 values is (0.3998 + 0.4665)/2 = 0.43315. So if we were to base our estimates on the first 10 samples, we would estimate the mean system lifetime to be 2.724 months, the probability that the system fails within a month to be 0.5, and the 10th percentile of system lifetimes to be 0.43315. Of course, 10 samples is not nearly enough for a reliable estimate. Based on all 1000 samples, the estimate of the mean lifetime was 2.29 months, the estimate of the probability of failure within a month was 0.278, and the 10th percentile was 0.516 months. Using Simulation to Estimate Bias Simulation can be used to estimate bias. Example 4.82 shows how. Example 4.82 If X1, ..., Xn is a random sample, then the sample standard deviation s is used to estimate the population standard deviation σ. However, s is biased for σ. If X1, ..., X6 is a simple random sample from a N(0,1) distribution, use simulation to estimate the bias in s. Estimate the standard deviation σs of s as well. Solution We will generate 1000 random samples of size 6 from N(0, 1), and for each one compute the sample standard deviation first 10 samples and for the last sample. . Table 4.6 (page 312) presents the results for the TABLE 4.6 Simulated data for Example 4.82 1 2 3 4 5 6 7 8 9 10 ⋮ 1000 −0.4326 −1.6656 0.1253 −1.7580 1.6867 1.3626 −2.2424 1.3765 −1.8045 0.3165 ⋮ 0.3274 The values 0.7160 1.5986 −2.0647 0.1575 0.3784 0.7469 −0.5719 −0.4187 0.5361 0.6007 ⋮ 0.1787 −0.6028 −0.9934 1.1889 −0.8496 0.3809 −2.1102 −1.9659 −0.5014 −0.9121 −0.5363 ⋮ 0.2006 0.8304 −0.0938 −0.4598 0.3291 0.4870 2.6734 0.1269 1.9869 1.4059 −0.2300 ⋮ −1.1602 −0.1342 0.2873 0.3694 −1.5780 0.9454 −0.5311 −0.2642 −0.0532 −1.2156 0.2626 ⋮ 1.1020 −0.3560 −1.8924 0.4906 −1.1100 −0.4602 1.1611 0.3721 −0.7086 −0.9619 0.0523 ⋮ 0.3173 0.6160 1.3206 1.1190 0.8733 0.7111 1.6629 1.0955 1.1228 1.2085 0.4092 ⋮ 0.7328 are a random sample from the population of all possible values of s that can be calculated from a normal sample of size 6. The sample mean is therefore an estimate of the population mean μs. Now the true standard deviation of the distribution from which the simulated data were generated is σ = 1, so the bias in s is μs − 1. We estimate the bias with . Page 312 The sample mean of the first 10 values of is 1.0139. Therefore if we were to base our results on the first 10 values, we would estimate the bias to be 1.0139 – 1 = 0.0139. Of course, 10 values is not enough to construct a reliable estimate. The sample mean of the 1000 values is . We estimate the bias to be 0.9601 – 1 = −0.0399. The sample standard deviation of the 1000 values is 0.3156. This is the estimate of σs. The Bootstrap In the examples discussed so far in this section, the distribution from which to generate the simulated data has been specified. In some cases, this distribution must be determined from data. Simulation methods in which the distribution to be sampled from is determined from data are called bootstrap methods. To illustrate, we present a variation on Example 4.82 in which the distribution sampled from is determined from data. Example 4.83 A sample of size 6 is taken from a normal distribution whose mean and variance are unknown. The sample values are 5.23, 1.93, 5.66, 3.28, 5.93, and 6.21. The sample mean is , and the sample standard deviation is s = 1.7137. The value of s will be used to estimate the unknown population standard deviation σ. Estimate the bias in s. Solution If we knew the population mean μ and standard deviation σ of the normal distribution from which the sample came, we could use the method of Example 4.82, simulating from a N(μ, σ) distribution. Since we don't know these values, we will estimate them with the sample values and s = 1.7137. We will proceed exactly as in Example 4.82, except that we will sample from a N(4.7067, 1.71372) distribution. Since this distribution was determined from the data, this is a bootstrap method. Page 313 We will generate 1000 random samples of size 6 from N(4.7067, 1.71372), and for each one compute the sample standard deviation . Table 4.7 presents the results for the first 10 samples and for the last sample. TABLE 4.7 1 2 3 4 5 6 7 8 9 10 ⋮ 1000 The values 2.3995 2.6197 3.0114 3.9375 5.8829 7.8915 4.2737 5.8602 5.7813 3.3690 ⋮ 2.0496 Simulated data for Example 4.83 4.8961 4.3102 5.2492 5.2217 5.3084 3.9731 5.5189 5.3280 4.9364 1.8618 ⋮ 6.3385 3.6221 3.2350 7.6990 1.9737 4.6003 5.1229 2.3314 5.5860 2.5893 2.7627 ⋮ 6.2414 6.9787 6.2619 6.0439 4.5434 2.6439 5.1749 5.1512 6.8256 3.7633 3.2837 ⋮ 5.1580 4.4311 4.4233 6.5965 3.0304 2.3589 3.5255 5.7752 7.5063 0.9065 3.9863 ⋮ 3.7213 4.5367 3.5903 3.7505 3.8632 2.3055 3.3330 4.0205 3.9393 3.8372 6.0382 ⋮ 8.4576 s* 1.5157 1.2663 1.7652 1.1415 1.6054 1.6884 1.2705 1.2400 1.7260 1.4110 ⋮ 2.2364 are a random sample from the population of all possible values of s that can be calculated from a normal sample of size 6. The sample mean is therefore an estimate of the population mean μs. Now the standard deviation of the population from which the simulated data were generated is σ* = 1.7137. We estimate the bias with . The sample mean of the first 10 values of is 1.4630. Therefore if we were to base our results on the first 10 values, we would estimate the bias to be 1.4630 – 1.7137 = −0.2507. Of course, 10 values are not enough to construct a reliable estimate. The sample mean of all the 1000 values of is 1.6188. W e estimate the bias to be 1.6188 – 1.7137 = −0.0949. Bootstrap results can sometimes be used to adjust estimates to make them more accurate. Example 4.84 shows how this can be done with the sample standard deviation. Example 4.84 In Example 4.83, a sample of size 6 was taken from an N(μ, σ2) population. The sample standard deviation s = 1.7137 is an estimate of the unknown population standard deviation σ. Use the results of the bootstrap in Example 4.83 to reduce the bias in this estimate. Solution We estimated the bias in s to be −0.0949. This means that on the average, the sample standard deviation computed from this N(μ, σ2) population will be less than the true standard Page 314 deviation σ by about −0.0949. We therefore adjust for the bias by adding 0.0949 to the estimate. The bias-corrected estimate of the population standard deviation is 1.7137 + 0.0949 = 1.81. Parametric and Nonparametric Bootstrap In Example 4.83, we knew that the sample came from a normal distribution, but we didn't know the mean and variance. We therefore used the data to estimate the parameters (i and σ This procedure is called the parametric bootstrap, because the data are used to estimate parameters. What if we hadn't known that the distribution was normal? Then we would have used the nonparametric bootstrap. In the nonparametric bootstrap, we simulate by sampling from the data itself. The nonparametric bootstrap is useful in constructing confidence intervals and in performing hypothesis tests. We will briefly describe the nonparametric bootstrap, and then present some applications in Sections 5.9 and 6.15. If we had a sample X1, ..., Xn from an unknown distribution, we would simulate samples X1i*, ..., Xni*. as follows. Imagine placing the values X1, ..., Xn in a box, and drawing out one value at random. Then replace the value and draw again. The second draw is also a draw from the sample X1, ..., Xn. Continue until n draws have been made. This is the first simulated sample, called a bootstrap sample: X*11, ..., Xn1*. Note that since the sampling is done with replacement, the bootstrap sample will probably contain some of the original sample items more than once, and others not at all. Now draw more bootstrap samples; as many as one would draw in any simulation, perhaps 1000 or more. Then proceed just as in any other simulation. For more information about the bootstrap and other simulation procedures, Efron and Tibshirani (1993) is an excellent source of information. Comparison of Simulation with Propagation of Error In Example 4.78, we used simulation to approximate the mean and standard deviation of a function of random variables R = XY/(X + Y). The method of propagation of error, discussed in Section 3.4, can be used for this purpose as well (see Example 3.20). Of course, simulation can do many things that propagation of error cannot do, such as estimate probabilities and determine whether a given function of random variables is normally distributed. But if what is needed is to estimate the standard deviation of a function of random variables, it is natural to ask whether simulation or propagation of error is the better technique. The answer is that each method has its advantages and disadvantages. To make our discussion of this question concrete, let X1, ..., Xn be independent random variables, and let U = U(X1, ..., Xn) be a function. We wish to estimate σU. The first thing that needs to be said is that in many cases, both methods work well and give similar results, so it is just a matter of convenience which is used. Simulation has one advantage, which is that it does not require that the standard deviations of X1, ..., Xn be small, as propagation of error does. Propagation of error has two big advantages, however. First, it is not necessary to know the distributions of X1, ..., Xn, as it is for simulation. Second, propagation of error can Page 315 pinpoint which of the Xs contributes most to the uncertainty in U, which simulation cannot easily do. Exercises for Section 4.12 1. Vendor A supplies parts, each of which has probability 0.03 of being defective. Vendor B also supplies parts, each of which has probability 0.05 of being defective. You receive a shipment of 100 parts from each vendor. a. Let X be the number of defective parts in the shipment from vendor A and let Y be the number of defective parts in the shipment from vendor B. What are the distributions of X and Y? b. Generate simulated samples of size 1000 from the distributions of X and Y. c. Use the samples to estimate the probability that the total number of defective parts is less than 10. d. Use the samples to estimate the probability that the shipment from vendor A contains more defective parts than the shipment from vendor B. e. Construct a normal probability plot for the total number of defective parts. Is this quantity approximately normally distributed? There are two competing designs for a certain semiconductor circuit. The lifetime of the first (in hours) is exponentially distributed with λ = 10−4, and the lifetime of the second is lognormally distributed with μ = 6 and σ2 = 5.4. a. Use a simulated sample of size 1000 to estimate the probability that a circuit with the first design lasts longer than one with the second design. b. 3. Estimate the probability that a circuit with the first design lasts more than twice as long as one with the second design. Rectangular plates are manufactured whose lengths are distributed N(2.0, 0.12) and whose widths are distributed N(3.0, 0.22). Assume the lengths and widths are independent. The area of a plate is given by A = XY. a. Use a simulated sample of size 1000 to estimate the mean and variance of A. b. Estimate the probability that P(5.9 < A < 6.1). c. 4. Construct a normal probability plot for the areas. Is the area of a plate approximately normally distributed? A cable is made up of four wires. The breaking strength of each wire is a normally distributed random variable with mean 10 kN and standard deviation 1 kN. The strength of the cable, using the brittle wire method, is estimated to be the strength of the weakest wire multiplied by the number of wires. a. Use simulated samples of size 1000 to estimate the mean strength of this type of cable. b. Estimate the median cable strength. c. Estimate the standard deviation of the cable strength. d. 5. 6. To be acceptable for a certain application, the probability that the cable breaks under a load of 28 kN must be less than 0.01. Does the cable appear to be acceptable? Explain. The lifetime of a laser (in hours) is lognormally distributed with μ = 8 and σ2 = 2.4. Two such lasers are operating independently. a. Use a simulated sample of size 1000 to estimate the probability that the sum of the two lifetimes is greater than 20,000 hours. b. Estimate the probability that both lasers last more than 3000 hours. c. Estimate the probability that both lasers fail before 10,000 hours. This exercise demonstrates that the value of π can be estimated with a simulation. In the figure, a circle with area equal to π/4 is inscribed in a square whose area is equal to 1. One hundred points have been randomly chosen from within the square. The probability that a point is inside the circle is equal to the fraction of the area of the square that is taken up by the circle, which is π/4. We can therefore estimate the value of π/4 by counting the number of points inside the circle, which is 79, and dividing by the total number of points, which is 100, to obtain the estimate π/4 ≈ 0.79. From this we conclude that π ≈ 4(0.79) = 3.16. This exercise presents a simulation experiment that is designed to estimate the value of π by generating 1000 points in the unit square. Page 316 a. Generate 1000 x coordinates value 0 and maximum value 1. b. Generate 1000 y coordinates , again using the uniform distribution with minimum value 0 and maximum value 1. c. Each point ( ) is inside the circle if its distance from the center (0.5, 0.5) is less than 0.5. For each pair ( ), determine whether its distance from the center is less than 0.5. This can be done by computing the value , which is the squared distance, and determining whether it is less than 0.25. . Use the uniform distribution with minimum How many of the points are inside the circle? What is your estimate of π?(Note: With only 1000 points, it is not unlikely for your estimate to be off by as much as 0.05 or more. A simulation with 10,000 or 100,000 points is much more likely to provide an estimate that is very close to the true value.) Application to mobile computer networks. Computer scientists often model the movement of a mobile computer as a random path within a rectangle. That is, two points are chosen at random within the rectangle, and the computer moves on a straight line from the first point to the second. In the study of mobile computer networks, it is important to know the mean length of a path (see the article "Stationary Distributions for Random Waypoint Models," W. Navidi and T. Camp, IEEE Transactions on Mobile Computing, 2004:99–108). It is very difficult to compute this mean directly, but it is easy to estimate it with a simulation. If the endpoints of a path are denoted (X1, Y1), and (X2, Y2), then the length of the path is . The mean length is estimated by generating endpoints ( ), and ( ) for many paths, computing the length of each, and taking the mean. This exercise presents a simulation experiment that is designed to estimate the mean distance between two points randomly chosen within a square of side 1. 7. a. Generate 1000 sets of endpoints ( ), and ( ). Use the uniform distribution with minimum value 0 and maximum value 1 for each coordinate of each point. The uniform distribution generates values that are equally likely to come from any part of the interval (0, 1). 8. 9. b. Compute the 1000 path lengths . c. Compute the sample mean path length. The true mean, to six significant digits, is 0.521405. How close did you come? d. Estimate the probability that a path is more than 1 unit long. Refer to Example 4.81 (page 310). In order to increase the lifetime of the system, the engineers have a choice between replacing component A with one whose lifetime is distributed Exp(1/2), or replacing component B with one whose lifetime is distributed Exp(1/3). a. Generate, by simulation, a large number (at least 1000) of system lifetimes under the assumption that component A is replaced. b. Generate, by simulation, a large number (at least 1000) of system lifetimes under the assumption that component B is replaced. c. If the goal is to maximize the mean system lifetime, which is the better choice? Explain. d. If the goal is to minimize the probability that the system fails within a month, which is the better choice? Explain. e. If the goal is to maximize the 10th percentile of the system lifetimes, which is the better choice? Explain. Page 317 A system consists of components A and B connected in series, as shown in the following schematic illustration. The lifetime in months of component A is lognormally distributed with μ = 1and σ = 0.5, and the lifetime in months of component B is lognormally distributed with μ = 2 and σ = 1. The system will function only so long as A and B both function. a. Generate, by simulation, a large number (at least 1000) of system lifetimes. b. Estimate the mean system lifetime. c. Estimate the probability that the system fails within 2 months. d. Estimate the 20th percentile of system lifetimes. e. Construct a normal probability plot of system lifetimes. Is the system lifetime approximately normally distributed? f. Construct a histogram of the system lifetimes. Is it skewed to the left, skewed to the right, or approximately symmetric? 10. A system consists of two subsystems connected in series, as shown in the following schematic illustration. Each subsystem consists of two components connected in parallel. The AB subsystem fails when both A and B have failed. The CD subsystem fails when both C and D have failed. The system fails as soon as one of the two subsystems fails. Assume that the lifetimes of the components, in months, have the following distributions: A: Exp(1), B: Exp(0.1), C: Exp(0.2), D: Exp(0.2). a. Generate, by simulation, a large number (at least 1000) of system lifetimes. b. Estimate the mean system lifetime. c. Estimate the median system lifetime. d. Estimate the probability that the system functions for more than 6 months. e. Estimate the 90th percentile of system lifetimes. f. Estimate the probability that the AB subsystem fails before the CD subsystem does. 11. (Continues Exercise 20 in Section 4.11.) The age of an ancient piece of organic matter can be estimated from the rate at which it emits beta particles as a result of carbon-14 decay. For example, if X is the number of particles emitted in 10 minutes by a 10,000-year-old bone fragment that contains 1 g of carbon, then X has a Poisson distribution with mean λ = 45.62. An archaeologist has found a small bone fragment that contains exactly 1 g of carbon. If t is the unknown age of the bone, in years, the archaeologist will count the number X of particles emitted in 10 minutes and compute an estimated age with the formula Unknown to the archaeologist, the bone is exactly 10,000 years old, so X has a Poisson distribution with λ = 45.62. a. Generate a simulated sample of 1000 values of X, and their corresponding values of t̂. b. Estimate the mean of . c. Estimate the standard deviation of . d. Estimate the probability that will be within 1000 years of the actual age of 10,000 e. Estimate the probability that will be more than 2000 years from the actual age of f. Construct a normal probability plot for . Is approximately normally distributed? 12. A random sample will be drawn from a normal distribution, for the purpose of estimating the population mean μ. Since μ is the median as well as the mean, it seems that both the sample median and the sample mean are reasonable estimators. This exercise is designed to determine which of these estimators has the smaller uncertainty. a. Generate a large number (at least 1000) of samples of size 5 from a N(0, 1) distribution. b. Compute the sample medians for the 1000 samples. c. Compute the mean and the standard deviation sm* of for the 1000 samples. . e. Compute the mean and standard deviation of Page 318 f. The true value of μ is 0. Estimate the bias and uncertainty (μm) in m. (Note: In fact, the median is unbiased, so your bias estimate should be close to 0.) g. Estimate the bias and uncertainty ( ) in . Is your bias estimate close to 0? Explain why it should be. Is your uncertainty estimate close to ? Explain why it should be. 13. A random sample of size 8 is taken from a Exp(λ) distribution, where λ is unknown. The sample values are 2.74, 6.41, 4.96, 1.65, 6.38, 0.19, 0.52, and 8.38. This exercise shows how to use the bootstrap to estimate the bias and uncertainty ( ) in . a. Compute for the given sample. b. Generate 1000 bootstrap samples of size 8 from an c. Compute the values of . Compute the sample mean e. Estimate the bias and uncertainty ( ) in . distribution. for each of the 1000 bootstrap samples. and the sample standard deviation of . Supplementary Exercises for Chapter 4 1. 2. An airplane has 100 seats for passengers. Assume that the probability that a person holding a ticket appears for the flight is 0.90. If the airline sells 105 tickets, what is the probability that everyone who appears for the flight will get a seat? The number of large cracks in a length of pavement along a certain street has a Poisson distribution with a mean of 1 crack per 100 m. a. What is the probability that there will be exactly 8 cracks in a 500 m length of pavement? b. What is the probability that there will be no cracks in a 100 m length of pavement? c. Let T be the distance in meters between two successive cracks. What is the probability density function of T? d. What is the probability that the distance between two successive cracks will be more than 50 m? Pea plants contain two genes for seed color, each of which may be Y (for yellow seeds) or G (for green seeds). Plants that contain one of each type of gene are called heterozygous. According to the Mendelian theory of genetics, if two heterozygous plants are crossed, each of their offspring will have probability 0.75 of having yellow seeds and probability 0.25 of having green seeds. a. Out of 10 offspring of heterozygous plants, what is the probability that exactly 3 have green seeds? b. Out of 10 offspring of heterozygous plants, what is the probability that more than 2 have green seeds? c. Out of 100 offspring of heterozygous plants, what is the probability that more than 30 have green seeds? d. Out of 100 offspring of heterozygous plants, what is the probability that between 30 and 35 inclusive have green seeds? e. Out of 100 offspring of heterozygous plants, what is the probability that fewer than 80 have yellow seeds? 3. 4. 5. A simple random sample X1, ..., Xn is drawn from a population, and the quantities ln X1, ..., ln Xn are plotted on a normal probability plot. The points approximately follow a straight line. True or false: a. X1, ..., Xn come from a population that is approximately lognormal. b. X1, ..., Xn come from a population that is approximately normal. c. ln X1, ..., ln Xn come from a population that is approximately lognormal. d. ln X1, ..., ln Xn come from a population that is approximately normal. The Environmental Protection Agency (EPA) has contracted with your company for equipment to monitor water quality for several lakes in your water district. A total of 10 devices will be used. Assume that each device has a probability of 0.01 of failure during the course of the monitoring period. a. What is the probability that none of the devices fail? b. What is the probability that two or more devices fail? Page 319 c. 6. 7. If the EPA requires the probability that none of the devices fail to be at least 0.95, what is the largest individual failure probability allowable? In the article "Occurrence and Distribution of Ammonium in Iowa Groundwater" (K. Schilling, Water Environment Research, 2002:177–186), ammonium concentrations (in mg/L) were measured at a large number of wells in the state of Iowa. The mean concentration was 0.71, the median was 0.22, and the standard deviation was 1.09. Is it possible to determine whether these concentrations are approximately normally distributed? If so, say whether they are normally distributed, and explain how you know. If not, describe the additional information you would need to determine whether they are normally distributed. Medication used to treat a certain condition is administered by syringe. The target dose in a particular application is μ. Because of the variations in the syringe, in reading the scale, and in mixing the fluid suspension, the actual dose administered is normally distributed with mean μ and variance σ2. a. What is the probability that the dose administered differs from the mean μ by less than σ? b. If X represents the dose administered, find the value of z so that P(X < μ + zσ) = 0.90. If the mean dose is 10 mg, the variance is 2.6 mg2, and a clinical overdose is defined as a dose larger than 15 mg, what is the probability that a patient will receive an overdose? 8. A certain type of plywood consists of five layers. The thicknesses of the layers are c. independent and normally distributed with mean 5 mm and standard deviation 0.2 mm. 9. a. Find the mean thickness of the plywood. b. Find the standard deviation of the thickness of the plywood. c. Find the probability that the plywood is less than 24 mm thick. The intake valve clearances on new engines of a certain type are normally distributed with mean 200 μm and standard deviation 10 μm. a. What is the probability that the clearance is greater than 215μm? b. What is the probability that the clearance is between 180 and 205 μm? c. An engine has six intake valves. What is the probability that exactly two of them have clearances greater than 215 μm? 10 The stiffness of a certain type of steel beam used in building construction has mean 30 kN/mm and standard deviation 2 kN/mm. a. Is it possible to compute the probability that the stiffness of a randomly chosen beam is greater than 32 kN/mm? If so, compute the probability. If not, explain why not. b. In a sample of 100 beams, is it possible to compute the probability that the sample mean stiffness of the beams is greater than 30.2 kN/mm? If so, compute the probability. If not, explain why not. 11. In a certain process, the probability of producing a defective component is 0.07. a. In a sample of 250 randomly chosen components, what is the probability that fewer than 20 of them are defective? b. In a sample of 10 randomly chosen components, what is the probability that one or more of them is defective? c. To what value must the probability of a defective component be reduced so that only 1% of lots of 250 components contain 20 or more that are defective? 12. A process that polishes a mirrored surface leaves an average of 2 small flaws per 5 m2 of surface. The number of flaws on an area of surface follows a Poisson distribution. a. What is the probability that a surface with area 3 m × 5 m will contain more than 5 flaws? b. What is the probability that a surface with area 2 m × 3 m will contain no flaws? c. What is the probability that 50 surfaces, each with dimensions 3 m × 6 m, will contain more than 350 flaws in total? 13. Yeast cells are suspended in a liquid medium. A 2 mL sample of the suspension is withdrawn. A total of 56 yeast cells are counted. a. Estimate the concentration of yeast cells per mL of suspension, and find the uncertainty. b. What volume of suspension must be withdrawn to reduce the uncertainty to 1 cell per mL? Page 320 14. A plate is attached to its base by 10 bolts. Each bolt is inspected before installation, and the probability of passing the inspection is 0.9. Only bolts that pass the inspection are installed. Let X denote the number of bolts that are inspected in order to attach one plate. a. Find P(X = 12). b. Find μX. c. Find σX. 15. Thicknesses of shims are normally distributed with mean 1.5 mm and standard deviation 0.2 mm. Three shims are stacked, one atop another. a. Find the probability that the stack is more than 5mmthick. b. Find the 80th percentile of the stack thickness. c. What is the minimum number of shims to be stacked so that the probability that the stack is more than 5 mm thick is at least 0.99? 16. The lifetime of a microprocessor is exponentially distributed with mean 3000 hours. a. What proportion of microprocessors will fail within 300 hours? b. What proportion of microprocessors will function for more than 6000 hours? c. A new microprocessor is installed alongside one that has been functioning for 1000 hours. Assume the two microprocessors function independently. What is the probability that the new one fails before the old one? 17. The lifetime of a bearing (in years) follows the Weibull distribution with parameters α = 1.5 and β = 0.8. a. What is the probability that a bearing lasts more than1year? b. What is the probability that a bearing lasts less than 2 years? 18. The length of time to perform an oil change at a certain shop is normally distributed with mean 29.5 minutes and standard deviation 3 minutes. What is the probability that a mechanic can complete 16 oil changes in an eight-hour day? 19. A cereal manufacturer claims that the gross weight (including packaging) of a box of cereal labeled as weighing 12 oz has a mean of 12.2 oz and a standard deviation of 0.1 oz. You gather 75 boxes and weigh them all together. Let S denote the total weight of the 75 boxes of cereal. a. If the claim is true, what is P(S ≤ 914.8)? b. Based on the answer to part (a), if the claim is true, is 914.8 oz an unusually small total weight for a sample of 75 boxes? c. If the total weight of the boxes were 914.8 oz, would you be convinced that the claim was false? Explain. d. If the claim is true, what is P(S ≤ 910.3)? e. Based on the answer to part (d), if the claim is true, is 910.3 oz an unusually small total weight for a sample of 75 boxes? f. If the total weight of the boxes were 910.3 oz, would you be convinced that the claim was false? Explain 20. Someone claims that the number of hits on his website has a Poisson distribution with mean 20 per hour. Let X be the number of hits in five hours. a. If the claim is true, what is P(X ≤ 95)? b. Based on the answer to part (a), if the claim is true, is 95 hits in a five-hour time period an unusually small number? c. If you observed 95 hits in a five-hour time period, would this be convincing evidence that the claim is false? Explain. d. If the claim is true, what is P(X ≤ 65)? e. Based on the answer to part (d), if the claim is true, is 65 hits in a five-hour time period an unusually small number? f. If you observed 65 hits in a five-hour time period, would this be convincing evidence that the claim is false? Explain. 21. A distribution sometimes used to model the largest item in a sample is the extreme value distribution. This distribution has cumulative distribution function Let X be a random variable with this distribution. a. Find P(X ≤ 0). b. Find P(X > ln 2). c. Find the median of X. Page 321 22. An alternative to the lognormal distribution for modeling highly skewed populations is the Pareto distribution with parameters θ and r. The probability density function is The parameters θ and r may be any positive numbers. Let X be a random variable with this distribution. a. Find the cumulative distribution function of X. b. Assume r > 1. Find μX. c. Assume r > 2. Find d. Show that if r ≤ 1, μX does not exist. e. Show that if r ≤ 2, does not exist. 23. A distribution that has been used to model tolerance levels in bioassays is the logistic distribution with parameters α and β. The cumulative distribution function of the logistic distribution is The parameter α may be any real number; the parameter β may be any positive number. Let X be a random variable with this distribution. a. Find the probability density function fX(x). b. Show that fX(x) is symmetric around α, that is, c. Explain why the symmetry described in part (b) shows that μX = α. You may assume that μX exists. forall x. 24. Radioactive mass 1 emits particles at a mean rate of λ1 per second, and radioactive mass 2 emits particles at a mean rate of λ2 per second. Mass 1 is selected with probability p, and mass 2 is selected with probability 1 − p. Let X be the time at which the first particle is emitted. It can be shown that X has a mixed exponential distribution with probability density function a. Find μX. b. Find the cumulative distribution function of X. c. Let λ1 = 2, λ2 = 1, and p = 0.5. Find P(X ≤ 2). d. Let λ1 = 2, λ2 = 1, and p = 0.5. Given that P(X ≤ 2), find the probability that mass 1 was selected. 25. Let X ∼ Geom(p). Let s ≥ 0 be an integer. a. Show that P(X > s) = (1 − p)s. (Hint: The probability that more than s trials are needed to obtain the first success is equal to the probability that the first s trials are all failures.) b. Let t ≥ 0 be an integer. Show that P(X > s + t | X > s) = P(X > t). This is the memoryless property .[Hint: P(X > s + t and X > s) = P(X > s + t).] c. A penny and a nickel are both fair coins. The penny is tossed three times and comes up tails each time. Now both coins will be tossed twice each, so that the penny will be tossed a total of five times and the nickel will be tossed twice. Use the memoryless property to compute the conditional probability that all five tosses of the penny will be tails, given that the first three tosses were tails. Then compute the probability that both tosses of the nickel will be tails. Are both probabilities the same? 26. A stick of length 1 is broken at a point chosen uniformly along its length. One piece is used as the length of a rectangle, and the other is used as the width. Find the mean area of a rectangle formed in this way. 27. Let X represent the lifetime of a component, in weeks. Let Y represent the lifetime of the component in days, so Y = 7X. Suppose X ∼ Exp(λ). a. Let FY be the cumulative distribution function of Y and let FX be the cumulative distribution function of X. Show that FY(y) = 1 e−−λy/7. [Hint: FY(y) = P(Y ≤ y) = P(7X ≤ y) = P(X ≤ y/7).] b. Show that Y ∼ Exp(λ/7). [Hint: Find the probability density function of Y by differentiating FY(y).] 28. Let X ∼ Bin(n, p). a. Show that if x is an integer between 1 and n inclusive, then Show that if X ∼ Bin(n, p), the most probable value for X is the greatest integer less than or equal to np + p. [Hint: Use part (a) to show that P(X = x) ≥ P(X = Page 322 1) if and only if x ≤ np + p.] 29. Let X ∼ Poisson(λ). a. b. Show that if x is a positive integer, then Show that if X ∼ Poisson(λ), the most probable value for X is the greatest integer less than or equal to λ. [Hint: Use part (a) to show that P(X = x) ≥ P(X = x − 1) if and only if x ≤ λ.] 30. Let Z ∼ N(0, 1), and let X = σZ + μ where μ and σ > 0 are constants. Let Φ represent the cumulative distribution function of Z, and let φ represent the probability density function, so a. Show that the cumulative distribution function of X is b. Differentiate FX(x) to show that X ∼ N(μ, σ2). c. Now let X = −σZ + μ. Compute the cumulative distribution function of X in terms of Φ, then differentiate it to show that X ∼ N(μ, σ2). . Page 323 Chapter 5 Confidence Intervals Introduction In Chapter 4 we discussed estimates for various parameters: for example, the sample mean as an estimate of a population mean, and the sample proportion p̂ as an estimate of a success probability p. These estimates are called point estimates, because they are single numbers, or points. An important thing to remember about point estimates is that they are almost never exactly equal to the true values they are estimating. They are almost always off—sometimes by a little, sometimes by a lot. In order for a point estimate to be useful, it is necessary to describe just how far off the true value it is likely to be. One way to do this is by reporting an estimate of the standard deviation, or uncertainty, in the point estimate. In this chapter, we will show that when the estimate comes from a normal distribution, we can obtain more information about its precision by computing a confidence interval. The following example presents the basic idea. Assume that a large number of independent measurements, all using the same procedure, are made on the diameter of a piston. The sample mean of the measurements is 14.0 cm, and the uncertainty in this quantity, which is the standard deviation of the sample mean, is 0.1 cm. Assume that the measurements are unbiased. The value 14.0 comes from a normal distribution, because it is the average of a large number of measurements. Now the true diameter of the piston will certainly not be exactly equal to the sample mean of 14.0 cm. However, because the sample mean comes from a normal distribution, we can use its standard deviation to determine how close it is likely to be to the true diameter. For example, it is very unlikely that the sample mean will differ from the true diameter by more than three standard deviations. Therefore we have a high level of confidence that the true diameter is in the interval (13.7, 14.3). On the other hand, it is not too unlikely for the sample mean to differ from the true value by more than one standard deviation. Therefore we have a lower level of confidence that the true diameter is in the interval (13.9, 14.1). The intervals (13.7, 14.3) and (13.9, 14.1) are confidence intervals for the true diameter of the piston. We will see in this chapter how to compute a quantitative measure of the level of confidence that we may have in these intervals, and in other intervals we may construct. Specifically, the results of Section 5.1 will show us that we may be 99.7% confident Page 324 that the true diameter of the piston is in the interval (13.7, 14.3), but only 68% confident that the true value is in the interval (13.9, 14.1). 5-1 Large-Sample Confidence Intervals for a Population Mean We begin with an example. A quality-control engineer wants to estimate the mean fill weight of boxes that have been filled with cereal by a certain machine on a certain day. He draws a simple random sample of 100 boxes from the population of boxes that have been filled by that machine on that day. He computes the sample mean fill weight to be , and the sample standard deviation to be s = 0.1 oz. Since the population mean will not be exactly equal to the sample mean of 12.05, it is best to construct a confidence interval around 12.05 that is likely to cover the population mean. We can then quantify our level of confidence that the population mean is actually covered by the interval. To see how to construct a confidence interval in this example, let μ represent the unknown population mean and let σ2 represent the unknown population variance. Let X1,..., X100 be the 100 fill weights of the sample boxes. The observed value of the sample mean is . Since is the mean of a large sample, the Central Limit Theorem specifies that it comes from a normal distribution whose mean is μ and whose standard deviation is . Figure 5.1 presents a normal curve, which represents the distribution of . The middle 95% of the curve, extending a distance on either side of population mean μ, is indicated. The observed value is a single draw from this distribution. We have no way to know from what part of the curve this particular value of was drawn. Figure 5.1 presents one possibility, which is that the sample mean lies within the middle 95% of the distribution. Ninety-five percent of all the samples that could have been drawn fall into this category. The horizontal line below the curve in Figure 5.1 is an interval around that is exactly the same length as the middle 95% of the distribution, namely, the interval . This interval is a 95% confidence interval for the population mean μ. It is clear that this interval covers the population mean μ. FIGURE 5.1 deviation . The sample mean is drawn from a normal distribution with mean μ and standard . For this particular sample, comes from the middle 95% of the distribution, so the 95% confidence interval succeeds in covering the population mean μ. In contrast, Figure 5.2 represents a sample whose mean lies outside the middle 95% of the curve. Only 5% of all the samples that could have been drawn fall into this category. For these more unusual samples the 95% confidence interval fails to cover the population mean μ. FIGURE 5.2 deviation The sample mean is drawn from a normal distribution with mean μ and standard . For this particular sample, comes from the outer 5% of the distribution, so the 95% confidence interval fails to cover the population mean μ. We will now compute a 95% confidence interval for the mean fill weight. The population standard deviation σ and thus for the mean fill weight. The are unknown. However, in this example, since the sample size is large, we may approximate σ with the sample standard deviation s = 0.1. We therefore compute a 95% confidence interval for the mean fill weight μ to be 12.05 ± (1.96) (0.01), or (12.0304, 12.0696). We can say that we are 95% confident, or confident at the 95% level, that the population mean fill weight lies between 12.0304 and 12.0696. Does this 95% confidence interval actually cover the population mean μ? It depends on whether this particular sample happened to be one whose mean came from the middle 95% of the distribution, or whether it was a sample whose mean was unusually large or small, in the outer 5% of the distribution. There is no way to know for sure into which category this particular sample falls. But imagine that the engineer were to repeat this procedure every day, drawing a large sample and computing the 95% confidence interval . In the long run, 95% of the samples he draws will have means in the middle 95% of the distribution, so 95% of the confidence intervals he computes will cover the population mean. To put it another way, a 95% confidence interval is computed by a procedure that succeeds in covering the population mean 95% of the time. Page 326 We can use this same reasoning to compute intervals with various confidence levels. For example, we can construct a 68% confidence interval as follows. We know that the middle 68% of the normal curve corresponds to the interval extending a distance on either side of the population mean μ. It follows that an interval of the same length around specifically , , will cover the population mean for 68% of the samples that could possibly be drawn. Therefore a 68% confidence interval for the mean fill weight of the boxes is 12.05 ± (1.0)(0.01), or (12.04, 12.06). Note that the 95% confidence interval is wider than the 68% confidence interval. This is intuitively plausible. In order to increase our confidence that we have covered the true population mean, the interval must be made wider, to provide a wider margin for error. To take two extreme cases, we have 100% confidence that the true population mean is in the infinitely wide interval (−∞, ∞), and 0% confidence that the true population mean is in the zero-width interval [12.05, 12.05] that contains the sample mean and no other point. We now illustrate how to find a confidence interval with any desired level of confidence. Specifically, let α be a number between 0 and 1, and let 100(1 − α)% denote the required confidence level. Figure 5.3 presents a normal curve representing the distribution of . Define zα/2 to be the z-score that cuts off an area of α/2 in the right-hand tail. For example, the z table (Table A.2) indicates that z.025 = 1.96, since 2.5% of the area under the standard normal curve is to the right of 1.96. Similarly, the quantity −zα/2 cuts off an area of α/2 in the left-hand tail. The middle 1 − α of the area under the curve corresponds to the interval reasoning shown in Figures 5.1 and 5.2, it follows that the interval . By the will cover the population mean μ for a proportion 1 − α of all the samples that could possibly be drawn. Therefore a level 100(1 − α)% confidence interval for μ is , or . FIGURE 5.3 deviation The sample mean is drawn from a normal distribution with mean μ and standard . The quantity zα/2 is the z-score that cuts off an area of α/2 in the right-hand tail. The quantity −zα/2 is the z-score that cuts off an area of α/2 in the left-hand tail. The interval will cover the population mean μ for a proportion 1 − α of all samples that could possibly be drawn. Therefore is a level 100(1 − α)% confidence interval for μ. . We note that even for large samples, the distribution of is only approximately normal, rather than exactly normal. Therefore the levels stated for confidence intervals are approximate. When the sample size is large enough for the Central Limit Theorem to be used, the distinction between approximate and exact levels is generally ignored in practice. Page 327 Summary Let X1, ..., Xn be a large (n > 30) random sample from a population with mean μ and standard deviation σ, so that for μ is is approximately normal. Then a level 100(1 − α)% confidence interval (5.1) where . When the value of σ is unknown, it can be replaced with the sample standard deviation s. In particular, ■ ■ ■ ■ is a 68% confidence interval for μ. is a 90% confidence interval for μ. is a 95% confidence interval for μ. is a 99% confidence interval for μ. ■ is a 99.7% confidence interval for μ. Example 5.1 The sample mean and standard deviation for the fill weights of 100 boxes are Find an 85% confidence interval for the mean fill weight of the boxes. and s = 0.1. Solution To find an 85% confidence interval, set 1 − α = 0.85 to obtain α = 0.15 and α/2 = 0.075. We then look in the table for z.075, the z-score that cuts off 7.5% of the area in the right-hand tail. We find z.075 = 1.44. We approximate . So the 85% confidence interval is 12.05±(1.44)(0.01). This can be written as 12.05±0.0144, or as (12.0356, 12.0644). Example 5.2 The article "A Non-Local Approach to Model the Combined Effects of Frequency Defects and Shot-Peening on the Fatigue Strength of a Pearlitic Steel" (B. Gerin, E. Pessard, et al., Theoretical and Applied Fracture Mechanics, 2018:19–32) reports that in a sample of 70 steel connecting rods subject to fatigue testing, the average fatigue strength, in MPa, was 408.2 with a standard deviation of 72.9. Find a 95% confidence interval for the mean fatigue strength under these conditions. Page 328 Solution First let's translate the problem into statistical language. We have a simple random sample X1, ..., X70 of strengths. The sample mean and standard deviation are and s = 72.9. The population mean is unknown, and denoted by μ. The confidence interval has the form , as specified in expression (5.1). Since we want a 95% confidence interval, the confidence level 1 − α is equal to 0.95. Thus α = 0.05, and zα/2 = z.025 = 1.96. We approximate σ with s = 72.9, and obtain . Thus the 95% confidence interval is 408.2±(1.96)(8.7132). This can be written as 408.20 ± 17.08, or as (391.12, 425.28). The following computer output (from MINITAB) presents the 95% confidence interval calculated in Example 5.2. Most of the output is self-explanatory. The quantity labeled "SE Mean" is the standard deviation of the sample mean, , approximated by . ("SE Mean" stands for standard error of the mean, which is another term for the standard deviation of the sample mean.) Example 5.3 Use the data in Example 5.2 to find an 80% confidence interval. Solution To find an 80% confidence interval, set 1 − α = 0.80 to obtain α = 0.20. Then look in the table for z.10, the z-score that cuts off 10% of the area in the right hand tail. The value is z.10 = 1.28. So the 80% confidence interval is 408.20 ± (1.28)(8.7132). This can be written as 408.20 ± 11.15, or as (397.05, 419.35). We have seen how to compute a confidence interval with a given confidence level. It is also possible to compute the level of a given confidence interval. The next example illustrates the method. Example 5.4 Based on the fatigue strength data presented in Example 5.2, an engineer reported a confidence interval of (393.86, 422.54) but neglected to specify the level. What is the level of this confidence interval? Solution The confidence interval has the form . We will solve for zα/2, then consult the z table to determine the value of α. Now , s = 72.9, and n = 70. The Page 329 upper confidence limit of 422.54 therefore satisfies the equation . Therefore zα/2 = 1.646. From the z table, we determine that α/2, the area to the right of 1.646, is approximately 0.05. The level is 100(1 − α)%, or 90%. More About Confidence Levels The confidence level of an interval measures the reliability of the method used to compute the interval. A level 100(1 − α)% confidence interval is one computed by a method that in the long run will succeed in covering the population mean a proportion 1 − α of all the times that it is used. In practice, when one computes a confidence interval, one must decide what level of confidence to use. This decision involves a trade-off, because intervals with greater confidence levels are less precise. For example, a 68% confidence interval specifies the population mean to within , while a 95% confidence interval specifies it only to within , and therefore has only about half the precision of the 68% confidence interval. Figure 5.4 (page 330) illustrates the trade-off between confidence and precision. One hundred samples were drawn from a population with mean μ. Figure 5.4b presents one hundred 95% confidence intervals, each based on one of these samples. The confidence intervals are all different, because each sample has a different mean . (They also have different values of s with which to approximate σ, but this has a much smaller effect.) About 95% of these intervals cover the population mean μ. Figure 5.4a presents 68% confidence intervals based on the same samples. These intervals are more precise (narrower), but many of them fail to cover the population mean. Figure 5.4c presents 99.7% confidence intervals. These intervals are very reliable. In the long run, only 3 in 1000 of these intervals will fail to cover the population mean. However, they are less precise (wider), and thus do not convey as much information. FIGURE 5.4 (a) One hundred 68% confidence intervals for a population mean, each computed from a different sample. Although precise, they fail to cover the population mean 32% of the time. This high failure rate makes the 68% confidence interval unacceptable for practical purposes. (b) One hundred 95% confidence intervals computed from these samples. This represents a good compromise between reliability and precision for many purposes. (c) One hundred 99.7% confidence intervals computed from these samples. These intervals fail to cover the population mean only three times in 1000. They are extremely reliable, but imprecise. The level of confidence most often used in practice is 95%. For many applications, this level provides a good compromise between reliability and precision. Confidence levels below 90% are rarely used. For some quality-assurance applications, where product reliability is extremely important, intervals with very high confidence levels, such as 99.7%, are used. Probability versus Confidence In the fill weight example discussed at the beginning of this section, a 95% confidence interval for the population mean μ was computed to be (12.304, 12.696). It is tempting to say that the probability is 95% that μ is between 12.304 and 12.696. This, however, is not correct. The term probability refers to random events, which can come out differently when experiments are repeated. The numbers 12.304 and 12.696 are fixed, not random. The population mean is also fixed. The mean fill weight is either in the interval 12.304 to 12.696, or it is not. There is no randomness involved. Therefore we say that we have 95% confidence (not probability) that the population mean is in this interval. On the other hand, let's say that we are discussing a method used to compute a 95% confidence interval. The method will succeed in covering the population mean 95% of the time, and fail the other 5% of the time. In this case, whether the population mean is covered or not is a random event, because it can vary from experiment to experiment. Therefore it is Page 330 correct to say that a method for computing a 95% confidence interval has probability 95% of covering the population mean. Example 5.5 A 90% confidence interval for the mean diameter (in cm) of steel rods manufactured on a certain extrusion machine is computed to be (14.73, 14.91). True or false: The probability that the mean diameter of rods manufactured by this process is between 14.73 and 14.91 is 90%. Page 331 Solution False. A specific confidence interval is given. The mean is either in the interval or it isn't. We are 90% confident that the population mean is between 14.73 and 14.91. The term probability is inappropriate. Example 5.6 An engineer plans to compute a 90% confidence interval for the mean diameter of steel rods. She will measure the diameters of a large sample of rods, compute and s, and then compute the interval . True or false: The probability that the population mean diameter will be in this interval is 90%. Solution True. What is described here is a method for computing a confidence interval, rather than a specific numerical value. It is correct to say that a method for computing a 90% confidence interval has probability 90% of covering the population mean. Example 5.7 A team of geologists plans to measure the weights of 250 rocks. After weighing each rock a large number of times, they will compute a 95% confidence interval for its weight. Assume there is no bias in the weighing procedure. What is the probability that more than 240 of the confidence intervals will cover the true weights of the rocks? Solution Here we are discussing 250 planned implementations of a method for computing a confidence interval, not 250 specific intervals that have already been computed. Therefore it is appropriate to compute the probability that a specified number of these intervals will cover the true weights of their respective rocks. Since the weighing procedure is unbiased, the true weight of a rock is equal to the population mean of its measurements. We may think of each of the 250 confidence intervals as a Bernoulli trial, with success occurring if the confidence interval covers its population mean. Since a 95% confidence interval is computed by a process that covers the population mean 95% of the time, the success probability for each Bernoulli trial is 0.95. Let Y represent the number of confidence intervals that cover the true weight. Then Y ∼ Bin(250, 0.95) ≈ N(237.5, 11.875). The standard deviation of Y is . Using the normal curve, the probability that Y > 240 is 0.1922. See Figure 5.5. Note that the continuity correction (see Section 4.11) has been used. FIGURE 5.5 Solution to Example 5.7. Page 332 Determining the Sample Size Needed for a Confidence Interval of Specified Width In Example 5.2, a 95% confidence interval was given by 12.68 ± 1.89, or (10.79, 14.57). This interval specifies the mean to within ± 1.89. Now assume that this interval is too wide to be useful. Assume that it is desirable to produce a 95% confidence interval that specifies the mean to within ± 0.50. To do this, the sample size must be increased. We show how to calculate the sample size needed to obtain a confidence interval of any specified width. It follows from expression (5.1) that the width of a confidence interval for a population mean based on a sample of size n drawn from a population with standard deviation σ is . If the confidence level 100(1 − α)% is specified, we can look up the value zα/2. If the population standard deviation σ is also specified, we can then compute the value of n needed to produce a specified width. In Example 5.2, the confidence level is 95% and the standard deviation is estimated to be 6.83. We look up zα/2 = z.025 = 1.96. The sample size necessary to produce a 95% confidence interval with width ±0.50 is found by solving the equation for n. We obtain n = 716.83, which we round up to n = 717. Example 5.8 In the fill weight example discussed earlier in this section, the sample standard deviation of weights from 100 boxes was s = 0.1 oz. How many boxes must be sampled to obtain a 99% confidence interval of width ±0.012 oz? Solution The level is 99%, so 1 − α = 0.99. Therefore α = 0.01 and zα/2 = 2.58. The value of σ is estimated with s = 0.1. The necessary . We obtain n ≈ 463. sample size is found by solving One-Sided Confidence Intervals The confidence intervals discussed so far have been two-sided, in that they specify both a lower and an upper confidence bound. Occasionally we are interested only in one of these bounds. In these cases, one-sided confidence intervals are appropriate. For example, assume that a reliability engineer wants to estimate the mean crushing strength of a certain type of concrete block, to determine the sorts of applications for which it will be appropriate. The engineer will probably be interested only in a lower bound for the strength, since specifications for various applications will generally specify only a minimum strength. Assume that a large sample has sample mean and standard deviation . Figure 5.6 shows how the idea behind the two-sided confidence interval can be adapted to produce a onesided confidence interval for the population mean μ. The normal curve represents the distribution of . For 95% of all the samples that could be drawn, , and therefore the interval covers μ. This interval will fail to cover μ only if the sample Page 333 mean is in the upper 5% of its distribution. The interval is a 95% one-sided confidence interval for μ, and the quantity is a 95% lower confidence bound for μ. FIGURE 5.6 deviation The sample mean is drawn from a normal distribution with mean μ and standard . For this particular sample, comes from the lower 95% of the distribution, so the succeeds in covering the population mean μ. 95% one-sided confidence interval By constructing a figure like Figure 5.6 with the lower 5% tail shaded, it can be seen that the quantity is a 95% upper confidence bound for μ. We now generalize the method to produce one-sided confidence intervals of any desired level. Define zα to be the zscore that cuts off an area α in the right-hand tail of the normal curve. For example, z.05 = 1.645. By the reasoning used to obtain a 95% confidence interval, it can be seen that a level 100(1 − α)% lower confidence bound for μ is given by , and a level 1 − α upper confidence bound for μ is given by . Summary Let X1, ..., Xn be a large (n > 30) random sample from a population with mean μ and standard deviation σ, so that bound for μ is is approximately normal. Then a level 100(1 − α)% lower confidence (5.2) and a level 100(1 − α)% upper confidence bound for μ is (5.3) where . When the value of σ is unknown, it can be replaced with the sample standard deviation s. In particular, ■ Page 334 is a 90% upper confidence bound for μ. ■ is a 95% upper confidence bound for μ. ■ is a 99% upper confidence bound for μ. The corresponding lower bounds are found by replacing the "+" with "−." Example 5.9 Refer to Example 5.2. Find both a 95% lower confidence bound and a 99% upper confidence bound for the mean fatigue strength. Solution The sample mean and standard deviation are and s = 72.88, respectively. The sample size is n = 70. We estimate . The 95% lower confidence bound is , and the 99% upper confidence bound is . In Example 5.2, the 95% two-sided confidence interval was computed to be (10.79, 14.57). The 95% lower confidence bound of 11.09, computed in Example 5.9, is greater than the lower bound of the two-sided confidence interval. The reason for this is that the two-sided interval can fail in two ways—the value of μ may be too high or too low. The two-sided 95% confidence interval is designed to fail 2.5% of the time on the high side and 2.5% of the time on the low side. In contrast, the 95% lower confidence bound never fails on the high side. It is therefore designed to fail 5% of the time on the low side, so its lower limit is greater than that of the twosided interval. Confidence Intervals Must Be Based on Random Samples The methods described in this section require that the data be a random sample from a population. When used for other samples, the results may not be meaningful. Following are two examples in which the assumption of random sampling is violated. Example 5.10 A chemical engineer wishes to estimate the mean yield of a new process. The process is run 100 times over a period of several days. Figure 5.7 presents the 100 yields plotted against time. Would it be appropriate to compute a confidence interval for the mean yield by calculating and s for the yields, and then using expression (5.1)? Solution No. Expression (5.1) is valid only when the data are a random sample from a population. Figure 5.7 shows a cyclic pattern. This might indicate that the yield on each run is influenced by the yield on the previous run, which would violate the assumption of Page 335 independence. Another possibility is that the yield is influenced by ambient conditions that fluctuate in a regular fashion. In either case, the data do not satisfy the conditions of a random sample, and expression (5.1) should not be used. FIGURE 5.7 Yields from 100 runs of a chemical process, plotted against time. There is a clear pattern, indicating that the data do not constitute a random sample. Example 5.11 The engineer in Example 5.10 is studying the yield of another process. Figure 5.8 presents yields from 100 runs of this process, plotted against time. Should expression (5.1) be used to compute a confidence interval for the mean yield of this process? FIGURE 5.8 Yields from 100 runs of a chemical process, plotted against time. There is an increasing trend with time, at least in the initial part of the plot, indicating that the data do not constitute a random sample. Solution No. As in Example 5.10, there is a pattern in time. In this case, the yields tend to increase with time, at least in the initial part of the plot. This might indicate a "learning effect"; as an operator becomes more experienced in running a process, the results get better. A more thorough analysis of the data might indicate a point in time where the increase appears to stop, in which case the succeeding portion of the data might be used to form a confidence interval. Page 336 Exercises for Section 5.1 1. Find the value of zα/2 to use in expression (5.1) to construct a confidence interval with level 2. a. 95% b. 98% c. 99% d. 80% Find the levels of the confidence intervals that have the following values of zα/2: 3. a. zα/2 = 1.96 b. zα/2 = 2.17 c. zα/2 = 1.28 d. zα/2 = 3.28 As the confidence level goes up, the reliability goes _____ and the precision goes _____. Options: up, down. 4. 5. 6. 7. 8. Three confidence intervals for a population mean are constructed, all from the same random sample. The confidence levels are 95%, 98%, and 99.9%. The confidence intervals are (6.4, 12.3), (5.1, 13.6), and (6.8, 11.9). Match each confidence interval to its level. A sample of 114 patients were given a drug to lower cholesterol. A 95% confidence interval for the mean reduction in cholesterol (in mmol/L) was (0.88, 1.02). a. What was the sample mean reduction? b. What was the sample standard deviation of the reduction amounts? The article "Modeling Arterial Signal Optimization with Enhanced Cell Transmission Formulations" (Z. Li, Journal of Transportation Engineering 2011:445–454) presents a new method for timing traffic signals in heavily traveled intersections. The effectiveness of the new method was evaluated in a simulation study. In 50 simulations, the mean improvement in traffic flow in a particular intersection was 654.1 vehicles per hour, with a standard deviation of 311.7 vehicles per hour. a. Find a 95% confidence interval for the improvement in traffic flow due to the new system. b. Find a 98% confidence interval for the improvement in traffic flow due to the new system. c. A traffic engineer states that the mean improvement is between 581.6 and 726.6 vehicles per hour. With what level of confidence can this statement be made? d. Approximately what sample size is needed so that a 95% confidence interval will specify the mean to within ±50 vehicles per hour? e. Approximately what sample size is needed so that a 98% confidence interval will specify the mean to within ±50 vehicles per hour? In a sample of 100 steel canisters, the mean wall thickness was 8.1 mm with a standard deviation of 0.5 mm. a. Find a 95% confidence interval for the mean wall thickness of this type of canister. b. Find a 99% confidence interval for the mean wall thickness of this type of canister. c. An engineer claims that the mean thickness is between 8.02 and 8.18 mm. With what level of confidence can this statement be made? d. How many canisters must be sampled so that a 95% confidence interval will specify the mean wall thickness to within ±0.08 mm? e. How many canisters must be sampled so that a 99% confidence interval will specify the mean wall thickness to within ±0.08 mm? The article "Application of Surgical Navigation to Total Hip Arthroplasty" (T. Ecker and S. Murphy, Journal of Engineering in Medicine, 2007: 699–712) reports that in a sample of 123 hip surgeries of a certain type, the average surgery time was 136.9 minutes with a standard deviation of 22.6 minutes. a. Find a 95% confidence interval for the mean surgery time for this procedure. b. Find a 99.5% confidence interval for the mean surgery time for this procedure. c. A surgeon claims that the mean surgery time is between 133.9 and 139.9 minutes. With what level of confidence can this statement be made? d. Approximately how many surgeries must be sampled so that a 95% confidence interval will specify the mean to within ±3 minutes? Page 337 Approximately how many surgeries must be sampled so that a 99% confidence interval will specify the mean to within ±3 minutes? A sample of 75 concrete blocks had a mean mass of 38.3 kg with a standard deviation of 0.6 kg. a. Find a 95% confidence interval for the mean mass of this type of concrete block. b. Find a 99% confidence interval for the mean mass of this type of concrete block. c. An engineer claims that the mean mass is between 38.2 and 38.4 kg. With what level of confidence can this statement be made? d. How many blocks must be sampled so that a 95% confidence interval will specify the mean mass to within ±0.1 kg? e. How many blocks must be sampled so that a 99% confidence interval will specify the mean mass to within ±0.1 kg? Oven thermostats were tested by setting them to 350°F and measuring the actual temperature of the oven. In a sample of 67 thermostats, the average temperature was 348.2°F and the standard deviation was 5.1°F. a. Find a 90% confidence interval for the mean oven temperature. b. Find a 95% confidence interval for the mean oven temperature. c. What is the confidence level of the interval (347.5, 348.9)? d. How many thermostats must be sampled so that a 90% confidence interval specifies the mean to within ±0.8°F? e. How many thermostats must be sampled so that a 95% confidence interval specifies the mean to within ±0.8°F? In a sample of 80 light bulbs, the mean lifetime was 1217 hours with a standard deviation of 52 hours. a. Find a 95% confidence interval for the mean lifetime of this type of light bulb. b. Find a 98% confidence interval for the mean lifetime of this type of light bulb. c. An engineer claims that the mean lifetime is between 1208 and 1226 hours. With what level of confidence can this statement be made? d. How many light bulbs must be sampled so that a 95% confidence interval will specify the mean lifetime to within ±8 hours? e. How many light bulbs must be sampled so that a 98% confidence interval will specify the mean lifetime to within ±8 hours? In a sample of 60 electric motors, the average efficiency (in percent) was 85 and the standard deviation was 2. a. Find a 95% confidence interval for the mean efficiency. b. Find a 99.5% confidence interval for the mean efficiency. c. What is the confidence level of the interval (84.63, 85.37)? d. How many motors must be sampled so that a 95% confidence interval specifies the mean to within ±0.35? e. How many motors must be sampled so that a 99.5% confidence interval specifies the mean to within ±0.35? The sugar content in a one-cup serving of a certain breakfast cereal was measured for a e. 9. 10. 11. 12. 13. sample of 140 servings. The average was 11.9 g and the standard deviation was 1.1 g. a. Find a 95% confidence interval for the mean sugar content. b. Find a 99% confidence interval for the mean sugar content. c. What is the confidence level of the interval (11.81, 11.99)? d. How large a sample is needed so that a 95% confidence interval specifies the mean to within ±0.1? e. How large a sample is needed so that a 99% confidence interval specifies the mean to within ±0.1? 14. Refer to Exercise 7. a. Find a 95% lower confidence bound for the mean wall thickness. b. Someone says that the mean thickness is less than 8.2 mm. With what level of confidence can this statement be made? 15. Refer to Exercise 8. a. Find a 98% lower confidence bound for the mean time. Page 338 b. 16. 17. 18. 19. 20. 21. 22. Someone says that the mean time is greater than 134.3 minutes. With what level of confidence can this statement be made? Refer to Exercise 9. a. Find a 95% lower confidence bound for the mean mass. b. An engineer claims that the mean mass is greater than 38.25 kg. With what level of confidence can this statement be made? Refer to Exercise 10. a. Find a 99% upper confidence bound for the mean temperature. b. The claim is made that the mean temperature is less than 349.5°F. With what level of confidence can this statement be made? Refer to Exercise 11. a. Find a 90% upper confidence interval for the mean lifetime. b. Someone says that the mean lifetime is less than 1220 hours. With what level of confidence can this statement be made? Refer to Exercise 12. a. Find a 98% lower confidence bound for the mean efficiency. b. The claim is made that the mean efficiency is greater than 84.6%. With what level of confidence can this statement be made? Refer to Exercise 13. a. Find a 95% upper confidence bound for the mean sugar content. b. The claim is made that the mean sugar content is greater than 11.7 g. With what level of confidence can this statement be made? An investigator computes a 95% confidence interval for a population mean on the basis of a sample of size 70. If she wishes to compute a 95% confidence interval that is half as wide, how large a sample does she need? A 95% confidence interval for a population mean is computed from a sample of size 400. Another 95% confidence interval will be computed from a sample of size 100 drawn from the same population. Choose the best answer to fill in the blank: The interval from the sample of size 400 will be approximately _____ as the interval from the sample of size 100. i. One-eighth as wide ii. One-fourth as wide iii. One-half as wide iv. The same width v. Twice as wide vi. Four times as wide vii. Eight times as wide 23. Based on a large sample of capacitors of a certain type, a 95% confidence interval for the mean capacitance, in μF, was computed to be (0.213, 0.241). Find a 90% confidence interval for the mean capacitance of this type of capacitor. 24. Sixty-four independent measurements were made of the speed of light. They averaged 299,795 km/s and had a standard deviation of 8 km/s. True or false: a. A 95% confidence interval for the speed of light is 299,795 ± 1.96 km/s. b. The probability is 95% that the speed of light is in the interval 299,795 ± 1.96. c. 95% of the measurements were in the interval 299, 795 ± 1.96. d. If a 65th measurement is made, the probability is 95% that it would fall in the interval 299,795 ± 1.96. 25. A large box contains 10,000 ball bearings. A random sample of 120 is chosen. The sample mean diameter is 10 mm, and the standard deviation is 0.24 mm. True or false: a. A 95% confidence interval for the mean diameter of the 120 bearings in the sample is . b. A 95% confidence interval for the mean diameter of the 10,000 bearings in the box is . c. d. 95% of the ball bearings had diameters in the interval . A 95% confidence interval for the mean diameter of the 10,000 bearings in the box is Page 339 . 26. Each day a quality engineer selects a random sample of 50 power supplies from the day's production, measures their output voltages, and computes a 90% confidence interval for the mean output voltage of all the power supplies manufactured that day. What is the probability that more than 15 of the confidence intervals constructed in the next 200 days will fail to cover the true mean? 27. Based on a sample of repair records, an engineer calculates a 95% confidence interval for the mean cost to repair a fiber-optic component to be ($140, $160). A supervisor summarizes this result in a report, saying, "We are 95% confident that the mean cost of repairs is less than $160." Is the supervisor underestimating the confidence, overestimating it, or getting it right? Explain. 28. A meteorologist measures the temperature in downtown Denver at noon on each day for one year. The 365 readings average 57°F and have standard deviation 20°F. The meteorologist computes a 95% confidence interval for the mean temperature at noon to be . Is this correct? Why or why not? 5-2 Confidence Intervals for Proportions The methods of Section 5.1, in particular expression (5.1), can be used to find confidence intervals for the mean of any population from which a large sample has been drawn. When the population has a Bernoulli distribution, this expression takes on a special form. We illustrate this with an example. In Example 5.2 (in Section 5.1), a confidence interval was constructed for the mean lifetime of a certain type of microdrill when drilling a low-carbon alloy steel. Now assume that a specification has been set that a drill should have a minimum lifetime of 10 holes drilled before failure. A sample of 144 microdrills is tested, and 120, or 83.3%, meet this specification. Let p represent the proportion of microdrills in the population that will meet the specification. We wish to find a 95% confidence interval for p. We begin by constructing an estimate for p. Let X represent the number of drills in the sample that meet the specification. Then X ∼ Bin(n, p), where n = 144 is the sample size. The estimate for p is . In this example, X = 120, so = 0.833. The uncertainty, or standard deviation of , is . Since the sample size is large, it follows from the Central Limit Theorem (Equation 4.58 in Section 4.11) that The reasoning illustrated by Figures 5.1 and 5.2 (in Section 5.1) shows that for 95% of all possible samples, the population proportion p satisfies the following inequality: (5.4) At first glance, expression (5.4) looks like a 95% confidence interval for p. However, the limits contain the unknown p, and so cannot be computed. The traditional approach is to replace p with , obtaining the confidence interval . It turns out that replacing the population proportion p with the sample proportion tends to make the confidence interval too short in some cases, even for some fairly large sample sizes. Recent research, involving simulation studies, has shown that this effect can be largely Page 340 compensated for by modifying both n and p slightly. Specifically, one should add 4 to the number of trials, and 2 to the number of successes. So in place of n we use ñ = n + 4, and in place of we use . A 95% confidence interval for p is thus given by . In this example, ñ = 148 and , so the 95% confidence interval is 0.8243 ± 0.0613, or (0.763, 0.886). We justified this confidence interval on the basis of the Central Limit Theorem, which requires n to be large. However, this method of computing confidence intervals is appropriate for any sample size n. When used with small samples, it may occasionally happen that the lower limit is less than 0 or that the upper limit is greater than 1. Since 0 < p < 1, a lower limit less than 0 should be replaced with 0, and an upper limit greater than 1 should be replaced with 1. Summary Let X be the number of successes in n independent Bernoulli trials with success probability p, so that X ∼ Bin(n, p). Define ñ = n + 4, and . Then a level 100(1 − α)% confidence interval for p is (5.5) If the lower limit is less than 0, replace it with 0. If the upper limit is greater than 1, replace it with 1. The confidence interval given by expression (5.5) is sometimes called the Agresti–Coull interval, after Alan Agresti and Brent Coull, who developed it. For more information on this confidence interval, consult the article "Approximate Is Better Than 'Exact' for Interval Estimation of Binomial Proportions" (A. Agresti and B. Coull, The American Statistician, 1998:119–126). Example 5.12 The anterior cruciate ligament (ACL) runs diagonally through the middle of the knee. The article "Return to Sport After Pediatric Anterior Cruciate Ligament Reconstruction and its Effect on Subsequent Anterior Cruciate Ligament Injury" (T. Dekker, J. Godin, et al., Journal of Bone and Joint Surgery 2017:897–904) reported results for 85 young athletes who suffered ACL injuries. Of the 85 injuries, 51 were to the left knee and 34 were to the right knee. Find a 90% confidence interval for the proportion of ACL injuries that are to the left knee. Solution The number of successes is X = 51, and the number of trials is n = 85. We therefore compute ñ = 85 + 4 = 89, , and . For a 90% confidence interval, the value of α/2 is 0.05, so zα/2 = 1.645. The 90% confidence interval is therefore 0.5955 ± (1.645) (0.0520), or (0.510, 0.681). Page 341 One-sided confidence intervals can be computed for proportions as well. They are analogous to the one-sided intervals for a population mean (Equations 5.2 and 5.3 in Section 5.1). The levels for one-sided confidence intervals are only roughly approximate for small samples. Summary Let X be the number of successes in n independent Bernoulli trials with success probability p, so that X ∼ Bin(n, p). Define ñ = n + 4, and . Then a level 100(1 − α)% lower confidence bound for p is (5.6) and level 100(1 − α)% upper confidence bound for p is (5.7) If the lower bound is less than 0, replace it with 0. If the upper bound is greater than 1, replace it with 1. Example 5.13 shows how to compute an approximate sample size necessary for a confidence interval to have a specified width when a preliminary value of is known. Example 5.13 In Example 5.12, what sample size is needed to obtain a 95% confidence interval with width ±0.08? Solution , where ñ = n + 4. We therefore A 95% confidence interval has width determine the sample size n by solving the equation the data in Example 5.12, . Substituting this value for − 141. . From and solving, we obtain n Sometimes we may wish to compute a necessary sample size without having a reliable estimate available. The quantity , which determines the width of the confidence interval, is maximized for . Since the width is greatest when compute a conservative sample size estimate by substituting Example 5.13. Example is greatest, we can and proceeding as in 5.14 In Example 5.12, how large a sample is needed to guarantee that the width of the 95% confidence interval will be no greater than ±0.08, if no preliminary sample has been taken? Page 342 Solution A 95% confidence interval has width . The widest the confidence interval could be, for a sample of size n, is . Solving the equation , or for n, we obtain n ≈ 147. Note that this estimate is somewhat larger than the one obtained in Example 5.13. The Traditional Method The method we have described was developed quite recently (although it was created by simplifying a much older method). Many people still use a more traditional method. The traditional method uses the actual sample size n in place of ñ, and the actual sample proportion in place of . Although this method is still widely used, it fails to achieve its stated coverage probability even for some fairly large values of n. This means that 100(1 − α)% confidence intervals computed by the traditional method will cover the true proportion less than 100(1 − α)% of the time. The traditional method cannot be used for small samples at all; one rule of thumb regarding the sample size is that both (the number of successes) and (the number of failures) should be greater than 10. Since the traditional method is still widely used, we summarize it in the following box. For very large sample sizes, the results of the traditional method are almost identical to those of the modern method. For small or moderately large sample sizes, the modern approach is better. Summary The Traditional Method for Computing Confidence Intervals for a Proportion (widely used but not recommended) Let be the proportion of successes in a large number n of independent Bernoulli trials with success probability p. Then the traditional level 100(1 − α)% confidence interval for p is (5.8) The method cannot be used unless the sample contains at least 10 successes and 10 failures. Exercises for Section 5.2 1. In a simple random sample of 70 automobiles registered in a certain state, 28 of them were found to have emission levels that exceed a state standard. a. What proportion of the automobiles in the sample had emission levels that exceed the standard? b. Find a 95% confidence interval for the proportion of automobiles in the state whose emission levels exceed the standard. c. Find a 98% confidence interval for the proportion of automobiles whose emission levels exceed the standard. d. How many automobiles must be sampled to specify the proportion that exceed the standard to within ±0.10 with 95% confidence? e. How many automobiles must be sampled to specify the proportion that exceed the standard to within ±0.10 with 98% confidence? Page 343 f. 2. Someone claims that less than half of the automobiles in the state exceed the standard. With what level of confidence can this statement be made? During a recent drought, a water utility in a certain town sampled 100 residential water bills and found that 73 of the residences had reduced their water consumption over that of the previous year. a. Find a 95% confidence interval for the proportion of residences that reduced their water consumption. b. Find a 99% confidence interval for the proportion of residences that reduced their 3. 4. 5. water consumption. c. Find the sample size needed for a 95% confidence interval to specify the proportion to within ±0.05. d. Find the sample size needed for a 99% confidence interval to specify the proportion to within ±0.05. e. Someone claims that more than 70% of residences reduced their water consumption. With what level of confidence can this statement be made? f. If 95% confidence intervals are computed for 200 towns, what is the probability that more than 192 of the confidence intervals cover the true proportions? A soft-drink manufacturer purchases aluminum cans from an outside vendor. A random sample of 70 cans is selected from a large shipment, and each is tested for strength by applying an increasing load to the side of the can until it punctures. Of the 70 cans, 52 meet the specification for puncture resistance. a. Find a 95% confidence interval for the proportion of cans in the shipment that meet the specification. b. Find a 90% confidence interval for the proportion of cans in the shipment that meet the specification. c. Find the sample size needed for a 95% confidence interval to specify the proportion to within ±0.05. d. Find the sample size needed for a 90% confidence interval to specify the proportion to within ±0.05. e. If a 90% confidence interval is computed each day for 300 days, what is the probability that more than 280 of the confidence intervals cover the true proportions? The National Center for Health Statistics interviewed 5409 adult smokers in 2015, and 2636 of them said they had tried to quit smoking during the past year. Consider this to be a random sample. a. Find a 95% confidence interval for the proportion of smokers who have tried to quit within the past year. b. Find a 99% confidence interval for the proportion of smokers who have tried to quit within the past year. c. Someone claims that the proportion is less than 0.5. With what level of confidence can this statement be made? d. Find the sample size necessary for a 95% confidence interval to specify the proportion to within ±0.01. e. Find the sample size necessary for a 99% confidence interval to specify the proportion to within ±0.01. High levels of blood sugar are associated with an increased risk of type 2 diabetes. A level higher than normal is referred to as "impaired fasting glucose." The article "Association of Low-Moderate Arsenic Exposure and Arsenic Metabolism with Incident Diabetes and Insulin Resistance in the Strong Heart Family Study" (M. Grau-Perez, C. Kuo, et al., Environmental Health Perspectives, 2017, online) reports a study in which 47 of 155 people with impaired fasting glucose had type 2 diabetes. Consider this to be a simple random sample. a. 6. 7. 8. Find a 95% confidence interval for the proportion of people with impaired fasting glucose who have type 2 diabetes. b. Find a 99% confidence interval for the proportion of people with impaired fasting glucose who have type 2 diabetes. c. A doctor claims that less than 35% of people with impaired fasting glucose have type 2 diabetes. With what level of confidence can this claim be made? Refer to Exercise 1. Find a 95% lower confidence bound for the proportion of automobiles whose emissions exceed the standard. Refer to Exercise 2. Find a 98% upper confidence bound for the proportion of residences that reduced their water consumption. Refer to Exercise 4. Find a 99% lower confidence bound for the proportion of smokers who have tried to quit smoking in the past year. Page 344 A random sample of 400 electronic components manufactured by a certain process are tested, and 30 are found to be defective. a. Let p represent the proportion of components manufactured by this process that are defective. Find a 95% confidence interval for p. b. How many components must be sampled so that the 95% confidence interval will specify the proportion defective to within ±0.02? c. (Hard) The company ships the components in lots of 200. Lots containing more than 20 defective components may be returned. Find a 95% confidence interval for the proportion of lots that will be returned. 10. Refer to Exercise 9. A device will be manufactured in which two of the components in Exercise 9 will be connected in series. The components function independently, and the device will function only if both components function. Let q be the probability that a device functions. Find a 95% confidence interval for q.(Hint: Express q in terms of p, and then use the result of Exercise 9a.) 11. When the light turns yellow, should you stop or go through it? The article "Evaluation of Driver Behavior in Type II Dilemma Zones at High-Speed Signalized Intersections" (D. Hurwitz, M. Knodler, and B. Nyquist, Journal of Transportation Engineering, 2011:277– 286) defines the "indecision zone" as the period when a vehicle is between 2.5 and 5.5 seconds away from an intersection. It presents observations of 710 vehicles passing through various intersections in Vermont for which the light turned yellow in the indecision zone. Of the 710 vehicles, 89 ran a red light. a. Find a 90% confidence interval for the proportion of vehicles that will run the red light when encountering a yellow light in the indecision zone. b. Find a 95% confidence interval for the proportion of vehicles that will run the red light when encountering a yellow light in the indecision zone. c. Find a 99% confidence interval for the proportion of vehicles that will run the red light when encountering a yellow light in the indecision zone. 12. Sleep apnea is a disorder in which there are pauses in breathing during sleep. People with this condition must wake up frequently to breathe. The article "Postoperative Complications in Patients With Obstructive Sleep Apnea Syndrome Undergoing Hip or Knee Replacement: A Case-Control Study" (R. Gupta, J. Parvizi, et al., Mayo Clinic Proceedings, 2001:897–905) reported that in a sample of 427 people 65 and over, 104 of them had sleep 9. 13. 14. 15. 16. apnea. a. Find a 95% confidence interval for the proportion of those 65 and over who have sleep apnea. b. Find a 99% confidence interval for the proportion of those 65 and over who have sleep apnea. c. Find the sample size needed for a 95% confidence interval to specify the proportion to within ±0.03. d. Find the sample size needed for a 99% confidence interval to specify the proportion to within ±0.03. A sociologist is interested in surveying workers in computer-related jobs to estimate the proportion of such workers who have changed jobs within the past year. a. In the absence of preliminary data, how large a sample must be taken to ensure that a 95% confidence interval will specify the proportion to within ±0.05? b. In a sample of 100 workers, 20 of them had changed jobs within the past year. Find a 95% confidence interval for the proportion of workers who have changed jobs within the past year. c. Based on the data in part (b), estimate the sample size needed so that the 95% confidence interval will specify the proportion to within ±0.05. Stainless steels can be susceptible to stress corrosion cracking under certain conditions. A materials engineer is interested in determining the proportion of steel alloy failures that are due to stress corrosion cracking. a. In the absence of preliminary data, how large a sample must be taken so as to be sure that a 98% confidence interval will specify the proportion to within ±0.05? b. In a sample of 200 failures, 30 of them were caused by stress corrosion cracking. Find a 98% confidence interval for the proportion of failures caused by stress corrosion cracking. c. Based on the data in part (b), estimate the sample size needed so that the 98% confidence interval will specify the proportion to within ±0.05. The article "A Music Key Detection Method Based on Pitch Class Distribution Theory" (J. Sun, H. Li, and L. Ma, International Journal of Knowledge-based and Intelligent Engineering Systems, 2011:165–175) describes a method of analyzing digital music files to determine the key in which the music is written. In a sample of 335 classical Page 345 music selections, the key was identified correctly in 293 of them. a. Find a 90% confidence interval for the proportion of pieces for which the key will be correctly identified. b. How many music pieces should be sampled to specify the proportion to within ±0.025 with 90% confidence? c. Another method of key detection is to be tested. At this point, there is no estimate of the proportion of the time this method will be identified correctly. Find a conservative estimate of the sample size needed so that the proportion will be specified to within ±0.03 with 90% confidence. A stock market analyst notices that in a certain year, the price of IBM stock increased on 131 out of 252 trading days. Can these data be used to find a 95% confidence interval for the proportion of days that IBM stock increases? Explain. 5-3 Small-Sample Confidence Population Mean Intervals for a The methods described in Section 5.1 for computing confidence intervals for a population mean require that the sample size be large. When the sample size is small, there are no good general methods for finding confidence intervals. However, when the population is approximately normal, a probability distribution called the Student's t distribution can be used to compute confidence intervals for a population mean. In this section, we describe this distribution and show how to use it. The Student's t Distribution If is the mean of a large sample of size n from a population with mean µ and variance σ2, then the Central Limit Theorem specifies that . The quantity then has a normal distribution with mean 0 and variance 1. In addition, the sample standard deviation s will almost certainly be close to the population standard deviation σ. For this reason the quantity is approximately normal with mean 0 and variance 1, so we can look up probabilities pertaining to this quantity in the standard normal table (z table). This enables us to compute confidence intervals of various levels for the population mean µ. What can we do if is the mean of a small sample? If the sample size is small, s may not be close to σ, and may not be approximately normal. If we know nothing about the population from which the small sample was drawn, there are no easy methods for computing confidence intervals. However, if the population is approximately normal, will be approximately normal even when the sample size is small. It turns out that we can still use the quantity , but since s is not necessarily close to σ, this quantity will not have a normal distribution. Instead, it has the Student's t distribution with n − 1 degrees of freedom, which we denote tn−1. The number of degrees of freedom for the t distribution is one less than the sample size. The Student's t distribution was discovered in 1908 by William Sealy Gossett, a statistician who worked for the Guinness Brewing Company in Dublin, Ireland. The management at Guinness considered the discovery to be proprietary information, and required him to use the pseudonym "Student." This had happened before; see Section 4.3. Page 346 Summary Let X1, ..., Xn be a small (e.g., n < 30) sample from a normal population with mean µ. Then the quantity has a Student's t distribution with n − 1 degrees of freedom, denoted tn−1. When n is large, the distribution of the quantity is very close to normal, so the normal curve can be used, rather than the Student's t. The probability density function of the Student's t distribution is different for different degrees of freedom. Figure 5.9 presents plots of the probability density function for several choices of degrees of freedom. The curves all have a shape similar to that of the normal, or z, curve with mean 0 and standard deviation 1. The t curves are more spread out, however. For example, the t curve with one degree of freedom corresponds to a sample size of 2. When drawing samples of size 2, it will frequently happen that the sample standard deviation s is much smaller than σ, which makes the value of quite large (either positive or negative). For this reason, the t curve with one degree of freedom has a lot of area in the tails. For larger sample sizes, the value of s is less likely to be far from σ, and the t curve is closer to the normal curve. With 10 degrees of freedom (corresponding to a sample size of 11), the difference between the t curve and the normal curve is not great. If a t curve with 30 degrees of freedom were plotted in Figure 5.9, it would be indistinguishable from the normal curve. Page 347 FIGURE 5.9 Plots of the probability density function of the Student's t curve for various degrees of freedom. The normal curve with mean 0 and variance 1 (z curve) is plotted for comparison. The t curves are more spread out than the normal, but the amount of extra spread decreases as the number of degrees of freedom increases. Table A.3 (in Appendix A), called a t table, provides probabilities associated with the Student's t distribution. We present some examples to show how to use the table. Example 5.15 A random sample of size 10 is to be drawn from a normal distribution with mean 4. The Student's t statistic is to be computed. What is the probability that t > 1.833? Solution This t statistic has 10 − 1 = 9 degrees of freedom. From the t table, P(t > 1.833) = 0.05. See Figure 5.10. FIGURE 5.10 Solution to Example 5.15. Example 5.16 Refer to Example 5.15. Find P(t > 1.5). Solution Looking across the row corresponding to 9 degrees of freedom, we see that the t table does not list the value 1.5. We find that P(t > 1.383) = 0.10 and P(t > 1.833) = 0.05. We conclude that 0.05 < P(t > 1.5) < 0.10. See Figure 5.11. If a more precise result were required, linear interpolation could be used as follows: A computer package gives the answer correct to three significant digits as 0.0839. FIGURE 5.11 Solution to Example 5.16. Example 5.17 Find the value or the t12 distribution whose upper-tail probability is 0.025. Page 348 Solution Look down the column headed "0.025" to the row corresponding to 12 degrees of freedom. The value for t12 is 2.179. Example 5.18 Find the value for the t14 distribution whose lower-tail probability is 0.01. Solution Look down the column headed "0.01" to the row corresponding to 14 degrees of freedom. The value for t14 is 2.624. This value cuts off an area, or probability, of 1% in the upper tail. The value whose lower-tail probability is 1% is −2.624. Don't Use the Student's t Statistic If the Sample Contains Outliers For the Student's t statistic to be valid, the sample must come from a population that is approximately normal. Such samples rarely contain outliers. Therefore, methods involving the Student's t statistic should not be used for samples that contain outliers. Confidence Intervals Using the Student's t Distribution When the sample size is small, and the population is approximately normal, we can use the Student's t distribution to compute confidence intervals. We illustrate this with an example. A metallurgist is studying a new welding process. He manufactures five welded joints and measures the yield strength of each. The five values (in ksi) are 56.3, 65.4, 58.7, 70.1, and 63.9. Assume that these values are a random sample from an approximately normal population. The task is to find a confidence interval for the mean strength of welds made by this process. When the sample size is large, we don't need to worry much about the nature of the population, because the Central Limit Theorem guarantees that the quantity will be approximately normally distributed. When the sample is small, however, the distribution of the population must be approximately normal. The confidence interval in this situation is constructed much like the ones in Section 5.1, except that the z-score is replaced with a value from the Student's t distribution. The quantity has a Student's t distribution with n − 1 degrees of freedom. Figure 5.12 shows the t4 distribution. From the Student's t table, we find that 95% of the area under the curve is contained between the values t = −2.776 and t = 2.776. It follows that for 95% of all the samples that might have been chosen, FIGURE 5.12 The Student's t distribution with four degrees of freedom. Ninety-five percent of the area falls between t = −2.776 and t = 2.776. Put another way, for 95% of all the samples that might have been chosen, it is the case Page 349 that Multiplying by −1 and adding µ: across the inequality, we obtain a 95% confidence interval for In this example, the sample mean is , and the sample standard deviation s = 5.4838. The sample size is n = 5. Substituting values for , s, and n, we find that a 95% confidence interval for µ is 62.88 − 6.81 11. The statement "µ > 11" is a hypothesis about the population mean µ. To determine just how certain we can be that a hypothesis such as this is true, we must perform a hypothesis test. A hypothesis test produces a number between 0 and 1 that measures the degree of certainty we may have in the truth of a hypothesis about a quantity such as a population mean or proportion. It turns out that hypothesis tests are closely related to confidence intervals. In general, whenever a confidence interval can be computed, a hypothesis test can also be performed, and vice versa. 6.1 Large-Sample Tests for a Population Mean We begin with an example. A certain type of automobile engine emits a mean of 100 mg of oxides of nitrogen (NOx) per second at 100 horsepower. A modification to the engine design has been proposed that may reduce NOx emissions. The new design will be put into production if it can be demonstrated that its mean emission rate is less than 100 mg/s. A sample of 50 modified engines are built and tested. The sample mean NOx emission is 92 mg/s, and the sample standard deviation is 21 mg/s. Page 403 The population in this case consists of the emission rates from the engines that would be built if this modified design is put into production. If there were no uncertainty in the sample mean, then we could conclude that the modification would reduce emissions—from 100 to 92 mg/s. Of course, there is uncertainty in the sample mean. The population mean will actually be somewhat higher or lower than 92. The manufacturers are concerned that the modified engines might not reduce emissions at all, that is, that the population mean might be 100 or more. They want to know whether this concern is justified. The question, therefore, is this: Is it plausible that this sample, with its mean of 92, could have come from a population whose mean is 100 or more? This is the sort of question that hypothesis tests are designed to address, and we will now construct a hypothesis test to address this question. We have observed a sample with mean 92. There are two possible interpretations of this observation: 1. The population mean is actually greater than or equal to 100, and the sample mean is lower than this only because of random variation from the population mean. Thus emissions will not go down if the new design is put into production, and the sample is misleading. 2. The population mean is actually less than 100, and the sample mean reflects this fact. Thus the sample represents a real difference that can be expected if the new design is put into production. These two explanations have standard names. The first is called the null hypothesis. In most situations, the null hypothesis says that the effect indicated by the sample is due only to random variation between the sample and the population. The second explanation is called the alternate hypothesis. The alternate hypothesis says that the effect indicated by the sample is real, in that it accurately represents the whole population. In our example, the engine manufacturers are concerned that the null hypothesis might be true. A hypothesis test assigns a quantitative measure to the plausibility of the null hypothesis. After performing a hypothesis test, we will be able to tell the manufacturers, in numerical terms, precisely how valid their concern is. To make things more precise, we express everything in symbols. The null hypothesis is denoted H0. The alternate hypothesis is denoted H1. As usual, the population mean is denoted µ. We have, therefore, In performing a hypothesis test, we essentially put the null hypothesis on trial. We begin by assuming that H0 is true, just as we begin a trial by assuming a defendant to be innocent. The random sample provides the evidence. The hypothesis test involves measuring the strength of the disagreement between the sample and H0 to produce a number between 0 and 1, called a Pvalue. The P-value measures the plausibility of H0. The smaller the P-value, the stronger the evidence is against H0. If the P-value is sufficiently small, we may be willing to Page 404 abandon our assumption that H0 is true and believe H1 instead. This is referred to as rejecting the null hypothesis. In this example, let X1, ..., X50 be the emissions rates measured from the 50 sample engines. The observed value of the sample mean is . We will also need to know the sample standard deviation, which is s = 21. We must assess the plausibility of H0, which says that the population mean is 100 or more, given that we have observed a sample from this population whose mean is only 92. We will do this in two steps, as follows: 1. We will compute the distribution of under the assumption that H0 is true. This distribution is called the null distribution of . 2. We will compute the P-value. This is the probability, under the assumption that H0 is true, of observing a value of whose disagreement with H0 is as least as great as that of the observed value of 92. To perform step 1, note that is the mean of a large sample, so the Central Limit Theorem specifies that it comes from a normal distribution whose mean is µ and whose variance is σ2/50, where σ2 is the population variance and 50 is the sample size. We must specify values for µ and for σ in order to determine the null distribution. Since we are assuming that H0 is true, we assume that µ ≥ 100. This does not provide a specific value for µ. We take as the assumed value for µ the value closest to the alternate hypothesis H1, for reasons that will be explained later in this section. Thus we assume µ = 100. We do not know the population standard deviation σ. However, since the sample is large, we may approximate σ with the sample standard deviation s = 21. Thus we have determined that under H0, has a normal distribution with mean 100 and standard deviation . The null distribution is . We are now ready for step 2. Figure 6.1 illustrates the null distribution. The number 92 indicates the point on the distribution corresponding to the observed value of . How plausible is it that a number sampled from this distribution would be as small as 92? This is measured by the P-value. The P-value is the probability that a number drawn from the null distribution would disagree with H0 at least as strongly as the observed value of , which is 92. Since H0 specifies that the mean of is greater than or equal to 100, values less than 92 are in greater disagreement with H0. The P-value, therefore, is the probability that a number drawn from an N(100, 2.972) distribution is less than or equal to 92. This probability is determined by computing the z-score: The null distribution of is N(100, 2.972). Thus if H0 is true, the probability that takes on a value as extreme as or more extreme than the observed value of 92 is 0.0036. This is the P-value. FIGURE 6.1 From the z table, the probability that a standard normal random variable z is less than or equal to −2.69 is 0.0036. The P-value for this test is 0.0036. The P-value, as promised, provides a quantitative measure of the plausibility of H0. But how do we interpret this quantity? The proper interpretation is rather subtle. The P-value tells us that if H0 were true, the probability of drawing a sample whose mean was as far from Page 405 H0 as the observed value of 92 is only 0.0036. Therefore, one of the following two conclusions is possible: 1. H0 is false. 2. H0 is true, which implies that of all the samples that might have been drawn, only 0.36% of them have a mean as small as or smaller than that of the sample actually drawn. In other words, our sample mean lies in the most extreme 0.36% of its distribution. In practice, events in the most extreme 0.36% of their distributions very seldom occur. Therefore we reject H0 and conclude that the new engines will lower emissions. The null hypothesis in this case specified only that µ ≥ 100. In assuming H0 to be true, why did we choose the value µ = 100, which is closest to H1? To give H0 a fair test, we must test it in its most plausible form. The most plausible value for µ is the value closest to . Now , so among the values µ ≥ 100, the closest to is µ = 100. This value is also the one closest to H1. This is typical. In practice, when it is of interest to perform a hypothesis test, the most plausible value for H0 will be the value closest to H1. It is natural to ask how small the P-value should be in order to reject H0. Some people use the "5% rule"; they reject H0 if P ≤ 0.05. However, there is no scientific justification for this or any other rule. We discuss this issue in more detail in Section 6.2. Note that the method we have just described uses the Central Limit Theorem. It follows that for this method to be valid, the sample size must be reasonably large, say 30 or more. Hypothesis tests that are sometimes valid for small samples are presented in Section 6.4. Finally, note that the calculation of the P-value was done by computing a z-score. For this reason, the z-score is called a test statistic. A test that uses a z-score as a test statistic is called a z test. Page 406 There are many kinds of hypothesis tests. All of them follow a basic series of steps, which are outlined in the following box. Steps in Performing a Hypothesis Test 1. Define H0 and H1. 2. Assume H0 to be true. 3. Compute a test statistic. A test statistic is a statistic that is used to assess the strength of the evidence against H0. 4. Compute the P-value of the test statistic. The P-value is the probability, assuming H0 to be true, that the test statistic would have a value whose disagreement with H0 is as great as or greater than that actually observed. The P-value is also called the observed significance level. State a conclusion about the strength of the evidence against H0. 5. Another Way to Express H0 We have mentioned that when assuming H0 to be true, we use the value closest to H1. Some authors consider this single value to be H0, so that, in the previous example, they would write H0: µ = 100 instead of H0: µ ≥ 100. There is an advantage to this notation, which is that it makes it clear which value is being used when H0 is assumed to be true. But there is a disadvantage when it comes to interpretation. Generally, the value closest to H1 is of no special interest. For example, in the emissions example just discussed, we are not specifically concerned with the possibility µ = 100, but with the possibility µ ≥ 100. The importance of rejecting H0 is not that we reject the single value µ = 100, but that we reject all values µ ≥ 100. For this reason, we choose to write H0: µ ≥ 100. Example 6.1 The article "Wear in Boundary Lubrication" (S. Hsu, R. Munro, and M. Shen, Journal of Engineering Tribology, 2002:427–441) discusses several experiments involving various lubricants. In one experiment, 45 steel balls lubricated with purified paraffin were subjected to a 40 kg load at 600 rpm for 60 minutes. The average wear, measured by the reduction in diameter, was 673.2 µm, and the standard deviation was 14.9 µm. Assume that the specification for a lubricant is that the mean wear be less than 675 µm. Find the P-value for testing H0: µ ≥ 675 versus H1: µ < 675. Solution First let's translate the problem into statistical language. We have a simple random sample X1, ..., X45 of wear diameters. The sample mean and standard deviation are and s = 14.9. The population mean is unknown and denoted by µ. Before getting into the construction of the test, we'll point out again that the basic issue is the uncertainty in the sample mean. If there were no uncertainty in the sample mean, we could conclude that the lubricant would meet the specification, since 673.2 < 675. The question is whether the uncertainty in the sample mean is large enough so that the population mean could plausibly be as high as 675. To perform the hypothesis test, we follow the steps given earlier. The null hypothesis is that the lubricant does not meet the specification, and that the difference between Page 407 the sample mean of 673.2 and 675 is due to chance. The alternate hypothesis is that the lubricant does indeed meet the specification. We assume H0 is true, so that the sample was drawn from a population with mean µ = 675 (the value closest to H1). We estimate the population standard deviation σ with the sample standard deviation s = 14.9. The test is based on population with mean 675 and standard deviation . Under H0, comes from a normal . The P-value is the probability of observing a sample mean less than or equal to 673.2. The test statistic is the zscore, which is The P-value is 0.209 (see Figure 6.2). Therefore if H0 is true, there is a 20.9% chance to observe a sample whose disagreement with H0 is as least as great as that which was actually observed. Since 0.209 is not a very small probability, we do not reject H0. Instead, we conclude that H0 is plausible. The data do not show conclusively that the lubricant meets the specification. Note that we are not concluding that H0 is true, only that it is plausible. We will discuss this distinction further in Section 6.2. The null distribution of is N(675, 2.222). Thus if H0 is true, the probability that takes on a value as extreme as or more extreme than the observed value of 673.2 is 0.209. This is the Pvalue. FIGURE 6.2 The following computer output (from MINITAB) presents the results of Example 6.1. One-Sample Z: Wear Test of mu = 675 vs < 675 The assumed standard deviation = 14.9 Variable N Wear Mean StDev 45 673.200 14.9 SE Mean 2.221 95% Upper Z P Bound 676.853 −0.81 0.209 The output states the null hypothesis as µ = 675 rather than µ ≥ 675. This reflects the fact that the value µ = 675 is used to construct the null distribution. The quantity "SE Mean" is the standard deviation of , estimated by . The output also provides a 95% upper confidence bound for µ. Page 408 In the examples shown so far, the null hypothesis specified that the population mean was less than or equal to something, or greater than or equal to something. In some cases, a null hypothesis specifies that the population mean is equal to a specific value. Example 6.2 provides an illustration. Example 6.2 A scale is to be calibrated by weighing a 1000 g test weight 60 times. The 60 scale readings have mean 1000.6 g and standard deviation 2 g. Find the P-value for testing H0: µ = 1000 versus H1: µ ≠ 1000. Solution Let µ denote the population mean reading. The null hypothesis says that the scale is in calibration, so that the population mean µ is equal to the true weight of 1000 g, and the difference between the sample mean reading and the true weight is due entirely to chance. The alternate hypothesis says that the scale is out of calibration. In this example, the null hypothesis specifies that µ is equal to a specific value, rather than greater than or equal to or less than or equal to. For this reason, values of that are either much larger or much smaller than µ will provide evidence against H0. In the previous examples, only the values of on one side of µ provided evidence against H0. We assume H0 is true, and that therefore the sample readings were drawn from a population with mean µ = 1000. We approximate the population standard deviation σ with s = 2. The null distribution of is normal with mean 1000 and standard deviation . The z-score of the observed value is Since H0 specifies µ = 1000, regions in both tails of the curve are in greater disagreement with H0 than the observed value of 1000.6. The P-value is the sum of the areas in both of these tails, which is 0.0204 (see Figure 6.3). Therefore, if H0 is true, the probability of a result as extreme as or more extreme than that observed is only 0.0204. The evidence against H0 is pretty strong. It would be prudent to reject H0 and to recalibrate the scale. The null distribution of is N(1000, 0.2582). Thus if H0 is true, the probability that takes on a value as extreme as or more extreme than the observed value of 1000.6 is 0.0204. This is the Pvalue. FIGURE 6.3 When H0 specifies a single value for µ, as in Example 6.2, both tails contribute to Page 409 the P-value, and the test is said to be a two-sided or two-tailed test. When H0 specifies only that µ is greater than or equal to, or less than or equal to a value, only one tail contributes to the Pvalue, and the test is called a one-sided or one-tailed test. We conclude this section by summarizing the procedure used to perform a large-sample hypothesis test for a population mean. Summary Let X1, ..., Xn be a large (e.g., n > 30) sample from a population with mean µ and standard deviation σ. To test a null hypothesis of the form H0: µ ≤ µ0, H0: µ ≥ µ0, or H0: µ = µ0: ■ Compute the z-score: ■ If σ is unknown it may be approximated with s. Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows: . Alternate Hypothesis P-value H1: µ > µ0 Area to the right of z H1: µ < µ0 Area to the left of z H1: µ ≠ µ0 Sum of the areas in the tails cut off by z and −z Exercises for Section 6.1 1. A sample of 50 copper wires had a mean resistance of 1.03 mΩ with a standard deviation of 0.1 mΩ. Let µ represent the mean resistance of copper wires of this type. a. Find the P-value for testing H0: µ ≤ 1 versus H1: µ > 1. b. 2. Either the mean resistance is greater than 1 mΩ, or the sample is in the most extreme _____% of its distribution. A sample of 65 electric motors had a mean efficiency of 0.595 with a standard deviation of 0.05. Let µ represent the mean efficiency of electric motors of this type. a. Find the P-value for testing H0: µ ≥ 0.6 versus H1: µ < 0.6. b. 3. Either the mean efficiency is less than 0.6, or the sample is in the most extreme _____% of its distribution. The article "Supply Voltage Quality in Low-Voltage Industrial Networks of Estonia" (T. Vinnal, K. Janson, et al., Estonian Journal of Engineering, 2012:102–126) presents voltage measurements for a sample of 66 industrial networks in Estonia. Assume the rated voltage for these networks is 232 V. The sample mean voltage was 231.7 V with a standard deviation of 2.19 V. Let µ represent the population mean voltage for these networks. a. Find the P-value for testing H0: µ = 232 versus H1: µ ≠ 232. b. 4. Either the mean voltage is not equal to 232, or the sample is in the most extreme _____% of its distribution. The pH of an acid solution used to etch aluminum varies somewhat from batch to batch. In a sample of 50 batches the mean pH was 2.6, with a standard deviation of 0.3. Let µ represent the mean pH for batches of this solution. Page 410 a. Find the P-value for testing H0: µ ≤ 2.5 versus H1: µ > 2.5. b. 5. Either the mean pH is greater than 2.5 mm, or the sample is in the most extreme _____% of its distribution. Recently many companies have been experimenting with telecommuting, allowing employees to work at home on their computers. Among other things, telecommuting is supposed to reduce the number of sick days taken. Suppose that at one firm, it is known that over the past few years employees have taken a mean of 5.4 sick days. This year, the firm introduces telecommuting. Management chooses a simple random sample of 80 employees to follow in detail, and, at the end of the year, these employees average 4.5 sick days with a standard deviation of 2.7 days. Let µ represent the mean number of sick days for all employees of the firm. a. Find the P-value for testing H0: µ ≥ 5.4 versus H1: µ < 5.4. b. 6. Do you believe it is plausible that the mean number of sick days is at least 5.4, or are you convinced that it is less than 5.4? Explain your reasoning. A certain type of stainless steel powder is supposed to have a mean particle diameter of µ = 15 µm. A random sample of 87 particles had a mean diameter of 15.2 µm, with a standard deviation of 1.8 µm. A test is made of H0: µ = 15 versus H1: µ ≠ 15. Find the P-value. Do you believe it is plausible that the mean diameter is 15 µm, or are you convinced that it differs from 15 µm? Explain your reasoning. In a test of corrosion resistance, a sample of 60 Incoloy steel specimens were immersed in acidified brine for four hours, after which each specimen had developed a number of corrosive pits. The maximum pit depth was measured for each specimen. The mean depth was 850 µm with a standard deviation of 153 µm. The specification is that the population mean depth µ is less than 900 µm. a. Find the P-value for testing H0: µ ≥ 900 versus H1: µ < 900. a. b. 7. b. 8. Do you believe it is plausible that the mean depth is at least 900µm, or are you convinced that it is less than 900µm? Explain. A process that manufactures steel bolts is supposed to be calibrated to produce bolts with a mean length of 5 cm. A sample of 100 bolts has a mean length of 5.02 cm with a standard deviation of 0.06 cm. Let µ be the mean length of bolts manufactured by this process. a. Find the P-value for testing H0: µ = 5 versus H1: µ ≠ 5. b. Do you believe it is plausible that the process is properly calibrated, or are you 9. convinced that it is out of calibration? Explain. The article "Predicting Profit Performance for Selecting Candidate International Construction Projects" (S. Han, D. Kim, and H. Kim, Journal of Construction Engineering and Management Science, 2007: 425–436) presents an analysis of the profit of international construction projects. In a sample of 126 projects, the average profit margin (in percent) was 8.24 with a standard deviation of 16.33. A test is made of H0: µ ≥ 10 versus H1: µ < 10. Find the P-value. Do you believe that it is plausible that the mean profit margin is at least 10%, or are you convinced that it is less than 10%? Explain your reasoning. 10. A new concrete mix is being designed to provide adequate compressive strength for concrete blocks. The specification for a particular application calls for the blocks to have a mean compressive strength µ greater than 1350 kPa. A sample of 100 blocks is produced and tested. Their mean compressive strength is 1356 kPa and their standard deviation is 70 kPa. A test is made of H0: µ ≤ 1350 versus H1: µ > 1350. a. b. Find the P-value. Do you believe it is plausible that the blocks do not meet the specification, or are you convinced that they do? Explain your reasoning. 11. Fill in the blank: If the null hypothesis is H0: µ ≤ 5, then the mean of under the null distribution is _____. i. 0 ii. 5 iii. Any number less than or equal to 5. iv. We can't tell unless we know H1. a. b. 12. Fill in the blank: In a test of H0: µ ≥ 10 versus H1: µ < 10, the sample mean was and the P-value was 0.04. This means that if µ = 10, and the experiment were repeated Page 411 100 times, we would expect to obtain a value of of 8 or less approximately _____ times. i. 10 ii. 2 iii. 4 iv. 8 13. Fill in the blank: In a test of H0: µ = 5 vs. H1: µ ≠ 5, the sample mean was , and the Pvalue was 0.22. This means that if µ = 5, and the experiment were repeated 100 times, we would expect to obtain a value of between 4 and 6 approximately _____ times. i. 4 ii. 22 iii. 78 iv. 6 14. An automotive engineer subjects a large number of brake pads to a stress test and measures the wear on each. She obtains mm and . Use this information to find the P-value for testing H0: µ = 7.0 versus H1: µ ≠ 7.0. 15. An engineer takes a large number of independent measurements of the length of a component and obtains and Use this information to find the Pvalue for testing H0: µ = 5.0 versus H1: µ ≠ 5.0. 16. The following MINITAB output presents the results of a hypothesis test for a population mean µ. One-Sample Z: X Test of mu = 73.5 vs not = 73.5 The assumed standard deviation = 2.3634 Variable N X a. b. c. d. Mean StDev SE Mean 14573.24612.3634 0.1963 95% CI Z P (72.8614, −1.29 0.196 73.6308) Is this a one-tailed or two-tailed test? What is the null hypothesis? What is the P-value? Use the output and an appropriate table to compute the P-value for the test of H0: µ ≥ 73.6 versus H1: µ < 73.6. e. Use the output and an appropriate table to compute a 99% confidence interval for µ. 17. The following MINITAB output presents the results of a hypothesis test for a population mean µ. One-Sample Z: X Test of mu = 45 vs > 45 The assumed standard deviation = 5.0921 Variable N X a. b. c. d. Mean StDev SE Mean 12246.25985.0921 0.4610 95% Lower Z P Bound 45.50152.73 0.003 Is this a one-tailed or two-tailed test? What is the null hypothesis? What is the P-value? Use the output and an appropriate table to compute the P-value for the test of H0: µ ≤ 46 vs. H1: µ > 46. e. Use the output and an appropriate table to compute a 98% confidence interval for µ. Page 412 18. The following MINITAB output presents the results of a hypothesis test for a population mean µ. Some of the numbers are missing. Fill them in. One-Sample Z: X Test of mu = 10.5 vs < 10.5 The assumed standard deviation = 2.2136 Variable N Mean StDev SE Mean X (a) (b) 2.2136 0.2767 95% Upper Bound Z P 10.6699 −1.03 (c) 19. The following MINITAB output presents the results of a hypothesis test for a population mean µ. Some of the numbers are missing. Fill them in. One-Sample Z: X Test of mu = 3.5 vs not = 3.5 The assumed standard deviation = 2.00819 Variable N Mean StDev SE Mean X 87 4.07114 2.00819 (a) 95% CI Z P (3.64916, 4.49312) (b) (c) 6.2 Drawing Conclusions from the Results of Hypothesis Tests Let's take a closer look at the conclusions reached in Examples 6.1 and 6.2 (in Section 6.1). In Example 6.2, we rejected H0; in other words, we concluded that H0 was false. In Example 6.1, we did not reject H0. However, we did not conclude that H0 was true. We could only conclude that H0 was plausible. In fact, the only two conclusions that can be reached in a hypothesis test are that H0 is false or that H0 is plausible. In particular, one can never conclude that H0 is true. To understand why, think of Example 6.1 again. The sample mean was , and the null hypothesis was µ ≥ 675. The conclusion was that 673.2 is close enough to 675 so that the null hypothesis might be true. But a sample mean of 673.2 obviously could not lead us to conclude that µ ≥ 675 is true, since 673.2 is less than 675. This is typical of many situations of interest. The test statistic is consistent with the alternate hypothesis and disagrees somewhat with the null. The only issue is whether the level of disagreement, measured with the P-value, is great enough to render the null hypothesis implausible. How do we know when to reject H0? The smaller the P-value, the less plausible H0 becomes. A common rule of thumb is to draw the line at 5%. According to this rule of thumb, if P ≤ 0.05, H0 is rejected; otherwise H0 is not rejected. In fact, there is no sharp dividing line between conclusive evidence against H0 and inconclusive evidence, just as there is no sharp dividing line between hot and cold weather. So while this rule of thumb is convenient, it has no real scientific justification. Page 413 Summary ■ The smaller the P-value, the more certain we can be that H0 is false. ■ The larger the P-value, the more plausible H0 becomes, but we can never be certain that H0 is true. ■ A rule of thumb suggests to reject H0 whenever P ≤ 0.05. While this rule is convenient, it has no scientific basis. Statistical Significance Whenever the P-value is less than a particular threshold, the result is said to be "statistically significant" at that level. So, for example, if P ≤ 0.05, the result is statistically significant at the 5% level; if P ≤ 0.01, the result is statistically significant at the 1% level, and so on. If a result is statistically significant at the 100α% level, we can also say that the null hypothesis is "rejected at level 100α%." Example 6.3 A hypothesis test is performed of the null hypothesis H0: µ = 0. The P-value turns out to be 0.03. Is the result statistically significant at the 10% level? The 5% level? The 1% level? Is the null hypothesis rejected at the 10% level? The 5% level? The 1% level? Solution The result is statistically significant at any level greater than or equal to 3%. Thus it is statistically significant at the 10% and 5% levels, but not at the 1% level. Similarly, we can reject the null hypothesis at any level greater than or equal to 3%, so H0 is rejected at the 10% and 5% levels, but not at the 1% level. Sometimes people report only that a test result was statistically significant at a certain level, without giving the P-value. It is common, for example, to read that a result was "statistically significant at the 5% level" or "statistically significant (P < 0.05)." This is a poor practice, for three reasons. First, it provides no way to tell whether the P-value was just barely less than 0.05, or whether it was a lot less. Second, reporting that a result was statistically significant at the 5% level implies that there is a big difference between a P-value just under 0.05 and one just above 0.05, when in fact there is little difference. Third, a report like this does not allow readers to decide for themselves whether the P-value is small enough to reject the null hypothesis. If a reader believes that the null hypothesis should not be rejected unless P < 0.01, then reporting only that P < 0.05 does not allow that reader to decide whether or not to reject H0. Reporting the P-value gives more information about the strength of the evidence against the null hypothesis and allows each reader to decide for himself or herself whether to reject. Software packages always output P-values; these should be included whenever the results of a hypothesis test are reported. Page 414 Summary Let α be any value between 0 and 1. Then, if P ≤ α, ■ ■ ■ The result of the test is said to be statistically significant at the 100α% level. The null hypothesis is rejected at the 100α% level. When reporting the result of a hypothesis test, report the P-value, rather than just comparing it to 5% or 1%. The P-value Is Not the Probability That H0 Is True Since the P-value is a probability, and since small P-values indicate that H0 is unlikely to be true, it is tempting to think that the P-value represents the probability that H0 is true. This is emphatically not the case. The concept of probability discussed here is useful only when applied to outcomes that can turn out in different ways when experiments are repeated. It makes sense to define the P-value as the probability of observing an extreme value of a statistic such as , since the value of could come out differently if the experiment were repeated. The null hypothesis, on the other hand, either is true or is not true. The truth or falsehood of H0 cannot be changed by repeating the experiment. It is therefore not correct to discuss the "probability" that H0 is true. At this point we must mention that there is a notion of probability, different from that which we discuss in this book, in which one can compute a probability that a statement such as a null hypothesis is true. This kind of probability, called subjective probability, plays an important role in the theory of Bayesian statistics. The kind of probability we discuss in this book is called frequentist probability. A good reference for Bayesian statistics is Lee (2013). Choose H0 to Answer the Right Question When performing a hypothesis test, it is important to choose H0 and H1 appropriately so that the result of the test can be useful in forming a conclusion. Examples 6.4 and 6.5 illustrate this. Example 6.4 Specifications for a water pipe call for a mean breaking strength µ of more than 2000 lb per linear foot. Engineers will perform a hypothesis test to decide whether or not to use a certain kind of pipe. They will select a random sample of 1 ft sections of pipe, measure their breaking strengths, and perform a hypothesis test. The pipe will not be used unless the engineers can conclude that µ > 2000. Assume they test H0: µ ≤ 2000 versus H1: µ > 2000. Will the engineers decide to use the pipe if H0 is rejected? What if H0 is not rejected? Solution If H0 is rejected, the engineers will conclude that µ > 2000, and they will use the pipe. If H0 is not rejected, the engineers will conclude that µ might be less than or equal to 2000, and they will not use the pipe. Page 415 In Example 6.4, the engineers' action with regard to using the pipe will differ depending on whether H0 is rejected or not. This is therefore a useful test to perform, and H0 and H1 have been specified correctly. Example 6.5 In Example 6.4, assume the engineers test H0: µ ≥ 2000 versus H1: µ < 2000. Will the engineers decide to use the pipe if H0 is rejected? What if H0 is not rejected? Solution If H0 is rejected, the engineers will conclude that µ < 2000, and they will not use the pipe. If H0 is not rejected, the engineers will conclude that µ might be greater than or equal to 2000, but that it also might not be. So again, they won't use the pipe. In Example 6.5, the engineers' action with regard to using the pipe will be the same—they won't use it—whether or not H0 is rejected. There is no point in performing this test. The hypotheses H0 and H1 have not been specified correctly. Final note: In a one-tailed test, the equality always goes with the null hypothesis. Thus if µ0 is the point that divides H0 from H1, we may have H0: µ ≤ µ0 or H0: µ ≥ µ0, but never H0: µ < µ0 or H0: µ > µ0. The reason for this is that when defining the null distribution, we represent H0 with the value of µ closest to H1. Without the equality, there is no value specified by H0 that is the closest to H1. Therefore the equality must go with H0. Statistical Significance Is Not the Same as Practical Significance When a result has a small P-value, we say that it is "statistically significant." In common usage, the word significant means "important." It is therefore tempting to think that statistically significant results must always be important. This is not the case. Sometimes statistically significant results do not have any scientific or practical importance. We will illustrate this with an example. Assume that a process used to manufacture synthetic fibers is known to produce fibers with a mean breaking strength of 50 N. A new process, which would require considerable retooling to implement, has been developed. In a sample of 1000 fibers produced by this new method, the average breaking strength was 50.1 N, and the standard deviation was 1 N. Can we conclude that the new process produces fibers with greater mean breaking strength? To answer this question, let µ be the mean breaking strength of fibers produced by the new process. We need to test H0: µ ≤ 50 versus H1: µ > 50. In this way, if we reject H0, we will conclude that the new process is better. Under H0, the sample mean with mean 50 and standard deviation has a normal distribution . The z-score is The P-value is 0.0008. This is very strong evidence against H0. The new process produces fibers with a greater mean breaking strength. Page 416 What practical conclusion should be drawn from this result? On the basis of the hypothesis test, we are quite sure that the new process is better. Would it be worthwhile to implement the new process? Probably not. The reason is that the difference between the old and new processes, although highly statistically significant, amounts to only 0.1 N. It is unlikely that this difference is large enough to matter. The lesson here is that a result can be statistically significant without being large enough to be of practical importance. How can this happen? A difference is statistically significant when it is large compared to its standard deviation. In the example, a difference of 0.1 N was statistically significant because the standard deviation was only 0.0316 N. When the standard deviation is very small, even a small difference can be statistically significant. The P-value does not measure practical significance. What it does measure is the degree of confidence we can have that the true value is really different from the value specified by the null hypothesis. When the P-value is small, then we can be confident that the true value is really different. This does not necessarily imply that the difference is large enough to be of practical importance. The Relationship Between Hypothesis Tests and Confidence Intervals Both confidence intervals and hypothesis tests are concerned with determining plausible values for a quantity such as a population mean µ. In a hypothesis test for a population mean µ, we specify a particular value of µ (the null hypothesis) and determine whether that value is plausible. In contrast, a confidence interval for a population mean µ can be thought of as the collection of all values for µ that meet a certain criterion of plausibility, specified by the confidence level 100(1 − α)%. In fact, the relationship between confidence intervals and hypothesis tests is very close. To be specific, the values contained within a two-sided level 100(1 − α)% confidence interval for a population mean µ are precisely those values for which the P-value of a two-tailed hypothesis test will be greater than α. To illustrate this, consider the following example (presented as Example 5.2 in Section 5.1). The sample mean lifetime of 50 microdrills was holes drilled and the standard deviation was s = 6.83. Setting α to 0.05 (5%), the 95% confidence interval for the population mean lifetime µ was computed to be (10.79, 14.57). Suppose we wanted to test the hypothesis that µ was equal to one of the endpoints of the confidence interval. For example, consider testing H0: µ = 10.79 versus H1: µ ≠ 10.79. Under H0, the observed value comes from a normal distribution with mean 10.79 and standard deviation . The z-score is (12.68−10.79)/0.9659 = 1.96. Since H0 specifies that µ is equal to 10.79, both tails contribute to the P-value, which is 0.05, and thus equal to α (see Figure 6.4). FIGURE 6.4 interval based on The sample mean is equal to 12.68. Since 10.79 is an endpoint of a 95% confidence , the P-value for testing H0: µ = 10.79 is equal to 0.05. Now consider testing the hypothesis H0: µ = 14.57 versus H1: µ ≠ 14.57, where 14.57 is the other endpoint of the confidence interval. This time we will obtain z = (12.68 − 14.57)/0.9659 = −1.96, and again the P-value is 0.05. It is easy to check that if we choose any value µ0 in the interval (10.79, 14.57) and test H0: µ = µ0 versus H1: µ ≠ µ0, the P-value will be greater than 0.05. On the other hand, if we choose µ0 < 10.79 or µ0 > 14.57, the P-value will be less than 0.05. Thus the 95% confidence interval consists of precisely those values of µ whose Page 417 P-values are greater than 0.05 in a hypothesis test. In this sense, the confidence interval contains all the values that are plausible for the population mean µ. It is easy to check that a one-sided level 100(1 − α)% confidence interval consists of all the values for which the P-value in a one-tailed test would be greater than α. For example, with , s = 6.83, and n = 50, the 95% lower confidence bound for the lifetime of the drills is 11.09. If µ0 > 11.09, then the P-value for testing H0: µ ≤ µ0 will be greater than 0.05. Similarly, the 95% upper confidence bound for the lifetimes of the drills is 14.27. If µ0 < 14.27, then the P-value for testing H0: µ ≥ µ0 will be greater than 0.05. Exercises for Section 6.2 1. 2. 3. For which P-value is the null hypothesis more plausible: P = 0.5 or P = 0.05? True or false: a. If we reject H0, then we conclude that H0 is false. b. If we do not reject H0, then we conclude that H0 is true. c. If we reject H0, then we conclude that H1 is true. d. If we do not reject H0, then we conclude that H1 is false. If P = 0.01, which is the best conclusion? i. H0 is definitely false. ii. H0 is definitely true. iii. There is a 1% probability that H0 is true. iv. H0 might be true, but it's unlikely. v. H0 might be false, but it's unlikely. vi. H0 is plausible. 4. If P = 0.50, which is the best conclusion? i. H0 is definitely false. ii. H0 is definitely true. iii. There is a 50% probability that H0 is true. iv. H0 is plausible, and H1 is false. v. 5. 6. Both H0 and H1 are plausible. True or false: If P = 0.02, then a. The result is statistically significant at the 5% level. b. The result is statistically significant at the 1% level. c. The null hypothesis is rejected at the 5% level. d. The null hypothesis is rejected at the 1% level. It is desired to check the calibration of a scale by weighing a standard 10-gram weight 100 times. Let µ be the population mean reading on the scale, so that the scale is in calibration if µ = 10 and out of calibration if µ ≠ 10. A test is made of the hypotheses H0: µ = 10 versus H1: µ ≠ 10. Consider three possible conclusions: (i) The scale is in calibration. (ii) The scale is not in calibration. (iii) The scale might be in calibration. a. Which of the three conclusions is best if H0 is rejected? b. Which of the three conclusions is best if H0 is not rejected? Page 418 7. Scores on a certain IQ test are known to have a mean of 100. A random sample of 60 students attend a series of coaching classes before taking the test. Let µ be the population mean IQ score that would occur if every student took the coaching classes. The classes are successful if µ > 100. A test is made of the hypotheses H0: µ ≤ 100 versus H1: µ > 100. Consider three possible conclusions: (i) The classes are successful. (ii) The classes are not successful. (iii) The classes might not be successful. a. Which of the three conclusions is best if H0 is rejected? b. Which of the three conclusions is best if H0 is not rejected? George performed a hypothesis test. Luis checked George's work by redoing the calculations. Both George and Luis agree that the result was statistically significant at the 5% level, but they got different P-values. George got a P-value of 0.20 and Luis got a Pvalue of 0.02. a. Is is possible that George's work is correct? Explain. b. Is is possible that Luis's work is correct? Explain. 9. The article "The Effect of Restricting Opening Hours on Alcohol-Related Violence" (S. Duailibi, W. Ponicki, et al., American Journal of Public Health, 2007:2276–2280) presented homicide rates for the years 1995–2005 for the town of Diadema, Brazil. In 2002, a law was passed requiring bars to close at 11 pm each night. After the law's passage, the homicide rate dropped by an average of 9 homicides per month, a statistically significant decrease. Which of the following is the best conclusion? i. It is reasonable to conclude that the law is responsible for a reduction of 9 homicides per month. ii. It is reasonable to conclude that the law is responsible for a reduction in homicides, but the actual amount might be somewhat more or less than 9 per month. iii. It is reasonable to conclude that the homicide rate decreased, but the law may not have anything to do with the decrease. iv. It is plausible that the homicide rate may not have decreased at all after the passage of the law. 10. Let µ be the radiation level to which a radiation worker is exposed during the course of a year. The Environmental Protection Agency has set the maximum safe level of exposure at 5 rem per year. If a hypothesis test is to be performed to determine whether a workplace is safe, which is the most appropriate null hypothesis: H0: µ ≤ 5, H0: µ ≥ 5, or H0: µ = 5? Explain. 11. In each of the following situations, state the most appropriate null hypothesis regarding the population mean µ. a. A new type of battery will be installed in heart pacemakers if it can be shown to have a mean lifetime greater than eight years. b. A new material for manufacturing tires will be used if it can be shown that the mean lifetime of tires will be more than 60,000 miles. c. A quality control inspector will recalibrate a flowmeter if the mean flow rate differs from 10 mL/s. 12. The installation of a radon abatement device is recommended in any home where the mean radon concentration is 4.0 picocuries per liter (pCi/L) or more, because it is thought that long-term exposure to sufficiently high doses of radon can increase the risk of cancer. Seventy-five measurements are made in a particular home. The mean concentration was 3.72 pCi/L and the standard deviation was 1.93 pCi/L. a. The home inspector who performed the test says that since the mean measurement is less than 4.0, radon abatement is not necessary. Explain why this reasoning is 8. incorrect. b. Because of health concerns, radon abatement is recommended whenever it is plausible that the mean radon concentration may be 4.0 pCi/L or more. State the appropriate null and alternate hypotheses for determining whether radon abatement is appropriate. c. Compute the P-value. Would you recommend radon abatement? Explain. 13. It is desired to check the calibration of a scale by weighing a standard 10 g weight 100 times. Let µ be the population mean reading on the scale, so that the scale is in calibration if µ = 10. A test is made of the hypotheses H0: µ = 10 versus H1: µ ≠ 10. Consider three possible conclusions: (i) The scale is in calibration. (ii) The scale is out of Page 419 calibration. (iii) The scale might be in calibration. a. Which of the three conclusions is best if H0 is rejected? b. Which of the three conclusions is best if H0 is not rejected? c. Is it possible to perform a hypothesis test in a way that makes it possible to demonstrate conclusively that the scale is in calibration? Explain. 14. A machine that fills cereal boxes is supposed to be calibrated so that the mean fill weight is 12 oz. Let µ denote the true mean fill weight. Assume that in a test of the hypotheses H0: µ = 12 versus H1: µ ≠ 12, the P-value is 0.30. a. Should H0 be rejected on the basis of this test? Explain. b. Can you conclude that the machine is calibrated to provide a mean fill weight of 12 oz? Explain. 15. A method of applying zinc plating to steel is supposed to produce a coating whose mean thickness is no greater than 7 microns. A quality inspector measures the thickness of 36 coated specimens and tests H0: µ ≤ 7 versus H1: µ > 7. She obtains a P-value of 0.40. Since P > 0.05, she concludes that the mean thickness is within the specification. Is this conclusion correct? Explain. 16. A sample of size n = 100 is used to test H0: µ ≤ 20 versus H1: µ > 20. The value of µ will not have practical significance unless µ > 25. The population standard deviation is σ = 10. The value of is 21. a. Assume the sample size is n = 100. Compute the P-value. Show that you do not reject H0. b. c. d. Assume the sample size is n = 1000. Compute the P-value. Show that you reject H0. Do you think the difference is likely to be of practical significance? Explain. Explain why a larger sample can be more likely to produce a statistically significant result that is not of practical significance. 17. A new method of postoperative treatment was evaluated for patients undergoing a certain surgical procedure. Under the old method, the mean length of hospital stay was 6.3 days. The sample mean for the new method was 6.1 days. A hypothesis test was performed in which the null hypothesis stated that the mean length of stay for the new method was greater than or equal to the mean length for the old method, and the alternate hypothesis stated that the mean stay for the new method was lower. The P-value was 0.002. True or false: a. Because the P-value is very small, we can conclude that the new method provides an important reduction in the mean length of hospital stay. b. Because the P-value is very small, we can conclude that the mean length of hospital stay is less for patients treated by the new method than for patients treated by the old method. 18. Fill in the blank: A 95% confidence interval for µ is (1.2, 2.0). Based on the data from which the confidence interval was constructed, someone wants to test H0: µ = 1.4 versus H1: µ ≠ 1.4. The P-value will be _____. i. Greater than 0.05 ii. Less than 0.05 iii. Equal to 0.05 19. Refer to Exercise 18. For which null hypothesis will P = 0.05? i. H0: µ = 1.2 ii. H0: µ ≤ 1.2 iii. H0: µ ≥ 1.2 20. A scientist computes a 90% confidence interval to be (4.38, 6.02). Using the same data, she also computes a 95% confidence interval to be (4.22, 6.18), and a 99% confidence interval to be (3.91, 6.49). Now she wants to test H0: µ = 4 versus H1: µ ≠ 4. Regarding the P-value, which one of the following statements is true? i. P > 0.10 ii. 0.05 < P < 0.10 iii. 0.01 < P < 0.05 iv. P < 0.01 21. The strength of a certain type of rubber is tested by subjecting pieces of the rubber to an abrasion test. For the rubber to be acceptable, the mean weight loss µ must be less than 3.5 mg. A large number of pieces of rubber that were cured in a certain way were subject to the abrasion test. A 95% upper confidence bound for the mean weight loss was computed from these data to be 3.45 mg. Someone suggests using these data to test H0: µ ≥ 3.5 Page 420 versus H1: µ < 3.5. a. Is it possible to determine from the confidence bound whether P < 0.05? Explain. b. Is it possible to determine from the confidence bound whether P < 0.01? Explain. 22. A shipment of fibers is not acceptable if the mean breaking strength of the fibers is less than 50 N. A large sample of fibers from this shipment was tested, and a 98% lower confidence bound for the mean breaking strength was computed to be 50.1 N. Someone suggests using these data to test the hypotheses H0: µ ≤ 50 versus H1: µ > 50. a. Is it possible to determine from the confidence bound whether P < 0.01? Explain. b. Is it possible to determine from the confidence bound whether P < 0.05? Explain. 23. Refer to Exercise 21. It is discovered that the mean of the sample used to compute the confidence bound is . Is it possible to determine whether P < 0.01? Explain. 24. Refer to Exercise 22. It is discovered that the standard deviation of the sample used to compute the confidence interval is 5 N. Is it possible to determine whether P < 0.01? Explain. 25. The following MINITAB output (first shown in Exercise 16 in Section 6.1) presents the results of a hypothesis test for a population mean µ. One-Sample Z: X Test of mu = 73.5 vs not = 73.5 The assumed standard deviation = 2.3634 Variable N Mean StDev SE Mean 95% CI Z P X 145 73.2461 2.3634 0.1963 (72.8614, 73.6308) −1.29 0.196 a. Can H0 be rejected at the 5% level? How can you tell? b. Someone asks you whether the null hypothesis H0: µ = 73 versus H1: µ ≠ 73 can be rejected at the 5% level. Can you answer without doing any calculations? How? 6.3 Tests for a Population Proportion A population proportion is simply a population mean for a population of 0s and 1s: a Bernoulli population. For this reason, hypothesis tests for proportions are similar to those discussed in Section 6.1 for population means. Here is an example. A supplier of semiconductor wafers claims that of all the wafers he supplies, no more than 10% are defective. A sample of 400 wafers is tested, and 50 of them, or 12.5%, are defective. Can we conclude that the claim is false? The hypothesis test here proceeds much like those in Section 6.1. What makes this problem distinct is that the sample consists of successes and failures, with "success" indicating a defective wafer. If the population proportion of defective wafers is denoted by p, then the supplier's claim is that p ≤ 0.1. Since our hypothesis concerns a population proportion, it is natural to base the test on the sample proportion . Making the reasonable assumption that the wafers are Page 421 sampled independently, it follows from the Central Limit Theorem, since the sample size is large, that (6.1) where n is the sample size, equal to 400. We must define the null hypothesis. The question asked is whether the data allow us to conclude that the supplier's claim is false. Therefore, the supplier's claim, which is that p ≤ 0.1, must be H0. Otherwise it would be impossible to prove the claim false, no matter what the data showed. The null and alternate hypotheses are To perform the hypothesis test, we assume H0 to be true and take p = 0.1. Substituting p = 0.1 and n = 400 in expression (6.1) yields the null distribution of : The standard deviation of is = 0.125. The z-score of is . The observed value of is 50/400 is The z table indicates that the probability that a standard normal random variable has a value greater than 1.67 is approximately 0.0475. The P-value is therefore 0.0475 (see Figure 6.5). The null distribution of is N(0.1, 0.0152). Thus if H0 is true, the probability that takes on a value as extreme as or more extreme than the observed value of 0.125 is 0.0475. This is the Pvalue. FIGURE 6.5 What do we conclude about H0? Either the supplier's claim is false, or we have observed a sample that is as extreme as all but 4.75% of the samples we might have drawn. Such a sample would be unusual, but not fantastically unlikely. There is every reason to be quite skeptical of the claim, but we probably shouldn't convict the supplier quite yet. If possible, it would be a good idea to sample more wafers. Note that under the commonly used rule of thumb, we would reject H0 and condemn the supplier, because P is less than 0.05. This example illustrates the weakness of this rule. If you do the calculations, you will find that if only 49 of the sample wafers had been defective rather than 50, the P-value would have risen to 0.0668, and the supplier would be off the hook. Page 422 Thus the fate of the supplier hangs on the outcome of one single wafer out of 400. It doesn't make sense to draw such a sharp line. It's better just to report the P-value and wait for more evidence before reaching a firm conclusion. The Sample Size Must Be Large The test just described requires that the sample proportion be approximately normally distributed. This assumption will be justified whenever both np0 > 10 and n(1−p0) > 10, where p0 is the population proportion specified in the null distribution. Then the z-score can be used as the test statistic, making this a z test. Example 6.6 In courses on surveying, field work is an important part of the curriculum. The article "Enhancing Civil Engineering Surveying Learning through Workshops" (J. Awange, A. Faisal Anwar, et al., Journal of Surveying Engineering, 2017, online) reports that in a sample of 67 students studying surveying, 45 said that field work improved their ability to handle unforeseen problems. Can we conclude that more than 65% of students find that field work improves their ability to handle unforeseen problems? Solution Let p denote the probability that a randomly chosen student believes that field work improves one's ability to handle unforeseen problems. The null and alternate hypotheses are The sample proportion is . Under the null hypothesis, normally distributed with mean 0.65 and standard deviation is . The z-score is The P-value is 0.3557 (see Figure 6.6). We cannot conclude that more than 65% of students find that field work improves their ability to handle unforeseen problems. FIGURE 6.6 The null distribution of is N(0.65, 0.05832). Thus if H0 is true, the probability that takes on a value as extreme as or more extreme than the observed value of 0.6716 is 0.3557. This is the Pvalue. The following computer output (from MINITAB) presents the results from Page 423 Example 6.6. Test and CI for One Proportion: UNF Test of p = 0.75 vs p > 0.75 Variable UNF X N 45 67 Sample p 0.671642 95% Lower Bound 0.577272 Z-Value 0.37 P-Value 0.356 The output contains a 95% lower confidence bound as well as the P-value. Relationship with Confidence Intervals for a Proportion A level 100(1 − α)% confidence interval for a population mean µ contains those values for a parameter for which the P-value of a hypothesis test will be greater than α. For the confidence intervals for a proportion presented in Section 5.2 and the hypothesis test presented here, this statement is only approximately true. The reason for this is that the methods presented in Section 5.2 are slight modifications (that are much easier to compute) of a more complicated confidence interval method for which the statement is exactly true. Summary Let X be the number of successes in n independent Bernoulli trials, each with success probability p; in other words, let X ∼ Bin(n, p). To test a null hypothesis of the form H0: p ≤ p0, H0: p ≥ p0, or H0: p = p0, assuming that both np0 and n(1 − p0) are greater than 10: ■ Compute the z-score: ■ Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows: . Alternate Hypothesis P-value H1: p > p0 Area to the right of z H1: p < p0 Area to the left of z H1: p ≠ p0 Sum of the areas in the tails cut off by z and −z Exercises for Section 6.3 1. Integrated circuits consist of electric channels that are etched onto silicon wafers. A certain proportion of circuits are defective because of "undercutting," which occurs when too much material is etched away so that the channels, which consist of the unetched portions of the wafers, are too narrow. A redesigned process, involving lower pressure in the etching chamber, is being investigated. The goal is to reduce the rate of undercutting to less than 5%. Out of the first 1000 circuits manufactured by the new process, only 35 show evidence of undercutting. Can you conclude that the goal has been met? Page 424 2. The article "HIV-positive Smokers Considering Quitting: Differences by Race/Ethnicity" (E. Lloyd-Richardson, C. Stanton, et al., Am J Health Behav, 2008:3–15) surveyed 444 HIV-positive smokers. Of these, 281 were male and 163 were female. Consider this to be a simple random sample. Can you conclude that more than 60% of HIV-positive smokers are male? 3. Do bathroom scales tend to underestimate a person's true weight? A 150 lb test weight was placed on each of 50 bathroom scales. The readings on 29 of the scales were too light, and the readings on the other 21 were too heavy. Can you conclude that more than half of bathroom scales underestimate weight? 4. The article "Evaluation of Criteria for Setting Speed Limits on Gravel Roads" (S. Dissanayake, Journal of Transportation Engineering, 2011:57–63) measured speeds of vehicles on several roads in the state of Kansas. On South Cedar Niles, 73 vehicles were observed, and 49 of them were exceeding the speed limit. Can you conclude that more than 60% of the vehicles on South Cedar Niles exceed the speed limit? 5. In a survey of 500 residents in a certain town, 274 said they were opposed to constructing a new shopping mall. Can you conclude that more than half of the residents in this town are opposed to constructing a new shopping mall? 6. The anterior cruciate ligament (ACL) runs diagonally in the middle of the knee. The article "Return to Sport After Pediatric Anterior Cruciate Ligament Reconstruction and Its Effect on Subsequent Anterior Cruciate Ligament Injury" (T. Dekker, J. Godin, et al., Journal of Bone and Joint Surgery, 2017:897–904) reported results for 85 young athletes who suffered anterior cruciate ligament (ACL) injuries. Of the 85 injuries, 51 were to the left knee and 34 were to the right knee. Can you conclude that more than half of ACL injuries are to the left knee? 7. The article "Developing a Tool to Measure the Factors Influencing Nurses' Enjoyment of Nursing" (L. Wilkes, M. Doull, et al., Journal of Clinical Nursing, 2016:1854–1860) reports that in a sample of 124 nurses, 54 said that a busy workload had a positive effect on their enjoyment of their job. Can you conclude that less than 45% of nurses feel that a busy workload has a positive effect on their enjoyment of their job? 8. A grinding machine will be qualified for a particular task if it can be shown to produce less than 8% defective parts. In a random sample of 300 parts, 12 were defective. On the basis of these data, can the machine be qualified? 9. Let A and B represent two variants (alleles) of the DNA at a certain locus on the genome. Assume that 40% of all the alleles in a certain population are type A and 30% are type B. The locus is said to be in Hardy-Weinberg equilibrium if the proportion of organisms that are of type AB is (0.40)(0.30) = 0.12. In a sample of 300 organisms, 42 are of type AB. Can you conclude that this locus is not in Hardy-Weinberg equilibrium? 10. A sociologist sampled 200 people who work in computer-related jobs, and found that 42 of them have changed jobs in the past six months. Can you conclude that more than 15% of computer-related workers have changed jobs in the past six months? 11. The General Social Survey asked a sample of 1294 people whether they performed any volunteer work during the past year. A total of 517 people said they did. Can you conclude that more than 35% of adults performed volunteer work during the past year? 12. The following MINITAB output presents the results of a hypothesis test for a population proportion p. Test and CI for One Proportion: X Test of p = 0.4 vs p < 0.4 Variable X N Sample p X 73 240 0.304167 Page 425 95% Upper Bound Z-Value P-Value 0.353013 −3.03 0.001 a. b. c. Is this a one-tailed or two-tailed test? What is the null hypothesis? Can H0 be rejected at the 2% level? How can you tell? d. Someone asks you whether the null hypothesis H0: p ≥ 0.45 versus H1: p < 0.45 can be rejected at the 2% level. Can you answer without doing any calculations? How? Use the output and an appropriate table to compute the P-value for the test of H0: p ≤ 0.25 versus H1: p > 0.25. e. f. Use the output and an appropriate table to compute a 90% confidence interval for p. 13. The following MINITAB output presents the results of a hypothesis test for a population proportion p. Some of the numbers are missing. Fill in the numbers for (a) through (c). Test and CI for One Proportion: X Test of p = 0.7 vs p < 0.7 Variable X N Sample p X 345 500 (a) (b) 95% Upper Bound Z-Value P-Value 0.724021 (b) (c) 6.4 Small-Sample Tests for a Population Mean In Section 6.1, we described a method for testing a hypothesis about a population mean, based on a large sample. A key step in the method is to approximate the population standard deviation σ with the sample standard deviation s. The normal curve is then used to find the P-value. When the sample size is small, s may not be close to σ, which invalidates this large-sample method. However, when the population is approximately normal, the Student's t distribution can be used. We illustrate with an example. Spacer collars for a transmission countershaft have a thickness specification of 38.98–39.02 mm. The process that manufactures the collars is supposed to be calibrated so that the mean thickness is 39.00 mm, which is in the center of the specification window. A sample of six collars is drawn and measured for thickness. The six thicknesses are 39.030, 38.997, 39.012, 39.008, 39.019, and 39.002. Assume that the population of thicknesses of the collars is approximately normal. Can we conclude that the process needs recalibration? Denoting the population mean by μ, the null and alternate hypotheses are Note that H0 specifies a single value for μ, since calibration requires that the mean be equal to the correct value. To construct the test statistic, note that since the population is assumed to follow a normal distribution, the quantity has a Student's t distribution with n − 1 = 5 degrees of freedom. This is the test statistic. Page 426 In this example the observed values of the sample mean and standard deviation are and s = 0.011928. The sample size is n = 6. The null hypothesis specifies that μ = 39. The value of the test statistic is therefore The P-value is the probability of observing a value of the test statistic whose disagreement with H0 is as great as or greater than that actually observed. Since H0 specifies that μ = 39.00, this is a two-tailed test, so values both above and below 39.00 disagree with H0. Therefore the P-value is the sum of the areas under the curve corresponding to t > 2.327 and t < −2.327. Figure 6.7 illustrates the null distribution and indicates the location of the test statistic. From the t table (Table A.3 in Appendix A) the row corresponding to 5 degrees of freedom indicates that the value t = ±2.015 cuts off an area of 0.05 in each tail, for a total of 0.10, and that the value t = ±2.571 cuts off an area of 0.025 in each tail, for a total of 0.05. Thus the P-value is between 0.05 and 0.10. (Software yields P = 0.0675.) While we cannot conclusively state that the process is out of calibration, it doesn't look too good. It would be prudent to recalibrate. FIGURE 6.7 The null distribution of is Student's t with five degrees of freedom. The observed value of t, corresponding to the observed values and s = 0.011928, is 2.327. If H0 is true, the probability that t takes on a value as extreme as or more extreme than that observed is between 0.05 and 0.10. Because H0 specified that μ was equal to a specific value, both tails of the curve contribute to the P-value. In this example, the test statistic was a t statistic rather than a z-score. For this reason, this test is referred to as a t test. Example 6.7 The article "Geographically Weighted Regression-Based Methods for Merging Satellite and Gauge Precipitation" (L. Chao, K. Zhang, et al., Journal of Hydrology, 2018:275–289) describes a method of measuring precipitation by combining satellite measurements with measurements made on the ground. Errors were recorded monthly for 16 months, and averaged 0.393 mm with a standard deviation of 0.368 mm. Can we conclude that the mean error is less than 0.6 mm? Solution Let μ denote the mean error. The null and alternate hypotheses are Page 427 Under H0, the test statistic has a Student's t distribution with 15 degrees of freedom. Substituting , s = 0.368, and n = 16, the value of the test statistic is Consulting the t-table, we find that the value t = −2.131 cuts off an area of 0.025 in the lefthand tail, and the value t = −2.602 cuts off an area of 0.01 (see Figure 6.8). We conclude that the P-value is between 0.01 and 0.025. (Software yields P = 0.0199.) FIGURE 6.8 Solution for Example 6.7. The null distribution is Student's t with 15 degrees of freedom. The observed value of t is −2.25. If H0 is true, the probability that t takes on a value as extreme as or more extreme than that observed is between 0.01 and 0.025. Software yields a value of 0.0199. The following computer output (from MINITAB) presents the results from Example 6.7. One-Sample T: Error Test of mu = 0.6 vs < 0.6 Variable N Mean StDev SE Mean 95% Upper Bound T P Error 16 0.393 0.368 0.0920 0.55428 −2.25 0.020 Note that the upper 95% confidence bound provided in the output is consistent with the alternate hypothesis. This indicates that the P-value is less than 5%. Use z, Not t, If σ Is Known Occasionally a small sample may be taken from a normal population whose standard deviation σ is known. In these cases, we do not use the Student's t curve, because we are not Page 428 approximating σ with s. Instead, we use the z table and perform a z test. Example 6.8 illustrates the method. Example 6.8 At the beginning of this section, we described a sample of six spacer collars, whose thicknesses (in mm) were 39.030, 38.997, 39.012, 39.008, 39.019, and 39.002. We denoted the population mean thickness by μ and tested the hypotheses Now assume that these six spacer collars were manufactured just after the machine that produces them had been moved to a new location. Assume that on the basis of a very large number of collars manufactured before the move, the population of collar thicknesses is known to be very close to normal, with standard deviation σ = 0.010 mm, and that it is reasonable to assume that the move has not changed this. On the basis of the given data, can we reject H0? We compute . We do not need the value of s, because we know that σ = 0.010. Since the population is normal, is normal even though the sample size is small. The null distribution is therefore The z-score is The P-value is 0.0054, so H0 can be rejected. Summary Let X1, ..., Xn be a sample from a normal population with mean μ and standard deviation σ, where σ is unknown. To test a null hypothesis of the form H0: μ ≤ μ0, H0: μ ≥ μ0, or H0: μ = μ0: ■ Compute the test statistic . ■ Compute the P-value. The P-value is an area under the Student's t curve with n − 1 degrees of freedom, which depends on the alternate hypothesis as follows: Alternate Hypothesis P-value H1: μ > μ0 Area to the right of t H1: μ < μ0 Area to the left of t H1: μ ≠ μ0 Sum of the areas in the tails cut off by t and −t ■ If σ is known, the test statistic is Exercises for Section 6.4 1. 2. 3. 4. , and a z test should be performed. Page 429 Each of the following hypothetical data sets represents some repeated weighings of a standard weight that is known to have a mass of 100 g. Assume that the readings are a random sample from a population that follows the normal curve. Perform a t test to see whether the scale is properly calibrated, if possible. If impossible, explain why. a. 100.02, 99.98, 100.03 b. 100.01 A geologist is making repeated measurements (in grams) on the mass of a rock. It is not known whether the measurements are a random sample from an approximately normal population. Below are three sets of replicate measurements, listed in the order they were made. For each set of readings, state whether the assumptions necessary for the validity of the t test appear to be met. If the assumptions are not met, explain why. a. 213.03 212.95 213.04 213.00 212.99 213.01 221.03 213.05 b. 213.05 213.00 212.94 213.09 212.98 213.02 213.06 212.99 c. 212.92 212.95 212.97 213.00 213.01 213.04 213.05 213.06 A new centrifugal pump is being considered for an application involving the pumping of ammonia. The specification is that the flow rate be more than 5 gallons per minute (gpm). In an initial study, eight runs were made. The average flow rate was 6.5 gpm and the standard deviation was 1.9 gpm. If the mean flow rate is found to meet the specification, the pump will be put into service. a. State the appropriate null and alternate hypotheses. b. Find the P-value. c. Should the pump be put into service? Explain. A certain manufactured product is supposed to contain 23% potassium by weight. A sample of 10 specimens of this product had an average percentage of 23.2 with a standard deviation of 0.2. If the mean percentage is found to differ from 23, the manufacturing process will be recalibrated. 5. 6. a. State the appropriate null and alternate hypotheses. b. Compute the P-value. c. Should the process be recalibrated? Explain. The article "Influence of Penetration Rate on Penetrometer Resistance" (G. Gagnon and J. Doubrough, Canadian Journal of Civil Engineering, 2011:741–750) describes a study in which twenty 2-L specimens of water drawn from a public works building in Bridgewater, Nova Scotia. The mean lead concentration was 6.7 μg/L with a standard deviation of 3.9 μg/L. a. The Health Canada guideline states that the concentration should be less than 10 5.5 Variable N Mean StDev SE Mean 95% Lower Bound T P X 5 5.92563 0.15755 0.07046 5.77542 6.04 0.002 a. b. c. Is this a one-tailed or two-tailed test? What is the null hypothesis? Can H0 be rejected at the 1% level? How can you tell? d. Use the output and an appropriate table to compute the P-value for the test of H0: μ ≥ 6.5 versus H1: μ < 6.5. e. Use the output and an appropriate table to compute a 99% confidence interval for μ. 13. The following MINITAB output presents the results of a hypothesis test for a population mean μ. Some of the numbers are missing. Fill in the numbers. One-Sample T: X Test of mu = 16 vs not = 16 Variable X N Mean 11 13.2874 StDev (a) SE Mean 1.8389 95% CI T P (b), (c) (d) 0.171 Page 431 6.5 Large-Sample Tests for the Difference Between Two Means We now investigate examples in which we wish to determine whether the means of two populations are equal. The data will consist of two samples, one from each population. The basic idea is quite simple. We will compute the difference of the sample means. If the difference is far from 0, we will conclude that the population means are different. If the difference is close to 0, we will conclude that the population means might be the same. As an example, suppose that a production manager for a manufacturer of industrial machinery is concerned that ball bearings produced in environments with low ambient temperatures may have smaller diameters than those produced under higher temperatures. To investigate this concern, she samples 120 ball bearings that were manufactured early in the morning, before the shop was fully heated, and finds their mean diameter to be 5.068 mm and their standard deviation to be 0.011 mm. She independently samples 65 ball bearings manufactured during the afternoon and finds their mean diameter to be 5.072 mm and their standard deviation to be 0.007 mm. Can she conclude that ball bearings manufactured in the morning have smaller diameters, on average, than ball bearings manufactured in the afternoon? We begin by translating the problem into statistical language. We have a simple random sample X1, ..., X120 of diameters of ball bearings manufactured in the morning, and another simple random sample Y1, ..., Y65 of diameters of ball bearings manufactured in the afternoon. Denote the population mean of diameters of bearings manufactured in the morning by μX, and the population mean of diameters of bearings manufactured in the afternoon by μY. Denote the corresponding standard deviations by σX and σY. These population means and standard deviations are unknown. The sample sizes are nX = 120 and nY = 65. We are interested in the difference μX − μY. We must now determine the null and alternate hypotheses. The question asked is whether we can conclude that the population mean for the morning bearings is less than that for the afternoon bearings. Therefore the null and alternate hypotheses are The test is based on . Since both sample sizes are large, and are both approximately normally distributed. Since the samples are independent, it follows that the null distribution of is (6.2) The observed values are for the sample means, and sX = 0.011 and sY = 0.007 for the sample standard deviations. Under H0, μX − μY = 0 (the value closest to H1). We approximate the population variances and with the sample variances and , respectively, and substitute nX = 120 and nY = 65 to compute the standard deviation of the null distribution, obtaining . The null distribution of is therefore The observed value of is 5.068 − 5.072 = −0.004. The z-score is The P-value is 0.0013. Figure 6.9 shows the null distribution and the location of the test statistic. The P-value is 0.0013, so the manager's suspicion is correct. The bearings manufactured in the morning have a smaller mean diameter. FIGURE 6.9 The null distribution of is N(0, 0.0013272). Thus if H0 is true, the probability that takes on a value as extreme as or more extreme than the observed value of −0.004 is 0.0013. This is the P-value. Note that we used the assumption that the samples were independent when computing the variance of . This is one condition that is usually easy to achieve in practice. Unless there is some fairly obvious connection between the items in the two samples, it is usually reasonable to assume that the samples are independent. Example 6.9 The article "Effect of Welding Procedure on Flux Cored Steel Wire Deposits" (N. Ramini de Rissone, I. de S. Bott, et al., Science and Technology of Welding and Joining, 2003:113–122) compares properties of welds made using carbon dioxide as a shielding gas with those of welds made using a mixture of argon and carbon dioxide. One property studied was the diameter of inclusions, which are particles embedded in the weld. A sample of 544 inclusions in welds made using argon shielding averaged 0.37 μm in diameter, with a standard deviation of 0.25 μm. A sample of 581 inclusions in welds made using carbon dioxide shielding averaged 0.40 μm in diameter, with a standard deviation of 0.26 μm. (Standard deviations were estimated from a graph.) Can you conclude that the mean diameters of inclusions differ between the two shielding gases? Solution Let denote the sample mean diameter for argon welds. Then sX = 0.25 and the Let sample size is nX = 544. Let denote the sample mean diameter for carbon dioxide welds. Then sY = 0.26 and the sample size is nY = 581. Let μX denote the population mean diameter for argon welds, and let μY denote the population mean diameter for carbon Page 433 dioxide welds. The null and alternate hypotheses are We have observed . This value was drawn from a normal population with mean μX − μY, and variance approximated by . Under H0, we take μX − μY = 0. Substituting values of sX, sY, nX, and nY, the standard deviation is . The null distribution of is therefore The z-score is This is a two-tailed test, and the P-value is 0.0488 (see Figure 6.10). A follower of the 5% rule would reject the null hypothesis. It is certainly reasonable to be skeptical about the truth of H0. FIGURE 6.10 Solution to Example 6.9. The following computer output (from MINITAB) presents the results of Example 6.9. Two-sample T for Argon vs CO2 Argon CO2 N 544 581 Mean 0.37 0.40 StDev 0.25 0.26 SE Mean 0.010719 0.010787 Difference = mu (Argon) − mu (CO2) Estimate for difference: 0.030000 95% confidence bound for difference: (−0.0598366, −0.000163) T-Test of difference = 0 (vs not = 0): T-Value = −1.97 P-Value = 0.049 DF = 1122 Note that the computer uses the t statistic rather than the z statistic for this test. Many computer packages use the t statistic whenever a sample standard deviation is used to estimate a population standard deviation. When the sample size is large, the difference between t and z is negligible for practical purposes. When using tables rather than a computer, the z-score has the Page 434 advantage that the P-value can be determined with greater precision with a z table than with a t table. The methods described in this section can be used to test a hypothesis that two population means differ by a specified constant. Example 6.10 shows how. Example 6.10 Refer to Example 6.9. Can you conclude that the mean diameter for carbon dioxide welds (μY) exceeds that for argon welds (μX) by more than 0.015 μm? Solution The null and alternate hypotheses are , sX = 0.25, sY = 0.26, nX = 544, and nY = 581. Under H0, we take μX − μY = −0.015. The null distribution of is given by expression (6.2) to be . The z-score is We observe This is a one-tailed test. The P-value is 0.1611. We cannot conclude that the mean diameter of inclusions from carbon dioxide welds exceeds that of argon welds by more than 0.015 μm. Summary Let and be large (e.g., nX > 30 and nY > 30) samples from populations with means μX and μY and standard deviations σX and σY, respectively. Assume the samples are drawn independently of each other. To test a null hypothesis of the form H0: μX − μY ≤ Δ0, H0: μX − μY ≥ Δ0, or H0: μX − μY = Δ0: ■ Compute the z-score: . If σX and σY are unknown they may be approximated with sX and sY, respectively. ■ Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows: Alternate Hypothesis P-value Hypothesis H1: μX − μY > Δ0 Area to the right of z Δ0 H1: μX − μY < Δ0 Area to the left of z Δ0 H1: μX − μY ≠ Sum of the areas in the tails cut off by z and −z Page 435 Exercises for Section 6.5 1. 2. 3. 4. 5. 6. The article "Capillary Leak Syndrome in Children with C4A-Deficiency Undergoing Cardiac Surgery with Cardiopulmonary Bypass: A Double-Blind, Randomised Controlled Study" (S. Zhang, S. Wang, et al., Lancet, 2005:556–562) presents the results of a study of the effectiveness of giving blood plasma containing complement component C4A to pediatric cardiopulmonary bypass patients. Of 58 patients receiving C4A-rich plasma, the average length of hospital stay was 8.5 days and the standard deviation was 1.9 days. Of 58 patients receiving C4A-free plasma, the average length of hospital stay was 11.9 days and the standard deviation was 3.6 days. Can you conclude that the mean hospital stay is shorter for patients receiving C4A-rich plasma? The article "Some Parameters of the Population Biology of Spotted Flounder (Ciutharus linguatula Linnaeus, 1758) in Edremit Bay (North Aegean Sea)" (D. Türker, B. Bayhan, et al., Turkish Journal of Veterinary and Animal Science, 2005:1013–1018) reports that a sample of 482 female spotted flounder had an average weight of 20.95 g with a standard deviation of 14.5 g, and a sample of 614 male spotted flounder had an average weight of 22.79 g with a standard deviation of 15.6 g. Can you conclude that the mean weight of male spotted flounder is greater than that of females? The article "Utility of Pretest Probability and Exercise Treadmill Test in Korean Women with Suspected Coronary Artery Disease" (Y. Kim, W. Shim, et al., Journal of Women's Health, 2015:617–622) presents a study of health outcomes in women with symptoms of heart disease. In a sample of 110 women whose test results suggested the presence of coronary artery disease, the mean peak systolic blood pressure was 169.9 mmHg, with a standard deviation of 24.8 mmHg. In a sample of 225 women whose test results suggested an absence of coronary artery disease, the mean peak systolic blood pressure was 163.3 mmHg, with a standard deviation of 25.8 mmHg. Can you conclude that the mean peak systolic blood pressure differs between these two groups of women? The article "Wired: Energy Drinks, Jock Identity, Masculine Norms, and Risk Taking" (K. Miller, Journal of American College Health, 2008:481–489) reports that in a sample of 413 male college students, the average number of energy drinks consumed per month was 2.49 with a standard deviation of 4.87, and in a sample of 382 female college students, the average was 1.22 with a standard deviation of 3.24. Can you conclude that the mean number of energy drinks is greater for male students than for female students? Do you prefer taking tests on paper or online? A college instructor gave identical tests to two randomly sampled groups of 50 students. One group took the test on paper and the other took it online. Those who took the test on paper had an average score of 68.4 with a standard deviation of 12.1. Those who took the test online had an average score of 71.3 with a standard deviation of 14.2. Can you conclude that the mean scores differ between online and paper tests? Two machines used to fill soft drink containers are being compared. The number of containers filled each minute is counted for 60 minutes for each machine. During the 60 minutes, machine 1 filled an average of 73.8 cans per minute with a standard deviation of 5.2 cans per minute, and machine 2 filled an average of 76.1 cans per minute with a standard deviation of 4.1 cans per minute. a. If the counts are made each minute for 60 consecutive minutes, what assumption necessary to the validity of a hypothesis test may be violated? b. Assuming that all necessary assumptions are met, perform a hypothesis test. Can you conclude that machine 2 is faster than machine 1? 7. Statistics instructor who teaches a lecture section of 160 students wants to determine whether students have more difficulty with one-tailed hypothesis tests or with two-tailed hypothesis tests. On the next exam, 80 of the students, chosen at random, get a version of the exam with a 10-point question that requires a one-tailed test. The other 80 students get a question that is identical except that it requires a two-tailed test. The one-tailed students average 7.79 points, and their standard deviation is 1.06 points. The two-tailed students average 7.64 points, and their standard deviation is 1.31 points. a. Can you conclude that the mean score μ1 on one-tailed hypothesis test questions is higher than the mean score μ2 on two-tailed hypothesis test questions? State the appropriate null and alternate hypotheses, and then compute the P-value. b. Can you conclude that the mean score μ1 on one-tailed hypothesis test questions differs from the mean score μ2 on two-tailed hypothesis test questions? State the Page 436 appropriate null and alternate hypotheses, and then compute the P-value. 8. Fifty specimens of a new computer chip were tested for speed in a certain application, along with 50 specimens of chips with the old design. The average speed, in MHz, for the new chips was 495.6, and the standard deviation was 19.4. The average speed for the old chips was 481.2, and the standard deviation was 14.3. a. Can you conclude that the mean speed for the new chips is greater than that of the old chips? State the appropriate null and alternate hypotheses, and then find the P-value. b. A sample of 60 even older chips had an average speed of 391.2 MHz with a standard deviation of 17.2 MHz. Someone claims that the new chips average more than 100 MHz faster than these very old ones. Do the data provide convincing evidence for this claim? State the appropriate null and alternate hypotheses, and then find the P-value. 9. Are low-fat diets or low-carb diets more effective for weight loss? This question was addressed in the article "Comparison of the Atkins, Zone, Ornish, and LEARN Diets for Change in Weight and Related Risk Factors Among Overweight Premenopausal Women: The A TO Z Weight Loss Study: A Randomized Trial" (C. Gardner, A. Kiazand, et al., Journal of the American Medical Association, 2007:969–977). A sample of 77 subjects went on a low-carbohydrate diet for six months. At the end of that time the sample mean weight loss was 4.7 kg with a sample standard deviation of 7.2 kg. A second sample of 79 subjects went on a low-fat diet. Their sample mean weight loss was 2.6 kg with a standard deviation of 5.9 kg. a. Can you conclude that the mean weight loss is greater for those on the lowcarbohydrate diet? b. Can you conclude that the mean weight loss on the low-carbohydrate diet is more than 1 kg greater than that of the low-fat diet? 10. In a certain supermarket, a sample of 60 customers who used a self-service checkout lane averaged 5.2 minutes of checkout time, with a standard deviation of 3.1 minutes. A sample of 72 customers who used a cashier averaged 6.1 minutes with a standard deviation of 2.8 minutes. a. Can you conclude that the mean checkout time is less for people who use the selfservice lane? b. Can you conclude that if everyone used the selfservice lane, the mean checkout time would decrease? Consider the number of items checked out when formulating your answer. 11. The article "Treadmill Exercise and Resistance Training in Patients With Peripheral Arterial Disease With and Without Intermittent Claudication. A Randomized Controlled Trial" (M. McDermott, P. Ades, et al., Journal of the American Medical Association, 2009:165–174) reported the results of a study to determine whether treadmill exercise could improve the walking ability of patients suffering from claudication, which is pain caused by insufficient blood flow to the muscles of the legs. A sample of 48 patients walked on a treadmill for six minutes every day. After six months, the mean distance walked in six minutes was 348 meters, with a standard deviation of 80 m. For a control group of 46 patients who did not walk on a treadmill, the mean distance was 309 m with a standard deviation of 89 m. Can you conclude that the mean distance walked for patients using a treadmill is greater than the mean for the controls? Use the α = 0.05 level of significance. 12. The following MINITAB output presents the results of a hypothesis test for the difference μX − μY between two population means: Two-sample T for X vs Y X Y N 135 180 Mean 3.94 4.43 StDev 2.65 2.38 Page 437 SE Mean 0.23 0.18 Difference = mu (X) − mu (Y) Estimate for difference: −0.484442 95% upper bound for difference: −0.007380 T-Test of difference = 0 (vs 0): Z = 1.13 P-Value = 0.129 Summary Let X ∼ Bin(nX, pX) and let Y ∼ Bin(nY, pY). Assume that there are at least 10 successes and 10 failures in each sample, and that X and Y are independent. To test a null hypothesis of the form H0: pX − pY ≤ 0, H0: pX − pY ≥ 0, or H0: pX − pY = 0: ■ Compute , and . ■ Compute the z-score: ■ Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows: . Alternate P-value Hypothesis H1: pX − pY > Area to the right of z 0 H1: pX − pY < Area to the left of z 0 H1: pX − pY ≠ Sum of the areas in the tails cut off by z and −z 0 Page 440 Exercises for Section 6.1 1. 2. 3. 4. Two extrusion machines that manufacture steel rods are being compared. In a Page 441 sample of 1000 rods taken from machine 1, 960 met specifications regarding length and diameter. In a sample of 600 rods taken from machine 2, 582 met the specifications. Machine 2 is more expensive to run, so it is decided that machine 1 will be used unless it can be convincingly shown that machine 2 produces a larger proportion of rods meeting specifications. a. State the appropriate null and alternate hypotheses for making the decision as to which machine to use. b. Compute the P-value. c. Which machine should be used? Resistors labeled as 100 Ω are purchased from two different vendors. The specification for this type of resistor is that its actual resistance be within 5% of its labeled resistance. In a sample of 180 resistors from vendor A, 150 of them met the specification. In a sample of 270 resistors purchased from vendor B, 233 of them met the specification. Vendor A is the current supplier, but if the data demonstrate convincingly that a greater proportion of the resistors from vendor B meet the specification, a change will be made. a. State the appropriate null and alternate hypotheses. b. Find the P-value. c. Should a change be made? The article "A Music Key Detection Method Based on Pitch Class Distribution Theory" (J. Sun, H. Li, and L. Ma, International Journal of Knowledge-based and Intelligent Engineering Systems, 2011:165–175) describes a method of analyzing digital music files to determine the key in which the music is written. In a sample of 307 pop music selections, the key was identified correctly in 245 of them. In a sample of 347 new-age selections, the key was identified correctly in 304 of them. Can you conclude that the method is more accurate for new-age songs than for pop songs? When the light turns yellow, should you stop or go through it? The article "Evaluation of Driver Behavior in Type II Dilemma Zones at High-Speed Signalized Intersections" (D. Hurwitz, M. Knodler, and B. Nyquist, Journal of Transportation Engineering, 2011:277– 286) defines the "indecision zone" as the period when a vehicle is between 2.5 and 5.5 seconds away from an intersection. At the intersection of Route 7 and North Shrewsbury in Clarendon, Vermont, 154 vehicles were observed to encounter a yellow light in the indecision zone, and 21 of them ran the red light. At the intersection of Route 62 and Paine Turnpike in Berlin, Vermont, 183 vehicles entered the intersection in the indecision zone, and 20 ran the red light. Can you conclude that the proportion of redlight runners differs between the two intersections? 5. The article "HIV-positive Smokers Considering Quitting: Differences by Race/Ethnicity" (E. Lloyd-Richardson, C. Stanton, et al., Am J Health Behav, 2008:3–15) reported that in a group of 230 European-American HIV-positive smokers, 102 of them had used a nicotine patch to try to quit smoking, and in a group of 72 Hispanic-American HIV-positive smokers, 20 had used a nicotine patch. Can you conclude that the proportion of patch users is greater among European-Americans? 6. In August and September 2005, Hurricanes Katrina and Rita caused extraordinary flooding in New Orleans, Louisiana. Many homes were severely damaged or destroyed; of those that survived, many required extensive cleaning. It was thought that cleaning flood-damaged homes might present a health hazard due to the large amounts of mold present in many of the homes. The article "Health Effects of Exposure to Water-Damaged New Orleans Homes Six Months After Hurricanes Katrina and Rita" (K. Cummings, J. Cox-Ganser, et al., American Journal of Public Health, 2008:869–875) reports that in a sample of 365 residents of Orleans Parish who had participated in the cleaning of one or more homes, 77 had experienced symptoms of wheezing, and in a sample of 179 residents who had not participated in cleaning, 23 reported wheezing symptoms (numbers read from a graph). Can you conclude that the frequency of wheezing symptoms is greater among those residents who participated in the cleaning of flood-damaged homes? 7. To test the effectiveness of protective packaging, a firm shipped out 1200 orders in regular light packaging and 1500 orders in heavy-duty packaging. Of the orders shipped in light packaging, 20 arrived in damaged condition, while of the orders shipped in heavy-duty packaging, 15 arrived in damaged condition. Can you conclude that heavy-duty Page 442 packaging reduces the proportion of damaged shipments? 8. Colonoscopy is a medical procedure that is designed to find and remove precancerous lesions in the colon before they become cancerous. The article "Association of Colonoscopy and Death from Colorectal Cancer" (N. Baxter, M. Goldwasser, et al., Annals of Internal Medicine, 2009:1–8) reports that in a sample of 10,292 people diagnosed with colorectal cancer, 7.0% had previously had a colonoscopy, and in a sample of 51,460 people without colorectal cancer, 9.8% had previously had a colonoscopy. Can you conclude that the percentage of people who have had colonoscopies is greater in those without colorectal cancer? 9. The article "Factors Associated with Exercise Behavior in People with Parkinson Disease" (T. Ellis, J. Cavanaugh, et al., Physical Therapy, 2011:1838–1848) reported a survey of patients with Parkinson's disease. Of the 245 patients who said that they exercised regularly, 76 reported falling in the previous six months. Of 96 patients who said they did not exercise regularly, 48 reported falling in the previous six months. Can you conclude that the proportion of patients who fall is less for those who exercise than for those who do not? 10. The article "Bioremediation of Soil Contaminated Crude Oil by Agaricomycetes" (Mohammad-Sichani, Assadi, et al., Journal of Environmental Health and Engineering, 2017, online) describes an experiment to determine the effectiveness of mushroom compost in removing petroleum contaminants from soil. Out of 150 seeds planted in soil containing 3% mushroom compost by weight, 74 germinated. Out of 150 seeds planted in soil 11. 12. 13. 14. containing 5% mushroom compost by weight, 86 germinated. Can you conclude that the proportion of seeds that germinate differs with the percent of mushroom compost in the soil? The article "Weekly Personal Ozone Exposure and Respiratory Health in a Panel of Greek Schoolchildren" (A. Karakatsani, E. Samoli, et al., Environmental Health Perspectives, 2017, online) describes a study of the health effects of ozone air pollution on children's health. Out of 70 children who lived in areas with low ozone levels, 30 missed at least one day of school during a five-week period. Out of 118 children who lived in areas with high ozone levels, 53 missed at least one day of school during the same five-week period. Can you conclude that children who live in high ozone areas are more likely to miss a day of school? In a study conducted by the U.S. Department of Health and Human Services, a sample of 546 boys aged 6-11 was weighed, and it was determined that 87 of them were overweight. A sample of 508 girls aged 6-11 was also weighed, and 74 of them were overweight. Can you conclude that the proportion of boys who are overweight differs from the proportion of girls who are overweight? In a sample of 1000 automobiles driven at high altitude, 17% exceeded a standard of 10 grams of particulate pollution per gallon of fuel consumed. The same cars were driven at sea level, and 11% of them exceeded the standard. Let p1 be the proportion of automobiles that exceed the standard at high altitude, and let p2 be the proportion that exceed the standard at sea level. Can the statistic be used to test H0: p1 − p2 = 0 versus H1: p1 − p2 ≠ 0? If so, perform the test and compute the P-value. If not, explain why not. The following MINITAB output presents the results of a hypothesis test for the difference p1 − p2 between two population proportions. Test and CI for Two Proportions Sample 1 2 X 41 37 N 97 61 Sample p 0.422680 0.606557 Difference = p (1) − p (2) Estimate for difference: −0.183877 95% CI for difference: (−0.341016, −0.026738) Test for difference = 0 (vs not = 0): Z = −2.25 P-Value = 0.024 a. b. c. Page 443 Is this a one-tailed or two-tailed test? What is the null hypothesis? Can H0 be rejected at the 5% level? How can you tell? 15. The following MINITAB output presents the results of a hypothesis test for the difference p1 − p2 between two population proportions. Some of the numbers are missing. Fill in the numbers for (a) through (d). Test and CI for Two Proportions Sample 1 2 X 101 (b) N 153 90 Sample p (a) 0.544444 Difference = p (1) − p (2) Estimate for difference: 0.115686 95% CI for difference: (−0.0116695, 0.243042) Test for difference = 0 (vs not = 0): Z = (c) P-Value = (d) 6.7 Small-Sample Tests for the Difference Between Two Means The t test can be used in some cases where samples are small, and thus where the Central Limit Theorem does not apply. We present an example. The article "The Achondroplasia Paternal Age Effect Is Not Explained By An Increase in Mutant Frequency" (I. Tiemann-Boege, W. Navidi, et al., Proceedings of the National Academy of Sciences, 2002:14952–14957) describes an experiment in which a number of DNA molecules is counted, and it needs to be determined whether these molecules contain a certain sequence of nucleotides. This is done by repeating the experiment with an added enzyme that digests the sequence of interest. If the mean count is lower with the enzyme present, then it can be concluded that the molecules being counted do indeed contain the sequence. Assume that in six identically prepared specimens without the enzyme present, the numbers of molecules counted are 33, 30, 26, 22, 37, and 34. Assume that in four identically prepared specimens with the enzyme present, the counts were 22, 29, 25, and 23. Can we conclude that the counts are lower when the enzyme is present? We have only a few observations for each process, so the Central Limit Theorem does not apply. If both populations are approximately normal, the Student's t distribution can be used to construct a hypothesis test. Let X1, ..., X6 represent the counts obtained without the enzyme, and let Y1, ..., Y4 represent the counts obtained with the enzyme. Let μX and μY be the means of the populations from which these samples are drawn; let nX and nY denote the sample sizes. The null and alternate hypotheses are Page 444 By assumption, both populations follow normal distributions. Therefore (as discussed in Section 5.6) the quantity (6.5) has an approximate Student's t distribution with ν degrees of freedom, where The observed values for the sample means and standard deviations are , sX = 5.538, and sY = 3.096. The sample sizes are nX = 6 and nY = 4. Substituting the values for the sample standard deviations and sample sizes, we compute ν = 7.89, which we round down to 7. Under H0, μX − μY = 0. The test statistic is therefore Under H0, the test statistic has approximately a Student's t distribution with seven degrees of freedom. Substituting values for , sX, sY, nX, and nY, we compute the value of the test statistic to be Consulting the t table with seven degrees of freedom, we find that the value cutting off 5% in the right-hand tail is 1.895, and the value cutting off 2.5% is 2.365. The P-value is therefore between 0.025 and 0.05 (see Figure 6.13). We conclude that the mean count is lower when the enzyme is present. FIGURE 6.13 The null distribution is Student's t with seven degrees of freedom. The observed value of the test statistic is 2.038. If H0 is true, the probability that t takes on a value as extreme as or more extreme than that observed is between 2.5% and 5%. Example 6.12 Good website design can make Web navigation easier. The article "The Implications of Visualization Ability and Structure Preview Design for Web Information Search Page 445 Tasks" (H. Zhang and G. Salvendy, International Journal of Human-Computer Interaction, 2001:75–95) presents a comparison of item recognition between two designs. A sample of 10 users using a conventional Web design averaged 32.3 items identified, with a standard deviation of 8.56. A sample of 10 users using a new structured Web design averaged 44.1 items identified, with a standard deviation of 10.09. Can we conclude that the mean number of items identified is greater with the new structured design? Solution Let be the sample mean for the structured Web design. Then sX = 10.09 and nX = 10. Let be the sample mean for the conventional Web design. Then sY = 8.56 and nY = 10. Let μX and μY denote the population mean measurements made by the structured and conventional methods, respectively. The null and alternate hypotheses are The test statistic is Substituting values for , sX, sY, nX, and nY, we compute the value of the test statistic to be t = 2.820. Under H0, this statistic has an approximate Student's t distribution, with the number of degrees of freedom given by Consulting the t table with 17 degrees of freedom, we find that the value cutting off 1% in the right-hand tail is 2.567, and the value cutting off 0.5% in the right-hand tail is 2.898. Therefore the area in the right-hand tail corresponding to values as extreme as or more extreme than the observed value of 2.820 is between 0.005 and 0.010. Therefore 0.005 < P < 0.01 (see Figure 6.14). There is strong evidence that the mean number of items identified is greater for the new design. FIGURE 6.14 Solution to Example 6.12. The P-value is the area in the right-hand tail, which is between 0.005 and 0.01. The following computer output (from MINITAB) presents the results from Page 446 Example 6.12. Two-Sample T-Test and CI: Struct, Conven Two-sample T for C1 vs C2 Struct Conven N 10 10 Mean 44.10 32.30 StDev 10.09 8.56 SE Mean 3.19074 2.70691 Difference = mu (Struct) − mu (Conven) Estimate for difference: 11.8000 95% lower bound for difference: 4.52100 T-Test of difference = 0 (vs >): T-Value = 2.82 P-Value = 0.006 DF = 17 Note that the 95% lower confidence bound is consistent with the alternate hypothesis. This indicates that the P-value is less than 5%. The methods described in this section can be used to test a hypothesis that two population means differ by a specified constant. Example 6.13 shows how. Example 6.13 Refer to Example 6.12. Can you conclude that the mean number of items identified with the new structured design exceeds that of the conventional design by more than 2? Solution The null and alternate hypotheses are We observe , sX = 10.09, sY = 8.56, nX = 10, and nY = 10. Under H0, we take μX − μY = 2. The test statistic is given by expression (6.5) to be Under H0, the test statistic has a Student's t distribution with 17 degrees of freedom. Note that the number of degrees of freedom is calculated in the same way as in Example 6.12. The value of the test statistic is t = 2.342. This is a one-tailed test. The P-value is between 0.01 and 0.025. We conclude that the mean number of items identified with the new structured design exceeds that of the conventional design by more than 2. Page 447 Summary Let and be samples from normal populations with means μX and μY and standard deviations σX and σY, respectively. Assume the samples are drawn independently of each other. If σX and σY are not known to be equal, then, to test a null hypothesis of the form H0: μX − MY ≤ Δ0, H0: MX − MY ≥ Δ0, or H0: MX − MY = Δ0: ■ Compute , rounded down to the nearest integer. ■ Compute the test statistic ■ Compute the P-value. The P-value is an area under the Student's t curve with ν degrees of freedom, which depends on the alternate hypothesis as follows: Alternate Hypothesis H1: μX − μY > Δ0 . P-value Area to the right of t Δ0 H1: μX − μY < Area to the left of t Δ0 H1: μX − μY ≠ Sum of the areas in the tails cut off by t and −t Δ0 When the Populations Have Equal Variances When the population variances are known to be nearly equal, the pooled variance (see Section 5.6) may be used. The pooled variance is given by The test statistic for testing any of the null hypotheses H0: μX − μY = 0, H0: μX − μY ≤ 0, or H0: μX − μY ≥ 0 is Under H0, the test statistic has a Student's t distribution with nX + nY − 2 degrees of freedom. Example 6.14 Two methods have been developed to determine the nickel content of steel. In a sample of five replications of the first method on a certain kind of steel, the average measurement (in percent) was and the standard deviation was sX = 0.042. The average of seven replications of the second method was and the standard deviation was sY Page 448 = 0.048. Assume that it is known that the population variances are nearly equal. Can we conclude that there is a difference in the mean measurements between the two methods? replications of the second method were Solution Substituting the sample sizes nX = 5 and nY = 7 along with the sample standard deviations sX = 0.042 and sY = 0.048, we compute the pooled standard deviation, obtaining sp = 0.0457. The value of the test statistic is therefore Under H0, the test statistic has the Student's t distribution with 10 degrees of freedom. Consulting the Student's t table, we find that the area under the curve in each tail is between 0.01 and 0.005. Since the null hypothesis stated that the means are equal, this is a two-tailed test, so the P-value is the sum of the areas in both tails. We conclude that 0.01 < P < 0.02 (see Figure 6.15). There does appear to be a difference in the mean measurements between the two methods. FIGURE 6.15 Solution to Example 6.14. The P-value is the sum of the areas in both tails, which is between 0.01 and 0.02. Don't Assume the Population Variances Are Equal Just Because the Sample Variances Are Close The major problem with using the pooled variance is that the assumption that the population variances are equal is very strict. The hypothesis test based on the pooled variance can be quite unreliable if it is used when the population variances are not equal. In practice, the population variances are almost always unknown, and it is usually impossible to be sure that they are equal. In situations where the sample variances are nearly equal, it is tempting to assume that the population variances are nearly equal as well. This assumption is not justified, however, because it is possible for the sample variances to be nearly equal even when the population variances are quite different. Computer packages often offer a choice of assuming variances to be equal or unequal. The best practice is to assume the variances to be unequal unless it is quite certain that they are equal. See the discussion in Section 5.6. Page 449 Summary Let and be samples from normal populations with means μX and μY and standard deviations σX and σY, respectively. Assume the samples are drawn independently of each other. If σX and σY are known to be equal, then, to test a null hypothesis of the form H0: μX − μY ≤ Δ0, H0: μX − μY ≥ Δ0, or H0: MX − MY = Δ0: ■ Compute . ■ Compute the test statistic ■ Compute the P-value. The P-value is an area under the Student's t curve with nX + nY − 2 degrees of freedom, which depends on the alternate hypothesis as follows: . Alternate P-value Hypothesis H1: μX − μY > Area to the right of t Δ0 H1: μX − μY < Area to the left of t Δ0 H1: μX − μY ≠ Sum of the areas in the tails cut off by t and −t Δ0 Exercises for Section 6.7 1. 2. A crayon manufacturer is comparing the effects of two kinds of yellow dye on the brittleness of crayons. Dye B is more expensive than dye A, but it is thought that it might produce a stronger crayon. Four crayons are tested with each kind of dye, and the impact strength (in joules) is measured for each. The results are as follows: Dye A: 1.0 2.0 1.2 3.0 Dye B: 3.0 3.2 2.6 3.4 a. Can you conclude that the mean strength of crayons made with dye B is greater than that of crayons made with dye A? b. Can you conclude that the mean strength of crayons made with dye B exceeds that of crayons made with dye A by more than 1 J? In a study of the relationship of the shape of a tablet to its dissolution time, 6 disk-shaped ibuprofen tablets and 8 oval-shaped ibuprofen tablets were dissolved in water. The dissolve times, in seconds, were as follows: Disk: 269.0 249.3 255.2 252.7 247.0 261.6 Oval: 268.8 260.0 273.5 253.9 278.5 289.4 261.6 280.2 Can you conclude that the mean dissolve times differ between the two shapes? 3. 4. The article "Influence of Penetration Rate on Penetrometer Resistance" (J. Oliveira, M. Almeida, et al., Journal of Geotechnical and Geoenvironmental Engineering, 2011:695– 703) presents measures of penetration resistance, expressed as a multiple of a standard quantity, for a certain fine-grained soil. Fifteen measurements taken at a depth of 1 m had a mean of 2.31 with a standard deviation of 0.89. Fifteen measurements taken at a depth of 2 m had a mean of 2.80 with a standard deviation of 1.10. Can you conclude that the penetration resistance differs between the two depths? The article "Time Series Analysis for Construction Productivity Experiments" (T. Abdelhamid and J. Everett, Journal of Construction Engineering and Page 450 Management, 1999:87–95) presents a study comparing the effectiveness of a video system that allows a crane operator to see the lifting point while operating the crane with the old system in which the operator relies on hand signals from a tagman. Three different lifts, A, B, and C, were studied. Lift A was of little difficulty, lift B was of moderate difficulty, and lift C was of high difficulty. Each lift was performed several times, both with the new video system and with the old tagman system. The time (in seconds) required to perform each lift was recorded. The following tables present the means, standard deviations, and sample sizes. Tagman Video Tagman Video Tagman Video a. 5. Mean 47.79 47.15 Low Difficulty Standard Deviation 2.19 2.65 Mean 69.33 58.50 Moderate Difficulty Standard Deviation Sample Size 6.26 12 5.59 24 Mean 109.71 84.52 High Difficulty Standard Deviation 17.02 13.51 Sample Size 14 40 Sample Size 17 29 Can you conclude that the mean time to perform a lift of low difficulty is less when using the video system than when using the tagman system? Explain. b. Can you conclude that the mean time to perform a lift of moderate difficulty is less when using the video system than when using the tagman system? Explain. c. Can you conclude that the mean time to perform a lift of high difficulty is less when using the video system than when using the tagman system? Explain. The Mastic tree (Pistacia lentiscus) is used in reforestation efforts in southeastern Spain. The article "Nutrient Deprivation Improves Field Performance of Woody Seedlings in a Degraded Semi-arid Shrubland" (R. Trubata, J. Cortina, and A. Vilagrosaa, Ecological Engineering, 2011:1164–1173) presents a study that investigated the effect of adding slow- release fertilizer to the usual solution on the growth of trees. Following are the heights, in cm, of 10 trees grown with the usual fertilizer (the control group), and 10 trees grown with the slow-release fertilizer (treatment). These data are consistent with the mean and standard deviation reported in the article. Can you conclude that the mean height of plants grown with slow-release fertilizer is greater than that of plants with the usual fertilizer? Usual 17.3 18.5 Slow-release 25.2 25.5 6. 9. 18.7 20.3 26.2 24.8 19.5 20.3 25.0 23.6 53 88 89 62 39 66 23 39 28 2 49 Since the same scale was used for both weights, and since both weights are similar, it is reasonable to assume that the variance of the weighing does not depend on the object being weighed. Can you conclude that the weights differ? It is thought that a new process for producing a certain chemical may be cheaper than the currently used process. Each process was run 6 times, and the cost of producing 100 L of the chemical was determined each time. The results, in dollars, were as follows: New Process: Old Process: 8. 19.5 20.3 25.2 24.1 Two weights, each labeled as weighing 100 g, are each weighed several times on the same scale. The results, in units of μg above 100 g, are as follows: First weight: Second weight: 7. 22.0 18.6 23.2 25.2 51 52 55 53 54 53 50 54 59 56 50 58 Can you conclude that the mean cost of the new method is less than that of the old method? King Tut was an ancient Egyptian ruler whose tomb was discovered and opened in 1923. Legend has it that the archaeologists who opened the tomb were subject to a "mummy's curse," which would shorten their life spans. The article "The Mummy's Curse: Page 451 Historical Cohort Study" (M. Nelson, British Medical Journal, 2002:1482–1484) presents an investigation of the mummy's curse. The article reports that 25 people exposed to the curse had a mean life span of 70.0 years with a standard deviation of 12.4 years, while a sample of 11 Westerners in Egypt at the time who were not exposed to the curse had a mean life span of 75.0 years with a standard deviation of 13.6 years. Can you conclude that the mean life span of those exposed to the mummy's curse is less than the mean of those not exposed? The article "Wind-Uplift Capacity of Residential Wood Roof-Sheathing Panels Retrofitted with Insulating Foam Adhesive" (P. Datin, D. Prevatt, and W. Pang, Journal of Architectural Engineering, 2011:144–154) presents a study of the failure pressures of roof panels. A sample of 15 panels constructed with 8-inch nail spacing on the intermediate framing members had a mean failure pressure of 8.38 kPa with a standard deviation of 0.96 kPa. A sample of 15 panels constructed with 6-inch nail spacing on the intermediate framing members had a mean failure pressure of 9.83 kPa with a standard deviation of 1.02 10. 11. 12. 13. 14. kPa. Can you conclude that 6-inch spacing provides a higher mean failure pressure? The article "Magma Interaction Processes Inferred from Fe-Ti Oxide Compositions in the Dölek and Sariçiçek Plutons, Eastern Turkey" (O. Karsli, F. Aydin, et al., Turkish Journal of Earth Sciences, 2008:297–315) presents chemical compositions (in weight-percent) for several rock specimens. Fourteen specimens (two outliers were removed) of limenite grain had an average iron oxide (Fe2O3) content of 9.30 with a standard deviation of 2.71, and seven specimens of limenite lamella had an average iron oxide content of 9.47 with a standard deviation of 2.22. Can you conclude that the mean iron oxide content differs between limenite grain and limenite lamella? The article "Structural Performance of Rounded Dovetail Connections Under Different Loading Conditions" (T. Tannert, H. Prion, and F. Lam, Can J Civ Eng, 2007:1600–1605) describes a study of the deformation properties of dovetail joints. In one experiment, 10 rounded dovetail connections and 10 double rounded dovetail connections were loaded until failure. The rounded connections had an average load at failure of 8.27 kN with a standard deviation of 0.62 kN. The double-rounded connections had an average load at failure of 6.11 kN with a standard deviation of 1.31 kN. Can you conclude that the mean load at failure is greater for rounded connections than for double-rounded connections? The article "Predicting Depressed Patients with Suicidal Ideation from ECG Recordings" (A. Khandoker, V. Luthra, et al., Med Biol Eng Comput, 2017:793–805) reports a study in which systolic blood pressure (in mmHg) was measured for 16 patients suffering from depression and for 29 controls. Those with depression averaged 112.50 with a standard deviation of 13.90, and the controls averaged 110.34 with a standard deviation of 8.65. Can you conclude that the mean systolic blood pressure is higher in those suffering from depression? In an experiment to test the effectiveness of a new sleeping aid, a sample of 12 patients took the new drug, and a sample of 14 patients took a commonly used drug. Of the patients taking the new drug, the average time to fall asleep was 27.3 minutes with a standard deviation of 5.2 minutes, and for the patients taking the commonly used drug the average time was 32.7 minutes with a standard deviation of 4.1 minutes. Can you conclude that the mean time to sleep is less for the new drug? A new post-surgical treatment was compared with a standard treatment. Seven subjects received the new treatment, while seven others (the controls) received the standard treatment. The recovery times, in days, are given below. Treatment: Control: 12 18 13 23 15 24 19 30 20 32 21 35 24 39 Can you conclude that the mean recovery time for those receiving the new treatment is less than the mean for those receiving the standard treatment? 15. In a study to compare the effectiveness of distance learning with traditional classroom instruction, 12 students took a business course online while 14 students took it in a classroom. The final exam scores for the online students averaged 76.8 with a standard deviation of 8.6, and the final exam scores for the classroom students averaged 80.1 with a standard deviation of 9.3. Can you conclude that the traditional classroom instruction was more effective? Page 452 16. The following MINITAB output presents the results of a hypothesis test for the difference μX − μY between two population means. Two-sample T for X vs Y X Y N 10 10 Mean 39.31 29.12 StDev 8.71 4.79 SE Mean 2.8 1.5 Difference = mu (X) − mu (Y) Estimate for difference: 10.1974 95% lower bound for difference: 4.6333 T-Test of difference = 0 (vs >): T-Value = 3.25 P-Value = 0.003 DF = 13 a. b. c. Is this a one-tailed or two-tailed test? What is the null hypothesis? Can H0 be rejected at the 1% level? How can you tell? 17. The following MINITAB output presents the results of a hypothesis test for the difference μX − μY between two population means. Some of the numbers are missing. Fill in the numbers for (a) through (d). Two-sample T for X vs Y X Y N 6 13 Mean 1.755 3.239 StDev 0.482 (b) SE Mean (a) 0.094 Difference = mu (X) − mu (Y) Estimate for difference: (c) 95% CI for difference: (−1.99996, −0.96791) T-Test of difference = 0 (vs not =): T-Value = (d) P-Value = 0.000 DF = 7 6.8 Tests with Paired Data We saw in Section 5.7 that it is sometimes better to design a two-sample experiment so that each item in one sample is paired with an item in the other. In this section, we present a method for testing hypotheses involving the difference between two population means on the basis of such paired data. We begin with an example. Particulate matter (PM) emissions from automobiles are a serious environmental concern. Eight vehicles were chosen at random from a fleet, and their emissions were measured under both highway driving and stop-and-go driving conditions. The differences (stop-and-go emission – highway emission) were computed as well. The results, in milligrams of particulates per gallon of fuel, were as follows: Vehicle 1 2 3 4 5 6 7 8 Stop-and-go 1500 870 1120 1250 3460 1110 1120 880 Highway 941 456 893 1060 3107 1339 1346 644 Difference 559 414 227 190 353 −229 −226 236 Can we conclude that the mean level of emissions is less for highway driving than for Page 453 stop-and-go driving? The basic idea behind the construction of the hypothesis test in this example is the same as the idea behind the construction of confidence intervals for paired data presented in Section 5.7. We treat the collection of differences as a single random sample from a population of differences. Denote the population mean of the differences by μD and the standard deviation by σD. There are only eight differences, which is a small sample. If we assume that the population of differences is approximately normal, we can use the Student's t test, as presented in Section 6.4. The observed value of the sample mean of the differences is . The sample standard deviation is sD = 284.1. The null and alternate hypotheses are The test statistic is The null distribution of the test statistic is Student's t with seven degrees of freedom. Figure 6.16 presents the null distribution and indicates the location of the test statistic. This is a one-tailed test. The t table indicates that 5% of the area in the tail is cut off by a t value of 1.895, very close to the observed value of 1.897. The P-value is approximately 0.05. The following computer output (from MINITAB) presents this result. Paired T-Test and CI: StopGo, Highway Paired T for StopGo - Highway StopGo Highway Difference N Mean StDev SE Mean 8 1413.75 850.780 300.796 8 1223.25 820.850 290.214 8 190.50 284.104 100.446 95% lower bound for mean difference: 0.197215 T-Test of mean difference = 0 (vs > 0): T-Value = 1.90 P-Value = 0.050 FIGURE 6.16 The null distribution of is t7. The observed value of t, corresponding to the observed values and sp = 284.1, is 1.897. If H0 is true, the probability that t takes on a value as extreme as or more extreme than that observed is very close to 0.05. Note that the 95% lower bound is just barely consistent with the alternate hypothesis. Page 454 This indicates that the P-value is just barely less than 0.05 (although it is given by 0.050 to two significant digits). Summary Let (X1, Y1), ..., (Xn, Yn) be a sample of ordered pairs whose differences D1,...,Dn are a sample from a normal population with mean μD. Let sD be the sample standard deviation of D1, ..., Dn. To test a null hypothesis of the form H0 : μD ≤ μ0, H0 : μD ≥ μ0, or H0 : μD = μ0: ■ Compute the test statistic . ■ Compute the P-value. The P-value is an area under the Student's t curve with n − 1 degrees of freedom, which depends on the alternate hypothesis as follows: Alternate P-value Hypothesis H1 : μD > μ0 Area to the right of t H1 : μD < μ0 Area to the left of t H1 : μD ≠ μ0 Sum of the areas in the tails cut off by t and −t ■ If the sample is large, the Di need not be normally distributed, the test statistic is , and a z test should be performed. Exercises for Section 6.8 1. The article "Improved Bioequivalence Assessment of Topical Dermatological Drug Products Using Dermatopharmacokinetics" (B. N'Dri-Stempfer, W. Navidi, R. Guy, and A. Bunge, Pharmaceutical Research, 2009:316–328) described a study comparing the amounts of econozole nitrate absorbed into human skin for several formulations of antifungal ointment. Both a brand name and generic drug were applied to the arms of 14 subjects, and the amounts absorbed, in μg/cm2, were measured. Following are the results. Can you conclude that the mean amount absorbed differs between the brand name and the generic drug? Brand Name 2.23 1.68 1.96 2.81 1.14 3.20 2.33 4.06 2.92 2.92 2.83 3.45 2.72 Generic 1.42 1.95 2.58 2.25 1.21 3.01 2.76 3.65 2.89 2.85 2.44 3.11 2.64 Difference 0.81 −0.27 −0.62 0.56 −0.07 0.19 −0.43 0.41 0.03 0.07 0.39 0.34 0.08 3.74 2.82 0.92 Page 455 2. The article "Prediction of Performance and Evaluation of Flexible Pavement Rehabilitation Strategies" (K. Lee, K. Wilson, and S. Hassan, Journal of Traffic and Transportation Engineering, 2017:178–184), describes methods used to rehabilitate Route 165 in the town of Exeter, Rhode Island. The following table presents monthly counts of two-axle and fiveaxle trucks during a recent year. The results are presented in the following table. Month January February March April May June July August September October November December 3. Two-axle 1168 1466 1389 1691 2116 2109 2432 2523 2408 1899 1296 1333 Five-axle 1301 1495 1966 2136 2015 2180 2115 2064 2234 2144 1781 1625 Difference −133 −29 −577 −445 101 −71 317 459 174 −245 −485 −292 Can you conclude that the monthly means differ between the two types of trucks? In an experiment to determine whether there is a systematic difference between the weights obtained with two different scales, 10 rock specimens were weighed, in grams, on each scale. The following data were obtained: Specimen 1 2 3 4 5 6 7 8 9 10 Weight on Scale 1 11.23 14.36 8.33 10.50 23.42 9.15 13.47 6.47 12.40 19.38 Weight on Scale 2 11.27 14.41 8.35 10.52 23.41 9.17 13.52 6.46 12.45 19.35 Can you conclude that the mean weight on scale 2 is greater than the mean weight on scale 1? 4. In an experiment to determine the effect of ambient temperature on the emissions of oxides of nitrogen (NOx) of diesel trucks, 10 trucks were run at temperatures of 40°F and 80°F. The emissions, in ppm, are presented in the following table. Truck 1 2 3 4 5 6 7 8 9 10 5. 40°F 0.8347 0.7532 0.8557 0.9012 0.7854 0.8629 0.8827 0.7403 0.7480 0.8486 Can you conclude that the mean emissions differ between the two temperatures? Two formulations of a certain coating, designed to inhibit corrosion, are being tested. For each of eight pipes, half the pipe is coated with formulation A and the other half is coated with formulation B. Each pipe is exposed to a salt environment for 500 hours. Afterward, the corrosion loss (in μm) is measured for each formulation on each pipe. Pipe 1 2 3 4 5 6 7 8 6. 80°F 0.8152 0.7652 0.8426 0.7971 0.7643 0.8195 0.7836 0.6945 0.7729 0.7947 A 197 161 144 162 185 154 136 130 B 204 182 140 178 183 163 156 143 Can you conclude that the mean amount of corrosion differs between the two formulations? Two microprocessors are compared on a sample of six benchmark codes to determine whether there is a difference in speed. The times (in seconds) used by each processor on each code are given in the following table. Code Processor A 1 27.2 2 18.1 3 27.2 4 19.7 5 24.5 6 22.1 Processor B 24.1 19.3 26.8 20.1 27.6 29.8 Can you conclude that the mean speeds of the two processors differ? 7. The compressive strength, in kilopascals, was measured for concrete blocks from five different batches of concrete, both three and six days after pouring. The data are presented in the following table. After 3 days After 6 days 8. Page 456 1 1341 1376 2 1316 1373 Batch 3 1352 1366 4 1355 1384 5 1327 1358 Can you conclude that the mean strength after six days is greater than the mean strength after three days? The article "Effect of Granular Subbase Thickness on Airfield Pavement Structural Response" (K. Gopalakrishnan and M. Thompson, Journal of Materials in Civil Engineering, 2008:331–342) presents a study of the effect of the subbase thickness (in mm) on the amount of surface deflection caused by aircraft landing on an airport runway. Two landing gears, one simulating a Boeing 747 aircraft, and the other a Boeing 777 aircraft, were trafficked across four test sections of runway. The results are presented in the following table. Section 1 4.01 4.57 Boeing 747 Boeing 777 9. 2 3.87 4.48 3 3.72 4.36 4 3.76 4.43 Can you conclude that the mean deflection is greater for the Boeing 777? A crossover trial is a type of experiment used to compare two drugs. Subjects take one drug for a period of time, then switch to the other. The responses of the subjects are then compared using matched pair methods. In an experiment to compare two pain relievers, seven subjects took one pain reliever for two weeks, then switched to the other. They rated their pain level from 1 to 10, with larger numbers representing higher levels of pain. The results were 1 Drug A Drug B 6 5 2 3 1 Subject 3 4 5 4 5 7 5 5 5 6 1 2 7 4 2 Can you conclude that the mean response differs between the two drugs? 10. A group of eight individuals with high cholesterol levels were given a new drug that was designed to lower cholesterol levels. Cholesterol levels, in mg/dL, were measured before and after treatment for each individual, with the following results: Subject 1 2 3 4 5 6 7 8 Before 283 299 274 284 248 275 293 277 After 215 206 187 212 178 212 192 196 a. Can you conclude that the mean cholesterol level after treatment is less than the mean before treatment? b. Can you conclude that the reduction in mean cholesterol level after treatment is greater than 75 mg/dL? 11. The management of a taxi cab company is trying to decide if they should switch from bias tires to radial tires to improve fuel economy. Each of 10 taxis was equipped with one of the two tire types and driven on a test course. Without changing drivers, tires were then switched to the other tire type and the test course was repeated. The fuel economy (in mpg) for the 10 cars is as follows: Page 457 Car 1 2 3 4 5 6 7 8 9 10 a. b. Radial 32.1 36.1 32.3 29.5 34.3 31.9 33.4 34.6 35.2 32.7 Bias 27.1 31.5 30.4 26.9 29.9 28.7 30.2 31.8 33.6 29.9 Because switching tires on the taxi fleet is expensive, management does not want to switch unless a hypothesis test provides strong evidence that the mileage will be improved. State the appropriate null and alternate hypotheses, and find the P-value. A cost-benefit analysis shows that it will be profitable to switch to radial tires if the mean mileage improvement is greater than 2 mpg. State the appropriate null and alternate hypotheses, and find the P-value, for a hypothesis test that is designed to form the basis for the decision whether to switch. 12. The following MINITAB output presents the results of a hypothesis test for the difference μX − μY between two population means. Paired T for X − Y X Y Difference N 12 12 12 Mean 134.233 100.601 33.6316 StDev 68.376 94.583 59.5113 SE Mean 19.739 27.304 17.1794 95% lower bound for mean difference: 2.7793 T-Test of mean difference = 0 (vs > 0): T-Value = 1.96 P-Value = 0.038 a. b. c. Is this a one-tailed or two-tailed test? What is the null hypothesis? Can H0 be rejected at the 1% level? How can you tell? d. Use the output and an appropriate table to compute a 98% confidence interval for μX − μY. 13. The following MINITAB output presents the results of a hypothesis test for the difference μX − μY between two population means. Some of the numbers are missing. Fill in the numbers for (a) through (d). Paired T for X − Y N Mean StDev SE Mean X 7 12.4141 2.9235 (a) Y 7 8.3476 (b) 1.0764 Difference 7 (c) 3.16758 1.19723 95% lower bound for mean difference: 1.74006 T-Test of mean difference = 0 (vs > 0): T-Value = (d) P-Value = 0.007 Page 458 6.9 Distribution-Free Tests The Student's t tests described in Sections 6.4 and 6.7 formally require that samples come from normal populations. Distribution-free tests get their name from the fact that the samples are not required to come from any specific distribution. While distributionfree tests do require assumptions for their validity, these assumptions are somewhat less restrictive than the assumptions needed for the t test. Distribution-free tests are sometimes called nonparametric tests. We discuss two distribution-free tests in this section. The first, called the Wilcoxon signedrank test, is a test for a population mean, analogous to the one-sample t test discussed in Section 6.4. The second, called the Wilcoxon rank-sum test, or the Mann–Whitney test, is analogous to the two-sample t test discussed in Section 6.7. The Wilcoxon Signed-Rank Test We illustrate this test with an example. The nickel content, in parts per thousand by weight, is measured for six welds. The results are 9.3, 0.9, 9.0, 21.7, 11.5, and 13.9. Let μ represent the mean nickel content for this type of weld. It is desired to test H0 : μ ≥ 12 versus H1 : μ < 12. The Student's t test is not appropriate, because there are two outliers, 0.9 and 21.7, which indicate that the population is not normal. The Wilcoxon signed-rank test can be used in this situation. This test does not require the population to be normal. It does, however, require that the population be continuous (rather than discrete), and that the probability density function be symmetric. (The normal is a special case of a continuous symmetric population.) The given sample clearly comes from a continuous population, and the presence of outliers on either side make it reasonable to assume that the population is approximately symmetric as well. We therefore proceed as follows. Under H0, the population mean is μ = 12. Since the population is assumed to be symmetric, the population median is 12 as well. To compute the rank-sum statistic, we begin by subtracting 12 from each sample observation to obtain differences. The difference closest to 0, ignoring sign, is assigned a rank of 1. The difference next closest to 0, again ignoring sign, is assigned a rank of 2, and so on. Finally, the ranks corresponding to negative differences are given negative signs. The following table shows the results. x 11.5 13.9 x − 12 −0.5 1.9 Signed Rank −1 2 9.3 9.0 21.7 0.9 −2.7 −3.0 9.7 −11.1 −3 −4 5 −6 Denote the sum of the positive ranks S+ and the sum of the absolute values of the negative ranks S−. Either S+ or S− may be used as a test statistic; we shall use S+. In this example S+ = 2 + 5 = 7, and S− = 1 + 3 + 4 + 6 = 14. Note that since the sample size is 6, by necessity S+ Page 459 + S− = 1 + 2 + 3 + 4 + 5 + 6 = 21. For any sample, it is the case that S+ + S− = 1 + 2 + ⋯ + n = n(n + 1)/2. In some cases, where there are many more positive ranks than negative ranks, it is easiest first to compute S− by summing the negative ranks and then to compute S+ = n(n + 1)/2 − S−. Figures 6.17 and 6.18 show how S+ can be used as a test statistic. In Figure 6.17, μ > 12. For this distribution, positive differences are more probable than negative differences and tend to be larger in magnitude as well. Therefore it is likely that the positive ranks will be greater both in number and in magnitude than the negative ranks, so S+ is likely to be large. In Figure 6.18, μ < 12, and the situation is reversed. Here the positive ranks are likely to be fewer in number and smaller in magnitude, so S+ is likely to be small. FIGURE 6.17 The true median is greater than 12. Sample observations are more likely to be above 12 than below 12. Furthermore, the observations above 12 will tend to have larger differences from 12 than the observations below 12. Therefore S+ is likely to be large. FIGURE 6.18 The true median is less than 12. Sample observations are more likely to be below 12 than above 12. Furthermore, the observations below 12 will tend to have larger differences from 12 than the observations above 12. Therefore S+ is likely to be small. We see that in general, large values of S+ will provide evidence against a null Page 460 hypothesis of the form $H_0 : \mu \le \mu_0$, while small values of S+ will provide evidence against a null hypothesis of the form $H_0 : \mu \ge \mu_0$. In this example, the null hypothesis is $H_0 : \mu \ge 12$, so a small value of S+ will provide evidence against it. We observe S+ = 7. The P-value is the probability of observing a value of S+ that is less than or equal to 7 when H0 is true. Table A.5 (in Appendix A) presents certain probabilities for the null distribution of S+. Consulting this table with the sample size n = 6, we find that the probability of observing a value of 4 or less is 0.1094. The probability of observing a value of 7 or less must be greater than this, so we conclude that P > 0.1094, and thus do not reject H0. Example 6.15 In the example discussed previously, the nickel content for six welds was measured to be 9.3, 0.9, 9.0, 21.7, 11.5, and 13.9. Use these data to test H0 : μ ≤ 5 versus H1 : μ > 5. Solution The table of differences and signed ranks is as follows: x 9.0 0.9 x−5 4.0 −4.1 Signed Rank −1 2 9.3 11.5 13.9 21.7 4.3 6.5 8.9 16.7 3 4 5 6 The observed value of S+ is 2 + 3 + 4 + 5 + 6 = 19. Since the null hypothesis is of the form μ ≤ μ0, large values of S+ provide evidence against H0. Therefore the P-value is the area in the right-hand tail of the null distribution, corresponding to values greater than or equal to 19. Consulting Table A.5 shows that the P-value is 0.0469. Example 6.16 Use the data in Example 6.15 to test H0 : μ = 16 versus H1 : μ ≠ 16. Solution The table of differences and signed ranks is as follows: x 13.9 11.5 21.7 9.3 9.0 0.9 x − 16 −2.1 −4.5 5.7 −6.7 −7.0 −15.1 Signed Rank −1 −2 3 −4 −5 −6 Since the null hypothesis is of the form μ = μ0, this is a two-tailed test. The Page 461 observed value of S+ is 3. Consulting Table A.5, we find that the area in the left-hand tail, corresponding to values less than or equal to 3, is 0.0781. The Pvalue is twice this amount, since it is the sum of areas in two equal tails. Thus the P-value is 2(0.0781) = 0.1562. Ties Sometimes two or more of the quantities to be ranked have exactly the same value. Such quantities are said to be tied. The standard method for dealing with ties is to assign to each tied observation the average of the ranks they would have received if they had differed slightly. For example, the quantities 3, 4, 4, 5, 7 would receive the ranks 1, 2.5, 2.5, 4, 5, and the quantities 12, 15, 16, 16, 16, 20 would receive the ranks 1, 2, 4, 4, 4, 6. Differences of Zero If the mean under H0 is μ0, and one of the observations is equal to μ0, then its difference is 0, which is neither positive nor negative. An observation that is equal to μ0 cannot receive a signed rank. The appropriate procedure is to drop such observations from the sample altogether, and to consider the sample size to be reduced by the number of these observations. Example 6.17 serves to illustrate this point. Example 6.17 Use the data in Example 6.15 to test H0 : μ = 9 versus H1 : μ ≠ 9. Solution. The table of differences and signed ranks is as follows: x 9.0 9.3 11.5 13.9 0.9 21.7 x−9 0.0 0.3 2.5 4.9 −8.1 12.7 Signed Rank − 1 2 3 −4 5 The value of the test statistic is S+ = 11. The sample size for the purposes of the test is 5, since the value 9.0 is not ranked. Entering Table A.5 with sample size 5, we find that if S+ = 12, the P-value would be 2(0.1562) = 0.3124. We conclude that for S+ = 11, P > 0.3124. Large-Sample Approximation When the sample size n is large, the test statistic S+ is approximately normally distributed. A rule of thumb is that the normal approximation is good if n > 20. It can be shown by advanced methods that under H0, S+ has mean n(n + 1)/4 and variance n(n + 1)(2n + 1)/24. The Wilcoxon signed-rank test is performed by computing the z-score of S+, and then using the normal table to find the P-value. The z-score is Page 462 Example 6.18 illustrates the method. Example 6.18 The article "Exact Evaluation of Batch-Ordering Inventory Policies in Two-Echelon Supply Chains with Periodic Review" (G. Chacon, Operations Research, 2001:79–98) presents an evaluation of a reorder point policy, which is a rule for determining when to restock an inventory. Costs for 32 scenarios are estimated. Let μ represent the mean cost. Test H0 : μ ≥ 70 versus H1 : μ < 70. The data, along with the differences and signed ranks, are presented in Table 6.1. TABLE 6.1 x 79.26 80.79 82.07 82.14 57.19 55.86 42.08 41.78 100.01 100.36 30.46 30.27 22.39 118.39 118.46 20.32 16.69 16.50 15.95 15.16 14.22 Data for Example 6.18 x − 70 9.26 10.79 12.07 12.14 −12.81 −14.14 −27.92 −28.22 30.01 30.36 −39.54 −39.73 −47.61 48.39 48.46 −49.68 −53.31 −53.50 −54.05 −54.84 −55.78 Signed Rank 1 2 3 4 −5 −6 −7 −8 9 10 −11 −12 −13 14 15 −16 −17 −18 −19 −20 −21 11.64 11.48 11.28 10.08 7.28 6.87 6.23 4.57 4.09 140.09 140.77 −58.36 −58.52 −58.72 −59.92 −62.72 −63.13 −63.77 −65.43 −65.91 70.09 70.77 −22 −23 −24 −25 −26 −27 −28 −29 −30 31 32 Solution The sample size is n = 32, so the mean is n(n + 1)/4 = 264 and the variance is n(n + 1)(2n + 1)/24 = 2860. The sum of the positive ranks is S+ = 121. We compute Since the null hypothesis is of the form H0 : μ ≥ μ0, small values of S+ provide evidence against H0. Thus the P-value is the area under the normal curve to the left of z = − 2.67. This area, and thus the P-value, is 0.0038. The Wilcoxon Rank-Sum Test The Wilcoxon rank-sum test, also called the Mann-Whitney test, can be used to test the difference in population means in certain cases where the populations are not normal. Two assumptions are necessary. First the populations must be continuous. Second, their Page 463 probability density functions must be identical in shape and size; the only possible difference between them being their location. To describe the test, let X1, ..., Xm be a random sample from one population and let Y1, ..., Yn be a random sample from the other. We adopt the notational convention that when the sample sizes are unequal, the smaller sample will be denoted X1, ..., Xm. Thus the sample sizes are m and n, with m ≤ n. Denote the population means by μX and μY, respectively. The test is performed by ordering the m + n values obtained by combining the two samples, and assigning ranks 1, 2, ..., m + n to them. The test statistic, denoted by W, is the sum of the ranks corresponding to X1, ..., Xm. Since the populations are identical with the possible exception of location, it follows that if μX < μY, the values in the X sample will tend to be smaller than those in the Y sample, so the rank sum W will tend to be smaller as well. By similar reasoning, if μX > μY, W will tend to be larger. We illustrate the test in Example 6.19. Example 6.19 Resistances, in mΩ, are measured for five wires of one type and six wires of another type. The results are as follows: X: 36 28 29 20 38 Y: 34 41 35 47 49 46 Use the Wilcoxon rank-sum test to test H0 : μX ≥ μY versus H1 : μX < μY. Solution We order the 11 values and assign the ranks. Value 20 28 29 34 35 36 38 41 46 47 49 Rank 1 2 3 4 5 6 7 8 9 10 11 Sample X X X Y Y X X Y Y Y Y The test statistic W is the sum of the ranks corresponding to the X values, so W = 1 +2+3+6+7 = 19. To determine the P-value, we consult Table A.6 (in Appendix A). We note that small values of W provide evidence against H0 : μX ≥ μY, so the P-value is the area in the left-hand tail of the null distribution. Entering the table with m = 5 and n = 6 we find that the area to the left of W = 19 is 0.0260. This is the P-value. Large-Sample Approximation When both sample sizes m and n are greater than 8, it can be shown by advanced methods that the null distribution of the test statistic W is approximately normal with mean m(m + n + 1)/2 and variance mn(m + n + 1)/12. In these cases the test is performed by computing the z-score of W, and then using the normal table to find the P-value. The z-score is Page 464 Example 6.20 illustrates the method. Example 6.20 The article "Cost Analysis Between SABER and Design Bid Build Contracting Methods" (E. Henry and H. Brothers, Journal of Construction Engineering and Management, 2001:359– 366) presents data on construction costs for 10 jobs bid by the traditional method (denoted X) and 19 jobs bid by an experimental system (denoted Y). The data, in units of dollars per square meter, and their ranks, are presented in Table 6.2. Test H0 : μX ≤ μY versus H1 : μX > μY. TABLE 6.2 Value 57 95 101 118 149 196 200 233 243 341 419 457 584 592 594 613 622 708 726 Data for Example 6.20 Rank 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 Sample Y Y Y Y Y Y Y Y Y X Y X Y Y X Y Y X Y 843 908 926 943 1048 1165 1293 1593 1952 2424 20 21 22 23 24 25 26 27 28 29 Y Y X Y X Y X X X X Solution The sum of the X ranks is W = 1 + 12 + 13 + 16 + 18 + 22 + 25 + 26 + 27 + 28 = 188. The sample sizes are m = 10 and n = 19. We use the normal approximation and compute Large values of W provide evidence against the null hypothesis. Therefore the P-value is the area under the normal curve to the right of z = 1.74. From the z table we find that the P-value is 0.0409. Page 465 Distribution-Free Methods Are Not Assumption-Free We have pointed out that the distribution-free methods presented here require certain assumptions for their validity. Unfortunately, this is sometimes forgotten in practice. It is tempting to turn automatically to a distribution-free procedure in any situation in which the Student's t test does not appear to be justified, and to assume that the results will always be valid. This is not the case. The necessary assumptions of symmetry for the signed-rank test and of identical shapes and spreads for the rank-sum test are actually rather restrictive. While these tests perform reasonably well under moderate violations of these assumptions, they are not universally applicable. Exercises for Section 6.9 1. The article "Wastewater Treatment Sludge as a Raw Material for the Production of Bacillus thuringiensis Based Biopesticides" (M. Tirado Montiel, R. Tyagi, and J. Valero, Water Research, 2001:3807–3816) presents measurements of total solids, in g/L, for seven sludge specimens. The results (rounded to the nearest gram) are 20, 5, 25, 43, 24, 21, and 32. Assume the distribution of total solids is approximately symmetric. a. Can you conclude that the mean concentration of total solids is greater than 14 g/L? Compute the appropriate test statistic and find the P-value. b. Can you conclude that the mean concentration of total solids is less than 30 g/L? Compute the appropriate test statistic and find the P-value. c. An environmental engineer claims that the mean concentration of total solids is equal to 18 g/L. Can you conclude that the claim is false? 2. The thicknesses of eight pads designed for use in aircraft engine mounts are measured. The results, in mm, are 41.83, 41.01, 42.68, 41.37, 41.83, 40.50, 41.70, and 41.42. Assume that the thicknesses are a sample from an approximately symmetric distribution. a. Can you conclude that the mean thickness is greater than 41 mm? Compute the appropriate test statistic and find the P-value. b. Can you conclude that the mean thickness is less than 41.8 mm? Compute the appropriate test statistic and find the P-value. c. The target thickness is 42 mm. Can you conclude that the mean thickness differs from the target value? Compute the appropriate test statistic and find the P-value. 3. The article "Reaction Modeling and Optimization Using Neural Networks and Genetic Algorithms: Case Study Involving TS-1-Catalyzed Hydroxylation of Benzene" (S. Nandi, P. Mukherjee, et al., Industrial and Engineering Chemistry Research, 2002:2159–2169) presents benzene conversions (in mole percent) for 24 different benzenehydroxylation reactions. The results are 52.3 30.3 14.3 41.1 40.3 30.1 28.8 63.0 33.4 67.8 80.8 36.2 78.6 26.8 34.6 72.3 37.3 40.0 38.1 81.2 19.0 33.6 59.4 a. a. Can you conclude that the mean conversion is less than 45? Compute the appropriate test statistic and find the P-value. b. Can you conclude that the mean conversion is greater than 30? Compute the appropriate test statistic and find the P-value. c. Can you conclude that the mean conversion differs from 55? Compute the appropriate test statistic and find the P-value. 4. The article "Abyssal Peridotites > 3,800 Ma from Southern West Greenland: Field Relationships, Petrography, Geochronology, Whole-Rock and Mineral Chemistry of Dunite and Harzburgite Inclusions in the Itsaq Gneiss Complex" (C. Friend, V. Bennett, and A. Nutman, Contributions to Mineral Petrology, 2002:71–92) presents silicon dioxide (SiO2) concentrations (in weight percent) for 10 dunites. The results are 40.57 43.76 a. b. 41.48 44.86 40.76 43.06 39.68 46.14 43.68 43.53 Can you conclude that the mean concentration is greater than 41? Page 466 Compute the appropriate test statistic and find the P-value. c. Can you conclude that the mean concentration is less than 43? Compute the appropriate test statistic and find the P-value. c. Can you conclude that the mean concentration differs from 44? Compute the appropriate test statistic and find the P-value. 5. This exercise shows that the signed-rank test can be used with paired data. Two gauges that measure tire tread depth are being compared. Ten different locations on a tire are measured once by each gauge. The results, in mm, are presented in the following table. Location 1 2 3 4 5 6 7 8 9 10 Gauge 1 3.95 3.23 3.60 3.49 3.89 3.76 3.45 3.01 3.82 3.44 Gauge 2 3.80 3.30 3.59 3.61 3.88 3.73 3.56 3.02 3.77 3.49 Difference 0.15 −0.07 0.01 −0.13 0.01 0.05 −0.11 −0.01 0.05 −0.05 Assume the differences are a sample from an approximately symmetric population with mean μ. Use the Wilcoxon signed-rank test to test H0 : μ = 0 versus H1 : μ ≠ 0. 6. The article "n-Nonane Hydroconversion on Ni and Pt Containing HMFI, HMOR and HBEA" (G. Kinger and H. Vinek, Applied Catalysis A: General, 2002:139–149) presents hydroconversion rates (in μmol/g · s) of n-nonane over both HMFI and HBEA catalysts. The results are as follows: HMFI: HBEA: 7. 0.43 0.73 1.91 1.24 2.56 2.93 3.72 6.19 11.00 Can you conclude that the mean rate differs between the two catalysts? A new postsurgical treatment is being compared with a standard treatment. Seven subjects receive the new treatment, while seven others (the controls) receive the standard treatment. The recovery times, in days, are as follows: Treatment (X): Control (Y): 8. 0.93 1 12 18 13 23 15 24 19 30 20 32 21 35 27 40 Can you conclude that the mean recovery time differs between the treatment and control? In an experiment to determine the effect of curing time on compressive strength of concrete blocks, two samples of 15 blocks each were prepared identically except for curing time. The blocks in one sample were cured for two days, while the blocks in the other were cured for six days. The compressive strengths of the blocks, in MPa, are as follows: Cured 2 days (X): 1326 1302 1314 1270 Cured 6 days (Y): 9. 1287 1306 1255 1328 1329 1291 1318 1255 1280 1296 1310 1387 1399 1321 1372 1301 1378 1364 1341 1376 1343 1332 1374 1397 1349 1396 Can you conclude that the mean strength is greater for blocks cured for six days? In a comparison of the effectiveness of distance learning with traditional classroom instruction, 12 students took a business administration course online, while 14 students took it in a classroom. The final exam scores were as follows. Online: 66 91 80 81 Classroom: 75 72 83 51 85 69 64 64 77 71 59 88 83 75 85 77 74 80 74 87 Can you conclude that the mean score differs between the two types of course? 10. A woman who has moved into a new house is trying to determine which of two routes to work has the shorter average driving time. Times in minutes for six trips on route A and five trips on route B are as follows: A: B: 16.0 17.2 15.7 16.9 16.4 16.1 15.9 19.8 16.2 16.7 16.3 Can you conclude that the mean time is less for route A? Page 467 6.10 Tests with Categorical Data In Section 4.1 we studied the Bernoulli trial, which is a process that results in one of two possible outcomes, labeled "success" and "failure." If a number of Bernoulli trials are conducted, and the number of successes is counted, we can test the null hypothesis that the success probability p is equal to a prespecified value p0. This was covered in Section 6.3. If two sets of Bernoulli trials are conducted, with success probability p1 for the first set and p2 for the second set, we can test the null hypothesis that p1 = p2. This was covered in Section 6.6. A generalization of the Bernoulli trial is the multinomial trial (see Section 4.4), which is an experiment that can result in any one of k outcomes, where k ≥ 2. The probabilities of the k outcomes are denoted p1, ..., pk. For example, the roll of a fair die is a multinomial trial with six outcomes 1, 2, 3, 4, 5, 6; and probabilities p1 = p2 = p3 = p4 = p5 = p6 = 1/6. In this section, we generalize the tests for a Bernoulli probability to multinomial trials. We begin with an example in which we test the null hypothesis that the multinomial probabilities p1, p2, ..., pk are equal to a prespecified set of values p01, p02, ..., p0k, so that the null hypothesis has the form H0 : p1 = p01, p2 = p02 ..., pk = p0k. Imagine that a gambler wants to test a die to see whether it deviates from fairness. Let pi be the probability that the number i comes up. The null hypothesis will state that the die is fair, so the probabilities specified under the null hypothesis are p01 = ⋯ = p06 = 1/6. The null hypothesis is H0 : p1 = ⋯ = p6 = 1/6. The gambler rolls the die 600 times and obtains the results shown in Table 6.3, in the column labeled "Observed." The results obtained are called the observed values. To test the null hypothesis, we construct a second column, labeled "Expected." This column contains the expected values. The expected value for a given outcome is the mean number of trials that would result in that outcome if H0 were true. To compute the expected values, let N be the total number of trials. (In the die example, N = 600.) When H0 is true, the probability that a trial results in outcome i is p0i, so the expected number of trials resulting in outcome i is Np0i. In the die example, the expected number of trials for each outcome is 100. TABLE 6.3 Category 1 2 3 Observed and expected values for 600 rolls of a die Observed 115 97 91 Expected 100 100 100 4 5 6 Total 101 110 86 600 100 100 100 600 The idea behind the hypothesis test is that if H0 is true, then the observed and Page 468 expected values are likely to be close to each other. Therefore we will construct a test statistic that measures the closeness of the observed to the expected values. The statistic is called the chi-square statistic. To define it, let k be the number of outcomes (k = 6 in the die example), and let Oi and Ei be the observed and expected numbers of trials, respectively, that result in outcome i. The chi-square statistic is (6.6) The larger the value of χ2, the stronger the evidence against H0. To determine the P-value for the test, we must know the null distribution of this test statistic. In general, we cannot determine the null distribution exactly. However, when the expected values are all sufficiently large, a good approximation is available. It is called the chi-square distribution with k − 1 degrees of freedom, denoted . Note that the number of degrees of freedom is one less than the number of categories. Use of the chi-square distribution is appropriate whenever all the expected values are greater than or equal to 5. A table for the chi-square distribution (Table A.7) is provided in Appendix A. The table provides values for certain quantiles, or upper percentage points, for a large number of choices of degrees of freedom. As an example, Figure 6.19 presents the probability density function of the distribution. The upper 5% of the distribution is shaded. To find the upper 5% point in the table, look under α = 0.05 and degrees of freedom ν = 10. The value is 18.307. FIGURE 6.19 Probability density function of the distribution. The upper 5% point is 18.307. [See the chi-square table (Table A.7) in Appendix A.] We now compute the value of the chi-square statistic for the data in Table 6.3. The number of degrees of freedom is 5 (one less than the number of outcomes). Using Equation (6.6), the value of the statistic is To determine the P-value for the test statistic, we first note that all the expected values are greater than or equal to 5, so use of the chi-square distribution is appropriate. We consult the chi-square table under five degrees of freedom. The upper 10% point is 9.236. We Page 469 conclude that P > 0.10. (See Figure 6.20.) There is no evidence to suggest that the die is not fair. FIGURE 6.20 Probability density function of the distribution. The observed value of the test statistic is 6.12. The upper 10% point is 9.236. Therefore the P-value is greater than 0.10. The test we have just described determines how well a given multinomial distribution fits the data. For this reason it is called a goodness-of-fit test. Example 6.21 Mega Millions is a multistate lottery in which players try to guess the numbers that will turn up in a drawing of numbered balls. One of the balls drawn is the Mega Ball. Matching the number drawn on the Mega Ball increases one's winnings. During a recent three-year period, the Mega Ball was drawn from a collection of 15 balls numbered 1 through 15, and a total of 344 drawings were made. For the purposes of this example, we grouped the numbers into five categories: 1–3, 4–6, and so on. If the lottery is fair, then the winning number is equally likely to occur in any category. Following are the observed number of draws in each category. Category 1–3 4–6 7–9 10–12 13–15 Total Observed 73 64 74 62 71 344 Can you conclude that the lottery is not fair? Solution Under the null hypothesis, each draw has probability 1/5 to fall in each category. Therefore each expected value is 344/5 = 68.8. We compute the test statistic: There are 5 categories, thus 4 degrees of freedom. Consulting the chi-square table, we find that P > 0.10. (Software yields P = 0.79.) There is no evidence that the lottery is not fair. The Chi-Square Test for Homogeneity In Example 6.21, we tested the null hypothesis that the probabilities of the outcomes for a multinomial trial were equal to a prespecified set of Page 470 values. Sometimes several multinomial trials are conducted, each with the same set of possible outcomes. The null hypothesis is that the probabilities of the outcomes are the same for each experiment. We present an example. Four machines manufacture cylindrical steel pins. The pins are subject to a diameter specification. A pin may meet the specification, or it may be too thin or too thick. Pins are sampled from each machine, and the number of pins in each category is counted. Table 6.4 presents the results. TABLE 6.4 Observed numbers of pins in various categories with regard to a diameter specification Machine 1 Machine 2 Machine 3 Machine 4 Total Too Thin 10 34 12 10 66 OK 102 161 79 60 402 Too Thick 8 5 9 10 32 Total 120 200 100 80 500 Table 6.4 is an example of a contingency table. Each row specifies a category regarding one criterion (machine, in this case), and each column specifies a category regarding another criterion (thickness, in this case). Each intersection of row and column is called a cell, so there are 12 cells in Table 6.4. The number in the cell at the intersection of row i and column j is the number of trials whose outcome was observed to fall into row category i and into column category j. This number is called the observed value for cell ij. Note that we have included the totals of the observed values for each row and column. These are called the marginal totals. The null hypothesis is that the proportion of pins that are too thin, OK, or too thick is the same for all machines. More generally, the null hypothesis says that no matter which row is chosen, the probabilities of the outcomes associated with the columns are the same. We will develop some notation with which to express H0 and to define the test statistic. Let I denote the number of rows in the table, and let J denote the number of columns. Let pij denote the probability that the outcome of a trial falls into column j given that it is in row i. Then the null hypothesis is (6.7) Let Oij denote the observed value in cell ij. Let Oi. denote the sum of the observed values in row i, let O.j denote the sum of the observed values in column j, and let O.. denote the sum of the observed values in all the cells (see Table 6.5). TABLE 6.5 Notation for observed values Row 1 Column 1 O11 Column 2 O12 ... ... Column J O1J Total O1. Row 2 O21 O22 ... O2J O2. ⋮ Row I ⋮ OI1 ⋮ OI2 ⋮ ... ⋮ OIJ ⋮ OI. Total O.1 O.2 ... O.J O.. To define a test statistic, we must compute an expected value for each cell in the table. Under H0, the probability that the outcome of a trial falls into column j is the same for each row i. The best estimate of this probability is the proportion of trials whose outcome falls into column j. This proportion is O.j/O... We need to compute the expected number of trials whose outcome falls into cell ij. We denote this expected value by Eij. It is equal to the proportion of trials whose outcome falls into column j, multiplied by the number Oi. of trials in row i. That is, Page 471 (6.8) The test statistic is based on the differences between the observed and expected values: (6.9) Under H0, this test statistic has a chi-square distribution with (I − 1)(J − 1) degrees of freedom. Use of the chi-square distribution is appropriate whenever the expected values are all greater than or equal to 5. Example 6.22 Use the data in Table 6.4 to test the null hypothesis that the proportions of pins that are too thin, OK, or too thick are the same for all the machines. Solution We begin by using Equation (6.8) to compute the expected values Eij. We show the calculations of E11 and E23 in detail: The complete table of expected values is as follows: Expected values for Table 6.4 Machine 1 Machine 2 Machine 3 Machine 4 Total Too Thin 15.84 26.40 13.20 10.56 66.00 OK 96.48 160.80 80.40 64.32 402.00 Too Thick 7.68 12.80 6.40 5.12 32.00 Total 120.00 200.00 100.00 80.00 500.00 We note that all the expected values are greater than 5. Therefore the chi- Page 472 square test is appropriate. We use Equation (6.9) to compute the value of the chi-square statistic: Since there are four rows and three columns, the number of degrees of freedom is (4 − 1)(3 − 1) = 6. To obtain the P-value, we consult the chi-square table (Table A.7). Looking under six degrees of freedom, we find that the upper 2.5% point is 14.449, and the upper 1% point is 16.812. Therefore 0.01 < P < 0.025. It is reasonable to conclude that the machines differ in the proportions of pins that are too thin, OK, or too thick. Note that the observed row and column totals are identical to the expected row and column totals. This is always the case. The following computer output (from MINITAB) presents the results of this hypothesis test. Chi-Square Test: Thin, OK, Thick Expected counts are printed below observed counts Chi-Square contributions are printed below expected counts 1 2 3 4 Total Thin OK Thick Total 10 102 8 120 15.84 96.48 7.68 2.153 0.316 0.013 34 161 5 200 26.40 160.80 12.80 2.188 0.000 4.753 12 79 9 100 13.20 80.40 6.40 0.109 0.024 1.056 10 60 10 80 10.56 64.32 5.12 0.030 0.290 4.651 66 402 32 500 Chi-Sq = 15.584 DF = 6, P-Value = 0.016 In the output, each cell (intersection of row and column) contains three numbers. The top number is the observed value, the middle number is the expected value, and the bottom number is the contribution (Oij − Eij)2/Eij made to the chi-square statistic from that cell. Page 473 The Chi-Square Test for Independence In Example 6.22 the column totals were random, while the row totals were presumably fixed in advance, since they represented numbers of items sampled from various machines. In some cases, both row and column totals are random. In either case, we can test the null hypothesis that the probabilities of the column outcomes are the same for each row outcome, and the test is exactly the same in both cases. We present an example where both row and column totals are random. Example 6.23 The cylindrical steel pins in Example 6.22 are subject to a length specification as well as a diameter specification. With respect to the length, a pin may meet the specification, or it may be too short or too long. A total of 1021 pins are sampled and categorized with respect to both length and diameter specification. The results are presented in the following table. Test the null hypothesis that the proportions of pins that are too thin, OK, or too thick with respect to the diameter specification do not depend on the classification with respect to the length specification. Observed Values for 1021 Steel Pins Length Too Short OK Too Long Total Diameter Too Thin OK Too Thick 13 117 4 62 664 59 5 68 8 80 849 92 Total 134 806 81 1021 Solution We begin by using Equation (6.8) to compute the expected values. The expected values are given in the following table. Expected Values for 1021 Steel Pins Diameter Length Too Short OK Too Long Too Thin 10.50 63.15 6.35 80.0 OK 111.43 670.22 67.36 849.0 Too Thick 12.07 72.63 7.30 92.0 Total 134.0 806.0 81.0 1021.0 We note that all the expected values are greater than or equal to 5. (One of the observed values is not; this doesn't matter.) Therefore the chi-square test is appropriate. We use Equation (6.9) to compute the value of the chi-square statistic: Page 474 Since there are three rows and three columns, the number of degrees of freedom is (3 − 1)(3 − 1) = 4. To obtain the P-value, we consult the chi-square table (Table A.7). Looking under four degrees of freedom, we find that the upper 10% point is 7.779. We conclude that P > 0.10. (Software yields P = 0.13.) There is no evidence that the length and thickness are related. Exercises for Section 6.10 1. 2. Fasteners are manufactured for an application involving aircraft. Each fastener is categorized as either conforming (suitable for its intended use), downgraded (unsuitable for its intended use but usable for another purpose), and scrap (not usable). It is thought that 85% of the fasteners are conforming, while 10% are downgraded and 5% are scrap. In a sample of 500 fasteners, 405 were conforming, 55 were downgraded, and 40 were scrap. Can you conclude that the true percentages differ from 85%, 10%, and 5%? a. State the appropriate null hypothesis. b. Compute the expected values under the null hypothesis. c. Compute the value of the chi-square statistic. d. Find the P-value. e. What do you conclude? At an assembly plant for light trucks, routine monitoring of the quality of welds yields the following data: Day Shift Evening Shift Night Shift High Quality 467 445 Number of Welds Moderate Quality 191 171 Low Quality 42 34 254 129 17 3. Can you conclude that the quality varies among shifts? a. State the appropriate null hypothesis. b. Compute the expected values under the null hypothesis. c. Compute the value of the chi-square statistic. d. Find the P-value. e. What do you conclude? The article "Inconsistent Health Perceptions for US Women and Men with Diabetes" (M. McCollum, L. Hansen, et al., Journal of Women's Health, 2007:1421–1428) presents results of a survey of adults with diabetes. Each respondent was categorized by gender and income level. The numbers in each category (calculated from percentages given in the article) are presented in the following table. Poor Men 156 Women 348 4. Low Income 253 433 Middle Income 513 592 High Income 604 511 Can you conclude that the proportions in the various income categories differ between men and women? The article "Analysis of Time Headways on Urban Roads: Case Study from Riyadh" (A. Al-Ghamdi, Journal of Transportation Engineering, 2001:289–294) presents a model for the time elapsed between the arrival of consecutive vehicles on urban roads. Following are 137 arrival times (in seconds) along with the values expected from a theoretical model. Time 0–2 2–4 4–6 6–8 8–10 10–12 12–18 18–22 > 22 Observed 18 28 14 7 11 11 8 3 Expected 13 18 16 13 11 9 20 8 9 Can you conclude that the theoretical model does not explain the observed values well? The article "Chronic Beryllium Disease and Sensitization at a Beryllium Processing Facility" (K. Rosenman, V. Hertzberg, et al., Environmental Health Perspectives, 2005:1366–1372) discusses the effects of exposure to beryllium in a cohort of workers. Workers were categorized by their duration of exposure (in years) and by their disease status (chronic beryllium disease, sensitization to beryllium, or no disease). The Page 475 results were as follows: 0.10 (see Figure 6.21 on page 478). We cannot conclude that the scale is reliable. FIGURE 6.21 Probability density function of the distribution. The observed value of the test statistic is 12.15. The lower 10% point is 8.547. Therefore the P-value is greater than 0.10. When the null hypothesis has the form, larger values of the test statistic provide stronger evidence against H0. Then the P-value is the area to the right of the observed value of the test statistic. When the null hypothesis has the form , both Page 478 large and small values of the test statistic provide evidence against H0. The P-value is then found by doubling the area in the tail containing the observed value of the test statistic. The F Test for Equality of Variance Sometimes it is desirable to test a null hypothesis that two populations have equal variances. In general there is no good way to do this. In the special case where both populations are normal, however, a method is available. Let X1, ..., Xm be a simple random sample from a population, and let Y1, ..., Yn population. Assume that the samples are chosen independently. The values of the means, μ1 and μ2, are irrelevant here; we are concerned only with the variances and . Note that the sample sizes, m and n, may be different. Let and be the sample variances. That is, Any of three null hypotheses may be tested. They are and The procedures for testing these hypotheses are similar, but not identical. We will describe the procedure for testing the null hypothesis , and then discuss how the procedure may be modified to test the other two hypotheses. The test statistic is the ratio of the two sample variances: (6.10) Page 479 When H0 is true, we assume that . When H0 is true, F is likely to be near 1. When H0 is false, and , (or equivalently, that are, on average, the same size) So when H0 is true, F is likely to be near 1. is likely to be larger than , and F is likely to be greater than 1. In order to use F as a test statistic, we must know its null distribution. The null distribution is called an F distribution, which we now describe. The F Distribution Statistics that have an F distribution are ratios of quantities, such as the ratio of the two sample variances in Equation (6.10). The F distribution therefore has two values for the degrees of freedom: one associated with the numerator, and one associated with the denominator. The degrees of freedom are indicated with subscripts under the letter F. For example, the symbol F3,16 denotes the F distribution with 3 degrees of freedom for the numerator and 16 degrees of freedom for the denominator. Note that the degrees of freedom for the numerator are always listed first. A table for the F distribution is provided (Table A.8 in Appendix A). The table provides values for certain quantiles, or upper percentage points, for a large number of choices for the degrees of freedom. As an example, Figure 6.22 presents the probability density function of the F3,16 distribution. The upper 5% of the distribution is shaded. To find the upper 5% point in the table, look under α = 0.050, and degrees of freedom ν1 = 3, ν2 = 16. The value is 3.24. FIGURE 6.22 Probability density function of the F3,16 distribution. The upper 5% point is 3.24. [See the F table (Table A.8) in Appendix A.] The F Statistic for Testing Equality of Variance The null distribution of the test statistic is Fm−1, n−1. The number of degrees of freedom for the numerator is one less than the sample size used to compute , and the number of degrees of freedom for the denominator is one less than the sample size used to compute . We illustrate the F test with an example. Example 6.25 In a series of experiments to determine the absorption rate of certain pesticides into skin, measured amounts of two pesticides were applied to several skin specimens. After a time, the amounts absorbed (in μg) were measured. For pesticide A, the variance of the amounts absorbed in 6 specimens was 2.3, while for pesticide B, the variance of the amounts absorbed in 10 specimens was 0.6. Assume that for each pesticide, the amounts absorbed are a simple random sample from a normal population. Can we conclude that the variance in the amount absorbed is greater for pesticide A than for pesticide B? Page 480 Solution Let be the population variance for pesticide A, and let be the population variance for pesticide B. The sample variances are and . The value of the test statistic is The null distribution of the test statistic is F5,9. If H0 is true, then will on the average be smaller than . It follows that the larger the value of F, the stronger the evidence against H0. Consulting the F table with five and nine degrees of freedom, we find that the upper 5% point is 3.48, while the upper 1% point is 6.06. We conclude that 0.01 < P < 0.05. There is reasonably strong evidence against the null hypothesis. See Figure 6.23. FIGURE 6.23 The observed value of the test statistic is 3.83. The upper 5% point of the F5,9 distribution is 3.48; the upper 1 % point is 6.06. Therefore the P-value is between 0.01 and 0.05. We now describe the modifications to the procedure shown in Example 6.25 that are necessary to test the other null hypotheses. To test one could in principle use the test statistic , with small values of the statistic providing evidence against H0. However, since the F table contains only large values (i.e., greater than 1) for the F statistic, it is easier to use the statistic . Under H0,the distribution of is Fn −1, m−1. Finally, we describe the method for testing the two-tailed hypothesis Page 481 For this hypothesis, both large and small values of the statistic provide evidence against H0. The procedure is to use either or , whichever is greater than 1. The P-value for the two-tailed test is twice the P-value for the one-tailed test. In other words, the P-value of the two-tailed test is twice the upper tail area of the F distribution. We illustrate with an example. Example 6.26 In Example 6.25, and , with a sample size of 6, and with a sample size of 10. Test the null hypothesis Solution The null hypothesis is equivalent to . Since , we use the test . In Example 6.25, we found that for the one-tailed test, 0.01 < P < 0.05. Therefore for the two-tailed test, 0.02 < P < 0.10. The following computer output (from MINITAB) presents the solution to Example 6.26. Test for Equal Variances F-Test (normal distribution) Test statistic = 3.83, p-value = 0.078 The F Test Is Sensitive to Departures from Normality The F test, like the t test, requires that the samples come from normal populations. Unlike the t test, the F test for comparing variances is fairly sensitive to this assumption. If the shapes of the populations differ much from the normal curve, the F test may give misleading results. For this reason, the F test for comparing variances must be used with caution. In Chapters 8 and 9, we will use the F distribution to perform certain hypothesis tests in the context of linear regression and analysis of variance. In these settings, the F test is less sensitive to violations of the normality assumption. The F Test Cannot Prove That Two Variances Are Equal In Section 6.7, we presented two versions of the t test for the difference between two means. One version is generally applicable, while the second version, which uses the pooled variance, is appropriate only when the population variances are equal. When deciding whether it is appropriate to assume population variances to be equal, it is tempting to perform an F test and assume the variances to be equal if the null hypothesis of equality is not rejected. Unfortunately this procedure is unreliable, for the basic reason that failure to reject the null hypothesis does not justify the assumption that the null hypothesis is true. In general, an assumption that population variances are equal cannot be justified by a hypothesis test. Page 482 Exercises for Section 6.11 1. A random sample of size 11 from a normal distribution has variance s2 = 96. Test H0 : σ2 ≤ 50 versus H1 : σ2 > 50. 2. A random sample of size 29 from a normal distribution has variance s2 = 24. Test H0 : σ2 ≥ 30 versus H1 : σ2 < 30. 3. Scores on an IQ test are normally distributed. A sample of 25 IQ scores had variance s2 = 64. The developer of the test claims that the population variance is σ2 = 225. Do these data provide sufficient evidence to contradict this claim? 4. A machine that fills beverage cans is supposed to put 12 ounces of beverage in each can. The variance of the amount in each can is 0.01. The machine is moved to a new location. To determine whether the variance has changed, 10 cans are filled. Following are the amounts in the 10 cans. Assume them to be a random sample from a normal population. 12.18 11.96 5. 6. 7. 8. 9. 11.77 12.03 12.09 12.36 12.03 12.28 11.87 11.85 Perform a hypothesis test to determine whether the variance differs from 0.01. What do you conclude? A sample of 25 one-year-old girls had a mean weight of 24.1 pounds with a standard deviation of 4.3 pounds. Assume that the population of weights is normally distributed. A pediatrician claims that the standard deviation of the weights of one-year-old girls is less than 5 pounds. Do the data provide convincing evidence that the pediatrician's claim is true? (Based on data from the National Health Statistics Reports.) The 2008 General Social Survey asked a large number of people how much time they spent watching TV each day. The mean number of hours was 2.98 with a standard deviation of 2.66. Assume that in a sample of 40 teenagers, the sample standard deviation of daily TV time is 1.9 hours, and that the population of TV watching times is normally distributed. Can you conclude that the population standard deviation of TV watching times for teenagers is less than 2.66? Scores on the math SAT are normally distributed. A sample of 20 SAT scores had standard deviation s = 87. Someone says that the scoring system for the SAT is designed so that the population standard deviation will be σ = 100. Do these data provide sufficient evidence to contradict this claim? One of the ways in which doctors try to determine how long a single dose of pain reliever will provide relief is to measure the drug's half-life, which is the length of time it takes for one-half of the dose to be eliminated from the body. A report of the National Institutes of Health states that the standard deviation of the half-life of the pain reliever oxycodone is σ = 1.43 hours. Assume that a sample of 25 patients is given the drug, and the sample standard deviation of the half-lives was s = 1.5 hours. Assume the population is normally distributed. Can you conclude that the true standard deviation is greater than the value reported by the National Institutes of Health? Find the upper 5% point of F7,20. 10. Find the upper 1 % point of F2,5. 11. An F test with five degrees of freedom in the numerator and seven degrees of freedom in the denominator produced a test statistic whose value was 7.46. a. What is the P-value if the test is one-tailed? b. What is the P-value if the test is two-tailed? 12. A broth used to manufacture a pharmaceutical product has its sugar content, in mg/mL, measured several times on each of three successive days. Day 1: Day 2: Day 3: a. b. 5.0 4.8 5.1 5.3 6.3 6.0 4.8 5.0 4.7 5.3 4.7 5.2 4.8 5.1 4.9 5.5 4.9 4.9 5.7 5.9 5.7 4.8 5.1 5.3 4.8 5.4 5.8 5.0 4.9 5.5 4.7 5.6 Can you conclude that the variability of the process is greater on the second day than on the first day? Can you conclude that the variability of the process is greater on the third day than on the second day? Page 483 13. Refer to Exercise 11 in Section 5.6. Can you conclude that the variance of the sodium content varies between the two brands? 14. Refer to Exercise 13 in Section 5.6. Can you conclude that the time to freeze-up is more variable in the seventh month than in the first month after installation? 6.12 Fixed-Level Testing Critical Points and Rejection Regions A hypothesis test measures the plausibility of the null hypothesis by producing a P-value. The smaller the P-value, the less plausible the null. We have pointed out that there is no scientifically valid dividing line between plausibility and implausibility, so it is impossible to specify a "correct" P-value below which we should reject H0. When possible, it is best simply to report the P-value, and not to make a firm decision whether or not to reject. Sometimes, however, a decision has to be made. For example, if items are sampled from an assembly line to test whether the mean diameter is within tolerance, a decision must be made whether to recalibrate the process. If a sample of parts is drawn from a shipment and checked for defects, a decision must be made whether to accept or to return the shipment. If a decision is going to be made on the basis of a hypothesis test, there is no choice but to pick a cutoff point for the P-value. When this is done, the test is referred to as a fixed-level test. Fixed-level testing is just like the hypothesis testing we have been discussing so far, except that a firm rule is set ahead of time for rejecting the null hypothesis. A value α, where 0 < α < 1, is chosen. Then the P-value is computed. If P ≤ α, the null hypothesis is rejected and the alternate hypothesis is taken as truth. If P > α, then the null hypothesis is considered to be plausible. The value of α is called the significance level, or, more simply, the level, of the test. Recall from Section 6.2 that if a test results in a P-value less than α, we say that the null hypothesis is rejected at level α (or 100α%), or that the result is statistically significant at level α (or 100α%). As we have mentioned, a common choice for α is 0.05. Summary To conduct a fixed-level test: ■ Choose a number α, where 0 < α < 1. This is called the significance level, or the level, of the test. ■ Compute the P-value in the usual way. ■ If P ≤ α, reject H0. If P > α, do not reject H0. Example 6.27 Refer to Example 6.1 in Section 6.1. The mean wear in a sample of 45 steel balls was μm, and the standard deviation was s = 14.9 μm. Let μ denote the population mean wear. A test of H0 : μ ≤ 675 versus H1 : μ > 675 yielded a P-value of 0.209. Can we reject H0 at the 25% level? Can we reject H0 at the 5% level? Page 484 Solution The P-value of 0.209 is less than 0.25, so if we had chosen a significance level of α = 0.25, we would reject H0. Thus we reject H0 at the 25% level. Since0.209 > 0.05, we do not reject H0 at the 5% level. In a fixed-level test, a critical point is a value of the test statistic that produces a P-value exactly equal to α. A critical point is a dividing line for the test statistic just as the significance level is a dividing line for the P-value. If the test statistic is on one side of the critical point, the P-value will be less than α, and H0 will be rejected. If the test statistic is on the other side of the critical point, the P-value will be greater than α, and H0 will not be rejected. The region on the side of the critical point that leads to rejection is called the rejection region. The critical point itself is also in the rejection region. Example 6.28 A new concrete mix is being evaluated. The plan is to sample 100 concrete blocks made with the new mix, compute the sample mean compressive strength , and then test H0 : μ ≤ 1350 versus H1 : μ > 1350, where the units are MPa. It is assumed from previous tests of this sort that the population standard deviation σ will be close to 70 MPa. Find the critical point and the rejection region if the test will be conducted at a significance level of 5%. Solution We will reject H0 if the P-value is less than or equal to 0.05. The P-value for this test will be the area to the right of the value of . Therefore the P-value will be less than 0.05, and H0 will be rejected, if the value of is in the upper 5% of the null distribution (see Figure 6.24). The rejection region therefore consists of the upper 5% of the null distribution. The critical point is the boundary of the upper 5%. The null distribution is normal, and from the z table we find that the z-score of the point that cuts off the upper 5% of the normal curve is z.05 = 1.645. Therefore we can express the critical point as z = 1.645, and the rejection region as z ≥ 1.645. It is often more convenient to express the critical point and rejection region in terms of , by converting the z-score to the original units. The null distribution has mean μ = 1350 and standard deviation . Therefore the critical point can be expressed as . The rejection region consists of all values of greater than or equal to 1361.5. FIGURE 6.24 The rejection region for this one-tailed test consists of the upper 5% of the null distribution. The critical point is 1361.5, on the boundary of the rejection region. Page 485 Example 6.29 In a hypothesis test to determine whether a scale is in calibration, the null hypothesis is H0 : μ = 1000 and the null distribution of is N(1000, 0.262). (This situation was presented in Example 6.2 in Section 6.1.) Find the rejection region if the test will be conducted at a significance level of 5%. Solution Since this is a two-tailed test, the rejection region is contained in both tails of the null distribution. Specifically, H0 will be rejected if is in either the upper or the lower 2.5% of the null distribution (see Figure 6.25). The z-scores that cut off the upper and lower 2.5% of the distribution are z = ± 1.96. Therefore the rejection region consists of all values of greater than or equal to 1000 + (1.96)(0.26) = 1000.51, along with all the values less than or equal to 1000 − (1.96)(0.26) = 999.49. Note that there are two critical points, 999.49 and 1000.51. FIGURE 6.25 The rejection region for this two-tailed test consists of both the lower and the upper 2.5% of the null distribution. There are two critical points, 999.49 and 1000.51. Type I and Type II Errors Since a fixed-level test results in a firm decision, there is a chance that the decision could be the wrong one. There are exactly two ways in which the decision can be wrong. One can reject H0 when it is in fact true. This is known as a type I error. Or, one can fail to reject H0 when it is false. This is known as a type II error. When designing experiments whose data will be analyzed with a fixed-level test, it is important to try to make the probabilities of type I and type II errors reasonably small. There is no use in conducting an experiment that has a high probability of leading to an incorrect decision. It turns out that it is easy to control the probability of a type I error, as shown by the following result. If α is the significance level that has been chosen for the test, then the probability of a type I error is never greater than α. We illustrate this fact with the following example. Let X1, ..., Xn be a large random sample from a population with mean μ and variance σ2. Then is normally distributed with Page 486 2 mean μ and variance σ /n. Assume that we are to test H0 : μ ≤ 0 versus H1 : μ > 0 at the fixed level α = 0.05. That is, we will reject H0 if P ≤ 0.05. The null distribution, shown in Figure 6.26, is normal with mean 0 and variance . Assume the null hypothesis is true. We will compute the probability of a type I error and show that it is no greater than 0.05. FIGURE 6.26 The null distribution with the rejection region for H0 : μ ≤ 0. A type I error will occur if we reject H0, which will occur if P ≤ 0.05, which in turn will occur if . Therefore the rejection region is the region . Now since H0 is true, μ ≤ 0. First, we'll consider the case where μ = 0. Then the distribution of is given by Figure 6.26. In this case, is given , so the probability of rejecting H0 and making a type I error is equal to 0.05. Next, consider the case where μ < 0. Then the distribution of is obtained by shifting the curve in Figure 6.26 to the left, so , and the probability of a type I error is less than 0.05. We could repeat this illustration using any number α in place of 0.05. We conclude that if H0 is true, the probability of a type I error is never greater than α. Furthermore, note that if μ is on the boundary of H0 (μ = 0 in this case), then the probability of a type I error is equal to α. We can therefore make the probability of a type I error as small as we please, because it is never greater than the significance level α that we choose. Unfortunately, as we will see in Section 6.13, the smaller we make the probability of a type I error, the larger the probability of a type II error becomes. The usual strategy is to begin by choosing a value for α so that the probability of a type I error will be reasonably small. As we have mentioned, a conventional choice for α is 0.05. Then one computes the probability of a type II error and hopes that it is not too large. If it is large, it can be reduced only by redesigning the experiment—for example by increasing the sample size. Calculating and controlling the size of the type II error is somewhat more difficult than calculating and controlling the size of the type I error. We will discuss this in Section 6.13. Summary When conducting a fixed-level test at significance level α, there are two types of errors that can be made. These are ■ Type I error: Reject H0 when it is true. ■ Type II error: Fail to reject H0 when it is false. The probability of a type I error is never greater than α. Exercises for Section 6.12 1. A hypothesis test is performed, and the P-value is 0.03. True or false: a. H0 is rejected at the 5% level. b. H0 is rejected at the 2% level. c. H0 is not rejected at the 10% level. Page 487 c. 2. 3. 4. 5. 6. A process for a certain type of ore is designed to reduce the concentration of impurities to less than 2%. It is known that the standard deviation of impurities for processed ore is 0.6%. Let μ represent the mean impurity level, in percent, for ore specimens treated by this process. The impurity of 80 ore specimens is measured, and a test of the hypothesis H0 : μ ≥ 2 versus H1 : μ < 2 will be performed. a. b. If the test is made at the 5% level, what is the rejection region? If the sample mean impurity level is 1.85, will H0 be rejected at the 10% level? c. If the sample mean pH is 1.85, will H0 be rejected at the 1% level? d. If the value 1.9 is a critical point, what is the level of the test? A new braking system is being evaluated for a certain type of car. The braking system will be installed if it can be conclusively demonstrated that the stopping distance under certain controlled conditions at a speed of 30 mi/h is less than 90 ft. It is known that under these conditions the standard deviation of stopping distance is approximately 5 ft. A sample of 150 stops will be made from a speed of 30 mi/h. Let μ represent the mean stopping distance for the new braking system. a. State the appropriate null and alternate hypotheses. b. Find the rejection region if the test is to be conducted at the 5% level. c. Someone suggests rejecting H0 if ft. Is this an appropriate rejection region, or is something wrong? If this is an appropriate rejection region, find the level of the test. Otherwise explain what is wrong. d. Someone suggests rejecting H0 if ft. Is this an appropriate rejection region, or is something wrong? If this is an appropriate rejection region, find the level of the test. Otherwise, explain what is wrong. e. Someone suggests rejecting H0 if or if ft. Is this an appropriate rejection region, or is something wrong? If this is an appropriate rejection region, find the level of the test. Otherwise, explain what is wrong. A test is made of the hypotheses H0 : μ ≤ 10 versus H1 : μ > 10. For each of the following situations, determine whether the decision was correct, a type I error occurred, or a type II error occurred. a. μ = 8, H0 is rejected. b. μ = 10, H0 is not rejected. c. μ = 14, H0 is not rejected. d. μ = 12, H0 is rejected. A vendor claims that no more than 10% of the parts she supplies are defective. Let p denote the actual proportion of parts that are defective. A test is made of the hypotheses H0 : p ≤ 0.10 versus H1 : p > 0.10. For each of the following situations, determine whether the decision was correct, a type I error occurred, or a type II error occurred. a. The claim is true, and H0 is rejected. b. The claim is false, and H0 is rejected. c. The claim is true, and H0 is not rejected. d. The claim is false, and H0 is not rejected. A company that manufactures steel wires guarantees that the mean breaking strength (in kN) of the wires is greater than 50. They measure the strengths for a sample of wires and test H0 : μ ≤ 50 versus H1 : μ > 50. a. If a type I error is made, what conclusion will be drawn regarding the mean breaking strength? b. If a type II error is made, what conclusion will be drawn regarding the mean breaking strength? Washers used in a certain application are supposed to have a thickness of 2 millimeters. A quality control engineer measures the thicknesses for a sample of washers and tests H0 : μ = 2 versus H1 : μ ≠ 2. a. If a type I error is made, what conclusion will be drawn regarding the mean washer thickness? b. If a type II error is made, what conclusion will be drawn regarding the mean washer thickness? Page 488 8. A hypothesis test is to be performed, and it is decided to reject the null hypothesis if P ≤ 0.10. If H0 is in fact true, what is the maximum probability that it will be rejected? A new process is being considered for the liquefaction of coal. The old process yielded a mean of 15 kg of distillate fuel per kilogram of hydrogen consumed in the process. Let μ represent the mean of the new process. A test of H0 : μ ≤ 15 versus H1 : μ > 15 will be performed. If H0 is rejected, the new process will be put into production. The new process will be put into production if it is in fact better than the old process. Which procedure provides a smaller probability that the process will be put into production? a. Assume that the new process is put into production. Which of the following results is the test likely to produce? Why? The test is conducted at the 5% level or the test is conducted at the 1% level? 10. Let μ denote the mean yield strength of a certain type of steel bar, in units of mPa. A test will be made of H0 : μ ≤ 250 versus H1 : μ > 250. Three rejection regions are considered: (i) , (ii) , and (iii) . The levels of the tests are 0.10, 0.05, and 0.01. Match each rejection region with its level. 11. It is desired to check the calibration of a scale by weighing a standard 10-gram weight 100 times. Let μ be the population mean reading on the scale, so that the scale is in calibration if μ = 10 and out of calibration if μ ≠ 10. A test is made of the hypotheses H0 : μ = 10 versus H1 : μ ≠ 10. a. Assume that the scale is in calibration, but the conclusion is reached that the scale is not in calibration. Which type of error is this? b. Assume that the scale is not in calibration. Is it possible to make a Type I error? Explain. c. Assume that the scale is not in calibration. Is it possible to make a Type II error? Explain. 12. Scores on a certain IQ test are known to have a mean of 100. A random sample of 60 students attend a series of coaching classes before taking the test. Let μ be the population mean IQ score that would occur if every student took the coaching classes. The classes are successful if they increase the mean IQ score. A test is made of the hypotheses H0 : μ ≤ 100 versus H1 : μ > 100. a. Assume that the classes are successful but the conclusion is reached that the classes might not be successful. Which type of error is this? b. Assume that the classes are not successful. Is it possible to make a Type I error? Explain. c. Assume that the classes are not successful. Is it possible to make a Type II error? Explain. 13. Find the rejection region for a 5% level test of μ ≥ 10 versus H1 : μ < 10 is . Find the rejection region for a 1% level test. 14. A coin has probability p of landing heads when tossed. A test will be made of the hypotheses H0 : p ≤ 0.1 versus H1 : p > 0.1, as follows. The coin will be tossed once. If it comes up heads, H0 will be rejected. If it comes up tails, H0 will not be rejected. a. If the true value of p is 0.1, what is the probability that the test results in a type I error? b. If the true value of p is 0.4, what is the probability that the test results in a Type II error? 6.13 Power A hypothesis test results in a type II error if H0 is not rejected when it is false. The power of a test is the probability of rejecting H0 when it is false. Therefore Power = 1 − P(type II error) To be useful, a test must have reasonably small probabilities of both type I and type II errors. The type I error is kept small by choosing a small value of α as the significance level. Then the power of the test is calculated. If the power is large, then the probability of a type II error is small as well, and the test is a useful one. Note that power calculations are generally done before data are collected. The purpose of a power calculation is to determine whether or not a hypothesis test, when performed, is likely to reject H0 in the event that H0 is false. As an example of a power calculation, Page 489 assume that a new chemical process has been developed that may increase the yield over that of the current process. The current process is known to have a mean yield of 80 and a standard deviation of 5, where the units are the percentage of a theoretical maximum. If the mean yield of the new process is shown to be greater than 80, the new process will be put into production. Let μ denote the mean yield of the new process. It is proposed to run the new process 50 times and then to test the hypothesis at a significance level of 5%. If H0 is rejected, it will be concluded that μ > 80, and the new process will be put into production. Let us assume that if the new process had a mean yield of 81, then it would be a substantial benefit to put this process into production. If it is in fact the case that μ = 81, what is the power of the test, that is, the probability that H0 will be rejected? Before presenting the solution, we note that in order to compute the power, it is necessary to specify a particular value of μ, in this case μ = 81, for the alternate hypothesis. The reason for this is that the power is different for different values of μ. We will see that if μ is close to H0, the power will be small, while if μ is far from H0, the power will be large. Computing the power involves two steps: 1. Compute the rejection region. 2. Compute the probability that the test statistic falls in the rejection region if the alternate hypothesis is true. This is the power. We'll begin to find the power of the test by computing the rejection region, using the method illustrated in Example 6.28 in Section 6.12. We must first find the null distribution. We know that the statistic has a normal distribution with mean μ and standard deviation , where n = 50 is the sample size. Under H0, we take μ = 80. We must now find an approximation for σ. In practice this can be a difficult problem, because the sample has not yet been drawn, so there is no sample standard deviation s. There are several ways in which it may be possible to approximate σ. Sometimes a small preliminary sample has been drawn, for example in a feasibility study, and the standard deviation of this sample may be a satisfactory approximation for σ. In other cases, a sample from a similar population may exist, whose standard deviation may be used. In this example, there is a long history of a currently used process, whose standard deviation is 5. Let's say that it is reasonable to assume that the standard deviation of the new process will be similar to that of the current process. We will therefore assume that the population standard deviation for the new process is σ = 5 and that . Figure 6.27 (page 490) presents the null distribution of . Since H0 specifies that μ ≤ 80, large values of disagree with H0, so the P-value will be the area to the right of the observed value of . The P-value will be less than or equal to 0.05 if falls into the upper 5% of the null distribution. This upper 5% is the rejection region. The critical point has a z-score of 1.645, so its value is 80 + (1.645)(0.707) = 81.16. We will reject H0 if . This is the rejection region. Page 490 FIGURE 6.27 The hypothesis test will be conducted at a significance level of 5%. The rejection region for this test is the region where . We are now ready to compute the power, which is the probability that will fall into the rejection region if the alternate hypothesis μ = 81 is true. Under this alternate hypothesis, the distribution of is normal with mean 81 and standard deviation 0.707. Figure 6.28 presents the alternate distribution and the null distribution on the same plot. Note that the alternate distribution is obtained by shifting the null distribution so that the mean becomes the alternate mean of 81 rather than the null mean of 80. Because the alternate distribution is shifted over, the probability that the test statistic falls into the rejection region is greater than it is under H0. To be specific, the z-score under H1 for the critical point 81.16 is z = (81.16 − 81)/0.707 = 0.23. The area to the right of z = 0.23 is 0.4090. This is the power of the test. FIGURE 6.28 The rejection region, consisting of the upper 5% of the null distribution, is shaded. The z-score of the critical point is z0 = 1.645 under the null distribution and z1 = 0.23 under the alternate. The power is the area of the rejection region under the alternate distribution, which is 0.4090. A power of 0.4090 is very low. It means that if the mean yield of new process is actually equal to 81, there is only a 41% chance that the proposed experiment will detect the improvement over the old process and allow the new process to be put into production. It would be unwise to invest time and money to run this experiment, since it has a large chance to fail. It is natural to wonder how large the power must be for a test to be worthwhile to perform. As with P-values, there is no scientifically valid dividing line between sufficient and insufficient power. In general, tests with power greater than 0.80 or perhaps 0.90 are considered acceptable, but there are no well-established rules of thumb. Page 491 We have mentioned that the power depends on the value of μ chosen to represent the alternate hypothesis and is larger when the value is far from the null mean. Example 6.30 illustrates this. Example 6.30 Find the power of the 5% level test of H0 : μ ≤ 80 versus H1 : μ > 80 for the mean yield of the new process under the alternative μ = 82, assuming n = 50 and σ = 5. Solution We have already completed the first step of the solution, which is to compute the rejection region. We will reject H0 if . Figure 6.29 presents the alternate and null distributions on the same plot. The z-score for the critical point of 81.16 under the alternate hypothesis is z = (81.16 − 82)/0.707 = −1.19. The area to the right of z = −1.19 is 0.8830. This is the power. FIGURE 6.29 The rejection region, consisting of the upper 5% of the null distribution, is shaded. The z-score of the critical point is z0 = 1.645 under the null distribution and z1 = −1.19 under the alternate. The power is the area of the rejection region under the alternate distribution, which is 0.8830. Since the alternate distribution is obtained by shifting the null distribution, the power depends on which alternate value is chosen for μ, and can range from barely greater than the significance level α all the way up to 1. If the alternate mean is chosen very close to the null mean, the alternate curve will be almost identical with the null, and the power will be very close to α. If the alternate mean is far from the null, almost all the area under the alternate curve will lie in the rejection region, and the power will be close to 1. When power is not large enough, it can be increased by increasing the sample size. When planning an experiment, one can determine the sample size necessary to achieve a desired power. Example 6.31 illustrates this. Example 6.31 In testing the hypothesis H0 : μ ≤ 80 versus H1 : μ > 80 regarding the mean yield of the new process, how many times must the new process be run so that a test conducted at a significance level of 5% will have power 0.90 against the alternative μ = 81, if it is assumed that σ = 5? Page 492 Solution Let n represent the necessary sample size. We first use the null distribution to express the critical point for the test in terms of n. The null distribution of is normal with mean 80 and standard deviation . Therefore the critical point is . Now, we use the alternate distribution to obtain a different expression for the critical point in terms of n. Refer to Figure 6.30. The power of the test is the area of the rejection region under the alternate curve. This area must be 0.90. Therefore the z-score for the critical point, under the alternate hypothesis, is z = −1.28. The critical point is therefore . We now have two different expressions for the critical point. Since there is only one critical point, these two expressions are equal. We therefore set them equal and solve for n. FIGURE 6.30 To achieve a power of 0.90 with a significance level of 0.05, the z-score for the critical point must be z0 = 1.645 under the null distribution and z1 = −1.28 under the alternate distribution. Solving for n yields n ≈ 214. The critical point can be computed by substituting this value for n into either side of the previous equation. The critical point is 80.56. Using a Computer to Calculate Power We have presented a method for calculating the power, and the sample size needed to attain a specified power, for a one-tailed large-sample test of a population mean. It is reasonably straightforward to extend this method to compute power and sample sizes for two-tailed tests and for tests for proportions. It is more difficult to compute power for a t test, F test, or chi-square test. Computer packages, however, can compute power and sample sizes for all these tests. We present some examples. Example 6.32 A pollster will conduct a survey of a random sample of voters in a community to estimate the proportion who support a measure on school bonds. Let p be the proportion of the population who support the measure. The pollster will test H0 : p = 0.50 versus H1 : p ≠ 0.50 at the 5% level. If 200 voters are sampled, what is the power of the test if the true value of p is 0.55? Page 493 Solution The following computer output (from MINITAB) presents the solution: Power and Sample Size Test for One Proportion Testing proportion = 0.5 (versus not = 0.5) Alpha = 0.05 Alternative Proportion 0.55 Sample Size 200 Power 0.292022 The first two lines of output state that this is a power calculation for a test for a single population proportion p. The next two lines state the null and alternate hypotheses, and the significance level of the test. Note that we have specified a two-tailed test with significance level α = 0.05. Next is the alternative proportion, which is the value of p (0.55) that we are assuming to be true when the power is calculated. The sample size has been specified to be 200, and the power is computed to be 0.292. Example 6.33 Refer to Example 6.32. How many voters must be sampled so that the power will be 0.8 when the true value of p = 0.55? Solution The following computer output (from MINITAB) presents the solution: Power and Sample Size Test for One Proportion Testing proportion = 0.5 (versus not = 0.5) Alpha = 0.05 Alternative Proportion 0.55 Sample Size 783 Target Power 0.8 Actual Power 0.800239 The needed sample size is 783. Note that the actual power is slightly higher than 0.80. Because the sample size is discrete, it is not possible to find a sample size that provides exactly the power requested (the target power). So MINITAB calculates the smallest sample size for which the power is greater than that requested. Example 6.34 Shipments of coffee beans are checked for moisture content. High moisture content indicates possible water contamination, leading to rejection of the shipment. Let μ represent the mean moisture content (in percent by weight) in a shipment. Five moisture measurements will be made on beans chosen at random from the shipment. A test of the hypothesis Page 494 H0 : μ ≤ 10 versus H1 : μ > 10 will be made at the 5% level, using the Student's t test. What is the power of the test if the true moisture content is 12% and the standard deviation is σ = 1.5%? Solution The following computer output (from MINITAB) presents the solution: Power and Sample Size 1-Sample t Test Testing mean = null (versus > null) Calculating power for mean = null + difference Alpha = 0.05 Assumed standard deviation = 1.5 Difference 2 Sample Size 5 Power 0.786485 The power depends only on the difference between the true mean and the null mean, which is 12 − 10 = 2, and not on the means themselves. The power is 0.786. Note that the output specifies that this is the power for a one-tailed test. Example 6.35 Refer to Example 6.34. Find the sample size needed so that the power will be at least 0.9. Solution The following computer output (from MINITAB) presents the solution: Power and Sample Size 1-Sample t Test Testing mean = null (versus > null) Calculating power for mean = null + difference Alpha = 0.05 Assumed standard deviation = 1.5 Difference 2 Sample Size 7 Target Power 0.9 Actual Power 0.926750 The smallest sample size for which the power is 0.9 or more is 7. The actual power is 0.927. To summarize, power calculations are important to ensure that experiments rejected. To make the Bonferroni adjustment, simply multiply the P-value by the number of tests performed. Here are two examples. Example 6.40 Four different coating formulations are tested to see if they reduce the wear on cam gears to a value below 100 μm. The null hypothesis H0 : μ ≥ 100 μm is tested for each formulation, and the results are The operator suspects that formulation C may be effective, but he knows that the P-value of 0.005 is unreliable, because several tests have been performed. Use the Bonferroni adjustment to produce a reliable P-value. Solution Four tests were performed, so the Bonferroni adjustment yields P = (4)(0.005) = 0.02 for formulation C. So the evidence is reasonably strong that formulation C is in fact effective. Example 6.41 In Example 6.40, assume the P-value for formulation C had been 0.03 instead of 0.005. What conclusion would you reach then? Solution The Bonferroni adjustment would yield P = (4)(0.03) = 0.12. This is probably not strong enough evidence to conclude that formulation C is in fact effective. Since the original P-value was small, however, it is likely that one would not want to give up on formulation C quite yet. The Bonferroni adjustment is conservative; in other words, the P-value it produces is never smaller than the true P-value. So when the Bonferroni-adjusted P-value is small, the null hypothesis can be rejected conclusively. Unfortunately, as Example 6.41 shows, there are many occasions in which the original P-value is small enough to arouse a strong suspicion that a null hypothesis may be false, but the Bonferroni adjustment does not allow the hypothesis to be rejected. When the Bonferroni-adjusted P-value is too large to reject a null hypothesis, yet the original P-value is small enough to lead one to suspect that the hypothesis is in fact false, often the best thing to do is to retest the hypothesis that appears to be false, using data from a new experiment. If the P-value is again small, this time without multiple tests, this provides real evidence against the null hypothesis. Real industrial processes are monitored frequently by sampling and testing process output to see whether it meets specifications. Every so often, the output appears to be outside the specifications. But in these cases, how do we know whether the process is really Page 500 malfunctioning (out of control) or whether the result is a type I error? This is a version of the multiple testing problem that has received much attention. The subject of statistical quality control (see Chapter 10) is dedicated in large part to finding ways to overcome the multiple testing problem. Exercises for Section 6.14 1. 2. 3. An agricultural scientist tests six types of fertilizer, labeled A, B, C, D, E, and F, to determine whether any of them produces an increase in the yield of lima beans over that obtained with the current fertilizer. For fertilizer C, the increase in yield is statistically significant at the 0.05 level. For the other five, the increase is not statistically significant. The scientist concludes that the yield obtained with fertilizer C is greater than that of the current fertilizer. Explain why this conclusion is not justified. Refer to Exercise 1. The P-value for fertilizer C was 0.03. Use the Bonferroni correction to produce a reliable P-value for this fertilizer. Can you reject H0? Six different settings are tried on a machine to see if any of them will reduce the proportion of defective parts. For each setting, an appropriate null hypothesis is tested to see if the proportion of defective parts has been reduced. The six P-values are 0.34, 0.27, 0.002, 0.45, 0.03, and 0.19. a. Find the Bonferroni-adjusted P-value for the setting whose P-value is 0.002. Can you 4. 5. 6. conclude that this setting reduces the proportion of defective parts? Explain. b. Find the Bonferroni-adjusted P-value for the setting whose P-value is 0.03. Can you conclude that this setting reduces the proportion of defective parts? Explain. Five different variations of a bolt-making process are run to see if any of them can increase the mean breaking strength of the bolts over that of the current process. The P-values are 0.13, 0.34, 0.03, 0.28, and 0.38. Of the following choices, which is the best thing to do next? i. Implement the process whose P-value was 0.03, since it performed the best. ii. Since none of the processes had Bonferroni-adjusted P-values less than 0.05, we should stick with the current process. iii. Rerun the process whose P-value was 0.03 to see if it remains small in the absence of multiple testing. iv. Rerun all the five variations again, to see if any of them produce a small P-value the second time around. Twenty formulations of a coating are being tested to see if any of them reduce gear wear. For the Bonferroni-adjusted P-value for a formulation to be 0.05, what must the original Pvalue be? Five new paint additives have been tested to see if any of them can reduce the mean drying time from the current value of 12 minutes. Ten specimens have been painted with each of the new types of paint, and the drying times (in minutes) have been measured. The results are as follows: 1 2 3 4 5 6 7 8 9 10 A 14.573 12.012 13.449 13.928 13.123 13.254 12.772 10.948 13.702 11.616 B 10.393 10.435 11.440 9.719 11.045 11.707 11.141 9.852 13.694 9.474 Additive C 15.497 9.162 11.394 10.766 11.025 10.636 15.066 11.991 13.395 8.276 D 10.350 7.324 10.338 11.600 10.725 12.240 10.249 9.326 10.774 11.803 E 11.263 10.848 11.499 10.493 10.409 13.196 12.259 11.056 For each additive, perform a hypothesis test of the null hypothesis H0 : μ ≥ 12 against the alternate H1 : μ< 12. You may assume that each population is approximately normal. a. b. What are the P-values for the five tests? On the basis of the results, which of the three following conclusions seems most appropriate? Explain your answer. Page 501 i. At least one of the new additives results in an improvement. 7. ii. None of the new additives result in an improvement. iii. Some of the new additives may result in improvement, but the evidence is inconclusive. Each day for 200 days, a quality engineer samples 144 fuses rated at 15 A and measures the amperage at which they burn out. He performs a hypothesis test of H0 : μ = 15 versus H1 : μ ≠ 15, where μ is the mean burnout amperage of the fuses manufactured that day. a. On 10 of the 200 days, H0 is rejected at the 5% level. Does this provide conclusive evidence that the mean burnout amperage was different from 15 A on at least one of the 200 days? Explain. b. Would the answer to part (a) be different if H0 had been rejected on 20 of the 200 days? Explain. 6.15 Using Simulation to Perform Hypothesis Tests If X1, ..., Xn are normally distributed random variables, with known standard deviations σ1, ..., σn, and U = U(X1, ..., Xn) is a function of X1, ..., Xn, then it will often, but not always, be the case that U is approximately normally distributed and that its standard deviation σU can be estimated. In these cases we can perform hypothesis tests on the mean μU of U. To determine whether U is approximately normally distributed and to estimate σU, simulation can be used. The method is similar to that described in Section 5.10. We illustrate with an example. Let R represent the measurement of the radius of a cylinder, and let H represent the measurement of the height. Assume that these measurements are both unbiased and normally distributed. Let V = πR2H denote the measurement of the volume of the cylinder that is computed from R and H. Now assume that R = 4.8 cm, H = 10.1 cm, and the uncertainties (standard deviations) are σR = 0.1 cm and σH = 0.2 cm. The measured volume is V = π(4.82)(10.1) = 731.06 cm3. Suppose we wish to determine whether we can conclude that the true volume of the cylinder is greater than 700 cm3.Let μV denote the mean of V. Since R and H are unbiased, with fairly small uncertainties, V is nearly unbiased (see the discussion on pages 180–181), so μV is close to the true volume of the cylinder. Therefore we can address the question concerning the true volume by performing a test of the hypotheses H0 : μV ≤ 700 versus H1 : μV > 700. Now if the measured volume V is normally distributed, and if the population standard deviation σV were known, we could perform the test as follows: Under H0, . We observed the value V = 731.06. The P-value for H0 : μV ≤ 700 is P(V ≥ 731.06), where the probability is computed under the assumption that . The z-score is (731.06 − 700)/σV. If σV were known, we could compute z, and then use the z table to find the P-value. To determine whether V is normally distributed, and to approximate σV, we first generate a large number N of values for the radius measurement. We know that the radius measurement is normally distributed with standard deviation σR = 0.1. We do not know the mean radius measurement, which is equal to the true radius, but we can approximate it with Page 502 the observed value 4.8. Therefore we generate from the distribution N(4.8, 0.12). Similarly, we generate simulated volume measurements from the distribution N(10.1, 0.22). Then we compute . A normal probability plot of the be used to determine whether V is approximately normal. Figure 6.31 presents a normal probability plot for a sample of 1000 values of can then . The normality assumption is satisfied. The sample standard deviation of the simulated values was 34.42 cm3. We therefore compute the z-score to be z = (731.06 − 700)/34.42 = 0.90. The P-value is 0.1841. FIGURE 6.31 Normal probability plot for 1000 simulated volumes. The assumption of normality appears to be justified. Testing Hypotheses with Bootstrap Confidence Intervals Given a level 100(1 − α)% confidence interval for a parameter, such as a population mean μ, we can reject at level 100α% the null hypothesis that the parameter is equal to any given value outside the interval, and cannot reject at level 100α% the null hypothesis that the parameter is equal to any given value inside the interval (see the discussion beginning on page 416). This idea can be applied to a bootstrap confidence interval to construct a fixed-level hypothesis test. We present an example. Example 6.42 In Section 5.10, an approximate 95% confidence interval for the mean mileage, in mpg, of a population of trucks was found by a bootstrap method to be (4.7643, 6.4757). Can we conclude at the 5% level that the population mean mileage differs from 5 mpg? From 7 mpg? Solution A 95% confidence interval, whether computed by the bootstrap or by other means, contains the values that are not rejected at the 5% level. Therefore we conclude at the 5% Page 503 level that the population mean differs from 7 mpg, but we cannot conclude at the 5% level that it differs from 5 mpg. Randomization Tests Randomization tests, also called permutation tests, were among the earliest methods developed to test the difference between two population means. While they require virtually no assumptions about the distribution of the data, they involve a great deal of computation, and so did not become truly feasible until recently. We present an example. A crop scientist wants to determine whether the yield of lettuce will be increased by using a fertilizer with a higher nitrogen content. She conducts an experiment involving 20 plots. Ten plots are chosen at random to be treated with fertilizer A, which has a low nitrogen content. The other 10 plots are treated with fertilizer B, which has a higher nitrogen content. The following table presents, for each plot, the treatment applied (A or B), and the yield, in number of heads of lettuce harvested. Plot Number 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 Treatment A A B B A A A B B A A B A A B B Yield 145 142 144 141 142 155 143 157 152 143 103 151 150 148 150 162 The null hypothesis is that there is no difference between the fertilizers with regard to yield; in other words, the yield for each plot would have been the same no matter which fertilizer it had received. For example the yield in plot 1 would have been 145 whether fertilizer A or B was used. If H0 is true, then the 20 observed yields are constants, and the yields associated with fertilizer A are a simple random sample of 10 of these 20 constant yields. Denote the mean of the 10 yields associated with fertilizer A by , and the mean of the 10 yields associated with fertilizer B by . Since the main interest in the experiment is to determine whether fertilizer B increases the yield, a reasonable test statistic is the difference . The observed value of this statistic is 151.5 − 141.5 = 10.0. The larger the value of the test statistic, the stronger the evidence against H0. The strength of the evidence is measured by the P-value. We now discuss how to calculate the P-value. The experiment involves a random choice of 10 plots out of 20 to receive fertilizer A. In general, the number of different choices of k items to be selected from a group of n items is denoted and is given by (see Equation 2.12 in Section 2.2 for a derivation) Therefore the number of possible choices for these 10 plots is . This means that there are 184,756 ways that the experiment could have come out; the actual experiment Page 504 consists of observing one of these ways chosen at random. The choice that was actually made provided a value of for the test statistic. Since, under H0, the yields do not depend on which fertilizer was used, we could in principle compute the value of the test statistic for each of the 184,756 possible outcomes of the experiment. The P-value is the probability, under H0, that the test statistic has a value greater than or equal to 10. This probability is equal to the proportion of the 184,756 possible outcomes of the experiment for which . Table 6.6 presents a partial listing of the possible outcomes of the experiment. TABLE 6.6 Possible outcomes of the randomized experiment Outcome Yields Assigned to A 1 103 141 142 142 143 143 144 144 145 148 2 103 141 142 142 143 143 144 144 145 149 ⋮ ⋮ 184,755 148 150 150 151 151 152 155 157 158 162 184,756 149 150 150 151 151 152 155 157 158 162 Yields Assigned to B 149 150 150 151 151 152 155 157 158 162 148 150 150 151 151 152 155 157 158 162 ⋮ 103 141 142 142 143 143 144 144 145 149 103 141 142 142 143 143 144 144 145 148 139.5 153.5 14.0 139.6 153.4 13.8 ⋮ ⋮ ⋮ 153.4 139.6 −13.8 153.5 139.5 −14.0 The exact P-value can be found by completing Table 6.6 and then determining the proportion of outcomes for which . This procedure is called a randomization test, or permutation test. Computing the exact P-value is an intensive task, even for a computer. An easier method, which is just as good in practice, is to work with a randomly generated collection of outcomes instead. This is done by generating a large number (1000 or more) randomly chosen subsets of 10 yields to assign to treatment A. Each chosen subset corresponds to one of the possible outcomes of the experiment, and for each subset, the value of the test statistic is computed. The P-value is approximated with the proportion of the randomly chosen outcomes for which the value of the test statistic is greater than or equal to the observed value of 10. Table 6.7 presents the first 5 and the last of 1000 randomly chosen outcomes for the experiment. Of the first five outcomes, none of them have values of greater than or equal to 10, so the estimate of the P-value based on these five is 0/5 = 0. Of course, five outcomes is not enough upon which to base a reliable conclusion. Out of the full set of 1000 outcomes, only 9 had values of greater than or equal to 10. We therefore estimate the P-value to be 0.009, which is small enough to conclusively reject the null hypothesis that there is no difference between the fertilizers. It seems reasonable to conclude that fertilizer B tends to increase the yield. TABLE 6.7 One thousand simulated outcomes of the randomized experiment Outcome Yields Assigned to A 1 157 151 144 150 142 150 155 144 143 141 2 143 103 158 151 142 151 155 150 148 141 3 162 158 144 141 148 155 103 143 144 157 4 145 151 143 141 150 142 162 148 149 158 5 152 148 144 142 157 155 162 103 150 151 ⋮ ⋮ 1000 144 152 143 155 142 148 143 145 158 151 Yields Assigned to B 145 148 142 143 103 152 158 149 162 151 142 144 149 144 143 162 157 150 152 145 143 150 142 152 145 150 142 149 151 151 144 155 157 103 152 150 144 151 143 142 145 150 158 149 144 143 141 143 151 142 ⋮ 144 103 149 142 150 162 150 141 151 157 147.70 145.30 −2.40 144.20 148.80 4.60 145.50 147.50 2.00 148.90 144.10 −4.80 146.40 146.60 0.20 ⋮ ⋮ ⋮ 148.10 144.90 −3.20 Randomized experiments like the one just described play a major role in scientific investigations and are discussed more fully in Chapter 9. When no outliers are present, it has been shown that the Student's t test for the difference between means (see Section 6.7) provides a good approximation to the randomization test when two treatments are being compared. Data from randomized experiments can generally be treated as though they consist of Page 505 random samples from different populations; this is the approach we will take in Chapter 9. Freedman, Pisani, and Purves (2007) contains a good discussion of this topic. Rank tests (see Section 6.9) are sometimes used for these experiments as well. Randomization tests can be used in some cases when the data consist of two samples from two populations, which is the situation discussed in Section 6.7. Thus randomization tests can be an alternative to the t test for the difference between means when outliers are present. More information on randomization tests can be found in Efron and Tibshirani (1993). Using Simulation to Estimate Power For some tests, it is difficult to calculate power with a formula. In these cases, simulation can often be used to estimate the power. Following is an example. Example 6.43 A new type of weld is being developed. If the mean fracture toughness of this weld is conclusively shown to be greater than 20 ft · lb, the weld will be used in a certain application. Assume that toughness is normally distributed with standard deviation equal to 4 ft · lb. Six welds will be made, and the fracture toughness of each will be measured. A Student's t test will be made of the null hypothesis H0 : μ ≤ 20 versus H1 : μ > 20. If the test is conducted at a significance level of 5%, what is the power of the test if the true mean toughness is 25 ft · lb? Solution Let X1, ..., X6 represent the six sample toughnesses, and let s represent their sample standard deviation. This is a sample from a N(25, 16) distribution. The test statistic is . Under H0, this statistic has a Student's t distribution with Page 506 five degrees of freedom. The null hypothesis will be rejected if the value of the test statistic is greater than t5,.05 = 2.015. The power, therefore, is equal to P(T > 2.015). It is not easy to calculate this probability directly, since in fact the null hypothesis is false, so T does not in fact have a Student's t distribution. We can, however, estimate it with a simulation experiment. We will generate 10,000 samples sample, we will compute the sample mean statistic , each from a N(25, 16) distribution. For each , the sample standard deviation , and the test . Since each simulated sample is drawn from the same distribution as are the actual toughnesses of the welds, each simulated sample is statistically equivalent to a sample of actual welds. We can therefore estimate the power simply by computing the proportion of simulated samples for which the null hypothesis is rejected, that is, for which the value of the test statistic exceeds 2.015. Table 6.8 presents the results for the first 10 samples and for the last one. The rightmost column contains a "1" if the value of the test statistic exceeds 2.015 and a "0" otherwise. TABLE 6.8 Simulated data for Example 6.43 I s* T* T* > 2.015 1 23.24 23.78 15.65 25.67 24.08 25.88 23.05 3.776 1.978 0 2 26.51 19.89 20.53 25.03 28.35 28.01 24.72 3.693 3.131 1 3 4 5 6 7 8 9 10 ⋮ 10,000 28.61 22.84 22.36 26.54 24.05 28.38 23.55 29.98 ⋮ 30.75 28.19 28.69 21.26 28.63 24.42 29.51 21.73 34.65 ⋮ 19.99 29.48 23.93 26.37 24.79 20.32 23.80 19.57 21.17 ⋮ 26.20 20.06 27.37 23.61 20.63 23.74 29.05 25.04 28.43 ⋮ 22.41 30.00 19.51 34.45 25.44 24.14 26.39 22.34 23.43 ⋮ 31.53 21.19 30.28 29.97 26.69 24.66 23.76 29.71 34.44 ⋮ 21.78 26.26 25.44 26.34 25.45 23.56 26.81 23.68 28.68 ⋮ 25.45 4.423 4.046 5.061 2.703 1.615 2.579 3.484 5.559 ⋮ 4.862 3.465 3.291 3.067 4.940 5.394 6.472 2.570 3.825 ⋮ 2.744 1 1 1 1 1 1 1 1 ⋮ 1 The null hypothesis is rejected for 9 of the first 10 samples. If we were to base our results on the first 10 samples, we would estimate the power to be 0.9. Of course, 10 samples is not enough. Out of all 10,000 samples, the null hypothesis was rejected for 8366 of them. The estimate of the power is therefore 0.8366. Exercises for Section 6.15 1. Refer to Exercise 6 in Section 5.10. Let μ represent the population mean compressive strength, in MPa. Consider the following null hypotheses: i. H0 : μ = 38.53 ii. H0 : μ = 38.35 iii. H0 : μ = 38.45 iv. H0 : μ = 38.55 a. b. Using the bootstrap data presented in Exercise 6 in Section 5.10, which of these null hypotheses can be rejected at the 5% level if a confidence interval is formed using method 1 on page 391? Using the bootstrap data presented in Exercise 6 in Section 5.10, which of these null hypotheses can be rejected at the 10% level if a confidence interval is formed using method 1 on page 391? Page 507 2. 3. Refer to Exercise 6 in Section 5.10. Let μ represent the population mean compressive strength, in MPa. Generate 1000 bootstrap samples. a. Using the bootstrap data you generated, which of these null hypotheses can be rejected at the 5% level, using method 1 on page 391? b. Using the bootstrap data you generated, which of these null hypotheses can be rejected at the 10% level, using method 1 on page 391? c. If a bootstrap experiment is performed twice on the same data, is it necessary that the results will agree? Explain. In the lettuce yield example presented on page 503, would it be a good idea to use the t test described in Section 6.7 to determine whether the fertilizers differ in their effects on yield? 4. Why or why not? It is suspected that using premium gasoline rather than regular will increase the mileage for automobiles with a particular engine design. Sixteen cars are used in a randomized experiment. Eight are randomly chosen to be tested with regular gasoline, while the other eight are tested with premium gasoline. The results, in mpg, are as follows: Regular: Premium: 29.1 28.4 28.3 30.9 27.1 30.8 17.3 27.6 16.3 30.2 32.0 27.4 35.3 29.9 35.6 29.7 a. 5. Under the null hypothesis that each car would get the same mileage with either type of gasoline, how many different outcomes are possible for this experiment? b. Let and denote the sample mean mileages for the regular and premium groups, respectively. Compute and . c. Perform a randomization test to determine whether it can be concluded that premium gasoline tends to increase mileage. Use the test statistic . Generate at least 1000 random outcomes, and estimate the P-value. d. Use the Student's t test described in Section 6.7 to test the null hypothesis that the mean mileage using regular is greater than or equal to the mean mileage for premium. Is this result reliable? Explain? For the lettuce yield data (page 503), it is thought that the yields from fertilizer A might have a larger variance than the yields from fertilizer B. a. Compute the sample variances and of the yields assigned to A and B, respectively, and the quotient . b. c. Someone suggests using the F test in Section 6.11 for this problem. Is this a good idea? Why or why not? Perform a randomization test of versus , using the test statistic , and a minimum of 1000 random outcomes. (Hint: Proceed just as in the example in the text, but for each outcome compute and 6. 7. rather than , . A fair amount of coding may be required, depending on the software used.) Refer to Exercise 4. Perform a randomization test to determine whether the mileage using regular gasoline has a greater variance than the mileage using premium gasoline. Generate at least 1000 random outcomes. A certain wastewater treatment method is supposed to neutralize the wastewater so that the mean pH is 7. Measurements of pH will be made on seven specimens of treated wastewater, and a test of the hypotheses H0 : μ = 7 versus H1 : μ ≠ 7 will be made using the Student's t test (Section 6.4). Assume that the true mean is μ = 6.5, the pH measurements are normally distributed with mean μ and standard deviation 0.5, and the test is made at the 5% level. a. Let X1, ..., X7 denote the pH measurements, let denote their mean, and let s denote their sample standard deviation. What is the test statistic? For what values of the test statistic will H0 be rejected? b. 8. Generate 10,000 samples from the true distribution of pH measurements. For each sample, compute the test statistic and determine whether H0 is rejected. Estimate the power of the test. This exercise requires ideas from Section 2.6. In a two-sample experiment, when each item in one sample is paired with an item in the other, the paired t test (Section 6.8) can be used to test hypotheses regarding the difference between two population means. If one Page 508 ignores the fact that the data are paired, one can use the two-sample t test (Section 6.7) as well. The question arises as to which test has the greater power. The following simulation experiment is designed to address this question. Let (X1, Y1), ..., (X8, Y8) be a random sample of eight pairs, with X1, ..., X8 drawn from an N(0, 1) population and Y1, ..., Y8 drawn from an N(1, 1) population. It is desired to test H0 : μX − μY = 0 versus H1 : μX − μY ≠ 0. Note that μX = 0 and μY = 1, so the true difference between the means is 1. Also note that the population variances are equal. If a test is to be made at the 5% significance level, which test has the greater power? Let Di = Xi − Yi for i = 1, ..., 10. The test statistic for the paired t test is , where sD is the standard deviation of the Di (see Section 6.8). Its null distribution is Student's t with seven degrees of freedom. Therefore the paired t test will reject H0 if , so the power is . For the two-sample t test when the population variances are equal, the test statistic is , where sp is the pooled standard deviation, which is equal in this case to

Solution These conclusions are much better justified than the ones in Example 7.1. To see why, look at the plot of the time versus volume in the following figure. This experiment has been designed so that time and volume are uncorrelated. It now appears that the time, but not the volume, has an effect on the percentage absorbed. Before making a final conclusion that increasing the time actually causes the percentage absorbed to increase, the scientist must make sure that there are no other potential confounders around. For example, if the ambient temperature varied with each replication of the experiment, and was highly correlated with time, then it might be the case that the temperature, rather than the time, was causing the percentage absorbed to vary. Page 526 Controlled Experiments Reduce the Risk of Confounding In Examples 7.1 and 7.2, the experimenter was able to reduce confounding by assigning values for volume and time so that these two variables were uncorrelated. This is a controlled experiment, because the experimenter could assign the values for these factors (see Section 1.1 for a more complete description of controlled experiments). In controlled experiments, confounding can often be avoided by choosing values for factors in a way so that the factors are uncorrelated. Observational studies are studies in which the values of factors cannot be chosen by the experimenter. Studies involving public health issues, such as the effect of environmental pollutants on human health, are usually observational, because experimenters cannot deliberately expose people to high levels of pollution. In these studies, confounding is often difficult to avoid. For example, people who live in areas with higher levels of pollution may tend to have lower socio-economic status, which may affect their health. Because confounding is difficult to avoid, observational studies must generally be repeated a number of times, under a variety of conditions, before reliable conclusions can be drawn. Inference on the Population Correlation The rest of this section uses some ideas from Section 2.6. When the points (x_i, y_i) are a random sample from a population of ordered pairs, then each point can be thought of as an observation of an ordered pair of random variables (X, Y). The correlation coefficient, or sample correlation, r is then an estimate of the population correlation $\rho X, Y$. If the random variables X and Y have a certain joint distribution called a bivariate normal distribution, then the sample correlation r can be used to construct confidence intervals and perform hypothesis tests on the population correlation. In practice, if X and Y are both normally distributed, then it is a virtual certainty that X and Y will be bivariate normal, so the confidence intervals and tests described subsequently will be valid. (While it is mathematically possible to construct two normal random variables that jointly are not bivariate normal, the conditions under which this occurs are almost never seen in practice.) Page 527 Confidence intervals, and most tests, on $\rho X, Y$ are based on the following result: Let X and Y be random variables with the bivariate normal distribution. Let ρ denote the population correlation between X and Y. Let $(x_1, y_1),\ldots, (x_n, y_n)$ be a random sample from the joint distribution of X and Y. Let r be the sample correlation of the n points. Then the quantity (7.4) is approximately normally distributed, with mean given by (7.5) and variance given by (7.6) Note that μW is a function of the population correlation ρ. To construct confidence intervals, we will need to solve Equation (7.5) for ρ. We obtain (7.7) Example 7.3 In a study of reaction times, the time to respond to a visual stimulus (x) and the time to respond to an auditory stimulus (y) were recorded for each of 10 subjects. Times were measured in ms. The results are presented in the following table. x 161 203 235 176 201 188 228 211 191 178 y 159 206 241 163 197 193 209 189 169 201 Find a 95% confidence interval for the correlation between the two reaction times. Solution Using Equation (7.1), we compute the sample correlation, obtaining r = 0.8159. Next we use Equation (7.4) to compute the quantity W: Page 528 Since W is normally distributed with known standard deviation (Equation 7.6), a 95% confidence interval for μW is given by To obtain a 95% confidence interval for ρ we transform the inequality using Equation (7.7), obtaining Example 7.4 Refer to Example 7.3. Find the P-value for testing H0: ρ ≤ 0.3 versus H1: ρ > 0.3. Solution Under H0 we take ρ = 0.3, so, using Equation (7.5), The standard deviation of W is . It follows that under H0, W ∼ N(0.3095, 0.37802). The observed value of W is W = 1.1444. The z-score is therefore The P-value is 0.0136. We conclude that ρ > 0.3. For testing null hypotheses of the form ρ = 0, ρ ≤ 0, or ρ ≥ 0, a somewhat simpler procedure is available. When ρ = 0, the quantity has a Student's t distribution with n − 2 degrees of freedom. Example 7.5 shows how to use U as a test statistic. Page 529 Example 7.5 Refer to Example 7.3. Test the hypothesis H0: ρ ≤ 0 versus H1: ρ > 0. Solution Under H0 we take ρ = 0, so the test statistic U has a Student's t distribution with n − 2 = 8 degrees of freedom. The sample correlation is r = 0.8159, so the value of U is Consulting the Student's t table with eight degrees of freedom, we find that the P-value is between 0.001 and 0.005. (Software yields P = 0.00200.) It is reasonable to conclude that ρ > 0. Exercises for Section 7.1 1. Compute the correlation coefficient for the following data set. x 1 2 3 4 5 6 7 y 2 1 4 3 7 5 6 For each of the following data sets, explain why the correlation coefficient is the same as for the data set in Exercise 1. a. b. c. 3. x 1 2 3 4 5 6 7 y 2 3 4 6 7 4 7 6 10 8 9 x 11 21 31 41 51 61 71 y 5 4 7 6 10 8 9 Each of the following scatterplots contains one outlier. For each, state whether the correlation coefficient with the outlier is greater than or less than the correlation coefficient without it, and explain briefly. Page 530 4. 5. 6. True or false, and explain briefly: a. If the correlation coefficient is positive, then above-average values of one variable are associated with above-average values of the other. b. If the correlation coefficient is negative, then below-average values of one variable are associated with below-average values of the other. c. If y is usually less than x, then the correlation between y and x will be negative. An investigator collected data on heights and weights of college students. The correlation between height and weight for men was about 0.6, and for women it was about the same. If men and women are taken together, will the correlation between height and weight be more than 0.6, less than 0.6, or about equal to 0.6? It might be helpful to make a rough scatterplot. In a study of ground motion caused by earthquakes, the peak velocity (in m/s) and peak acceleration (in m/s2) were recorded for five earthquakes. The results are presented in the following table. Velocity 1.54 1.60 0.95 1.30 2.92 Acceleration 7.64 8.04 8.04 6.37 3.25 a. b. c. 7. Compute the correlation coefficient between peak velocity and peak acceleration. Construct a scatterplot for these data. Is the correlation coefficient an appropriate summary for these data? Explain why or why not. Someone suggests converting the units from meters to centimeters and from seconds to minutes. What effect would this have on the correlation? A chemical engineer is studying the effect of temperature and stirring rate on the yield of a certain product. The process is run 16 times, at the settings indicated in the following table. The units for yield are percent of a theoretical maximum. Temperature (°C) 110 110 111 111 Stirring Rate (rpm) 30 32 34 36 Yield (%) 70.27 72.29 72.57 74.69 112 112 114 114 117 117 122 122 130 130 143 143 38 40 42 44 46 48 50 52 54 56 58 60 76.09 73.14 75.61 69.56 74.41 73.49 79.18 75.44 81.71 83.03 76.98 80.99 a. 8. Compute the correlation between temperature and yield, between stirring rate and yield, and between temperature and stirring rate. b. Do these data provide good evidence that increasing the temperature causes the yield to increase, within the range of the data? Or might the result be due to confounding? Explain. c. Do these data provide good evidence that increasing the stirring rate causes the yield to increase, within the range of the data? Or might the result be due to confounding? Explain. Another chemical engineer is studying the same process as in Exercise 7, and uses the following experimental matrix. Temperature (°C) 110 110 110 110 121 121 121 121 132 132 132 132 143 143 143 143 Stirring Rate (rpm) 30 40 50 60 30 40 50 60 30 40 50 60 30 40 50 60 Yield (%) 70.27 74.95 77.91 82.69 73.43 73.14 78.27 74.89 69.07 70.83 79.18 78.10 73.71 77.70 74.31 80.99 Page 531 a. 9. Compute the correlation between temperature and yield, between stirring rate and yield, and between temperature and stirring rate. b. Do these data provide good evidence that the yield is unaffected by temperature, within the range of the data? Or might the result be due to confounding? Explain. c. Do these data provide good evidence that increasing the stirring rate causes the yield to increase, within the range of the data? Or might the result be due to confounding? Explain. d. Which experimental design is better, this one or the one in Exercise 7? Explain. A blood pressure measurement consists of two numbers: the systolic pressure, which is the maximum pressure taken when the heart is contracting, and the diastolic pressure, which is the minimum pressure taken at the beginning of the heartbeat. Blood pressures were measured, in millimeters of mercury (mmHg), for a sample of 14 adults. The following table presents the results. Assume that blood pressures follow a bivariate normal distribution. Systolic 134 115 113 123 119 118 130 116 133 112 107 110 108 105 Diastolic 87 83 77 77 69 88 76 70 91 75 71 74 69 66 Find a 95% confidence interval for ρ, the correlation between systolic and diastolic blood pressure. b. Can you conclude that ρ > 0.5? c. Can you conclude that ρ > 0? In a sample of 300 steel rods, the correlation coefficient between diameter and length was r = 0.15. a. Find the P-value for testing H0: ρ ≤ 0 vs. H1: ρ > 0. Can you conclude that ρ > 0? b. Does the result in part (a) allow you to conclude that there is a strong correlation between eccentricity and smoothness? Explain. 11. The article "Drift in Posturography Systems Equipped with a Piezoelectric Force Platform: Analysis and Numerical Compensation" (L. Quagliarella, N. Sasanelli, and V. Monaco, IEEE Transactions on Instrumentation and Measurement, 2008: 997–1004) reported the results of an experiment to determine the effect of load on the drift in signals derived from a piezoelectric force plate. The correlation coefficient y between output and time under a load of 588 N was −0.9515. Measurements were taken 100 times per second for 300 seconds, for a total of 30,000 measurements. Find a 95% confidence interval for the population correlation ρ. 12. Phonics is an instructional method in which children are taught to connect sounds with letters or groups of letters. The article "Predictive Accuracy of Nonsense Word Fluency for English Language Learners" (M. Vanderwood, D. Linklater, and K. Healy, School Psychology Review, 2008:5–17) reports that in a sample of 134 English-learning students, the correlation between the score on a phonics test given in first grade and a reading comprehension score given in third grade was r = 0.25. Can you conclude that there is a positive correlation between phonics test score and the reading comprehension score? 13. The article "'Little Ice Age' Proxy Glacier Mall Balance Records Reconstructed from Tree Rings in the Mt. Waddington Area, British Columbia Coast Mountains, Canada" (S. Larocque and D. Smith, The Holocene, 2005:748–757) evaluates the use of tree ring widths to estimate changes in the masses of glaciers. For the Sentinel glacier, the net mass balance (change in mass between the end of one summer and the end of the next summer) was measured for 23 years. During the same time period, the tree ring index for white bark pine trees was measured, and the sample correlation between net mass balance and tree ring index was r = −0.509. Can you conclude that the population correlation ρ differs from 0? 14. A scatterplot contains four points: (−2,−2), (−1,−1), (0,0), and (1,1). A fifth point, (2,y), is to be added to the plot. Let r represent the correlation between x and y. a. Find the value of y so that r = 1. b. Find the value of y so that r = 0. c. Find the value of y so that r = 0.5. d. Find the value of y so that r = −0.5. e. Give a geometric argument to show that there is no value of y for which r = −1. Page 532 7.2 The Least-Squares Line When two variables have a linear relationship, the scatterplot tends to be clustered around a line known as the least-squares line (see Figure 7.1 and 7.2 in Section 7.1). In this section we will learn how to compute the least-squares line and how it can be used to draw conclusions from data. We begin by describing a hypothetical experiment. Springs are used in applications for their ability to extend (stretch) under load. The stiffness of a spring is measured by the "spring constant," which is the length that the spring will be extended by one unit of force or load.1 To make sure that a given spring functions appropriately, it is necessary to estimate its spring constant with good accuracy and precision. In our hypothetical experiment, a spring is hung vertically with the top end fixed, and weights are hung one at a time from the other end. After each weight is hung, the length of the spring is measured. Let $x_1,\ldots, x_n$ represent the weights, and let l_i represent the length of the spring under the load x_i. Hooke's law states that (7.8) where β0 is the length of the spring when unloaded and β1 is the spring constant. Let yi be the measured length of the spring under load xi. Because of measurement error, yi will differ from the true length li. We write (7.9) where εi is the error in the ith measurement. Combining (7.8) and (7.9), we obtain (7.10) In Equation (7.10) yi is called the dependent variable, xi is called the independent variable, β0 and β1 are the regression coefficients, and εi is called the error. Equation (7.10) is called a linear model. Table 7.1 presents the results of the hypothetical experiment, and Figure 7.9 presents the scatterplot of y versus x. We wish to use these data to estimate the spring constant β1 and the unloaded length β0. If there were no measurement error, the points would lie on a straight line with slope β1 and intercept β0, and these quantities would be easy to determine. Because of measurement error, β0 and β1 cannot be determined exactly, but they can be estimated by calculating the least-squares line. TABLE 7.1 Measured lengths of a spring under various loads Weight (lb) x 0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 FIGURE 7.9 Measured Length (in.) y 5.06 5.01 5.12 5.13 5.14 5.16 5.25 5.19 5.24 5.46 5.40 5.57 5.47 5.53 5.61 5.59 5.61 5.75 5.68 5.80 Plot of measured lengths of a spring versus load. Page 533 We write the equation of the line as (7.11) The quantities and are called the least-squares coefficients. The coefficient , the slope of the least-squares line, is an estimate of the true spring constant β1, and the coefficient , the intercept of the least-squares line, is an estimate of the true unloaded length β0. FIGURE 7.10 Plot of measured lengths of a spring versus load. The least-squares line is superimposed. The vertical distance from a data point (xi, yi) to the point on the line is the ith residual ei. The least-squares line is the line that minimizes the sum of the squared residuals. The least-squares line is the line that fits the data "best." How do we define "best"? To understand the Page 533 how "best." For each data point (xi, yi), the vertical distance to the point on the least-squares line is (see Figure 7.10). The quantity is called the fitted value, and the quantity ei is called the residual associated with the point (xi, yi). The residual ei is the difference between the value yi observed in the data and the fitted value predicted by the least-squares line. This is the vertical distance from the point to the line. Points above the least-squares line have positive residuals, and points below the least-squares line have negative residuals. The closer the residuals are to 0, the closer the fitted values are to the observations and the better the line fits the data. We define the least-squares line to be the line for which the sum of the squared residuals is minimized. In this sense, the least-squares line fits the data better than any other line. In the Hooke's law example, there is only one independent variable (weight), since it is reasonable to assume that the only variable affecting the length of the spring is the amount of weight hung from it. In other cases, we may need to use several independent variables. Page 534 For example, to predict the yield of a certain crop, we might need to know the amount of fertilizer used, the amount of water applied, and various measurements of chemical properties of the soil. Linear models with only one independent variable are known as simple linear regression models. Linear models with more than one independent variable are called multiple regression models. This chapter covers simple linear regression. Multiple regression is covered in Chapter 8. Computing the Equation of the Least-Squares Line To compute the equation of the least-squares line, we must determine the values for the slope and the intercept that minimize the sum of the squared residuals first express ei in terms of and . To do this, we : (7.12) Therefore and are the quantities that minimize the sum (7.13) These quantities are (7.14) (7.15) Derivations of these results are provided at the end of this section. Page 535 Computing Formulas The quantities and need to be computed in order to determine the equation of the least-squares line, and as we will soon see, the quantity needs to be computed in order to determine how well the line fits the data. When computing these quantities by hand, there are alternate formulas that are often easier to use. They are given in the following box. Computing Formulas The expressions on the right are equivalent to those on the left, and are often easier to compute: (7.16) (7.17) (7.18) Example 7.6 Using the Hooke's law data in Table 7.1, compute the least-squares estimates of the spring constant and the unloaded length of the spring. Write the equation of the least-squares line. Solution The estimate of the spring constant is , and the estimate of the unloaded length is . From Table 7.1 we compute: Page 536 Using Equations (7.14) and (7.15), we compute The equation of the least-squares line is and we obtain . Substituting the computed values for Using the equation of the least-squares line, we can compute the fitted values and the residuals for each point (xi, yi) in the Hooke's law data set. The results are presented in Table 7.2. The point whose residual is shown in Figure 7.10 (page 534) is the one where x = 2.2. TABLE 7.2 Measured lengths of a spring under various loads, with fitted values and residuals Weight x 0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 Measured Length y 5.06 5.01 5.12 5.13 5.14 5.16 5.25 5.19 5.24 5.46 5.40 5.57 5.47 5.53 5.61 5.59 5.61 5.75 5.68 5.80 Fitted Value 5.00 5.04 5.08 5.12 5.16 5.20 5.25 5.29 5.33 5.37 5.41 5.45 5.49 5.53 5.57 5.61 5.65 5.70 5.74 5.78 Residual e 0.06 −0.03 0.04 0.01 −0.02 −0.04 0.00 −0.10 −0.09 0.09 −0.01 0.12 −0.02 0.00 0.04 −0.02 −0.04 0.05 −0.06 0.02 In the Hooke's law example, the quantity β0 + β1x represents the true length of the spring under a load x. Since and are estimates of the true values β0 and β1, the quantity is an estimate of β0 + β1x. Examples 7.7 and 7.8 illustrate this. Example 7.7 Using the Hooke's law data, estimate the length of the spring under a load of 1.3 lb. Solution In Example 7.6, we computed and . The equation of the least-squares line is y = 4.9997 + 0.2046x. Using the value x = 1.3, we estimate the length of the spring under a load of 1.3 lb to be Example 7.8 Using the Hooke's law data, estimate the length of the spring under a load of 1.4 lb. Solution The estimate is in. In Example 7.8, note that the measured length at a load of 1.4 was 5.19 in. (see Table 7.2). But the least-squares estimate of 5.29 in. is based on all the data and is more precise Page 537 (has smaller uncertainty). We will learn how to compute uncertainties for the estimates in Section 7.3. Interpreting the Slope of the Least-Squares Line If the x-values of two points on a line differ by 1, their y-values will differ by an amount equal to the slope of the line. For example, if a line has a slope of 4, then two points whose x-values differ by 1 will have y-values that differ by 4. This fact enables us to interpret the slope of the leastsquares regression line. If the values of the explanatory variable for two individuals differ by 1, their predicted values will differ by . If the values of the explanatory variable differ by an amount d, then their predicted values will differ by . We have worded the interpretation of the slope very carefully. The slope is the estimated difference in y values for two different individuals whose x values differ by 1. This does not mean that changing the value of x for a particular individual will cause that individual's y value to change. The following example will help make this clear. Example 7.9 A study is done in which a sample of men were weighed, and then each man was tested to see how much weight he could lift. The explanatory variable (x) was the man's weight, and the outcome (y) was the amount he could lift. The least-squares line was computed to be y = 50 + 0.6x. Joe is a weightlifter. After looking at the equation of the least-squares line, he reasons as follows: "The slope of the least-squares regression line is 0.6. Therefore, if I gain 10 pounds, I'll be able to lift 6 pounds more, because (0.6)(10) = 6." Is he right? Solution No, he is not right. You can't improve your weightlifting ability simply by putting on weight. What the slope of 0.6 tells us is that if two men have weights that differ by 10 pounds, then on the average the heavier man will be able to lift 6 pounds more than the lighter man. This does not mean that an individual man can increase his weightlifting ability by increasing his weight. The Estimates Are Not the Same as the True Values It is important to understand the difference between the least-squares estimates and , and the true values β0 and β1. The true values are constants whose values are unknown. The estimates are quantities that are computed from the data. We may use the estimates as approximations for the true values. In principle, an experiment such as the Hooke's law experiment could be repeated many times. The true values β0 and β1 would remain constant over the replications of the experiment. But each replication would produce different data, and thus different values of the estimates and . Therefore and are random variables, since their values vary from experiment to experiment. To make full use of these estimates, we will need to be able to compute their standard deviations. We will discuss this topic in Section 7.3. Page 538 The Residuals Are Not the Same as the Errors A collection of points $(x_1, y_1),\ldots, (x_n, y_n)$ follows a linear model if the x and y coordinates are related through the equation yi = β0 + β1xi + εi. It is important to understand the difference between the residuals ei and the errors εi. Each residual ei is the difference between an observed, or measured, value yi and the fitted value estimated from the least- squares line. Since the values yi are known and the values can be computed from the data, the residuals can be computed. In contrast, the errors εi are the differences between the yi and the values β0+β1xi. Since the true values β0 and β1 are unknown, the errors are unknown as well. Another way to think of the distinction is that the residuals are the vertical distances from the observed values yi to the least-squares line , and the errors are the distances from the yi to the true line y = β0+β1x. Summary Given points $(x_1, y_1),\ldots, (x_n, y_n)$: ■ The quantities intercept β0. ■ For any x and can be thought of as estimates of a true slope β1 and a true is an estimate of the quantity β0 + β1x. Don't Extrapolate Outside the Range of the Data What if we wanted to estimate the length of the spring under a load of 100 lb? The least-squares estimate is 4.9997 + (0.2046)(100) = 25.46 in. Should we believe this? No. None of the weights in the data set were this large. It is likely that the spring would be stretched out of shape, so Hooke's law would not hold. For many variables, linear relationships hold within a certain range, but not outside it. If we extrapolate a least-squares line outside the range of the data, therefore, there is no guarantee that it will properly describe the relationship. If we want to know how the spring will respond to a load of 100 lb, we must include weights of 100 lb or more in the data set. Summary Do not extrapolate a fitted line (such as the least-squares line) outside the range of the data. The linear relationship may not hold there. Don't Use the Least-Squares Line When the Data Aren't Linear In Section 7.1, we learned that the correlation coefficient should be used only when the relationship between x and y is linear. The same holds true for the least-squares line. Page 539 When the scatterplot follows a curved pattern, it does not make sense to summarize it with a straight line. To illustrate this, Figure 7.11 presents a plot of the relationship between the height y of an object released from a height of 256 ft and the time x since its release. The relationship between x and y is nonlinear. The least-squares line does not fit the data well. FIGURE 7.11 The relationship between the height of a free-falling object and the time in free fall is not linear. The least-squares line does not fit the data well and should not be used to predict the height of the object at a given time. In some cases the least-squares line can be used for nonlinear data, after a process known as variable transformation has been applied. This topic is discussed in Section 7.4. Another Look at the Least-Squares Line The expression (7.14) for can be rewritten in a way that provides a useful interpretation. Starting with the definition of the correlation coefficient r (Equation 7.2 in Section 7.1) and multiplying both sides by yields the result (7.19) Equation (7.19) allows us to interpret the slope of the least-squares line in terms of the correlation coefficient. The units of , the slope of the least-squares line, must be units of y/x. The correlation coefficient r is a unitless number that measures the strength of the linear relationship between x and y. Equation (7.19) shows that the slope is proportional to the correlation coefficient, where the constant of proportionality is the quantity sy/sx that adjusts for the units in which x and y are measured. Using Equation (7.19), we can write the equation of the least-squares line in a useful form: Substituting for in the equation for the least-squares line and rearranging terms yields (7.20) Page 540 Combining Equations (7.19) and (7.20) yields (7.21) Thus the least-squares line is the line that passes through the center of mass of the scatterplot , with slope . Measuring Goodness-of-Fit A goodness-of-fit statistic is a quantity that measures how well a model explains a given set of data. A linear model fits well if there is a strong linear relationship between x and y. We mentioned in Section 7.1 that the correlation coefficient r measures the strength of the linear relationship between x and y. Therefore r is a goodness-of-fit statistic for the linear model. We will now describe how r measures the goodness-of-fit. Figure 7.12 presents Galton's data on forearm lengths versus heights. The points on the scatterplot are (xi, yi) where xi is the height of the ith man and yi is the length of his forearm. Both the least-squares line and the horizontal line are superimposed. Now imagine that we must predict the length of one of the forearms. If we have no knowledge of the man's height, we must predict his forearm length to be the average . Our prediction error is . If we predict the length of each forearm this way, the sum of squared prediction errors will be . If, on the other hand, we know the Page 541 height of each man before predicting the length of his forearm, we can use the leastsquares line, and we will predict the ith forearm length to be . The prediction error will be the residual , and the sum of squared prediction errors is . The strength of the linear relationship can be measured by computing the reduction in sum of squared prediction errors obtained by using rather than . This is the difference . The bigger this difference is, the more tightly clustered the points are around the least-squares line and the stronger the linear relationship is between x and y. Thus is a goodness-of-fit statistic. FIGURE 7.12 Heights and forearm lengths of men. The least-squares line and the horizontal line are superimposed. There is a problem with using as a goodness-of-fit statistic, however. This quantity has units, namely the squared units of y. We could not use this statistic to compare the goodness-of-fit of two models fit to different data sets, since the units would be different. We need to use a goodness-of-fit statistic that is unitless, so that we can measure goodness-of-fit on an absolute scale. This is where the correlation coefficient r comes in. It is shown at the end of this section that (7.22) The quantity r2, the square of the correlation coefficient, is called the coefficient of determination. It is the reduction in the sum of the squared prediction errors obtained by using rather than , expressed as a fraction of the sum of squared prediction errors , obtained by using . This interpretation of r2 is important to know. In Chapter 8, we will see how it can be generalized to provide a measure of the goodness-of-fit of linear relationships involving several variables. For a visual interpretation of r2, look at Figure 7.12. For each point (xi, yi) on the scatterplot, the quantity is the vertical distance from the point to the horizontal line . The quantity is the vertical distance from the point to the least-squares line. Thus the quantity measures the overall spread of the points around the line and the quantity measures the overall spread of the points around the least-squares line. The quantity therefore measures the reduction in the spread of the points obtained by using the least-squares line rather than . The coefficient of determination r2 expresses this reduction as a proportion of the spread around . The sums of squares appearing in this discussion are used so often that statisticians have given them names. They call the error sum of squares and the total sum of squares. Their difference is called the regression sum of squares. Clearly, the following relationship holds: Using the preceding terminology, we can write Equation (7.22) as Page 542 Since the total sum of squares is just the sample variance of the y without dividing by n − 1, statisticians (and others) often refer to r2 as the proportion of the variance in y explained by regression. Derivation of the Least-Squares Coefficients and We derive Equations (7.14) and (7.15). The least-squares coefficients and are the quantities that minimize the sum We compute these values by taking partial derivatives of S with respect to setting them equal to 0. Thus and and are the quantities that solve the simultaneous equations (7.23) (7.24) These equations can be written as a system of two linear equations in two unknowns: (7.25) (7.26) We solve Equation (7.25) for , obtaining This establishes Equation (7.15). Now substitute for in Equation (7.26) to obtain (7.27) Solving Equation (7.27) for , we obtain Page 543 To establish Equation (7.14), we will show that . Examples 7.16 and 7.18.) Also and that Derivation of Equation (7.22) We first show that (7.28) This result is known as the analysis of variance identity. To derive it, we begin by adding and subtracting from the left-hand side: Now we need only to show that . Since and (7.29) Page 544 Therefore (7.30) Now , so Substituting into Equation (7.30), we obtain This establishes the analysis of variance identity. To derive Equation (7.22), Equation (7.29) implies that (7.31) Square both sides of Equation (7.31) and sum. Using the identity , so Substituting and canceling, we obtain By the analysis of variance identity, Therefore Page 545 Exercises for Section 7.2 1. Each month for several months, the average temperature in °C (x) and the number of pounds of steam (y) consumed by a certain chemical plant were measured. The least-squares line computed from the resulting data is y = 245.82 + 1.13x. a. Predict the number of pounds of steam consumed in a month where the average temperature is 65°C. b. If two months differ in their average temperatures by 5°C, by how much do you predict 2. 3. 4. 5. 6. the number of pounds of steam consumed to differ? In a study of the relationship between the Brinell hardness (x) and tensile strength in ksi (y) of specimens of cold drawn copper, the least-squares line was y = −196.32 + 2.42x. a. Predict the tensile strength of a specimen whose Brinell hardness is 102.7. b. If two specimens differ in their Brinell hardness by 3, by how much do you predict their tensile strengths to differ? A least-squares line is fit to a set of points. If the total sum of squares is , and the error sum of squares is , compute the coefficient of determination r 2. A least-squares line is fit to a set of points. If the total sum of squares is , and the error sum of squares is , compute the coefficient of determination r 2. In Galton's height data (Figure 7.1, in Section 7.1), the least-squares line for predicting forearm length (y) from height (x) is y = −0.2967 + 0.2738x. a. Predict the forearm length of a man whose height is 70 in. b. How tall must a man be so that we would predict his forearm length to be 19 in.? c. All the men in a certain group have heights greater than the height computed in part (b). Can you conclude that all their forearms will be at least 19 in. long? Explain. In a study of reaction to degree of warping, in mm, of a copper plate (y) to temperature in °C (x), the following summary statistics were calculated: n = 40, Compute the correlation r between the degree of warping and the temperature. Compute the error sum of squares, the regression sum of squares, and the total sum of squares. c. Compute the least-squares line for predicting warping from temperature. d. Predict the warping at a temperature of 40°C. e. At what temperature will we predict the warping to be 0.5 mm? f. Assume it is important that the warping not exceed 0.5 mm. An engineer suggests that if the temperature is kept below the level computed in part (e), we can be sure that the warping will not exceed 0.5 mm. Is this a correct conclusion? Explain. Moisture content in percent by volume (x) and conductivity in mS/m (y) were measured for 50 soil specimens. The means and standard deviations were . The correlation between conductivity and moisture was computed to be r = 0.85. Find the equation of the least-squares line for predicting soil conductivity from moisture content. Curing times in days (x) and compressive strengths in MPa (y) were recorded for several concrete specimens. The means and standard deviations of the x and y values were , sy = 100. The correlation between curing time and compressive strength was computed to be r = 0.7. Find the equation of the least-squares line to predict compressive strength from curing time. Foot ulcers are a common problem for people with diabetes. Higher skin temperatures on the foot indicate an increased risk of ulcers. The article "An Intelligent Insole for Diabetic a. b. 7. 8. 9. Patients with the Loss of Protective Sensation" (Kimberly Anderson, M.S. Thesis, Colorado School of Mines), reports measurements of temperatures, in °F, of both feet for 18 Page 546 diabetic patients. The results are presented in the following table. Left Foot 80 85 75 89 87 78 89 76 89 87 78 80 87 86 76 88 Right Foot 80 85 80 86 82 78 81 82 85 80 89 a. Construct a scatterplot of the right foot temperature (y) versus the left foot temperature (x). Verify that a linear model is appropriate. b. Compute the least-squares line for predicting the right foot temperature from the left foot temperature. c. If the left foot temperatures of two patients differ by 2 degrees, by how much would you predict their right foot temperatures to differ? d. Predict the right foot temperature for a patient whose left foot temperature is 81 degrees. 10. A blood pressure measurement consists of two numbers: the systolic pressure, which is the maximum pressure taken when the heart is contracting, and the diastolic pressure, which is the minimum pressure taken at the beginning of the heartbeat. Blood pressures were measured, in mmHg for a sample of 16 adults. The following table presents the results. Systolic 134 115 113 123 Diastolic 87 83 77 77 119 118 130 116 133 112 107 110 108 105 157 154 69 88 76 70 91 75 71 74 69 66 103 94 a. Construct a scatterplot of diastolic pressure (y) versus systolic pressure (x). Verify that a linear model is appropriate. b. Compute the least-squares line for predicting the diastolic pressure from the systolic pressure. c. If the systolic pressures of two patients differ by 10 mHg, by how much would you predict their diastolic pressures to differ? d. Predict the diastolic pressure for a patient whose systolic pressure is 125 mmHg. 11. Structural engineers use wireless sensor networks to monitor the condition of dams and bridges. The article "Statistical Analysis of Vibration Modes of a Suspension Bridge Using Spatially Dense Wireless Sensor Network" (S. Pakzad and G. Fenves, Journal of Structural Engineering, 2009:863–872) describes an experiment in which accelerometers were placed on the Golden Gate Bridge for the purpose of estimating vibration modes. For 18 vertical modes, the system was underdamped (damping ratio < 1). Following are the damping ratios and frequencies for those modes. Damping Ratio 0.3 0.3 0.3 0.4 0.4 0.4 0.5 0.5 0.5 0.6 0.6 Frequency (Hz) 2.72 2.84 3.77 2.07 2.20 2.34 2.61 1.80 1.93 1.53 0.77 1.26 0.6 0.7 0.7 0.8 0.8 a. b. c. 1.66 0.89 1.00 0.66 1.13 0.37 Construct a scatterplot of frequency (y) versus damping ratio (x). Verify that a linear model is appropriate. Compute the least-squares line for predicting frequency from damping ratio. If two modes differ in damping ratio by 0.2, by how much would you predict their frequencies to differ? Page 547 d. e. Predict the frequency for modes with damping ratio 0.75. Should the equation be used to predict the frequency for modes that are overdamped (damping ratio > 1)? Explain why or why not. f. For what damping ratio would you predict a frequency of 2.0? 12. The article "Effect of Environmental Factors on Steel Plate Corrosion Under Marine Immersion Conditions" (C. Soares, Y. Garbatov, and A. Zayed, Corrosion Engineering, Science and Technology, 2011:524–541) describes an experiment in which nine steel specimens were submerged in seawater at various temperatures, and the corrosion rates were measured. The results are presented in the following table (obtained by digitizing a graph). Temperature (°C) 26.6 26.0 27.4 21.7 14.9 11.3 15.0 8.7 8.2 a. b. c. d. e. Corrosion (mm/yr) 1.58 1.45 1.13 0.96 0.99 1.05 0.82 0.68 0.56 Construct a scatterplot of corrosion (y) versus temperature (x). Verify that a linear model is appropriate. Compute the least-squares line for predicting corrosion from temperature. Two steel specimens whose temperatures differ by 10°C are submerged in seawater. By how much would you predict their corrosion rates to differ? Predict the corrosion rate for steel submerged in seawater at a temperature of 20°C. Compute the fitted values. Compute the residuals. Which point has the residual with the largest magnitude? g. h. Compute the correlation between temperature and corrosion rate. Compute the regression sum of squares, the error sum of squares, and the total sum of squares. i. Divide the regression sum of squares by the total sum of squares. What is the relationship between this quantity and the correlation coefficient? 13. Concrete expands both horizontally and vertically over time. The article "Evaluation of the Expansion Attained to Date by Concrete Affected by Alkali-Silica Reaction. Part III: Application to Existing Structures" (M. Bérubé, N. Smaoui, et al., Canadian Journal of Civil Engineering, 2005:463–479), reports measurements of horizontal and vertical expansion (in units of parts per hundred thousand) made at several locations on a bridge in Quebec City in Canada. The results are presented in the following table. Horizontal 20 15 43 5 18 24 32 10 21 a. b. c. d. e. Vertical 55 80 58 68 57 69 63 Construct a scatterplot of vertical expansion (y) versus horizontal expansion (x). Verify that a linear model is appropriate. Compute the least-squares line for predicting vertical expansion (y) from horizontal expansion (x). Compute the fitted value and residual for each point. If the horizontal expansion increases by 10 parts per hundred thousand, by how much would you predict the vertical expansion to increase or decrease? Predict the vertical expansion for a location where the horizontal expansion is 30 parts per hundred thousand. Can the least-squares line be used to predict the vertical expansion in a location where the horizontal expansion is 60 parts per hundred thousand? If so, predict it. If not, explain why not. Page 548 f. For what horizontal expansion would you predict a vertical expansion of 65 parts per hundred thousand? 14. An engineer wants to predict the value for y when x = 4.5, using the following data set. x 1 2 3 4 5 6 7 8 9 10 1.3 2.9 4.5 8.7 12.0 0.26 0.83 1.06 1.50 2.16 2.48 Construct a scatterplot of the points (x, y). Should the least-squares line be used to predict the value of y when x = 4.5? If so, compute the least-squares line and the predicted value. If not, explain why not. c. Compute the natural logarithm of each value of y. x 1 2 3 4 5 6 7 8 9 10 y 0.2 0.3 0.5 0.5 2 u = ln y −1.61 −1.20 −0.69 −0.69 5.6 7 8 9 10 1.3 2.9 4.5 8.7 12.0 d. e. f. Let z denote the natural logarithm of each of the points (x, z), where z = ln y. Use the least-squares line to predict the value of z when x = 4.5. Is this an appropriate method of prediction? Explain why or why not. e. Let denote the predicted value of z computed in part (d). Let . Explain why is a reasonable predictor of the value of y when x = 4.5. 15. A simple random sample of 100 men aged 25–34 averaged 70 inches in height, and had a standard deviation of 3 inches. Their incomes averaged $34,900 and had a standard deviation of $17,200. Fill in the blank: From the least-squares line, we would predict that the income of a man 70 inches tall would be _____. i. less than $34,900. ii. greater than $34,900. iii. equal to $34,900. iv. We cannot tell unless we know the correlation. 16. A mixture of sucrose and water was heated on a hot plate, and the temperature (in °C) was recorded each minute for 20 minutes by three thermocouples. The results are shown in the following table. a. b. Time 0 1 2 3 4 5 6 7 8 9 10 11 12 T1 T2 T3 20 18 29 32 37 36 46 46 56 58 62 64 79 18 22 22 37 45 46 54 64 69 65 80 21 11 26 35 33 35 44 41 43 63 68 62 65 80 13 14 15 16 17 18 19 20 84 82 87 98 103 103 102 74 87 93 90 98 91 103 103 107 104 a. Construct a scatterplot of temperature (y) versus time (x) using T1 as the value for temperature. b. Compute the least-squares line for estimating temperature as a function of time, using T2 as the value for temperature. c. Compute the least-squares line for estimating the temperature as a function of time, using T3 as the value for temperature. d. It is desired to compute a single line to estimate temperature as a function of time. One person suggests averaging the three slope estimates to obtain a single slope estimate, and averaging the three intercept estimates to obtain a single intercept estimate. Find the equation of the line that results from this method. e. Someone else suggests averaging the three temperature measurements at each time to obtain Compute the least-squares line using as the value for temperature. f. Are the results of parts (d) and (e) different? 17. Carrie and Ryan have both computed the least-squares line using data for which the standard deviation of the x-values are and the standard deviation of the y-values are equal. Carrie gets a value of 0.5 for the slope, and Ryan gets a value of 2. One of them is right. Which one? Page 549 7.3 Uncertainties in the Least-Squares Coefficients In Section 7.2, the linear model was presented (Equation 7.10): Here εi is the error in the ith observation yi. In practice, εi represents the accumulation of error from many sources. For example, in the Hooke's law data, εi can be affected by errors in measuring the length of the spring, errors in measuring the weights of the loads placed on the spring, variations in the elasticity of the spring due to changes in ambient temperature or metal fatigue, and so on. If there were no error, the points would lie exactly on the least-squares line, and the slope and intercept of the least-squares line would equal the true values β0 and β1. Because of error, the points are scattered around the line, and the quantities and do not equal β0 and β1. Each time the process is repeated, the values of εi, and thus the values of and , will be different. In other words, εi, and , and specific, the errors εi create uncertainty in the estimates are random variables. To be more and . It is intuitively clear that if the εi tend to be small in magnitude, the points will be tightly clustered around the line, and the uncertainty in the least-squares estimates and will be small. On the other hand, if the εi tend to be large in magnitude, the points will be widely scattered around the line, and the uncertainties (standard deviations) in the least-squares estimates and will be larger. Assume we have n data points $(x_1, y_1),\ldots, (x_n, y_n)$, and we plan to fit the least-squares line. In order for the estimates and to be useful, we need to estimate just how large their uncertainties are. In order to do this, we need to know something about the nature of the errors εi. We will begin by studying the simplest situation, in which four important assumptions are satisfied. These are given in the following box. Assumptions for Errors in Linear Models In the simplest situation, the following assumptions are satisfied: 1. The errors ε1,..., εn are random and independent. In particular, the magnitude of any error εi does not influence the value of the next error εi+1. 2. The errors ε1,..., εn all have mean 0. 3. The errors ε1,..., εn all have the same variance, which we denote by σ2. 4. The errors ε1,..., εn are normally distributed. These assumptions are restrictive, so it is worthwhile to discuss briefly the degree to which it is acceptable to violate them in practice. When the sample size is large, the normality assumption (4) becomes less important. Mild violations of the assumption of constant variance (3) do not matter too much, but severe violations should be corrected. In Section 7.4, we discuss methods to correct certain violations of these assumptions. Page 550 Under these assumptions, the effect of the εi is largely governed by the magnitude of the variance σ2, since it is this variance that determines how large the errors are likely to be. Therefore, in order to estimate the uncertainties in and , we must first estimate the error variance σ2. This can be done by first computing the fitted values and then the residuals ei = yi − . The magnitude of the variance is reflected in the degree of spread of the points around the least-squares line. Specifically, the vertical distance from each data point (xi, yi) to the least-squares line is given by the residual ei (see Figure 7.10 in Section 7.2). The spread of the points around the line can be measured by the sum of the squared residuals σ2 is the quantity s2 . The estimate of the error variance given by (7.32) The estimate of the error variance is thus the average of the squared residuals, except that we divide by n − 2 rather than n. The reason for this is that since the least-squares line minimizes the sum , the residuals tend to be a little smaller than the errors εi. It turns out that dividing by n − 2 rather than n appropriately compensates for this. There is an equivalent formula for s2, involving the correlation coefficient r, that is often easier to calculate. (7.33) We present a brief derivation of this result. Equation (7.22) (in Section 7.2) shows that . Then , and it follows that Under assumptions 1 through 4, the observations yi are also random variables. In fact, since yi = β0 + β1xi + εi, it follows that yi has a normal distribution with mean β0 + β1xi and variance σ2. In particular, β1 represents the change in the mean of y associated with an increase of one unit in the value of x. Summary In the linear model yi = β0 + β1xi + εi, under assumptions 1 through 4, the observations y1,..., yn are independent random variables that follow the normal distribution. The mean and variance of yi are given by The slope β1 represents the change in the mean of y associated with an increase of one unit in the value of x. Page 551 We can now calculate the means and standard deviations of and . The standard deviations of these estimates are of course the uncertainties. Both and and . The standard can be expressed as linear combinations of the yi, so their means can be found using Equation (2.51) and their standard deviations can be found using Equation (2.55) (both equations in Section 2.5). Specifically, algebraic manipulation of Equations (7.14) and (7.15) (in Section 7.2) yields (7.34) (7.35) Using the fact that each of the yi has mean β0 + β1xi and variance σ2, Equations (2.51) and (2.55) yield the following results, after further manipulation: Now and are unbiased, since their means are equal to the true values. They are also normally distributed, because they are linear combinations of the independent normal random variables yi. In practice, when computing the standard deviations, we usually don't know the value of σ, so we approximate it with s. Summary Under assumptions 1 through 4 (page 549), ■ The quantities and are normally distributed random variables. ■ The means of ■ The standard deviations of and are the true values β0 and respectively. and are estimated with (7.36) and (7.37) is an estimate of the error standard deviation σ. where Example Page 552 7.10 For the Hooke's law data, compute s, , and unloaded length, and find their uncertainties. . Estimate the spring constant and the Solution In Example 7.6 (in Section 7.2) we computed , and . The correlation is , . Now compute . , The More Spread in the x Values, the Better (Within Reason) In the expressions for both of the uncertainties quantity and in Equations (7.36) and (7.37), the appears in a denominator. This quantity measures the spread in the x values; when divided by the constant n − 1, it is just the sample variance of the x values. It follows that other things being equal, an experiment performed with more widely spread out x values will result in smaller uncertainties for and , and thus more precise estimation of the true values β0 and β1. Of course, it is important not to use x values so large or so small that they are outside the range for which the linear model holds. Summary When one is able to choose the x values, it is best to spread them out widely. The more spread out the x values, the smaller the uncertainties in and . Specifically, the uncertainty in is inversely proportional to , or equivalently, to the sample standard deviation of $x_1, x_2,\ldots, x_n$. Caution: If the range of x values extends beyond the range where the linear model holds, the results will not be valid. There are two other ways to improve the accuracy of the estimated regression line. First, one can increase the size of the sum by taking more observations, thus adding more terms to the sum. And second, one can decrease the size of the error variance σ2, for example, by measuring more precisely. These two methods usually add to the cost of a project, however, while simply choosing more widely spread x values often does not. Page 553 Example 7.11 Two engineers are conducting independent experiments to estimate a spring constant for a particular spring. The first engineer suggests measuring the length of the spring with no load, and then applying loads of 1, 2, 3, and 4 lb. The second engineer suggests using loads of 0, 2, 4, 6, and 8 lb. Which result will be more precise? By what factor? Solution The sample standard deviation of the numbers 0, 2, 4, 6, 8 is twice as great as the sample standard deviation of the numbers 0, 1, 2, 3, 4. Therefore the uncertainty for the first engineer is twice as large as for the second engineer, so the second engineer's estimate is twice as precise. We have made two assumptions in the solution to this example. First, we assumed that the error variance σ2 is the same for both engineers. If they are both using the same apparatus and the same measurement procedure, this could be a safe assumption. But if one engineer is able to measure more precisely, this needs to be taken into account. Second, we have assumed that a load of 8 lb is within the elastic zone of the spring, so that the linear model applies throughout the range of the data. Inferences on the Slope and Intercept Given a scatterplot with points $(x_1, y_1),\ldots, (x_n, y_n)$, we can compute the slope and intercept of the least-squares line. We consider these to be estimates of a true slope β1 and intercept β0. We will now explain how to use these estimates to find confidence intervals for, and to test hypotheses about, the true values β1 and β0. It turns out that the methods for a population mean, based on the Student's t distribution, can be easily adapted for this purpose. We have seen that under assumptions 1 through 4, and are normally distributed with means β0 and β1, and standard deviations that are estimated by and and . The quantities have Student's t distributions with n − 2 degrees of freedom. The number of degrees of freedom is n − 2 because in the computation of and we divide the sum of squared residuals by n − 2. When the sample size n is large enough, the normal distribution is nearly indistinguishable from the Student's t and may be used instead. However, most software packages use the Student's t distribution regardless of sample size. Summary Under assumptions 1 through 4, the quantities and have Student's t distributions with n − 2 degrees of freedom. Confidence intervals for β0 and β1 can be derived in exactly the same way as the Student's t based confidence interval for a population mean. Let tn−2, α/2 denote the point on the Page 554 Student's t curve with n − 2 degrees of freedom that cuts off an area of α/2 in the right-hand tail. Level 100(1 − α)% confidence intervals for β0 and β1 are given by (7.38) where We illustrate the preceding method with some examples. Example 7.12 Find a 95% confidence interval for the spring constant in the Hooke's law data. Solution The spring constant is β1. We have previously computed Section 7.2) and (Example 7.6 in (Example 7.10). The number of degrees of freedom is n − 2 = 20 − 2 = 18, so the t value for a 95% confidence interval is t18,.025 = 2.101. The confidence interval for β1 is therefore We are 95% confident that the increase in the length of the spring that will result from an increase of 1 lb in the load is between 0.181 and 0.228 in. Of course, this confidence interval is valid only within the range of the data (0 to 3.8 lb). Example 7.13 In the Hooke's law data, find a 99% confidence interval for the unloaded length of the spring. Solution The unloaded length of the spring is β0. We have previously computed (Example 7.6) and (Example 7.10). The number of degrees of freedom is n − 2 = 20 − 2 = 18, so the t value for a 99% confidence interval is t18,.005 = 2.878. The confidence interval for β0 is therefore We are 99% confident that the unloaded length of the spring is between 4.928 and 5.071 in. We can perform hypothesis tests on β0 and β1 as well. We present some examples. Page 555 Example 7.14 The manufacturer of the spring in the Hooke's law data claims that the spring constant β1 is at least 0.215 in./lb. We have estimated the spring constant to be . Can we conclude that the manufacturer's claim is false? Solution This calls for a hypothesis test. The null and alternate hypotheses are The quantity has a Student's t distribution with n − 2 = 20 − 2 = 18 degrees of freedom. Under H0, we take β1 = 0.215. The test statistic is therefore We have previously computed and . The value of the test statistic is therefore Consulting the Student's t table, we find that the P-value is between 0.10 and 0.25. (Software yields P = 0.181.) We cannot reject the manufacturer's claim on the basis of these data. Example 7.15 Can we conclude from the Hooke's law data that the unloaded length of the spring is more than 4.9 in.? Solution This requires a hypothesis test. The null and alternate hypotheses are The quantity has a Student's t distribution with n − 2 = 20 − 2 = 18 degrees of freedom. Under H0, we take β0 = 4.9. The test statistic is therefore Page 556 We have previously computed and . The value of the test statistic is therefore Consulting the Student's t table, we find that the P-value is less than 0.0005. (Software yields P = 0.000402.) We can conclude that the unloaded length of the spring is more than 4.9 in. The most commonly tested null hypothesis is H0: β1 = 0. If this hypothesis is true, then there is no tendency for y either to increase or decrease as x increases. This implies that x and y have no linear relationship. In general, if the hypothesis that β1 = 0 is not rejected, the linear model should not be used to predict y from x. Example 7.16 The ability of a welded joint to elongate under stress is affected by the chemical composition of the weld metal. In an experiment to determine the effect of carbon content (x) on elongation (y), 39 welds were stressed until fracture, and both carbon content (in parts per thousand) and elongation (in percent) were measured. The following summary statistics were calculated: Assuming that x and y follow a linear model, compute the estimated change in elongation due to an increase of one part per thousand in carbon content. Should we use the linear model to predict elongation from carbon content? Solution The linear model is y = β0+βxx+ε, and the change in elongation (y) due to a one part per thousand increase in carbon content (x) is β1. The null and alternate hypotheses are The null hypothesis says that increasing the carbon content does not affect the elongation, while the alternate hypothesis says that is does. The quantity has a Student's t distribution with n − 2 = 39 − 2 = 37 degrees of freedom. Under H0, β1 = 0. The test statistic is therefore Page 557 We compute and : The value of the test statistic is The t table shows that the P-value is greater than 0.20. (Software yields P = 0.272.) We cannot conclude that the linear model is useful for predicting elongation from carbon content. Inferences on the Mean Response In Example 7.8 (Section 7.2), we estimated the length of a spring under a load of 1.4 lb to be 5.29 in. Since this estimate was based on measurements that were subject to uncertainty, the estimate itself is subject to uncertainty. For the estimate to be more useful, we should construct a confidence interval around it to reflect its uncertainty. We now describe how to do this, for the general case where the load on the spring is x lb. If a measurement y were taken of the length of the spring under a load of x lb, the mean of y would be the true length (or "mean response") β0 + β1x, where β1 is the true spring constant and β0 is the true unloaded length of the spring. We estimate this length with and . Since are normally distributed with means β0 and β1, respectively, it follows that is normally distributed with mean β0 + β1x. To use to find a confidence interval, we must know its standard deviation. The standard deviation can be derived by expressing as a linear combination of the yi and using Equation (2.55) (in Section 2.5). Equations (7.34) and (7.35) express the yi. Since and as linear combinations of , these equations, after some algebraic manipulation, yield (7.39) Equation (2.55) now can be used to derive an expression for the standard deviation of . The standard deviation depends on the error variance σ2. Since in practice we don't usually know the value of σ, we approximate it with s. The standard deviation of is approximated by (7.40) Page 558 The quantity has a Student's t distribution with n − 2 degrees of freedom. We can now provide the expression for a confidence interval for the mean response. A level 100(1 − α)% confidence interval for the quantity β0 + β1x is given by (7.41) where Example . 7.17 Using the Hooke's law data, compute a 95% confidence interval for the length of a spring under a load of 1.4 lb. Solution We will calculate , and , and use expression (7.41). The number of points is n = 20. In Example 7.10, we computed s = 0.0575. In Example 7.6 (in Section 7.2), we computed , and . Using x = 1.4, we now compute Using Equation (7.40) with x = 1.4, we obtain The number of degrees of freedom is n − 2 = 20 − 2 = 18. We find that the t value is t18,.025 = 2.101. Substituting into expression (7.41) we determine the 95% confidence interval for the length β0 + β1 (1.4) to be Example 7.18 In a study of the relationship between the permeability (y) of human skin and its electrical resistance (x), the data presented in the following table were obtained for 50 skin specimens, each 2.54 cm2 in area. Here permeability is measured in μm/h and resistance is measured in kΩ. Using a linear model, find a 95% confidence interval for the mean permeability for skin specimens with resistance 25 kΩ.(From the article "Multi-Species Assessment of Electrical Resistance as a Skin Integrity Marker for In Vitro Percutaneous Absorption Studies," D. J. Davies, R. J. Ward, and J. R. Heylings, Toxicology in Vitro, 2004:351–358; values obtained by digitizing a graph.) Page 559 Resistance 10.09 11.37 12.08 12.25 13.08 13.52 13.75 14.19 15.13 15.13 16.07 16.51 17.18 18.34 18.84 18.34 18.17 18.67 20.28 20.17 20.17 19.94 21.72 20.94 21.44 22.05 21.66 21.72 22.66 22.10 22.82 23.99 24.82 25.70 25.98 Permeability 11.58 13.89 11.77 9.02 9.65 9.91 12.42 9.93 10.08 5.42 12.99 10.49 8.13 5.78 7.47 7.93 9.95 9.73 14.33 7.52 5.96 8.10 10.44 7.30 7.56 7.58 6.49 5.90 7.01 9.14 8.69 4.66 8.88 5.92 7.01 26.37 26.42 26.75 26.92 27.80 27.80 28.63 28.47 28.19 28.97 29.85 30.02 31.79 34.28 34.61 6.66 5.35 4.05 7.16 7.07 6.47 6.50 5.30 4.93 4.36 4.28 4.88 6.02 4.67 6.12 Solution We calculate the following quantities (the computing formulas on page 535 may be used): The estimate of the mean permeability for skin specimens with a resistance of 25 kΩ is The standard deviation of ŷ is estimated to be There are n−2 = 50−2 = 48 degrees of freedom. The t value is therefore t48,0.25 = 2.011. (This value is not found in Table A.3 but can be obtained on many calculators or with computer software. Alternatively, since there are more than 30 degrees of freedom, one could use z = 1.96.) The 95% confidence interval is Page 560 Hypothesis tests on the mean response can be conducted using a Student's t distribution. Following is an example. Example 7.19 Refer to Example 7.18. Let μ0 represent the mean permeability of skin whose resistance is 15 kΩ.Test H0 : μ0 ≤ 9 versus H1 : μ0 > 9. Solution Since μ0 is the mean permeability of skin whose resistance is 15 kΩ, μ0 = β0+β1(15). Now let ŷ = β0 + β1(15). The quantity has a Student's t distribution with n − 2 = 48 degrees of freedom. Under H0,wetake μ0 = 9. The test statistic is therefore We compute ŷ and sŷ : The value of the test statistic is There are n − 2 = 50 − 2 = 48 degrees of freedom. This number of degrees of freedom is not found in Table A.3; however, the P-value can be determined with a calculator or computer software to be 0.0126. Alternatively, since the number of degrees of freedom is greater than 30, one can use the z table (Table A.2) to approximate the P-value as 0.0104. It is reasonable to conclude that the mean permeability is greater than 9 μm/h. Prediction Intervals for Future Observations In Example 7.18 we found a confidence interval for the mean permeability of skin specimens with a resistance of 25 kΩ. Here is a somewhat different question: Assume we wish to predict the permeability of a particular skin specimen whose resistance is 25 kΩ, rather than the mean permeability of all such specimens. Using values calculated in Example 7.18, we predict this specimen's permeability to be . This prediction is the same as the estimate of the mean permeability for all skin specimens with a resistance of 25. Now we wish to put an interval around this prediction to indicate its uncertainty. To compute this prediction interval, we must determine the uncertainty in the prediction. The mean permeability of skin specimens with a resistance of 25 is β0 + β1 (25). The actual permeability of a particular specimen is equal to β0 + β1(25) + ε, where ε represents the random difference between the permeability of the particular specimen and the mean Page 561 permeability of all specimens whose resistance is 25. The error in predicting the permeability of the particular specimen with ŷ is the prediction error (7.42) The uncertainty in the prediction of the strength of the particular weld is the standard deviation of this prediction error. We briefly show how to compute this standard deviation. The quantity β0 + β1 (25) is constant and does not affect the standard deviation. The quantities ŷ and ε are independent, since ŷ is calculated from the data in Example 7.18, while ε applies to a specimen that is not part of that data set. It follows that the standard deviation of the prediction error (expression 7.42) is approximated by Using Equation (7.40) to substitute for sŷ yields (7.43) The appropriate expression for the prediction interval can now be determined. A level 100(1 − α)% prediction interval for the quantity β0 + β1x is given by (7.44) where . Note that the prediction interval is wider than the confidence interval, because the value 1 is added to the quantity under the square root to account for the additional uncertainty. Example 7.20 For the permeability data in Example 7.18, find a 95% prediction interval for the permeability of a particular skin whose resistance is 25 kΩ. Solution The predicted permeability is , which we have calculated in Example 7.18 to be 6.875. Using the quantities presented in Example 7.18, we compute the value of spred to be There are n − 2 = 50 − 2 = 48 degrees of freedom. The t value is therefore t48025 = 2.011. (This value is not found in Table A.3 but can be obtained on many calculators or with computer software. Alternatively, since there are more than 30 degrees of freedom, one could use Page 562 z = 1.96.) The 95% prediction interval is Both the confidence intervals and the prediction intervals described here are specific to a given x value. In the preceding examples, we took x = 25. By computing the intervals for many values of x and connecting the points with a smooth curve, we obtain confidence bands or prediction bands, respectively. Figure 7.13 illustrates 95% confidence bands and prediction bands for the data presented in Example 7.18. For any given resistance, the 95% confidence or prediction bands can be read off the figure. FIGURE 7.13 Permeability versus resistance for 50 skin specimens. In both plots, the dotted line is the least-squares line. Left: The two solid curves are the 95% confidence bands. Given any resistance, we are 95% confident that the mean permeability for skin specimens with that resistance lies between the upper and lower confidence limits. Right: The two solid curves are the 95% prediction bands. Given any specific skin specimen, we are 95% confident that the permeability for that particular skin specimen lies between the upper and lower prediction limits corresponding to the resistance of that skin specimen. Confidence and prediction bands provide a nice visual presentation of the way in which the uncertainty depends on the value of the independent variable. Note that both the confidence interval and the prediction interval are narrowest when , and increases in width as x moves away from . This is due to the term appearing in a numerator in the expressions for sŷ and spred. We conclude that predictions based on the least-squares line are more precise near the center of the scatterplot and are less precise near the edges. Note that the confidence bands indicate confidence intervals for individual values of x. They do not provide a confidence region for the true line y = β0+β1x. In other words, we cannot be 95% confident that the true line lies between the 95% confidence bands. Interpreting Computer Output Nowadays, least-squares calculations are usually done on a computer. The following output (from MINITAB) is for the Hooke's law data. Page 563 We will now explain the labeled quantities in the output: (1) This is the equation of the least-squares line. (2) Coef: The coefficients and . (3) SE Coef: The standard deviations . ("SE" stands for standard error, another term for standard deviation.) and (4) T: The values of the Student's t statistics for testing the hypotheses β0 = 0 and β1 = 0. The t statistic is equal to the coefficient divided by its standard deviation. (5) P: The P-values for the tests of the hypotheses β0 = 0 and β1 = 0. The more important Pvalue is that for β1. If this P-value is not small enough to reject the hypothesis that β1 = 0, the linear model is not useful for predicting y from x. In this example, the P-values are extremely small, indicating that neither β0 nor β1 is equal to 0. Page 564 (6) S: The estimate s of the error standard deviation. (7) R-Sq: This is r2, the square of the correlation coefficient r, also called the coefficient of determination. (8) Analysis of Variance: This table is not so important in simple linear regression, where there is only one independent variable. It is more important in multiple regression, where there are several independent variables. However, it is worth noting that the three numbers in the column labeled "SS" are the regression sum of squares , the error sum of squares , and their sum, the total sum of squares . (9) Unusual Observations: Here MINITAB tries to alert you to data points that may violate some of the assumptions 1 through 4 previously discussed. MINITAB is conservative and will often list several such points even when the data are well described by a linear model. In Section 7.4, we will learn some graphical methods for checking the assumptions of the linear model. (10) Predicted Values for New Observations: These are confidence intervals and prediction intervals for values of x that are specified by the user. Here we specified x = 1.2 for the weight. The "Fit" is the fitted value , and "SE Fit" is the standard deviation sŷ. Then come the 95% confidence and prediction intervals, respectively. (11) Values of Predictors for New Observations: This is simply a list of the x values for which confidence and prediction intervals have been calculated. It shows that these intervals refer to a weight of x = 1.2. Exercises for Section 7.3 1. A chemical reaction is run 12 times, and the temperature xi (in °C) and the yield yi (in percent of a theoretical maximum) is recorded each time. The following summary statistics are recorded: Let β0 represent the hypothetical yield at a temperature of 0°C, and let β1 represent the increase in yield caused by an increase in temperature of 1°C. Assume that assumptions 1 through 4 on page 549 hold. a. Compute the least-squares estimates and . b. c. d. 2. Compute the error variance estimate s2. Find 95% confidence intervals for β0 and β1. A chemical engineer claims that the yield increases by more than 0.5 for each 1°C increase in temperature. Do the data provide sufficient evidence for you to conclude that this claim is false? e. Find a 95% confidence interval for the mean yield at a temperature of 40°C. f. Find a 95% prediction interval for the yield of a particular reaction at a temperature of 40°C. Structural engineers use wireless sensor networks to monitor the condition of dams and bridges. The article "Statistical Analysis of Vibration Modes of a Suspension Bridge Using Spatially Dense Wireless Sensor Network" (S. Pakzad and G. Fenves, Journal of Structural Engineering, 2009:863–872) describes an experiment in which accelerometers were placed on the Golden Gate Bridge for the purpose of estimating vibration modes. The following output (from MINITAB) describes the fit of a linear model that predicts the Page 565 frequency (in Hz) in terms of the damping ratio for overdamped (damping ratio > 1) modes. There are n = 7 observations. The regression equation is Frequency = 0.773 − 0.280 Damping Ratio Predictor Coef SE Coef T P Constant 0.77289 0.14534 5.3176 0.003 Damping −0.27985 0.079258 −3.5309 0.017 Ratio a. b. How many degrees of freedom are there for the Student's t statistics? Find a 98% confidence interval for β1. c. Find a 98% confidence interval for β0. d. 3. Someone claims that the frequency decreases by 0.6 Hz if the damping ratio increases by 1. Use the given output to perform a hypothesis test to determine whether this claim is plausible. Ozone (O3) is a major component of air pollution in many cities. Atmospheric ozone levels are influenced by many factors, including weather. In one study, the mean percent relative humidity (x) and the mean ozone levels (y) were measured for 120 days in a western city. Mean ozone levels were measured in ppb. The following output (from MINITAB) describes the fit of a linear model to these data. Assume that assumptions 1 through 4 on page 549 hold. The regression equation is Ozone = 88.8 − 0.752 Humidity Predictor Coef SE Coef T P Constant 88.761 7.288 12.18 0.000 Humidity −0.7524 0.13024 −5.78 0.000 S = 11.43 R-Sq = 22.0°% R-Sq(adj) = 21.4% Predicted Values for New Observations New Obs Fit SE Fit 95.0% CI 95.0% PI 1 43.62 1.20 (41.23 46.00) (20.86, 66.37) Values of Predictors for New Observations New Obs Humidity 1 60.0 a. b. c. d. e. f. What are the slope and intercept of the least-squares line? Is the linear model useful for predicting ozone levels from relative humidity? Explain. Predict the ozone level for a day when the relative humidity is 50%. What is the correlation between relative humidity and ozone level? The output provides a 95% confidence interval for the mean ozone level for days where the relative humidity is 60%. There are n = 120 observations in this data set. Using the value "SE Fit," find a 90% confidence interval. Upon learning that the relative humidity on a certain day is 60%, someone predicts that the ozone level that day will be 80 ppb. Is this a reasonable prediction? If so, explain why. If not, give a reasonable range of predicted values. Page 566 4. In an study similar to the one in Exercise 3, the relative humidity and ozone levels were measured for 120 days in another city. The MINITAB output follows. Assume that assumptions 1 through 4 on page 549 hold. The regression equation is Ozone = 29.7 − 0.135 Humidity Predictor Coef SE Coef T P Constant 29.703 2.066 14.38 0.000 Humidity −0.13468 0.03798 −3.55 0.001 S = 6.26 R−Sq = 9.6 % R-Sq(adj) = 8.9 % a. b. c. 5. What is the slope of the least-squares line? Find a 95% confidence interval for the slope. Perform a test of the null hypothesis that the slope is greater than or equal to −0.1. What is the P-value? Refer to Exercises 3 and 4. An atmospheric scientist notices that the slope of the leastsquares line in the study described in Exercise 4 differs from the one in the study described in Exercise 3. He wishes to test the hypothesis that the effect of humidity on ozone level differs between the two cities. Let β_A denote the change in ozone level associated with an increase of 1 percent relative humidity for the city in Exercise 3, and β_B denote the corresponding increase for the city in Exercise 4. a. Express the null hypothesis to be tested in terms of β_A and β_B. b. Let $\hat{\beta}_A$ and $\hat{\beta}_B$ denote the slopes of the least-squares lines. Assume these slopes are independent. There are 120 observations in each data set. Test the null hypothesis in part (a). Can you conclude that the effect of humidity differs between the two cities? 6. Cardiologists use the short-range scaling exponent α1, which measures the randomness of heart rate patterns, as a tool to assess risk of heart attack. The article "Applying Fractal Analysis to Short Sets of Heart Rate Variability Data" (M. Peña et al., Med Biol Eng Comput, 2009:709–717) compared values of α1 computed from long series of measurements (approximately 40,000 heartbeats) with those estimated from the first 300 beats to determine how well the long-term measurement (y) could be predicted the short-term one (x). Following are the data (obtained by digitizing a graph). Short 0.54 1.02 1.40 0.88 1.68 1.16 0.82 0.93 1.26 1.18 0.81 1.28 1.18 0.71 1.10 0.81 1.07 1.61 1.66 1.54 1.61 1.49 1.65 1.33 0.98 0.79 1.27 1.16 1.34 1.08 1.14 0.91 1.03 0.82 1.34 Long 0.70 0.79 0.81 0.90 1.05 1.05 1.05 0.57 1.10 0.79 1.19 1.20 1.23 1.23 1.24 1.27 1.29 1.30 1.42 1.42 1.42 1.44 1.44 1.45 1.46 1.47 1.10 1.12 1.13 1.14 1.14 1.15 1.16 1.16 1.18 1.31 1.23 1.30 0.68 1.16 0.93 1.26 1.18 1.36 1.73 1.35 1.57 1.40 1.60 a. b. c. d. e. f. Compute the least-squares line for predicting the long-term measurement from the short-term measurement. Compute the error standard deviation estimate s. Compute a 95% confidence interval for the slope. Find a 95% confidence interval for the mean long-term measurement for those with short-term measurements of 1.2. Can you conclude that the mean long-term measurement for those with short-term measurements of 1.2 is greater than 1.2? Perform a hypothesis test and report the Pvalue. Find a 95% prediction interval for the long-term measurement for a particular individual whose short-term measurement is 1.2. Page 567 g. 7. The purpose of a short-term measurement is to substitute for a long-term measurement. For this purpose, which do you think is more relevant, the confidence interval or the prediction interval? Explain. How much noise are streets where cars travel faster? The article "Modeling of Urban Area Stop-and-Go Traffic Noise" (P. Pamanikabud and C. Tharasawatipipat, Journal of Transportation Engineering, 1999:152–159) reports noise levels in decibels and average speed in kilometers per hour for a sample of roads. The following table presents the results. Speed 28 26 30 32 38 73 Noise 78.1 79.6 81.0 29.07 30.28 30.25 29.03 33.17 78.7 78.6 78.5 79.6 a. b. c. d. e. Compute the least-squares line for predicting noise level (y) from speed (x). Compute the error standard deviation estimate s. Construct a 95% confidence interval for the slope. Find a 95% confidence interval for the mean noise level for streets whose average speed is 30 kilometers per hour. Can you conclude that the mean noise level for streets whose average speed is 30 kilometers per hour is greater than 78 decibels? Perform a hypothesis test and report the P-value. f. Find a 95% prediction interval for the mean noise level for streets whose average speed is 30 kilometers per hour. 8. The article "Application of Radial Basis Function Neural Networks in Optimization of Hard Turning of AISID2 Cold-Worked Tool Steel With a Ceramic Tool" (S. Basak, U. Dixit, and J. Davim, Journal of Engineering Manufacture, 2007:987–998) presents the results of an experiment in which the surface roughness (in μm) was measured for 27 D2 steel specimens and compared with the roughness predicted by a neural network model. The results are presented in the following table. True Value (x) 0.45 0.82 0.54 0.41 0.77 0.79 0.25 0.62 0.91 0.52 1.02 0.60 0.58 0.87 1.06 0.45 1.09 Predicted Value (y) 0.42 0.70 0.52 0.39 0.74 0.78 0.27 0.60 0.87 0.51 0.91 0.71 0.50 0.91 0.74 0.52 0.97 1.35 0.57 1.14 0.74 0.62 1.15 1.27 1.31 1.33 1.46 1.29 0.55 1.01 0.63 1.06 1.31 1.40 1.50 1.16 a. b. c. To check the accuracy of the prediction method, the linear model $y = \beta_0 + \beta_1 x + \varepsilon$ is fit. If the prediction method is accurate, the value of β_0 will be 0 and the value of β_1 will be 1. a. Compute the least-squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. b. Can you reject the null hypothesis H0 : β0 = 0? c. Can you reject the null hypothesis H0 : β1 = 1? d. Do the data provide sufficient evidence to conclude that the prediction method is not accurate? e. Compute a 95% confidence interval for the mean prediction when the true roughness is 0.8 μm. f. Someone claims that when the true roughness is 0.8 μm, the mean prediction is only 0.75; 3? Perform the appropriate hypothesis test. d. 3. Can you conclude that β2 < −0.1? Perform the appropriate hypothesis test. 4. The data used to fit the model in Exercise 1 are presented in the following table, along with the residuals and the fitted values. Plot the residuals versus the fitted values. Does the plot indicate that the linear model is reasonable? Explain. Manganese Thickness Residual Fitted Value 47.7 7.4 8.0 −0.110 47.81 50.9 8.8 10.0 0.091 51.609 50.8 8.1 7.1 −0.517 50.51 ... 49.7 7.2 7.2 0.214 47.486 −0.464 47.564 45.0 7.3 11.8 −0.864 45.864 47.6 7.3 8.0 1.621 47.479 45.7 7.3 11.8 ... 49.560 Page 614 4. The article "Application of Analysis of Variance to Wet Clutch Engagement" (M. Mansouri, M. Khonsari, et al., Proceedings of the Institution of Mechanical Engineers, 2002:117–125) presents the following fitted model for predicting clutch engagement time in seconds (y) from engagement starting speed in m/s (x1), maximum drive torque in N · m (x2), system inertia in kg · m2 (x3), and applied force rate in kN/s (x4): The sum of squares for regression was SSR = 1.08613 and the sum of squares for error was SSE = 0.036310. There were 44 degrees of freedom for error. a. 5. 6. Predict the clutch engagement time when the starting speed is 20 m/s, the maximum drive torque is 17 N · m, the system inertia is 0.006 kg · m2, and the applied force rate is 10 kN/s. b. Is it possible to predict the change in engagement time associated with an increase of 2 m/s in starting speed? If so, find the predicted change. If not, explain why not. c. Is it possible to predict the change in engagement time associated with an increase of 2 N · m in maximum drive torque? If so, find the predicted change. If not, explain why not. d. Compute the coefficient of determination R2. e. Compute the F statistic for testing the null hypothesis that all the coefficients are equal to 0. Can this hypothesis be rejected? In the article "Application of Statistical Design in the Leaching Study of Low-Grade Manganese Ore Using Aqueous Sulfur Dioxide" (P. Naik, L. Sukla, and S. Das, Separation Science and Technology, 2002:1375–1389), a fitted model for predicting the extraction of manganese in % (y) from particle size in mm (x1), the amount of sulfur dioxide in multiples of the stoichiometric quantity needed for the dissolution of manganese (x2), and the duration of leaching in minutes (x3) is given as There were a total of n = 27 observations, with SSE = 209.55 and SST = 6777.5. a. Predict the extraction percent when the particle size is 3 mm, the amount of sulfur dioxide is 1.5, and the duration of leaching is 20 minutes. b. Is it possible to predict the change in extraction percent when the duration of leaching increases by one minute? If so, find the predicted change. If not, explain why not. c. Compute the coefficient of determination R2. d. Compute the F statistic for testing the null hypothesis that all the coefficients are equal to 0. Can this hypothesis be rejected? The article "Earthmoving Productivity Estimation Using Linear Regression Techniques" (S. Smith, Journal of Construction Engineering and Management, 1999:133–141) presents the following linear model to predict earth-moving productivity (in m3 moved per hour): Page 615 where x1 = number of trucks x2 = number of buckets per load x3 = x1x2 x4 = haul length (m) x5 = match factor (ratio of hauling capacity to loading capacity) x6 = truck travel time, in s If the bucket volume increases by 1 m3, while other independent variables are unchanged, can you determine the change in the predicted productivity? If so, determine it. If not, state what other information you would need to determine it. b. If the haul length increases by 1 m, can you determine the change in the predicted productivity? If so, determine it. If not, state what other information you would need to determine it. 7. In a study of the lung function of children, the volume of air exhaled under force in one second is called FEV1. (FEV1 stands for forced expiratory volume in one second.) Measurements were made on a group of children each year for two years. A linear model was fit to predict this year's FEV1 as a function of last year's FEV1 (in liters), the child's gender (0 = Male, 1 = Female), the child's height (in m), and the ambient atmospheric pressure (in mm). The following MINITAB output presents the results of fitting the model FEV1 = β0 + β1 Last FEV1 + β2 Gender + β3 Height + β4 Pressure + ε The regression equation is FEV1 = −0.219 + 0.779 Last FEV − 0.108 Gender + 1.354 Height − 0.00134 Pressure Predictor Coef SE Coef T P Constant −0.21947 0.4503 −0.49 0.627 Last FEV 0.779 0.04909 15.87 0.000 FEV Gender −0.10827 0.0352 −3.08 0.002 Height 1.35356 0.2880 4.70 0.000 Pressure −0.0013431 0.0004722 −2.84 0.005 S= R−Sq = 93.5% R-Sq(adj) = 0.22039 = 93.3% Analysis of Variance Source DF SS MS F P Regression 4111.31 27.826572.890.000 Residual 1607.7716 0.048572 Error Total 164119.08 a. b. Predict the FEV1 for a boy who is 1.4 m tall, if the measurement was taken at a pressure of 730 mm and last year's measurement was 2.113 L. c. If two girls differ in height by 5 cm, by how much would you expect their FEV1 measurements to differ, other things being equal? c. The constant term β0 is estimated to be negative. But FEV1 must always be positive. Is something wrong? Explain. Page 616 8. Refer to Exercise 7. a. Find a 95% confidence interval for the coefficient of Last FEV. b. Find a 98% confidence interval for the coefficient of Height. c. Can you conclude that β2 < −0.08? Perform the appropriate hypothesis test. d. 9. Can you conclude that β3 > 0.5? Perform the appropriate hypothesis test. The article "Drying of Pulps in Sprouted Bed: Effect of Composition on Dryer Performance" (M. Medeiros, S. Rocha, et al., Drying Technology, 2002:865–881) presents measurements of pH, viscosity (in kg/m · s), density (in g/cm3), and BRIX (in percent). The following MINITAB output presents the results of fitting the model a. Predict the pH for a pulp with a viscosity of 1500 kg/m · s, a density of 1.04 g/cm3, and a BRIX of 17.5%. b. If two pulps differ in density by 0.01 g/cm3, by how much would you expect them to differ in pH, other things being equal? The constant term β0 is estimated to be negative. But pulp pH must always be positive. Is something wrong? Explain. Page 617 d. Find a 95% confidence interval for the mean pH of pulps with viscosity 1200 kg/m · s, density 1.08 g/cm3,and BRIX 18.0%. e. Find a 95% prediction interval for the pH of a pulp with viscosity 1000 kg/m · s, density 1.05 g/cm3, and BRIX 19.0%. f. Pulp A has viscosity 2000, density 1.03, and BRIX 20.0. Pulp B has viscosity 1000, density 1.05, and BRIX 19.0. Which pulp will have its pH predicted with greater precision? Explain. 10. A scientist has measured quantities y, x1, and x2. She believes that y is related to x1 and x2 through the equation , where δ is a random error that is always positive. Find a transformation of the data that will enable her to use a linear model to estimate β1 and β2. 11. The following MINITAB output is for a multiple regression. Something went wrong with the printer, so some of the numbers are missing. Fill in the missing numbers. Predictor Coef SE Coef T P Constant (a) 1.4553 5.91 0.000 X1 1.2127 7.8369 (d) 5.1 0.118 3.2109 (c)0.035 0.8943 −3.56 0.005 R= S= Sq(adj) = 71.4% 0.82936 Source Regression Residual Error Total 78.0% DF SS MS F P (e) (f)8.1292 11.8180.001 106.8784 (g) 13 (h) Page 618 13. The article "Evaluating Vent Manifold Inerting Requirements: Flash Point Modeling for Organic Acid-Water Mixtures" (R. Garland and M. Malcolm, Process Safety Progress, 2002:254-260) presents a model to predict the flash point (in °F) of a mixture of water, acetic acid, propionic acid, and butyric acid from the concentrations (in weight %) of the three acids. The results are as follows. The variable "Butyric Acid * Acetic Acid" is the interaction between butyric acid concentration and acetic acid concentration. Predictor Coef SE Coef T P Constant 267.53 11.306 23.660.000 Acetic Acid −1.5926 0.1295 −12.30 0.000 Propionic −1.3897 0.1260 −11.03 0.000 Acid Butyric −1.0934 0.1164 −9.39 0.000 Acid Butyric −0.002658 0.001145 −2.32 0.034 Acid*Acetic Acid a. Predict the flash point for a mixture that is 30% acetic acid, 35% propionic acid, and 30% butyric acid. (Note:In the model, 30% is represented by 30, not by 0.30.) b. Someone asks by how much the predicted flash point will change if the concentration of propionic acid is increased by 10% while the other concentrations are kept constant. Is it possible to answer this question? If so, answer it. If not, explain why not. c. Someone asks by how much the predicted flash point will change if the concentration of acetic acid is increased by 10% while the other concentrations are kept constant. Is it possible to answer this question? If so, answer it. If not, explain why not. 14. In the article "Low-Temperature Heat Capacity and Thermodynamic Properties of 1, 1, 1-trifluoro-2, 2-dichloroethane" (R. Varushchenko and A. Druzhinina, Fluid Phase Equilibria, 2002:109–119), the relationship between vapor pressure (p) and heat capacity (t) is given as , where δ is a random error that is always positive. Express this relationship as a linear model by using an appropriate transformation. 15. The following data were collected in an experiment to study the relationship between extrusion pressure (in KPa) and wear (in mg). x 150 175 200 225 250 275 y 10.4 12.4 14.9 15.0 13.9 11.9 The least-squares quadratic model is y = −32.445714 + 0.43154286x − 0.000982857x2. a. Using this equation, compute the residuals. b. Compute the error sum of squares SSE and the total sum of squares SST. c. Compute the error variance estimate s2. d. Compute the coefficient of determination R2. e. Compute the value of the F statistic for the hypothesis H0 : β1 = β2 = 0. How many degrees of freedom does this statistic have? f. Can the hypothesis H0 : β1 = β2 = 0 be rejected at the 5% level? Explain. Page 619 16. The following data were collected in an experiment to study the relationship between the speed of a cutting tool in m/s (x) and the lifetime of the tool in hours (y). x 1 1.5 2 2.5 3 y 99 96 88 76 66 The least-squares quadratic model is y = 101.4000 + 3.371429x − 5.142857x2. a. Using this equation, compute the residuals. b. Compute the error sum of squares SSE and the total sum of squares SST. c. Compute the error variance estimate s2. d. Compute the coefficient of determination R2. e. Compute the value of the F statistic for the hypothesis H0 : β1 = β2 = 0. How many degrees of freedom does this statistic have? f. Can the hypothesis H0 : β1 = β2 = 0 be rejected at the 5% level? Explain. 17. The November 24, 2001, issue of The Economist published economic data for 15 industrialized nations. Included were the percent changes in gross domestic product (GDP), industrial production (IP), consumer prices (CP), and producer prices (PP) from Fall 2000 to Fall 2001, and the unemployment rate in Fall 2001 (UNEMP). An economist wants to construct a model to predict GDP from the other variables. A fit of the model yields the following output: The regression equation is GDP = 1.19 + 0.17 IP + 0.18 UNEMP + 0.18 CP − 0.18 PP Predictor Coef SE Coef T P Constant 1.18957 0.42180 2.82 0.018 IP 0.17326 0.041962 4.13 0.002 UNEMP 0.17918 0.045895 3.90 0.003 CP 0.17591 0.11365 1.55 0.153 PP −0.18393 0.068808 −2.67 0.023 a. Predict the percent change in GDP for a country with IP = 0.5, UNEMP = 5.7, CP = 3.0, and PP = 4.1. b. If two countries differ in unemployment rate by 1%, by how much would you expect the percent change in GDP to differ, other things being equal? c. CP and PP are both measures of the inflation rate. Which one is more useful in predicting GDP? Explain. d. The producer price index for Sweden in September 2000 was 4.0, and for Austria it was 6.0. Other things being equal, for which country would you expect the percent change in GDP to be larger? Explain. 18. The article "Multiple Linear Regression for Lake Ice and Lake Temperature Characteristics" (S. Gao and H. Stefan, Journal of Cold Regions Engineering, 1999:59–77) presents data on maximum ice thickness in mm (y), average number of days per year of ice cover (x1), number of days the bottom temperature is lower than 8°C(x2), Page 620 and the average snow depth in mm (x3) for 13 lakes in Minnesota. The data are presented in the following table. y 730 760 850 840 720 730 840 730 650 850 740 720 710 a. x1 x2 x3 152 173 166 161 152 153 166 157 136 142 151 145 147 y x1 x2 x3 198 201 202 202 198 205 204 204 172 218 207 209 190 91 81 69 72 91 91 70 90 47 59 88 60 63 Fit the model y = β0 + β1x1 + β2x2 + β3x3 + ε. For each coefficient, find the P-value for testing the null hypothesis that the coefficient is equal to 0. b. If two lakes differ by 2 in the average number of days per year of ice cover, with other variables being equal, by how much would you expect their maximum ice thicknesses to differ? c. Do lakes with greater average snow depth tend to have greater or lesser maximum ice thickness? Explain. 19. In an experiment to estimate the acceleration of an object down an inclined plane, the object is released and its distance in meters (y) from the top of the plane is measured every 0.1 second from time t = 0.1 to t = 1.0. The data are presented in the following table. t 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 y 0.03 0.1 0.27 0.47 0.73 1.07 1.46 1.89 2.39 2.95 The data follow the quadratic model y = β0 + β1t + β2t2 + ε, where β0 represents the initial position of the object, β1 represents the initial velocity of the object, and β2 = a/2, where a is the acceleration of the object, assumed to be constant. In a perfect experiment, both the position and velocity of the object would be zero at time 0. However, due to experimental error, it is possible that the position and velocity at t = 0 are nonzero. a. Fit the quadratic model y = β0 + β1t + β2t2 + ε. b. Find a 95% confidence interval for β2. c. d. Find a 95% confidence interval for the acceleration a. Compute the P-value for each coefficient. Can you conclude that the initial position was not zero? Explain. Can you conclude that the initial velocity was not zero? Explain. Page 621 8.2 Confounding and Collinearity The subtitle of this section is: Fitting separate models to each variable is not the same as fitting the multivariate model. To illustrate what we are talking about, we review the gas well data, first described in Exercise 17 in Section 7.3. A total of 255 gas wells received hydraulic fracturing in order to increase production. In this treatment, a mixture of fluid and sand is pumped into the well. The sand holds open the cracks in the rock, thus increasing the flow of gas. The main questions are these: Does increasing the volume of fluid pumped increase the production of the well? Does increasing the volume of sand increase the production of the well? Other things being equal, deeper wells produce more gas, because they provide more surface through which the gas can permeate. For this reason, it is appropriate to express all variables in units per foot of depth of the well. Thus production is measured in units of ft3 of gas per ft of depth, fluid is measured in units of gal/ft, and sand is measured in units of lb/ft. We showed in Figure 7.22 (in Section 7.4) that a log transformation was needed to obtain homoscedasticity in the plot of production versus fluid. It turns out that a log transform is required for the sand variable as well. Figure 8.4 shows the scatterplots of ln Production versus ln Fluid and ln Production versus ln Sand. Both fluid and sand appear to be strongly related to production. FIGURE 8.4 (a) A plot of the log of production versus the log of volume of fluid for 255 gas wells, with the least-squares line superimposed. (b) A plot of the log of production versus the log of weight of sand for the same 255 wells. There appear to be strong linear relationships between the log of production and both the log of fluid and the log of sand. To confirm the result that is apparent from the plots in Figure 8.4, we fit two Page 622 simple linear regression models: The output (from MINITAB) for these models is as follows: Both ln Fluid and ln Sand have coefficients that are definitely different from 0 (the P-values for both are ≈ 0). We therefore might be tempted to conclude immediately that increasing either the volume of fluid or the volume of sand pumped into a well will increase production. But first we must consider the possibility of confounding. The issue of confounding arises this way. Fluid and sand are pumped in together in a single mixture. It is logical to expect that wells that get more fluid also tend to get more sand. If this is true, then confounding is a possibility. Figure 8.5 presents the scatterplot of ln Fluid versus ln Sand. Sure enough, the amount of sand pumped into a well is highly correlated with the amount of fluid pumped in. It is quite possible, therefore, that either of the two univariate results previously presented may represent a confounding rather than a real relationship. If production depends only on the volume of fluid, there will still be a relationship in the data between production and sand. If production depends only on the volume of sand, there will still be a relationship in the data between production and fluid. Multiple regression provides a way to resolve the issue. The following output (from MINITAB) is for the model (8.11) Page 623 FIGURE 8.5 Scatterplot of ln Fluid versus ln Sand for 255 gas wells. There is clearly a strong linear relationship. Therefore apparent relationships between either fluid or sand and production may represent a confounding rather than a causal relationship. The regression equation is ln Prod = −0.729 + 0.670 ln Fluid + 0.148 ln Sand Predictor Coef SE Coef T P Constant −0.7288 0.6719 −1.08 0.279 ln Fluid 0.6701 0.1687 3.97 0.000 ln Sand 0.1481 0.1714 0.86 0.389 S = 0.38 R−Sq = 28.4% R−Sq(adj) = 0.7463 27.8% We can see that the coefficient of ln Fluid is significantly different from 0, but the coefficient of ln Sand is not. If we assume that there is no other confounding going on (e.g., with the location of the wells), we can conclude that increasing the amount of fluid tends to increase production, but it is not clear that increasing the amount of sand has an effect. Therefore, one might increase the amount of fluid, but it might not be necessary to add more sand to it. A final note: None of the models have a particularly high value of R2. This indicates that there are other important factors affecting production that have not been included in the models. In a more complete analysis, one would attempt to identify and measure some of these factors in order to build a model with greater predictive power. Page 624 Collinearity When two independent variables are very strongly correlated, multiple regression may not be able to determine which is the important one. In this case, the variables are said to be collinear.The word collinear means to lie on the same line, and when two variables are highly correlated, their scatterplot is approximately a straight line. The word multicollinearity is sometimes used as well. When collinearity is present, the set of independent variables is sometimes said to be ill-conditioned. Table 8.2 presents some hypothetical data that illustrate the phenomenon of collinearity. TABLE 8.2 Collinear data x1 X2 0.1 0.2 0.6 1.4 2.0 2.2 2.1 2.8 3.6 4.2 4.5 4.7 5.3 6.1 6.8 7.5 8.2 8.5 9.4 0.3 0.2 1.4 3.4 5.2 5.5 5.3 7.4 9.4 10.3 11.4 11.3 13.6 15.3 17.4 18.5 21.3 23.3 3.6 0.3 6.0 10.6 8.4 11.8 12.7 6.8 9.9 16.7 16.3 19.9 22.2 20.5 21.1 31.0 28.8 32.4 35.0 First we fit the simple linear models Page 625 The output (from MINITAB) shows that both x1 and x2 have a strong linear relationship with y. The values of r2 are both around 0.96, so the correlations r between x1 and y and between x2 and y are both around 0.98. Page 625 Figure 8.6 presents the scatterplot of x2 versus x1. There is clearly a strong linear relationship, so we suspect that y may really have a relationship with only one of these variables, with the other being a confounder. FIGURE 8.6 The independent variables x1 and x2 are collinear, because they have a strong linear relationship. We therefore fit the multiple regression model Page 626 The output (from MINITAB) is as follows: The regression equation is Y = 2.72 − 0.49 X1 + 1.62 X2 Predictor Coef SE Coef T P Constant 2.7248 0.8488 3.21 0.005 X1 −0.490 4.660 −0.11 0.914 X2 1.617 1.791 0.90 0.379 S = 2.091 R−Sq = 96.4% R−Sq(adj) = 96.0% Surprisingly, the output appears to indicate that neither x1 nor x2 is linearly related to y, since both have large P-values. What is happening is that the linear relationship between x1 and x2 is so strong that it is simply impossible to determine which of the two is responsible for the linear relationship with y. Seen in this light, the large P-values make sense. It is possible that the coefficient of x1 is 0 and that only x2 has a real relationship with y. It is equally possible that the coefficient of x2 is 0 and that only x1 has a real relationship with y. Therefore the P-value for x1 must be large. Likewise, it is possible that the coefficient of x2 is 0 and that only x1 has a real relationship with y. Therefore the P-value for x2 must be large also. In general, there is not much that can be done when variables are collinear. The only way to fix the situation is to collect more data, including some values for the independent variables that are not on the same straight line. Then multiple regression will be able to determine which of the variables are really important. Exercises for Section 8.2 1. In an experiment to determine factors related to weld toughness, the Charpy V-notch impact toughness in ft · lb (y) was measured for 22 welds at 0°C, along with the lateral expansion at the notch in % (x1), and the brittle fracture surface in % (x2). The data are presented in the following table. y 32 39 20 21 25 20 32 29 27 43 22 22 x1 20.0 23.0 12.8 16.0 10.2 16.1 20.0 22.6 14.8 23.5 11.1 16.2 16.0 20.4 12.8 17.0 16.9 25.1 ... a. b. Fit the model y = β0 + β1 x1 + ε. For each coefficient, test the null hypothesis that it is equal to 0. Fit the model y = β0 + β1 x2 + ε. For each coefficient, test the null hypothesis that it is equal to 0. c. Fit the model y = β0 + β1 x1 + β2x2 + ε. For each coefficient, test the null hypothesis that it is equal to 0. d. Which of the models in parts (a) through (c) is the best of the three? Why do you think so? 2. In a laboratory test of a new engine design, the emissions rate (in mg/s of oxides of nitrogen, NOx) was measured as a function of engine speed (in rpm), engine torque (in ft · lb), and total horsepower. (From "In-Use Emissions from Heavy-Duty Diesel Page 627 Vehicles," J. Yanowitz, Ph.D. thesis, Colorado School of Mines, 2001.) MINITAB output is presented for the following three models: a. 2. 43 24 36 36 29 30 31 36 34 30 The regression equation is NOx = − 321 + 0.378 Speed − 0.160 Torque Predictor Coef SE Coef T P Constant −320.59 98.14 −3.27 0.003 Speed 0.378200.06861 5.51 0.000 Torque −0.16047 0.06082 −2.64 0.013 S = R−Sq = 51.6% R−Sq(adj) = 67.13 48.3% The regression equation is NOx = −380 + 0.416 Speed − 0.520 HP Predictor Coef SE Coef T P Constant −380.1 104.8 −3.63 0.001 Speed 0.41641 0.07510 5.54 0.000 HP −0.5198 0.1980 −2.63 0.014 S = R−Sq = 51.5% R−Sq(adj) = 67.19 48.2% The regression equation is NOx = −302 + 0.366 Speed − 0.211 Torque + 0.16 HP Predictor Coef SE Coef T P Constant −301.8 347.3 −0.87 0.392 Speed 0.3660 0.2257 1.62 0.116 Torque −0.2106 0.8884 −0.24 0.814 HP 0.164 2.889 0.06 0.955 S = R−Sq = 51.6% R−Sq(adj) = 68.31 46.4% 3. Of the variables Speed, Torque, and HP, which two are most nearly collinear? How can you tell? Two chemical engineers, A and B, are working independently to develop a model to predict the viscosity of a product (y) from the pH (x1) and the concentration of a certain catalyst (x2). Each engineer has sent you output summarizing the analysis. Engineer A Predictor Coef SE Coef T P Constant 199.2 0.5047 394.7 0.000 pH −1.569 0.4558 −3.44 0.007 Concent. −4.730 0.5857 −8.08 0.000 Engineer B Predictor Coef SE Coef T P Constant 199.0 0.548 363.1 0.000 pH −1.256 Concent. −3.636 Page 628 SE Coef 0.548 1.983 1.952 T 363.1 −0.63 −1.86 P 0.000 0.544 0.112 The engineers have also sent you the following scatterplots of pH versus concentration, but forgot to put their names on them. 4. a. Which plot came from which engineer? How do you know? b. Which engineer's experiment produced the more reliable results? Explain. The article "Influence of Freezing Temperature on Hydraulic Conductivity of Silty Clay" (J. Konrad and M. Samson, Journal of Geotechnical and Geoenvironmental Engineering, 2000:180–187) describes a study of factors affecting hydraulic conductivity of soils. The measurements of hydraulic conductivity in units of 10−8 cm/s (y), initial void ratio (x1), and thawed void ratio (x2) for 12 specimens of silty clay are presented in the following table. a. b. c. d. y 1.01 1.12 1.04 1.30 1.01 1.04 0.955 1.15 1.23 1.28 1.23 1.30 x1 1.04 0.88 0.85 0.88 0.85 0.89 0.90 0.88 0.90 x2 0.81 0.85 0.88 0.79 0.84 0.85 0.88 0.86 0.85 0.89 0.90 0.88 0.90 0.92 Fit the model y = β0 + β1x1 + ε. For each coefficient, test the null hypothesis that it is equal to 0. Fit the model y = β0 + β1x2 + ε. For each coefficient, test the null hypothesis that it is equal to 0. Fit the model y = β0 + β1x1 + β2x2 + ε. For each coefficient, test the null hypothesis that it is equal to 0. Which of the models in parts (a) to (c) is the best of the three? Why do you think so? Page 629 5. Refer to Exercise 12 in Section 7.4. a. Divide the data into two groups: points where R1 < 4 in one group, points where R1 ≥ 4 in the other. Compute the least-squares line for predicting R2 from R1 for each group. (You already did this if you did Exercise 12c in Section 7.4.) b. For one of the two groups, the relationship is clearly nonlinear. For this group, fit a c. d. quadratic model (i.e., using R1 and as independent variables), a cubic model, and a quartic model. Compute the P-values for each of the coefficients in each of the three models in part (b). Compute the correlation coefficient between and , and make a scatterplot of the points . On the basis of the correlation coefficient and the scatterplot, explain why the P-values are much different for the quartic model than for the cubic model. 6. f. Which of the three models in part (b) is most appropriate? Why? The following table lists values for three variables measured for 60 consecutive eruptions of the geyser Old Faithful in Yellowstone National Park. They are the duration of the eruption (x1), the duration of the dormant period immediately before the eruption (x2), and the duration of the dormant period immediately after the eruption (y). All the times are in minutes. (Variable x2 was presented in Supplementary Exercise 3 in Chapter 7.) e. 6. x1 x2 3.5 4.1 2.3 4.7 1.7 4.9 1.7 4.6 3.4 4.3 1.7 3.9 3.7 3.1 4.0 1.8 4.1 1.8 3.2 1.9 4.6 2.0 4.5 3.9 4.3 80 84 50 93 55 76 58 74 75 80 56 80 69 57 90 42 91 51 79 53 82 51 76 82 84 53 86 51 85 45 88 51 80 49 82 75 73 67 68 86 72 75 75 66 84 70 79 60 86 71 67 81 76 83 76 55 73 56 83 57 y 80 84 50 93 55 76 58 74 75 80 56 80 69 57 90 42 91 51 79 53 82 51 76 82 84 53 86 51 85 45 88 51 80 49 82 75 73 67 68 86 72 75 75 66 84 70 79 60 86 71 67 81 76 83 76 55 73 56 83 57 ... a. Compute the least-squares line for predicting the duration of the dormant period following an eruption (y) from the duration of the eruption (x1). b. Is there a linear relationship between the duration of an eruption and the waiting time until the next eruption occurs? Page 630 8.3 Model Selection There are many situations in which a large number of independent variables have been measured, and we need to decide which of them to include in a model. This is the problem of model selection, and it is a difficult one. In practice, model selection often proceeds by ad hoc methods, guided by whatever physical intuition may be available. We will not attempt a complete discussion of this extensive and difficult topic. Instead, we will be content to state some basic principles and to present some examples. An advanced reference such as Miller (2002) can be consulted for information on specific methods. Good model selection rests on a basic principle known as Occam's razor. This principle is stated as follows: Occam's Razor The best scientific model is the simplest model that explains the observed facts. In terms of linear models, Occam's razor implies the principle of parsimony: The Principle of Parsimony A model should contain the smallest number of variables necessary to fit the data. There are some exceptions to the principle of parsimony: 1. 2. 3. A linear model should always contain an intercept, unless physical theory dictates otherwise. If a power xn of a variable is included in a model, all lower powers x, x2,..., xn−1 should be included as well, unless physical theory dictates otherwise. If a product xixj of two variables is included in a model, then the variables xi and xj should be included separately as well, unless physical theory dictates otherwise. Models that contain only the variables that are needed to fit the data are called parsimonious models. Much of the practical work of multiple regression involves the development of parsimonious models. We illustrate the principle of parsimony with the following example. The data in Table 8.3 were taken from the article "Capacities and Performance Page 631 Crushers" (S. Sastri, Minerals and Metallurgical Processing, 1994:80–86). Feed rates and amounts of power drawn were measured for several industrial jaw crushers. TABLE 8.3 Feed rates and power for industrial jaw crushers Feed Rate (100 tons/h) 0.10 1.55 3.00 3.64 0.38 1.59 4.73 0.20 2.91 0.36 0.14 0.91 4.27 3.36 0.20 0.97 0.59 0.27 0.55 0.68 0.27 3.64 1.36 2.36 1.09 0.91 2.55 Power (kW) 11 60 40 150 69 77 83 15 84 30 16 30 150 60 14 47 37 13 24 49 15 150 90 45 55 100 ... The following output presents the results for fitting the model (8.12) The regression equation is Power = 21.0 + 24.6 FeedRate Predictor Coef SE Coef T P Constant 21.0288.038 2.62 0.015 FeedRate 24.5953.338 7.37 0.000 S= R−Sq = 68.5% R−Sq(adj) = 26.20 67.2%% From the output, we see that the fitted model is (8.13) and that the coefficient for FeedRate is significantly different from 0 (t = 7.37, P ≈ 0). We wonder whether a quadratic model might fit better than this linear one. We fit (8.14) Page 632 The results of the fit are as follows. The regression equation is Power = 19.3 + 27.5 FeedRate − 0.64 FeedRate^2 Predictor Coef SE Coef T P Constant 19.3411.56 1.67 0.107 FeedRate 27.4714.31 1.92 0.067 FeedRate^2 −0.6387 3.090 −0.21 0.838 S = 26.72 R−Sq = 68.5% R−Sq(adj) = 65.9% The most important point to notice is that the P-value for the coefficient of FeedRate2 is large (0.838). Recall that this P-value is for the test of the null hypothesis that the coefficient is equal to 0. Thus the data provide no evidence that the coefficient of FeedRate2 is different from 0. Note also that including FeedRate2 in the model increases the value of the goodness-of-fit statistic R2 only slightly, in fact so slightly that the first three decimal places are unchanged. It follows that there is no evidence that the quadratic model fits the data better than the linear model, so by the principle of parsimony, we should prefer the linear model. Figure 8.7 provides a graphical illustration of the principle of parsimony. The scatterplot of power versus feed rate is presented, and both the least-squares line (8.13) and the fitted quadratic model (8.14) are superimposed. Even though the coefficients of the models are different, we can see that the two curves are almost identical. There is no reason to include the quadratic term in the model. It makes the model more complicated, without improving the fit. FIGURE 8.7 Scatterplot of power versus feed rate for 27 industrial jaw crushers. The least-squares line and best fitting quadratic model are both superimposed. The two curves are practically identical, which reflects the fact that the coefficient of FeedRate2 in the quadratic model does not differ significantly from 0. Page 633 Determining Whether Variables Can Be Dropped from a Model It often happens that one has formed a model that contains a large number of independent variables, and one wishes to determine whether a given subset of them may be dropped from the model without significantly reducing the accuracy of the model. To be specific, assume that we know that the model (8.15) is correct, in that it represents the true relationship between the x variables and y. We will call this model the "full" model. We wish to test the null hypothesis If H0 is true, the model will remain correct if we drop the variables xk+1,..., xp, so we can replace the full model with the following reduced model: (8.16) To develop a test statistic for H0, we begin by computing the error sum of squares for both the full and the reduced models. We'll call them SSEfull and SSEreduced. The number of degrees of freedom for SSEfull is n − p − 1, and the number of degrees of freedom for SSEreduced is n − k − 1. Now since the full model is correct, we know that the quantity SSEfull/(n − p − 1) is an estimate of the error variance σ2; in fact it is just s2. If H0 is true, then the reduced model is also correct, so the quantity SSEreduced/(n − k − 1) is also an estimate of the error variance. Intuitively, SSEfull is close to (n − p − 1)σ2, and if H0 is true, SSEreduced is close to (n − k − 1)σ2. It follows that if H0 is true, the difference (SSEreduced − SSEfull) is close to (p − k)σ2, so the quotient (SSEreduced − SSEfull)/(p − k) is close to σ2. The test statistic is (8.17) Now if H0 is true, both numerator and denominator of f are estimates of σ2, so f is likely to be near 1. If H0 is false, the quantity SSEreduced tends to be larger, so the value of f tends to be larger. The statistic f is an F statistic; its null distribution is Fp−k, n−p−1. The method we have just described is very useful in practice for developing parsimonious models by removing unnecessary variables. However, the conditions under which it is formally valid are seldom met in practice. First, it is rarely the case that the full model is correct; there will be nonrandom quantities that affect the value of the dependent variable y that are not accounted for by the independent variables. Second, for the method to be formally valid, the Page 634 subset of variables to be dropped must be determined independently of the data. This is usually not the case. More often, a large model is fit, some of the variables are seen to have fairly large P-values, and the F test is used to decide whether to drop them from the model. As we have said, this is a useful technique in practice, but, like most methods of model selection, it should be seen as an informal tool rather than a rigorous theory-based procedure. We illustrate the method with an example. In mobile ad hoc computer networks, messages must be forwarded from computer to computer until they reach their destinations. The data overhead is the number of bytes of information that must be transmitted along with the messages to get them to the right places. A successful protocol will generally have a low data overhead. The overhead is affected by several features of the network, including the speed with which the computers are moving, the length of time they pause at each destination, and the link change rate. The link change rate for a given computer is the rate at which other computers in the network enter and leave the transmission range of the given computer. Table 8.4 presents average speed, pause time, link change rate (LCR), and data overhead for 25 simulated computer networks. These data were generated for a study published in the article "Metrics to Enable Adaptive Protocols for Mobile Ad Hoc Networks" (J. Boleng, W. Navidi, and T. Camp, Proceedings of the 2002 International Conference on Wireless Networks, 2002:293–298). TABLE 8.4 Data overhead, speed, pause time, and link change rate for a mobile computer network Speed (m/s) 5 5 5 5 5 10 10 10 10 10 20 20 20 20 20 30 30 30 30 30 40 40 40 40 40 Pause Time (s) 10 20 30 40 50 10 20 30 40 50 10 20 30 40 50 10 20 30 40 50 10 20 30 40 50 LCR (100/s) 9.426 8.318 7.366 6.744 5.933 12.587 10.964 9.426 8.318 7.366 ... Data Overhead (kB) 428.90 443.68 452.38 461.24 475.07 446.06 465.89 477.07 488.73 498.77 ... To begin, we fit the following model, which includes the data, namely, Page 635 This model is fit to the data. Speed, Pause, Speed*Pause as independent variables, and Overhead as dependent variable... The regression equation is Overhead = 368 + 3.48 LCR + 3.04 Speed + 2.29 Pause − 0.0122 Speed*Pause − 0.1041 LCR^2 − 0.0132 Pause^2 Predictor Coef SE Coef T P Constant 367.96 19.40 18.96 0.000 LCR 3.477 2.129 1.63 0.121 Speed 3.044 1.591 1.91 0.073 Pause 2.2924 0.6984 3.28 0.004 Speed*Pa −0.01222 0.01534 −0.80 0.437 LCR^2 −0.10412 0.03192 −3.26 0.005 Speed^2 −0.03131 −0.010 0.01318 0.01045 −1.26 0.224 S= R−Sq = 97.2% R−Sq(adj) = 5.72344 96.1% Analysis of Variance Source DF SS MS F P Regression 7 19567.5 2795.4 85.33 0.000 Residual Error 17 556.9 32.8 Total 24 20124.3 The P-values of Speed*Pause, Speed2, and Pause2 are fairly large. We will leave LCR in the model for now, because LCR2 has a very small P-value, and therefore should stay in the model. We will use the F test to determine whether the reduced model obtained by dropping Speed · Pause, Speed2, and Pause2 is a reasonable one. First, from the output for the full model, we note that SSEfull = 556.9, and it has 17 degrees of freedom. The number of independent variables in the full model is p = 7. We now fit the reduced model (8.18) The output from fitting this model is as follows: The regression equation is Overhead = 359 + 6.69 LCR + 0.777 Speed + 1.67 Pause − 0.156 LCR^2 Predictor Coef SE Coef T P Constant 359.22 13.01 27.61 0.000 LCR 6.695 1.156 5.79 0.000 Speed 0.7766 0.2054 3.78 0.001 Pause 1.6729 0.1826 9.16 0.000 LCR^2 −0.15567 0.02142 −7.27 0.000 S= 6.44304 −1.5572 0.02148 R−Sq = 95.9% R−Sq(adj) = 95.1% Analysis of Variance Source DF SS MS F P Regression 419294.1 4823.5 116.2 0.000 Residual Error 20 830.3 41.5 Total 2420124.3 The P-values for the variables in this model are all quite small. Page 636 From the output for this reduced model, we note that SSEreduced = 830.3. The number of variables in this reduced model is k = 4. Now we can compute the F statistic. Using Equation (8.17), we compute The null distribution is F3, 17,Fromthe F table (Table A.8 in Appendix A), we find that 0.05 < P < 0.10. According to the 5% rule of thumb, since P > 0.05, the reduced model is plausible, but only barely so. Rather than settle for a barely plausible model, it is wise to explore further, to look for a slightly less reduced model that has a larger P-value. To do this, we note that of the three variables we dropped, the variable Speed2 had the smallest P-value in the full model. This is an indication that this might be the most important of the variables we dropped, so we will put it back in the model. We thus fit a second reduced model, which is as follows: The results from fitting this model are as follows. The regression equation is Overhead = 373 + 4.80 LCR + 1.99 Speed + 1.45 Pause − 0.123 LCR^2 − 0.0212 Speed^2 Predictor Coef SE Coef T P Constant 372.60 16.93 22.00 0.000 LCR 4.799 1.935 2.48 0.023 Speed 1.993 1.023 1.95 0.066 Pause 1.4479 0.2587 5.60 0.000 LCR^2 −0.12345 0.03040 −3.63 0.002 Speed^2 −0.02120 0.01746 −1.21 0.240 S= 6.36809 R−Sq = 96.2% R−Sq(adj) = 95.2% Analysis of Variance Source DF SS MS F P Regression 5 19353.8 3870.8 95.45 0.000 Residual Error 19 770.5 40.6 Total 24 20124.3 Note that the P-value for Speed2 in this model is large (0.240). This is not good. In general, we do not want to add a variable whose coefficient might be equal to 0. So we probably won't want to stick with this model. Let's compute the value of the F statistic anyway, just for practice. The value of SSEreduced in this model is 770.5. The number of variables is k = 5. The value of the F statistic, using Equation (8.17), is Page 637 The null distribution is F2, 17,Fromthe F table (Table A.8), we again find that 0.05 < P < 0.10 (software yields P = 0.0633), so the reduced model is barely plausible at best. We chose to put Speed2 back into the model because it had the smallest P-value among the variables we originally dropped. But as we have seen, this does not guarantee that it will have a small P-value when it is put back into the reduced model. Perhaps one of the other variables we dropped will do better. Of the three variables originally dropped, the one with the second smallest P value was Pause2. We try replacing Speed2 in the preceding model with Pause2. So we now fit a third reduced model. The results from fitting this model are as follows. The regression equation is Overhead = 345 + 6.484 LCR + 0.707 Speed + 2.85 Pause − 0.145 LCR^2 − 0.0183 Pause^2 Predictor Coef SE Coef T P Constant 345.42 13.19 26.20 0.000 LCR 6.484 1.050 6.17 0.000 Speed 0.7072 0.1883 3.76 0.001 Pause 2.8537 0.5337 5.35 0.000 LCR^2 −0.14482 0.01996 −7.25 0.000 Pause^2 −0.018334 0.007879 −2.33 0.031 S= R−Sq = 96.8% R − Sq(adj) = 5.83154 95.9% Analysis of Variance Source DF SS MS F P Regression 5 19478.2 3895.6 114.55 0.000 Residual Error 19 646.1 34.0 Total 24 20124.3 This model looks good, at least at first. All the variables have small P-values. We'll compute the F statistic to see if this model is plausible. The value of SSEreduced in this model is 646.1. The number of independent variables is k = 5. The value of the F statistic, using Equation (8.17), is The null distribution is F2, 17,Fromthe F table (Table A.8), we find that the 0.10 point on this F distribution is 2.64. The P-value is therefore much larger than 0.10. (Software yields P = 0.283.) Thus this model is clearly plausible. We have obtained a good parsimonious model. It is important to realize that the above procedure could have been carried out somewhat differently, with different choices for variables to drop or to add in to the model. We might have come up with a different final model that fit the data just as well. In general, there are many models that fit a given data set about equally well, there is no single "correct" model. There are a few tools, however, that can make the process somewhat more systematic. One of them is best subsets regression. Page 638 Best Subsets Regression mentioned, methods of model selection are often rather informal and ad hoc. As we have There are a few tools, however, that can make the process somewhat more systematic. One of them is best subsets regression. The concept is quite simple. Assume that there are p independent variables, x1, ..., xp, that are available to be put into the model. Let's assume that we wish to find a good model that contains exactly four independent variables. We can simply fit every possible model containing four of the variables, and rank them in order of their goodness-of-fit, as measured by the coefficient of determination R2. The subset of four variables that yields the largest value of R2 is the "best" subset of size 4. One can repeat the process for subsets of other sizes, finding the best subsets of size 1, 2, ..., p. These best subsets can then be examined to see which provide a good fit, while being parsimonious. The best subsets procedure is computationally intensive. When there are a lot of potential independent variables, there are a lot of models to fit. However, for most data sets, computers today are powerful enough to handle 30 or more independent variables, which is enough to cover many situations in practice. The following MINITAB output is from the best subsets procedure, applied to the data in Table 8.4. There are a total of seven independent variables being considered: Speed, Pause, LCR, Speed · Pause, Speed2, Pause2, and LCR2. In this output, both the best and the second-best subsets are presented, for sizes 1 through 7. We emphasize that the term best means only that the model has the largest value of R2, Page 639 and does not guarantee that it is best in any practical sense. We'll explain the output column by column. The first column, labeled "Vars," presents the number of variables in the model. Thus the first row of the table describes the best model that can be made with one independent variable, and the second row describes the second-best such model. The third and fourth rows describe the best and second-best models that can be made with two variables, and so on. The second column presents the coefficient of determination, R2, for each model. Note that the value of R2 for the best subset increases as the number of variables increases. It is a mathematical fact that the best subset of k + 1 variables will always have at least as large an R2 as the best subset of k variables. We will skip over the next two columns for the moment. The column labeled "s" presents the estimate of the error standard deviation. It is the square root of the estimate s2 (Equation 8.8 in Section 8.1). Finally, the columns on the right represent the independent variables that are candidates for inclusion into the model. The name of each variable is written vertically above its column. An "X" in the column means that the variable is included in the model. Thus, the best model containing four variables is the one with the variables Speed, Pause, LCR, and LCR2. Looking at the best subsets output, it is important to note how little difference there is in the fit between the best and second-best models of each size (except for size 1). It is also important to realize that the value of R2 is a random quantity; it depends on the data. If the process were repeated and new data obtained, the values of R2 for the various models would be somewhat different, and different models would be "best." For this reason, one should not use this procedure, or any other, to choose a single model. Instead, one should realize that there will be many models that fit the data about equally well. Nevertheless, methods have been developed to choose a single model, presumably the "best" of the "best." We describe two of them here, with a caution not to take them too seriously. We begin by noting that if we simply choose the model with the highest value of R2, we will always pick the one that contains all the variables, since the value of R2 necessarily increases as the number of variables in the model increases. The methods for selecting a single model involve statistics that adjust the value of R2, so as to eliminate this feature. The first is the adjusted R2. Let n denote the number of observations, and let k denote the number of independent variables in the model. The adjusted R2 is defined as follows: (8.18) The adjusted R2 is always smaller than R2, since a positive quantity is subtracted from R2. As the number of variables k increases, R2 will increase, but the amount subtracted from it will increase as well. The value of k for which the value of adjusted R2 is a maximum can be used to determine the number of variables in the model, and the best subset of that size can be chosen as the model. In the preceding output, we can see that the adjusted R2 reaches its maximum (96.2%) at the six-variable model containing the variables Speed, Pause, LCR, Speed2, Pause2, and LCR2. Another commonly used statistic is Mallows' Cp. To compute this quantity, let n be the number of observations, let p be the total number of independent variables under Page 640 consideration, and let k be the number of independent variables in a subset. As before, let SSEfull denote the error sum of squares for the full model containing all p variables, and let SSEreduced denote the error sum of squares for the model containing only the subset of k variables. Mallows' Cp is defined as (8.19) Models that contain all the relevant variables will have a value of Cp close to k + 1. Models with Cp much greater than k + 1 are missing some important variables. The model chosen is the model with the minimum value of Cp. In the preceding output, the minimum value of Cp is 5.4 for the six-variable model, both criteria yield the same six-variable model chosen by the adjusted R2 criterion. Its value of Cp is 6.6. Finally, we point out that our ad hoc procedure using the F test yielded the five-variable model containing the variables Speed, Pause, LCR, Pause2, and LCR2.The output shows that this model is the best five-variable model in terms of R2. Its adjusted R2 is 95.9%, and its Cp value is 6.7, both of which are close to their optimum values. In practice, there is no clear reason to prefer the model chosen by adjusted R2 and Mallows' Cp to this one, or vice versa. Stepwise Regression Stepwise regression is perhaps the most widely used model selection technique. Its main advantage over best subsets regression is that it is less computationally intensive, so it can be used in situations where there are a very large number of candidate independent variables and too many possible subsets for every one of them to be examined. The version of stepwise regression that we will describe is based on the P-values of the t statistics for the independent variables. An equivalent version is based on the F statistic (which is the square of the t statistic). Before running the algorithm, the user chooses two threshold P-values, αin and αout,with αin ≤ αout. Stepwise regression begins with a step called a forward selection step, in which the independent variable with the smallest P-value is selected, provided that it satisfies P < αin. This variable is entered into the model, creating a model with a single independent variable. Call this variable x1. In the next step, also a forward selection step, the remaining variables are examined one at a time as candidates for the second variable in the model. The one with the smallest Pvalue is added to the model, again provided that P < αin. Now it is possible that adding the second variable to the model has increased the P-value of the first variable. In the next step, called a backward elimination step, the first variable is dropped from the model if its P-value has grown to exceed the value αout. The algorithm then continues by alternating forward selection steps with backward elimination steps: at each forward selection step adding the variable with the smallest P-value if P < αin, and at each backward elimination step dropping the variable with the largest P-value if P > αout. The algorithm terminates when no variables meet the criteria for being added to or dropped from the model. Page 641 The following output is from the MINITAB stepwise regression procedure, applied to the data in Table 8.4. The threshold P-values are αin = αout = 0.15. There are a total of seven independent variables being considered: Speed, Pause, LCR, Speed · Pause, Speed2, Pause2, and LCR2. Alpha-to-Enter = 0.15 Alpha-to-Remove = 0.15 Response is Overhead on 7 predictors, with N = 25 Step 1 2 3 4 5 Constant 452.2 437.3 410.7 388.4 338.5 Speed*Pause 0.0470 0.0355 0.0355 0.0304 0.0146 T-Value 8.03 6.00 6.96 5.21 3.52 P-Value 0.000 0.000 0.000 0.000 0.002 Pause 0.74 3.02 3.75 3.24 T-Value 3.40 3.78 4.20 6.25 P-Value 0.003 0.001 0.000 0.000 Pause^2 −0.0320 −0.0442 −0.0256 T-Value −2.94 −3.39 −3.19 P-Value 0.008 0.003 0.005 LCR 0.69 6.97 T-Value 1.62 6.95 P-Value 0.121 0.000 LCR^2 −0.139 T-Value −6.46 P-Value 0.000 S 15.2 11.8 10.4 9.99 R-Sq 73.70 82.74 87.77 89.19 96.62 R-Sq(adj) 72.55 81.18 86.02 87.02 95.73 Mallows C-p 140.6 87.0 58.1 51.4 7.8 In step 1, the variable Speed · Pause had the smallest P-value (0.000) among the seven, so it was the first variable in the model. In step 2, Pause had the smallest P-value (0.003) among the remaining variables, so it was added next. The P-value for Speed · Pause remained at 0.000 after the addition of Pause to the model; since it did not rise to a value greater than αout = 0.15, it is not dropped. In steps 3, 4, and 5, Pause2, LCR, and LCR2 are added in turn. At no point does the P-value of a variable in the model exceed the threshold αout = 0.15, so no variables are dropped. After five steps, none of the variables remaining have P-values less than αin = 0.15, so the algorithm terminates. The final model contains the variables Speed · Pause, Pause, Pause2, LCR, and LCR2. Page 642 The model chosen by stepwise regression is a five-variable model. Comparison with the best subsets output shows that it is not one of the best two five-variable models in terms of R2. Still, the model fits well, and in terms of fit alone, it is reasonable model. We point out that this model has the undesirable feature that it contains the interaction term Speed · Pause without containing the main effect Speed. This points out a weakness of all automatic variable selection procedures, including stepwise regression and best subsets. They operate on the basis of goodness-of-fit alone, and are not able to take into account relationships among independent variables that may be important to consider. Model Selection Procedures Sometimes Find Models When They Shouldn't When constructing a model to predict the value of a dependent variable, it might seem reasonable to try to start with as many candidate independent variables as possible, so that a model selection procedure has a very large number of models to choose from. Unfortunately, this is not a good idea, as we will now demonstrate. A correlation coefficient can be computed between any two variables. Sometimes, two variables that have no real relationship will be strongly correlated, just by chance. For example, the statistician George Udny Yule noticed that the annual birthrate in Great Britain was almost perfectly correlated (r = −0.98) with the annual production of pig iron in the United States for the years 1875–1920. Yet no one would suggest trying to predict one of these variables from the other. This illustrates a difficulty shared by all model selection procedures. The more candidate independent variables that are provided, the more likely it becomes that some of them will exhibit meaningless correlations with the dependent variable, just by chance. We illustrate this phenomenon with a simulation. We generated a simple random sample y1, ..., y30 of size 30 from a N(0, 1) distribution; we will denote this sample by y. Then we generated 20 more independent samples of size 30 from a N(0, 1) distribution; we will denote these samples by x1, ..., x20. To make the notation clear, the sample xi contains 30 values xi1, ..., xi30. We then applied both stepwise regression and best subsets regression to these simulated data. None of the xi are related to y; they were all generated independently. Therefore the ideal output from a model selection procedure would be to produce a model with no dependent variables at all. The actual behavior was quite different. The following two MINITAB outputs are for the stepwise regression and best subsets procedures. The stepwise regression method recommends a model containing six variables, with an adjusted R2 of 41.89%. The best subsets procedure produces the best-fitting model for each number of variables from 1 to 20. Using the adjusted R2 criterion, the best subsets procedure recommends a 12-variable model, with an adjusted R2 of 51.0%. Using the minimum Mallows' Cp criterion, the "best" model is a five-variable model. Anyone taking this output at face value would believe that some of the independent variables might be useful in predicting the dependent variable. But none of them are. All the apparent relationships are due entirely to chance. Page 643 Stepwise Regression: Y versus X1, X2, ... Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15 Response is Y on 20 predictors, with N = 30 Step 1 2 3 4 5 6 Constant 0.14173 0.11689 0.12016 0.13756 0.00336 0.00589 X15 −0.38 −0.38 −0.28 −0.32 −0.28 −0.30 T-Value −2.08 −2.19 −1.60 −1.87 −1.69 −1.89 P-Value 0.047 0.037 0.122 0.073 0.105 0.071 X6 0.39 0.55 0.57 0.52 0.52 T-Value 2.04 2.76 2.99 3.15 2.87 P-Value 0.051 0.010 0.006 0.004 0.009 X16 −0.43 −0.43 −0.45 −0.40 T-Value −2.02 −2.08 −2.60 −2.05 P-Value 0.054 0.048 0.016 0.051 X12 0.33 0.42 0.40 ... S 1.01 1.04 0.998 0.954 0.928 0.889 R-Sq 13.33 24.92 34.75 42.15 49.23 53.91 R-Sq(adj) 10.24 19.36 27.22 32.90 38.66 41.89 Mallows C-p 6.5 5.3 1.7 1.0 0.4 0.7 C-p Best Subsets Regression: Y versus X1, X2, ... Response is Y Page 644 X X X X X X X X X X X 1 1 1 1 1 1 1 1 1 1 2 Vars R-Sq R-Sq(adj) Mallows Cp S 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 ... 1 15.2 12.2 5.8 1.0091 X ... 2 27.1 21.6 2.9 0.95669 X X ... 3 37.0 29.7 1.3 0.90616 X X X ... 20 54.2 0.0 21.0 1.3200 X X X X X X X X X X X X X X X X X X X X ... In the case of a data set such as this, in which none of the independent variables are related to the dependent variable, model selection procedures may still find models that seem useful. Dividing the data into two pieces, then using one (the training set) to fit the model and the other (the test set) to evaluate it, can be helpful. A method known as crossvalidation employs a more advanced version of this idea. A procedure called the lasso, which is one of a class of methods known as shrinkage methods, has been shown to be useful in some situations. Finally, if it is possible to repeat the experiment to collect more data, the model constructed from the original data can be fit to the new data. If it fits well, the evidence of a real relationship becomes more convincing. More information on these methods can be found in James, Witten, et al., (2013). We summarize our discussion of model selection by emphasizing four points. Page 645 Summary When selecting a regression model, keep the following in mind: ■ When there is little or no physical theory to rely on, many different models will fit the data about equally well. ■ The methods for choosing a model involve statistics (R2,the F statistic, Cp), whose values depend on the data. Therefore if the experiment is repeated, these statistics will come out differently, and different models may appear to be "best." ■ Some or all of the independent variables in a selected model may not really be related to the dependent variable. Whenever possible, experiments should be repeated to test these apparent relationships. ■ Model selection is an art, not a science. Exercises for Section 8.3 1. True or false: a. b. c. For any set of data, there is always one best model. When there is no physical theory to specify a model, there is usually no best model, but many that are about equally good. Model selection methods such as best subsets and stepwise regression, when properly used, are scientifically designed to find the best available model. d. Model selection methods such as best subsets and stepwise regression, when properly used, can suggest models that fit the data well. The article "Experimental Design Approach for the Optimization of the Separation of Enantiomers in Preparative Liquid Chromatography" (S. Lai and Z. Lin, Separation Science and Technology, 2002: 847–875) describes an experiment involving a chemical process designed to separate enantiomers. A model was fit to estimate the cycle time (y) in terms of the flow rate (x1), sample concentration (x2), and mobile-phase composition (x3). The results of a least-squares fit are presented in the following table. (The article did not provide the value of the t statistic for the constant term.) Predictor Constant x1 x2 x3 Coefficient 1.603 −0.619 T P −22.289 0.000 x1x2 0.272 0.057 0.105 −0.022 8.542 1.802 −3.300 −0.115 0.910 x1x2 0.304 0.003 x1x3 0.168 0.059 3.016 0.006 x2x3 0.036 1.018 0.343 Of the following, which is the best next step in the analysis? Page 646 i. Nothing needs to be done. This model is fine. ii. Drop , and from the model, and then perform an F test. iii. Drop x1x3 and x2x3 from the model, and then perform an F test. iv. Drop x1 and 3. from the model, and then perform an F test. v. Add cubic terms , , and to the model to try to improve the fit. 3. The article referred to in Exercise 2, a model was fit to investigate the relationship between the independent variables given in Exercise 2 and the amount of S-isomer collected. The results of a least-squares fit are presented in the following table. (The article did not provide the value of the t statistic for the constant term.) Predictor Constant x1 Coefficient 3.367 −0.018 T P −1.7 0.127 x2 1.396 135.987 0.000 x3 0.104 10.098 0.000 x1x2 0.017 −0.023 −0.030 −0.006 1.471 −0.009 −538 −0.460 0.184 0.394 0.039 0.655 x1x3 0.01 0.943 0.377 x2x3 0.035 4.194 0.004 Of the following, which is the best next step in the analysis? Explain your reasoning. i. Drop , and from the model, and then perform an F test. ii. Nothing needs to be done. This model is fine. iii. Add cubic terms , , and to the model to try to improve the fit. iv. Drop x1x2,x1x3,and x2x3 from the model, and then perform an F test. v. Drop x1,x2andx1x3 from the model, and then perform an F test. 4. An engineer measures a dependent variable y and independent variables x1, x2,and x3. MINITAB output for the model y = β0 + β1x1 + β2x2 + β3x3 + ε is presented below. The regression equation is Y = 0.367 + 1.61 X1 − 5.51 X2 + 1.27 X3 Predictor Coef SE Coef T P Constant 0.3692 0.9231 0.40 0.698 X1 1.6121 1.3395 1.21 0.254 X2 5.5049 1.4959 3.68 0.004 X3 1.2646 1.9760 0.64 0.537 Of the following, which is the best next step in the analysis? Explain your reasoning. i. Add interaction terms x1x2 and x2x3 to try to find more variables to put into the model. ii. Add the interaction term x1x3 to try to find another variable to put into the model. iii. Nothing needs to be done. This model is fine. iv. Drop x1 and x3 from the model, and then perform an F test. v. Drop X2 from the model. vi. Drop the intercept (Constant). Since its P-value is large, it is not useful. Page 647 5. The article "Simultaneous Optimization of Mechanical Properties of Steel by Maximizing Exponential Desirability Functions" (K. J. Kim and D. K. J. Lin, Journal of the Royal Statistical Society Series C, Applied Statistics, 2000: 311–325) presents measurements on 72 steel plates. The following MINITAB output presents the results of a study to determine the relationship between yield strength (in kg/mm2), and the proportion of carbon, manganese, and silicon, each measured in percent. The model fit is The regression equation is Yield Strength = 24.677 − 19.402 Carbon + 14.720 Manganese + 70.720 Silicon Predictor Coef StDev T P Constant 24.677 5.5589 4.21 0.000 Carbon −19.402 28.455 −0.68 0.498 Manganese 14.720 5.6237 2.62 0.011 Silicon 70.720 45.675 1.55 0.126 Of the following, which is the best next step in the analysis? Explain your reasoning. i. 6. Add interaction terms Carbon · Manganese and Manganese · Silicon to try to find more variables to put into the model. ii. Add the interaction term Carbon · Silicon to try to find another variable to put into the model. iii. Nothing needs to be done. This model is fine. iv. Drop Carbon and Silicon, and then perform an F test. v. Drop Carbon from the model. The following MINITAB output is from a best subsets regression involving five dependent variables X1,..., X5.The two models of each size with the highest values of R2 are listed. Best Subsets Regression: Y versus X1, X2, X3, X4, X5 Response is Y Mallows XXXXX Vars R-Sq R-Sq(adj) C-p S 12345 1 77.3 77.1 133.6 1.4051 X 1 10.2 9.3 811.7 2.7940 X 2 89.3 89.0 14.6 0.97126 X X 2 77.7 77.3 130.5 1.3966 X X 3 90.5 90.2 4.8 0.91630 X X X 3 89.4 89.1 14.6 0.96763 X X X 4 90.7 90.3 4.3 0.91446 X X X X 4 90.6 90.2 5.3 0.91942 X X X X 5 90.7 90.2 6.0 0.91805 X X X X X a. Which variables are in the model selected by the minimum Cp criterion? b. Which variables are in the model selected by the adjusted R2 criterion? c. Are there any other good models? Page 648 7. The following is supposed to be the result of a best subsets regression involving five independent variables X1,..., X5. The two models of each size with the highest values of R2 are listed. Something is wrong. What is it? Best Subsets Regression Response is Y Adj. XXXXX Vars R-Sq R-Sq C-p s 12345 1 69.1 68.0 101.4336.79 X 1 60.859.4 135.4379.11 X 2 80.679.2 55.9271.60 X X 2 79.577.9 60.7279.59 X X 3 93.892.7 13.8184.27 X X X 3 93.792.7 18.8978.8 X X X 4 91.490.4 5.5159.59 X X X X 4 90.188.9 5.6159.81 X X X X 5 94.293.0 6.0157.88 X X X X X 8. The article "Effect of Granular Subbase Thickness on Airfield Pavement Response" (K. Gopalakrishnan and M. Thompson, Journal of Materials in Civil Engineering, 2008:331–342) presents a study of the amount of surface deflection caused by aircraft landing on an airport runway. A load of 160 kN was applied to a runway surface, and the amount of deflection in mm (y) was measured at various distances in m (x) from the point of application. The results are presented in the following table. x 0.000 0.305 0.610 0.914 1.219 1.524 1.830 y 3.24 2.36 1.42 0.87 0.54 0.34 0.24 Fit the linear model y = β0 + β1x + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. b. Fit the quadratic model y = β0 + β1x + β2x2 + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. c. Fit the cubic model y = β0 + β1x + β2x2 + β3x3 + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. d. Which of the models in parts (a) through (c) is the most appropriate? Explain. e. Using the most appropriate model, estimate the deflection at a distance of 1 m. (Continues Exercise 7 in Section 8.1.) To try to improve the prediction of FEV1, additional independent variables are included in the model. These new variables are Weight (in kg), the product (interaction) of Height and Weight, and the ambient temperature (in °C). The following MINITAB output presents results of fitting the model a. 9. Page 649 The regression equation is FEV1 = −0.257 + 0.778 Last FEV − 0.105 Gender + 1.213 Height − 0.00624 Weight + 0.00386 Height*Weight − 0.00740 Temp − 0.00148 Pressure Predictor Coef SE Coef T P Constant −0.2565 0.7602 −0.34 0.736 Last FEV 0.77818 0.05270 14.770.000 Gender −0.10479 0.03647 −2.87 0.005 Height 1.2128 0.4270 2.840.005 Weight −0.0062446 0.01351 −0.46 0.645 Height*Weight 0.0038642 0.0084114 0.460.647 Temp −0.007404 0.009313 −0.79 0.428 Pressure −0.0014773 0.0005170 −2.86 0.005 S = 0.22189 R-Sq = 93.5% R-Sq(adj) = 93.2% Analysis of Variance Source DF SS MS F P Regression 7 111.35 15.907323.06 0.000 Residual Error 157 7.7302 0.049237 Total 164 119.08 a. The following MINITAB output, reproduced from Exercise 7 in Section 8.1, is for a reduced model in which Weight, Height*Weight, and Temp have been dropped. Compute the F statistic for testing the plausibility of the reduced model. b. The regression equation is FEV1 = −0.219 + 0.779 Last FEV − 0.108 Gender + 1.354 Height − 0.00134 Pressure Predictor Coef SE Coef T P Constant −0.21947 0.4503 −0.49 0.627 Last FEV 0.779 0.04909 15.870.000 Gender −0.10827 0.0352 −3.08 0.002 Height 1.3536 0.2880 4.700.000 Pressure −0.0013431 0.0004722 −2.84 0.005 S= R-Sq = 93.5% R-Sq(adj) 0.22039 = 93.3% Analysis of Variance Source DF SS MS F P Regression 4 111.31 27.826572.89 0.000 Residual 160 7.7716 0.048572 Error Total 164 119.08 b. c. d. How many degrees of freedom does the F statistic have? Find the P-value for the F statistic. Is the reduced model plausible? Someone claims that since each of the variables being dropped had large P-values, the reduced model must be plausible, and it was not necessary to perform an F test. Is this correct? Explain why or why not. Page 650 The total sum of squares is the same in both models, even though the independent variables are different. Is there a mistake? Explain. 10. In a study to determine the effect of vehicle weight in tons (x1) and engine displacement in in3 (x2) on fuel economy in miles per gallon (y), these measurements were taken for ten automobiles. The full quadratic model was fit to the data, and the sum of squares for error was SSE = 62.068. Then the reduced model y = β0 + β1x1 + β2x2 + ε was fit, and the sum of squares for error was SSE = 66.984. Is it reasonable to use the reduced model, rather than the full quadratic model, to predict fuel economy? Explain. 11. In a study of the causes of bearing wear, a machine was run 24 times, with various loads (denoted x1), oil viscosities (x2), and ambient temperatures (x3). The wear, denoted y, was modeled as y = β0 + β1x1 + β2x2 + β3x3 + ε. When this model was fit to the data, the sum of squares for error was SSE = 9.37. Then the reduced model y = β0 + β1x1 + β2x2 + β3x3 was fit, and the sum of squares for error was SSE = 27.49. Is it reasonable to use the reduced model, rather than the model containing all the interactions, to predict wear? Explain. 12. In rock blasting, the peak particle velocity (PPV) depends both on the distance from the blast and on the amount of charge. The article "Prediction of Particle Velocity Caused by Blasting for an Infrastructure Excavation Covering Granite Bedrock" (A. Kahriman, Mineral Resources Engineering, 2001:205–218) presents data on PPV, scaled distance (which is equal to the distance divided by the square root of the charge), and the amount of charge. The following table presents the values of PPV, scaled distance, and amount of charge for 15 blasts. PPV (mm/s) 1.4 15.7 2.54 1.14 0.889 1.65 1.4 26.8 1.02 4.57 6.6 1.02 3.94 1.4 1.4 Scaled Distance (m/kg0.5) 47.33 9.6 15.8 24.3 23.0 12.7 39.3 8.0 29.94 10.9 8.63 28.64 18.21 33.0 34.0 Amount of Charge (kg) 4.2 92.0 40.0 48.7 95.7 67.7 13.0 70.0 13.5 41.0 108.8 27.43 59.1 11.5 175.0 a. Fit the model ln PPV = β0 + β1 ln Scaled Distance + β2 ln Charge + ε. Compute the Pvalue for testing H0 : βi = 0 for β0, β1, and β2. b. The article claims that the model ln PPV = β0 + β1 ln Scaled Distance + ε is appropriate. Fit this model. Compute the P-value for testing H0 : βi = 0 for β0, β1, and β2. c. Which model do you prefer? Why? 13. The article "Ultimate Load Analysis of Plate Reinforced Concrete Beams" (N. Subedi and P. Baglin, Engineering Structures, 2001:1068–1079) presents theoretical and measured ultimate strengths (in kN) for a sample of steel-reinforced concrete beams. The results are presented in the following table (two outliers have been deleted). Let y denote the measured strength, x the theoretical strength, and t the true strength, which is unknown. Assume that y = t + ε, where ε is the measurement error. It is uncertain whether t is related to x by a linear model t = β0 + β1x or by a quadratic model t = β0 + β1x + β2x2. Theoretical 991 785 1195 1021 1285 1167 1519 1314 1743 791 1516 1071 1480 1622 2032 2032 660 565 738 682 a. b. c. d. Measured 1118 902 1373 1196 1609 1413 1668 1491 1952 844 1550 1167 1609 1756 2119 2237 640 530 893 775 Fit the linear model y = β0 + β1x + ε. For each coefficient, find the P-value for the null hypothesis that the coefficient is equal to 0. Fit the quadratic model y = β0 + β1x + β2x2 + ε. For each coefficient, find the P-value for the null hypothesis that the coefficient is equal to 0. Plot the residuals versus the fitted values for the linear model. Plot the residuals versus the fitted values for the quadratic model. Based on the results in parts (a) through (d), which model seems more appropriate? Explain. f. Using the more appropriate model, estimate the true strength if the theoretical strength is 1500. g. Using the more appropriate model, find a 95% confidence interval for the true strength if the theoretical strength is 1500. 14. The article "Permanent Deformation Characterization of Subgrade Soils from RLT Test" (A. Puppala, L. Mohammad, et al., Journal of Materials in Civil Engineering, 1999:274– 282) presents measurements of plastic strains (in percent) on soils at various confining and deviator stresses in kPa (y), the confining stress (x1), and the deviatoric stress (x2) for tests on a sandy soil. e. y 0.01 0.02 0.05 0.09 0.003 0.006 0.05 0.23 0.003 0.0008 0.01 0.07 0.002 0.0003 0.0009 0.01 0.007 0.0003 0.005 0.03 a. b. c. x1 21 21 21 21 35 35 35 35 70 70 70 105 105 105 105 140 140 140 140 x2 21 35 52.5 70 35 70 105 140 105 140 210 105 140 210 280 140 210 280 Page 652 Fit the model y = β0 + β1x1 + β2x2 + ε. Plot the residuals versus the fitted values. Does the model seem appropriate? Fit the model ln y = β0 + β1 ln x1 + β2 ln x2 + ε. Plot the residuals versus the fitted values. Does the model seem appropriate? Use the more appropriate of the models from (a) and (b) to predict y when x1 = 50 and x2 = 100. d. Is the model you used in part (c) improved by including an interaction term? Explain. 15. The article "Vehicle-Arrival Characteristics at Urban Uncontrolled Intersections" (V. Rengaraju and V. Rao, Journal of Transportation Engineering, 1995:317–323) presents data on traffic characteristics at 10 intersections in Madras, India. The following table provides data on road width in m (x1), traffic volume in vehicles per lane per hour (x2), and median speed in km/h (y). x 35.0 37.5 26.5 33.0 22.5 26.5 27.5 28.0 23.5 24.5 x1 x2 76 78 76 60 65 72 92 50 86 370 475 507 654 917 842 723 1039 1120 Fit the model y = β0 + β1x1 + β2x2 + ε. Find the P-values for testing that the coefficients are equal to 0. c. Fit the model y = β0 + β1x1 + β2x2 + ε. Find the P-values for testing that the coefficients are equal to 0. d. Which of the models in parts (a) through (c) do you think is best? Why? 16. The following table presents measurements of mean noise levels in dBA (y), roadway width in m (x1), and mean speed in km/h (x2), for 10 locations in Bangkok, Thailand, as reported in the article "Modeling of Urban Area Stop-and-Go Traffic Noise" (P. Pamanikabud and C. Tharasawatipipat, Journal of Transportation Engineering, 1999:152–159). a. y 78.1 78.1 79.6 81.0 78.7 78.1 78.6 78.5 78.4 x1 x2 6.0 10.0 12.0 6.0 12.0 6.5 6.5 12.0 6.5 6.5 30.61 36.55 36.22 38.73 29.07 28.26 30.28 30.25 29.03 79.6 65.33 Construct a good linear model to predict mean noise levels using roadway width, mean speed, or both, as predictors. Provide the standard deviations of the coefficient estimates and the P-values for testing that they are different from 0. Explain how you chose your model. 17. The article "Modeling Resilient Modulus and Temperature Correction for Saudi Roads" (H. Wahhab, I. Asi, and R. Ramadhan, Journal of Materials in Civil Engineering, 2001:298– 305) describes a study designed to predict the resilient modulus of pavement from physical properties. The following table presents data for the resilient modulus at 40°C in 106 kPa (y), the surface area of the aggregate in m2/kg (x1), and the softening point of the asphalt in °C (x2). Page 653 y 1.48 1.70 2.03 2.86 2.43 3.06 4.14 1.29 3.53 1.04 1.88 1.90 1.76 2.82 1.00 The full quadratic model is x1 5.77 7.45 8.14 8.73 7.12 6.89 8.64 6.58 9.10 8.06 5.93 8.17 9.84 7.77 7.70 6.64 6.65 3.66 2.64 1.07 0.87 x2 60.5 74.9 67.4 77.7 71.7 75.0 66.3 62.4 68.9 68.5 69.5 68.9 63.2 62.1 68.9 72.2 63.4 68.0 69.0 3.25 5.75 3.94 1. Which submodel of this full model do you believe is most appropriate? Justify your answer by fitting two or more models and comparing the results. 18. The article "Models for Assessing Hoisting Times of Tower Cranes" (A. Leung and C. Tam, Journal of Construction Engineering and Management, 1999: 385–391) presents a model constructed by a stepwise regression procedure to predict the time needed for a tower crane hoisting operation. Twenty variables were considered, and the stepwise procedure chose a nine-variable model. The adjusted R2 for the selected model was 0.73. True or false: a. The value 0.73 is a reliable measure of the goodness of fit of the selected model. b. The value 0.73 may exaggerate the goodness of fit of the model. c. A stepwise regression procedure selects only variables that are of some use in predicting the value of a dependent variable. b. It is possible for a variable that is of no use in predicting the value of a dependent variable to be part of a model selected by a stepwise regression procedure. Supplementary Exercises for Chapter 8 1. The article "Advances in Oxygen Equivalence Equations for Predicting the Properties of Titanium Welds" (D. Harwig, W. Ittiwattana, and H. Castner, The Welding Journal, 2001:126s–136s) reports an experiment to predict various properties of titanium welds. Among other properties, the elongation (in %) was measured, along with the oxygen content and nitrogen content (both in percent). The following MINITAB output presents results of fitting the model The regression equation is Elongation = 46.80 − 130.11 Oxygen − 807.1 Nitrogen + 3580.5 Oxy*Nit Predictor Coef SE T P Coef Constant 46.802 3.702 12.64 0.000 Oxygen −130.11 20.467 −6.36 0.000 Nitrogen −807.10 158.03 −5.107 0.000 Oxy*Nit 3580.5 958.05 3.737 0.001 S= R-Sq = 74.5% R-Sq(adj) = 2.809 72.3% Analysis of Variance Source DF SS MS F P Regression 3 805.43 268.4834.03 0.000 Residual 35 276.11 7.89 Error Total 38 1081.54 a. 2. Page 654 Predict the elongation for a weld with an oxygen content of 0.15% and a nitrogen content of 0.01%. b. If two welds both have a nitrogen content of 0.006%, and their oxygen content differs by 0.05%, what would you predict their difference in elongation to be? c. Two welds have identical oxygen contents, and nitrogen contents that differ by 0.005%. Is this enough information to predict their difference in elongation? If so, predict the elongation. If not, explain what additional information is needed. Refer to Exercise 1. a. Find a 95% confidence interval for the coefficient of Oxygen. b. Find a 99% confidence interval for the coefficient of Nitrogen. c. Find a 98% confidence interval for the coefficient of the interaction term Oxygen · Nitrogen. d. Can you conclude that β1 < − 75? Find the P-value. e. Can you conclude that β2 > − 1000? Find the P-value. c. 3. The following MINITAB output is for a multiple regression. Some of the numbers got smudged, becoming illegible. Fill in the missing numbers. Predictor Coef SE T P Coef Constant (a)0.3501 0.59 0.568 X1 1.8515 (b) 2.31 0.040 X2 2.7241 0.7124 (c) 0.002 S = (d) R-Sq = 83.4% R-Sq(adj) = 80.6% Analysis of Variance Source DF SS MS F P Regression (e) (f) (g) (h)0.000 Residual 12 17.28 1.44 Error Total (i)104.09 4. 5. An engineer tries three different methods for selecting a linear model. First she uses an informal method based on the F statistic, as described in Section 8.3. Then she runs the best subsets routine, and finds the model with the best adjusted R2 and the one with the best Mallows' Cp. It turns out that all three methods select the same model. The engineer says that since all three methods agree, this model must be the best one. One of her colleagues says that other models might be equally good. Who is right? Explain. In a simulation of 30 mobile computer networks, the average speed, pause time, and number of neighbors were measured. A "neighbor" is a computer within the transmission range of another. The data are presented in the following table. Page 655 Neighbors 10 17 8 46 10 29 9 30 13 17 6 44 10 19 7.30 10 18 Speed 5 10 10 20 20 30 30 40 Pause 0 30 0 7.19 9.36 8.40 7.73 8.27 7.30 7 7.14 7.72 7.05 8.02 8.08 8.27 7.86 7.68 7.21 7.45 7.08 7.32 6.99 a. b. c. f. e. 6. 40 5 10 10 20 20 30 30 40 40 50 10 10 10 10 10 10 40 40 40 10 10 10 10 20 20 20 20 20 20 20 20 20 30 30 30 30 30 40 40 40 40 40 40 50 50 50 50 50 50 Fit the model with Neighbors as the dependent variable, and independent variables Speed, Pause, Speed · Pause, Speed2, and Pause2. Construct a reduced model by dropping any variables whose P-values are large, and test the plausibility of the model with an F test. Plot the residuals versus the fitted values for the reduced model. Are there any indications that the model is inappropriate? If so, what are they? Someone suggests that a model containing Pause and Pause2 as the only dependent variables is adequate. Do you agree? Why or why not? Using a best subsets software package, find the two models with the highest R2 value for each model size from one to five variables. Compute Cp and adjusted R2 for each model. d. Which model is selected by minimum Cp? By adjusted R2? Are they the same? The data in Table SE6 (page 656) consist of yield measurements from many runs of a chemical reaction. The quantities varied were the temperature in °C (x1), the concentration of the primary reactant in % (x2), and the duration of the reaction in hours (x3). The dependent variable (y) is the fraction converted to the desired product. a. Fit the linear model y = β0 + β1x1 + β2x2 + β3x3 + ε. b. Two of the variables in this model have coefficients significantly different from 0 at the 15% level. Fit a linear regression model containing these two variables. Compute the product (interaction) of the two variables referred to in part (b). Fit the model that contains the two variables along with the interaction term. Based on the results in parts (a) through (c), specify a model that appears to be good for predicting y from x1, x2, and x3. c. d. e. Might it be possible to construct an equally good or better model in another way? Page 656 TABLE SE6 Data for Exercise 6 x1 x2 x3 50 90 50 70 60 70 80 70 60 70 80 70 70 70 70 80 70 70 70 80 60 70 60 70 80 70 80 60 70 60 70 70 80 80 70 70 70 70 70 70 60 70 70 60 80 70 80 60 70 60 70 60 60 70 80 70 70 70 80 70 80 70 70 60 70 70 70 70 70 80 70 80 70 60 70 70 70 60 50 70 70 70 70 70 70 70 70 70 60 70 70 70 70 70 70 70 70 70 80 70 70 70 70 70 70 70 70 70 70 70 70 70 70 70 70 70 70 70 20 26 32 32 20 26 26 32 21 20 26 21 26 32 26 26 20 32 32 26 21 17 34 27 24 4.0 8.0 6.5 5.5 6.5 6.5 0.5 5.5 5.5 6.5 4.5 7.0 8.5 5.5 6.5 4.0 7.0 8.5 4.0 6.5 7.5 5.5 7.5 6.0 7.5 5.5 7.5 6.0 7.0 6.5 7.0 7.0 6.0 8.5 5.5 6.0 7.5 6.5 6.5 5.5 7.0 6.5 7.0 6.0 7.5 6.5 5.5 6.5 6.0 9.5 5.5 6.0 10.0 4.5 4.0 7.0 2.0 8.5 5.5 2.5 y 27.464 49.303 40.152 38.947 48.391 38.021 33.611 36.478 33.776 35.092 31.307 37.863 41.109 28.605 35.917 33.489 31.381 38.067 31.278 32.172 26.103 31.533 35.877 37.637 40.263 36.694 45.620 38.571 19.163 31.962 23.147 47.278 32.725 28.735 70 80 50 50 80 50 80 60 60 60 60 60 70 60 70 60 60 60 60 60 60 60 70 60 60 60 60 60 80 70 60 70 60 60 60 60 70 70 60 70 60 70 60 60 60 70 60 70 60 70 50 60 70 60 70 60 80 70 70 70 70 80 70 70 70 80 70 80 70 70 80 80 70 70 70 70 70 80 70 70 32 26 24 21 30 22 26 25 24 18 25 20 25 33 31 27 19 15 32 27 32 28 25 20 27 28 25 21 32 26 25 21 26 24 31 20 34 32 20 21 17 28 25 20 25 20 26 20 24 25 6.5 5.5 6.5 7.0 6.5 7.5 6.5 5.5 6.0 7.5 6.0 7.0 6.5 6.0 7.5 6.0 5.5 6.0 6.0 8.0 6.5 6.0 6.5 5.5 7.5 6.5 8.5 6.5 5.5 6.0 5.5 5.5 7.5 6.5 5.5 4.5 5.5 7.5 6.0 4.5 6.0 7.0 5.5 6.5 5.5 6.5 8.0 6.5 4.0 6.5 7.0 6.0 4.5 6.5 7.0 4.5 6.5 8.5 6.5 6.5 5.5 6.5 8.5 6.5 5.5 7.5 9.5 5.5 6.0 5.5 6.5 5.5 6.0 5.5 4.5 7.0 6.5 6.0 7.0 6.5 5.5 6.5 6.0 7.0 5.5 7.5 7.0 5.5 6.0 7.0 6.0 7.0 5.5 7.5 6.5 7.0 5.5 6.0 7.5 5.5 6.0 7.0 6.5 6.0 7.0 7.0 6.5 6.0 6.0 7.0 7.0 6.5 6.0 6.0 7.5 6.5 7.0 6.0 7.5 6.0 7.0 5.5 7.0 6.5 7.0 6.5 6.0 7.0 6.5 7.0 7.0 7.0 7.0 6.5 7.5 6.5 7.0 6.5 7.0 6.5 7.0 6.5 7.0 7.0 7.0 7.5 6.5 7.0 6.5 6.0 7.0 4.5 6.5 7.5 7.0 5.5 7.0 6.0 4.0 6.0 7.0 6.5 6.5 6.0 6.5 7.5 6.0 6.5 7.0 5.5 6.5 6.0 7.0 6.0 6.5 6.0 7.0 6.0 7.0 7.0 6.5 6.5 5.5 6.0 7.5 5.5 6.5 6.5 7.0 7.5 5.5 7.0 7.0 5.0 7.0 7.0 7.0 5.5 7.5 7.5 7.5 7.5 8.0 9.5 7.5 6.5 5.5 6.0 6.5 6.5 38.241 34.635 44.963 30.012 41.077 41.964 44.152 29.901 26.706 28.602 33.401 41.324 24.000 38.158 25.412 37.671 27.979 31.079 30.778 28.221 30.495 38.710 27.581 38.705 40.525 29.420 37.898 40.340 27.891 38.259 35.091 34.372 26.481 36.739 36.185 38.725 32.707 32.563 36.006 33.127 7. 70 29.5 0 32.941 70 50 60 60 60 60 60 60 60 60 60 50 60 60 60 60 60 60 70 19 24 26 28 22 28 25 20 26 25 29 30 26 24 25 23 26 24 26 26 4.5 6.5 7.5 5.0 6.0 7.5 5.5 7.0 4.5 8.5 9.5 4.0 6.0 7.5 6.0 7.0 6.0 6.0 7.0 5.5 6.5 7.0 6.0 5.5 6.0 7.5 6.5 4.0 6.0 6.5 6.0 8.0 6.5 6.0 5.5 5.0 8.0 33.650 34.192 24.115 37.614 29.612 39.196 36.974 28.334 33.767 38.358 33.381 37.672 36.615 39.351 38.611 36.460 23.449 23.027 31.372 In a study to predict temperature from air pressure in psi (y) from temperature in °F (x), 19 measurements were made of temperature and pressure. The results are presented in the following table. a. Fit the linear model y = β0 + β1x + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. b. Fit the quadratic model y = β0 + β1x + β2x2 + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. c. Fit the cubic model y = β0 + β1x + β2x2 + β3x3 + ε. For each model, which model is the most appropriate. Page 657 Linear Model Residual Fit −56.2 125.6 −4.4 130.1 85.1 8.4 179.8 21.4 207.2 28.6 234.7 46.9 260.9 47.2 288.1 8.5 314.4 −7.1 342.0 −47.1 139.3 Quadratic Model Residual Fit 11 258.2 −7.4 126.5 4.9 183.4 −3.6 232.2 −8.2 271.5 8.1 299.7 15.7 319.6 −7.0 329.8 4.0 330.9 −1.3 93.6 Cubic Model Residual Fit 3.3 66.1 −6.7 125.7 9.2 179.0 0.9 227.8 −5.9 269.2 7.3 300.5 12.1 323.2 −11.7 334.6 0.3 333.9 −4.1 86.4 −1.6 38.0 35.7 34.1 34.6 1.0 −23.2 −50.7 −72.9 8. 166.2 220.9 247.8 273.1 301.1 328.2 355.3 368.4 382.1 9.1 5.9 −3.2 −2.1 9.9 −2.4 4.1 −5.0 −5.7 155.5 253.0 286.7 311.3 325.8 331.6 327.9 322.7 314.9 12.1 9.5 −2.4 −4.5 5.5 −6.7 35 −2.1 21.1 152.5 249.9 286.0 313.7 330.2 335.9 328.5 319.9 307.1 The voltage output (y) of a battery was measured over a range of temperatures (x) from 0°C to 50°C. The following figure is a scatterplot of voltage versus temperature, with three fitted curves superimposed. The curves are the linear model y = β0 + β1x + ε, the quadratic model y = β0 + β1x = β2x2 + ε, and the cubic model y = β0 + β1x + β2x2 + β3x3 + ε. Based on the plot, which of the models should be used to describe the data? Explain. i. The linear model. ii. The quadratic model. iii. The cubic model. iv. All three appear to be about equally good. Page 658 9. Refer to Exercise 2 in Section 8.2. a. Using each of the three models in turn, predict the NOx emission when Speed = 1500, Torque = 400, and HP = 150. b. Using each of the three models in turn, predict the NOx emission when Speed = 1600, Torque = 300, and HP = 100. c. Using each of the three models in turn, predict the NOx emission when Speed = 1400, Torque = 200, and HP = 75. d. Which model or models appear to be best? Choose one of the answers, and explain. i. ii. iii. iv. The model with Speed and Torque as independent variables is the best. The model with Speed and HP as independent variables is the best. The model with Speed, Torque, and HP as independent variables is the best. The model with Speed and Torque and the model with Speed and HP are about equally good; both are better than the model with Speed, Torque, and HP. v. The model with Speed and Torque and the model with Speed, Torque, and HP are about equally good; both are better than the model with Speed and HP. vi. The model with Speed and HP and the model with Speed, Torque, and HP are about equally good; both are better than the model with Speed and Torque. vii. All three models are about equally good. 10. This exercise illustrates a reason for the exceptions to the rule of parsimony (see page 630). a. A scientist fits the model Y = β1C + ε, where C represents temperature in °C and Y can represent any outcome. Note that the model has no intercept. Now convert °C to °F (C = 0.556F − 17.78). Does the model have a linear term now? c. Assume that x and z are two different units that can be used to measure the same quantity, and that z = a + bx, where a ≠ 0. (°C and °F are an example.) Show that the no-intercept models y = βx and y = β2z cannot both be correct, so that the validity of a no-intercept model depends on the zero point of the units for the independent variable. d. Let x and z be as in part (c). Show that the models y = β0 + β2x2 and y = β0 + β2z2 cannot both be correct, and, thus, that the validity of such a model depends on the zero point of the units for the independent variable. 11. The data presented in the following table give the tensile strength in psi (y) of paper as a function of the percentage of hardwood content (x). Hardwood Content 1.0 1.5 2.0 3.0 4.0 4.5 5.0 5.5 6.0 6.5 7.0 8.0 Tensile Strength 26.8 29.5 36.6 37.8 38.2 41.5 44.8 44.7 48.5 50.1 52.1 56.1 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 63.1 62.0 62.5 58.0 52.9 38.2 32.9 21.9 Page 659 a. Fit polynomial models of degrees 1, 2, and so on to predict tensile strength from hardwood content. For each one, use the F test to compare it with the model of degree one less. Stop when the P-value of the F test is greater than 0.05. What is the degree of the polynomial model chosen by this method? b. Using the model from part (a), estimate the hardwood concentration that produces the highest tensile strength. 12. The article "Enthalpies and Entropies of Transfer of Electrolytes and Ions from Water to Mixed Aqueous Organic Solvents" (G. Hefter, Y. Marcus, and W. Waghorne, Chemical Reviews, 2002:2773–2836) presents measurements of entropy and enthalpy changes for many salts under a variety of conditions. The following table presents the results for enthropies of transfer (in J/K · mol) from water to water + methanol of NaCl (table salt) over a range of concentrations of methanol: Concentration (%) 5 10 20 30 40 50 60 70 80 90 100 a. Entropy 1 −1 −7 −17 −28 −39 −52 −65 −80 −98 −121 Fit polynomial models of degrees 1, 2, and 3 to predict entropy (y) from the concentration (x). b. Which degree polynomial is the most appropriate? Explain. c. Using the most appropriate model, find 99% confidence intervals for each of the coefficients. 13. The following table presents measured data on the durability of its paint and that of its competitors. The company measured the lifetimes of three samples of each type of paint used in several American cities. The results are given in Table SE13. TABLE SE13 Data for Exercise 13 Avg. Temp (°F) City January July Atlanta, GA 41.9 78.6 Boston, MA 29.6 73.5 Kansas City, 28.4 80.9 KS Minneapolis, 11.2 73.1 MN Dallas, TX 45.0 86.3 Denver, CO 29.5 73.3 Miami, FL 67.1 82.4 Phoenix, AZ 52.3 92.3 San 48.5 62.2 Francisco, CA Seattle, WA 40.6 65.3 Washington, 35.2 78.9 DC Mean Annual Precipitation (in.) 48.6 43.8 29.3 Lifetime (years) Sponsor's Competitor's Paint Paint 11.5 10.7 12.3 10.8 11.1 10.2 11.7 10.1 11.9 12.5 10.7 11.4 10.9 12.3 13.4 12.8 11.8 11.2 11.5 10.7 12.3 10.8 11.1 10.2 11.7 10.1 11.9 12.5 10.7 11.4 10.9 12.3 11.0 12.2 13.4 7.1 10.7 11.2 15.2 8.7 11.1 16.7 38.9 39.4 14.2 14.1 13.6 12.6 15.4 13.6 14.1 12.6 11.5 12.0 11.9 10.9 11.4 10.6 14.2 7.9 11.8 17.2 12.0 13.8 9.4 12.4 15.9 10.6 13.4 8.1 10.9 15.8 9.2 10.1 11.4 14.4 13.2 8.6 7.6 10.1 9.9 15.4 14.9 Page 660 a. Prior testing suggests that the most important factors that influence the lifetimes of paint coatings are the minimum temperature (estimated by the average January temperature), the maximum temperature (estimated by the average July temperature), and the annual precipitation. Using these variables, and products and powers of these variables, construct a good model for predicting the lifetime of the sponsor's paint and a good model (perhaps different) for predicting the lifetime of the competitor's paint. b. Using the models developed in part (a), compute the expected lifetimes for these two paints for someone living in Cheyenne, Wyoming, where the January mean temperature is 26.1°F, the July mean temperature is 68.9°F, and the mean annual precipitation is 13.3 in. 14. The article "Two Different Approaches for RDC Modelling When Simulating a Solvent Deasphalting Plant" (J. Aparicio, M. Heronimo, et al., Computers and Chemical Engineering, 2002:1369–1377) reports flow rate (in dm3/h) and specific gravity measurements for a sample of paraffinic hydrocarbons. The natural logs of the flow rates (y) and the specific gravity measurements (x) are presented in the following table. y −1.204 −0.580 x 0.8139 0.8171 0.049 0.673 1.311 1.959 2.614 3.270 0.8202 0.8233 0.8264 0.8294 0.8323 0.8352 Fit the linear model y = β0 + β1x + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. b. Fit the quadratic model y = β0 + β1x + β2x2 + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. c. Fit the cubic model y = β0 + β1x + β2x2 + β3x3 + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. d. Which of the models in parts (a) through (c) is the most appropriate? Explain. e. Using the most appropriate model, estimate the flow rate when the specific gravity is 0.83. 15. The article "Measurements of the Thermal Conductivity and Thermal Diffusivity of Polymer Melts with the Short-Hot-Wire Method" (X. Zhang, W. Hendro, et al., International Journal of Thermophysics, 2002:1077–1090) reports measurements of the thermal conductivity (in W · m−1 · K−1) and diffusivity of several polymers at several temperatures (in 1000°C). The following table presents results for the thermal conductivity of polycarbonate. a. Cond. 0.236 0.241 0.244 0.251 0.259 0.257 0.257 0.261 0.254 0.256 0.251 0.249 0.249 0.230 0.230 0.228 Temp. 0.028 0.038 0.061 0.083 0.107 0.119 0.130 0.146 0.159 0.169 0.181 0.204 0.215 0.225 0.237 0.248 a. b. Denoting conductivity by y and temperature by x, fit the linear model y = β0 + β1x + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. Fit the quadratic model y = β0 + β1x + β2x2 + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. Page 661 Fit the cubic model y = β0 + β1x + β2x2 + β3x3 + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. d. Fit the quartic model y = β0 + β1x + β2x2 + β3x3 + β4x4 + ε. For each coefficient, test the hypothesis that the coefficient is equal to 0. e. Which of the models in parts (a) through (d) is the most appropriate? Explain. f. Using the most appropriate model, estimate the conductivity at a temperature of 120°C. 16. The article "Electrical Impedance Variation with Water Saturation in Rock" (Q. Su, Q. Feng, and Z. Shang, Geophysics, 2000:68–75) reports measurements of permeabilities (in 10−3μm2), porosities (in percent), and surface area per unit volume of pore space (in 104 cm−1) for several rock samples. The results are presented in the following table, denoting ln Permeability by y, porosity by x1, and surface area per unit volume by x2. c. y −0.27 2.58 3.18 1.70 −1.17 −0.27 −0.53 −0.29 4.94 1.94 3.74 0.58 −0.56 −0.49 0.01 −1.71 −0.12 −0.92 2.18 4.46 2.11 −0.04 x1 x2 19.83 17.93 21.27 18.67 7.98 10.16 17.86 5.43 18.48 17.49 14.18 23.88 10.52 18.92 18.55 13.72 9.12 14.39 11.38 16.59 16.77 18.55 18.02 9.55 10.97 31.02 28.12 52.35 32.82 57.66 21.10 9.15 11.72 5.43 20.03 16.11 5.72 10.16 17.78 40.28 53.67 26.75 75.62 9.95 7.88 88.10 10.95 a. Fit the model y = β0 + β1x1 + β2x2 + β3x1x2 + ε. Compute the analysis of variance table. b. Fit the model y = β0 + β1x1 + β2x2 + ε. Compute the analysis of variance table. c. Fit the model y = β0 + β1x1 + ε. Compute the analysis of variance table. d. Compute the F statistics for comparing the models in parts (b) and (c) with the model in part (a). Which model do you prefer? Why? 17. The article "Groundwater Electromagnetic Imaging in Complex Geological and Topographical Regions: A Case Study of a Tectonic Boundary in the French Alps" (S. Houtot, P. Tarits, et al., Geophysics, 2002:1048–1060) presents measurements of concentrations of several chemicals (in mmol/L) and electrical conductivity (in 10−2 S/m) for several water samples in various locations near Gittaz Lake in the French Alps. The results for magnesium and calcium are presented in the following table. Two outliers have been deleted. a. Conductivity 2.77 3.03 3.09 3.29 3.37 0.88 0.77 0.97 1.10 1.11 2.57 3.27 2.28 3.32 3.93 4.26 a. Magnesium 0.037 0.041 0.215 0.166 0.100 0.031 0.012 0.027 0.039 0.168 0.172 0.044 0.069 0.188 Calcium 1.342 1.500 1.332 1.609 1.627 0.382 0.364 0.467 0.487 0.497 1.093 1.480 1.093 1.754 1.974 2.103 Page 662 To predict conductivity (y) from the concentrations of magnesium (x1) and calcium (x2), fit the full quadratic model y = β0 + β1x1 + β2x2 + β3x12 + β4x22 + β5x1x2 + ε. Compute the analysis of variance table. b. Use the F test to investigate some submodels of the full quadratic model. State which model you prefer and why. c. Use a best subsets routine to find the submodels with the maximum adjusted R2 and the minimum Mallows Cp Are they the same model? Comment on the appropriateness of this (these) model(s). 18. The article "Low-Temperature Heat Capacity and Thermodynamic Properties of 1,1,1-trifluoro-2, 2-dichloroethane" (R. Varushchenko and A. Druzhinina, Fluid Phase Equilibria, 2002:109–119) presents measurements of the molar heat capacity (y) of 1,1,1-trifluoro-2,2-dichloroethane (in J · K−1 · mol−1) at several temperatures (x) in units of 10 K. The results for every tenth measurement are presented in the following table. y 5.7037 16.707 29.717 41.005 48.822 55.334 60.354 62.077 71.328 75.822 80.029 a. x 1.044 1.687 2.531 3.604 4.669 5.722 6.765 7.798 9.241 10.214 11.266 Fit the simple linear model y = β0 + β1x + ε. Make a residual plot, and comment on the appropriateness of the model. b. Fit the simple linear model y = β0 + β1 ln x + ε. Make a residual plot, and comment on the appropriateness of the model. c. Compute the coefficients and their standard deviations for polynomials of degrees 2, 3, 4, and 5. Make residual plots for each. d. The article cited at the beginning of this exercise recommends the quartic model y = β0 + β1x + β2x2 + β3x3 + β4x4 + ε. Does this seem reasonable? Why or why not? 19. The article "Lead Dissolution from Lead Smelter Slags Using Magnesium Chloride Solutions" (A. Xenidis, T. Lillis, and I. Hallikia, The AusIMM Proceedings, 1999:37–14) discusses an investigation of leaching rates of lead in solutions of magnesium chloride. The data in the following table (read from a graph) present the percentage of lead that has been extracted at various times (in minutes). Time (t) 4 8 16 30 60 120 Percent extracted (y) 1.2 1.6 2.3 2.8 3.6 4.4 a. b. The article suggests fitting a quadratic model y = β0 + β1t + β2t2 + ε to these data. Fit this model, and compute the standard deviations of the coefficients. The reaction rate at time t is given by the derivative dy/dt = β1 + 2β2t. Estimate the time at which the reaction rate will be equal to 0.05. The reaction rate at t = 0 is equal to β1. Find a 95% confidence interval for the reaction rate at t = 0. d. Can you conclude that the reaction rate is decreasing with time? Explain. 20. The article "The Ball-on-Three-Ball Test for Tensile Strength: Refined Methodology and Results for Three Hohokam Ceramic Types" (M. Beck, American Antiquity, 2002:558–569) discusses the strength of ancient ceramics. The following table presents measured weights (in g), thicknesses (in mm), and loads (in kg) required to crack the specimen for a collection of specimens dated between A.D. 1100 and 1300 from the Middle Gila River, in Arizona. Page 663 a. Weight (x1) Thickness (x2) 12.7 12.9 17.8 18.5 13.4 15.2 13.2 18.3 16.2 14.7 18.2 14.8 17.7 16.0 17.2 14.1 16.1 5.69 5.05 6.53 6.51 5.92 5.88 4.09 6.14 5.73 5.47 7.32 4.91 6.72 5.85 6.18 5.13 5.71 Load (y) 20 16 20 36 27 35 15 18 24 21 30 20 24 23 21 13 21 Fit the model y = β0 + β1x1 + β2x2 + ε. Drop the variable whose coefficient has the larger P-value, and refit. Plot the residuals versus the fitted values from the model in part (b). Are there any indications that the model is not appropriate? 21. Piecewise linear model: Let x be a known constant, and suppose that a dependent variable y is related to an independent variable x1 as follows: b. c. In other words, y and x1 are linearly related, but different lines are appropriate depending on whether x1 is above or below . Define a new independent variable x2 by and . Find a multiple regression model involving y, x1, and x2 that expresses the relationship described here. 22. The article "Seismic Hazard in Greece Based on Different Strong Motion Parameters" (S. Koutrakis, G. Karakaisis, et al., Journal of Earthquake Engineering, 2002:75–109) presents a study of seismic events in Greece during the period 1978–1997. Of interest is the duration of "strong ground motion," which is the length of time that the acceleration of the ground exceeds a specified value. For each event, measurements of the duration of strong ground motion were made at one or more locations. Table SE22 (page 664) presents, for each of 121 such measurements, the data for the duration of time y (in seconds) that the ground acceleration exceeded twice the acceleration due to gravity, the magnitude m of the earthquake, the distance d (in km) of the measurement from the epicenter, and two indicators of the soil type s1 and s2, defined as follows: s1 = 1 if the soil consists of soft alluvial deposits, s1 = 0 otherwise, and s2 = 1 if the soil consists of tertiary or older rock, s2 = 0 otherwise. Cases where both s1 = 0 and s2 = 0 correspond to intermediate soil conditions. The article presents repeated measurements at some locations, which we have not included here. Page 664 TABLE SE22 y 8.82 4.08 15.90 6.04 0.15 5.06 0.01 4.13 0.02 2.14 4.41 17.19 5.14 0.05 20.00 12.04 0.87 0.62 8.10 1.30 11.92 3.93 2.00 0.43 14.22 0.06 1.48 3.27 6.36 1.98 0.42 5.56 ... Data for Exercise 22 m 6.4 5.2 6.9 5.8 4.9 6.2 6.6 5.1 5.3 4.5 5.2 5.9 5.5 4.9 5.8 6.1 5.0 4.8 5.7 ... d 30 6 12 34 5 10 15 58 8 ... s1 s2 ... Use the data in Table SE22 to construct a linear model to predict duration y from some or all of the variables m, d, s1, and s2. Be sure to consider transformations of the variables, as well as powers of and interactions between the independent variables. Describe the steps taken to construct your model. Plot the residuals versus the fitted values to verify that your model satisfies the necessary assumptions. In addition, note that the data are presented in chronological order, reading down the columns. Make a plot to determine whether time should be included as an independent variable. Page 665 23. The article "Estimating Resource Requirements at Conceptual Design Stage Using Neural Networks" (A. Elazouni, I. Nosair, et al., Journal of Computing in Civil Engineering, 1997:217–223) suggests that certain resource requirements in the construction of concrete silos can be predicted from a model. These include the quantity of concrete in m3 (y), the number of crew-days of labor (z), or the number of concrete mixer hours (w) needed for a particular job. Table SE23A defines 23 potential independent variables that can be used to predict y, z, or w. Values of the dependent and independent variables, collected on 28 construction jobs, are presented in Table SE23B (page 665) and Table SE23C (page 667). Unless otherwise stated, lengths are in meters, areas in m2, and volumes in m3. TABLE SE23A Descriptions of Variables for Exercise 23 x1 Number of bins x2 Maximum required concrete per hour (m3/h) x3 Height x4 Sliding rate of the slipform (m/day) x5 Number of construction stages x6 Perimeter of slipform x7 Volume of silo complex x8 Surface area of silo walls x9 Volume of one bin x10 Wall-to-floor areas x11 Number of lifting jacks x12 Length-to-thickness ratio x13 Breadth-to-thickness ratio x14 Perimeter of complex x15 Mixer capacity x16 Density of stored material x17 Waste percent in reinforcing steel x18 Waste percent in concrete x19 Number of workers in concrete crew x20 Wall thickness (cm) x21 Number of reinforcing steel crews x22 Number of workers in forms crew x23 Length-to-breadth ratio TABLE SE23B y z w x1 x2 x3 x4 x5 x6 x7 ... Data for Exercise 23 ... TABLE SE23C ... Using best subsets regression, find the model that is best for predicting y according to the adjusted R2 criterion. Using best subsets regression, find the model that is best for predicting y according to the minimum Mallows Cp criterion. Find a model for predicting y using stepwise regression. Explain the criterion you are using to determine which variables to add to or drop from the model. Using best subsets regression, find the model that is best for predicting z according to the adjusted R2 criterion. Using best subsets regression, find the model that is best for predicting z according to the minimum Mallows Cp criterion. Find a model for predicting z using stepwise regression. Explain the criterion you are using to determine which variables to add to or drop from the model. Using best subsets regression, find the model that is best for predicting w according to the adjusted R2 criterion. Using best subsets regression, find the model that is best for predicting w according to the minimum Mallows Cp criterion. Find a model for predicting w using stepwise regression. Explain the criterion you are using to determine which variables to add to or drop from the model. Page 667 24. The article referred to in Exercise 23 presents values for the dependent and independent variables for 10 additional construction jobs. These values are presented in Tables SE24A and SE24B (page 668). a. Using the equation constructed in part (a) of Exercise 23, predict the concrete quantity (y) for each of these 10 jobs. b. Denoting the predicted values by ŷ1, ..., ŷ10 and the observed values by y1, ..., y10, compute the quantities yi − ŷi. These are the prediction errors. c. Now compute the fitted values ŷ1, ..., ŷ28 from the data in Exercise 23. Using the observed values y1, ..., y28 from those data, compute the residuals yi − ŷi. d. On the whole, which are larger, the residuals or the prediction errors? Why will this be true in general? TABLE SE24A Page 668 Data for Exercise 24 y x1 x2 x3 ... TABLE SE24B Data for Exercise 24 x12 x13 x14 x15 x16 x17 x18 x19 x20 x21 x22 x23 ... Page 669 Chapter 9 Factorial Experiments Introduction Experiments are essential to the development and improvement of engineering and scientific methods. Only through experimentation can different variants of a method be compared to see which are most effective. To be useful, an experiment must be designed properly, and the data it produces must be analyzed correctly. In this chapter we will discuss the design of and the analysis of data from a class of experiments known as factorial experiments. 9.1 One-Factor Experiments We begin with an example. The article "An Investigation of the CaCO3-CaF2-K2SiO3-SiO2-Fe Flux Systems Using the Submerged Arc Welding Process on HSLA-100 and AISI-1081 Steels" (G. Fredrickson, M.S. Thesis, Colorado School of Mines, 1992) describes an experiment in which welding fluxes with differing chemical compositions were prepared. Several welds using each flux were made on AISI-1018 steel base plates. The results of hardness measurements, on the Brinell scale, of five welds using each of four different fluxes are presented in Table 9.1. TABLE 9.1 Brinell hardness of welds using four different fluxes Sample Values 264 256 260 254 267 265 279 258 262 264 239 267 277 Sample Mean 253.8 263.2 271.0 262.0 Sample Standard Deviation 9.7570 5.4037 8.7178 7.4498 Figure 9.1 presents dotplots for the hardnesses using the four fluxes. Each sample mean is marked with an "X." It is clear that the sample means differ. In particular, the welds made using flux C have the largest sample mean and those using flux A have the lowest. Of course, there is uncertainty in the sample means, and the question is whether the sample means differ from each other by a greater amount than could be accounted for by uncertainty alone. Another way to phrase the question is this: Can we conclude that there are differences in the population means among the four flux types? FIGURE 9.1 Dotplots for each sample in Table 9.1. Each sample mean is marked with an "X." The sample means differ somewhat, but the sample values overlap considerably. This is an example of a factorial experiment. In general a factorial experiment involves several variables. One variable is the response variable, which is sometimes called the outcome variable or the dependent variable. The other variables are called factors. The question addressed by a factorial experiment is whether varying the levels of the factors produces a difference in the mean of the response variable. In the experiment described in Table 9.1, the hardness is the response, and there is one factor: flux type. Since there is only one factor, this is a one-factor experiment. There are four different values for the flux-type factor in this experiment. These different values are called the levels of the factor and can also be called treatments. Finally, the objects upon which measurements are made are called experimental units. The units assigned to a given treatment are called replicates. In the preceding experiment, the welds are the experimental units, and there are five replicates for each treatment. In this welding experiment, the four particular flux compositions were chosen deliberately by the experimenter, rather than at random from a larger population of fluxes. Such an experiment is said to follow a fixed effects model. In some experiments, treatments are chosen at random from a population of possible treatments. In this case the experiment is said to follow a random effects model. The methods of analysis for these two models are essentially the same, although the conclusions to be drawn from them differ. We will focus on fixed effects Page 671 models. Later in this section, we will discuss some of the differences between fixed and random effects models. Completely Randomized Experiments In this welding experiment, a total of 20 welds were produced, five with each of the four fluxes. Each weld was produced on a different steel base plate. Therefore, to run the experiment, the experimenter had to choose, from a total of 20 base plates, a group of 5 to be welded with flux A, another group of 5 to be welded with flux B, and so on. The best way to assign the base plates to the fluxes is at random. In this way, the experimental design will not favor any one treatment over another. For example, the experimenter could number the plates from 1 to 20, and then generate a random ordering of the integers from 1 to 20. The plates whose numbers correspond to the first five numbers on the list are assigned to flux A, and so on. This is an example of a completely randomized experiment. Definition A factorial experiment in which experimental units are assigned to treatments at random, with all possible assignments being equally likely, is called a completely randomized experiment. In many situations, the results of an experiment can be affected by the order in which the observations are taken. For example, the performance of a machine used to make measurements may change over time, due, for example, to calibration drift, or to warm-up effects. In cases like this, the ideal procedure is to take the observations in random order. This requires switching from treatment to treatment as observations are taken, rather than running all the observations that correspond to a given treatment consecutively. In some cases changing treatments involves considerable time or expense, so it is not feasible to switch back and forth. In these cases, the treatments should be run in a random order, with all the observations corresponding to the first randomly chosen treatment being run first, and so on. In a completely randomized experiment, it is appropriate to think of each treatment as representing a population, and the responses observed for the units assigned to that treatment as a simple random sample from that population. The data from the experiment thus consist of several random samples, each from a different population. The population means are called treatment means. The questions of interest concern the treatment means—whether they are all equal, and if not, which ones are different, how big the differences are, and so on. One-Way Analysis of Variance To make a formal determination as to whether the treatment means differ, a hypothesis test is needed. We begin by introducing the notation. We have I samples, each from a different treatment. The treatment means are denoted It is not necessary that the sample sizes be equal, although it is desirable, as we will Page 672 discuss later in this section. The sample sizes are denoted The total number in all the samples combined is denoted by N. The hypotheses we wish to test are H0 : μ1 = ⋯ = μI versus H1: two or more of the μi are different If there were only two samples, we could use the two-sample t test (Section 6.7) to test the null hypothesis. Since there are more than two samples, we use a method known as one-way analysis of variance (ANOVA). To define the test statistic for one-way ANOVA, we first develop the notation for the sample observations. Since there are several samples, we use a double subscript to denote the observations. Specifically, we let Xij denote the jth observation in the ith sample. The sample mean of the ith sample is denoted . (9.1) The sample grand mean, denoted , is the average of all the sampled items taken together: (9.2) With a little algebra, it can be shown that the sample grand mean is also a weighted average of the sample means: (9.3) Example 9.1 For the data in Table 9.1, find I, J1, ... , JI, N, X23, , . Solution There are four samples, so I = 4. Each sample contains five observations, so J1 = J2 = J3 = J4 = 5. The total number of observations is N = 20. The quantity X23 is the third observation in the second sample, which is 267. The quantity This value is is the sample mean of the third sample. . Finally, we use Equation (9.3) to compute the sample grand mean . Figure 9.2 presents the idea behind one-way ANOVA. The figure illustrates Page 673 several hypothetical samples from different treatments, along with their sample means and the sample grand mean. The sample means are spread out around the sample grand mean. One-way ANOVA provides a way to measure this spread. If the sample means are highly spread out, then it is likely that the treatment means are different, and we will reject H0. FIGURE 9.2 The variation of the sample means around the sample grand mean can be due both to random uncertainty and to differences among the treatment means. The variation within a given sample around its own sample mean is due only to random uncertainty. The variation of the sample means around the sample grand mean is measured by a quantity called the treatment sum of squares (SSTr for short), which is given by (9.4) Each term in SSTr involves the distance from the sample means to the sample grand mean. Note that each squared distance is multiplied by the sample size corresponding to its sample mean, so that the means for the larger samples count more. SSTr provides an indication of how different the treatment means are from each other. If SSTr is large, then the sample means are spread out widely, and it is reasonable to conclude that the treatment means differ and to reject H0. If on the other hand SSTr is small, then the sample means are all close to the sample grand mean and therefore to each other, so it is plausible that the treatment means are equal. An equivalent formula for SSTr, which is a bit easier to compute by hand, is (9.5) In order to determine whether SSTr is large enough to reject H0, we compare it to another sum of squares, called the error sum of squares (SSE for short). SSE measures the variation in the individual sample points around their respective sample means. This variation is measured by summing the squares of the distances from each point to its own sample mean. SSE is given by (9.6) The quantities are called the residuals, so SSE is the sum of the squared Page 674 residuals. SSE, unlike SSTr, depends only on the distances of the sample points from their own means and is not affected by the location of treatment means relative to one another. SSE therefore measures only the underlying random variation in the process being studied. It is analogous to the error sum of squares in regression. An equivalent formula for SSE, which is a bit easier to compute by hand, is (9.7) Another equivalent formula for SSE is based on the sample variances. Let denote the sample variance of the ith sample. Then (9.8) It follows from Equation (9.8) that . Substituting into Equation (9.6) yields (9.9) Example 9.2 For the data in Table 9.1, compute SSTr and SSE. Solution The sample means are presented in Table 9.1. They are The sample grand mean was computed in Example 9.1 to be Equation (9.4) to calculate SSTr. . We now use To compute SSE we will use Equation (9.9), since the sample standard deviations si have already been presented in Table 9.1. We can use SSTr and SSE to construct a test statistic, provided the following two assumptions are met. Page 675 Assumptions for One-Way ANOVA The standard one-way ANOVA hypothesis tests are valid under the following conditions: 1. 2. The treatment populations must be normal. The treatment populations must all have the same variance, which we will denote by σ2. Before presenting the test statistic, we will explain how it works. If the two assumptions for one-way ANOVA are approximately met, we can compute the means of SSE and SSTr. The mean of SSTr depends on whether H0 is true, because SSTr tends to be smaller when H0 is true and larger when H0 is false. The mean of SSTr satisfies the condition (9.10) (9.11) The likely size of SSE, and thus its mean, does not depend on whether H0 is true. The mean of SSE is given by (9.12) Derivations of Equations (9.10) and (9.12) are given at the end of this section. The quantities I − 1 and N − I are the degrees of freedom for SSTr and SSE, respectively. When a sum of squares is divided by its degrees of freedom, the quantity obtained is called a mean square. The treatment mean square is denoted MSTr, and the error mean square is denoted MSE. They are defined by (9.13) It follows from Equations (9.10) through (9.13) that (9.14) (9.15) (9.16) Equations (9.14) and (9.16) show that when H0 is true, MSTr and MSE have the same mean. Therefore, when H0 is true, we would expect their quotient to be near 1. This quotient is in fact the test statistic. The test statistic for testing H0 : μ1 = ⋯ = μI is (9.17) When H0 is true, the numerator and denominator of F are on average the same size, so F tends to be near 1. In fact, when H0 is true, this test statistic has an F distribution with I − 1 and N − I degrees of freedom, denoted FI−1,N−I. When H0 is false, MSTr tends to be larger, but MSE does not, so F tends to be greater than 1. Page 676 Summary The F test for One-Way ANOVA To test H0: μ1 = ⋯ = μI versus H1 : two or more of the μi are different: 1. Compute . 2. . 3. 4. 5. Compute Compute the test statistic: and . . Find the P-value by consulting the F table (Table A.8 in Appendix A) with I − 1 and N − I degrees of freedom. We now apply the method of analysis of variance to the example with which we introduced this section. Example 9.3 For the data in Table 9.1, compute MSTr, MSE, and F. Find the P-value for testing the null hypothesis that all the means are equal. What do you conclude? Solution From Example 9.2, SSTr = 743.4 and SSE = 1023.6. We have I = 4 samples and N = 20 observations in all the samples taken together. Using Equation (9.13), The value of the test statistic F is therefore To find the P-value, we consult the F table (Table A.8). The degrees of freedom are 4 − 1 = 3 for the numerator and 20 − 4 = 16 for the denominator. Under H0, F has an F3,16 distribution. Looking at the F table under 3 and 16 degrees of freedom, we find that the upper 5% point is 3.24 and the upper 1% point is 5.29. Therefore the P-value is between 0.01 and 0.05 (see Figure 9.3); a computer software package gives a value of 0.029 accurate to two significant digits). It is reasonable to conclude that the population means are not all equal, and, thus, that flux composition does affect hardness. Page 677 FIGURE 9.3 The observed value of the test statistic is 3.87. The upper 5% point of the F3,16 distribution is 3.24. The upper 1% point of the F3,16 distribution is 5.29. Therefore the P-value is between 0.01 and 0.05. A computer software package gives a value of 0.029. Confidence Intervals for the Treatment Means The observations on the ith treatment are assumed to be a simple random sample from a normal population with mean μi and variance σ2. To construct a confidence interval for μi, the first step is to estimate the population variance σ2. One way to do this would be to use the sample variance of the observations on the ith treatment. However, since we assume that all observations for all treatments have the same variance, it is better to combine all the sample variances into one "pooled" estimate. To do this, note that SSE is a weighted sum of the sample variances (Equation 9.9) and MSE is the weighted average (Equation 9.13). The quantity MSE is therefore the pooled estimate of the variance σ2. Since is the sample mean of Ji observations, the variance of is σ2/Ji, estimated with MSE/Ji. The number of degrees of freedom for MSE is N − I. The quantity has a Student's t distribution with N − I degrees of freedom. A confidence interval for μi can therefore be constructed by the method described in Section 5.3. A level 100(1 − α)% confidence interval for μi is given by (9.18) Example 9.4 Find a 95% confidence interval for the mean hardness of welds produced with flux A. Solution From Table 9.1, . The value of MSE was computed in Example 9.3 to be 63.975. There are I = 4 treatments, J1 = 5 observations for flux A, and N = 20 observations altogether. From the Student's t table we obtain t16, .025 = 2.120. The 95% confidence interval is therefore Page 678 The ANOVA Table The results of an analysis of variance are usually summarized in an analysis of variance (ANOVA) table. This table is much like the analysis of variance table produced in multiple regression. The following output (from MINITAB) shows the analysis of variance for the weld data presented in Table 9.1. In the ANOVA table, the column labeled "DF" presents the number of degrees of freedom for both the treatment ("Factor") and error ("Error") sum of squares. The column labeled "SS" presents SSTr (in the row labeled "Factor") and SSE (in the row labeled "Error"). The row labeled "Total" contains the total sum of squares, which is the sum of SSTr and SSE. The column labeled "MS" presents the mean squares MSTr and MSE. The column labeled "F" presents the F statistic for testing the null hypothesis that all the population means are equal. Finally, the column labeled "P" presents the P-value for the F test. Below the ANOVA table, the value "S" is the pooled estimate of the error standard deviation σ, computed by taking the square root of MSE. The quantity "R-sq" is R2, the coefficient of determination, which is equal to the quotient SSTr/SST. This is analogous to the multiple regression situation (see Equation 8.9 in Section 8.1). The value "R-Sq(adj)" is the adjusted R2, equal to R2 − [(I − 1)/(N − I)](1 Page 679 − R2), again analogous to multiple regression. The quantities R2 and adjusted R2 are not used as much in analysis of variance as they are in multiple regression. Finally, sample means and standard deviations are presented for each treatment group, along with a graphic that illustrates a 95% confidence interval for each treatment mean. Example 9.5 In the article "Review of Development and Application of CRSTER and MPTER Models" (R. Wilson, Atmospheric Environment, 1993:41–57), several measurements of the maximum hourly concentrations (in μg/m3) of SO2 are presented for each of four power plants. The results are as follows (two outliers have been deleted): Plant 1: Plant 2: Plant 3: Plant 4: 438 857 925 893 619 732 638 1014 1153 883 1053 786 1179 786 891 917 695 675 595 The following output (from MINITAB) presents results for a one-way ANOVA. Can you conclude that the maximum hourly concentrations differ among the plants? Solution Page 680 In the ANOVA table, the P-value for the null hypothesis that all treatment means are equal is 0.006. Therefore we conclude that not all the treatment means are equal. Checking the Assumptions As previously mentioned, the methods of analysis of variance require the assumptions that the observations on each treatment are a sample from a normal population and that the normal populations all have the same variance. A good way to check the normality assumption is with a normal probability plot. If the sample sizes are large enough, one can construct a separate probability plot for each sample. This is rarely the case in practice. When the sample sizes are not large enough for individual probability plots to be informative, the residuals can all be plotted together in a single plot. When the assumptions of normality and constant variance are satisfied, these residuals will be normally distributed with mean zero and should plot approximately on a straight line. Figure 9.4 presents a normal probability plot of the residuals for the weld data of Table 9.1. There is no evidence of a serious violation of the assumption of normality. FIGURE 9.4 Probability plot for the residuals from the weld data. There is no evidence of a serious violation of the assumption of normality. The assumption of equal variances can be difficult to check, because with only a few observations in each sample, the sample standard deviations can differ greatly (by a factor of 2 or more) even when the assumption holds. For the weld data, the sample standard deviations range from 5.4037 to 9.7570. It is reasonable to proceed as though the variances were equal. The spreads of the observations within the various samples can be checked visually by making a residual plot. This is done by plotting the residuals versus the fitted values, which are the sample means . If the spreads differ considerably among the samples, Page 681 the assumption of equal variances is suspect. If one or more of the samples contain outliers, the assumption of normality is suspect as well. Figure 9.5 presents a residual plot for the weld data. There are no serious outliers, and the spreads do not differ greatly among samples. FIGURE 9.5 Residual plot of the values versus the fitted values for the weld data. The spreads do not differ greatly from sample to sample, and there are no serious outliers. Balanced versus Unbalanced Designs When equal numbers of units are assigned to each treatment, the design is said to be balanced. Although one-way analysis of variance can be used with both balanced and unbalanced designs, balanced designs offer a big advantage. A balanced design is much less sensitive to violations of the assumption of equality of variance than an unbalanced one. Since moderate departures from this assumption can be difficult to detect, it is best to use a balanced design whenever possible, so that undetected violations of the assumption will not seriously compromise the validity of the results. When a balanced design is impossible to achieve, a slightly unbalanced design is preferable to a severely unbalanced one. Summary ■ ■ ■ With a balanced design, the effect of unequal variances is generally not great. With an unbalanced design, the effect of unequal variances can be substantial. The more unbalanced the design, the greater the effect of unequal variances. The Analysis of Variance Identity In both linear regression and analysis of variance, a quantity called the total sum of Page 682 squares is obtained by subtracting the sample grand mean from each observation, squaring these deviations, and then summing them. An analysis of variance identity is an equation that expresses the total sum of squares as a sum of other sums of squares. We have presented analysis of variance identities for simple linear regression (at the end of Section 7.2) and for multiple regression (Equation 8.7 in Section 8.1). The total sum of squares for one-way ANOVA is given by (9.19) An equivalent formula is given by (9.20) Examining Equations (9.5), (9.7), and (9.20) shows that the total sum of squares is equal to the treatment sum of squares plus the error sum of squares. This is the analysis of variance identity for one-way analysis of variance. The Analysis of Variance Identity (9.21) An Alternate Parameterization Our presentation of one-way analysis of variance, as a method to compare several treatment means by using random samples drawn from each treatment population, is one natural way to view the subject. There is another way to express these same ideas, in somewhat different notation, that is sometimes useful. For each observation Xij, define εij = Xij − μi, the difference between the observation and its mean. By analogy with linear regression, the quantities εij are called errors. It is clearly true that (9.22) Now since Xij is normally distributed with mean μi and variance σ2, it follows that is normally distributed with mean 0 and variance σ2. In a single-factor experiment, we are interested in determining whether the treatment means are all equal. Given treatment means μ1, ... , μI, the quantity (9.23) is the average of all the treatment means. The quantity μ is called the population Page 683 grand mean. The ith treatment effect, denoted αi, is the difference between the ith treatment mean and the population grand mean: (9.24) It follows from the definition of αi that . We can now decompose the treatment means as follows: (9.25) Combining Equations (9.22) and (9.25) yields the one-way analysis of variance model: (9.26) The null hypothesis H0 : μ1 = ⋯ = μI is equivalent to H0 : α1 = ⋯ = αI = 0. In one-way ANOVA, it is possible to work with the treatment means μi, as we have done, rather than with the treatment effects αi. In multi-factor experiments, however, the treatment means by themselves are not sufficient and must be decomposed in a manner analogous to the one described here. We will discuss this further in Section 9.3. Power When designing a factorial experiment, it is important that the F test have good power, that is, a large probability of rejecting the null hypothesis of equality if in fact the treatment means are not all equal. An experiment with low power is not of much use, since it is unlikely to detect a difference in treatments even if one exists. In what follows, we will assume that the experiment is balanced and that the assumptions of normality and equal variance hold. Assume the number of levels is fixed at I. The power of any test depends first on the rejection criterion: The larger the level at which one is willing to reject, the greater the power. The 5% level is the one most often used in practice. Once the rejection level is set, the power of the F test depends on three quantities: (1) the spread in the true means as measured by the quantity where αi is the ith treatment effect, (2) the error standard deviation σ, and (3) the sample size J. Note that if the null hypothesis is true, then . The larger is, the farther from the truth is the null hypothesis, and the larger is the power, which is the probability that the null hypothesis is rejected. A power calculation can serve either of two purposes: to determine the sample size for each treatment necessary to achieve a desired power, or to determine how much power one has with a given sample size. In a traditional power calculation, one specifies the quantity that one wishes to detect and the value of σ one expects to encounter. Then one can compute the power for a given sample size, or the sample size needed to achieve a given power. In practice, one rarely knows how to specify a value for , but one can often specify the size of a difference between the largest and smallest treatment means that one wishes to detect. For example, in the weld experiment, a metallurgist might be able to specify that a difference of 10 or more between the largest and smallest treatment means is scientifically important, but it is unlikely that she could specify a scientifically important value for . In MINITAB, one can specify the size of a scientifically important difference Page 684 between the largest and smallest treatment means and compute the sample size necessary to guarantee that the power to detect that difference will be at least a specified amount. We present an example. Example 9.6 A metallurgist wants to repeat the weld experiment with four different fluxes and wants the design to be sensitive enough so that it is likely to detect a difference of 10 or more in Brinell hardness at the 5% level. He assumes that the error standard deviation will be about the same as the value of 7.998 calculated in the experiment we have been discussing. The following output (from MINITAB) shows the result of a power calculation for an experiment with five observations per treatment. What is the power? What recommendation would you give the metallurgist regarding the usefulness of this proposed experiment? One-way ANOVA Alpha = 0.05 Assumed standard deviation = 7.998 Number of Levels = 4 SS Means 50 Sample Size 5 Power 0.281722 Maximum Difference 10 The sample size is for each level. Solution The power is 0.281772. This means that the probability that the proposed experiment will detect a difference of 10 between the largest and smallest treatment means may be no more than 0.28. The appropriate recommendation is not to run this experiment; it has too little chance of success. Instead, the sample size necessary to provide adequate power should be calculated, and if feasible, an experiment of that size should be run. Example 9.7 The metallurgist in Example 9.6 has taken your advice and has computed the sample size necessary to provide a power of 0.90 to detect a difference of 10 at the 5% level. The results (from MINITAB) follow. What is the power? How many observations will be necessary at each level? How many observations will be necessary in total? One-way ANOVA Page 685 Alpha = 0.05 Assumed standard deviation = 7.998 Number of Levels = 4 SS Means 50 Sample Size 20 Target Power 0.9 Actual Power 0.914048 Maximum Difference 10 The sample size is for each level. Solution The needed sample size is 20 per level; with four levels there will be 80 observations in total. Note that the actual power of the experiment is approximately 0.914, which is higher than the "target power" of 0.90 that was requested. The reason for this is that the power provided by a sample size of 19 per level could be somewhat less than 0.90; a sample size of 20 is the smallest that is guaranteed to provide a power of 0.90 or more. Random Effects Models In many factorial experiments, the treatments are chosen deliberately by the experiment. These experiments are said to follow a fixed effects model. In some cases, the treatments are chosen at random from a population of possible treatments. In these cases the experiments are said to follow a random effects model. In a fixed effects model, the interest is on the specific treatments chosen for the experiment. In a random effects model, the interest is in the whole population of possible treatments, and there is no particular interest in the ones that happened to be chosen for the experiment. The article describing the weld experiment states that the treatments were chosen deliberately and do not represent a random sample from a larger population of flux compositions. This experiment therefore follows a fixed effects model. The four power plants in Example 9.5 are a sample of convenience; they are plants at which measurements were readily available. In some cases it is appropriate to treat a sample of convenience as if it were a simple random sample (see the discussion in Section 1.1). If these conditions hold, then the power plant experiment may be considered to follow a random effects model; otherwise it must be treated as a fixed effects model. There is an important difference in interpretation between the results of a fixed effects model and those of a random effects model. In a fixed effects model, the only conclusions that can be drawn are conclusions about the treatments actually used in the experiment. In a random effects model, however, since the treatments are a simple random sample from a population of treatments, conclusions can be drawn concerning the whole population, including treatments not actually used in the experiment. This difference in interpretations results in a difference in the null hypotheses to be tested. In the fixed effects model, the null hypothesis of interest is H0 : μ1 = ⋯ = μI. In the random effects model, the null hypothesis of interest is H0 : the treatment means are equal for every treatment in the population. In the random effects model, the assumption is made that the population of treatment means is normal. Interestingly enough, although the null hypothesis for the random effects model differs from that of the fixed effects model, the hypothesis test is exactly the same. The F test previously described is used for the random effects model as well as for the fixed effects model. Example 9.8 In Example 9.5, assume that it is reasonable to treat the four power plants as a random sample from a large population of power plants, and furthermore, assume that the SO2 concentrations in the population of plants are normally distributed. Can we conclude that there are differences in SO2 concentrations among the power plants in the population? Solution This is a random effects model, so we can use the F test to test the null hypothesis that all the treatment means in the population are the same. The results of the F test are shown in Example 9.5. The P-value is 0.006. We therefore reject the null hypothesis and conclude that there are differences in mean SO2 concentrations among the power plants in the population. Derivations of Equations (9.10) and (9.12) In what follows it will be easier to use the notation E(⋅) to denote the mean of a quantity and V(⋅) to denote the variance. So, for example, E(SSE) = μSSE, E(SSTr) = μSSTr, and V(Xij) denotes the variance of Xij. We will show that , whether or not the population means are equal. This is Equation (9.12). We begin by adding and subtracting the treatment mean μi from each term in to obtain Multiplying out, we obtain Now . Substituting into the middle term of the right-hand side of (9.27) yields Since , this simplifies to Page 687 (9.28) Taking means of both sides of (9.28) yields (9.29) Now . The population variances are all equal; denote their common value by σ2. It follows that Substituting into (9.29) yields This completes the derivation of E(SSE). We now show that under the assumption that the treatment means are all equal to a common value denoted by μ. This is Equation (9.10). We begin by adding and subtracting the common treatment mean μ from each term in to obtain Multiplying out, we obtain (9.30) Now so and Substituting into the right-hand side of (9.30), we obtain Page 688 Since , we obtain Taking means of both sides yields (9.31) Now Substituting into (9.31) yields Exercises for Section 9.1 1. A study is made of the effect of curing temperature on the compressive strength of a certain type of concrete. Five concrete specimens are cured at each of four temperatures, and the compressive strength of each specimen is measured (in MPa). The results are as follows: Temperature (°C) 0 10 20 30 Strengths 31.2 29.6 30.8 30.0 31.4 30.1 29.7 31.1 35.8 35.0 37.0 37.5 33.6 39.4 38.3 36.4 Strengths 29.6 30.8 27.3 31.1 36.8 35.6 ... a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that the mean strength differs with temperature? 2. The article "Nutrient Deprivation Improves Field Performance of Woody Seedlings in a Degraded Semi-arid Shrubland" (R. Trubata, J. Cortina, and A. Vilagrosa, Ecological Engineering, 2011:1164–1173) presents a study that investigated the effect of varying the type of fertilizer on the height of certain Mediterranean woody tree species. In one experiment, three samples, each consisting of ten trees, were grown with three different fertilizers. One, the control group, was grown with a standard fertilizer. Another was grown with a fertilizer containing only half the nutrients of the standard fertilizer. The third was grown with the standard fertilizer to which a commercial slow-release fertilizer had been added. Following are the heights of the trees after one year. These data are consistent with the means and standard deviations reported in the article. Fertilizer Height Control 17.9 12.2 14.9 13.8 26.1 15.4 20.3 16.9 20.8 14.8 Deficient 7.0 6.9 13.3 11.1 11.0 16.5 12.7 12.4 17.1 9.0 Slow-release 19.8 20.3 16.1 17.9 12.4 12.5 17.4 19.9 27.3 14.4 release 3. a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that the heights differ among the types of fertilizer? The removal of ammoniacal nitrogen is an important aspect of treatment of leachate at landfill sites. The rate of removal (in percent per day) is recorded for several days for each of several treatment methods. The results are presented in the following table. (Based on the article "Removal of Ammoniacal Nitrogen from Landfill Leachate by Irrigation onto Vegetated Treatment Planes," S. Tyrrel, P. Leeds-Harrison, and K. Harrison, Water Research, 2002:291–299.) Treatment A B C D E 4. 5.21 5.59 6.24 6.85 4.04 Rate of Removal 4.65 2.69 7.57 5.94 6.41 9.18 4.94 3.29 4.52 5.16 3.75 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that the treatment methods differ in their rates of removal? The antibiotic gentamicin sulphate is often blended with acrylic bone cement to help prevent infection following joint replacement surgery. The article "Incorporation of Large Amounts of Gentamicin Sulphate Into Acrylic Bone Cement: Effect on Handling and Mechanical Properties, Antibiotic Release, and Biofilm Formation" (N. Dunne, P. McAfee, et al., Journal of Engineering in Medicine, 2008:355–365) presents a study of the effect of the amount of antibiotic added on various properties of the cement. Following are measurements of the setting time of the cement, for six levels of antibiotic amount, and three replications per level. The measurements are consistent with means and standard deviations presented in the article. Antibiotic per 40 g Cement Setting Time (min) 0 g 12.7 14.1 13.2 0.5 g 13.5 14.5 14.6 1g 12.7 13.4 13.2 2g 12.7 13.6 14.1 3g 13.4 13.5 14.3 4g 14.5 13.5 14.9 5. a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that there are differences among the mean setting times? The article "Influence of Age on Masonry Bond Strength and Mortar Microstructure" (H. Sugo, A. Page, and S. Lawrence, Can J Civ Eng, 2007:1433–442) investigates the effect of age on tensile strength of mortar. Several specimens of various ages were loaded until failure, and the maximum load (in MPa) was recorded for each. The results are presented in the following table. Page 690 Age (days) Maximum load (MPa) 3 1.69, 1.69, 1.97, 2.09, 1.81, 1.53, 1.63, 1.70, 1.73, 1.72, 1.48, 1.15 7 1.82, 1.86, 1.72, 1.73, 1.70, 1.44, 2.00, 1.78, 1.47, 1.32, 1.87, 1.57 28 2.76, 2.60, 2.38, 2.06, 1.81, 2.76, 2.41, 2.29, 2.00, 2.15, 0.97, 1.91, 2.26 90 1.18, 1.46, 2.02, 2.16, 1.79, 1.74, 2.08, 1.99, 1.63, 1.95, 1.66, 2.34 180 2.60, 2.28, 2.42, 2.26, 2.53, 1.66, 2.33, 2.02, 2.28, 2.18, 2.27, 2.24, 1.81, 1.93 365 2.16, 1.51, 2.44, 2.13, 2.01, 2.09, 2.18, 2.48, 1.99, 1.75, 2.14, 1.56, 1.94, 1.75 6. a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that there are differences among the mean maximum loads? Archaeologists can determine the diets of ancient civilizations by measuring the ratio of carbon-13 to carbon-12 in bones found at burial sites. Large amounts of carbon-13 suggest a diet rich in grasses such as maize, while small amounts suggest a diet based on herbaceous plants. The article "Climate and Diet in Fremont Prehistory: Economic Variability and Abandonment of Maize Agriculture in the Great Salt Lake Basin" (J. Coltrain and S. Leavitt, American Antiquity, 2002:453–485) reports ratios, as a difference from a standard in units of parts per thousand, for bones from individuals in several age groups. The data are presented in the following table. Ratio Age Group (years) 0–11 17.2 18.4 17.9 12–24 14.8 17.6 18.3 25–45 18.4 13.0 14.8 19.2 15.4 13.2 46+ 15.5 18.2 12.7 7. 15.1 18.2 10.4 16.7 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that the concentration ratios differ among the age groups? The article "Secretion of Parathyroid Hormone Oscillates Depending on the Change in Serum Ionized Calcium During Hemodialysis and May Affect Bone Metabolism" (T. Kitahara, K. Ueki, et al., Nephron Clinical Practice, 2005: c9–c17) presents measurements of basal ionized calcium (Ca) levels for four groups of patients with differing levels of basal intact parathyroid hormone (PTH). The following results are consistent with means and standard deviations presented in the article. Group Ca levels (mM) I 1.23 1.02 1.33 1.49 II 1.18 1.16 1.07 1.27 1.26 1.33 III 1.04 1.32 1.29 IV 1.35 1.67 1.38 8. 16.6 19.0 18.3 13.6 13.5 15.8 19.1 19.1 13.4 17.2 10.0 11.3 10.2 17.0 18.9 19.2 18.4 12.8 17.6 18.8 17.9 18.5 17.5 18.3 15.2 10.8 19.8 17.3 13.4 1.27 1.31 1.26 1.36 1.11 1.51 1.51 1.36 1.34 1.27 1.43 1.21 1.29 1.26 0.95 1.05 1.01 1.26 1.37 1.09 1.28 1.33 0.98 0.99 1.24 1.12 1.32 1.38 1.08 1.65 1.14 1.44 1.37 1.11 0.82 1.31 1.09 1.29 1.32 0.95 1.21 1.22 1.13 1.51 0.72 1.33 1.46 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that there are differences among the mean Ca levels? The article "Impact of Free Calcium Oxide Content of Fly Ash on Dust and Sulfur Dioxide Emissions in a Lignite-Fired Power Plant" (D. Sotiropoulos, A. Georgakopoulos, and N. Kolovos, Journal of Air and Waste Management, 2005:1042–1049) presents measurements of dust emissions, in mg/m3, for four power plants. Thirty measurements were taken for each plant. The sample means and standard deviations are presented in the following table: Plant 1 Plant 2 Plant 3 Plant 4 a. b. 9. Mean 211.50 214.00 211.75 236.08 Standard Deviation 24.85 35.26 33.53 23.09 Sample Size 30 30 30 30 Construct an ANOVA table. You may give a range for the P-value. Can you conclude that there are differences among the mean emission levels? The article "Experimental and Statistical Investigation of Self-Consolidating Concrete Mixture Constituents for Pre-stressed Bridge Girder Fabrication" (E. Torres, J. Seo, and R. Lederle, Journal of Materials in Civil Engineering 2017, online) presents measurements of compressive strength (in MPa) for concrete manufactured in three different plants. The results are presented in the following table. Plant Compressive Strength A 44.42 41.00 48.45 46.58 44.71 47.98 49.20 58.14 48.60 48.22 49.94 B 36.00 38.06 42.66 48.25 49.05 47.98 40.81 46.11 49.29 40.47 C 57.28 61.58 58.72 62.58 46.38 46.98 49.17 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that the mean emissions differ with mixing temperature? 10. An experiment to compare the lifetimes of four different brands of spark plug was carried out. Five plugs of each brand were used, and the number of miles until failure was recorded for each. Following is part of the MINITAB output for a one-way ANOVA. One-way Analysis of Variance Analysis of Variance Source DF Brand 3 Error (b) Total 19 SS 176.482 (c) 235.958 MS (a) (d) F (e) P (f) Fill in the missing numbers for (a) through (f) in the table. You may give a range for (f). 11. Refer to Exercise 10. It is plausible that the brands of spark plug all have the same mean lifetime? Page 692 12. The article "Characterization of Effects of Thermal Property of Aggregate on the Carbon Footprint of Asphalt Industries in China" (A. Jamshidi, K. Kurumisawa, et al., Journal of Traffic and Transportation Engineering 2017: 118–130) presents the results measurements of CO2 emissions produced during the manufacture of asphalt. Three measurements were taken at each of three mixing temperatures. The results are presented in the following table. Temperature (°C) 160 145 130 Emissions (100kt) 9.52 14.10 7.43 8.37 6.54 12.40 7.24 10.70 5.66 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that the mean emissions differ with mixing temperature? 13. An experiment was performed to determine whether the annealing temperature of ductile iron affects its tensile strength. Five specimens were annealed at each of four temperatures. The tensile strength (in ksi) was measured for each. The results are presented in the following table. Temperature (°C) 750 800 850 900 19.72 16.01 16.66 16.93 Sample Values 20.88 19.63 18.68 20.04 18.10 20.28 17.38 14.49 16.21 14.49 16.15 15.53 17.89 20.53 15.58 13.25 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that there are differences among the mean strengths? 14. Refer to Exercise 12. a. Compute the quantity , the estimate of the error standard deviation σ. b. Assuming s to be the error standard deviation, find the sample size necessary in each treatment to provide a power of 0.90 to detect a maximum difference of 5 in the treatment means at the 5% level. c. Using a more conservative estimate of 1.5s as the error standard deviation, find the sample size necessary in each treatment to provide a power of 0.90 to detect a maximum difference of 5 in the treatment means at the 5% level. 15. Refer to Exercise 13. a. Compute the quantity , the estimate of the error standard deviation σ. b. Assuming s to be the error standard deviation, find the sample size necessary in each treatment to provide a power of 0.90 to detect a maximum difference of 2 in the treatment means at the 5% level. c. Using a more conservative estimate of 1.5s as the error standard deviation, find the sample size necessary in each treatment to provide a power of 0.90 to detect a maximum difference of 2 in the treatment means at the 5% level. 16. The article "The Lubrication of Metal-on-Metal Total Hip Joints: A Slide Down the Stribeck Curve" (S. Smith, D. Dowson, and A. Goldsmith, Proceedings of the Institution of Mechanical Engineers, 2001:483–493) presents results from wear tests done on Page 693 metal artificial hip joints. Joints with several different diameters were tested. The data presented in the following table on head roughness are consistent with the means and standard deviations reported in the article. Diameter (mm) 16 28 36 Head Roughness (nm) 0.83 2.25 0.20 2.78 3.93 2.72 2.48 3.80 5.99 5.32 4.59 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that mean roughness varies with diameter? Explain. 17. The article "Multi-objective Scheduling Problems: Determination of Pruned Pareto Sets" (H. Taboada and D. Coit, IIE Transactions, 2008:552–564), presents examples in a discussion of optimization methods for industrial scheduling and production planning. In one example, seven different jobs were performed on each of five machines. The means and standard deviations of the processing times for each machine are presented in the following table. Machine A B C D E Mean 25.43 23.71 44.57 23.14 58.00 SD 10.67 13.92 15.90 12.75 19.11 Sample Size 7 7 7 7 7 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that there are differences among the mean processing times? 18. The article "Withdrawal Strength of Threaded Nails" (D. Rammer, S. Winistorfer, and D. Bender, Journal of Structural Engineering, 2001:442–449) describes an experiment comparing the withdrawal strengths for several types of nails. The data presented in the following table are consistent with means and standard deviations reported in the article for three types of nails: annularly threaded, helically threaded, and smooth shank. All nails had diameters within 0.1 cm of each other, and all were driven into the same type of lumber. Nail Type Withdrawal Strength (N/mm) Annularly 36.57 29.67 43.38 26.94 12.03 21.66 41.79 31.50 35.84 40.81 Threaded Helically 14.66 24.22 23.83 21.80 27.22 38.25 28.15 36.35 23.89 28.44 threaded Smooth 12.61 25.71 17.69 24.69 26.48 19.35 28.60 42.17 25.11 19.98 shank Construct an ANOVA table. You may give a range for the P-value. Can you conclude that the mean withdrawal strength is different for different nail types? 19. The article "Solid-Phase Chemical Fractionation of Selected Trace Metals in Some Northern Kentucky Soils" (A. Karathanasis and J. Pils, Soil and Sediment Contamination, 2005:293–308) presents pH measurements of soil specimens taken from three different types of Page 694 soils. The results in the following table are consistent with means and standard deviations reported in the article. a. b. Soil Type Alluvium Glacial Till Residuum pH Measurements 6.53, 6.03, 6.75, 6.82, 6.24 6.07, 6.07, 5.36, 5.57, 5.48, 5.27, 5.80, 5.03, 6.65 6.03, 6.16, 6.63, 6.13, 6.05, 5.68, 6.25, 5.43, 6.46, 6.91, 5.75, 6.53 a. Construct an ANOVA table. You may give a range for the P-value. b. Can you conclude that there are differences among the mean pH levels? 20. The following MINITAB output presents a power calculation. Alpha = 0.05 SS Means 20000 Assumed standard deviation = 142.6 Sample Size 14 Target Power 0.85 Actual Power 0.864138 Number of Levels = 4 Maximum Difference 200 The sample size is for each level. a. b. c. d. What is the power requested by the experimenter? To guarantee a power of 0.864138, how many observations must be taken at each level? What is the difference between means that can be detected with a power of at least 0.864138? Is this difference to detect a maximum difference of 250 greater than 0.864138 or less than 0.864138? Explain. 9.2 Pairwise Comparisons Experiments in One-Factor In a one-way ANOVA, an F test is used to test the null hypothesis that all the treatment means are equal. If this hypothesis is rejected, we can conclude that the treatment means are not all the same. But the test does not tell us which ones are different from the rest. Sometimes an experimenter has in mind two specific treatments, i and j, and wants to study the difference μi − μj. In this case a method known as Fisher's least significant difference (LSD) method is appropriate and can be used to construct confidence intervals for μi − μj or to test the null hypothesis that μi − μj = 0. At other times, an experimenter may want to determine all the pairs of means that can be concluded to differ from each other. In this case a type of procedure called a multiple comparisons method must be used. We will discuss two methods of multiple comparisons, the Bonferroni method and the Tukey–Kramer method. Page 695 Fisher's Least Significant Difference (LSD) Method We begin by describing Fisher's LSD method for constructing confidence intervals. The confidence interval for the difference μi − μj is centered at the difference in sample means . To determine how wide to make the confidence interval, it is necessary to estimate the standard deviation of . Let Ji and Jj be the sample sizes at levels i and j, respectively. Since by assumption all observations are normally distributed with variance σ2, it follows that is normally distributed with mean μi − μj and variance σ2(1/Ji+1/Jj). The variance σ2 is estimated with MSE, for reasons explained previously in the discussion about confidence intervals for the treatment means (Section 9.1). Now the quantity has a Student's t distribution with N − I degrees of freedom. (The value N − I is the number of degrees of freedom used in computing MSE; see Equation 9.13.) The quantity is called the least significant difference. This quantity forms the basis for confidence intervals and hypothesis tests. Fisher's Least Significant Difference Method for Confidence Intervals and Hypothesis Tests The Fisher's least significant difference confidence interval, at level 100(1 − α)%, for the difference μi − μj is (9.32) To test the null hypothesis H0 : μi − μj = 0, the test statistic is (9.33) If H0 is true, this statistic has a Student's t distribution with N − I degrees of freedom. Specifically, if (9.34) then H0 is rejected at level α. The reason that the quantity is called the least significant difference is that the null hypothesis of equal means is rejected at level α whenever the difference in sample means exceeds this value. When the design is balanced, with all sample sizes equal to J, the least significant difference is equal to for all pairs of means. Example 9.9 In the weld experiment discussed in Section 9.1, hardness measurements were made for five welds from each of four fluxes A, B, C, and D. The sample mean hardness values were , , , and . The following output (from MINITAB) presents the ANOVA table. One-way ANOVA: A, B, C, D Source Factor Error Total S = 7.998 DF 3 16 19 SS 743.40 1023.60 1767.00 R-Sq = 42.07% MS 247.800 63.975 F 3.87 P 0.029 R-Sq(adj) = 31.21% Before the experiment was performed, the carbon contents of the fluxes were measured. Flux B had the lowest carbon content (2.67% by weight), and flux C had the highest (5.05% by weight). The experimenter is therefore particularly interested in comparing the hardnesses obtained with these two fluxes. Find a 95% confidence interval for the difference in mean hardness between welds produced with flux B and those produced with flux C. Can we conclude that the two means differ? Solution We use expression (9.32). The sample means are 271.0 for flux C and 263.2 for flux B. The preceding output gives the quantity MSE as 63.975. (This value was also computed in Example 9.3.) The sample sizes are both equal to 5. There are I = 4 levels and N = 20 observations in total. For a 95% confidence interval, we consult the t table to find the value t16,.025 = 2.120. The 95% confidence interval is therefore or (−2.92, 18.52). To perform a test of the null hypothesis that the two treatment means are equal, we compute the value of the test statistic (expression 9.33) and obtain Consulting the t table with N − I = 16 degrees of freedom, we find that P is between 2(0.05) = 0.10 and 2(0.10) = 0.20 (note that this is a two-tailed test; software yields P = 0.143). We cannot conclude that the treatment means differ. If it is desired to perform a fixed-level test at level α = 0.05 as an alternative to Page 697 computing the P-value, the critical t value is t16,.025 = 2.120. The left-hand side of the inequality (9.34) is |271.0 − 263.2| = 7.8. The right-hand side is . Since 7.8 does not exceed 10.72, we do not reject H0 at the 5% level. The following output (from MINITAB) presents 95% Fisher LSD confidence intervals for each difference between treatment means in the weld experiment. The values labeled "Center" are the differences between pairs of treatment means. The quantities labeled "Lower" and "Upper" are the lower and upper bounds, respectively, of the confidence interval. Of particular note is the simultaneous confidence level of 81.11%. This indicates that although we are 95% confident that any given confidence interval contains its true difference in means, we are only 81.11% confident that all the confidence intervals contain their true differences. In Example 9.9, a single test was performed on the difference between two specific means. What if we wanted to test every pair of means, to see which ones we could conclude to be different? It might seem reasonable to perform the LSD test on each pair. However, this is not appropriate, because when several tests are performed, the likelihood of rejecting a true null hypothesis increases. This is the multiple testing problem, which is discussed in some detail in Section 6.14. This problem is revealed in the preceding output, which shows that the Page 698 confidence is only 81.11% that all the 95% confidence intervals contain their true values. When several confidence intervals or hypothesis tests are to be considered simultaneously, the confidence intervals must be wider, and the criterion for rejecting the null hypothesis more strict, than in situations where only a single interval or test is involved. In these situations, multiple comparisons methods are used to produce simultaneous confidence intervals and simultaneous hypothesis tests. If level 100(1 − α)% simultaneous confidence intervals are constructed for differences between every pair of means, then we are confident at the 100(1 − α)% level that every confidence interval contains the true difference. If simultaneous hypothesis tests are conducted for all null hypotheses of the form H0 : μi − μj = 0, then we may reject, at level α, every null hypothesis whose P-value is less than α. The Bonferroni Method of Multiple Comparisons The Bonferroni method, discussed in Section 6.14, is a general method, valid anytime that several confidence intervals or tests are considered simultaneously. The method is simple to apply. Let C be the number of pairs of differences to be compared. For example, if there are I treatments, and all pairs of differences are to be compared, then C = I(I − 1)/2. The Bonferroni method is the same as the LSD method, except that α is replaced with α/C. The Bonferroni Method for Simultaneous Confidence Intervals and Hypothesis Tests Assume that C differences of the form μi − μj are to be considered. The Bonferroni simultaneous confidence intervals, at level 100(1 − α)%, for the C differences μi − μj are (9.35) We are 100(1 − α)% confident that the Bonferroni confidence intervals contain the true value of the difference μi − μj for all C pairs under consideration. To test C null hypotheses of the form H0 : μi − μj = 0, the test statistics are To find the P-value for each test, consult the Student's t table with N − I degrees of freedom, and multiply the P-value found there by C. Specifically, if then H0 is rejected at level α. Example 9.10 Page 699 Example 9.10 For the weld data discussed in Example 9.9, use the Bonferroni method to determine which pairs of fluxes, if any, can be concluded, at the 5% level, to differ in their effect on hardness. Solution There are I = 4levels, with J = 5 observations at each level, for a total of N = 20 observations in all. With four levels, there are a total of C = (4)(3)/2 = 6 pairs of means to compare. To test at the α = 5% level, we compute α/(2C) = 0.004167. The critical t value is t16,.004167. This value is not in the table; it is between t16,.005 = 2.921 and t16,.001 = 3.686. Using computer software, we calculated t16,.004167 = 3.0083. Without software, one could roughly approximate this value by interpolation. Now MSE = 63.975 (see Example 9.9), so . The four sample means are as follows: Flux Mean hardness A B C D 253.8 263.2 271.0 262.0 There is only one pair of sample means, 271.0 and 253.8, whose difference is greater than 15.22. We therefore conclude that welds produced with flux A have different mean hardness than welds produced with flux C. None of the other differences are significant at the 5% level. Although easy to use, the Bonferroni method has the disadvantage that as the number of pairs C becomes large, the confidence intervals become very wide, and the hypothesis tests have low power. The reason for this is that the Bonferroni method is a general method, not specifically designed for analysis of variance or for normal populations. In many cases C is fairly large, in particular it is often desired to compare all pairs of means. In these cases, a method called the Tukey–Kramer method is superior, because it is designed for multiple comparisons of means of normal populations. We now describe this method. The Tukey–Kramer Method of Multiple Comparisons The Tukey–Kramer method is based on a distribution called the Studentized range distribution, rather than on the Student's t distribution. The Studentized range distribution has two values for degrees of freedom, which for the Tukey–Kramer method are I and N − I. (In comparison, the F test uses I − 1 and N − I degrees of freedom.) The Tukey–Kramer method uses the 1 − α quantile of the Studentized range distribution with I and N − I degrees of freedom; this quantity is denoted qI,N−I,α. Table A.9 (in Appendix A) presents values of qI,N−I,α for various values of I, N, and α. The mechanics of the Tukey–Kramer method are the same as those for the LSD method, except that is replaced with . The quantity is sometimes called the honestly significant difference (HSD), in contrast to Fisher's least significant difference. Page 700 The Tukey–Kramer Method for Simultaneous Confidence Intervals and Hypothesis Tests The Tukey–Kramer level 100(1 − α)% simultaneous confidence intervals for all differences μi − μj are (9.36) We are 100(1 − α)% confident that the Tukey–Kramer confidence intervals contain the true value of the difference μi − μj for every i and j. To test all null hypotheses H0 : μi − μj = 0 simultaneously, the test statistics are The P-value for each test is found by consulting the Studentized range table (Table A.9) with I and N − I degrees of freedom. For every pair of levels i and j for which the null hypothesis H0 : μi − μj = 0 is rejected at level α. A note on terminology: When the design is balanced, with all sample sizes equal to J, the quantity is equal for all pairs of levels. In this case, the method is often simply called Tukey's method. Example 9.11 For the weld data in Table 9.1 (in Section 9.1), use the Tukey–Kramer method to determine which pairs of fluxes, if any, can be concluded, at the 5% level, to differ in their effect on hardness. Solution There are I = 4levels, with J = 5 observations at each level, for a total of N = 20 observations in all. To test at level α = 0.05, we consult the Studentized range table (Table A.9) to find q4,16,.05 = 4.05. The value of MSE is 63.975 (see Example 9.9). Therefore . The four sample means are as follows: Flux Mean hardness A B C D 253.8 263.2 271.0 262.0 There is only one pair of sample means, 271.0 and 253.8, whose difference is greater than 14.49. We therefore conclude that welds Page 701 produced with flux A have a different mean hardness than welds produced with flux C. None of the other differences are significant at the 5% level. Comparing the results of Example 9.11 with those of Example 9.10 shows that in this case the Tukey–Kramer method is slightly more powerful than the Bonferroni method, since its critical value is only 14.49 while that of the Bonferroni method was 15.22. When all possible pairs are compared, as in this example, the Tukey–Kramer method is always more powerful than the Bonferroni method. When only a few of the possible pairs are to be compared, the Bonferroni method is sometimes more powerful. Sometimes only a single test is of interest, but the difference that is tested is chosen by examining the sample means and choosing two whose difference is large. In these cases a multiple comparisons method should be used, even though only one test is being performed. Example 9.12 illustrates the idea. Example 9.12 An engineer examines the weld data in Table 9.1 and notices that the two treatments with the largest difference in sample means are flux A and flux C. He decides to test the null hypothesis that the hardness for welds produced with flux A differs from that for welds produced with flux C. Since he will only perform one test, he uses the Fisher LSD method rather than the Bonferroni or Tukey–Kramer method. Explain why this is wrong. Solution The engineer has examined every pair of means and has chosen the two whose difference is largest. Although he is formally performing only one test, he has chosen that test by comparing every pair of sample means. For this reason he should use a multiple comparisons procedure, such as the Bonferroni or Tukey–Kramer method. The following output (from MINITAB) presents the Tukey–Kramer 95% simultaneous confidence intervals for the weld data. Page 702 The values labeled "Center" are the differences between pairs of treatment means. The quantities labeled "Lower" and "Upper" are the lower and upper bounds, respectively, of the confidence interval. We are 95% confident that every one of these confidence intervals contains the true difference in treatment means. Note that the "Individual confidence level" is 98.87%. This means that we are 98.87% confident that any one specific confidence interval contains its true value. Finally we point out that because the confidence level for the Tukey–Kramer intervals is higher than that for the Fisher LSD intervals, the Tukey–Kramer intervals are wider. Example 9.13 In Example 9.5 (in Section 9.1), several measurements of the maximum hourly concentrations (in μg/m3) of SO2 were presented for each of four power plants, and it was concluded that the mean concentrations at the four plants were not all the same. The following output (from MINITAB) presents the Tukey–Kramer 95% simultaneous confidence intervals for mean concentrations at the four plants. Which pairs of plants, if any, can you conclude with 95% confidence to have differing means? Page 703 Solution Among the simultaneous confidence intervals, there are two that do not contain 0. These are the intervals for μ1 − μ2 and for μ1 − μ3. Therefore we conclude that the mean concentrations differ between plants 1 and 2 and between plants 1 and 3. Exercises for Section 9.2 1. The article "Organic Recycling for Soil Quality Conservation in a Sub-Tropical Plateau Region" (K. Chakrabarti, B. Sarkar, et al., J. Agronomy and Crop Science, 2000:137–142) reports an experiment in which soil specimens were treated with six different treatments, with two replicates per treatment, and the acid phosphate activity (in μmol p-nitrophenol released per gram of oven-dry soil per hour) was recorded. An ANOVA table for a one-way ANOVA follows. One-way ANOVA: Treatments A, B, C, D, E, F Source DF SS MS Treatment 5 1.18547 0.23709 Error 6 0.03050 0.00508 Total 11 1.21597 The treatment means were F 46.64 P 0.000 Page 704 Treatment A B C D E F Mean 0.99 1.99 1.405 1.63 1.395 1.22 a. Can you conclude that there are differences in acid phosphate activity among the treatments? b. Use the Tukey–Kramer method to determine which pairs of treatment means, if any, are different at the 5% level. c. Use the Bonferroni method to determine which pairs of treatment means, if any, are different at the 5% level. d. Which method is more powerful in this case? 2. Use the Tukey–Kramer method to test the hypothesis that the two treatment means are equal? e. The experimenter notices that treatment A had the smallest sample mean, while treatment B had the largest. Of the Fisher LSD, Bonferroni, and Tukey–Kramer method, which, if any, can be used to test the hypothesis that these two treatment means are equal? The article "Optimum Design of an A-pillar Trim with Rib Structures for Occupant Head Protection" (H. Kim and S. Kang, Proceedings of the Institution of Mechanical Engineers, 2001:1161–1169) discusses a study in which several types of A-pillars were compared to determine which provided the greatest protection to occupants of automobiles during a collision. Following is a one-way ANOVA table, where the treatments are three levels of longitudinal spacing of the rib (the article also discussed two insignificant factors, which are omitted here). There were nine replicates at each level. The response is the head injury criterion (HIC), which is a unitless quantity that measures the impact energy absorption of the pillar. One-way ANOVA: Spacing Source Spacing Error Total DF 2 24 26 SS 50946.6 120550.9 171497.4 MS 25473.3 5023.0 F 5.071 P 0.015 The treatment means were Treatment Mean A B C 930.87 873.14 979.41 a. Can you conclude that the longitudinal spacing affects the absorption of impact energy? Use the Tukey–Kramer method to determine which pairs of treatment means, if any, are different at the 5% level. c. 3. Use the Bonferroni method to determine which pairs of treatment means, if any, are different at the 5% level. d. Which method is more powerful in this case, the Tukey–Kramer method or the Bonferroni method? Acrylic resins used in the fabrication of dentures should not absorb much water, since water sorption reduces strength. The article "Reinforcement of Acrylic Resin for Provisional Fixed Restorations. Part III: Effects of Addition of Titania and Zirconia Mixtures on Some Mechanical and Physical Properties" (W. Panyayong, Y. Oshida, et al., Bio-Medical Materials and Engineering, 2002:353–366) describes a study of the effect on water sorption of adding titanium dioxide (TiO2) and zirconium dioxide (ZrO2) to a standard acrylic resin. Twelve specimens from each of several formulations, containing various amounts of TiO2 and ZrO2, were immersed in water for one week, and the water sorption (in μg/mm2) was measured in each. The results are presented in the following table. Page 705 Formulation A (control) B C D E F G H Volume % TiO2 ZrO2 Mean 0 2 0 1 0.5 1 2 1 14 12 8 0 0.5 0 1 0.5 1 1 2 2 16.05 2 2 14.23 3 15.37 3 2 2 14.50 2 15.5 13.98 0.96 0.97 0.64 Use the Bonferroni method to determine which of the noncontrol formulations (B through H) differ, at the 5% level, in their mean water sorption from the control formulation A. b. Repeat part (a) using the Tukey–Kramer method. c. Which method is more powerful for these comparisons? Why? 4. Refer to Exercise 2 in Section 9.1. a. Use the Bonferroni method to determine which, if any, differ from the mean of the control group at the 5% level. b. Use the Tukey–Kramer method to determine which, if any, differ from the mean of the control group at the 5% level. c. Which is the more powerful method to find the treatments whose mean differs from that of the control group, the Bonferroni method or the Tukey–Kramer method? Refer to Exercise 13 in Section 9.1. a. Use the Bonferroni method to determine which pairs, if any, are different at the 5% level. b. Use the Tukey–Kramer method to find all the pairs of treatments whose means differ at the 5% level. c. Which is the more powerful method to find all the pairs of treatments whose means are different, the Bonferroni method or the Tukey–Kramer method? Refer to Exercise 1 in Section 9.1. A scientist wants to determine whether the mean strength of specimens cured at 30°C differs from the mean strengths of specimens cured at 0°C, 10°C, and 20°C. a. Use the Bonferroni method to determine which of the means, if any, for 0°C, 10°C, and 20°C differ from the mean for 30°C. Use the 5% level. b. Use the Tukey–Kramer method to determine which of the means, if any, 0°C, 10°C, and 20°C differ from the mean for 30°C. Use the 5% level. c. Which is the more powerful method to find the treatments whose means differ from that of the 30°C temperature, the Bonferroni method or the Tukey–Kramer method? Refer to Exercise 13 in Section 9.1. A metallurgist wants to determine whether the mean tensile strength for specimens annealed at 900°C differs from the mean strengths for specimens annealed at 750°C, 800°C, and 850°C. a. Use the Bonferroni method to determine which of the means, if any, for 750°C, 800°C, and 850°C differ from the mean for 900°C. b. Use the Tukey–Kramer method to determine which of the means, if any, for 750°C, 800°C, and 850°Cdiffer from the mean for 900°C. c. Which is the more powerful method to find the treatments whose means differ from that of the 900°C mean, the Bonferroni method or the Tukey–Kramer method? Refer to Exercise 3 in Section 9.1. Use the Tukey–Kramer method to determine which pairs of treatments, if any, differ at the 5% level. 8. Refer to Exercise 5 in Section 9.1. Page 706 a. Use the Fisher LSD method to find a 95% confidence interval for the difference between the means for treatments B and D. b. Use the Tukey–Kramer method to determine which pairs of treatments, if any, differ at the 5% level. 9. Refer to Exercise 5 in Section 9.1. a. Use the Fisher LSD method to find a 95% confidence interval for the difference between the means for specimens aged 3 days and specimens aged 365 days. b. Use the Tukey–Kramer method to determine which pairs of temperatures, if any, differ at the 5% level. 10. Refer to Exercise 9 in Section 9.1. a. Use the Fisher LSD method to find a 95% confidence interval for the difference between the means for plants A and C. b. Use the Tukey–Kramer method to determine which pairs of temperatures, if any, differ at the 5% level. 11. Refer to Exercise 16 in Section 9.1. a. Use the Fisher LSD method to find a 95% confidence interval for the difference between the means for a diameter of 16 and a diameter of 36. b. Use the Tukey–Kramer method to determine which pairs of diameters, if any, differ at the 5% level. 12. Refer to Exercise 18 in Section 9.1. a. Use the Fisher LSD method to find a 95% confidence interval for the difference between the means for annularly threaded and smooth shank nails. b. Use the Tukey–Kramer method to determine which pairs of nail types, if any, differ at the 5% level. 13. In an experiment to determine the effect of catalyst on the yield of a certain reaction, the mean yields for reactions run with each of four catalysts were , , and . Assume that five runs were made with each catalyst. a. If MSE = 3.85, compute the value of the F statistic for testing the null hypothesis that all four catalysts have the same mean yield. Can this null hypothesis be rejected at the 5% level? b. Use the Tukey–Kramer method to determine which pairs of catalysts, if any, may be concluded to differ at the 5% level. 14. In an experiment to determine the effect of curing time on the compressive strength of a certain type of concrete, the mean strengths, in MPa, for specimens cured for each of four curing times were , and . Assume that four specimens were cured for each curing time. a. If MSE = 875.2, compute the value of the F statistic for testing the null hypothesis that all four curing times have the same mean strength. Can this null hypothesis be rejected at the 5% level? b. Use the Tukey–Kramer method to determine which pairs of curing times, if any, may be concluded to differ at the 5% level. 15. For some data sets, the F statistic will reject the null hypothesis of no difference in mean yields, but the Tukey–Kramer method will not find any pair of means that can be concluded to differ. For the four sample means given in Exercise 13, assuming a sample size of 5 for each treatment, find a value of MSE so that the F statistic rejects the null hypothesis of no difference at the 5% level, while the Tukey–Kramer method does not find any pair of means to differ at the 5% level. 16. For some data sets, the F statistic will reject the null hypothesis of no difference in mean yields, but the Tukey–Kramer method will not find any pair of means that can be concluded to differ. For the four sample means given in Exercise 14, assuming a sample size of 4 for each treatment, find a value of MSE so that the F statistic rejects the null hypothesis of no difference at the 5% level, while the Tukey–Kramer method does not find any pair of means to differ at the 5% level. Page 707 a. 13. 14. 15. 16. 9.3 Two-Factor Experiments In one-factor experiments, discussed in Sections 9.1 and 9.2, the purpose is to determine whether varying the level of a single factor affects the response. Many experiments involve varying several factors, each of which may affect the response. In this section, we will discuss the case in which there are two factors. The experiments, naturally enough, are called two-factor experiments. We illustrate with an example. A chemical engineer is studying the effects of various reagents and catalysts on the yield of a certain process. Yield is expressed as a percentage of a theoretical maximum. Four runs of the process were made for each combination of three reagents and four catalysts. The results are presented in Table 9.2. In this experiment there are two factors, the catalyst and the reagent. The catalyst is called the row factor, since its value varies from row to row in the table, while the reagent is called the column factor. These designations are arbitrary, in that the table could just as easily have been presented with the rows representing the reagents and the columns representing the catalysts. TABLE 9.2 Yields for runs of a chemical process with various combinations of reagent and catalyst Catalyst A B C D 86.8 71.9 65.5 63.9 82.4 72.1 72.4 70.4 86.7 80.0 76.6 77.2 83.5 77.4 66.7 81.5 93.4 74.5 66.7 73.7 Reagent 2 85.2 94.8 87.1 71.9 77.1 76.7 81.6 84.2 3 83.1 86.1 84.9 77.9 87.5 72.7 79.8 89.6 82.7 77.8 75.7 85.1 77.3 83.0 90.1 ... 94.8 78.9 76.1 ... In general, there are I levels of the row factor and J levels of the column factor. (In Table 9.2, I = 4 and J = 3.) There are therefore IJ different combinations of the two factors. The terminology for these factor combinations is not standardized. We will refer to each combination of factors as a treatment, but some authors use the term treatment combination. Recall that the units assigned to a given treatment are called replicates. When the number of replicates is the same for each treatment, we will denote this number by K. Thus in Table 9.2, K = 4. When observations are taken on every possible treatment, the design is called a complete design or a full factorial design. Incomplete designs, in which there are no data for one or more treatments, can be difficult to interpret, except for some special cases. When possible, complete designs should be used. When the number of replicates is the same for each treatment, the design is said to be balanced. For one-factor experiments, we did not need to assume that the design was balanced. With two-factor experiments, unbalanced designs are more difficult to analyze than balanced designs. We will restrict our discussion to balanced designs. As with one-factor experiments, the factors may be fixed or random. The methods that we will describe apply to models where both effects are fixed. Later we will briefly describe models where one or both factors are random. In a completely randomized design, each treatment represents a population, and the observations on that treatment are a simple random sample from that population. We will denote the sample values for the treatment corresponding to the ith level of the row factor and the jth level of the column factor by Xij1,..., XijK. We will denote the population mean outcome for this treatment by μij. The values μij are often called the treatment means. In general, the purpose of a two-factor experiment is to determine whether the treatment means are affected by varying either the row factor, the column factor, or both. The method of analysis appropriate for two-factor experiments is called two-way analysis of variance. Parameterization for Two-Way Analysis of Variance In a two-way analysis of variance, we wish to determine whether varying the level of the row or column factors changes the value of μij. To do this, we must express μij in terms of parameters that describe the row and column factors separately. We'll begin this task by describing some notation for the averages of the treatment means for the different levels of the row and column factors. For any level i of the row factor, the average of all the treatment means μij in the ith row is denoted . We express in terms of the treatment means as follows: (9.37) Similarly, for any level j of the column factor, the average of all the treatment means μij in the jth column is denoted . We express in terms of the treatment means as follows: (9.38) Finally, we define the population grand mean, denoted by μ, which represents the average of all the treatment means μij. The population grand mean can also be expressed as the average of the quantities or of the quantities : (9.39) Table 9.3 illustrates the relationships among μij, TABLE 9.3 , and μ. Treatment means and their averages across rows and down columns Column Level 2 ⋯ μ12 ⋯ Row Level 1 1 μ11 2 μ21 μ22 ⋯ μ2J ⋮ I ⋮ μI1 ⋮ μI2 ⋯ ⋯ ⋮ μIJ Column Mean Using the quantities J μ1J ⋯ Row Mean ⋮ μ , and μ, we can decompose the treatment mean μij as follows: (9.40) Equation (9.40) expresses the treatment mean μij as a sum of four terms. In practice, simpler notation is used for the three rightmost terms in Equation (9.40): (9.41) (9.42) (9.43) Page 709 Each of quantities μ, αi, βj, and γij has an important interpretation: ■ ■ ■ ■ The quantity μ is the population grand mean, which is the average of all the treatment means. The quantity is called the ith row effect. It is the difference between the average treatment mean for the ith level of the row factor and the population grand mean. The value of αi indicates the degree to which the ith level of the row factor tends to produce outcomes that are larger or smaller than the population grand mean. The quantity is called the jth column effect. It is the difference between the average treatment mean for the jth level of the column factor and the population grand mean. The value of βj indicates the degree to which the jth level of the column factor tends to produce outcomes that are larger or smaller than the population grand mean. The quantity is called the ij interaction. The effect of a level of a row (or column) factor may depend on which level of the column (or row) factor it is paired with. The interaction terms measure the degree to which this occurs. For example, assume that level 1 of the row factor tends to produce a large outcome when paired with column level 1, but a small outcome when paired with column level 2. In this case γ1, 1 would be positive, and γ1, 2 would be negative. Both row effects and column effects are called main effects to distinguish them from the interactions. Note that there are J row effects, one for each level of the row factor, J column effects, one for each level of the column factor, and IJ interactions, one for each treatment. Furthermore, it follows from the definitions of quantities and μ in Equations (9.37) through (9.39) that the row effects, column effects, and interactions satisfy the following constraints: (9.44) We now can express the treatment means μij in terms of αi, βj, and γij. From Equation (9.40) it follows that (9.45) For each observation Xijk, define εijk = Xijk − μij, the difference between the observation Page 710 and its treatment mean. The quantities εijk are called errors. It follows that (9.46) Combining Equations (9.46) and (9.45) yields the two-way ANOVA model: (9.47) When the interactions γij are all equal to 0, the additive model is said to apply. Under the additive model, Equation (9.45) becomes (9.48) and Equation (9.47) becomes (9.49) Under the additive model, the treatment mean μij is equal to the population grand mean μ, plus an amount αi that results from using row level i plus an amount βj that results from using column level j. In other words, the combined effect of using row level i along with column level j is found by adding the individual main effects of the two levels. When some or all of the interactions are not equal to 0, the additive model does not hold, and the combined effect of a row level and a column level cannot be determined from their individual main effects. We will now show how to estimate the parameters for the full two-way model (9.47). The procedure for the additive model is exactly the same, except that the interactions γij are not estimated. The procedure is straightforward. We first define some notation for various averages of the data Xijk, using the data in Table 9.2 as an example. Table 9.4 presents the average yield for the four runs for each reagent and catalyst in Table 9.2. TABLE 9.4 Average yields for runs of a chemical process using different combinations of reagent and catalyst Catalyst A B C D Column Mean 1 84.85 75.35 70.30 73.18 75.92 Reagent 2 89.13 79.40 76.65 81.10 81.57 3 85.28 84.65 78.20 77.23 81.34 Row Mean 86.42 79.80 75.05 77.17 Sample Grand Mean Each number in the body of Table 9.4 is the average of the four numbers in the corresponding cell of Table 9.2. These are called the cell means. They are denoted and are defined by (9.50) Averaging the cell means across the rows produces the row means : Page 711 (9.51) Averaging the cell means down the columns produces the column means : (9.52) The sample grand mean can be found by computing the average of the row means, the average of the column means, the average of the cell means, or the average of all the observations: (9.53) Now we describe how to estimate the parameters in the two-way ANOVA model. The fundamental idea is that the best estimate of the treatment mean is the cell mean , which is the average of the sample observations having that treatment. It follows that the best estimate of the quantity is the row mean , the best estimate of the quantity is the column mean , and the best estimate of the population grand mean μ is the sample grand mean . We estimate the row effects αi, the column effects βj, and the interactions γij by substituting these estimates into Equations (9.41) through (9.43). (9.54) (9.55) (9.56) The row effects, column effects, and interactions satisfy constraints given in Equation (9.44). By performing some algebra, it can be shown that their estimates satisfy the same constraints: (9.57) Example 9.14 Compute the estimated row effects, column effects, and interactions for the data in Table 9.2. Solution Using the quantities in Table 9.4 and Equations (9.54) through (9.56), we compute Page 712 Using Two-Way ANOVA to Test Hypotheses A two-way analysis of variance is designed to address three main questions: 1. Does the additive model hold? 2. If so, is the mean outcome the same for all levels of the row factor? 3. If so, is the mean outcome the same for all levels of the column factor? In general, we ask questions 2 and 3 only when we believe that the additive model may hold. We will discuss this further later in this section. The three questions are addressed by performing hypothesis tests. The null hypotheses for these tests are as follows: 1. To test whether the additive model holds, we test the null hypothesis that all the interactions are equal to 0: If this null hypothesis is true, the additive model holds. 2. 3. If this null hypothesis is true, the additive model holds. To test whether the mean outcome is the same for all levels of the row factor, we test the null hypothesis that all the row effects are equal to 0: If this null hypothesis is true, then the mean outcome is the same for all levels of the row factor. To test whether the mean outcome is the same for all levels of the column factor, we test the null hypothesis that all the column effects are equal to 0: If this null hypothesis is true, then the mean outcome is the same for all levels of the column factor. We now describe the standard tests for these null hypotheses. For the tests to be valid, the following conditions must hold: Page 713 Assumptions for Two-Way ANOVA The standard two-way ANOVA hypothesis tests are valid under the following conditions: 1. The design must be complete. 2. The design must be balanced. 3. The number of replicates per treatment, K, must be at least 2. 4. Within any treatment, the observations Xij1,..., XijK are a simple random sample from a normal population. 5. The population variance is the same for all treatments. We denote this variance by σ2. Just as in one-way ANOVA, the standard tests for these null hypotheses are based on sums of squares. Specifically, they are the row sum of squares (SSA), the column sum of squares (SSB), the interaction sum of squares (SSAB), and the error sum of squares (SSE). Also of interest is the total sum of squares (SST), which is equal to the sum of the others. Formulas for these sums of squares are as follows: (9.58) (9.59) (9.60) (9.61) (9.62) It can be seen from the rightmost expressions in Equations (9.58) through (9.62) that the total sum of squares, SST, is equal to the sum of the others. This is the analysis of variance identity for two-way ANOVA. Page 714 The Analysis of Variance Identity (9.63) Along with each sum of squares is a quantity known as its degrees of freedom. The sums of squares and their degrees of freedom are generally presented in an ANOVA table. Table 9.5 presents the degrees of freedom for each sum of squares, along with the computationally most convenient formula. We point out that the degrees of freedom for SST is the sum of the degrees of freedom for the other sums of squares. TABLE 9.5 Source Rows (SSA) Columns (SSB) ANOVA table for two-way ANOVA Degrees of Freedom I−1 Sum of Squares J−1 Interactions (SSAB) (I − 1)(J − 1) Error (SSE) IJ(K − 1) Total (SST) IJK − 1 Note that the magnitude of SSA depends on the magnitude of the estimated row effects . Therefore when the true row effects αi are equal to 0, SSA will tend to be smaller, and when some of the true row effects are not equal to 0, SSA will tend to be larger. We will therefore reject H0: α1 = ⋯ = αI = 0 when SSA is sufficiently large. Similarly, SSB will tend to be smaller when the true column effects βj, are all equal to 0 and larger when some column effects are not zero, and SSAB will tend to be smaller when the true interactions γij are all equal to 0 and larger when some interactions are not zero. We will therefore reject H0: β1 = ⋯ = βJ = 0 when SSB is sufficiently large, and we will reject H0: γ11 = ⋯ = γIJ = 0 when SSAB is sufficiently large. We can determine whether SSA, SSB, and SSAB are sufficiently large by comparing them to the error sum of squares, SSE. As in one-way ANOVA (Section 9.1), SSE depends only on the distances between the observations and their own cell means. SSE therefore Page 715 measures only the random variation inherent in the process and is not affected by the values of the row effects, column effects, or interactions. To compare SSA, SSB, and SSAB with SSE, we first divide each sum of squares by its degrees of freedom, producing quantities known as mean squares. The mean squares, denoted MSA, MSB, MSAB, and MSE, are defined as follows: (9.64) The test statistics for the three null hypotheses are the quotients of MSA, MSB, and MSAB with MSE. The null distributions of these test statistics are F distributions. Specifically, ■ Under H0: α1 = ⋯ = αI = 0, the statistic distribution. ■ ■ Under H0: β1 = ⋯ = βJ = 0, the statistic distribution. Under H0: γ11 = ⋯ = γIJ = 0, the statistic distribution. In practice, the sums of squares, mean squares, and test statistics are usually calculated with the use of a computer. The following output (from MINITAB) presents the ANOVA table for the data in Table 9.2. Two-way ANOVA: Yield versus Catalyst, Reagent Source Catalyst Reagent Interaction Error Total DF 3 2 6 36 47 SS 877.56 327.14 156.98 1125.33 2487.02 MS 292.521 163.570 26.164 31.259 F 9.36 5.23 0.84 P 0.000 0.010 0.550 S = 5.591 R-sq = 54.75% R-Sq(adj) = 40.93% The labels DF, SS, MS, F, and P refer to degrees of freedom, sum of squares, mean square, F statistic, and P-value, respectively. As in one-way ANOVA, the mean square for error (MSE) is an estimate of the error variance σ2 and the quantity labeled "S" is the square root of MSE and is an estimate of the error standard deviation σ. The quantities "R-sq" and "R-sq(adj)" are computed with formulas analogous to those in one-way ANOVA. Page 716 Example 9.15 Use the preceding ANOVA table to determine whether the additive model is plausible for the yield data. If the additive model is plausible, can we conclude that either the catalyst or the reagent affects the yield? Solution We first check to see if the additive model is plausible. The P-value for the interactions is 0.55, which is not small. We therefore do not reject the null hypothesis that all the interactions are equal to 0, and we conclude that the additive model is plausible. Since the additive model is plausible, we now ask whether the row or column factors affect the outcome. We see from the table that the P-value for the row effects (Catalyst) is approximately 0, so we conclude that the catalyst does affect the yield. Similarly, the P-value for the column effects (Reagent) is small (0.010), so we conclude that the reagent affects the yield as well. Example 9.16 The article "Uncertainty in Measurements of Dermal Absorption of Pesticides" (W. Navidi and A. Bunge, Risk Analysis, 2002:1175–1182) describes an experiment in which a pesticide was applied to skin at various concentrations and for various lengths of time. The outcome is the amount of the pesticide that was absorbed into the skin. The following output (from MINITAB) presents the ANOVA table. Is the additive model plausible? If so, do either the concentration or the duration affect the amount absorbed? Two-way ANOVA: Absorbed versus Concentration, Duration Source Concent Duration Interaction Error Total DF 2 2 4 27 35 SS 49.991 19.157 0.337 6.250 75.735 MS 24.996 9.579 0.084 0.231 F 107.99 41.38 0.36 P 0.000 0.000 0.832 Solution The P-value for the interaction is 0.832, so we conclude that the additive model is plausible. The P-values for both concentration and dose are very small. Therefore we can conclude that both concentration and duration affect the amount absorbed. Checking the Assumptions A residual plot can be used to check the assumption of equal variances, and a normal probability plot of the residuals can be used to check normality. The residual plot plots the residuals versus the fitted values, which are the sample means . Figures 9.6 and Page 717 9.7 present both a normal probability plot and a residual plot for the yield data found in Table 9.2. The assumptions appear to be well satisfied. FIGURE 9.6 Normal probability plot for the residuals from the yield data. There is no evidence of a strong departure from normality. FIGURE 9.7 Residual plot for the yield data. There is no evidence against the assumption of equal variances. Don't Interpret the Main Effects When the Additive Model Doesn't Hold When the interactions are small enough so that the additive model is plausible, interpretation of the main effects is fairly straightforward, as shown in Examples 9.15 and 9.16. When the additive model does not hold, however, it is not always easy to interpret the main Page 718 effects. Here is a hypothetical example to illustrate the point. Assume that a process is run under conditions obtained by varying two factors at two levels each. Two runs are made at each of the four combinations of row and column levels. The yield of the process is measured each time, with the results presented in the following table. Column Level Row Level 1 2 1 51, 49 43, 41 2 43, 41 51, 49 Clearly, if it is desired to maximize yield, the row and column factors matter—we want either row level 1 paired with column level 1 or row level 2 paired with column level 2. Now look at the following ANOVA table. Source Row Column Interaction Error Total DF 1 1 1 4 7 SS 0.0000 0.0000 128.00 8.0000 136.00 MS 0.0000 0.0000 128.00 2.0000 F 0.00 0.00 64.00 P 1.000 1.000 0.001 The main effects sum of squares for both the row and column main effects are equal to 0, and their P-values are equal to 1, which is as large as a P-value can be. If we follow the procedure used in Examples 9.15 and 9.16, we would conclude that neither the row factor nor the column factor affects the yield. But it is clear from the data that the row and column factors do affect the yield. What is happening is that the row and column factors do not matter on the average. Level 1 of the row factor is better if level 1 of the column factor is used, and level 2 of the row factor is better if level 2 of the column factor is used. When averaged over the two levels of the column factor, the levels of the row factor have the same mean yield. Similarly, the column levels have the same mean yield when averaged over the levels of the row factor. When the effects of the row levels depend on which column levels they are paired with, and vice versa, the main effects can be misleading. It is the P-value for the interactions that tells us not to try to interpret the main effects. This P-value is quite small, so we reject the additive model. Then we know that some of the interactions are nonzero, so the effects of the row levels depend on the column levels, and vice versa. For this reason, when the additive model is rejected, we should not try to interpret the main effects. We need to look at the cell means themselves in order to determine how various combinations of row and column levels affect the outcome. Page 719 Summary In a two-way analysis of variance: ■ If the additive model is not rejected, then hypothesis tests for the main effects can be used to determine whether the row or column factors affect the outcome. ■ If the additive model is rejected, then hypothesis tests for the main effects should not be used. Instead, the cell means must be examined to determine how various combinations of row and column levels affect the outcome. Example 9.17 The thickness of the silicon dioxide layer on a semiconductor wafer is crucial to its performance. In the article "Virgin Versus Recycled Wafers for Furnace Qualification: Is the Expense Justified?" (V. Czitrom and J. Reece, Statistical Case Studies for Process Improvement, SIAM-ASA, 1997:87–103), oxide layer thicknesses were measured for three types of wafers: virgin wafers, wafers recycled in-house, and wafers recycled by an external supplier. In addition, several furnace locations were used to grow the oxide layer. A two-way ANOVA for three runs at one wafer site for the three types of wafers at three furnace locations was performed. The data are presented in the following table, followed by the results (from MINITAB). Furnace Location 1 1 1 2 2 2 3 3 3 Wafer Type Oxide Layer Thickness (Å) Virgin 90.1 90.7 89.4 In-house 90.4 88.8 90.6 External 92.6 90.0 93.3 Virgin 91.9 88.6 89.7 In-house 90.3 91.9 91.5 External 88.3 88.2 89.4 Virgin 88.1 90.2 86.6 In-house 91.0 90.4 90.2 External 91.5 89.8 89.8 Two-way ANOVA for Thickness versus Wafer, Location Source DF SS Wafer 2 5.8756 Location 2 4.1089 Interaction 4 21.349 Error 18 25.573 Total 26 56.907 MS 2.9378 2.0544 5.3372 1.4207 F 2.07 1.45 3.76 P 0.155 0.262 0.022 Since recycled wafers are cheaper, the company hopes that there is no Page 720 difference in the oxide layer thickness among the three types of chips. If possible, determine whether the data are consistent with the hypothesis of no difference. If not possible, explain why not. Solution The P-value for the interactions is 0.022, which is small. Therefore the additive model is not plausible, so we cannot interpret the main effects. A good thing to do is to make a table of the cell means. Table 9.6 presents the sample mean for each treatment. TABLE 9.6 Sample means for each treatment Furnace Location 1 2 3 Column Mean Virgin 90.067 90.067 88.300 89.478 Wafer Type In-House 89.933 91.233 90.533 90.566 External 91.967 88.633 90.367 90.322 Row Mean 90.656 89.978 89.733 From Table 9.6, it can be seen that the thicknesses do vary among wafer types, but no one wafer type consistently produces the thickest, or the thinnest, oxide layer. For example, at furnace location 1 the externally recycled wafers produce the thickest layer while the in-house recycled wafers produce the thinnest. At furnace location 2 the order is reversed: The in-house wafers produce the thickest layer while the external ones produce the thinnest. This is due to the interaction of furnace location and wafer type. A Two-Way ANOVA Is Not the Same as Two One-Way ANOVAs Example 9.17 presented a two-way ANOVA with three row levels and three column levels, for a total of nine treatments. If separate one-way ANOVAs were run on the row and column factors separately, there would be only six treatments. This means that in practice, running separate oneway ANOVAs on each factor may be less costly than running a two-way ANOVA. Unfortunately, this "one-at-a-time" design is sometimes used in practice for this reason. It is important to realize that running separate one-way analyses on the individual factors can give results that are misleading when interactions are present. To see this, look at Table 9.6. Assume that an engineer is trying to find the combination of furnace and location that will produce the thinnest oxide layer. He first runs the process once at each furnace location, using in-house recycled wafers, because those wafers are the ones currently being used in production. Furnace location 1 produces the thinnest layer for in-house wafers. Now the engineer runs the process once for each wafer type, all at location 1, which was the best for the in-house wafers. Of the three wafer types, in-house wafers produce the thinnest layer at location 1. So the conclusion drawn from the one-at-a-time analysis is that the thinnest layers are produced by the combination of in-house wafers at furnace location 1. A look at Table 9.6 shows that the conclusion Page 721 is false. There are two combinations of furnace location and wafer type that produce thinner layers than this. The one-at-a-time method assumes that the wafer that produces the thinnest layers at one location will produce the thinnest layers at all locations, and that the location that produces the thinnest layers for one wafer type will produce the thinnest layers for all types. This is equivalent to assuming that there are no interactions between the factors, which in the case of the wafers and locations is incorrect. In summary, the one-at-a-time method fails because it cannot detect interactions between the factors. Summary ■ ■ When there are two factors, a two-factor design must be used. Examining one factor at a time cannot reveal interactions between the factors. Interaction Plots Interaction plots can help to visualize interactions. Figure 9.8 presents an interaction plot for the wafer data. We describe the method by which this plot was constructed. The vertical axis represents the response, which is layer thickness. One factor is chosen to be represented on the horizontal axis. We chose furnace location; it would have been equally acceptable to have chosen wafer type. Now we proceed through the levels of the wafer-type factor. We'll start with external wafers. The three cell means for external wafers, as shown in Table 9.6, are 91.967, 88.633, and 90.367, corresponding to furnace locations 1, 2, and 3, respectively. These values are plotted above their respective furnace locations and are connected with line segments. This procedure is repeated for the other two wafer types to complete the plot. FIGURE 9.8 Interaction plot for the wafer data. The lines are far from parallel, indicating substantial interaction between the factors. For the wafer data, the means for external wafers follow a substantially different Page 722 pattern than those for the other two wafer types. This is the source of the significant interaction and is the reason that the main effects of wafer and furnace type cannot be easily interpreted. In comparison, for perfectly additive data, for which the interaction estimates are equal to 0, the line segments in the interaction plot are parallel. Figure 9.9 illustrates this hypothetical case. FIGURE 9.9 Interaction plot for hypothetical data with interaction estimates equal to 0. The line segments are parallel. Figure 9.10 presents an interaction plot for the yield data. The cell means were presented in Table 9.4. The lines are not parallel, but their slopes match better than those for the wafer data. This indicates that the interaction estimates are nonzero, but are smaller than those for Page 723 the wafer data. In fact, the P-value for the test of the null hypothesis of no interaction was 0.550. (see the ANOVA table on page 715). The deviation from parallelism exhibited in Figure 9.10 is therefore small enough to be consistent with the hypothesis of no interaction. FIGURE 9.10 Interaction plot for yield data. Multiple Comparisons in Two-Way ANOVA An F test is used to test the null hypothesis that all the row effects (or all the column effects) are equal to 0. If the null hypothesis is rejected, we can conclude that some of the row effects (or column effects) differ from each other. But the hypothesis test does not tell us which ones are different from the rest. If the additive model is plausible, then a method of multiple comparisons known as Tukey's method (related to the Tukey–Kramer method described in Section 9.2) can be applied to determine for which pairs the row effects or column effects can be concluded to differ from one another. The method is described in the following box. Tukey's Method for Simultaneous Confidence Intervals and Hypothesis Tests in Two-Way ANOVA Let I be the number of levels of the row factor, J be the number of levels of the column factor, and K be the sample size for each treatment. Then, if the additive model is plausible, the Tukey level 100(1 − α)% simultaneous confidence intervals for all differences αi − αj (or all differences βi − βj) are We are 100(1 − α)% confident that the Tukey confidence intervals contain the true value of the difference αi − (or βi − βj) for every i and j. For every pair of levels i and j for which , the null hypothesis H0: αi − αj = 0 is rejected at level α. For every pair of levels i and j for which , the null hypothesis H0: βi − βj = 0 is rejected at level α. Example 9.18 In Example 9.14, the main effects and interactions were computed for the yield data in Table 9.2. An ANOVA table for these data was presented on page 715. If appropriate, use Tukey's method to determine which pairs of catalysts and which pairs of reagents can be concluded to differ, at the 5% level, in their effect on yield. Page 724 Solution From the ANOVA table, the P-value for interactions is 0.550. Therefore the additive model is plausible, so it is appropriate to use Tukey's method. Catalyst is the row factor and reagent is the column factor, so I = 4, J = 3, and K = 4. From the ANOVA table, MSE = 31.259. We first find all pairs for which the row effects differ at the 5% level. For the row effects, we should use the value q4,36,.05. This value is not found in the Studentized range table (Table A.9 in Appendix A). We will therefore use the value q4,30,.05 = 3.85, which is close to (just slightly greater than) q4,36,.05. In Example 9.14, the estimated row effects were computed to be We compute The pairs of row effects whose differences are greater than 6.21 are and and , and . We conclude that the mean yield of catalyst A differs from the mean yields of catalysts B, C, and D, and that catalyst B differs from catalysts C and D. We now find all pairs for which the column effects differ at the 5% level. For the column effects, we should use the value q3,36,.05, but since this value is not found in the Studentized range table, we will use the value q3,30,05 = 3.49. We compute . In Example 9.14, the estimated column effects were computed to be The pairs of column effects whose differences are greater than 4.88 are and and . We conclude that the mean yield of reagent 1 differs from the mean yields of reagents 2 and 3. There is no evidence that the mean yields of reagents 2 and 3 differ from each other. Two-Way ANOVA when K = 1 The F tests we have presented require the assumption that the sample size K for each treatment be at least 2. The reason for this is that when K = 1, the error sum of squares (SSE) will be equal to 0, since Xijk = X̄ij. for each i and j. In addition, the degrees of freedom for SSE, which is IJ(K − 1), is equal to 0 when K = 1. When K = 1, a two-way ANOVA cannot be performed unless it is certain that the additive model holds. In this case, since the interactions are assumed to be zero, the mean square for interaction (MSAB; see Equation 9.64) and its degrees of freedom can be used in place of MSE to test the main row and column effects. Random Factors Our discussion of two-factor experiments has focused on the case where both factors are fixed. Such an experiment is said to follow a fixed effects model. Experiments can also be designed in which one or both factors are random. If both factors are random, the experiment is Page 725 said to follow a random effects model. If one factor is fixed and one is random, the experiment is said to follow a mixed model. In the one-factor case, the analysis is the same for both fixed and random effects models, while the null hypothesis being tested differs. In the two-factor case, both the methods of analysis and the null hypotheses differ among fixed effects models, random effects models, and mixed models. Methods for models in which one or more effects are random can be found in more advanced texts, such as Hocking (2014). Unbalanced Designs We have assumed that the design is balanced, that is, that the number of replications is the same for each treatment. The methods described here do not apply to unbalanced designs. However, unbalanced designs that are complete may be analyzed with the methods of multiple regression. An advanced text such as Draper and Smith (1998) may be consulted for details. Exercises for Section 9.3 1. 2. To assess the effect of piston ring type and oil type on piston ring wear, three types of piston ring and four types of oil were studied. Three replications of an experiment, in which the number of milligrams of material lost from the ring in four hours of running was measured, were carried out for each of the 12 combinations of oil type and piston ring type. With oil type as the row effect and piston ring type as the column effect, the following sums of squares were observed: SSA = 1.0926, SSB = 0.9340, SSAB = 0.2485, SSE = 1.7034. a. How many degrees of freedom are there for the effect of oil type? b. How many degrees of freedom are there for the effect of piston ring type? c. How many degrees of freedom are there for interactions? d. How many degrees of freedom are there for error? e. Construct an ANOVA table. You may give ranges for the P-values. f. Is the additive model plausible? Provide the value of the test statistic and the P-value. g. Is it plausible that the main effects of oil type are all equal to 0? Provide the value of the test statistic and the P-value. h. Is it plausible that the main effects of piston ring type are all equal to 0? Provide the value of the test statistic and the P-value. A machine shop has three machines used in precision grinding of cam rollers. Three machinists are employed to grind rollers on the machines. In an experiment to determine whether there are differences in output among the machines or their operators, each operator worked on each machine on four different days. The outcome measured was the daily production of parts that met specifications. With the operator as the row effect and the machine as the column effect, the following sums of squares were observed: SSA = 3147.0, SSB = 136.5, SSAB = 411.7, SSE = 1522.0. a. How many degrees of freedom are there for the operator effect? b. How many degrees of freedom are there for the machine effect? c. How many degrees of freedom are there for interactions? d. How many degrees of freedom are there for error? e. Construct an ANOVA table. You may give ranges for the P-values. f. Is the additive model plausible? Provide the value of the test statistic and the P-value. g. Is it plausible that the main effects of operator are all equal to 0? Provide the value of the test statistic and the P-value. h. Is it plausible that the main effects of machine are all equal to 0? Provide the value of the test statistic and the P-value. 3. 4. 5. Page 726 To test the effect of alloy and mold temperature on tensile strength, an experiment to determine the tensile strength of castings involved three different alloys and five different mold temperatures. Four specimens of each alloy were cast at each mold temperature. With mold temperature as the row factor and alloy as the column factor, the sums of squares were: SSA = 69,738, SSB = 8958, SSAB = 7275, and SST = 201,816. a. Construct an ANOVA table. You may give ranges for the P-values. b. Is the additive model plausible? Explain. c. Is it plausible that the main effects of mold temperature are all equal to 0? Provide the value of the test statistic and the P-value. d. Is it plausible that the main effects of alloy are all equal to 0? Provide the value of the test statistic and the P-value. The effect of curing pressure on bond strength was tested for four different adhesives. There were three levels of curing pressure. Five replications were performed for each combination of curing pressure and adhesive. With adhesive as the row factor and curing pressure as the column factor, the sums of squares were: SSA = 155.7, SSB = 287.9, SSAB = 156.7, and SST = 997.3. a. Construct an ANOVA table. You may give ranges for the P-values. b. Is the additive model plausible? Explain. c. Is it plausible that the main effects of curing pressure are all equal to 0? Provide the value of the test statistic and the P-value. d. Is it plausible that the main effects of adhesive are all equal to 0? Provide the value of the test statistic and the P-value. The article "Change in Creep Behavior of Plexiform Bone with Phosphate Ion Treatment" (R. Regimbal, C. DePaula, and N. Guzelsu, Bio-Medical Materials and Engineering, 2003:11–25) describes an experiment to study the effects of saline and phosphate ion solutions on mechanical properties of plexiform bone. The following table presents the yield stress measurements for six specimens treated with either saline (NaCl) or phosphate ion (Na2HPO4) solution, at a temperature of either 25°C or 37°C. (The article presents means and standard deviations only; the values in the table are consistent with these.) Solution Temperature Yield Stress (MPa) NaCl 25°C 138.40 130.89 94.646 96.653 116.90 88.215 NaCl 37°C 92.312 147.28 116.48 88.802 114.37 90.737 Na2HPO4 25°C 120.18 129.43 139.76 132.75 137.23 121.73 Na2HPO4 37°C 123.50 128.94 102.86 99.941 161.68 136.44 Estimate all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of the test statistic and the P-value. Can the effect of solution (NaCl versus Na2HPO4) on yield stress be described by interpreting the main effects of solution? If so, interpret the main effects, including the appropriate test statistic and P-value. If not, explain why not. e. Can the effect of temperature on yield stress be described by interpreting the main effects of temperature? If so, interpret the main effects, including the appropriate test statistic and P-value. If not, explain why not. The article "Variance Reduction Techniques: Experimental Comparison and Analysis for Single Systems" (I. Sabuncuoglu, M. Fadiloglu, and S. Çelik, IIE Transactions, 2008:538– 551) describes a study of methods for reducing variance in estimators of the mean inventory on hand. Two systems, the serial line system and the inventory system, were studied, along with two schemes for proportional sampling. The results given below (in percent) are consistent with the means and standard deviations presented in the article. Page 727 System Scheme Reduction Serial A 6.4 5.8 5.1 4.4 7.0 8.4 8.5 7.5 7.0 7.9 Line Serial B 4.7 7.3 8.3 5.3 10.6 4.5 8.2 10.8 8.5 5.7 Line Inventory A 11.0 8.9 9.3 9.2 7.9 9.7 9.0 12.5 6.7 9.8 Inventory B 8.9 7.0 10.7 10.3 6.2 12.2 7.0 9.5 10.7 8.6 a. b. c. d. 7. Estimate all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of a test statistic and the P-value. Can the effect of system on reduction be described by interpreting the main effects of system? If so, interpret the main effects, including the appropriate test statistic and P-value. If not, explain why not. e. Can the effect of scheme on reduction be described by interpreting the main effects of scheme? If so, interpret the main effects, including the appropriate test statistic and Pvalue. If not, explain why not. The effect of curing pressure on bond strength (in MPa) was tested for two different adhesives. There were three levels of curing pressure. Three replications were performed for each combination of curing pressure and adhesive. The results are presented in the following table. Adhesive A A A B B B a. b. Curing Pressure Low Medium High Low Medium High Bond Strength 8.1 8.8 6.3 6.6 6.4 8.1 3.5 4.1 2.6 5.1 6.0 5.7 2.9 5.2 5.6 4.5 6.0 8.3 2 Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of the test statistic and the P-value. c. 8. Can the effect of adhesive on the bond strength be described by interpreting the main effects of adhesive? If so, interpret the main effects. If not, explain why not. d. Can the effect of curing pressure on the bond strength be described by interpreting the main effects of curing pressure? If so, interpret the main effects. If not, explain why not. Adding glass particles to clay brick may improve the structural properties of the brick. The article "Effects of Waste Glass Additions on the Properties and Durability of Fired Clay Brick" (S. Chidiac and L. Federico, Can J Civ Eng, 2007:1458–1466) describes experiments in which the compressive strength (in MPa) was measured for bricks with varying amounts of glass content and glass particle size. The results in the following table are consistent with means and standard deviations presented in the article. Glass Content (%) 5 5 10 10 15 15 a. b. Size Coarse 78.7 Fine 73.0 Coarse 80.1 Fine 76.2 Coarse 90.3 Fine 141.1 Strength (MPa) 70.8 78.6 81.7 79.2 90.1 71.4 93.8 82.7 76.9 76.5 84.3 77.7 80.1 121.2 81.4 61.2 95.8 103.1 99.5 73.3 144.1 122.4 134.5 124.9 Estimate all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Page 728 Is the additive model plausible? Provide the value of the test statistic and the P-value. c. Can the effect of glass content on strength be described by interpreting the main effects of glass content? If so, interpret the main effects, including the appropriate test statistic and P-value. If not, explain why not. d. Can the effect of particle size on strength be described by interpreting the main effects of particle size? If so, interpret the main effects, including the appropriate test statistic and P-value. If not, explain why not. 9. The article "Application of Radial Basis Function Neural Networks in Optimization of Hard Turning of AISID2 Cold-Worked Tool Steel With a Ceramic Tool" (S. Basak, U. Dixit, and J. Davim, Journal of Engineering Manufacture, 2007:987–998) presents the results of an experiment in which tool wear was computed for various values of three factors. We consider two of those factors, cutting speed and cutting time. The results are presented in the following table. c. d. 9. Speed (m/min) 80 80 80 150 Time (min) Wear (mm) 5 5 6 5 4 10 8 8 8 8 15 11 10 9 9 10 5 9 11 9 8 10 3 8 9 150 150 220 220 220 10 15 5 10 15 14 16 34 60 65 14 15 33 59 64 15 26 19 29 31 13 24 21 31 33 13 24 18 27 32 18 28 75 18 25 20 31 80 a. b. c. d. Estimate all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of a test statistic and the P-value. Can the effect of speed on wear be described by interpreting the main effects of speed? If so, interpret the main effects, including the appropriate test statistic and P-value. If not, explain why not. e. Can the effect of time on wear be described by interpreting the main effects of time? If so, interpret the main effects, including the appropriate test statistic and P-value. If not, explain why not. 10. The lifetime of a tool was investigated under three settings for feed rate and three settings for speed. Four tools were tested under each combination of settings. The results (in hours) were as follows. Feed Rate Light Light Light Medium Medium Medium Heavy Heavy Heavy a. b. Speed Slow Medium Fast Slow Medium Fast Slow Medium Fast 60.6 57.8 56.5 51.2 49.6 45.7 44.8 46.6 37.2 Lifetime 57.0 61.4 59.4 62.8 52.3 58.1 53.1 48.3 48.1 49.8 46.8 45.0 46.7 41.9 41.4 38.3 32.8 39.9 59.7 58.2 53.9 51.6 51.1 49.2 51.3 37.9 35.9 Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of the test statistic and the P-value. Page 729 c. Can the effect of feed rate on lifetime be described by interpreting the main effects of feed rate? If so, interpret the main effects, using multiple comparisons at the 5% level if necessary. If not, explain why not. d. Can the effect of the speed on lifetime be described by interpreting the main effects of distance? If so, interpret the main effects, using multiple comparisons at the 5% level if necessary. If not, explain why not. 11. Artificial joints consist of a ceramic ball mounted on a taper. The article "Friction in Orthopaedic Zirconia Taper Assemblies" (W. Macdonald, A. Aspenberg, et al., Proceedings of the Institution of Mechanical Engineers, 2000: 685–692) presents data on the coefficient of friction for a push-on load of 2 kN for taper assemblies made from two zirconium alloys and employing three different neck lengths. Five measurements were made for each combination of material and neck length. The results presented in the following table are consistent with the cell means and standard deviations presented in the article. Taper Material CPTi-ZrO2 Neck Length Short Coefficient of Friction 0.254 0.195 0.281 0.289 0.220 CPTi-ZrO2 Medium 0.196 0.220 0.185 0.259 0.197 CPTi-ZrO2 Long 0.329 0.481 0.320 0.296 0.178 TiAlloy-ZrO2 Short 0.150 0.118 0.158 0.175 0.131 TiAlloy-ZrO2 Medium 0.180 0.184 0.154 0.156 0.177 TiAlloy-ZrO2 Long 0.178 0 198 0.201 0.199 0.210 a. b. c. Compute the main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of the test statistic, its null distribution, and the P-value. d. Can the effect of material on the coefficient of friction be described by interpreting the main effects of material? If so, interpret the main effects. If not, explain why not. e. Can the effect of neck length on the coefficient of friction be described by interpreting the main effects of neck length? If so, interpret the main effects, using multiple comparisons at the 5% level if necessary. If not, explain why not. 12. The article "Anodic Fenton Treatment of Treflan MTF" (D. Saltmiras and A. Lemley, Journal of Environmental Science and Health, 2001:261–274) describes a two-factor experiment designed to study the sorption of the herbicide trifluralin. The factors are the initial trifluralin concentration and the Fe2:H2O2 delivery ratio. There were three replications for each treatment. The results presented in the following table are consistent with the means and standard deviations reported in the article. Initial Concentration (M) 15 15 15 15 40 40 40 40 100 Delivery Ratio Sorption (%) 1:0 10.90 8.47 12.43 1:1 3.33 2.40 2.67 1:5 0.79 0.76 0.84 1:10 0.54 0.69 0.57 1:0 6.84 7.68 6.79 1:1 1.72 1.55 1.82 1:5 0.68 0.83 0.89 1:10 0.58 1.13 1.28 1:0 6.61 6.66 7.43 100 100 100 1:1 1:5 1:10 1.25 1.46 1.17 1.27 0.93 0.67 1.49 1.16 0.80 Page 730 a. b. c. Estimate all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of the test statistic, its null distribution, and the P-value. 13. Refer to Exercise 12. The treatments with a delivery ratio of 1:0 were controls, or blanks. It was discovered after the experiment that the high apparent levels of sorption in these controls was largely due to volatility of the trifluralin. Eliminate the control treatments from the data. a. Estimate all main effects and interactions. b. Construct an ANOVA table. You may give ranges for the P-values. c. Is the additive model plausible? Provide the value of the test statistic, its null distribution, and the P-value. d. Construct an interaction plot. Explain how the plot illustrates the degree to which interactions are present. 14. The article "Use of Taguchi Methods and Multiple Regression Analysis for Optimal Process Development of High Energy Electron Beam Case Hardening of Cast Iron" (M. Jean and Y. Tzeng, Surface Engineering, 2003:150–156) describes a factorial experiment designed to determine factors in a high-energy electron beam process that affect hardness in metals. Results for two factors, each with three levels, are presented in the following table. Factor A is the travel speed in mm/s, and factor B is accelerating voltage in volts. The outcome is Vickers hardness. There were six replications for each treatment. In the article, a total of seven factors were studied; the two presented here are those that were found to be the most significant. A 10 10 10 20 20 20 30 30 30 a. B 10 25 50 10 25 50 10 25 50 875 712 568 876 889 756 901 789 792 896 719 546 835 876 732 926 801 786 Hardness 921 686 698 621 559 757 868 812 849 768 723 681 893 856 776 845 775 706 642 632 723 796 706 723 832 827 675 613 645 734 772 615 712 841 831 568 Estimate all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of the test statistic and the P-value. Can the effect of travel speed on the hardness be described by interpreting the main effects of travel speed? If so, interpret the main effects, using multiple comparisons at the 5% level if necessary. If not, explain why not. e. Can the effect of accelerating voltage on the hardness be described by interpreting the main effects of accelerating voltage? If so, interpret the main effects, using multiple comparisons at the 5% level if necessary. If not, explain why not. 15. The article "T-Bracing for Stability of Compression Webs in Wood Trusses" (R. Leichti, I. Hofaker, et al., Journal of Structural Engineering, 2002:374–381) presents results of experiments in which critical buckling loads (in kN) for T-braced assemblies were estimated by a finite-element method. The following table presents data in which the factors are the length of the side member and its method of attachment. There were 10 replications for each combination of factors. The data are consistent with the means and standard deviations given in the article. Page 731 c. d. Attachment Adhesive Adhesive Adhesive Nail Nail Nail Length Critical Buckling Load Quarter 7.90 8.71 7.72 8.88 8.55 6.95 7.07 Half 14.07 13.82 14.77 13.19 11.98 12.72 9.48 Full 26.80 28.57 24.82 23.51 27.57 25.96 24.28 Quarter 6.92 5.38 5.38 5.89 6.07 6.37 7.14 Half 9.67 9.17 10.39 10.90 10.06 9.86 10.41 Full 20.63 21.15 24.75 20.76 21.64 21.47 25.25 7.97 8.10 8.88 11.70 14.81 13.46 28.55 23.82 24.86 5.84 5.65 5.73 8.79 7.86 13.59 13.09 12.09 25.68 21.64 28.16 6.71 4.36 6.78 10.24 9.31 11.99 22.52 20.45 20.38 a. b. c. Compute all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of a test statistic and the P-value. d. Can the effect of attachment method (nail versus adhesive) on the critical buckling load be described by interpreting the main effects of attachment method? If so, interpret the main effects. If not, explain why not. e. Can the effect of side member length on the critical buckling load be described by interpreting the main effects of side member length? If so, interpret the main effects, using multiple comparisons at the 5% level if necessary. If not, explain why not. 16. The article referred to in Exercise 15 also presents measurements of Young's modulus for side members of T-braced assemblies. The following table presents data in which the factors are the length of the side member and its method of attachment. There were 10 replications for each combination of factors. The data (in kN/mm2) are consistent with the means and standard deviations given in the article. Attachment Adhesive Adhesive Adhesive Nail Nail Nail Length Young's Modulus Quarter 9.56 10.67 8.82 8.40 9.23 8.20 10.23 Half 8.74 9.24 10.77 9.10 8.00 11.14 10.00 Full 9.84 9.80 8.31 7.37 10.12 9.18 8.93 Quarter 10.24 9.38 9.38 7.48 9.23 9.64 8.45 9.58 9.17 8.65 8.12 7.57 9.79 7.89 8.86 8.05 8.13 9.07 8.07 Nail Half 9.74 7.96 9.34 8.32 9.64 8.21 10.43 8.75 9.37 9.48 9.42 8.48 9.76 9.45 8.68 9.55 10.20 9.66 9.84 8.89 10.80 8.97 a. b. c. d. Compute all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Provide the value of the test statistic and the P-value. Can the effect of attachment method (nail versus adhesive) on Young's modulus be described by interpreting the main effects of attachment method? If so, interpret the main effects. If not, explain why not. e. Can the effect of side member length on Young's modulus be described by interpreting the main effects of side member length? If so, interpret the main effects, using multiple comparisons at the 5% level if necessary. If not, explain why not. 17. Each of three operators made two weighings of several silicon wafers. Results are presented in the following table for three of the wafers. All the wafers had weights very close to 54 g, so the weights are reported in units of μg above 54 g. (Based on the article "Revelation of a Microbalance Warmup Effect," J. Buckner, B. Chin, et al., Statistical Case Studies for Industrial Process Improvement, SIAM-ASA, 1997:39–45.) Wafer 1 2 3 Operator 1 11 15 210 208 111 113 Operator 2 10 6 205 201 102 105 Operator 3 14 10 208 207 108 111 Page 732 a. Construct an ANOVA table. You may give ranges for the P-values. b. Can it be determined from the ANOVA table whether there is a difference in the measured weights among the operators? If so, provide the value of the test statistic and the P-value. If not, explain why not. 18. Refer to Exercise 17. It turns out that the measurements of operator 2 were taken in the morning, shortly after the balance had been powered up. A new policy was instituted to leave the balance powered up continuously. The three operators then made two weighings of three different wafers. The results are presented in the following table. a. Wafer 1 2 3 Operator 1 152 156 443 440 229 227 Operator 2 156 155 442 439 229 232 Operator 3 152 157 435 439 225 228 Construct an ANOVA table. You may give ranges for the P-values. b. Compare the ANOVA table in part (a) with the ANOVA table in part (a) of Exercise 17. Would you recommend leaving the balance powered up continuously? Explain your reasoning. 19. The article "Cellulose Acetate Microspheres Prepared by O/W Emulsification and Solvent a. b. Evaporation Method" (K. Soppimath, A. Kulkarni, et al., Journal of Microencapsulation, 2001:811–817) describes a study of the effects of the concentrations of polyvinyl alcohol (PVAL) and dichloromethane (DCM) on the encapsulation efficiency in a process that produces microspheres containing the drug ibuprofen. There were three concentrations of PVAL (measured in units of % w/v) and three of DCM (in mL). The results presented in the following table are consistent with the means and standard deviations presented in the article. PVAL DCM = 50 DCM = 40 DCM = 30 0.5 98.983 99.268 95.149 96.810 94.572 86.718 75.288 74.949 72.363 1.0 89.827 94.136 96.537 82.352 79.156 80.891 76.625 76.941 72.635 2.0 95.095 95.153 92.353 86.153 91.653 87.994 80.059 79.200 77.141 a. b. Construct an ANOVA table. You may give ranges for the P-values. Discuss the relationships among PVAL concentration, DCM concentration, and encapsulation efficiency. 9.4 Randomized Complete Block Designs In some experiments, there are factors that vary and have an effect on the response, but whose effects are not of interest to the experimenter. For example, in one commonly occurring situation, it is impossible to complete an experiment in a single day, so the observations have to be spread out over several days. If conditions that can affect the outcome vary from day to day, then the day becomes a factor in the experiment, even though there may be no interest in estimating its effect. For a more specific example, imagine that three types of fertilizer are to be evaluated for their effect on yield of fruit in an orange grove, and that three replicates will be performed, for a total of nine observations. An area is divided into nine plots, in three rows of three plots each. Now assume there is a water gradient along the plot area, so that the rows receive differing amounts of water. The amount of water is now a factor in the experiment, even though there is no interest in estimating the effect of water amount on the yield of oranges. If the water factor is ignored, a one-factor experiment could be carried out with fertilizer as the only factor. Each of the three fertilizers would be assigned to three of the plots. In a completely randomized experiment, the treatments would be assigned to the plots at random. Figure 9.11 presents two possible random arrangements. In the arrangement on the left, the plots with fertilizer A get more water than those with the other two fertilizers. In the plot on the right, the plots with fertilizer C get the most water. When the treatments for one factor are assigned completely at random, it is likely that they will not be distributed evenly over the levels of another factor. FIGURE 9.11 Two possible arrangements for three fertilizers, A, B, and C, assigned to nine plots completely at random. It is likely that the amounts of water will differ for the different fertilizers. If the amount of water in fact has a negligible effect on the response, then the completely randomized one-factor design is appropriate. There is no reason to worry about a factor that does not affect the response. But now assume that the water level does have a substantial impact on the response. Then Figure 9.11 shows that in any one experiment, the estimated effects of the treatments are likely to be thrown off the mark, or biased, by the differing levels of water. Different arrangements of the treatments bias the estimates in different directions. If the experiment is repeated several times, the estimates are likely to vary greatly from repetition to repetition. For this reason, the completely randomized one-factor design produces estimated effects that have large uncertainties. A better design for this experiment is a two-factor design, with water as the second factor. Since the effects of water are not of interest, water is called a blocking factor, rather than a treatment factor. In the two-factor experiment, there are nine treatment–block combinations, corresponding to the three fertilizer treatment levels and the three water block levels. With nine experimental units (the nine plots), it is necessary to assign one plot to each combination of fertilizer and water. Figure 9.12 presents two possible arrangements. FIGURE 9.12 Two possible arrangements for three fertilizers, A, B, and C, with the restriction that each fertilizer must appear at each water level (block). The distribution of water levels is always the same for each fertilizer. In the two-factor design, each treatment appears equally often (once, in this Page 734 example) in each block. As a result, the effect of the blocking factor does not contribute to uncertainty in the estimate of the main effects of the treatment factor. Because each treatment must appear equally often in each block, the only randomization in the assignment of treatments to experimental units is the order in which the treatments appear in each block. This is not a completely randomized design; it is a design in which the treatments are randomized within blocks. Since every possible combination of treatments and blocks is included in the experiment, the design is complete. For this reason the design is called a randomized complete block design. Randomized complete block designs can be constructed with several treatment factors and several blocking factors. We will restrict our discussion to the case where there is one treatment factor and one blocking factor. The data from a randomized complete block design are analyzed with a two-way ANOVA, in the same way that data from any complete, balanced two-factor design would be. There is one important consideration, however. The only effects of interest are the main effects of the treatment factor. In order to interpret these main effects, there must be no interaction between treatment and blocking factors. Example 9.19 Three fertilizers are studied for their effect on yield in an orange grove. Nine plots of land are used, divided into blocks of three plots each. A randomized complete block design is used, with each fertilizer applied once in each block. The results, in pounds of harvested fruit, are presented in the following table, followed by MINITAB output for the two-way ANOVA. Can we conclude that the mean yields differ among fertilizers? What assumption is made about interactions between fertilizer and plot? How is the sum of squares for error computed? Fertilizer A B C Plot 1 430 367 320 Plot 2 542 463 421 Plot 3 287 253 207 Two-way ANOVA: Yield versus Block, Fertilizer Source Fertilizer Block Error Total DF 2 2 4 8 SS 16213.6 77046.9 651.778 93912.2 MS 8106.778 38523.44 162.9444 F 49.75 236.4 P 0.001 0.000 Solution The P-value for the fertilizer factor is 0.001, so we conclude that fertilizer does have an effect on yield. The assumption is made that there is no interaction between the fertilizer Page 735 and the blocking factor (plot), so that the main effects of fertilizer can be interpreted. Since there is only one observation for each treatment–block combination (i.e., K = 1), the sum of squares for error (SSE) reported in the output is really SSAB, the sum of squares for interaction, and the error mean square (MSE) is actually MSAB. (See the discussion on page 724.) A closer look at the ANOVA table in Example 9.19 shows that in this experiment, blocking was necessary to detect the fertilizer effect. To see this, consider the experiment to be a one-factor experiment. The sum of squares for error (SSE) would then be the sum of SSE for the blocked design plus the sum of squares for blocks, or 651.778 + 77,046.9 = 77,698.7. The degrees of freedom for error would be equal to the sum of the degrees of freedom for error in the blocked design plus the degrees of freedom for blocks, or 2 + 4 = 6. The error mean square (MSE) would then be 77,698.7/6 ≈ 12,950 rather than 162.9444, and the F statistic for the fertilizer effect would be less than 1, which would result in a failure to detect an effect. In general, using a blocked design reduces the degrees of freedom for error, which by itself tends to reduce the power to detect an effect. However, unless the blocking factor has very little effect on the response, this will usually be more than offset by a reduction in the sum of squares for error. Failing to include a blocking factor that affects the response can reduce the power greatly, while including a blocking factor that does not affect the response reduces the power only modestly in most cases. For this reason it is a good idea to use a blocked design whenever it is thought to be possible that the blocking factor is related to the response. Summary ■ ■ ■ ■ A two-factor randomized complete block design is a complete balanced two-factor design in which the effects of one factor (the treatment factor) are of interest, while the effects of the other factor (the blocking factor) are not of interest. The blocking factor is included to reduce the uncertainty in the main effect estimates of the treatment factor. Since the object of a randomized complete block design is to estimate the main effects of the treatment factor, there must be no interaction between the treatment factor and the blocking factor. A two-way analysis of variance is used to estimate effects and to perform hypothesis tests on the main effects of the treatment factor. A randomized complete block design provides a great advantage over a completely randomized design when the blocking factor strongly affects the response and provides a relatively small disadvantage when the blocking factor has little or no effect. Therefore, when in doubt, it is a good idea to use a blocked design. Example 9.20 The article "Experimental Design for Process Settings in Aircraft Manufacturing" Page 736 (R. Sauter and R. Lenth, Statistical Case Studies: A Collaboration Between Academe and Industry, SIAM–ASA, 1998:151–157) describes an experiment in which the quality of holes drilled in metal aircraft parts was studied. One important indicator of hole quality is "excess diameter," which is the difference between the diameter of the drill bit and the diameter of the hole. Small excess diameters are better than large ones. Assume we are interested in the effect of the rotational speed of the drill on the excess diameter of the hole. Holes will be drilled in six test pieces (coupons), at three speeds: 6000, 10,000, and 15,000 rpm. The excess diameter can be affected not only by the speed of the drill, but also by the physical properties of the test coupon. Describe an appropriate design for this experiment. Solution A randomized complete block design is appropriate, with drill speed as the treatment factor, and test coupon as the blocking factor. Since six observations can be made in each block, each drill speed should be used twice in each block. The order of the speeds within each block should be chosen at random. Example 9.21 The design suggested in Example 9.20 has been adopted, and the experiment has been carried out. The results (from MINITAB) follow. Does the output indicate any violation of necessary assumptions? What do you conclude regarding the effect of drill speed on excess diameter? Two-way ANOVA: Excess Diameter versus Block, Speed Source Block Speed Interaction Error Total DF 5 2 10 18 35 SS 0.20156 0.07835 0.16272 0.67105 1.11368 MS 0.0403117 0.0391750 0.0162717 0.0372806 F 1.08 1.05 0.44 P 0.404 0.370 0.909 S = 0.1931 R-Sq = 39.74% R-Sq(adj) = 0.00% Solution In a randomized complete block design, there must be no interaction between the treatment factor and the blocking factor, so that the main effect of the treatment factor may be Page 737 interpreted. The P-value for interactions is 0.909, which is consistent with the hypothesis of no interactions. Therefore there is no indication in the output of any violation of assumptions. The P-value for the main effect of speed is 0.370, which is not small. Therefore we cannot conclude that excess hole diameter is affected by drill speed. Example 9.22 shows that a paired design (see Section 6.8), in which a t test is used to compare two population means, is a special case of a randomized block design. Example 9.22 A tire manufacturer wants to compare the tread wear of tires made from a new material with that of tires made from a conventional material. There are 10 tires of each type. Each tire will be mounted on the front wheel of a front-wheel drive car and driven for 40,000 miles. Then the tread wear will be measured for each tire. Describe an appropriate design for this experiment. Solution The response is the tread wear after 40,000 miles. There is one factor of interest: the type of tire. Since the cars may differ in the amounts of wear they produce, the car is a factor as well, but its effect is not of interest. A randomized complete block design is appropriate, in which one tire of each type is mounted on the front wheels of each car. You may note that the randomized complete block design in Example 9.22 is the same design that is used when comparing two population means with a paired t test, as discussed in Section 6.8. The paired design described there is a special case of a randomized complete block design, in which the treatment factor has only two levels, and each level appears once in each block. In fact, a two-way analysis of variance applied to data from such a design is equivalent to the paired t test. Multiple Comparisons in Randomized Complete Block Designs Once an ANOVA table has been constructed, then if the F test shows that the treatment main effects are not all the same, a method of multiple comparisons may be used to determine which pairs of effects may be concluded to differ. We describe Tukey's method, which is a special case of the Tukey–Kramer method described in Section 9.2. The degrees of freedom, and the mean square used, differ depending on whether each treatment appears only once, or more than once, in each block. Page 738 Tukey's Method for Multiple Comparisons in Randomized Complete Block Designs In a randomized complete block design, with I treatment levels, J block levels, and treatment main effects α1,..., αI: If each treatment appears only once in each block, then the null hypothesis H0: αi − αj = 0 is rejected at level α for every pair of treatments i and j for which ■ where MSAB is the mean square for interaction. If each treatment appears K > 1 times in each block, then the null hypothesis H0: αi − αj = 0 is rejected at level α for every pair of treatments i and j for which ■ where MSE is the mean square for error. For more information on randomized block designs, a text on design of experiments, such as Montgomery (2013a), can be consulted. Exercises for Section 9.4 1. Recycling newsprint is important in reducing waste. The article "The Influence of Newspaper Aging on Optical Properties of its De-inked Pulp" (M. Rahmaninia, A. Latibari, et al., Turkish J Eng Env Sci, 2008:35–39) presents the results of an experiment to determine the effects of the age of newsprint on the brightness (in percent) of the recycled paper. Four aging periods were considered, along with five concentrations of sodium hydroxide (NaOH), used to de-ink the paper. The effect of the sodium hydroxide concentration is not of interest. The results are presented in the following table. Age (Months) 0 4 8 12 a. b. c. d. 2. 0% 54.6 45.6 46.1 44.0 NaOH Concentration 0.5% 1% 1.5% 54.3 55.5 56.3 44.1 43.7 45.6 45.9 46.4 45.0 44.1 45.2 43.7 2% 56.6 47.1 47.9 46.5 Identify the blocking factor and the treatment factor. Construct an ANOVA table. You may give ranges for the P-values. Can you conclude that brightness differs with age? Explain. Which pairs of ages, if any, can you conclude to have differing brightnesses? Use the 5% level. Page 739 A study was done to see which of four machines is fastest in performing a certain task. There are three operators; each performs the task twice on each machine. A randomized block design is employed. The MINITAB output follows. Source Machine Block Interaction Error Total DF (i) (iv) (vii) (xi) (xii) SS 257.678 592.428 (viii) 215.836 1096.646 MS (ii) (v) (ix) 17.986 F (iii) (vi) (x) a. b. 3. 4. 5. P 0.021 0.000 0.933 Fill in the missing numbers (i) through (xii) in the output. Does the output indicate that the assumptions for the randomized block design are satisfied? Explain. c. Can you conclude that there are differences among the machines? Explain. Four lighting methods were used in each of three rooms. For each method and each room, the illuminance (in lux) was measured in three separate occasions, resulting in three replications of the experiment. The only effect of interest is the lighting type; the room is a blocking factor. The following sums of squares were calculated: sum of squares for blocks = 11432, sum of squares for treatments = 9943, sum of squares for interactions = 6135, total sum of squares = 51376. a. Construct an ANOVA table. You may give ranges for the P-values. b. Are the assumptions for a randomized complete block design satisfied? Explain. c. Does the ANOVA table provide evidence that lighting type affects illuminance? Explain. Three different corrosion-resistant coatings are being considered for use on iron pipes. Ten pieces of pipe are marked off in six equal segments. For each pipe, two segments received coating A, two received coating B, and the remaining two received coating C. The pipes were placed in a corrosive environment for a period of time; then the depth of the deepest pit (in mm) caused by corrosion was measured for each segment on each pipe. The effect of interest is the coating; the pipe is a blocking factor, and there were two replications on each pipe. The following sums of squares were calculated: sum of squares for blocks = 11.2, sum of squares for treatments = 4.8, sum of squares for interactions = 18.4, total sum of squares = 44.7. a. Construct an ANOVA table. You may give ranges for the P-values. b. Are the assumptions for a randomized complete block design satisfied? Explain. c. Does the ANOVA table provide evidence that mean pit depth differs among coatings? Explain. The article "Genotype-Environment Interactions and Phenotypic Stability Analyses of Linseed in Ethiopia" (W. Adguna and M. Labuschagne, Plant Breeding, 2002:66–71) describes a study in which seed yields of 10 varieties of linseed were compared. Each variety was grown on six different plots. The yields, in kilograms per hectare, are presented in the following table. Page 740 Plot Variety A 1 2032 2 1377 3 1343 4 1366 5 1276 6 1209 B C D E F G H I J 6. 1815 1739 1812 1781 1703 1476 1745 1679 1903 1126 1311 1313 1271 1089 1333 1308 1216 1382 1338 1340 1044 1308 1256 1162 1190 1326 1373 1188 1250 1245 1220 1385 1363 1269 1271 1609 1566 1473 1090 1371 1079 1056 1251 1506 1396 1454 1617 1280 1361 1318 1096 1325 1368 1366 a. Construct an ANOVA table. You may give ranges for the P-values. b. Can you conclude that the varieties have differing mean yields? The article "Sprinkler Technologies, Soil Infiltration, and Runoff" (D. DeBoer and S. Chu, Journal of Irrigation and Drainage Engineering, 2001:234–239) presents a study of the runoff depth (in mm) for various sprinkler types. Each of four sprinklers was tested on each of four days, with two replications per day (there were three replications on a few of the days; these are omitted). It is of interest to determine whether runoff depth varies with sprinkler type; variation from one day to another is not of interest. The data are presented in the following table. Sprinkler A B C D Day 1 8.3 5.5 6.5 9.5 1.8 1.2 0.7 0.8 Day 2 7.8 4.5 3.7 3.6 0.5 0.3 0.1 0.5 Day 3 10.7 9.8 7.7 10.6 1.7 1.9 0.1 0.5 Day 4 10.6 6.6 3.6 6.7 2.2 2.1 0.3 0.5 a. b. c. d. 7. Identify the blocking factor and the treatment factor. Construct an ANOVA table. You may give ranges for the P-values. Are the assumptions of a randomized complete block design met? Explain. Can you conclude that there are differences in mean runoff depth between some pairs of sprinklers? Explain. e. Which pairs of sprinklers, if any, can you conclude, at the 5% level, to have differing mean runoff depths? The article "Bromate Surveys in French Drinking Waterworks" (B. Legube, B. Parinet, et al., Ozone Science and Engineering, 2002:293–304) presents measurements of bromine concentrations (in μg/L) at several waterworks. The measurements made at 15 different times at each of four waterworks are presented in the following table. (The article also presented some additional measurements made at several other waterworks.) It is of interest to determine whether bromine concentrations vary among waterworks; it is not of interest to determine whether concentrations vary over time. Waterworks 1 2 3 4 5 6 Time 7 8 9 10 11 12 13 14 15 A B C D 29 24 25 31 9 29 17 37 7 21 20 34 35 24 24 30 40 20 19 39 53 25 19 41 38 15 17 34 38 14 23 34 41 8 22 29 34 12 27 33 42 14 17 33 35 35 33 34 38 32 33 16 35 38 39 31 36 33 37 16 Page 741 Construct an ANOVA table. You may give ranges for the P-values. Can you conclude that bromine concentration varies among waterworks? Which pairs of waterworks, if any, can you conclude, at the 5% level, to have differing bromine concentrations? d. Someone suggests that these data could have been analyzed with a one-way ANOVA, ignoring the time factor, with 15 observations for each of the four waterworks. Does the ANOVA table support this suggestion? Explain. The article "Multi-objective Scheduling Problems: Determination of Pruned Pareto Sets" (H. Taboada and D. Coit, IIE Transactions, 2008:552–564), presents examples in a discussion of optimization methods for industrial scheduling and production planning. In one example, seven different jobs were performed on each of five machines. The cost of each job on each machine is presented in the following table. Assume that it is of interest to determine whether costs differ between machines, but that it is not of interest whether costs differ between jobs. a. b. c. 8. Machine A B C D E a. b. c. 9. 1 16 22 12 18 14 2 24 18 12 28 14 3 18 30 15 16 11 Job 4 22 20 18 26 16 5 26 28 22 34 18 6 20 18 12 18 10 7 22 26 16 18 12 Identify the blocking factor and the treatment factor. Construct an ANOVA table. You may give ranges for the P-values. Can you conclude that there are differences in costs between some pairs of machines? Explain. d. Which pairs of machines, if any, can you conclude, at the 5% level, to have differing mean costs? You have been given the task of designing a study concerning the lifetimes of five different types of electric motor. The initial question to be addressed is whether there are differences in mean lifetime among the five types. There are 20 motors, four of each type, available for testing. A maximum of five motors can be tested each day. The ambient temperature differs from day to day, and this can affect motor lifetime. a. Describe how you would choose the five motors to test each day. Would you use a completely randomized design? Would you use any randomization at all? b. If Xij represents the measured lifetime of a motor of type i tested on day j, express the test statistic for testing the null hypothesis of equal lifetimes in terms of the Xij. 10. An engineering professor wants to determine which subject engineering students find most difficult among statistics, physics, and chemistry. She obtains the final exam grades for four students who took all three courses last semester and who were in the same sections of each class. The results are presented in the following table. Course Statistics Physics Chemistry a. b. 1 82 75 93 Student 2 3 94 78 70 81 82 80 4 70 83 70 The professor proposes a randomized complete block design, with the students as the blocks. Give a reason that this is likely not to be appropriate. Describe the features of the data in the preceding table that suggest that the assumptions of the randomized complete block design are violated. Page 742 9.5 2p Factorial Experiments When an experimenter wants to study several factors simultaneously, the number of different treatments can become quite large. In these cases, preliminary experiments are often performed in which each factor has only two levels. One level is designated as the "high" level, and the other is designated as the "low" level. If there are p factors, there are then 2p different treatments. Such experiments are called 2p factorial experiments. Often, the purpose of a 2p experiment is to determine which factors have an important effect on the outcome. Once this is determined, more elaborate experiments can be performed, in which the factors previously found to be important are varied over several levels. We will begin by describing 23 factorial experiments. Notation for 23 Factorial Experiments In a 23 factorial design, there are three factors and 23 = 8 treatments. The main effect of a factor is defined to be the difference between the mean response when the factor is at its high level and the mean response when the factor is at its low level. It is common to denote the main effects by A, B, and C. As with any factorial experiment, there can be interactions between the factors. With three factors, there are three two-way interactions, one for each pair of factors, and one threeway interaction. The two-way interactions are denoted by AB, AC, and BC, and the three-way interaction by ABC. The treatments are traditionally denoted with lowercase letters, with a letter indicating that a factor is at its high level. For example, ab denotes the treatment in which the first two factors are at their high level and the third factor is at its low level. The symbol "1" is used to denote the treatment in which all factors are at their low levels. Estimating Effects in a 23 Factorial Experiment Assume that there are K replicates for each treatment in a 23 factorial experiment. For each treatment, the cell mean is the average of the K observations for that treatment. The formulas for the effect estimates can be easily obtained from the 23 sign table, presented as Table 9.7. The signs are placed in the table as follows. For the main effects A, B, C, the sign is + for treatments in which the factor is at its high level, and − for treatments where the factor is at its low level. So for the main effect A, the sign is + for treatments a, ab, ac, and abc, and − for the rest. For the interactions, the signs are computed by taking the product of the signs in the corresponding main effects columns. For example, the signs for the two-way interaction AB are the products of the signs in columns A and B, and the signs for the three-way interaction ABC are the products of the signs in columns A and B and C. TABLE 9.7 Treatment 1 a b ab c ac bc abc Sign table for a 23 factorial experiment Cell Mean A − + − + − + − + B − − + + − − + + C − − − − + + + + AB + − − + + − − + AC + − + − − + − + BC + + − − − − + + ABC − + + − + − − + Estimating main effects and interactions is done with the use of the sign table. We illustrate how to estimate the main effect of factor A. Factor A is at its high level in the rows of the table where there is a "+" sign in column A. Each of the cell means , , , and Page 743 is an average response for runs made with factor A at its high level. We estimate the mean response for factor A at its high level to be the average of these cell means. Similarly, each row with a "—" sign in column A represents a treatment with factor A set to its low level. We estimate the mean response for factor A at its low level to be the average of the cell means in these rows. The estimate of the main effect of factor A is the difference in the estimated mean response between its high and low levels. The quantity inside the parentheses is called the contrast for factor A. It is computed by adding and subtracting the cell means, using the signs in the appropriate column of the sign table. Note that the number of plus signs is the same as the number of minus signs, so the sum of the coefficients is equal to 0. The effect estimate is obtained by dividing the contrast by half the number of treatments, which is 23/2, or 4. Estimates for other main effects and interactions are computed in an analogous manner. To illustrate, we present the effect estimates for the main effect C and for the two-way interaction AB: Page 744 Summary The contrast for any main effect or interaction is obtained by adding and subtracting the cell means, using the signs in the appropriate column of the sign table. For a 23 factorial experiment, (9.65) Example 9.23 A 23 factorial experiment was conducted to estimate the effects of three factors on the yield of a chemical reaction. The factors were A: catalyst concentration (low or high), B: reagent (standard formulation or new formulation), and C: stirring rate (slow or fast). Three replicates were obtained for each treatment. The yields, presented in the following table, are measured as a percent of a theoretical maximum. Estimate all effects and interactions. Treatment 1 a b ab c ac bc abc 71.67 78.46 77.14 79.72 72.65 80.10 80.20 75.58 Yield 70.55 75.42 78.25 76.17 71.03 73.91 73.49 80.28 67.40 81.77 78.33 78.41 73.54 74.81 74.86 71.64 Cell Mean 69.8733 78.5500 77.9067 78.1000 72.4067 76.2733 76.1833 75.8333 Solution We use the sign table (Table 9.7) to find the appropriate sums and differences of the cell means. We present the calculations for the main effect A, the two-way interaction BC, and the three-way interaction ABC: We present all the estimated effects in the following table, rounded off to the same Page 745 precision as the data: Term A B C AB AC BC ABC Effect 3.10 2.73 −0.93 −3.18 −1.34 −1.06 1.07 For each effect, we can test the null hypothesis that the effect is equal to 0. When the null hypothesis is rejected, this provides evidence that the factors involved actually affect the outcome. To test these null hypotheses, an ANOVA table is constructed containing the appropriate sums of squares. For the tests we present to be valid, the number of replicates must be the same for each treatment and must be at least 2. In addition, the observations in each treatment must constitute a random sample from a normal population, and the populations must all have the same variance. We compute the error sum of squares (SSE) by adding the sums of squared deviations from the sample means for all the treatments. To express this in an equation, let denote the sample variances of the observations in each of the eight treatments, and let K be the number of replicates per treatment. Then (9.66) Each main effect and interaction has its own sum of squares as well. These are easy to compute. The sum of squares for any effect or interaction is computed by squaring its contrast, multiplying by the number of replicates K, and dividing by the total number of treatments, which is 23 = 8. (9.67) When using Equation (9.67), it is best to keep as many digits in the effect estimates as possible, in order to obtain maximum precision in the sum of squares. For presentation in a table, effect estimates and sums of squares may be rounded to the same precision as the data. The sums of squares for the effects and interactions have one degree of freedom each. The error sum of squares has 8(K − 1) degrees of freedom. The method for computing mean squares and F statistics is the same as the one presented in Section 9.3 for a two-way ANOVA table. Each mean square is equal to its sum of squares divided by its degrees of freedom. The test statistic for testing the null hypothesis that an effect or interaction is equal to 0 is Page 746 computed by dividing the mean square for the effect estimate by the mean square for error. When the null hypothesis is true, the test statistic has an F1, 8(K−1) distribution. Example 9.24 Refer to Example 9.23. Construct an ANOVA table. For each effect and interaction, test the null hypothesis that it is equal to 0. Which factors, if any, seem most likely to have an effect on the outcome? Solution The ANOVA table follows. The sums of squares for the effects and interactions were computed by using Equation (9.67). The error sum of squares was computed by applying Equation (9.66) to the data in Example 9.23. Each F statistic is the quotient of the mean square with the mean square for error. Each F statistic has 1 and 16 degrees of freedom. Source A B C AB AC BC Effect 3.10 2.73 −0.93 −3.18 −1.34 −1.06 Sum of Squares df 57.54 1 44.72 1 5.23 1 60.48 1 10.75 1 6.76 1 Mean Square 57.54 44.72 5.23 60.48 10.75 6.76 F 7.34 5.70 0.67 7.71 1.37 0.86 P 0.015 0.030 0.426 0.013 0.259 0.367 ABC Error Total 1.07 6.83 1 125.48 16 317.78 23 6.83 0.87 0.365 7.84 The main effects of factors A and B,aswellasthe AB interaction, have fairly small Pvalues. This suggests that these effects are not equal to 0 and that factors A and B do affect the outcome. There is no evidence that the main effect of factor C, or any of its interactions, differ from 0. Further experiments might focus on factors A and B. Perhaps a two-way ANOVA would be conducted, with each of the factors A and B evaluated at several levels, to get more detailed information about their effects on the outcome. Interpreting Computer Output In practice, analyses of factorial designs are usually carried out on a computer. The following output (from MINITAB) presents the results of the analysis described in Examples 9.23 and 9.24. Page 747 Factorial Fit: Yield versus A, B, C Estimated Effects and Coefficients for Yield (coded units) Term Constant A B C A*B A*C B*C A*B*C S = 2.80040 Effect Coef SE Coef T 75.641 0.5716 132.33 3.097 1.548 0.5716 2.71 2.730 1.365 0.5716 2.39 −0.933 −0.467 0.5716 −0.82 −3.175 −1.587 0.5716 −2.78 −1.338 −0.669 0.5716 −1.17 −1.062 −0.531 0.5716 −0.93 1.067 0.533 0.5716 0.93 R-Sq = 60.51% R-Sq(adj) = 43.24% P 0.000 0.015 0.030 0.426 0.013 0.259 0.367 0.365 Analysis of Variance for Yield (coded units) Source Main Effects 2-Way Interactions 3-Way Interactions Residual Error Pure Error Total DF 3 3 1 16 16 23 Seq SS 107.480 77.993 6.827 125.476 125.476 317.776 Adj SS 107.480 77.993 6.827 125.476 125.476 Adj MS F P 35.827 4.57 0.017 25.998 3.32 0.047 6.827 0.87 0.365 7.842 7.842 The table at the top of the output presents estimated effects and coefficients. The phrase "coded units" means that the values 1 and −1, rather than the actual values, are used to represent the high and low levels of each factor. The estimated effects are listed in the column labeled "Effect." In the next column are the estimated coefficients, each of which is equal to one-half the corresponding effect. While the effect represents the difference in the mean response between the high and low levels of a factor, the coefficient represents the difference between the mean response at the high level and the grand mean response, which is half as much. The coefficient labeled "Constant" is the mean of all the observations, that is, the sample grand mean. Every coefficient estimate has the same standard deviation, which is shown in the column labeled "SE Coef." MINITAB uses the Student's t statistic, rather than the F statistic, to test the hypotheses that the effects are equal to zero. The column labeled "T" presents the value of the Student's t statistic, which is equal to the quotient of the coefficient estimate (Coef) and its standard deviation. Under the null hypothesis, the t statistic has a Student's t distribution with 8(K − 1) degrees of freedom. The P-values are presented in the column labeled "P." The t test performed by MINITAB is equivalent to the F test described in Example 9.24. The t8(K−1) statistic Page 748 can be computed by taking the square root of the F1,8(K−1) statistic and applying the sign of the effect estimate. The P-values are identical. We'll discuss the analysis of variance table next. The column labeled "DF" presents degrees of freedom. The columns labeled "Seq SS" (sequential sum of squares) and "Adj SS" (adjusted sum of squares) will be identical in all the examples we will consider and will contain sums of squares. The column labeled "Adj MS" contains mean squares, or sums of squares divided by their degrees of freedom. We will now explain the rows involving error. The row labeled "Pure Error" is concerned with the error sum of squares (SSE) (Equation 9.66). There are 8(K − 1) = 16 degrees of freedom (DF) for pure error. The sum of squares for pure error, found in each of the next two columns is the error sum of squares (SSE). Under the column "Adj MS" is the mean square for error. The row above the pure error row is labeled "Residual Error." The sum of squares for residual error is equal to the sum of squares for pure error, plus the sums of squares for any main effects or interactions that are not included in the model. The degrees of freedom for the residual error sum of squares is equal to the degrees of freedom for pure error, plus the degrees of freedom (one each) for each main effect or interaction not included in the model. Since in this example all main effects and interactions are included in the model, the residual error sum of squares and its degrees of freedom are equal to the corresponding quantities for pure error. The row labeled "Total" contains the total sum of squares (SST). The total sum of squares and its degrees of freedom are equal to the sums of the corresponding quantities for all the effects, interactions, and residual error. Going back to the top of the table, the first row is labeled "Main Effects." There are three degrees of freedom for main effects, because there are three main effects (A, B, and C), with one degree of freedom each. The sequential sum of squares is the sum of the sums of squares for each of the three main effects. The mean square (Adj MS) is the sum of squares divided by its degrees of freedom. The column labeled "F" presents the F statistic for testing the null hypothesis that all the main effects are equal to zero. The value of the F statistic (4.57) is equal to the quotient of the mean square for main effects (35.827) and the mean square for (pure) error (7.842). The degrees of freedom for the F statistic are 3 and 16, corresponding to the degrees of freedom for the two mean squares. The column labeled "P" presents the P-value for the F test. In this case the P-value is 0.017, which indicates that not all the main effects are zero. The rows labeled "2-Way Interactions" and "3-Way Interactions" are analogous to the row for main effects. The P-value for two-way interactions is 0.047, which is reasonably strong evidence that at least some of the two-way interactions are not equal to zero. Since there is only one three-way interaction (A * B * C), the P-value in the row labeled "3-Way Interactions" is the same (0.365) as the P-value in the table at the top of the MINITAB output for A * B * C. Recall that the hypothesis tests are performed under the assumption that all the observations have the same standard deviation σ. The quantity labeled "S" is the estimate of σ and is equal to the square root of the mean square for error (MSE). The quantities "R-sq" and "R-sq(adj)" are the coefficients of determination R2 and the adjusted R2, respectively, and are computed by methods analogous to those in one-way ANOVA. Page 749 Estimating Effects in a 2p Factorial Experiment A sign table can be used to obtain the formulas for computing effect estimates in any 2p factorial experiment. The method is analogous to the 23 case. The treatments are listed in a column. The sign for any main effect is + in the rows corresponding to treatments where the factor is at its high level, and − in rows corresponding to treatments where the factor is at its low level. Signs for the interactions are found by multiplying the signs corresponding to the factors in the interaction. The estimate for any effect or interaction is found by adding and subtracting the cell means for the treatments, using the signs in the appropriate columns, to compute a contrast. The contrast is then divided by half the number of treatments, or 2p−1, to obtain the effect estimate. Summary For a 2p factorial experiment: (9.68) As an example, Table 9.8 (page 750) presents a sign table for a 25 factorial experiment. We list the signs for the main effects and for selected interactions. TABLE 9.8 Sign table for the main effects and selected interactions for a 25 factorial experiment Treatment 1 a b ab c ac bc abc d ad bd abd cd acd bcd abcd e ae be abe ce ace bce abce de ade bde A B C D E + − + − + − + − + − + − + − + − + − + − + − + − + − + − + − + − + − − + + − − + + − − + + − − + + − − + + − − + + − − + + − − + + − − + − − − − + + + + − − − − + + + + − − − − + + + + − − − − + + + + − − − − − − − − + + + + + + + + − − − − − − − − + + + + + + + + − − − − − − − − − − − − − − − − + + + + + + + + + + + + + + + + AB + − − + + − − + + − − + + − − + + − − + + − − + + − − + + − − + CDE − − − − + + + + + + + + − − − − + + + + − − − − − − − − + + + + ABDE + − − + − + + − − + + − + − − + − + + − + − − + + − − + − + + − ABCDE − + + − + − − + + − − + − + + − + − − + − + + − − + + − + − − + abde cde acde bcde abcde + − + − + − + − − + − + + + + + − − + + + − + − − + Sums of squares are computed by a method analogous to that for a 23 experiment. To compute the error sum of squares (SSE), let s1, ..., s2p be the sample variances of the observations in each of the 2p treatments. Then The degrees of freedom for error is 2p(K − 1), where K is the number of replicates per treatment. The sum of squares for each effect and interaction is equal to the square of the contrast, multiplied by the number of replicates K and divided by the number of treatments 2p. The sums of squares for the effects and interactions have one degree of freedom each. (9.69) F statistics for main effects and interactions are computed by dividing the sum of squares for the effect by the mean square for error. The null distribution of the F statistic is F1, 2p(K−1). Factorial Experiments without Replication When the number of factors p is large, it is often not feasible to perform more than one replicate for each treatment. In this case, it is not possible to compute SSE, so the hypothesis tests previously described cannot be performed. If it is reasonable to assume that some of the higherorder interactions are equal to 0, then the sums of squares for those interactions can be added together and treated like an error sum of squares. Then the main effects and lower order interactions can be tested. Page 750 Example 9.25 A 25 factorial experiment was conducted to estimate the effects of five factors on the quality of lightbulbs manufactured by a certain process. The factors were A: plant (1 or 2), B: machine type (low or high speed), C: shift (day or evening), D: lead wire material (standard or new), and E: method of loading materials into the assembler (manual or automatic). One replicate was obtained for each treatment. Table 9.9 presents the results. Compute estimates of the main effects and interactions, and their sums of squares. Assume that the third-, fourth-, and fifth-order interactions are negligible, and add their sums of squares to use as a substitute for an error sum of squares. Use this substitute to test hypotheses concerning the main effects and second-order interactions. Page 751 TABLE 9.9 Treatment 1 a b ab c ac bc abc d ad bd abd cd acd bcd abcd e ae be abe ce ace bce abce de ade bde abde cde acde bcde Outcome 32.07 39.27 34.81 43.07 31.55 36.51 28.80 43.05 35.64 35.91 47.75 51.47 33.16 35.32 48.26 53.28 25.10 39.25 37.77 46.69 32.55 32.56 28.99 48.92 40.60 37.57 47.22 56.87 34.51 36.67 45.15 abcde 48.72 Solution We compute the effects, using the rules for adding and subtracting observations given by the sign table, and the sums of squares, using Equation (9.69). See Table 9.10. TABLE 9.10 Term A B C D E AB AC AD AE BC BD BE CD CE DE ABC ABD ABE ACD ACE ADE BCD BCE BDE CDE ABCD ABCE ABDE ACDE BCDE Effect 6.33 9.54 −2.07 6.70 0.58 2.84 0.18 −3.39 0.60 −0.49 4.13 0.65 −0.18 −0.81 0.24 1.35 −0.29 0.76 0.11 −0.69 −0.45 0.76 −0.82 −2.17 −1.25 −2.83 0.39 0.22 0.18 −0.25 Sum of Squares 320.05 727.52 34.16 358.72 2.66 64.52 0.27 91.67 2.83 1.95 136.54 3.42 0.26 5.23 0.46 14.47 0.67 4.59 0.088 3.75 1.60 4.67 5.43 37.63 12.48 63.96 1.22 0.37 0.24 0.52 −1.73 ABCDE 23.80 Note that none of the three-, four-, or five-way interactions are among the larger effects. If some of them were, it would not be wise to combine their sums of squares. As it is, we add the sums of squares of the three-, four-, and five-way interactions. The results are presented in the following output (from MINITAB). Factorial Fit: Response versus A, B, C, D, E Estimated Effects and Coefficients for Response (coded units) Term Constant A B C D E A*B A*C A*D A*E B*C B*D B*E C*D C*E D*E S = 3.31179 Effect Coef 39.658 6.325 3.163 9.536 4.768 −2.066 −1.033 6.696 3.348 0.576 0.288 2.840 1.420 0.183 0.091 −3.385 −1.693 0.595 0.298 −0.494 −0.247 4.131 2.066 0.654 0.327 −0.179 −0.089 −0.809 −0.404 0.239 0.119 R-Sq = 90.89% Page 752 SE Coef T P 0.5854 67.74 0.000 0.5854 5.40 0.000 0.5854 8.14 0.000 0.5854 −1.76 0.097 0.5854 5.72 0.000 0.5854 0.49 0.629 0.5854 2.43 0.027 0.5854 0.16 0.878 0.5854 −2.89 0.011 0.5854 0.51 0.618 0.5854 −0.42 0.679 0.5854 3.53 0.003 0.5854 0.56 0.584 0.5854 −0.15 0.881 0.5854 −0.69 0.500 0.5854 0.20 0.841 R-Sq(adj) = 82.34% Analysis of Variance for Response (coded units) Source DF Seq SS Adj SS Adj MS F P Main Effects 5 1443.1 1443.1 288.62 26.31 0.000 2-Way Interactions 10 307.1 307.1 30.71 2.80 0.032 Residual Error 16 175.5 175.5 10.97 Total 31 1925.7 The estimates have not changed for the main effects or two-way interactions. The residual error sum of squares (175.5) in the analysis of variance table is found by adding the sums of squares for all the higher-order interactions that were dropped from the model. The number of degrees of freedom (16) is equal to the sum of the degrees of freedom (one each) for the 16 higher-order interactions. There is no sum of squares for pure error (SSE), because there is only one replicate per treatment. The residual error sum of squares is used as a substitute for SSE to compute all the quantities that require an error sum of squares. We conclude from the output that factors A, B, and D are likely to affect the outcome. There seem to be interactions between some pairs of these factors as well. It might be appropriate to plan further experiments to focus on factors A, B, and D. Using Probability Plots to Detect Large Effects An informal method that has been suggested to help determine which effects are large is to plot the effect and interaction estimates on a normal probability plot. If in fact none of the factors affect the outcome, then the effect and interaction estimates form a simple random sample from a normal population and should lie approximately on a straight line. In many cases, most of the estimates will fall approximately on a line, and a few will plot far from the line. The Page 753 main effects and interactions whose estimates plot far from the line are the ones most likely to be important. Figure 9.13 presents a normal probability plot of the main effect and interaction estimates from the data in Example 9.25. It is clear from the plot that the main effects of factors A, B, and D, and the AB and BD interactions, stand out from the rest. FIGURE 9.13 Normal probability plot of the effect estimates from the data in Example 9.25. The main effects of factors A, B,and D stand out as being larger than the rest. Fractional Factorial Experiments When the number of factors is large, it may not be feasible to perform even one replicate for each treatment. In these cases, observations may be taken only for some fraction of the treatments. If these treatments are chosen correctly, it is still possible to obtain information about the factors. When each factor has two levels, the fraction must always be a power of 2, i.e., one-half, one-quarter, etc. An experiment in which half the treatments are used is called a half-replicate; if one-quarter of the treatments are used, it is a quarter-replicate, and so on. A half-replicate of a 2p experiment is often denoted 2p−1, to indicate that while there are p factors, there are only 2p −1 treatments being considered. Similarly, a quarter-replicate is often denoted 2p−2. We will focus on half-replicate experiments. We present a method for choosing a half-replicate of a 25 experiment. Such an experiment will have 16 treatments chosen from the 32 in the 25 experiment. To choose the 16 treatments, start with a sign table for a 24 design that shows the signs for the main effects and the highestorder interaction. This is presented as Table 9.11 (page 754). TABLE 9.11 Sign table for the main effects and four-way interaction in a 24 factorial experiment Treatment 1 a b ab c ac bc abc d ad bd abd cd acd bcd abcd A − + − + − + − + − + − + − + − + B − − + + − − + + − − + + − − + + C − − − − + + + + − − − − + + + + D − − − − − − − − + + + + + + + + ABCD + − − + − + + − − + + − + − − + Table 9.11 has the right number of treatments (16), but only four factors. To transform it into a half-replicate for a 25 design, we must introduce a fifth factor, E. We do this by replacing the highest-order interaction by E. This establishes the signs for the main effect of E. Then in each row where the sign for E is +, we add the letter e to the treatment, indicating that Page 754 factor E is to be set to its high level for that treatment. Where the sign for E is −, factor E is set to its low level. The resulting design is called the principal fraction of the 25 design. Table 9.12 presents the signs for the main effects and selected interactions of this design. TABLE 9.12 Sign table for the main effects and selected interactions for the principal fraction of a 25 factorial experiment Treatment e a b abc c ace bce abc d ade bde abd cde acd bcd abcde A − + − + − + − + − + − + − + − + B − − + + − − + + − − + + − − + + C − − − − + + + + − − − − + + + + D − − − − − − − − + + + + + + + + E = ABCD + − − + − + + − − + + − + − − + AB + − − + + − − + + − − + + − − + CDE + + + + − − − − − − − − + + + + ACDE − − + + − − + + + + − − + + − − + There is a price to be paid for using only half of the treatments. To see this, note Page 755 that in Table 9.12 the AB interaction has the same signs as the CDE interaction, and the ACDE interaction has the same signs as the main effect for B. When two effects have the same signs, they are said to be aliased. In fact, the main effects and interactions in a half-fraction form pairs in which each member of the pair is aliased with the other. The alias pairs for this half-fraction of the 25 design are {A, BCDE} {AB, CDE} {BD, ACE} {B, ACDE} {AC, BDE} {BE, ACD} {C, ABDE} {AD, BCE} {CD, ABE} {D, ABCE} {AE, BCD} {CE, ABD} {E, ABCD} {BC, ADE} {DE, ABC} When two effects are aliased, their effect estimates are the same, because they involve the same signs. In fact, when the principal fraction of a design is used, the estimate of any effect actually represents the sum of that effect and its alias. Therefore for the principal fraction of a 25 design, each main effect estimate actually represents the sum of the main effect plus its aliased four-way interaction, and each two-way interaction estimate represents the sum of the two-way interaction and its aliased three-way interaction. In many cases, it is reasonable to assume that the higher-order interactions are small. In the 5 2 half-replicate, if the four-way interactions are negligible, the main effect estimates will be accurate. If in addition the three-way interactions are negligible, the two-way interaction estimates will be accurate as well. In a fractional design without replication, there is often no good way to compute an error sum of squares, and therefore no rigorous way to test the hypotheses that the effects are equal to 0. In many cases, the purpose of a fractional design is simply to identify a few factors that appear to have the greatest impact on the outcome. This information may then be used to design more elaborate experiments to investigate these factors. For this purpose, it may be enough simply to choose those factors whose effects or two-way interactions are unusually large, without performing hypothesis tests. This can be done by listing the estimates in decreasing order, and then looking to see if there are a few that are noticeably larger than the rest. Another method is to plot the effect and interaction estimates on a normal probability plot, as previously discussed. Example 9.26 In an emulsion liquid membrane system, an emulsion (internal phase) is dispersed into an external liquid medium containing a contaminant. The contaminant is removed from the external liquid through mass transfer into the emulsion. Internal phase leakage occurs when portions of the extracted material spill into the external liquid. In the article "Leakage and Swell in Emulsion Liquid Membrane Systems: Batch Experiments" (R. Pfeiffer, W. Navidi, and A. Bunge, Separation Science and Technology, 2003:519–539), the effects of five factors were studied to determine the effect on leakage in a certain system. The five factors were A: surfactant concentration, B: internal phase lithium hydroxide concentration, C: membrane phase, D: internal phase volume fraction, and E: extraction vessel stirring rate. A half-fraction of a 25 design was used. The data are presented in the following table (in the actual experiment, each point actually represented the average of two measurements). Page 756 Leakage is measured in units of percent. Assume that the third-, fourth-, and fifthorder interactions are negligible. Estimate the main effects and two-way interactions. Which, if any, stand out as being noticeably larger than the rest? Treatment e Leakage 0.61 a b abc c ace bce abc d ade bde abd cde acd bcd abcde 0.13 2.23 0.095 0.35 0.075 7.31 0.080 2.03 0.64 11.72 0.56 1.45 0.31 1.33 6.24 Solution Using the sign table (Table 9.12), we compute estimates for the main effects and two-way interactions, shown in the following table. Term A B C D E AB AC AD AE BC BD BE CD CE DE Effect −2.36 3.00 −0.11 1.68 2.64 −1.54 1.43 0.17 −1.15 0.20 0.86 2.65 −1.30 0.61 1.32 Note that we do not bother to compute sums of squares for the estimates, because we have no SSE to compare them to. To determine informally which effects may be most worthy of further investigation, we rank the estimates in order of their absolute values: B: 3.00, BE: 2.65, E: 2.64, A: −2.36, D: 1.68, and so forth. It seems reasonable to decide that there is a fairly wide gap between the A and D effects, and therefore that factors A, B, and E are most likely to be important. Exercises for Section 9.5 1. 2. 3. Construct a sign table for the principal fraction for a 24 design. Then indicate all the alias pairs. Give an example of a factorial experiment in which failure to randomize can produce incorrect results. A chemical reaction was run using two levels each of temperature (A), reagent concentration (B), and pH (C). For each factor, the high level is denoted 1, and the low level is denoted −1. The reaction was run twice for each combination of levels, and the yield (in percent) was recorded. The results were as follows. Page 757 A −1 1 −1 1 −1 1 −1 1 B −1 −1 1 1 −1 −1 1 1 C −1 −1 −1 −1 1 1 1 1 Yields 74, 71 73, 74 78, 74 86, 89 71, 68 77, 77 75, 85, 92 Mean Yield 72.5 75.5 76.0 87.5 69.5 77.0 80.0 87.0 Compute estimates of the main effects and interactions, and the sum of squares and Pvalue for each. a. b. Which main effects and interactions, if any, are important? Other things being equal, will the mean yield be higher when the temperature is high or low? Explain. The article "Efficient Pyruvate Production by a Multi-Vitamin Auxotroph of Torulopsis glabrata: Key Role and Optimization of Vitamin Levels" (Y. Li, J. Chen, et al., Applied Microbiology and Biotechnology, 2001:680–685) investigates the effects of the levels of several vitamins in a cell culture on the yield (in g/L) of pyruvate, a useful organic acid. The data in the following table are presented as two replicates of a 23 design. The factors are A: nicotinic acid, B: thiamine, and C: biotin. (Two statistically insignificant factors have been dropped. In the article, each factor was tested at four levels; we have collapsed these to two.) A −1 B −1 C −1 Yields 0.55, 0.49 Mean Yield 0.520 1 −1 1 −1 1 1 −1 1 −1 −1 1 −1 1 −1 −1 1 1 −1 1 1 1 0.60, 0.42 0.37, 0.28 0.30, 0.28 0.54, 0.54 0.54, 0.47 0.44, 0.33 0.36, 0.20 0.510 0.325 0.290 0.540 0.505 0.385 0.280 Compute estimates of the main effects and interactions, and the sum of squares and Pvalue for each. a. b. Is the additive model appropriate? c. What conclusions about the factors can be drawn from these results? The article cited in Exercise 4 also investigated the effect of the factors on glucose consumption (in g/L). A single measurement is provided for each combination of factors (in the article, there was some replication). The results are presented in the following table. a. 5. A −1 1 −1 1 −1 1 −1 1 B −1 −1 1 1 −1 −1 1 1 C −1 −1 −1 −1 1 1 1 1 Glucose Consumption 68.0 77.5 98.0 98.0 74.0 77.0 97.0 98.0 Compute estimates of the main effects and the interactions. Page 758 b. c. d. 6. Is it possible to compute an error sum of squares? Explain. Are any of the interactions among the larger effects? If so, which ones? Assume that it is known from past experience that the additive model holds. Add the sums of squares for the interactions, and use that result in place of an error sum of squares to test the hypotheses that the main effects are equal to 0. A metal casting process for the production of turbine blades was studied. Three factors were varied. They were A: the temperature of the metal, B: the temperature of the mold, and C: the pour speed. The outcome was the thickness of the blades, in mm. The results are presented in the following table. A −1 1 −1 1 −1 1 −1 1 B −1 −1 1 1 −1 −1 1 1 C −1 −1 −1 −1 1 1 1 1 Thickness 4.61 4.51 4.60 1 −1 1 −1 1 −1 1 −1 1 1 −1 −1 1 1 1 1 4.54 4.61 4.61 4.48 4.51 a. b. c. 7. Compute estimates of the main effects and the interactions. Is it possible to compute an error sum of squares? Explain. Plot the estimates on a normal probability plot. Does the plot show that some of the factors influence the thickness? Explain. The article "An Investigation into the Ball Burnishing of Aluminium Alloy 6061-T6" (M. El-Axir, J Engineering Manufacture, 2007:1733–1742) presents the results of study that investigated the effects of three burnishing factors on the reduction in diameter of the workpiece (in μm). The factors are A: Burnishing speed, B: Burnishing force, and C: Burnishing feed. The results presented in the following table form a 23 factorial design (some additional results are omitted). A −1 1 −1 1 −1 1 −1 1 a. b. 8. 9. B −1 −1 1 1 −1 −1 1 1 C −1 −1 −1 −1 1 1 1 1 Reduction 570 353 778 769 544 319 651 625 Compute estimates of the main effects and the interactions. Is it possible to compute an error sum of squares? Explain. Are any of the interactions among the larger effects? If so, which ones? Someone claims that the additive model holds. Do the results tend to support this statement? Explain. Page 759 In a 2p design with one replicate per treatment, it sometimes happens that the observation for one of the treatments is missing, due to experimental error or to some other cause. When this happens, one approach is to replace the missing value with the value that makes the highest-order interaction equal to 0. Refer to Exercise 7. Assume the observation for the treatment where A, B, and C are all at their low level (−1) is missing. a. What value for this observation makes the three-way interaction equal to 0? b. Using this value, compute estimates for the main effects and the interactions. Safety considerations are important in the design of automobiles. The article "An Optimum Design Methodology Development Using a Statistical Technique for Vehicle Occupant Safety" (J. Hong, M. Mun, and S. Song, Proceedings of the Institution of Mechanical Engineers, 2001:795–801) presents results from an occupant simulation study. The outcome variable is chest acceleration (in g) 3 ms after impact. Four factors were considered. They were A: the airbag vent characteristic, B: the airbag inflator trigger time, C: the airbag inflator mass flow rate, and D: the stress-strain relationship of knee foam. The results (part of a larger study) are presented in the following table. There is one replicate per treatment. Treatment 1 a b ab c ac bc abc d ad bd abd cd acd bcd abcd Outcome 85.2 79.2 84.3 89.0 66.0 69.0 68.5 76.4 85.0 82.0 84.7 82.2 62.6 65.4 66.3 69.0 a. b. Compute estimates of the main effects and the interactions. If you were to design a follow-up study, which factor or factors would you focus on? Explain. 10. In a small-disc test a small, disc-shaped portion of a component is loaded until failure. The article "Optimizing the Sensitivity of the Small-Disc Creep Test to Damage and Test Conditions" (M. Evans and D. Wang, J. Strain Analysis, 2007:389–413) presents the results of a factorial experiment to estimate the effects of properties of the disc on the time to failure (in ms). The data in the following table are presented as a 25 design. The factors are A: hole diameter, B: disc diameter, C: disc thickness, D: punch head radius, and E: friction coefficient. Two other factors discussed in the article are not considered here. Treatment 1 a b ab c ac bc abc d ad bd abd cd acd bcd abcd e ae be abe ce ace bce abce de ade bde abde cde acde bcde abcde Outcome 2486.8 1328.1 2470.2 1303.2 6817.4 3845.2 6413.8 3827.2 2912.3 1507.2 2885.3 1491.8 7723.0 4289.3 7952.8 4505.5 2508.6 1319.4 2446.8 1303.3 6864.7 3875.0 6994.2 3961.2 2915.0 1536.7 2872.8 1477.9 7731.6 4345.1 7969.1 4494.5 a. b. Compute estimates of the main effects and the interactions. If you were to design a follow-up study, which factors would you focus on? Why? Page 760 11. The article "Factorial Design for Column Flotation of Phosphate Wastes" (N. AbdelKhalek, Particulate Science and Technology, 2000:57–70) describes a 23 factorial design to investigate the effect of superficial air velocity (A), frothier concentration (B), and superficial wash water velocity (C) on the percent recovery of P2O5. There were two replicates. The data are presented in the following table. A −1 1 −1 1 −1 1 −1 1 B −1 −1 1 1 −1 −1 1 1 C −1 −1 −1 −1 1 1 1 1 Percent 56.30 70.10 65.60 80.20 Recovery 54.85 72.70 63.60 78.80 −1 1 −1 1 −1 1 1 1 −1 1 1 1 50.30 65.30 60.50 70.63 48.95 66.00 59.50 69.86 a. Compute estimates of the main effects and interactions, along with their sums of squares and P-values. b. Which factors seem to be most important? Do the important factors interact? Explain. 12, The article "An Application of Fractional Factorial Designs" (M. Kilgo, Quality Engineering, 1988:19–23) describes a25−1 design (half-replicate of a 25 design) involving the use of carbon dioxide (CO2) at high pressure to extract oil from peanuts. The outcomes were the solubility of the peanut oil in the CO2 (in mg oil/liter CO2), and the yield of peanut oil (in percent). The five factors were A: CO2 pressure, B: CO2 temperature, C: peanut moisture, D: CO2 flow rate, and E: peanut particle size. The results are presented in the following table. Treatment e a b abe c ace bce abc d ade bde abd cde acd bcd abcde a. b. c. d. Solubility 29.2 23.0 37.0 139.7 23.3 38.3 42.6 141.4 22.4 37.2 31.3 48.6 22.9 36.2 33.6 172.6 Yield 63 21 36 99 24 66 71 54 23 74 80 33 63 21 44 96 Assuming third- and higher-order interactions to be negligible, compute the main effects and interactions for the solubility outcome. Plot the estimates on a normal probability plot. Does the plot show that some factors influence the solubility? If so, which ones? Assuming third- and higher-order interactions to be negligible, compute the main effects and interactions for the yield outcome. Plot the estimates on a normal probability plot. Does the plot show that some of the factors influence the yield? If so, which ones? 13. In a 25−1 design (such as the one in Exercise 12) what does the estimate of the main effect of factor A actually represent? Page 761 i. The main effect of A. ii. The sum of the main effect of A and the BCDE interaction. iii. The difference between the main effect of A and the BCDE interaction. iv. The interaction between A and BCDE. Supplementary Exercises for Chapter 9 1. The article "Gypsum Effect on the Aggregate Size and Geometry of Three Sodic Soils Under Reclamation" (I. Lebron, D. Suarez, and T. Yoshida, Journal of the Soil Science Society of America, 2002:92–98) reports on an experiment in which gypsum was added in various amounts to soil samples before leaching. One of the outcomes of interest was the pH of the soil. Gypsum was added in four different amounts. Three soil samples received each amount added. The pH measurements of the samples are presented in the following table. Gypsum (g/kg) 0.00 0.11 0.19 0.38 2. 7.68 7.85 7.87 8.00 Can you conclude that the pH differs with the amount of gypsum added? Provide the value of the test statistic and the P-value. The article referred to in Exercise 1 also considered the effect of gypsum on the electric conductivity (in dS m−1) of soil. Two types of soil were each treated with three different amounts of gypsum, with two replicates for each soil-gypsum combination. The data are presented in the following table. Gypsum (g/kg) 0.00 0.27 0.46 a. 3. 7.88 7.81 7.84 7.80 pH 7.72 7.64 7.63 7.73 Soil Type Las Animas Madera 1.52 1.05 1.01 0.92 1.49 0.91 1.12 0.92 0.99 0.92 0.88 0.92 Is there convincing evidence of an interaction between the amount of gypsum and soil type? b. Can you conclude that the conductivity differs among the soil types? c. Can you conclude that the conductivity differs with the amount of gypsum added? Penicillin is produced by the Penicillium fungus, which is grown in a broth whose sugar content must be carefully controlled. Several samples of broth were taken on each of three successive days, and the amount of dissolved sugars (in mg/mL) was measured on each sample. The results were as follows: Day 1: Day 2: Day 3: 4.8 5.4 5.7 5.1 5.0 5.1 5.1 5.0 5.3 4.8 5.1 5.5 5.2 5.2 5.3 4.9 5.1 5.5 5.1 5.3 5.5 5.0 5.3 5.1 4.9 5.2 5.6 5.0 5.2 5.3 4.8 5.1 5.2 4.8 5.4 5.5 5.1 5.2 5.3 5.0 5.1 5.2 Can you conclude that the mean sugar concentration differs among the three days? Page 762 4. The following MINITAB output is for a two-way ANOVA. Something went wrong with the printer, and some of the numbers weren't printed. Two-way Analysis of Variance Analysis of Variance Source Row Column Interaction Error Total 5. Variance DF 3 2 6 (a) 23 SS 145.375 15.042 (b) (c) 217.870 MS (d) (e) 4.2000 (f) F (g) (h) (i) F (j) 6 (k) (l) Fill in the missing numbers in the table for (a) through (l). You may give ranges for the P-values. An experiment was performed to determine whether different types of chocolate take different amounts of time to dissolve. Forty people were divided into five groups. Each group was assigned a certain type of chocolate. Each person dissolved one piece of chocolate, and the dissolve time (in seconds) was recorded. For comparison, each person in each group also dissolved one piece of butterscotch candy; these pieces were identical for all groups. The data, which include the group, the dissolve times for both chocolate and butterscotch, the difference between the dissolve times, and the ratio of the dissolve times, are presented in the following table. Note that the design is slightly unbalanced; group 3 has nine people and group 5 has only seven. Page 763 Group 1 1 1 1 1 1 1 Chocolate 135 865 122 110 71 81 2405 Butterscotch 60 635 63 75 37 58 1105 Difference 75 230 59 35 34 23 1300 Ratio 2.25 1.36 1.94 1.47 1.92 1.40 2.18 1 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 a. b. c. 242 42 30 104 124 75 80 255 71 51 47 90 65 27 105 90 54 93 48 85 82 46 64 125 69 73 105 99 45 137 170 153 49 135 38 30 110 118 40 91 122 71 41 15 82 71 28 112 51 50 71 45 71 121 73 60 150 150 ...

Construct a one-way ANOVA table, using the dissolve times for chocolate. Can you conclude that the mean dissolve time for chocolate differs among groups? Explain. d. Do these data appear to satisfy the assumptions for a oneway ANOVA? Explain. e. Someone else suggests using the differences (Chocolate − Butterscotch). Do these data appear to satisfy the assumptions for a one-way ANOVA? Explain. Perform a one-way analysis of variance using the ratios. Can you conclude that the 6. mean ratio of dissolve times differs for different types of chocolate? The article "Stability of Silico-Ferrite of Calcium and Aluminum (SFCA) in Air-Solid Solution Limits Between 1240°C and 1390°C and Phase Relationships within the Fe2O3CaO-Al2O3-SiO2 (FCAS) System" (T. Patrick and M. Pownceby, Metallurgical and Materials Transactions B, 2002:79–90) investigates properties of silico-ferrites of calcium and aluminum (SFCA). The data in the following table present the ratio of the weights of Fe2O3 and CaO for SFCA specimens with several different weight percents of Al2O3 and C4S3. Al2O3(%) 1.0 1.0 1.0 5.0 5.0 5.0 10.0 10.0 10.0 a. b. C4S3 Fe2O3/CaO Low (3%−6%) Medium (7%−10%) High (11%−14%) Low (3%−6%) Medium (7%−10%) High (11%−14%) Low (3%−6%) Medium (7%−10%) High (11%−14%) 7.25 6.03 5.10 6.92 5.74 4.84 6.50 5.37 4.52 6.92 5.78 4.65 6.38 5.26 4.65 6.18 5.14 4.33 6.60 5.54 4.71 6.29 5.04 4.50 6.06 4.88 4.67 6.72 5.01 4.43 5.81 4.84 4.67 5.61 4.46 3.99 Estimate all main effects and interactions. Construct an ANOVA table. You may give ranges for the P-values. Page 764 c. Do the data indicate that there are any interactions between the weight percent of Al2O3 and the weight percent of C4S3? Explain. d. Do the data convincingly demonstrate that the Fe2O3/CaO ratio depends on the weight percent of Al2O3? Explain. e. Do the data convincingly demonstrate that the Fe2O3/CaO ratio depends on the weight percent of C4S3? Explain. A component can be manufactured according to either of two designs and with either a more expensive or a less expensive material. Several components are manufactured with each combination of design and material, and the lifetimes of each are measured (in hours). A two-way analysis of variance was performed to estimate the effects of design and material on component lifetime. The cell means and main effect estimates are presented in the following table. Main Effects More expensive Less expensive Design 1 Design 2 14 −14 −16 16 ANOVA table Source Material Design Interaction Error Total 8. DF 1 1 1 8 11 SS 2352.0 3072.0 2700.0 1800.0 9924.0 MS 2352.0 3072.0 2700.0 225.00 F 10.45 13.65 12.00 P 0.012 0.006 0.009 The process engineer recommends that design 2 should be used along with the more expensive material. He argues that the main effects of both design 2 and the more expensive material are positive, so using this combination will result in the longest component life. Do you agree with the recommendation? Why or why not? The article "Case Study Based Instruction of DOE and SPC" (J. Brady and T. Allen, The American Statistician, 2002:312–315) presents the results of a 24−1 factorial experiment to investigate the effects of four factors on the yield of a process that manufactures printed circuit boards. The factors were A: transistor power output (upper or lower specification limit), B: transistor mounting approach (screwed or soldered), C: transistor heat sink type (current or alternative configuration), and D: screw position on the frequency adjustor (onehalf or two turns). The results are presented in the following table. The yield is a percent of a theoretical maximum. A −1 1 −1 1 −1 1 −1 1 B −1 −1 1 1 −1 −1 1 1 C −1 −1 −1 −1 1 1 1 1 D −1 1 1 −1 1 −1 −1 1 Yield 79.8 69.0 72.3 71.2 91.3 95.4 92.7 91.5 a. Page 765 a. Estimate the main effects of each of the four factors. Assuming all interactions to be negligible, pool the sums of squares for interaction to use in place of an error sum of squares. c. Which of the four factors, if any, can you conclude to have an effect on the yield? What is the P-value of the relevant test? The article "Combined Analysis of Real-Time Kinematic GPS Equipment and Its Users for Height Determination" (W. Featherstone and M. Stewart, Journal of Surveying Engineering, 2001:31–51) presents a study of the accuracy of global positioning system (GPS) equipment in measuring heights. Three types of equipment were studied, and each was used to make measurements at four different base stations (in the article a fifth station was included, for which the results differed considerably from the other four). There were 60 measurements made with each piece of equipment at each base. The means and standard deviations of the measurement errors (in mm) are presented in the following table for each combination of equipment type and base station. b. 9. Instrument A Instrument B Instrument C Standard Standard Standard Mean Deviation Mean Deviation Mean Deviation Base 3 15 −24 18 −6 18 0 Base 14 26 −13 13 −2 16 1 Base 2 1 26 −22 39 4 29 2 Base 8 34 −17 26 15 18 3 Construct an ANOVA table. You may give ranges for the P-values. The question of interest is whether the mean error differs among instruments. It is not of interest to determine whether the error differs among base stations. For this reason, a surveyor suggests treating this as a randomized complete block design, with the base stations as the blocks. Is this appropriate? Explain. 10. Vermont maple sugar producers sponsored a testing program to determine the benefit of a potential new fertilizer regimen. A random sample of 27 maple trees in Vermont were chosen and treated with one of three levels of fertilizer suggested by the chemical producer. In this experimental setup, nine trees (three in each of three climatic zones) were treated with each fertilizer level and the amount of sap produced (in mL) by the trees in the subsequent season was measured. The results are presented in the following table. a. b. Low fertilizer Medium fertilizer High fertilizer Southern Zone Central Zone Northern Zone 76.2 80.4 74.2 79.4 87.9 86.9 84.5 85.2 80.1 87.0 95.1 93.0 98.2 94.7 96.2 88.4 90.4 92.2 84.2 87.5 83.1 90.3 89.9 93.2 81.4 84.7 82.2 Estimate the main effects of fertilizer levels and climatic zone, and their interactions. Construct an ANOVA table. You may give ranges for the P-values. c. Test the hypothesis that there is no interaction between fertilizer levels and climatic zone. d. Test the hypothesis that there is no difference in sap production for the three fertilizer levels. 11. A civil engineer is interested in several designs for a drainage canal used to divert floodwaters from around a city. The drainage times of a reservoir attached to each of five different channel designs obtained from a series of experiments using similar initial flow conditions are given in the following table. Page 766 Channel Type 1 2 3 4 5 41.4 37.7 32.6 27.3 44.9 Drainage time (min) 43.4 50.0 49.5 52.1 33.7 34.8 29.9 32.3 37.2 33.1 24.8 37.1 a. Can you conclude that there is a difference in the mean drainage times for the different channel designs? b. Which pairs of designs, if any, can you conclude to differ in their mean drainage times? 12. A process that manufactures vinyl for automobile seat covers was studied. Three factors were varied: the proportion of a certain plasticizer (A), the rate of extrusion (B), and the temperature of drying (C). The outcome of interest was the thickness of the vinyl (in mils). A 23 factorial design with four replicates was employed. The results are presented in the following table. (Based on the article "Split-Plot Designs and Estimation Methods for Mixture Experiments with Process Variables," S. Kowalski, J. Cornell, and G. Vining, Technometrics, 2002:72–79.) A −1 1 −1 1 −1 1 −1 1 a. B −1 −1 1 1 −1 −1 1 1 C −1 −1 −1 −1 1 1 1 1 Estimate all main effects and interactions. b. Thickness 7 5 6 7 6 5 5 5 8 5 6 6 7 6 5 8 6 4 5 6 5 7 11 7 5 6 9 5 10 8 9 b. Construct an ANOVA table. You may give ranges for the P-values. c. Is the additive model appropriate? Explain. d. What conclusions about the factors can be drawn from these results? 13. In the article "Occurrence and Distribution of Ammonium in Iowa Groundwater" (K. Schilling, Water Environment Research, 2002:177–186), ammonium concentrations (in mg/L) were measured at a large number of wells in the state of Iowa. These included five types of bedrock wells. The number of wells of each type, along with the mean and standard deviation of the concentrations in those wells, is presented in the following table. Well Type Cretaceous Mississippian Devonian Silurian CambrianOrdovician Sample Size 53 57 66 67 51 Mean 0.75 0.90 0.68 0.50 0.82 Standard Deviation 0.90 0.92 1.03 0.97 0.89 Can you conclude that the mean concentration differs among the five types of wells? 14. The article "Enthalpies and Entropies of Transfer of Electrolytes and Ions from Water to Mixed Aqueous Organic Solvents" (G. Hefter, Y. Marcus, and W. Waghorne, Chemical Reviews, 2002:2773–2836) presents measurements of entropy and enthalpy Page 767 changes for many salts under a variety of conditions. The following table presents the results for enthalpy of transfer (in kJ/mol) from water to water + methanol of NaCl (table salt) for several concentrations of methanol. Four independent measurements were made at each concentration. Concentration (%) 5 10 20 30 1.62 2.69 3.56 3.35 Enthalpy 1.60 1.62 2.66 2.72 3.45 3.65 3.18 3.40 1.66 2.73 3.52 3.06 a. Is it plausible that the enthalpy is the same at all concentrations? Explain. b. Which pairs of concentrations, if any, can you conclude to have differing enthalpies? 15. Refer to Exercise 11. a. Compute the quantity , the estimate of the error standard deviation σ. b. Assuming s to be the error standard deviation, find the sample size necessary in each treatment to provide a power of 0.90 to detect a maximum difference of 10 in the treatment means at the 5% level. c. Using a more conservative estimate of 1.5s as the error standard deviation, find the sample size necessary in each treatment to provide a power of 0.90 to detect a maximum difference of 10 in the treatment means at the 5% level. 16. Refer to Exercise 14. a. Compute the quantity , the estimate of the error standard deviation σ. b. Assuming s to be the error standard deviation, find the sample size necessary in each treatment to provide a power of 0.80 to detect a maximum difference of 0.1 in the treatment means at the 5% level. c. Using a more conservative estimate of 1.5s as the error standard deviation, find the sample size necessary in each treatment to provide a power of 0.80 to detect a maximum difference of 0.1 in the treatment means at the 5% level. 17. The article "Factorial Experiments in the Optimization of Alkaline Wastewater Pretreatment" (M. Prisciandaro, A. Del Borghi, and F. Veglio, Industrial Engineering and Chemistry Research, 2002:5034–5041) presents the results of several experiments to investigate methods of treating alkaline wastewater. One experiment was an unreplicated 24 design. The four factors were A: concentration of sulfuric acid, B: temperature, C: time, and D: concentration of calcium chloride. The outcome variable is the amount of precipitate in kg/m3. The results are presented in the following table. A −1 1 −1 1 −1 1 −1 1 −1 1 −1 1 −1 1 −1 1 B −1 −1 1 1 −1 −1 1 1 −1 −1 1 1 −1 −1 1 1 C −1 −1 −1 −1 1 1 1 1 −1 −1 −1 −1 1 1 1 1 D −1 −1 −1 −1 −1 −1 −1 −1 1 1 1 1 1 1 1 1 Outcome 6.4 12.9 8.6 12.9 7.4 12.0 10.7 15.0 11.9 13.1 12.1 16.4 15.5 18.3 Page 768 a. b. c. d. Estimate all main effects and interactions. Which effects seem to be larger than the others? Assume that all third- and higher-order interactions are equal to 0, and add their sums of squares. Use the result in place of an error sum of squares to compute F statistics and P-values for the main effects. Which factors can you conclude to have an effect on the outcome? The article described some replicates of the experiment, in which the error mean square was found to be 1.04, with four degrees of freedom. Using this value, compute F statistics and P-values for all main effects and interactions. e. Do the results of part (d) help to justify the assumption that the third- and higher-order interactions are equal to 0? Explain. f. Using the results of part (d), which factors can you conclude to have an effect on the outcome? 18. The Williamsburg Bridge is a suspension bridge that spans the East River, connecting the boroughs of Brooklyn and Manhattan in New York City. An assessment of the strengths of its cables is reported in the article "Estimating Strength of the Williamsburg Bridge Cables" (R. Perry, The American Statistician, 2002:211–217). Each suspension cable consists of 7696 wires. From one of the cables, wires were sampled from 128 points. These points came from four locations along the length of the cable (I, II, III, IV). At each location there were eight equally spaced points around the circumference of the cable (A, B, C, D, E, F, G, H). At each of the eight points, wires were sampled from four depths: (1) the external surface of the cable, (2) two inches deep, (3) four inches deep, and (4) seven inches deep (the cable is 9.625 inches in radius). Under assumptions made in the article, it is appropriate to consider this as a two-factor experiment with circumferential position and depth as the factors, and with location providing four replicates for each combination of these factors. The minimum breaking strength (in lbf) is presented in the following table for each of the 128 points. Circumference A A A A B B B B C C C C D D D D E E E E Depth 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 I 6250 6650 5390 6510 6200 6430 5710 6510 5940 6220 6430 6190 5560 5800 6140 5330 ... Location II III IV ... Page 769 H H H H 1 2 3 4 5950 6180 6560 6790 5950 6560 6960 6450 5730 6450 6670 5870 6790 6550 6790 ... Construct an ANOVA table. You may give ranges for the P-values. a. b. Can you conclude that there are interactions between circumferential position and depth? Explain. c. Can you conclude that the strength varies with circumferential position? Explain. d. Can you conclude that the strength varies with depth? Explain. 19. In the article "Nitrate Contamination of Alluvial Groundwaters in the Nakdong River Basin, Korea" (J. Min, S. Yun, et al., Geosciences Journal, 2002:35–46), several chemical properties were measured for water samples taken from irrigation wells at three locations. The following table presents the means, standard deviations, and sample sizes for pH measurements. a. b. Location Upstream Midstream Downstream Mean 6.0 6.2 6.4 SD 0.2 0.4 0.6 Sample Size 49 31 30 Do the data prove conclusively that the pH differs at the different locations? 20. The article cited in Exercise 19 provides measures of electrical conductivity (in μS/cm). The results are presented in the following table. Location Upstream Midstream Downstream Mean 463 363 647 SD 208 98 878 Sample Size 49 31 30 Can a one-way analysis of variance be used to determine whether conductivity varies with location? Or is one of the necessary assumptions violated? Explain. 21. The article "Factorial Experiments in the Optimization of Alkaline Wastewater Pretreatment" (M. Prisciandaro, A. Del Borghi, and F. Veglio, Industrial Engineering and Chemistry Research, 2002:5034–5041) presents the results of an experiment to Page 770 investigate the effects of the concentrations of sulfuric acid (H2SO4) and calcium chloride (CaCl2) on the amount of black mud precipitate in the treatment of alkaline wastewater. There were three levels of each concentration, and two replicates of the experiment were made at each combination of levels. The results are presented in the following table (all measurements are in units of kg/m3). H2SO4 CaCl2 110 110 110 123 123 123 136 136 136 15 30 45 15 30 45 15 30 45 Precipitate 100.2 98.2 175.8 176.2 216.5 206.0 110.5 105.1 184.7 189.0 234.0 222.0 106.5 107.0 181.7 189.0 ... b. c. a. Construct an ANOVA table. You may give ranges for the P-values. Is the additive model plausible? Explain. Can you conclude that H2SO4 concentration affects the amount of precipitate? Explain. d. Can you conclude that CaCl2 concentration affects the amount of precipitate? Explain. 22. Fluid inclusions are microscopic volumes of fluid that are trapped in rock during rock formation. The article "Fluid Inclusion Study of Metamorphic Gold-Quartz Veins in Northwestern Nevada, U.S.A.: Characteristics of Tectonically Induced Fluid" (S. Cheong, Geosciences Journal, 2002:103–115) describes the geochemical properties of fluid inclusions in several different veins in northwest Nevada. The following table presents data on the maximum salinity (% NaCl by weight) of inclusions in several rock samples from several areas. Area Humboldt Range Santa Rosa Range Ten Mile Antelope Range Pine Forest Range 9.2 5.2 7.9 6.7 10.5 10.0 6.1 6.7 8.4 6.7 Salinity 11.2 8.8 8.3 9.5 7.3 9.9 17.5 15.3 Can you conclude that the salinity differs among the areas? 10.4 20.0 7.0 23. The article "Effect of Microstructure and Weathering on the Strength Anisotropy of Porous Rhyolite" (Y. Matsukura, K. Hashizume, and C. Oguchi, Engineering Geology, 2002:39– 17) investigated the relationship between the angle between cleavage and flow structure and the strength of porous rhyolite. Strengths (in MPa) were measured for a number of specimens cut at various angles. The mean and standard deviation of the strengths for each angle are presented in the following table. Page 771 Angle 0° 15° 30° 45° 60° 75° 90° Mean 22.9 22.9 19.7 14.9 13.5 11.9 14.3 Standard Deviation 2.98 1.16 3.00 2.99 2.33 2.10 3.95 Sample Size 12 6 4 5 7 6 6 Can you conclude that strength varies with the angle? 24. The article "Influence of Supplemental Acetate on Bioremediation for Dissolved Polycyclic Aromatic Hydrocarbons" (T. Ebihara and P. Bishop, Journal of Environmental Engineering, 2002:505–513) describes experiments in which water containing dissolved polyaromatic hydrocarbons (PAH) was fed into sand columns. PAH concentrations were measured at various depths after 25, 45, and 90 days. Assume that three independent measurements were made at each depth at each time. The data presented in the following table are naphthalene concentrations (in mg/L) that are consistent with means and standard deviations reported in the article. Depth 0 5 15 30 50 75 a. b. 25 days 11.15 11.39 11.36 14.40 11.78 11.92 11.51 11.01 11.09 12.77 12.18 11.65 11.71 11.29 11.20 11.18 11.45 11.47 45 days 9.28 8.15 8.59 9.44 9.34 9.33 9.34 9.11 8.94 9.37 9.27 9.05 9.25 8.97 9.29 9.09 8.66 8.78 90 days 7.68 7.59 7.41 7.53 7.92 7.12 7.43 7.47 7.53 7.60 7.48 7.84 7.76 7.84 7.68 7.72 7.61 7.74 Construct an ANOVA table. You may give ranges for the P-values. Perform a test to determine whether the additive model is plausible. Provide the value of the test statistic and the P-value. Page 772 Chapter 10 Statistical Quality Control Introduction As the marketplace for industrial goods has become more global, manufacturers have realized that the quality and reliability of their products must be as high as possible for them to be competitive. It is now generally recognized that the most cost-effective way to maintain high quality is through constant monitoring of the production process. This monitoring is often done by sampling units of production and measuring some quality characteristic. Because the units are sampled from some larger population, these methods are inherently statistical in nature. One of the early pioneers in the area of statistical quality control was Dr. Walter A. Shewart of the Bell Telephone Laboratories. In 1924, he developed the modern control chart, which remains one of the most widely used tools for quality control to this day. After World War II, W. Edwards Deming was instrumental in stimulating interest in quality control, first in Japan, and then in the United States and other countries. The Japanese scientist Genichi Taguchi played a major role as well, developing methods of experimental design with a view toward improving quality. In this chapter, we will focus on the Shewart control charts and on cumulative sum (CUSUM) charts, since these are among the most powerful of the commonly used tools for statistical quality control. 10.1 Basic Ideas The basic principle of control charts is that in any process there will always be variation in the output. Some of this variation will be due to causes that are inherent in the process and are difficult or impossible to specify. These causes are called common causes or chance causes. When common causes are the only causes of variation, the process is said to be in a state of statistical control, or, more simply, in control. Page 773 Sometimes special factors are present that produce additional variability. Machines that are malfunctioning, operator error, fluctuations in ambient conditions, and variations in the properties of raw materials are among the most common of these factors. These are called special causes or assignable causes. Special causes generally produce a higher level of variability than do common causes; this variability is considered to be unacceptable. When a process is operating in the presence of one or more special causes, it is said to be out of statistical control. Control charts enable the quality engineer to decide whether a process appears to be in control, or whether one or more special causes are present. If the process is found to be out of control, the nature of the special cause must be determined and corrected, so as to return the process to a state of statistical control. There are several types of control charts; which ones are used depend on whether the quality characteristic being measured is a continuous variable, a binary variable, or a count variable. For example, when monitoring a process that manufactures aluminum beverage cans, the height of each can in a sample might be measured. Height is a continuous variable. In some situations, it might be sufficient simply to determine whether the height falls within some specification limits. In this case the quality measurement takes on one of only two values: conforming (within the limits) or nonconforming (not within the limits). This measurement is a binary variable, since it has two possible values. Finally, we might be interested in counting the number of flaws on the surface of the can. This is a count variable. Control charts used for continuous variables are called variables control charts. Examples include the chart, the R chart, and the S chart. Control charts used for binary or count variables are called attribute control charts. The p chart is most commonly used for binary variables, while the c chart is commonly used for count variables. Collecting Data–Rational Subgroups Data to be used in the construction of a control chart are collected in a number of samples, taken over a period of time. These samples are called rational subgroups. There are many different strategies for choosing rational subgroups. The basic principle to be followed is that all the variability within the units in a rational subgroup should be due to common causes, and none should be due to special causes. In general, a good way to choose rational subgroups is to decide which special causes are most important to detect, and then choose the rational subgroups to provide the best chance to detect them. The two most commonly used methods are 1. Sample at regular time intervals, with all the items in each sample manufactured near the time the sampling is done. 2. Sample at regular time intervals, with the items in each sample drawn from all the units produced since the last sample was taken. For variables data, the number of units in each sample is typically small, often between three and eight. The number of samples should be at least 20. In general, many small samples taken frequently are better than a few large samples taken less frequently. For binary and for count data, samples must in general be larger. Page 774 Control versus Capability It is important to understand the difference between process control and process capability. A process is in control if there are no special causes operating. The distinguishing feature of a process that is in control is that the values of the quality characteristic vary without any trend or pattern, since the common causes do not change over time. However, it is quite possible for a process to be in control, and yet to be producing output that does not meet a given specification. For example, assume that a process produces steel rods whose lengths vary randomly between 19.9 and 20.1 cm, with no apparent pattern to the fluctuation. This process is in a state of control. However, if the design specification calls for a length between 21 and 21.2 cm, very little of the output would meet the specification. The ability of a process to produce output that meets a given specification is called the capability of the process. We will discuss the measurement of process capability in Section 10.5. Process Control Must Be Done Continually There are three basic phases to the use of control charts. First, data are collected. Second, these data are plotted to determine whether the process is in control. Third, once the process is brought into control, its capability may be assessed. Of course, a process that is in control and capable at a given time may go out of control at a later time, as special causes re-occur. For this reason processes must be continually monitored. Similarities Between Control Charts and Hypothesis Tests Control charts function in many ways like hypothesis tests. The null hypothesis is that the process is in control. The control chart presents data that provide evidence about the truth of this hypothesis. If the evidence against the null hypothesis is sufficiently strong, the process is declared out of control. Understanding how to use control charts involves knowing what data to collect and knowing how to organize those data to measure the strength of the evidence against the hypothesis that the process is in control. Exercises for Section 10.1 1. 2. 3. Indicate whether each of the following quality characteristics is a continuous, binary, or count variable. a. The number of flaws in a plate glass window. b. The length time taken to perform a final inspection of a finished product. c. Whether the breaking strength of a bolt meets a specification. d. The diameter of a rivet head. True or false: a. Control charts are used to determine whether special causes are operating. b. If no special causes are operating, then most of the output will meet specifications. c. Variability due to common causes does not increase or decrease much over short periods of time. d. Variability within the items sampled in a rational subgroup is due to special causes. e. If a process is in a state of statistical control, there will be almost no variation in the output. Fill in the blank. The choices are: is in control; has high capability. Page 775 a. 4. 5. If the variability in a process is approximately constant over time, the process ______. b. If most units produced conform to specifications, the process ______. Fill in the blank: Once a process has been brought into a state of statistical control, ______. i. it must still be monitored continually. ii. monitoring can be stopped for a while, since it is unlikely that the process will go out of control again right away. iii. the process need not be monitored again, unless it is redesigned. True or false: a. When a process is in a state of statistical control, then most of the output will meet specifications. b. When a process is out of control, an unacceptably large proportion of the output will not meet specifications. c. When a process is in a state of statistical control, all the variation in the process is due to causes that are inherent in the process itself. d. When a process is out of control, some of the variation in the process is due to causes 6. that are outside of the process. Fill in the blank: When sampling units for rational subgroups, ______ i. it is more important to choose large samples than to sample frequently, since large samples provide more precise information about the process. ii. it is more important to sample frequently than to choose large samples, so that special causes can be detected more quickly. 10.2 Control Charts for Variables When a quality measurement is made on a continuous scale, the data are called variables data. For these data an R chart or S chart is first used to control the variability in the process, and then an -chart is used to control the process mean. The methods described in this section assume that the measurements follow an approximately normal distribution. We illustrate with an example. The quality engineer in charge of a salt packaging process is concerned about the moisture content in packages of salt. To determine whether the process is in statistical control, it is first necessary to define the rational subgroups, and then to collect some data. Assume that for the salt packaging process, the primary concern is that variation in the ambient humidity in the plant may be causing variation in the mean moisture content in the packages over time. Recall that rational subgroups should be chosen so that the variation within each sample is due only to common causes, not to special causes. Therefore a good choice for the rational subgroups in this case is to draw samples of several packages each at regular time intervals. The packages in each sample will be produced as close to each other in time as possible. In this way, the ambient humidity will be nearly the same for each package in the sample, so the within-group variation will not be affected by this special cause. Assume that five packages of salt are sampled every 15 minutes for eight hours, and that the moisture content in each package is measured as a percentage of total weight. The data are presented in Table 10.1 (page 776). Since moisture is measured on a continuous scale, these are variables data. For each sample, we compute the sample mean, sample standard deviation s, and the sample range R (the difference Page 776 between the largest and smallest value). The last row of the table contains the mean of the sample means , the mean of the sample ranges , and the mean of the sample standard deviations . TABLE 10.1 Moisture content for salt packages, as a percentage of total weight Sample Values 2.66 1.88 2.21 2.28 2.34 2.47 2.47 2.23 2.10 2.43 2.32 2.38 2.56 2.42 2.56 2.51 1.63 2.95 2.12 2.26 2.61 2.54 2.28 2.36 2.67 Mean 2.308 2.498 2.394 2.346 2.498 2.402 Range (R) 0.780 0.350 0.570 0.480 0.280 1.320 SD(s) 0.303 0.149 0.230 0.196 0.111 0.525 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 2.58 2.31 3.03 2.86 2.71 2.95 3.14 2.85 2.82 3.17 2.81 2.99 3.11 2.83 2.76 2.54 ... 0.231 We assume that each of the 32 samples in Table 10.1 is a sample from a normal population with mean μ and standard deviation σ. The quantity μ is called the process mean, and σ is called the process standard deviation. The idea behind control charts is that each value of approximates the process mean during the time its sample was taken, while the values of R and s can be used to approximate the process standard deviation. If the process is in control, then the process mean and standard deviation are the same for each sample. If the process is out of control, the process mean μ or the process standard deviation σ, or both, will differ from sample to sample. Therefore the values of , R, and s will vary less when the process is in control than when the process is out of control. If the process is in control, the values of , R, and s will almost Page 777 always be contained within computable limits, called control limits. If the process is out of control, the values of , R, and s will be more likely to exceed these limits. A control chart plots the values of , R, or s, along with the control limits, so that it can be easily seen whether the variation is large enough to conclude that the process is out of control. Now let's see how to determine whether the salt packaging process is in a state of statistical control with respect to moisture content. Since the variation within each sample is supposed to be due only to common causes, this variation should not be too different from one sample to another. Therefore the first thing to do is to check to make sure that the amount of variation within each sample, measured either by the sample range or the sample standard deviation, does not vary too much from sample to sample. For this purpose the R chart can be used to assess variation in the sample range, or the S chart can be used to assess variation in the sample standard deviation. We will discuss the R chart first, since it is the more traditional. We will discuss the S chart at the end of this section. Figure 10.1 presents the R chart for the moisture data. The horizontal axis represents the samples, numbered from 1 to 32. The sample ranges are plotted on the vertical axis. Most important are the three horizontal lines. The line in the center is plotted at the value and is called the center line. The upper and lower lines indicate the 3σ upper and lower control limits (UCL and LCL, respectively). The control limits are drawn so that when the process is in control, almost all the points will lie within the limits. A point plotting outside the control limits is evidence that the process is out of control. FIGURE 10.1 R chart for the moisture data. To understand where the control limits are plotted, assume that the 32 sample ranges come from a population with mean μR and standard deviation σR. The values of μR and σR will not be known exactly, but it is known that for most populations, it is unusual to observe a value that differs from the mean by more than three standard deviations. For this reason, it is conventional to plot the control limits at points that approximate the values μR ± 3σR. It can be shown by advanced methods that the quantities μR ± 3σR can be estimated with multiples of ; these multiples are denoted D3 and D4. The quantity μR − 3σR is estimated with ; these , and the quantity μR + 3σR is estimated with . The quantities D3 and D4 are constants whose values depend on the sample size n. A brief table of values of D3 and D4 follows. A more extensive tabulation is provided in Table A.10 (in Appendix A). Note that for sample sizes of 6 or less, the Page 778 value of D3 is 0. For these small sample sizes, the quantity μR − 3σR is negative. In these cases the lower control limit is set to 0, because it is impossible for the range to be negative. n D3 D4 2 0 3.267 3 0 2.575 4 0 2.282 5 0 2.114 6 0 2.004 7 0.076 1.924 8 0.136 1.864 Example 10.1 Compute the 3σ R chart upper and lower control limits for the moisture data in Table 10.1. Solution The value of is 0.6066 (Table 10.1). The sample size is n = 5. From the table, D3 = 0 and D4 = 2.114. Therefore the upper control limit is (2.114)(0.6066) = 1.282, and the lower control limit is (0)(0.6066) = 0. Summary In an R chart, the center line and the 3σ upper and lower control limits are given by The values D3 and D4 depend on the sample size. Values are tabulated in Table A.10. Once the control limits have been calculated and the points plotted, the R chart can be used to assess whether the process is in control with respect to variation. Figure 10.1 shows that the range for sample number 6 is above the upper control limit, providing evidence that a special cause was operating and that the process variation is not in control. The appropriate action is to determine the nature of the special cause, and then delete the out-of-control sample and recompute the control limits. Assume it is discovered that a technician neglected to close a vent, causing greater than usual variation in moisture content during the time period when the sample was chosen. Retraining the technician will correct that special cause. We delete sample 6 from the data and recompute the R chart. The results are shown in Figure 10.2. The process variation is now in control. FIGURE 10.2 R chart for the moisture data, after deleting the out-of-control sample. Now that the process variation has been brought into control, we can assess whether the process mean is in control by plotting the chart. The chart is presented in Figure 10.3. The sample means are plotted on the vertical axis. Note that sample 6 has not been used in this chart since it had to be deleted in order to bring the process variation under control. Like all Page 779 control charts, the chart has a center line and upper and lower control limits. FIGURE 10.3 chart for the moisture data. Sample 6 has been deleted to bring the process variation under control. However, the chart shows that the process mean is out of control. To compute the center line and the control limits, we can assume that the process standard deviation is the same for all samples, since the R chart has been used to bring the process variation into control. If the process mean μ is in control as well, then it too is the same for all samples. In that case the 32 sample means are drawn from a normal population with mean and standard deviation , where n is the sample size, equal to 5 in this case. Ideally, we would like to plot the center line at μ and the 3σ control limits at the values of μ and μ ± 3σ. However, the values of μ and σ are usually unknown and have to be estimated from the data. We estimate μ with , the average of the sample means. The center line is therefore plotted at . The quantity σ can be estimated by using either the average range or by using the sample standard deviations. We will use here and discuss the methods based on the standard deviation at the end of the section, in conjunction with the discussion of S charts. It can be shown by advanced methods that the quantity can be estimated with , where A2 is a constant whose value depends on the sample size. A short table of values of A2 follows. A more extensive tabulation is provided in Table A.10. n 2 3 4 5 6 7 8 A2 1.880 1.023 0.729 0.577 0.483 0.419 0.373 Summary In an chart, when is used to estimate σ, the center line and the 3σ upper and lower Page 780 control limits are given by The value A2 depends on the sample size. Values are tabulated in Table A.10. Example 10.2 Compute the 3σ chart upper and lower control limits for the moisture data in Table 10.1. Solution With sample 6 deleted, the value of is 2.658, and the value of is 0.5836. The sample size is n = 5. From the table, A2 = 0.577. Therefore the upper control limit is 2.658 + (0.577)(0.5836) = 2.995, and the lower control limit is 2.658 − (0.577)(0.5836) = 2.321. The chart clearly shows that the process mean is not in control, as there are several points plotting outside the control limits. The production manager installs a hygrometer to monitor the ambient humidity and determines that the fluctuations in moisture content are caused by fluctuations in ambient humidity. A dehumidifier is installed to stabilize the ambient humidity. After this special cause is remedied, more data are collected, and a new R chart and chart are constructed. Figure 10.4 presents the results. The process is now in a state of statistical control. Of course, the process must be continually monitored, since new special causes are bound to crop up from time to time and will need to be detected and corrected. Note that while control charts can detect the presence of a special cause, they cannot determine its nature, nor how to correct it. It is necessary for the process engineer to have a good understanding of the process, so that special causes detected by control charts can be diagnosed and corrected. Page 781 FIGURE 10.4 R chart and chart after special cause is remedied. The process is now in a state of Summary The steps in using the R chart and chart are 1. Choose rational subgroups. 2. Compute the R chart. 3. Determine the special causes for any out-of-control points. 4. Recompute the R chart, omitting samples that resulted in out-of-control points. 5. Once the R chart indicates a state of control, compute the chart, omitting samples that resulted in out-of-control points on the R chart. 6. If the chart indicates that the process is not in control, identify and correct any special causes. 7. Continue to monitor and R. Control Chart Performance There is a close connection between control charts and hypothesis tests. The null Page 782 hypothesis is that the process is in a state of statistical control. A point plotting outside the 3σ control limits presents evidence against the null hypothesis. As with any hypothesis test, it is possible to make an error. For example, a point will occasionally plot outside the 3σ limits even when the process is in control. This is called a false alarm. It can also happen that a process that is not in control may not exhibit any points outside the control limits, especially if it is not observed for a long enough time. This is called a failure to detect. It is desirable for these errors to occur as infrequently as possible. We describe the frequency with which these errors occur with a quantity called the average run length (ARL). The ARL is the number of samples that must be observed, on average, before a point plots outside the control limits. We would like the ARL to be large when the process is in control, and small when the process is out of control. We can compute the ARL for an chart if we assume that process mean μ and process standard deviation σ are known. Then the center line is located at the process mean μ and the control limits are at . We must assume, as is always the case with the chart, that the quantity being measured is approximately normally distributed. Examples 10.3 through 10.6 show how to compute the ARL. Example For an in 10.3 chart with control limits at , compute the ARL for a process that is in control. Solution Let be the mean of a sample. Then . The probability that a point plots outside the control limits is equal to . This probability is equal to 0.00135 + 0.00135 = 0.0027 (see Figure 10.5). Therefore, on the average, 27 out of every 10,000 points will plot outside the control limits. This is equivalent to 1 every 10,000/27 = 370.4 points. The average run length is therefore equal to 370.4. FIGURE 10.5 The probability that a point plots outside the 3σ control limits, when the process is in control, is 0.0027 (0.00135 + 0.00135). The result of Example 10.3 can be interpreted as follows: If a process is in control, we expect to observe about 370 samples, on the average, before finding one that plots Page 783 outside the control limits, causing a false alarm. Note also that the ARL in Example 10.3 was 10,000/27, which is equal to 1/0.0027, where 0.0027 is the probability that any given sample plots outside the control limits. This is true in general. Summary The average run length (ARL) is the number of samples that will be observed, on the average, before a point plots outside the control limits. If p is the probability that any given point plots outside the control limits, then If a process is in control, the ARL will be large, but less than 370.4. Example 10.4 shows how to compute the ARL for a situation where the process shifts to an out-of-control condition. Example 10.4 A process has mean μ = 3 and standard deviation σ = 1. Samples of size n = 4 are taken. If a special cause shifts the process mean to a value of 3.5, find the ARL. Solution We first compute the probability p that a given point plots outside the control limits. Then ARL = 1/p. The control limits are plotted on the basis of a process that is in control. Therefore they are at , where μ = 3 and . The lower control limit is thus at 1.5, and the upper control limit is at 4.5. If process mean has shifted, then . (10.1) If μ = 0 the ARL is ... the upper control limit (1.5) and the upper control limit is ... below the mean, indicated by the fact that z = 2. The lower limit is now below the mean. The probability that a point plots below the lower control limit is 0.0228 (0 + 0.0228). Page 784 Example 10.5 Refer to Example 10.4. An upward shift to what value can be detected with an ARL of 20? Solution Let m be the new mean to which the process has shifted. Since we have specified an upward shift, m > 3. In Example 10.4 we computed the control limits to be 1.5 and 4.5. If ARL = 20, then p = 1/ARL = 1/20 = 0.05. Since m > 3, m is closer to 4.5 than to 1.5. We will begin by assuming that the area to the left of 1.5 is negligible and that the area to the right of 4.5 is equal to 0.05. The z-score of 4.5 is then 1.645, so (4.5 − m)/0.5 = 1.645. Solving for m, we have m = 3.68. We finish by checking our assumption that the area to the left of 1.5 is negligible. With m = 3.68, the z-score for 1.5 is (1.5 − 3.68)/0.5 = −4.36. The area to the left of 1.5 is indeed negligible. FIGURE 10.7 Solution to Example 10.5. Example 10.6 Refer to Example 10.4. If the sample size remains at n = 4, what must the value of the process standard deviation σ be to produce an ARL of 10 when the process mean shifts to 3.5? Solution Let σ denote the new process standard deviation. The new control limits are 3 ± 3σ/2. If the process mean shifts to 3.5, then ... The probability that a point plots outside the control limits is equal to 1/ARL = 1/10 = 0.10 (see Figure 10.8). The process mean, 3.5, is closer to 3 + 3σ/2 than to 3 − 3σ/2. We will assume that the area to the left of 3 − 3σ/2 is negligible and that the area to the right of 3 + 3σ/2 is equal to 0.10. The z-score for 3 + 3σ/2 is then 1.28, so Solving for σ, we obtain σ = 0.58. We finish by checking that the area to the left of 3 − 3σ/2 is negligible. Substituting σ = 0.58, we obtain 3 − 3σ/2 = 2.13. The z-score is (2.13 − 3.5)/(0.58/2) = −4.72. The area to the left of 3 − 3σ/2 is indeed negligible. FIGURE 10.8 Solution to Example 10.6. Examples 10.4 through 10.6 show that charts do not usually detect small shifts quickly. In other words, the ARL is high when shifts in the process mean are small. In principle, one could reduce the ARL by moving the control limits closer to the centerline. This would reduce the size of the shift needed to detect an out-of-control condition, so that changes in the process mean would be detected more quickly. However, there is a trade-off. The false alarm rate would increase as well, because shifts outside the control limits would be more likely to occur by chance. The situation is much like that in fixed-level hypothesis testing. The null hypothesis is that the process is in control. The control chart performs a hypothesis test on each sample. When a point plots outside the control limits, the null hypothesis is rejected. With the control limits at , a type I error (rejection of a true null hypothesis) will occur about once in every 370 samples. The price to pay for this low false alarm rate is lack of power to reject the null hypothesis when it is false. Moving the control limits closer together is not the answer. Although it will increase the power, it will also increase the false alarm rate. Two of the ways in which practitioners have attempted to improve their ability to detect small shifts quickly are by using the Western Electric rules to interpret the control chart and by using CUSUM charts. The Western Electric rules are described next. CUSUM charts are described in Section 10.4. The Western Electric Rules Figure 10.9 (page 786) presents an chart. While none of the points fall outside the 3σ control limits, the process is clearly not in a state of control, since there is a nonrandom pattern to the sample means. In recognition of the fact that a process can fail to be in control even when no points plot outside the control limits, engineers at the Western Electric company in 1956 suggested a list of conditions, any one of which could be used as evidence that a process is out of control. The idea behind these conditions is that if a trend or pattern in the control chart persists for long enough, it can indicate the absence of control, even if no point plots outside the 3σ control limits. To apply the Western Electric rules, it is necessary to compute the 1σ and 2σ control limits. The 1σ control limits are given by , and the 2σ control limits are given by . Page 786 FIGURE 10.9 This chart exhibits nonrandom patterns, indicating a lack of statistical control, even though no points plot outside the 3σ control limits. The 1σ and 2σ control limits are shown on this plot, so that the Western Electric Rules can be applied. The Western Electric Rules Any one of the following conditions is evidence that a process is out of control: 1. Any point plotting outside the 3σ control limits. 2. Two out of three consecutive points plotting above the upper 2σ limit, or two out of three consecutive points plotting below the lower 2σ limit. 3. Four out of five consecutive points plotting above the upper 1σ limit, or four out of five consecutive points plotting below the lower 1σ limit. 4. Eight consecutive points plotting on the same side of the center line. In Figure 10.9, the Western Electric rules indicate that the process is out of control at sample number 8, at which time four out of five consecutive points have plotted above the upper 1σ control limit. For more information on using the Western Electric rules to interpret control charts, see Montgomery (2013b). The S chart The S chart is an alternative to the R chart. Both the S chart and the R chart are used to control the variability in a process. While the R chart assesses variability with the sample range, the S chart uses the sample standard deviation. Figure 10.10 presents the S chart for the moisture data in Table 10.1. Page 787 FIGURE 10.10 S chart for the moisture data. Compare with Figure 10.1. Note that the S chart for the moisture data is similar in appearance to the R chart (Figure 10.1) for the same data. Like the R chart, the S chart indicates that the variation was out of control in sample 6. To understand where the control limits are plotted, assume that the 32 sample standard deviations come from a population with mean μs and standard deviation σs. Ideally we would like to plot the center line at μs and the control limits at μs ± 3σs. These quantities are typically unknown. We approximate μs with , the average of the sample standard deviations. Thus the center line is plotted at . It can be shown by advanced methods that the quantities μs ± 3σs can be estimated with multiples of ; these multiples are denoted B3 and B4. The quantity μs − 3σs is , while the quantity μs + 3σs is estimated with estimated with . The quantities B3 and B4 are constants whose values depend on the sample size n. A brief table of values of B3 and B4 follows. A more extensive tabulation is provided in Table A.10 (Appendix A). Note that for samples of size 5 or less, the value of B3 is 0. For samples this small, the value of μs − 3σs is negative. In these cases the lower control limit is set to 0, because it is impossible for a standard deviation to be negative. n 2 3 4 5 6 7 8 B3 0 0 0 0 0.030 0.118 0.185 B4 3.267 2.568 2.266 2.089 1.970 1.882 1.815 Example 10.7 Compute the center line and the 3σ S chart upper and lower control limits for the moisture data in Table 10.1. Solution The value of is 0.2445 (Table 10.1). The sample size is n = 5. From the table immediately preceding, B3 = 0 and B4 = 2.089. Therefore the upper control limit is (2.089)(0.2445) = 0.5108, and the lower control limit is (0)(0.2445) = 0. Page 788 Summary In an S chart, the center line and the 3σ upper and lower control limits are given by The values B3 and B4 depend on the sample size. Values are tabulated in Table A.10. The S chart in Figure 10.10 shows that the process is out of control in sample 6. We delete this sample and recompute the S chart. Figure 10.11 presents the results. The variation is now in control. Note that this S chart is similar in appearance to the R chart in Figure 10.2. FIGURE 10.11 S chart for the moisture data, after deleting the out-of-control sample. Compare with Figure 10.2. Once the variation is in control, we compute the chart to assess the process mean. Recall that for the chart, the center line is at , and the upper and lower control limits would ideally be located a distance above and below the center line. Since we used the R chart to assess the process variation, we estimated the quantity with a multiple of . Specifically, we estimated with , where A2 is a constant whose value depends on the sample size. If we use the S chart to assess the process variation, we estimate with a multiple of . Specifically, we estimate with , where A3 is a constant whose value Page 789 depends on the sample size. A brief table of values of A3 follows. A more extensive tabulation is provided in Table A.10. n 2 3 4 5 6 7 8 A3 2.659 1.954 1.628 1.427 1.287 1.182 1.099 Summary In an chart, when is used to estimate , the center line and the 3σ upper and lower control limits are given by The value A3 depends on the sample size. Values are tabulated in Table A.10. If Western Electric rules are to be used, 1σ and 2σ control limits must be computed. The 1σ limits are . Example 10.8 Compute the chart for the moisture data in Table 10.1. Solution With sample 6 deleted, the value of is 2.658, and the value of is 0.2354. The sample size is n = 5. From the table, A3 = 1.427. Therefore the upper control limit is 2.658 + (1.427) (0.2354) = 2.994, and the lower control limit is 2.658 − (1.427)(0.2354) = 2.322. The chart for the moisture data with sample 6 deleted is shown in Figure 10.12. The control limits are very similar to those calculated from the sample ranges, as shown in Figure 10.3. Figure 10.12 indicates that the process is out of control. After taking corrective Page 790 action, a new S chart and chart are constructed. Figure 10.13 presents the results. The process is now in a state of statistical control. FIGURE 10.12 chart for the moisture data. The control limits are based on the sample standard deviations rather than the sample ranges. Compare with Figure 10.3. FIGURE 10.13 S chart and chart after special cause is remedied. The process is now in a state of statistical control. Compare with Figure 10.4. In summary, the S chart is an alternative to the R chart, to be used in combination with the chart. For the moisture data, it turned out that the two charts gave very similar results. This is true in many cases, but it will sometimes happen that the results differ. Which is Better, the S Chart or the R Chart? Both the R chart and S chart have the same purpose: to estimate the process standard deviation and to determine whether it is in control. It seems more natural to estimate the process standard deviation with the sample standard deviation s than with the range R. In fact, when the population is normal, s is a more precise estimate of the process standard deviation than is R, because it has a smaller uncertainty. To see this intuitively, note that the computation of s involves all the measurements in each sample, while the computation of R involves only two measurements (the largest and the smallest). It turns out that the improvement in Page 791 precision obtained with s as opposed to R increases as the sample size increases. It follows that the S chart is a better choice, especially for larger sample sizes (greater than 5 or so). The R chart is still widely used, largely through tradition. At one time, the R chart had the advantage that the sample range required less arithmetic to compute than did the standard deviation. Now that most calculations are done electronically, this advantage no longer holds. So the S chart is in general the better choice. Samples of Size 1 Sometimes it is necessary to define rational subgroups in such a way that each sample can contain only one value. For example, if the production rate is very slow, it may not be convenient to wait to accumulate samples larger than n = 1. It is then impossible to compute a sample range or a sample standard deviation for a sample of size 1, so R charts and S charts cannot be used. Several other methods are available. One method is the CUSUM chart, discussed in Section 10.4. Exercises for Section 10.2 1. 2. The quality-control plan for a certain production process involves taking samples of size 4. The results from the last 30 samples can be summarized as follows: a. Compute the 3σ control limits for the R chart. b. Compute the 3σ control limits for the S chart. c. Using the sample ranges, compute the 3σ control limits for the chart. d. Using the sample standard deviations, compute the 3σ control limits for the chart. The following chart depicts the last 50 samples taken from the output of a process. Using the Western Electric rules, is the process detected to be out of control at any time? If so, specify at which sample the process is first detected to be out of control and which rule is violated. Page 791 3. The thickness, in mm, of metal washers is measured on samples of size 5. The following table presents the means, ranges, and standard deviations for 20 consecutive samples. Page 792 Sample 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 2.49 2.45 2.51 2.53 2.49 2.44 2.44 2.42 2.42 2.47 2.54 2.45 2.54 2.55 2.50 2.53 2.58 2.59 2.60 2.56 R 0.12 0.17 0.13 0.25 0.11 0.11 0.12 0.18 0.08 0.06 0.19 0.09 0.21 0.10 0.25 0.11 0.12 0.18 0.08 0.06 0.19 0.09 s 0.05 0.06 0.06 0.10 0.05 0.04 0.06 0.07 0.03 0.04 0.07 0.03 0.04 0.07 0.05 0.10 0.04 0.03 0.07 0.07 The means are , , and . a. Calculate the 3σ control limits for the R chart. b. Is the variance under control? If not, delete the samples that are out of control and recompute and . b. Based on the sample range R, calculate the 3σ control limits for the chart. Based on the 4. 5. 6. 7. 8. the 3σ limits, is the process mean in control? If not, when is it first detected to be out of control? c. Based on the Western Electric rules, is the process mean in control? If not, when is it first detected to be out of control? Repeat Exercise 3, using the S chart in place of the R chart. A process has mean 12 and standard deviation 3. The process is monitored by taking samples of size 5 at regular intervals. The process is declared to be out of control if a point plots outside the 3σ control limits on an chart. a. If the process mean shifts to 14, what is the average number of samples that will be drawn before the shift is detected on an chart? b. An upward shift to what value will be detected with an ARL of 4? c. If the sample size remains at 5, to what value must the standard deviation be reduced to produce an ARL of 4 when the process mean shifts to 14? d. If the standard deviation remains at 3, what sample size must be used to produce an ARL no greater than 4 when the process mean shifts to 14? A process has mean 8 and standard deviation 2. The process is monitored by taking samples of size 4 at regular intervals. The process is declared to be out of control if a point plots outside the 3σ control limits on an chart. a. If the process mean shifts to 9, what is the average number of samples that will be drawn before the shift is detected on the chart? b. An upward shift to what value will be detected with an ARL of 8? c. If the sample size remains at 4, to what value must the standard deviation be reduced to produce an ARL of 8 when the process mean shifts to 9? d. If the standard deviation remains at 2, what sample size must be used to produce an ARL no greater than 8 when the process mean shifts to 9? A process is monitored by taking samples at regular intervals and is declared to be out of control if a point plots outside the 3σ control limits on an chart. Assume the process is in control. a. What is the probability that a false alarm will occur within the next 50 samples? b. What is the probability that a false alarm will occur within the next 100 samples? c. What is the probability that there will be no false alarm within the next 200 samples? d. Fill in the blank: The probability is 0.5 that there will be a false alarm within the next _____ samples. Samples of eight bolts are taken periodically, and their diameters (in mm) are measured. The following table presents the means, ranges, and standard deviations for 25 consecutive samples. Page 793 Sample 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 R 9.99 10.02 10.10 9.90 9.92 10.05 9.97 9.93 10.01 9.87 9.90 9.97 10.02 9.99 9.99 10.04 10.07 10.04 9.95 9.98 9.98 10.02 9.94 10.04 0.28 0.43 0.16 0.26 0.22 0.40 0.08 0.48 0.25 0.29 0.39 0.27 0.20 0.37 0.20 0.26 0.23 0.35 0.25 0.15 0.30 0.14 0.24 0.13 0.24 0.07 0.15 0.03 0.15 0.09 0.12 0.14 0.06 0.11 0.14 0.08 0.13 0.06 0.09 0.08 0.11 0.10 0.13 0.09 0.12 0.09 0.06 0.06 0.07 0.07 The means are , , and . a. Calculate the 3σ control limits for the R chart. Is the variance under control? If not, delete the samples that are out of control and recompute and . b. Based on the sample range R, calculate the 3σ control limits for the chart. Based on the 3σ limits, is the process mean in control? If not, when is it first detected to be out of control? c. Based on the Western Electric rules, is the process mean in control? If not, when is it first detected to be out of control? 9. Repeat Exercise 8, using the S chart in place of the R chart. 10. A certain type of integrated circuit is connected to its frame by five wires. Thirty samples of five units each were taken, and the pull strength (in grams) of one wire on each unit was measured. The data are presented in Table E10 on page 794. The means are , , and . a. Compute the 3σ limits for the R chart. Is the variance out of control at any point? If so, delete the samples that are out of control and recompute and . b. Compute the 3σ limits for the chart. On the basis of the 3σ limits, is the process mean in control? If not, at what point is it first detected to be out of control? c. On the basis of the Western Electric rules, is the process mean in control? If not, when is it first detected to be out of control? TABLE E10 Sample 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 Sample Values 9.8 10.4 9.9 10.5 10.7 10.0 10.8 10.8 8.8 9.9 11.3 10.3 9.9 10.1 9.8 10.1 10.3 9.0 10.1 9.4 9.9 10.1 10.1 9.9 9.4 9.6 9.8 10.1 9.7 10.3 10.2 10.4 9.5 9.3 9.5 8.9 9.8 9.3 10.1 10.2 10.8 11.0 10.0 10.0 11.0 10.0 9.8 10.1 9.0 9.8 9.2 10.1 10.0 9.6 10.0 10.5 9.6 10.5 10.2 9.7 9.4 9.4 10.5 10.1 9.9 9.5 9.4 10.2 10.9 9.1 10.2 10.1 9.8 10.6 9.2 10.2 10.0 9.5 10.4 10.3 9.7 9.3 9.8 9.4 10.0 9.7 10.2 10.5 9.7 10.2 9.7 10.0 9.8 9.9 10.8 11.0 9.7 10.3 9.5 10.3 10.2 10.2 9.7 10.7 9.8 9.9 10.0 9.8 9.9 10.1 9.8 10.4 10.7 9.7 9.6 11.0 9.9 10.1 9.9 10.2 9.6 10.2 10.4 9.9 9.8 9.7 9.9 10.2 10.5 9.6 10.1 9.6 9.6 10.0 10.0 9.8 9.6 10.2 9.8 10.8 9.6 9.9 10.2 10.2 10.1 10.5 10.3 10.0 9.9 9.9 10.5 10.2 9.8 10.6 10.1 10.3 10.1 10.4 9.9 10.7 10.1 9.5 9.5 10.3 10.0 9.8 10.4 10.0 9.7 10.0 10.5 10.1 9.6 10.1 9.9 10.1 9.9 10.2 10.3 9.9 9.6 9.5 10.1 9.8 10.0 9.9 10.7 10.3 10.9 10.5 10.2 10.7 10.1 10.4 10.1 9.8 10.2 9.9 10.5 9.4 10.1 10.1 9.8 10.6 9.8 9.8 10.5 10.2 10.4 9.8 10.1 10.2 10.5 10.2 9.9 10.1 10.5 10.0 9.7 9.4 10.3 9.4 9.8 10.8 10.4 9.7 9.9 10.4 9.9 9.8 10.7 10.1 10.3 10.4 10.0 9.7 10.5 10.3 10.5 10.6 10.2 9.9 10.1 9.7 9.7 10.8 9.9 10.4 10.7 10.4 10.6 10.0 10.1 10.0 10.5 9.8 9.7 10.4 9.1 10.3 10.7 10.2 10.6 10.8 R 0.4 0.6 0.3 0.9 0.8 1.0 0.9 0.5 1.7 0.5 1.1 1.0 0.9 1.0 1.2 0.5 1.4 1.6 1.1 1.5 1.7 1.2 0.7 0.9 0.9 0.4 0.7 0.6 1.6 0.8 0.20 0.23 0.12 0.34 0.30 0.40 0.35 0.19 0.68 0.21 0.47 0.38 0.35 0.42 0.52 0.19 0.58 0.62 0.46 0.57 0.69 0.44 0.30 0.33 0.38 0.16 0.25 0.29 0.67 0.29 0.23 0.20 0.46 0.11 11. Repeat Exercise 10, using the S chart in place of the R chart. 12. Copper wires are coated with a thin plastic coating. Samples of four wires are taken every hour, and the thickness of the coating (in mils) is measured. The data from the last 30 samples are presented in Table E12 on page 795. The means are , , and . a. Compute the 3σ limits for the R chart. Is the variance out of control at any point? If so, delete the samples that are out of control and recompute and . b. Compute the 3σ limits for the chart. On the basis of the 3σ limits, is the process mean in control? If not, at what point is it first detected to be out of control? c. On the basis of the Western Electric rules, is the process mean in control? If not, when is it first detected to be out of control? TABLE E12 Sample 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 Sample Values 146.0 147.1 148.7 151.3 146.4 150.2 147.8 137.1 151.1 151.3 151.5 150.4 149.0 153.5 150.6 143.0 151.0 143.8 143.0 148.8 153.0 151.6 149.4 149.8 155.8 153.9 Sample Values 147.4 151.0 147.5 151.4 148.4 149.6 150.0 152.4 142.0 152.5 146.3 146.7 148.3 145.5 151.3 154.0 152.9 151.5 148.0 150.4 149.0 146.8 149.5 149.8 151.5 152.5 150.1 150.3 151.3 146.8 140.6 145.6 150.0 146.3 149.9 148.6 152.2 149.2 152.0 146.3 151.2 146.7 147.8 153.7 151.7 149.9 152.3 149.1 150.1 150.3 147.4 145.7 148.5 149.9 150.3 149.3 146.1 149.5 149.1 147.8 155.4 145.1 155.0 154.6 149.3 155.1 148.1 148.6 148.0 152.4 150.2 144.7 152.3 147.2 147.2 146.3 148.4 145.0 153.2 153.4 153.4 150.2 153.1 152.0 153.3 147.6 153.2 150.1 152.6 142.4 147.5 150.1 149.9 151.2 150.1 148.5 149.0 149.5 149.1 152.4 151.7 146.8 143.7 155.0 155.0 155.3 149.2 151.8 157.8 148.6 146.1 150.0 148.7 152.4 147.3 152.2 146.1 145.4 149.5 152.8 150.0 149.0 148.6 153.9 150.125 148.850 150.475 149.175 150.275 147.875 147.450 148.100 149.325 151.275 151.200 150.900 151.625 152.700 151.100 147.575 149.550 151.175 150.400 147.950 150.525 151.225 151.350 147.850 148.475 150.375 147.825 150.925 151.900 149.325 152.625 151.275 151.325 152.700 151.100 150.750 150.175 13. Repeat Exercise 12, using the S chart in place of the R chart. R 9.2 4.3 5.7 4.2 6.5 12.6 4.0 19.5 5.2 6.7 4.9 5.9 3.0 4.1 8.9 2.6 8.9 7.0 4.1 4.1 7.7 11.0 13.4 1.7 8.4 6.0 5.4 1.7 6.6 9.1 4.2 1.97 2.65 1.81 2.52 5.37 1.66 8.02 2.27 2.29 2.02 2.01 2.50 1.38 2.06 4.20 1.28 3.59 4.20 3.75 2.19 5.12 5.64 0.72 3.93 2.77 2.30 1.97 2.88 2.51 3.38 Page 795 10.3 Control Charts for Attributes The p Chart The p chart is used when the quality characteristic being measured on each unit has only two possible values, usually "defective" and "not defective." In each sample, the proportion of defectives is calculated; these sample proportions are then plotted. We will now describe how the center line and control limits are calculated. Let p be the probability that a given unit is defective. If the process is in control, this probability is constant over time. Let k be the number of samples. We will assume that all samples are the same size, and we will denote this size by n. Let Xi be the number of defective units in the ith sample, and let be the proportion of defective items in the ith sample. Now Xi ~ Bin(n, p), and if np > 10, it is approximately true that (see page 295). Since has mean μ = p and standard deviation , it Page 796 follows that the center line should be at p, and the 3σ control limits should be at . Usually p is not known and is estimated with , the average of the sample proportions . Summary In a p chart, where the number of items in each sample is n, the center line and the 3σ upper and lower control limits are given by These control limits will be valid if . We illustrate these ideas with Example 10.9. Example 10.9 In the production of silicon wafers, 30 lots of size 500 are sampled, and the proportion of defective wafers is calculated for each sample. Table 10.2 presents the results. Compute the center line and 3σ control limits for the p chart. Plot the chart. Does the process appear to be in control? TABLE 10.2 Sample 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 Number and proportion defective, for Example 10.9 Number Defective 17 26 31 25 26 29 36 26 25 21 18 33 29 17 28 26 19 31 27 24 28 32 30 24 24 27 28 22 26 23 Proportion Defective 0.034 0.052 0.062 0.050 0.052 0.058 0.072 0.052 0.050 0.042 0.036 0.066 0.058 0.034 0.056 0.052 0.038 0.062 0.054 0.048 0.044 0.048 0.060 0.050 0.052 0.056 0.044 0.062 0.036 0.046 Solution Page 797 The average of the 30 sample proportions is . The center line is therefore plotted at 0.050867. The upper control limit is therefore 0.0803, and the lower control limit is 0.0214. Figure 10.14 presents the p chart. The process appears to be in control. FIGURE 10.14 p chart for the data in Table 10.2 The sample size needed to construct a p chart is usually much larger than that needed for an chart. The reason is that the sample size must be large enough so that there will be several defective items in most of the samples. If defective items are not common, the sample size must be quite large. Interpreting Out-of-Control Signals in Attribute Charts When an attribute control chart is used to monitor the frequency of defective units, a point plotting above the upper control limit requires quite a different response than a point plotting below the lower control limit. Both conditions indicate that a special cause has changed the proportion of defective units. A point plotting above the upper control limit indicates that the proportion of defective units has increased, so action must be taken to identify and remove the special cause. A point plotting below the lower control limit, however, indicates that the special cause has decreased the proportion of defective units. The special cause still needs to be identified, but in this case, action should be taken to make it continue, so that the proportion of defective items can be decreased permanently. The C Chart The c chart is used when the quality measurement is a count of the number of defects, or flaws, in a given unit. A unit may be a single item, or it may be a group of items large enough so that the expected number of flaws is sufficiently large. Use of the c chart requires that the number of defects follow a Poisson distribution. Assume that k units are sampled, and let ci denote the total number of defects in the ith unit. Let λ denote the mean total number of flaws per unit. Then ci ~ Poisson(λ). If the process is in control, the value of λ is constant over time. Now if λ is reasonably large, say λ > 10, then ci ~ N(λ, λ), approximately (see page 299). Note that Page 798 the value of λ can in principle be made large enough by choosing a sufficiently large number of items per unit. The c chart is constructed by plotting the values ci. Since ci has mean λ and standard deviation equal to , the center line should be plotted at λ and the 3σ control limits should be plotted at . Usually the value of λ is unknown and has to be estimated from the data. The appropriate estimate is , the average number of defects per unit. Summary In a c chart, the center line and the 3σ upper and lower control limits are given by These control limits will be valid if . Example 10.10 illustrates these ideas. Example 10.10 Rolls of sheet aluminum, used to manufacture cans, are examined for surface flaws. Table 10.3 presents the numbers of flaws in 40 samples of 100 m2 each. Compute the center line and 3σ control limits for the c chart. Plot the chart. Does the process appear to be in control?' TABLE 10.3 Sample 1 2 3 4 Number of flaws, for Example 10.10 Number of Flaws (c) 16 12 9 13 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 15 13 11 15 12 14 11 10 9 9 14 10 12 9 13 15 13 14 10 12 16 13 12 17 15 13 15 14 13 10 11 12 11 14 Solution The average of the 40 counts is . The center line is therefore plotted at 12.275. The 3σ upper and lower control limits are therefore 22.7857, and the lower control limit is 1.7643. Figure 10.15 presents the c chart. The process appears to be in control. Page 799 FIGURE 10.15 c chart for the data in Table 10.3 Exercises for Section 10.3 1. 2. 3. 4. 5. A process is monitored for defective items by taking a sample of 200 items each day and calculating the proportion that are defective. Let pi be the proportion of defective items in the ith sample. For the last 30 samples, the sum of the proportions is . Calculate the center line and the 3σ upper and lower control limits for a p chart. The target fill weight for a box of cereal is 350 g. Each day a sample of 300 boxes is taken, and the number that are underweight is counted. The number of underweight boxes for each of the last 25 days is as follows: 2312 1919 201921 2726 232622 25 3030 2225 272935 3943 413929 a. Compute the upper and lower 3σ limits for a p chart. b. Is the process in control? If not, when is it first detected to be out of control? A process is monitored for defective items by periodically taking a sample of 100 items and counting the number that are defective. In the last 50 samples, there were a total of 622 defective items. Is this enough information to compute the 3σ control limits for a p chart? If so, compute the limits. If not, state what additional information would be required. Refer to Exercise 3. In the last 50 samples, there were a total of 622 defective items. The largest number of defectives in any sample was 24, while the smallest number was 6. Is this enough information to determine whether the process was out of control at any time during the last 50 samples? If so, state whether or not the process was out of control. If not, state what additional information would be required to make the determination. A newly designed quality-control program for a certain process involves sampling 20 items each day and counting the number of defective items. The numbers of defectives in the first 10 samples are 0, 0, 1, 0, 1, 0, 0, 0, 1, 0. A member of the quality-control team asks for advice, expressing concern that the number of defectives are too small to construct an accurate p chart. Which of the following is the best advice? Nothing needs to be changed. An accurate p chart can be constructed when the number of defective items is this small. Since the proportion of items that are defective is so small, it isn't necessary to construct a p chart for this process. iii. Increase the value of p to increase the number of defectives per sample. iv. Increase the sample size to increase the number of defectives per sample. A process that produces mirrors for automobiles is monitored by taking samples of 1500 mirrors and counting the total number of visual flaws on all the sample mirrors. Let c be the sample number of flaws on the mirrors in the ith sample. For the last 70 samples, the quantity has been calculated. Compute the center line and the 3σ upper and lower control limits for a c chart. Page 800 i. ii. 6. 7. 8. Refer to Exercise 6. The number of flaws in the 34th sample was 27. Is it possible to determine whether the process was in control at this time? If so, state whether or not the process was in control. If not, state what additional information would be required to make the determination. Each hour, a 10 m2 section of fabric is inspected for flaws. The numbers of flaws observed for the last 20 hours are as follows: 38 353549 3348 4047 45 46 41 533641 5163 3558 55 57 a. Compute the upper and lower 3σ limits for a c chart. b. Is the process in control? If not, when is it first detected to be out of control? 10.4 The CUSUM Chart One purpose of an chart is to detect a shift in the process mean. Unless a shift is fairly large, however, it may be some time before a point plots outside the 3σ control limits. Example 10.4 (in Section 10.2) showed that when a process mean shifts by an amount equal to , the average run length (ARL) is approximately 44, which means that on the average 44 samples must be observed before the process is judged to be out of control. The Western Electric rules (Section 10.2) provide one method for reducing the ARL. CUSUM charts provide another. One way that small shifts manifest themselves is with a run of points above or below the center line. The Western Electric rules are designed to respond to runs. Another way to detect smaller shifts is with cumulative sums. Imagine that a process mean shifts upward slightly. There will then be a tendency for points to plot above the center line. If we add the deviations from the center line as we go along, and plot the cumulative sums, the points will drift upward and will exceed a control limit much sooner than they would in an chart. We now describe how to plot the points in a CUSUM chart. We assume that we have m samples of size n, with sample means . To begin, a target value μ must be specified for the process mean. Often μ is taken to be the value . Then an estimate of , the standard deviation of the sample means, is needed. This can be obtained either with sample ranges, using the estimate , or with sample standard deviations, using the estimate . If there is only one item per sample (n = 1), then an external estimate is needed. Even a rough guess can produce good results, so the CUSUM procedure can be useful when n = 1. Finally two constants, usually called k and h, must be specified. Larger values for these constants result in longer average run lengths, and thus fewer false alarms, but also result in longer waiting times to discover that a process is out of control. The values k = 0.5 and h = 4 or 5 are often used, because they provide a reasonably long ARL when the process is in control but still have fairly good power to detect a shift of magnitude or more in the process mean. For each sample, the quantity is the deviation from the target value. We define two cumulative sums, SH and SL. The sum SH is always either positive or zero and signals that the process mean has become greater than the target value. The sum SL is always either negative or zero and signals that the process mean has become less than the target value. Both Page 801 these sums are computed recursively: in other words, the current value in the sequence is used to compute the next value. The initial values of SH and SL are (10.2) For i ≥ 1 the values are (10.3) (10.4) If SH(i) > h for some i, it is concluded that the process mean has become greater than the target value. If SL(i) < −h for some i, it is concluded that the process mean has become less than the target value. Figure 10.16 presents a CUSUM chart for the data in Figure 10.9 (in Section 10.2). The values k = 0.5 and h = 4 were used. The value 2.952 is the quantity . The CUSUM chart indicates an out-of-control condition on the tenth sample. For these data, the CUSUM chart performs about as well as the Western Electric rules, which determined that the process was out of control at the eighth sample (see Figure 10.9). FIGURE 10.16 CUSUM chart for the data in Figure 10.9. Summary In a CUSUM chart, two cumulative sums, SH and SL, are plotted. The initial values are SH0 = SL0 = 0. For i ≥ 1, The constants k and h must be specified. Good results are often obtained for the values k = 0.5 and h = 4 or 5. If for any i, SHi > h or SLi < −h, the process is judged to be out of control. There are several other methods for constructing CUSUM charts, which are Page 802 equivalent, or nearly equivalent, to the method presented here. Some people define the deviations to be the z-scores , and then use zi in place of Xi − μ, and h in place of in the formulas for SH and SL. With this definition, the control limits are plotted at ±h rather than . Other methods for graphing the CUSUM chart are available as well. The most common alternative is the "V-mask" approach. A text on statistical quality control, such as Montgomery (2013b), can be consulted for further information. Exercises for Section 10.4 1. 2. 3. Refer to Exercise 3 in Section 10.2. a. Delete any samples necessary to bring the process variation under control. (You did this already if you did Exercise 3 in Section 10.2.) b. Use to estimate (is the difference between and the 1σ control limit on an chart). c. Construct a CUSUM chart, using for the target mean μ, and the estimate of found in part (b) for the standard deviation. Use the values k = 0.5 and h = 4. d. Is the process mean in control? If not, when is it first detected to be out of control? e. Construct an chart, and use the Western Electric rules to determine whether the process mean is in control. (You did this already if you did Exercise 3 in Section 10.2.) Do the Western Electric rules give the same results as the CUSUM chart? If not, how are they different? Refer to Exercise 8 in Section 10.2. a. Delete any samples necessary to bring the process variation under control. (You did this already if you did Exercise 8 in Section 10.2.) b. Use to estimate (is the difference between and the 1σ control limit on an chart). c. Construct a CUSUM chart, using for the target mean μ, and the estimate of found in part (b) for the standard deviation. Use the values k = 0.5 and h = 4. d. Is the process mean in control? If not, when is it first detected to be out of control? e. Construct an chart, and use the Western Electric rules to determine whether the process mean is in control. (You did this already if you did Exercise 8 in Section 10.2.) Do the Western Electric rules give the same results as the CUSUM chart? If not, how are they different? Refer to Exercise 10 in Section 10.2. a. Delete any samples necessary to bring the process variation under control. (You did this already if you did Exercise 10 in Section 10.2.) b. Use to estimate (is the difference between and the 1σ control limit on an chart). c. Construct a CUSUM chart, using for the target mean μ, and the estimate of found in part (b) for the standard deviation. Use the values k = 0.5 and h = 4. d. Is the process mean in control? If not, when is it first detected to be out of control? e. Construct an chart, and use the Western Electric rules to determine whether the process mean is in control. (You did this already if you did Exercise 10 in Section 10.2.) Do the Western Electric rules give the same results as the CUSUM chart? If not, how are they different? Refer to Exercise 12 in Section 10.2. a. Delete any samples necessary to bring the process variation under control. (You did this already if you did Exercise 12 in Section 10.2.) b. Use to estimate (is the difference between and the 1σ control limit on an chart). Page 803 Construct a CUSUM chart, using for the target mean μ,and the estimate of found in part (b) for the standard deviation. Use the values k = 0.5 and h = 4. d. Is the process mean in control? If not, when is it first detected to be out of control? e. Construct an chart, and use the Western Electric rules to determine whether the process mean is in control. (You did this already if you did Exercise 12 in Section 10.2.) Do the Western Electric rules give the same results as the CUSUM chart? If not, how are they different? Concrete blocks to be used in a certain application are supposed to have a mean compressive strength of 1500 MPa. Samples of size 1 are used for quality control. The compressive strengths of the last 40 samples are given in the following table. c. 5. Sample 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 Strength 1487 1463 1499 1502 1473 1520 1520 1495 1503 1499 1497 1516 1489 1545 1498 1503 1522 1502 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 6. 1499 1484 1574 1514 1515 1487 1458 1521 1517 1496 1511 1489 1527 1542 1565 1549 1497 1488 Previous results suggest that a value of σ = 15 is reasonable for this process. a. Using the value 1500 for the target mean μ, and the values k = 0.5 and h = 4, construct a CUSUM chart. b. Is the process mean in control? If not, when is it first detected to be out of control? A quality-control apprentice is preparing a CUSUM chart. The values calculated for SL and SH are presented in the following table. Three of the values have been calculated incorrectly. Which are they? Sample 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 SL 0 0 −1.3280 −1.4364 −2.0464 −1.6370 −0.8234 SH 0 0 0 0 0 0.2767 0 0 0 0 0 0 0 0 0 0 0 0 0 0 −0.4528 0 0.2371 0.7104 0 0 0 0 0 −0.2036 0 0.1106 0.7836 0.0097 0 0.2775 0.5842 0.3750 0.4658 0.1866 0.3277 0 −0.7345 Page 804 10.5 Process Capability Once a process is in a state of statistical control, it is important to evaluate its ability to produce output that conforms to design specifications. We consider variables data, and we assume that the quality characteristic of interest follows a normal distribution. The first step in assessing process capability is to estimate the process mean and standard deviation. These estimates are denoted and , respectively. The data used to calculate and are usually taken from control charts at a time when the process is in a state of control. The process mean is estimated with . The process standard deviation can be estimated by using either the average sample range or the average sample standard deviation . Specifically, it has been shown that can be computed either by dividing by a constant called d2, or by dividing by a constant called c4. The values of the constants d2 and c4 depend on the sample size. Values are tabulated in Table A.10 in Appendix A. Summary If a quality characteristic from a process in a state of control is normally distributed, then the process mean and standard deviation can be estimated from control chart data as follows: The values of d2 and c4 depend on the sample size. Values are tabulated in Table A.10. Note that the process standard deviation σ is not the same quantity that is used to compute the 3σ control limits on the chart. The control limits are , where is the standard deviation of the sample mean. The process standard deviation σ is the standard deviation of the quality characteristic of individual units. They are related by , where n is the sample size. To be fit for use, a quality characteristic must fall between a lower specification limit (LSL) and an upper specification limit (USL). Sometimes there is only one limit; this situation will be discussed at the end of this section. The specification limits are determined by design requirements. They are not the control limits found on control charts. We will assume that the process mean falls between the LSL and the USL. We will discuss two indices of process capability, Cpk and Cp. The index Cpk describes the capability of the process as it is, while Cp describes the potential capability of the process. Note that the process capability index Cp has no relation to the quantity called Mallows' Cp that is used for linear model selection (see Chapter 8). It is a coincidence that the two quantities have the same name. Page 805 The index Cpk is defined to be the distance from the process mean to the nearest specification limit, divided by 3 times the process standard deviation. . Figure 10.17 presents an illustration where is closer to the upper FIGURE 10.17 The normal curve represents the population of units produced by the process. The process mean is closer to the upper specification limit (USL) than to the lower specification limit (LSL). The index Cpk is therefore equal to . Definition The index Cpk is equal either to whichever is less. By convention, the minimum acceptable value for Cpk is 1. That is, a process is considered to be minimally capable if the process mean is three standard deviations from the nearest specification limit. A Cpk value of 1.33, indicating that the process mean is four standard deviations from the nearest specification limit, is generally considered good. Example 10.11 The design specifications for a piston rod used in an automatic transmission call for the rod length to be between 71.4 and 72.8 mm. The process is monitored with an chart and an S chart, using samples of size n = 5. These show the process to be in control. The values of and are and . Compute the value of Cpk. Is the process capability acceptable? Solution We estimate . To compute , we find, from Table A.10, that c4 = 0.9400 when the sample size is 5. Therefore limits are LSL = 71.4 mm and USL = 72.8 mm. The value is closer to the USL than to the Since Cpk < 1, the process capability is not acceptable. Page 806 Example 10.12 Refer to Example 10.11. Assume that it is possible to adjust the process mean to any desired value. To what value should it be set to maximize Cpk? What will the value of Cpk be? Solution We are given that . The value of Cpk will be maximized if the process mean is adjusted to the midpoint between the specification limits; that is, if μ = 72.1. The process standard deviation is estimated with . Therefore the maximum value of Cpk is (72.1 − 71.4)/(3(0.2128) = 1.0965. The process capability is acceptable. The capability that can potentially be achieved by shifting the process mean to the midpoint between the upper and lower specification limits is called the process capability index, denoted Cp. If the process mean is at the midpoint between LSL and USL, then the distance from the mean to either specification limit is equal to one-half the distance between the specification limits, that is μ − LSL = USL − μ = (USL − LSL)/2 (see Figure 10.18). It follows that FIGURE 10.18 A process has maximum capability when the process mean is at the midpoint between the specification limits. In this case μ − LSL = USL − μ = (USL − LSL)/2. Example 10.13 Specifications for the output voltage of a certain electric circuit are 48 to 52 V. The process is in control with V. Compute the process capability index Cp. Page 807 Solution The process capability is potentially good. Estimating the Proportion of Nonconforming Units from Process Capability Many people use the value of Cp to try to estimate the proportion of units that will be nonconforming. For example, if Cp = 1, then the specification limits are equal to and , respectively. Therefore a unit will be nonconforming only if it is more than three standard deviations from the process mean. Now for a normal population, the proportion of items that are more than three standard deviations from the mean is equal to 0.0027. Therefore it is often stated that a process with Cp = 1 will produce 27 nonconforming parts per 10,000. The problem with this is that the normality assumption is only approximate for real processes. The approximation may be very good near the middle of the curve, but it is often not good in the tails. Therefore the true proportion of nonconforming parts may be quite different from that predicted from Cp, especially when the proportion is very small. In general, estimates of small probabilities that are based on a normal approximation are extremely crude at best. Six-Sigma Quality The term "six-sigma quality" has become quite prevalent in discussions of quality control during the last few years. A process is said to have six-sigma quality if the process capability index Cp has a value of 2.0 or greater. Equivalently, a process has six-sigma quality if the difference USL − LSL is at least 12σ. When a process has six-sigma quality, then if the process mean is optimally adjusted, it is six standard deviations from each specification limit. In this case the proportion of nonconforming units will be virtually zero. An important feature of a six-sigma process is that it can withstand moderate shifts in process mean without significant deterioration in capability. For example, even if the process mean shifts by 3σ in one direction or the other, it is still 3σ from the nearest specification limit, so the capability index will still be acceptable. Example 10.14 Refer to Example 10.13. To what value must the process standard deviation be reduced in order for the process to attain six-sigma quality? Page 808 Solution To attain six-sigma quality, the value of Cp must be at least 2.0. The value of σ for which this occurs is found by setting Cp = 2.0 and solving for σ. We obtain from which σ = 0.33. One-Sided Tolerances Some characteristics have only one specification limit. For example, strengths usually have a lower specification limit but no upper limit, since for most applications a part cannot be too strong. The analog of Cpk when there is only a lower specification limit is the lower capability index Cpl; when there is only an upper limit, it is the upper capability index Cpu. Each of these quantities is defined to be the difference between the estimated process mean and the specification limit, divided by . Summary If a process has only a lower specification limit (LSL), then the lower capability index If a process has only an upper specification limit (USL), then the upper capability index There is no analog for Cp for processes with only one specification limit. Exercises for Section 10.5 1. The thickness specification for aluminum sheets is 0.246–0.254 mm. Data from an chart, based on samples of size 6, that shows that the process is in control, yield values of and . a. 2. 3. 4. Compute the value of Cpk for this process. b. Is the process capability acceptable? Explain. The specification for the diameters of ball bearings is 15.40–15.60 mm. Data from an chart, based on samples of size 8, that shows that the process is in control, yield values of and . a. Compute the value of Cpk for this process. b. Is the process capability acceptable? Explain. Refer to Exercise 2. a. To what value should the process mean be set to maximize the process capability? Page 809 b. c. d. Is it possible to make the process capability acceptable simply by adjusting the process mean? Explain. When the process mean is adjusted to its optimum value, what value must be attained by the process standard deviation so that the process capability is acceptable? When the process mean is adjusted to its optimum value, what value must be attained 5. by the process standard deviation so that the process has six-sigma quality? A process has a process capability index of Cp = 1.2. a. b. c. Assume the process mean is set to its optimal value. Express the upper and lower specification limits in terms of the process mean and standard deviation. Using the normal curve, estimate the proportion of units that will be nonconforming. Is it likely or unlikely that the true proportion of nonconforming units will be quite different from the estimate in part (b)? Explain. Supplementary Exercises for Chapter 10 1. 2. 3. A process is monitored for defective items by taking a sample of 300 items each day and calculating the proportion that are defective. Let pi be the proportion of defective items in the ith sample. For the last 100 samples, the sum of the proportions is . Calculate the center line and the 3σ upper and lower control limits for a p chart. Someone constructs an chart where the control limits are at rather than at . a. If the process is in control, what is the ARL for this chart? b. If the process mean shifts by , what is the ARL for this chart? c. In units of , how large an upward shift can be detected with an ARL of 10? Samples of three resistors are taken periodically, and the resistances, in ohms, are measured. The following table presents the means, ranges, and standard deviations for 30 consecutive samples. Sample 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 5.114 5.144 5.220 5.196 5.176 5.222 5.209 5.212 5.238 5.152 5.163 5.221 5.144 5.098 5.070 5.029 5.045 5.008 R 0.146 0.158 0.057 0.158 0.158 0.158 0.030 0.118 0.099 0.157 0.104 0.051 0.105 0.132 0.123 0.083 0.073 0.161 0.138 s 0.077 0.085 0.031 0.081 0.099 0.017 0.059 0.053 0.085 0.054 0.026 0.051 0.071 0.026 0.042 0.038 0.087 0.071 19 20 21 22 23 24 25 26 27 28 29 30 The means are 6. 7. , , and . 0.082 0.109 0.066 0.078 0.063 0.072 0.124 0.106 0.135 0.120 0.122 0.016 0.038 0.062 0.032 0.028 0.036 0.039 0.071 0.052 0.076 0.063 0.063 0.062 0.009 0.067 0.042 0.034 0.072 0.043 0.032 0.038 0.041 0.044 0.066 0.041 0.068 0.062 0.074 0.026 0.076 Page 810 Compute the 3σ limits for the R chart. Is the variance out of control at any point? If so, delete the samples that are out of control and recompute and . b. Compute the 3σ limits for the chart. On the basis of the 3σ limits, is the process mean in control? If not, at what point is it first detected to be out of control? c. On the basis of the Western Electric rules, is the process mean in control? If not, when is it first detected to be out of control? Repeat Exercise 3, using the S chart in place of the R chart. Refer to Exercise 3. a. Delete any samples necessary to bring the process variation under control. (You did this already if you did Exercise 3.) b. Use to estimate (is the difference between and the 1σ control limit on an chart). c. Construct a CUSUM chart, using for the target mean μ,and the estimate of found in part (b) for the standard deviation. Use the values k = 0.5 and h = 4. d. Is the process mean in control? If not, when is it first detected to be out of control? e. Construct an chart, and use the Western Electric rules to determine whether the process mean is in control. (You did this already if you did Exercise 3.) Do the Western Electric rules give the same results as the CUSUM chart? If not, how are they different? A process is monitored for defects by taking a sample of 15 each hour and counting the total number of flaws in the sample items. The total number of flaws over the last 50 samples is 1085. a. Compute the center line and upper and lower 3σ control limits. b. The tenth sample had five flaws. Was the process out of control at that time? Explain. To set up a p chart to monitor a process that produces computer chips, samples of 300 chips are taken each hour, and the number of defective chips in each sample is counted. The numbers of defective chips for each of the last 20 days are as follows: 4. 5.029 5.038 4.962 5.033 4.993 4.961 4.976 5.005 5.022 5.077 5.033 5.068 12 1311 1015 14 1 10 9 15 8 1311 9 1612 1920 18 9 a. Compute the upper and lower 3σ limits for a p chart. b. At which sample is the process first detected to be out of control? c. Suppose that the special cause that resulted in the out-of-control condition is determined. Should this cause be remedied? Explain. Page 811 Appendix A Tables ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ Table A.1: Cumulative Binomial Distribution Table A.2: Cumulative Normal Distribution Table A.3: Upper Percentage Points for the Student's t Distribution Table A.4: Tolerance Factors for the Normal Distribution Table A.5: Critical Points for the Wilcoxon Signed-Rank Test Table A.6: Critical Points for the Wilcoxon Rank-Sum Test Table A.7: Upper Percentage Points for the χ2 Distribution Table A.8: Upper Percentage Points for the F Distribution Table A.9: Upper Percentage Points for the Studentized Range Table A.10: Control Chart Constants Page 812 TABLE A.1 n 2 3 4 5 x 0 1 2 0 1 2 3 0 1 2 3 4 0 1 2 3 4 Cumulative binomial distribution 0.05 0.902 0.997 1.000 0.857 0.993 1.000 1.000 0.815 0.986 1.000 1.000 1.000 0.774 0.977 0.999 1.000 1.000 p 0.10 0.810 0.990 1.000 0.729 0.972 0.999 1.000 0.656 0.948 0.996 1.000 1.000 0.590 0.919 0.991 1.000 1.000 0.20 0.640 0.960 1.000 0.512 0.896 0.992 1.000 0.410 0.819 0.973 0.998 1.000 0.328 0.737 0.942 0.993 1.000 0.25 0.563 0.938 1.000 0.422 0.844 0.984 1.000 0.316 0.738 0.949 0.996 1.000 0.237 0.633 0.896 0.984 0.999 0.30 0.490 0.910 1.000 0.343 0.784 0.973 1.000 0.240 0.652 0.916 0.992 1.000 0.168 0.528 0.837 0.969 0.998 0.40 0.360 0.840 1.000 0.216 0.648 0.936 1.000 0.130 0.475 0.821 0.974 1.000 0.078 0.337 0.683 0.913 0.990 0.50 0.250 0.750 1.000 0.125 0.500 0.875 1.000 0.063 0.313 0.688 0.938 1.000 0.031 0.187 0.500 0.813 0.969 0.60 0.160 0.640 1.000 0.064 0.352 0.784 1.000 0.026 0.179 0.525 0.870 1.000 0.010 0.087 0.317 0.663 0.922 0.70 0.090 0.510 1.000 0.027 0.216 0.657 1.000 0.008 0.084 0.348 0.760 1.000 0.002 0.031 0.163 0.472 0.832 0.75 0.063 0.438 1.000 0.016 0.156 0.578 1.000 0.004 0.051 0.262 0.684 1.000 0.001 0.016 0.104 0.367 0.763 0.80 0.040 0.360 1.000 0.008 0.104 0.488 1.000 0.002 0.027 0.181 0.590 1.000 0.000 0.007 0.058 0.263 0.672 0.90 0.010 0.190 1.000 0.001 0.028 0.271 1.000 0.000 0.004 0.052 0.344 1.000 0.000 0.000 0.009 0.081 0.410 0.95 0.003 0.098 1.000 0.000 0.007 0.143 1.000 0.000 0.000 0.014 0.185 1.000 0.000 0.000 0.001 0.023 0.226 1.000 0.000 0.000 0.000 0.000 0.000 Page 813 0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7 8 n x 6 7 3 4 5 0 1 2 3 4 5 6 7 0.735 0.967 0.998 1.000 1.000 1.000 1.000 0.698 0.956 0.996 1.000 1.000 1.000 1.000 1.000 0.531 0.886 0.984 0.999 1.000 1.000 1.000 0.478 0.850 0.974 0.997 1.000 1.000 1.000 1.000 0.262 0.655 0.901 0.983 0.998 1.000 1.000 0.210 0.577 0.852 0.967 0.995 1.000 1.000 1.000 0.178 0.534 0.831 0.962 0.995 1.000 1.000 0.133 0.445 0.756 0.929 0.987 0.999 1.000 1.000 0.118 0.420 0.744 0.930 0.989 1.000 1.000 0.082 0.329 0.647 0.874 0.971 0.996 1.000 1.000 0.047 0.233 0.544 0.821 0.959 0.996 1.000 0.028 0.159 0.420 0.710 0.904 0.981 0.998 1.000 0.016 0.109 0.344 0.656 0.891 0.984 1.000 0.008 0.063 0.227 0.500 0.773 0.938 0.992 1.000 0.004 0.041 0.179 0.456 0.767 0.953 1.000 0.002 0.019 0.096 0.290 0.580 0.841 0.972 1.000 0.001 0.011 0.070 0.256 0.580 0.882 1.000 0.000 0.004 0.029 0.126 0.353 0.671 0.918 1.000 0.000 0.005 0.038 0.169 0.466 0.822 1.000 0.000 0.001 0.013 0.071 0.244 0.555 0.867 1.000 0.000 0.002 0.017 0.099 0.345 0.738 1.000 0.000 0.000 0.005 0.033 0.148 0.423 0.790 1.000 0.000 0.000 0.001 0.016 0.114 0.469 1.000 0.000 0.000 0.000 0.003 0.026 0.150 0.522 1.000 0.000 0.000 0.000 0.002 0.033 0.265 1.000 0.000 0.000 0.000 0.000 0.004 0.044 0.302 1.000 0.663 0.943 0.994 1.000 1.000 1.000 1.000 1.000 1.000 0.430 0.813 0.962 0.995 1.000 1.000 1.000 1.000 1.000 0.168 0.503 0.797 0.944 0.990 0.999 1.000 1.000 1.000 0.100 0.367 0.679 0.886 0.973 0.996 1.000 1.000 1.000 0.058 0.255 0.552 0.806 0.942 0.989 0.999 1.000 1.000 0.017 0.106 0.315 0.594 0.826 0.950 0.991 0.999 1.000 0.004 0.035 0.145 0.363 0.637 0.855 0.965 0.996 1.000 0.001 0.009 0.050 0.174 0.406 0.685 0.894 0.983 1.000 0.000 0.001 0.011 0.058 0.194 0.448 0.745 0.942 1.000 0.000 0.000 0.004 0.027 0.114 0.321 0.633 0.900 1.000 0.000 0.000 0.001 0.010 0.056 0.203 0.497 0.832 1.000 0.000 0.000 0.000 0.000 0.005 0.038 0.187 0.570 1.000 0.000 0.000 0.000 0.000 0.000 0.006 0.057 0.337 1.000 0 1 2 3 4 5 6 7 8 9 0.630 0.929 0.992 0.999 1.000 1.000 1.000 1.000 1.000 1.000 0.387 0.775 0.947 0.992 0.999 1.000 1.000 1.000 1.000 1.000 0.134 0.436 0.738 0.914 0.980 0.997 1.000 1.000 1.000 1.000 0.075 0.300 0.601 0.834 0.951 0.990 0.999 1.000 1.000 1.000 0.040 0.196 0.463 0.730 0.901 0.975 0.996 1.000 1.000 1.000 0.010 0.071 0.232 0.483 0.733 0.901 0.975 0.996 1.000 1.000 0.002 0.020 0.090 0.254 0.500 0.746 0.910 0.980 0.998 1.000 0.000 0.004 0.025 0.099 0.267 0.517 0.768 0.929 0.990 1.000 0.000 0.000 0.004 0.025 0.099 0.270 0.537 0.804 0.960 1.000 0.000 0.000 0.001 0.010 0.049 0.166 0.399 0.700 0.925 1.000 0.000 0.000 0.000 0.003 0.020 0.086 0.262 0.564 0.866 1.000 0.000 0.000 0.000 0.000 0.001 0.008 0.053 0.225 0.613 1.000 0.000 0.000 0.000 0.000 0.000 0.001 0.008 0.071 0.370 1.000 0 1 2 3 4 5 6 7 8 9 10 0.599 0.914 0.988 0.999 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.349 0.736 0.930 0.987 0.998 1.000 1.000 1.000 1.000 1.000 1.000 0.107 0.376 0.678 0.879 0.967 0.994 0.999 1.000 1.000 1.000 1.000 0.056 0.244 0.526 0.776 0.922 0.980 0.996 1.000 1.000 1.000 1.000 0.028 0.149 0.383 0.650 0.850 0.953 0.989 0.998 1.000 1.000 1.000 0.006 0.046 0.167 0.382 0.633 0.834 0.945 0.988 0.998 1.000 1.000 0.001 0.011 0.055 0.172 0.377 0.623 0.828 0.945 0.989 0.999 1.000 0.000 0.002 0.012 0.055 0.166 0.367 0.618 0.833 0.954 0.994 1.000 0.000 0.000 0.002 0.011 0.047 0.150 0.350 0.617 0.851 0.972 1.000 0.000 0.000 0.000 0.004 0.020 0.078 0.224 0.474 0.756 0.944 1.000 0.000 0.000 0.000 0.001 0.006 0.033 0.121 0.322 0.624 0.893 1.000 0.000 0.000 0.000 0.000 0.000 0.002 0.013 0.070 0.264 0.651 1.000 0.000 0.000 0.000 0.000 0.000 0.000 0.001 0.012 0.086 0.401 1.000 Page 814

0.0444 0.0222 0.0571 0.0286 0.0714 0.0357 0.0179 0.0833 0.0476 0.0238 0.0583 0.0333 0.0167 0.0083 0.0667 0.0424 0.0242 4 4 5 6 7 7 6 5 5 6 7 7 6 7 7 6 11 10 13 12 11 10 15 14 13 12 11 10 16 15 14 13 12 11 20 19 18 17 16 15 21 20 19 18 17 16 22 21 29 30 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 36 37 38 39 40 41 40 35 36 37 38 39 40 41 42 43 44 43 44 0.0121 0.0571 0.0286 0.0143 0.0556 0.0317 0.0159 0.0079 0.0571 0.0333 0.0190 0.0095 0.0048 0.0545 0.0364 0.0212 0.0121 0.0061 0.0030 0.0545 0.0364 0.0242 0.0141 0.0081 0.0040 0.0754 0.0476 0.0278 0.0159 0.0079 0.0040 0.0628 0.0411 0.0260 0.0152 0.0087 0.0043 0.0530 0.0366 8 6 6 7 8 7 20 19 18 17 16 24 23 22 21 20 19 18 17 29 28 27 26 25 24 23 30 29 28 27 26 25 24 32 31 30 29 28 27 26 25 24 32 31 30 29 28 27 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 37 36 35 34 33 32 31 30 29 28 42 41 40 39 38 37 36 35 34 33 32 31 30 29 28 27 26 48 47 46 45 44 43 42 41 40 39 38 37 36 35 34 33 32 31 30 29 0.0454 0.0364 0.0242 0.0227 0.0152 0.0111 0.0080 0.0040 0.0758 0.0556 0.0288 0.0212 0.0118 0.0061 0.0027 0.0242 0.0152 0.0088 0.0051 0.0025 0.0637 0.0466 0.0326 0.0225 0.0148 0.0093 0.0054 0.0031 0.0660 0.0465 0.0325 0.0206 0.0130 0.0076 0.0043 0.0507 0.0367 0.0256 0.0175 0.0111 0.0070 0.0041 0.0539 0.0406 0.0296 0.0213 0.0147 0.0100 0.0063 0.0040 0.0641 0.0487 0.0265 0.0189 0.0131 34 33 32 42 41 39 38 36 35 34 32 51 50 49 46 45 44 43 42 41 0.0469 0.0270 0.0103 0.0047 0.0524 0.0415 0.0325 0.0249 0.0190 0.0103 0.0062 0.0035 When m and n are both greater than 8, compute and use the z table (Table A.2). Page 824 TABLE A.7 Upper percentage points for the χ^2 distribution α ν 1 2 3 4 5 6 7 8 9 0.995 0.000 0.010 0.072 0.207 0.412 0.676 0.989 1.344 1.735 0.99 0.000 0.020 0.115 0.297 0.554 0.872 1.239 1.646 2.088 0.975 0.001 0.051 0.216 0.484 0.831 1.237 1.690 2.180 2.700 0.95 0.004 0.103 0.352 0.711 1.145 1.635 2.167 2.733 3.325 0.90 0.016 0.211 0.584 1.064 1.610 2.204 2.833 3.490 4.168 0.10 2.706 4.605 6.251 7.779 9.236 10.645 12.017 13.362 14.684 0.05 3.841 5.991 7.815 9.488 11.070 12.592 14.067 15.507 16.919 0.025 5.024 7.378 9.348 11.143 12.833 14.449 16.013 17.535 19.023 0.01 6.635 9.210 11.345 13.277 15.086 16.812 18.475 20.090 21.666 0.005 7.879 10.597 12.838 14.860 16.750 18.548 20.278 21.955 23.589 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 2.156 2.603 3.074 3.565 4.075 4.601 5.142 5.697 6.265 6.844 7.434 8.034 8.643 9.260 9.886 10.520 11.160 11.808 12.461 13.121 13.787 14.458 15.134 15.815 16.501 17.192 17.887 18.586 19.289 19.996 20.707 2.558 3.053 3.571 4.107 4.660 5.229 5.812 6.408 7.015 7.633 8.260 8.897 9.542 10.196 10.856 11.524 12.198 12.879 13.565 14.256 14.953 15.655 16.362 17.074 17.789 18.509 19.233 19.960 20.691 21.426 22.164 For ν > 40, TABLE A.8 3.247 3.816 4.404 5.009 5.629 6.262 6.908 7.564 8.231 8.907 9.591 10.283 10.982 11.689 12.401 13.120 13.844 14.573 15.308 16.047 16.791 17.539 18.291 19.047 19.806 20.569 21.336 22.106 22.878 23.654 24.433 3.940 4.575 5.226 5.892 6.571 7.261 7.962 8.672 9.390 10.117 10.851 11.591 12.338 13.091 13.848 14.611 15.379 16.151 16.928 17.708 18.493 19.281 20.072 20.867 21.664 22.465 23.269 24.075 24.884 25.695 26.509 4.865 5.578 6.304 7.042 7.790 8.547 9.312 10.085 10.865 11.651 12.443 13.240 14.041 14.848 15.659 16.473 17.292 18.114 18.939 19.768 20.599 21.434 22.271 23.110 23.952 24.797 25.643 26.492 27.343 28.196 29.051 15.987 17.275 18.549 19.812 21.064 22.307 23.542 24.769 25.989 27.204 28.412 29.615 30.813 32.007 33.196 34.382 35.563 36.741 37.916 39.087 40.256 41.422 42.585 43.745 44.903 46.059 47.212 48.363 49.513 50.660 51.805 18.307 19.675 21.026 22.362 23.685 24.996 26.296 27.587 28.869 30.144 31.410 32.671 33.924 35.172 36.415 37.652 38.885 40.113 41.337 42.557 43.773 44.985 46.194 47.400 48.602 49.802 50.998 52.192 53.384 54.572 55.758 20.483 21.920 23.337 24.736 26.119 27.488 28.845 30.191 31.526 32.852 34.170 35.479 36.781 38.076 39.364 40.646 41.923 43.195 44.461 45.722 46.979 48.232 49.480 50.725 51.966 53.203 54.437 55.668 56.896 58.120 59.342 . Upper percentage points for the F distribution 23.209 24.725 26.217 27.688 29.141 30.578 32.000 33.409 34.805 36.191 37.566 38.932 40.289 41.638 42.980 44.314 45.642 46.963 48.278 49.588 50.892 52.191 53.486 54.776 56.061 57.342 58.619 59.893 61.162 62.428 63.691 25.188 26.757 28.300 29.819 31.319 32.801 34.267 35.718 37.156 38.582 39.997 41.401 42.796 44.181 45.559 46.928 48.290 49.645 50.993 52.336 53.672 55.003 56.328 57.648 58.964 60.275 61.581 62.883 64.181 65.476 66.766 Page 825 ν1 ν2 1 2 3 4 5 6 7 8 α 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 1 39.86 161.45 4052.18 405284 8.53 18.51 98.50 998.50 5.54 10.13 34.12 167.03 4.54 7.71 21.20 74.14 4.06 6.61 16.26 47.18 3.78 5.99 13.75 35.51 3.59 5.59 12.25 29.25 3.46 5.32 11.26 25.41 2 49.50 199.50 4999.50 500012 9.00 19.00 99.00 999.00 5.46 9.55 30.82 148.50 4.32 6.94 18.00 61.25 3.78 5.79 13.27 37.12 3.46 5.14 10.92 27.00 3.26 4.74 9.55 21.69 3.11 4.46 8.65 18.49 3 53.59 215.71 5403.35 540382 9.16 19.16 99.17 999.17 5.39 9.28 29.46 141.11 4.19 6.59 16.69 56.18 3.62 5.41 12.06 33.20 3.29 4.76 9.78 23.70 3.07 4.35 8.45 18.77 2.92 4.07 7.59 15.83 4 55.83 224.58 5624.58 562501 9.24 19.25 99.25 999.25 5.34 9.12 28.71 137.10 4.11 6.39 15.98 53.44 3.52 5.19 11.39 31.09 3.18 4.53 9.15 21.92 2.96 4.12 7.85 17.20 2.81 3.84 7.01 14.39 5 57.24 230.16 5763.65 576405 9.29 19.30 99.30 999.30 5.31 9.01 28.24 134.58 4.05 6.26 15.52 51.71 3.45 5.05 10.97 29.75 3.11 4.39 8.75 20.80 2.88 3.97 7.46 16.21 2.73 3.69 6.63 13.49 6 58.20 233.99 5858.99 585938 9.33 19.33 99.33 999.33 5.28 8.94 27.91 132.85 4.01 6.16 15.21 50.53 3.40 4.95 10.67 28.83 3.05 4.28 8.47 20.03 2.83 3.87 7.19 15.52 2.67 3.58 6.37 12.86 7 58.91 236.77 5928.36 592874 9.35 19.35 99.36 999.36 5.27 8.89 27.67 131.58 3.98 6.09 14.98 49.66 3.37 4.88 10.46 28.16 3.01 4.21 8.26 19.46 2.78 3.79 6.99 15.02 2.62 3.50 6.18 12.40 8 59.44 238.88 5981.07 599144 9.37 19.37 99.37 999.37 5.25 8.85 27.49 130.62 3.95 6.04 14.80 49.00 3.34 4.82 10.29 27.65 2.98 4.15 8.10 19.03 2.75 3.73 6.84 14.63 2.59 3.44 6.03 12.05 9 59.86 240.54 6022.47 603040 9.38 19.38 99.39 999.39 5.24 8.81 27.35 129.86 3.94 6.00 14.66 48.47 3.32 4.77 10.16 27.24 2.96 4.10 7.98 18.69 2.72 3.68 6.72 14.33 2.56 3.39 5.91 11.77 10 60.19 241.88 6055.85 605621 9.39 19.40 99.40 999.40 5.23 8.79 27.23 129.25 3.92 5.96 14.55 48.05 3.30 4.74 10.05 26.92 2.94 4.06 7.87 18.41 2.70 3.64 6.62 14.08 2.54 3.35 5.81 11.54 Page 826 1 2 3 4 5 6 7 8 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 12 60.71 243.91 6106.32 611276 9.41 19.41 99.42 999.42 5.22 8.74 27.05 128.32 3.90 5.91 14.37 47.41 3.27 4.68 9.89 26.42 2.90 4.00 7.72 17.99 2.67 3.57 6.47 13.71 2.50 3.28 5.67 11.19 15 61.22 245.95 6157.29 615764 9.42 19.43 99.43 999.43 5.20 8.70 26.87 127.37 3.87 5.86 14.20 46.76 3.24 4.62 9.72 25.91 2.87 3.94 7.56 17.56 2.63 3.51 6.31 13.32 2.46 3.22 5.52 10.84 20 61.74 248.01 6208.73 621362 9.44 19.45 99.45 999.45 5.18 8.66 26.69 126.42 3.84 5.80 14.02 46.10 3.21 4.56 9.55 25.39 2.84 3.87 7.40 17.12 2.59 3.44 6.16 12.93 2.42 3.15 5.36 10.48 25 62.05 249.26 6239.83 624430 9.45 19.46 99.46 999.46 5.17 8.63 26.58 125.84 3.83 5.77 13.91 45.70 3.19 4.52 9.45 25.08 2.81 3.83 7.30 16.85 2.57 3.40 6.06 12.69 2.40 3.11 5.26 10.26 30 62.26 250.10 6260.65 626486 9.46 19.46 99.47 999.47 5.17 8.62 26.50 125.45 3.82 5.75 13.84 45.43 3.17 4.50 9.38 24.87 2.80 3.81 7.23 16.67 2.56 3.38 5.99 12.53 2.38 3.08 5.20 10.11 40 62.53 251.14 6286.78 659725 9.47 19.47 99.47 999.47 5.16 8.59 26.41 124.96 3.80 5.72 13.75 45.09 3.16 4.46 9.29 24.60 2.78 3.77 7.14 16.44 2.54 3.34 5.91 12.33 2.36 3.04 5.12 9.92 50 62.69 251.77 6302.52 660511 9.47 19.48 99.48 999.48 5.15 8.58 26.35 124.66 3.80 5.70 13.69 44.88 3.15 4.44 9.24 24.44 2.77 3.75 7.09 16.31 2.52 3.32 5.86 12.20 2.35 3.02 5.07 9.80 60 62.79 252.20 6313.03 661190 9.47 19.48 99.48 999.48 5.15 8.57 26.32 124.47 3.79 5.69 13.65 44.75 3.14 4.43 9.20 24.33 2.76 3.74 7.06 16.21 2.51 3.30 5.82 12.12 2.34 3.01 5.03 9.73 120 63.06 253.25 6339.39 663921 9.48 19.49 99.49 999.49 5.14 8.55 26.22 123.97 3.78 5.66 13.56 44.40 3.12 4.40 9.11 24.06 2.74 3.70 6.97 15.98 2.49 3.27 5.74 11.91 2.32 2.97 4.95 9.53 ∞ 63.33 254.31 6365.76 666622 9.49 19.50 99.50 999.50 5.13 8.53 26.13 123.50 3.76 5.63 13.46 44.05 3.10 4.36 9.02 23.79 2.72 3.67 6.88 15.75 2.47 3.23 5.65 11.70 2.29 2.93 4.86 9.33 Page 827 α 10 0.100 0.050 0.010 0.001 11 0.100 0.050 0.010 0.001 12 0.100 0.050 0.010 0.001 13 0.100 0.050 0.010 0.001 14 0.100 0.050 0.010 0.001 15 0.100 0.050 0.010 0.001 16 0.100 0.050 0.010 0.001 17 0.100 0.050 0.010 0.001 18 0.100 0.050 0.010 0.001 19 0.100 0.050 0.010 0.001 20 0.100 0.050 0.010 0.001 3.29 4.96 10.04 21.04 3.23 4.84 9.65 19.69 3.18 4.75 9.33 18.64 3.14 4.67 9.07 17.82 3.10 4.60 8.86 17.14 3.07 4.54 8.68 16.59 3.05 4.49 8.53 16.12 3.03 4.45 8.40 15.72 3.01 4.41 8.29 15.38 2.99 4.38 8.18 15.08 2.97 4.35 8.10 14.82 2.92 4.10 7.56 14.91 2.86 3.98 7.21 13.81 2.81 3.89 6.93 12.97 2.76 3.81 6.70 12.31 2.73 3.74 6.51 11.78 2.70 3.68 6.36 11.34 2.67 3.63 6.23 10.97 2.64 3.59 6.11 10.66 2.62 3.55 6.01 10.39 2.61 3.52 5.93 10.16 2.59 3.49 5.85 9.95 2.73 3.71 6.55 12.55 2.66 3.59 6.22 11.56 2.61 3.49 5.95 10.80 2.56 3.41 5.74 10.21 2.52 3.34 5.56 9.73 2.49 3.29 5.42 9.34 2.46 3.24 5.29 9.01 2.44 3.20 5.18 8.73 2.42 3.16 5.09 8.49 2.40 3.13 5.01 8.28 2.38 3.10 4.94 8.10 2.61 3.48 5.99 11.28 2.54 3.36 5.67 10.35 2.48 3.26 5.41 9.63 2.43 3.18 5.21 9.07 2.39 3.11 5.04 8.62 2.36 3.06 4.89 8.25 2.33 3.01 4.77 7.94 2.31 2.96 4.67 7.68 2.29 2.93 4.58 7.46 2.27 2.90 4.50 7.27 2.25 2.87 4.43 7.10 2.52 3.33 5.64 10.48 2.45 3.20 5.32 9.58 2.39 3.11 5.06 8.89 2.35 3.03 4.86 8.35 2.31 2.96 4.69 7.92 2.27 2.90 4.56 7.57 2.24 2.85 4.44 7.27 2.22 2.81 4.34 7.02 2.20 2.77 4.25 6.81 2.18 2.74 4.17 6.62 2.16 2.71 4.10 6.46 2.46 3.22 5.39 9.93 2.39 3.09 5.07 9.05 2.33 3.00 4.82 8.38 2.28 2.92 4.62 7.86 2.24 2.85 4.46 7.44 2.21 2.79 4.32 7.09 2.18 2.74 4.20 6.80 2.15 2.70 4.10 6.56 2.13 2.66 4.01 6.35 2.11 2.63 3.94 6.18 2.09 2.60 3.87 6.02 2.41 3.14 5.20 9.52 2.34 3.01 4.89 8.66 2.28 2.91 4.64 8.00 2.23 2.83 4.44 7.49 2.19 2.76 4.28 7.08 2.16 2.71 4.14 6.74 2.13 2.66 4.03 6.46 2.10 2.61 3.93 6.22 2.08 2.58 3.84 6.02 2.06 2.54 3.77 5.85 2.04 2.51 3.70 5.69 2.38 3.07 5.06 9.20 2.30 2.95 4.74 8.35 2.24 2.85 4.50 7.71 2.20 2.77 4.30 7.21 2.15 2.70 4.14 6.80 2.12 2.64 4.00 6.47 2.09 2.59 3.89 6.19 2.06 2.55 3.79 5.96 2.04 2.51 3.71 5.76 2.02 2.48 3.63 5.59 2.00 2.45 3.56 5.44 2.35 3.02 4.94 8.96 2.27 2.90 4.63 8.12 2.21 2.80 4.39 7.48 2.16 2.71 4.19 6.98 2.12 2.65 4.03 6.58 2.09 2.59 3.89 6.26 2.06 2.54 3.78 5.98 2.03 2.49 3.68 5.75 2.00 2.46 3.60 5.56 1.98 2.42 3.52 5.39 1.96 2.39 3.46 5.24 Page 828 10 11 12 13 14 15 16 17 18 19 20 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 0.100 0.050 0.010 0.001 2.32 2.98 4.85 8.75 2.25 2.85 4.54 7.92 2.19 2.75 4.30 7.29 2.14 2.67 4.10 6.80 2.10 2.60 3.94 6.40 2.06 2.54 3.80 6.08 2.03 2.49 3.69 5.81 2.00 2.45 3.59 5.58 1.98 2.41 3.51 5.39 1.96 2.38 3.43 5.22 1.94 2.35 3.37 5.08 2.28 2.91 4.71 8.45 2.21 2.79 4.40 7.63 2.15 2.69 4.16 7.00 2.10 2.60 3.96 6.52 2.05 2.53 3.80 6.13 2.02 2.48 3.67 5.81 1.99 2.42 3.55 5.55 1.96 2.38 3.46 5.32 1.93 2.34 3.37 5.13 1.91 2.31 3.30 4.97 1.89 2.28 3.23 4.82 2.24 2.85 4.56 8.13 2.17 2.72 4.25 7.32 2.10 2.62 4.01 6.71 2.05 2.53 3.82 6.23 2.01 2.46 3.66 5.85 1.97 2.40 3.52 5.54 1.94 2.35 3.41 5.27 1.91 2.31 3.31 5.05 1.89 2.27 3.23 4.87 1.86 2.23 3.15 4.70 1.84 2.20 3.09 4.56 2.20 2.77 4.41 7.80 2.12 2.65 4.10 7.01 2.06 2.54 3.86 6.40 2.01 2.46 3.66 5.93 1.96 2.39 3.51 5.56 1.92 2.33 3.37 5.25 1.89 2.28 3.26 4.99 1.86 2.23 3.16 4.78 1.84 2.19 3.08 4.59 1.81 2.16 3.00 4.43 1.79 2.12 2.94 4.29 2.17 2.73 4.31 7.60 2.10 2.60 4.01 6.81 2.03 2.50 3.76 6.22 1.98 2.41 3.57 5.75 1.93 2.34 3.41 5.38 1.89 2.28 3.28 5.07 1.86 2.23 3.16 4.82 1.83 2.18 3.07 4.60 1.80 2.14 2.98 4.42 1.78 2.11 2.91 4.26 1.76 2.07 2.84 4.12 2.16 2.71 4.25 7.47 2.08 2.58 3.94 6.68 2.01 2.47 3.70 6.09 1.96 2.38 3.51 5.63 1.92 2.31 3.35 5.25 1.87 2.25 3.22 4.95 1.84 2.19 3.10 4.70 1.81 2.15 3.00 4.48 1.78 2.11 2.92 4.30 1.76 2.07 2.84 4.14 1.74 2.04 2.78 4.00 2.15 2.70 4.20 7.37 2.07 2.57 3.89 6.59 2.00 2.46 3.66 6.00 1.94 2.37 3.47 5.54 1.90 2.30 3.30 5.17 1.86 2.24 3.18 4.87 1.82 2.18 3.05 4.62 1.79 2.14 2.96 4.41 1.77 2.10 2.87 4.22 1.74 2.06 2.80 4.06 1.72 2.03 2.73 3.92 2.14 2.69 4.17 7.30 2.06 2.55 3.86 6.52 1.99 2.44 3.62 5.93 1.93 2.35 3.43 5.47 1.89 2.28 3.27 5.10 1.85 2.22 3.13 4.80 1.81 2.16 3.02 4.54 1.78 2.12 2.92 4.33 1.75 2.08 2.84 4.15 1.73 2.04 2.76 3.99 1.71 2.01 2.69 3.86 2.11 2.66 4.08 7.12 2.03 2.53 3.78 6.35 1.96 2.42 3.54 5.76 1.90 2.33 3.34 5.30 1.86 2.25 3.18 4.94 1.82 2.18 3.05 4.64 1.78 2.13 2.93 4.39 1.75 2.08 2.83 4.17 1.72 2.04 2.75 3.99 1.70 2.00 2.67 3.84 1.68 1.97 2.61 3.70 Page 829 α 21 0.100 0.050 0.010 0.001 22 0.100 0.050 0.010 0.001 23 0.100 0.050 0.010 0.001 24 0.100 0.050 0.010 0.001 25 0.100 0.050 0.010 0.001 26 0.100 0.050 0.010 0.001 27 0.100 0.050 0.010 0.001 28 0.100 0.050 0.010 0.001 29 0.100 0.050 0.010 0.001 30 0.100 0.050 0.010 0.001 31 0.100 0.050 0.010 0.001 2.96 4.32 8.02 14.59 2.95 4.30 7.95 14.38 2.94 4.28 7.88 14.20 2.93 4.26 7.82 14.03 2.92 4.24 7.77 13.88 2.91 4.23 7.72 13.74 2.90 4.21 7.68 13.61 2.89 4.20 7.64 13.50 2.89 4.18 7.60 13.39 2.88 4.17 7.56 13.29 2.87 4.16 7.53 13.20 2.57 3.47 5.78 9.77 2.56 3.44 5.72 9.61 2.55 3.42 5.66 9.47 2.54 3.40 5.61 9.34 2.53 3.39 5.57 9.22 2.52 3.37 5.53 9.12 2.51 3.35 5.49 9.02 2.50 3.34 5.45 8.93 2.50 3.33 5.42 8.85 2.49 3.32 5.39 8.77 2.48 3.30 5.36 8.70 2.36 3.07 4.87 7.94 2.35 3.05 4.82 7.80 2.34 3.03 4.76 7.67 2.33 3.01 4.72 7.55 2.32 2.99 4.68 7.45 2.31 2.98 4.64 7.36 2.30 2.96 4.60 7.27 2.29 2.95 4.57 7.19 2.28 2.93 4.54 7.12 2.28 2.92 4.51 7.05 2.27 2.91 4.48 6.99 2.23 2.84 4.37 6.95 2.22 2.82 4.31 6.81 2.21 2.80 4.26 6.70 2.19 2.78 4.22 6.59 2.18 2.76 4.18 6.49 2.17 2.74 4.14 6.41 2.17 2.73 4.11 6.33 2.16 2.71 4.07 6.25 2.15 2.70 4.04 6.19 2.14 2.69 4.02 6.12 2.14 2.68 3.99 6.07 2.14 2.68 4.04 6.32 2.13 2.66 3.99 6.19 2.11 2.64 3.94 6.08 2.10 2.62 3.90 5.98 2.09 2.60 3.85 5.89 2.08 2.59 3.82 5.80 2.07 2.57 3.78 5.73 2.06 2.56 3.75 5.66 2.06 2.55 3.73 5.59 2.05 2.53 3.70 5.53 2.04 2.52 3.67 5.48 2.08 2.57 3.81 5.88 2.06 2.55 3.76 5.76 2.05 2.53 3.71 5.65 2.04 2.51 3.67 5.55 2.02 2.49 3.63 5.46 2.01 2.47 3.59 5.38 2.00 2.46 3.56 5.31 2.00 2.45 3.53 5.24 1.99 2.43 3.50 5.18 1.98 2.42 3.47 5.12 1.97 2.41 3.45 5.07 2.02 2.49 3.64 5.56 2.01 2.46 3.59 5.44 1.99 2.44 3.54 5.33 1.98 2.42 3.50 5.23 1.97 2.40 3.46 5.15 1.96 2.39 3.42 5.07 1.95 2.37 3.39 5.00 1.94 2.36 3.36 4.93 1.93 2.35 3.33 4.87 1.93 2.33 3.30 4.82 1.92 2.32 3.28 4.77 1.98 2.42 3.51 5.31 1.97 2.40 3.45 5.19 1.95 2.37 3.41 5.09 1.94 2.36 3.36 4.99 1.93 2.34 3.32 4.91 1.92 2.32 3.29 4.83 1.91 2.31 3.26 4.76 1.90 2.29 3.23 4.69 1.89 2.28 3.20 4.64 1.88 2.27 3.17 4.58 1.88 2.25 3.15 4.53 Page 830 21 0.100 0.050 0.010 0.001 22 0.100 0.050 0.010 0.001 23 0.100 0.050 0.010 0.001 24 0.100 0.050 0.010 0.001 25 0.100 0.050 0.010 0.001 26 0.100 0.050 0.010 0.001 27 0.100 0.050 0.010 0.001 28 0.100 0.050 0.010 0.001 29 0.100 0.050 0.010 0.001 30 0.100 0.050 0.010 0.001 31 0.100 0.050 0.010 0.001 1.95 2.37 3.40 5.11 1.93 2.34 3.35 4.99 1.92 2.32 3.30 4.89 1.91 2.30 3.26 4.80 1.89 2.28 3.22 4.71 1.88 2.27 3.18 4.64 1.87 2.25 3.15 4.57 1.87 2.24 3.12 4.50 1.86 2.22 3.09 4.45 1.85 2.21 3.07 4.39 1.84 2.20 3.04 4.34 1.92 2.32 3.31 4.95 1.90 2.30 3.26 4.83 1.89 2.27 3.21 4.73 1.88 2.25 3.17 4.64 1.87 2.24 3.13 4.56 1.86 2.22 3.09 4.48 1.85 2.20 3.06 4.41 1.84 2.19 3.03 4.35 1.83 2.18 3.00 4.29 1.82 2.16 2.98 4.24 1.81 2.15 2.96 4.19 1.90 2.28 3.24 4.83 1.88 2.26 3.18 4.70 1.87 2.24 3.14 4.60 1.85 2.22 3.09 4.51 1.84 2.20 3.06 4.42 1.83 2.18 3.02 4.35 1.82 2.17 2.99 4.28 1.81 2.15 2.96 4.22 1.80 2.14 2.93 4.16 1.79 2.13 2.91 4.11 1.78 2.11 2.88 4.06 1.87 2.25 3.17 4.70 1.86 2.23 3.12 4.58 1.84 2.20 3.07 4.48 1.83 2.18 3.03 4.39 1.82 2.16 2.99 4.31 1.81 2.15 2.96 4.24 1.80 2.13 2.93 4.17 1.79 2.12 2.90 4.11 1.78 2.10 2.87 4.05 1.77 2.09 2.84 4.00 1.76 2.08 2.82 3.95 1.83 2.18 3.03 4.44 1.81 2.15 2.98 4.33 1.80 2.13 2.93 4.23 1.78 2.11 2.89 4.14 1.77 2.09 2.85 4.06 1.76 2.07 2.81 3.99 1.75 2.06 2.78 3.92 1.74 2.04 2.75 3.86 1.73 2.03 2.73 3.80 1.72 2.01 2.70 3.75 1.71 2.00 2.68 3.70 1.79 2.12 2.88 4.17 1.78 2.10 2.83 4.06 1.76 2.07 2.78 3.96 1.74 2.05 2.74 3.87 1.73 2.03 2.70 3.79 1.72 2.01 2.66 3.72 1.71 1.99 2.63 3.66 1.70 1.98 2.60 3.60 1.69 1.96 2.57 3.54 1.68 1.95 2.55 3.49 1.67 1.94 2.52 3.44 1.76 2.08 2.79 4.03 1.74 2.05 2.73 3.92 1.73 2.03 2.69 3.82 1.71 2.01 2.64 3.74 1.70 1.99 2.60 3.66 1.69 1.97 2.57 3.59 1.68 1.95 2.54 3.52 1.67 1.94 2.51 3.46 1.66 1.92 2.48 3.41 1.65 1.91 2.45 3.36 1.64 1.90 2.43 3.31 Page 831 32 0.100 0.050 0.010 0.001 33 0.100 0.050 0.010 0.001 ... 1.74 2.05 2.74 3.95 ... 1.81 2.15 2.94 4.14 1.80 2.14 2.92 4.10 ... 1.87 2.24 3.13 4.56 ... 1.91 2.30 3.26 4.85 ... 1.96 2.38 3.43 5.07 ... 2.04 2.51 3.67 5.48 ...

Page 832 ... Page 833 TABLE A.9 Upper percentage points for the Studentized range $q_{ν1,ν2}$ ... Page 834 ... Page 835 ... TABLE A.10 Control chart constants A2 A3 B3 B4 D3 D4 c4 d2 ... 2 1.880 2.659 0.000 3.267 0.000 3.267 0.7979 1.128 ... For n > 25: ... Page 836

Appendix B Partial Derivatives This appendix presents the mechanics of computing partial derivatives, which are needed in Section 3.4. We begin by recalling that a derivative specifies the rate of change of one variable with respect to another. For example, the volume v of a sphere whose radius is r is given by $v = 4\pi r^3/3$. If r is allowed to increase (or decrease), the rate at which v increases (or decreases) is given by the derivative of v with respect to r: $dv/dr = 12\pi r^2$. Partial derivatives are needed when the quantity whose rate of change is to be calculated is a function of more than one variable. Here is an example: The volume of a cylinder with radius r and height h is given by $v = \pi r^2 h$. If either r or h changes, v will change as well. Now imagine that h is constant, and r is allowed to increase. The rate of increase in v is the partial derivative of v with respect to r. This derivative is denoted $\partial v/\partial r$, and it is computed exactly like the ordinary derivative of v with respect to r, treating h as a constant: $\partial v/\partial r = 2\pi rh$. Now assume that r is constant and h is increasing. The rate of increase in v is the partial derivative of v with respect to h, denoted $\partial v/\partial h$. It is computed exactly like the ordinary derivative of v with respect to h, treating r as a constant: $\partial v/\partial h = \pi r^2$. If v is a function of several variables, $v = f(x_1, x_2, ..., x_n)$, the partial derivative of v with respect to any one of the variables $x_1, x_2, ..., x_n$ is computed just like the ordinary derivative, holding the other variables constant. Examples B.1 and B.2 show that computing partial derivatives is no more difficult than computing ordinary derivatives. Example B.1 Let $v = 12x^2y + 3xy^2$. Find the partial derivatives of v with respect to x and y. Solution To compute $\partial v/\partial x$, hold y constant, and compute the derivative with respect to x. The result is Page 837 To compute $\partial v/\partial y$, hold x constant and compute the derivative with respect to y. The result is Example B.2 . Find the partial derivatives of v with respect to x, y, and z. Let Solution To compute $\partial v/\partial x$, hold both y and z constant and compute the derivative with respect to x, using the quotient rule: Similarly, we compute the partial derivatives of v with respect to y and z: Exercises for Appendix B In the following exercises, compute all partial derivatives. 1. $v = 3x + 2xy^4$ 2. 3. $v = \cos x \sin y$ 2 4. $v = e^{xy}$ 5. $v = e^x(\cos y + \sin z)$ 6. 7. 8. 9. 10. 2 3 $z = \ln(x^2 + y^2)$ 3 $v = e^{y^2}\cos(xz) + \ln(x^2y + z)$ Page 838

Appendix C Bibliography Agresti, A. (2013). Categorical Data Analysis, 3rd ed. John Wiley & Sons, New York. An authoritative and comprehensive treatment of the subject. Belsley, D., Kuh, E., and Welsch, R. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, New York. A presentation of methods for evaluating the reliability of regression estimates. Bevington, P., and Robinson, D. (2003). Data Reduction and Error Analysis for the Physical Sciences, 3rd ed. McGraw-Hill, Boston. An introduction to data analysis, with emphasis on propagation of error and calculation. Bickel, P., and Doksum, K. (2015). Mathematical Statistics: Basic Ideas and Selected Topics, Vol. I, 2nd ed. CRC Press, Boca Raton, FL. A thorough treatment of the mathematical principles of statistics, at a fairly advanced level. Box, G., and Draper, N. (1987). Empirical Model-Building and Response Surfaces. John Wiley & Sons, New York. An excellent practical introduction to the fitting of curves and higher-order surfaces to data. Box, G., Hunter, W., and Hunter, J. (2005). Statistics for Experimenters, 2nd ed. John Wiley & Sons, New York. A very intuitive and practical introduction to the basic principles of data analysis and experimental design. Brockwell, R., and Davis, R. (2016). Introduction to Time Series and Forecasting, 3rd ed. Springer International Publishing, Switzerland. An excellent introductory text at the undergraduate level, more rigorous than Chatfield (2003). Casella, G., and Berger, R. (2002). Statistical Inference, 2nd ed. Duxbury, Pacific Grove, CA. A fairly rigorous development of the theory of statistics. Chatfield, C. (1983). Statistics for Technology, 3rd ed., revised. Chapman and Hall/CRC, Boca Raton, FL. A clear and concise introduction to basic principles of statistics, oriented toward engineers and scientists. Chatfield, C. (2003). An Analysis of Time Series: An Introduction, 6th ed. CRC Press, Boca Raton, FL. An intuitive presentation, at a somewhat less advanced level than Brockwell and Davis (2016). Page 839 Cochran, W. (1977). Sampling Techniques, 3rd ed. John Wiley & Sons, New York. A comprehensive account of sampling theory. Cook, D., and Weisberg, S. (1994). An Introduction to Regression Graphics. John Wiley & Sons, New York. A presentation of interactive methods for analyzing data with linear models. DeGroot, M., and Schervish, M. (2012). Probability and Statistics, 4th ed. Addison-Wesley, Reading, MA. A very readable introduction at a somewhat higher mathematical level than this book. Draper, N., and Smith, H. (1998). Applied Regression Analysis, 3rd ed. John Wiley & Sons, New York. An extensive and authoritative treatment of linear regression. Efron, B., and Tibshirani, R. (1994). An Introduction to the Bootstrap. Chapman and Hall, New York. A clear and comprehensive introduction to bootstrap methods. Freedman, D., Pisani, R., and Purves, R. (2007). Statistics, 4th ed. Norton, New York. An excellent intuitive introduction to the fundamental principles of statistics. Hocking, R. (2014). Methods and Applications of Linear Models: Regression and the Analysis of Variance, 3rd ed. John Wiley & Sons, New York. A thorough treatment of the theory and applications of regression and analysis of variance. James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer, New York. A practical introduction to statistical learning. Kenett, R., and Zacks, S. (2014). Modern Industrial Statistics: With Applications in R, MINITAB, and JMP. Brooks/Cole, Pacific Grove, California. An up-to-date treatment of the subject with emphasis on industrial engineering. Larsen, R., and Marx, M. (2012). An Introduction to Mathematical Statistics and Its Applications, 5th ed. Pearson, Boston, MA. An introduction to statistics at a higher mathematical level than that of this book. Contains many good examples. Lee, P. (2013). Bayesian Statistics: An Introduction, 4th ed. Wiley, New York. A clear and basic introduction to statistical methods that are based on the subjective view of probability. Lehmann, E. (1998). Nonparametrics: Statistical Methods Based on Ranks. Pearson, New York. Thorough presentation of basic distribution-free methods. Miller, A. (2002). Subset Selection in Regression, 2nd ed. Chapman and Hall, London. A strong and concise treatment of the basic principles of model selection. Miller, R. (1997). Beyond ANOVA: The Basics of Applied Statistics. Chapman and Hall/CRC, Boca Raton, FL. A very practical and intuitive treatment of methods useful in analyzing real data, when standard assumptions may not be satisfied. Page 840 Montgomery, D. (2013a). Design and Analysis of Experiments, 8th ed. John Wiley & Sons, New York. A thorough exposition of the methods of factorial experiments, focusing on engineering applications. Montgomery, D. (2013b). Introduction to Statistical Quality Control, 7th ed. John Wiley & Sons, New York. A comprehensive and readable introduction to the subject. Mood, A., Graybill, F., and Boes, D. (1974). Introduction to the Theory of Statistics, 3rd ed. McGraw-Hill, Boston. A classic introduction to mathematical statistics and an excellent reference. Mosteller, F., and Tukey, J. (1977). Data Analysis and Regression. Addison-Wesley, Reading, MA. An intuitive and philosophical presentation of some very practical ideas. Rice, J. (2006). Mathematical Statistics and Data Analysis, 3rd ed. Wadsworth, Belmont, CA. A good blend of theory and practice, at a somewhat higher level than this book. Ross, S. (2014). A First Course in Probability, 9th ed. Pearson, Boston, MA. A mathematically sophisticated introduction to probability. Ross, S. (2014). Introduction to Probability and Statistics for Engineers and Scientists, 5th ed. Harcourt/Academic Press, San Diego. An introduction at a somewhat higher mathematical level than this book. Salsburg, D. (2001). The Lady Tasting Tea. W. H. Freeman and Company, New York. An insightful discussion of the influence of statistics on 20th century science, with many fascinating anecdotes about famous statisticians. The story in Section 4.3 about Student's discovery that the number of particles in a suspension follows a Poisson distribution can be found in this book. Taylor, J. (1997). An Introduction to Error Analysis, 2nd ed. University Science Books, Sausalito, CA. A thorough treatment of propagation of error, along with a selection of other topics in data analysis. Tufte, E. (2001). The Visual Display of Quantitative Information, 2nd ed. Graphics Press, Cheshire, CT. A clear and compelling demonstration of the principles of effective statistical graphics, containing numerous examples. Tukey, J. (1977). Exploratory Data Analysis. Addison-Wesley, Reading, MA. A wealth of techniques for summarizing and describing data. Wackerly, D., Mendenhall, W., and Scheaffer, R. (2008). Mathematical Statistics with Applications, 7th ed. Duxbury, Pacific Grove, CA. An introduction to statistics at a somewhat higher mathematical level than this book. Weisberg, S. (2013). Applied Linear Regression, 4th ed. John Wiley & Sons, New York. A concise introduction to the application of linear regression models, including diagnostics, model building, and interpretation of output. Page 841

Answers to Odd-Numbered Exercises Section 1.1 1. (a) The population consists of all the times the process could be run. It is conceptual. (b) The population consists of all the registered voters in the state. It is tangible. (c) The population consists of all people with high cholesterol levels. It is tangible. (d) The population consists of all concrete specimens that could be made from the new formulation. It is conceptual. 3. (a) False (b) True 5. (a) No. What is important is the population proportion of defectives; the sample percentage is only an approximation. The population proportion for the new process may in fact be greater or less than that of the old process. (b) No. The population proportion for the new process may be 12% or more, even though the sample proportion was only 11%. 7. (b) No. (c) Finding two defective circuits in the sample. A good knowledge of the process that generated the data. 9. (a) A controlled experiment. (b) Yes, because it is based on a controlled experiment rather than an observational study. Section 1.2 1. 3. 5. 7. 9. 11. 13. False No. In the sample 1, 2, 4 the mean is 7/3, which does not appear at all. The sample size can be any odd number. Yes. If all the numbers on the list are the same, the standard deviation will equal 0. The mean and standard deviation both increase by 5%. 169.6 cm (a) All would be divided by 2.54. (b) Not exactly the same, because the measurements would be a little different the second time. 15. (a) The tertiles are 45 and 77.5. (b) The quintiles are 32, 47.5, 75, and 85.5. Page 842 Section 1.3 1. (a) Stem Leaf 0 011112335677 1 235579 2 468 3 1257 4 14699 5 5 6 16 7 9 8 0 9 1 10 0 11 12 7 13 7 (b) Here is one histogram. Other choices for the end points are possible. (c) (d) The boxplot shows one outlier. 3. Page 843 Stem Leaf 1 1588 2 00003468 3 0234588 4 0346 5 2235666689 6 00233459 7 113558 8 568 9 1225 10 1 11 12 2 13 06 14 15 16 17 1 18 6 19 9 20 21 22 23 3 There are 23 stems in this plot. An advantage of this plot over the one in Figure 1.6 is that the values are given to the tenths digit instead of to the ones digit. A disadvantage is that there are too many stems, and many of them are empty. 5. (a) Page 844 (b) (c) The yields for catalyst B are considerably more spread out than those for catalyst A. The median yield for catalyst A is greater than the median for catalyst B. The median yield for B is closer to the first quartile than the third, but the lower whisker is longer than the upper one, so the median is approximately equidistant from the extremes of the data. Thus the yields for catalyst B are approximately symmetric. The largest yield for catalyst A is an outlier; the remaining yields for catalyst A are approximately symmetric. 7. (a) Closest to 50% (b) 240–260 mg/dL (a) (b) (c) 0.7 9. (c) The shape of the histograms are the same. 11. (a) (b) The value 100 is an outlier. Page 845 13. (a) iii. (b) 5.86 (c) 4.60 (b) 3.86 (b) Yes. (c) No. The minimum value of −2.235 is an "outlier," since it is more than 1.5 times the interquartile range below the first quartile. The lower whisker should extend to the smallest point that is not an outlier, but the value of this point is not given. 15. (a) (b) The boxplot indicates that the value 470 is an outlier. (c) (d) The dotplot indicates that the value 384 is detached from the bulk of the data, and thus could be considered to be an outlier. 19. (a) The mean is 0.7619; the median is 1. (b) Skewed to the left. (c) Skewed to the right. Page 846 21. (a) The relationship is nonlinear. (b) The relationship is approximately linear. (c) Yes. It would be easier to work with x and ln y, because the relationship is approximately linear. Supplementary Exercises for Chapter 1 1. (a) The mean will be divided by 2.2. (b) The standard deviation will be divided by 2.2. 3. (a) False (b) True (c) False (d) True 5. (a) It is not possible to tell how much the mean changes. (b) If there are more than two numbers on the list, the median is unchanged. If there are only two numbers on the list, the median is changed, but we cannot tell by how much. (c) It is not possible to tell how much the standard deviation changes. 7. (a) The mean decreases by 0.774. (b) The value of the mean after the change is 25 − 0.774 = 24.226. (c) The median is unchanged. (d) It is not possible to tell by how much the standard deviation changes. 9. Statement (i) is true. 11. (a) Incorrect (b) Correct (c) Incorrect (d) Correct Page 847 13. (a) Skewed to the left. The 85th percentile is much closer to the median (50th percentile) than the 15th percentile is. Therefore the histogram is likely to have a longer left-hand tail than right-hand tail. (b) Skewed to the right. The 15th percentile is much closer to the median (50th percentile) than the 85th percentile is. Therefore the histogram is likely to have a longer right-hand tail than left-hand tail. 15. (a) (b) 0.14 (c) Approximately symmetric (d) The data on the raw scale are skewed so much to the right that it is impossible to see the features of the histogram. Page 848 17. (b) 3.35 (c) 1.88 (d) 7.70 19. (a) (b) Each sample contains one outlier. (c) In the Sacaton boxplot, the median is about midway between the first and third quartiles, suggesting that the data between these quartiles are fairly symmetric. The upper whisker of the box is much longer than the lower whisker, and there is an outlier on the upper side. This indicates that the data as a whole are skewed to the right. In the Gila Plain boxplot data, the median is about midway between the first and third quartiles, suggesting that the data between these quartiles are fairly symmetric. The upper whisker is slightly longer than the lower whisker, and there is an outlier on the upper side. This suggests that the data as a whole are somewhat skewed to the right. In the Casa Grande boxplot, the median is very close to the first quartile. This suggests that there are several values very close to each other about one-fourth of the way through the data. The two whiskers are of about equal length, which suggests that the tails are about equal, except for the outlier on the upper side. Section 2.1 1. 0.8 3. 0.13 (a) {TTTT, TTTF, TTFT, TTFF, TFTT, TFTF, TFFT, TFFF, FTTT, FTTF, FTFT, FTFF, FFTT, FFTF, FFFT, FFFF} (b) 1/8 (c) 1/4 (d) 5/16 Page 849 5. (a) {1, 2, 31, 32, 41, 42, 341, 342, 431, 432} (b) {1, 2} (c) {341, 342, 431, 432} (d) {31, 32, 341, 342, 431, 432} (e) {1, 31, 41, 341, 431} (f) A and E are mutually exclusive because they have no outcomes in common. B and E are not mutually exclusive because they both contain the outcomes 341 and 431. 7. (a) 0.6 (b) 0.9 (c) 0.4 (d) 0.3 9. (a) 0.88 (b) 0.12 (c) 0.07 11. (a) 0.40 (b) 0.79 (c) 0.57 (d) 0.79 (e) 0.89 (f) 0.63 13. (a) 0.5 (b) 0.1 (c) 0.4 15. (a) 0.95 11. (a) False (b) True (c) False (d) True Section 2.2 1. (a) 64 (b) 8 3. 5. (a) 24 (b) 2.6 … 6 7. 6,306 9. (a) 1/9 (b) 1.08 (c) 2.8211 × 10^{12} (b) 34,650 (c) 0.0033 … 11. (a) 5.7×10^{-5} … 15. (a) 0.9 (b) 0.1 (c) 0.5 (d) 0.5 (e) … 17. 0.9 17. 0.8 Section 2.3 1. 0.25 (a) 1/3 (b) 5/14 5. 7. 0.2 (a) 0.20 (b) 0.95 … 9. (a) 0.5 (b) 0.3 (c) 0.2 (d) 0.4 … (b) 0.3 (c) 0.2 (d) 0.1 (e) 0.4 11. (a) (b) 0.12 (c) 0.9 (d) 0.7 (e) 0.7333 … 13. 0.0625 … Page 850 21. (a) That the gauges fail independently. (b) One cause of failure, a fire, will cause both gauges to fail. Therefore, they do not fail independently. (c) Too low. The correct calculation would use P(second gauge fails|first gauge fails) in place of P(second gauge fails). Because there is a chance that both gauges fail together in a fire, the condition that the first gauge fails makes it more likely that the second gauge fails as well. Therefore P(second gauge fails|first gauge fails) > P(second gauge fails). 23. (a) 3/10 (b) 2/9 (c) 1/15 (d) 7/30 (e) 3/10 (f) No. P(B) ≠ P(B|A) [or P(A ∩ B) ≠ P(A)P(B)]. 25. n = 10,000. The two components are a simple random sample from the population. When the population is large, the items in a simple random sample are nearly independent. 27. (a) 0.99 (b) 0.01 (b) 0.0033 31. (a) 9/10 (b) 14/45 (c) 1/3 (b) 1/4 (c) 1/4 (d) 1/4 (e) Yes (f) No … 33. (a) 0.10 (b) 0.02 … 0.9801 35. 0.9125 37. (a) 0.9904 (b) 0.1 (c) 0.2154 (d) 7.39 41. (a) P(A) = 1 − P(A) … (b) Since A and B are independent … P(B ∩ A) = P(B)P(A) … Therefore, B and A are independent. … 0.40 Section 2.4 1. (a) Discrete (b) Continuous (c) Discrete (d) Continuous 3. (a) 2.3 (b) 1.81 (c) 1.345 (d) … (e) 0.9 (f) 0.3 … 13.45 5. (a) 0.85 (c) 0.05 … 7. (a) c = 0.1 (b) 0.2 … (a) 0.4 (b) 0.9 (c) 2.2 (d) 0.96 … 0.0655 … The probabilities of 4 and 1 for p(x) are much smaller than the observed proportions. (d) No, this is not right. The data are a simple random sample, and the model represents the population. Simple random samples generally do not reflect the population exactly. Page 851 11. (a) … (b) 0.09 (c) 0.147 (d) 0.0133 … (a) 10 months (b) 10 months (c) … 13. (a) 0.0016 (b) … 15. (a) 67/256 … 17. (a) 1/2 … 19. (a) 0.1 … 21. … 23. … 25. … Page 852 … Page 853 … Page 854 … Page 855 … Page 856 … Page 857 … Page 858 … Page 859 … Page 860 … Page 861 … Page 862 … Page 863 … Page 864 … Page 865 … Page 866 … Page 867 … Page 868 … Section 6.1 1. (a) 0.0014 (b) If the mean number of sick days were 5.4, the probability is only 0.0014 of observing a sample with a mean less than or equal to the mean of 4.5 that was actually observed. Since this is a very small probability, we are convinced that the mean number of sick days is less than 5.4. 3. (a) 0.0057 (b) If the mean depth were 900 μm or more, the probability is only 0.0057 of observing a sample mean as small as the value of 850 that was actually observed. Therefore we are convinced that the mean depth is not 900 μm or more, but instead is less than 900 μm. (a) P = 0.1131 11. 13. 15. 17. (b) If the profit margin were 10%, the probability is 0.1131 of observing a sample with a mean less than or equal to the mean of 8.24 that was actually observed. A probability of 0.1131 is not small enough to reject H0, so it is plausible that the mean profit margin is 10% or more.

[illegible]

strong evidence that the population medians were in fact different. (c) No, the X sample is heavily skewed to the right, while the Y sample is strongly bimodal. It does not seem reasonable to assume that these samples came from populations of the same shape. 19. (a) Let μA be the mean thrust/weight ratio for fuel A, and let μB be the mean thrust/weight ratio for fuel B. The appropriate null and alternate hypotheses are H0: μA − μB ≤ 0 vs. H1: μA − μB > 0. (b) Yes. t29 = 2.0339, 0.025 < P < 0.05 (P = 0.02560). 21. (a) Yes. (b) The conclusion is not justified. The engineer is concluding that H0 is true because the test failed to reject. 23. No. . Section 7.1 1. 3. 0.8214 (a) The correlation coefficient is appropriate. The points are approximately clustered around a line. (b) The correlation coefficient is not appropriate. The relationship is curved, not linear. 5. 7. (c) The correlation coefficient is not appropriate. The plot contains outliers. More than 0.6 (a) Between temperature and yield, r = 0.7323; between temperature and stirring rate, r = 0.7513; between temperature and stirring rate, r = 0.9064. (b) No, the result might be due to confounding, since the correlation between temperature and stirring rate is far from 0. (c) No, the result might be due to confounding, since the correlation between temperature and stirring rate is far from 0. 9. (a) (0.2508, 0.8933) (b) No. z = 0.99, P = 0.1611 (c) Yes, t12 = 3.2989, 0.001 < P < 0.005 (P = 0.003177). 11. (−0.95256, −0.95042) 13. Yes, t21 = −2.710, 0.01 < P < 0.02 (P = 0.0131). Section 7.2 1. (a) 319.27 lb 3. 5. (b) 5.65 lb 0.8492 (a) 18.869 in. (b) 70.477 in. 7. (c) No, some of the men whose points lie below the least-squares line will have shorter arms. y = 19.499 + 1.3458x. Page 875 9. (a) The linear model is appropriate. (b) y = 33.775 + 0.59299x (c) 1.186 degrees (d) 81.81 degrees 11. (a) The linear model is appropriate. (b) y = 4.3416 − 4.9051x (c) 0.9810 Hz (d) 0.6627 Hz (e) No, because all the modes in the data set have damping ratios less than 1. (f) 0.47738 13. (a) The linear model is appropriate. Page 876 (b) y = 76.5388 − 0.5606x. (c) Fitted Values 52.433 73.736 66.448 63.084 58.600 70.933 64.766 (d) Decrease by 5.606 parts per hundred thousand. (e) 59.721 (f) No. That value is outside the range of the data. (g) 20.583 15. (iii) 17. Carrie Section 7.3 1. (a) (b) 17.996 (c) For β0:(−0.744, 15.991), for β1 (0.208, 0.451) (d) Yes. t10 = −3.119, 0.005 < P < 0.01 (P = 0.00545). (e) (16.722, 24.896) (f) (10.512, 31.106) Residuals 2.567 6.264 −8.448 4.916 −1.600 −1.933 −1.766 3. (a) The slope is −0.7524; the intercept is 88.761. (b) Yes, the P-value for the slope is ≈ 0, so humidity is related to ozone level. (c) 51.14 ppb (d) −0.469 (e) (41.6, 45.6) (f) 5. No. A reasonable range of predicted values is given by the 95% prediction interval, which is (20.86, 66.37). (a) H0: βA − βB = 0 7. (b) Yes. z = −4.55, P ≈ 0. (a) y = 71.436 + 0.23924x (b) 0.29837 (c) (0.16687, 0.31162) (d) (78.322, 78.905) (e) Yes, t6 = 5.151, 0.001 < P < 0.005(P = 0.00106) 9. (f) (77.827, 79.400) (a) y = −0.32584 + 0.22345x. (b) For β0, (−2.031, 1.379), for β1, (0.146, 0.301). (c) 4.14 (d) (3.727, 4.559) (e) (1.585, 6.701) 11. The confidence interval for 20 Ω would be the shortest. The confidence interval for 15 Ω would be the longest. 13. 1.388 15. (a) 0.256 (b) 0.80 (c) 1.13448 (d) 0.001 17. (a) 553.71 (b) 162.06 (c) Below (d) There is a greater amount of vertical spread on the right side of the plot than on the left. Section 7.4 1. (a) ln y = −0.4442 + 0.79833 ln x (b) 330.95 (c) 231.76 3. (d) (53.19, 1009.89) (a) y = 20.162 + 1.269x Page 877 (b) There is no apparent pattern to the residual plot. The linear model looks fine. (c) 5. The residuals increase over time. The linear model is not appropriate as is. Time, or other variables related to time, must be included in the model. (a) y = 30.715 + 0.37279x (b) 0.94789 (c) The relationship does not appear to be linear. (d) False Page 878 7. (a) y = −235.32 + 0.695x. (b) The residual plot shows a pattern, with positive residuals at the higher and lower fitted values, and negative residuals in the middle. The model is not appropriate. (c) ln y = −0.0745 + 0.925 ln x. (d) The residual plot shows no obvious pattern. The model is appropriate. 9. (e) The log model is more appropriate. The 95% prediction interval is (197.26, 1559.76). (a) California is an outlier. (b) y = 30.1218 + 0.5958x (c) y = −22.7145 + 1.3046x (d) Yes. The slope and intercept change substantially. Page 879 11. (a) The least-squares line is y = 0.833 + 0.235x. (b) The least-squares line is y = 0.199 + 1.207 ln x. (c) The least-squares line is ln y = −0.0679 + 0.137x. Page 880 (d) The least-squares line is . (e) The model y = 0.199 + 1.207 ln x fits best. Its residual plot shows the least pattern. (f) The residuals show no pattern with time. (g) 2.14 (h) (1.689, 2.594) 13. (a) The model is log10 y = β0 + β1 log10 x+ε. Note that the natural log (ln) could be used in place of log10, but common logs are more convenient since partial pressures are expressed as powers of 10. Page 881 (b) The least-squares line is log10 y = − 3.277 − 0.225 log10 x. The linear model appears to fit quite well. (c) The theory says that the coefficient β1 of log10 x in the linear model is equal to −0.25. The estimated value is . We determine whether the data are consistent with the theory by testing the hypotheses H0: β1 = −0.25 vs. H1: β1 ≠ −0.25. The value of the test statistic is t21 = 0.821, so 0.20 < P < 0.50. We do not reject H0,so the data are consistent with the theory. 15. (a) y = 2049.87 − 4.270x (b) (12, 2046) and (13, 1954) are outliers. The least-squares line with (12, 2046) deleted is y = 2021.85 − 2.861x. The least-squares line with (13, 1954) deleted is y = 2069.30 − 5.236x.The least-squares line with both outliers deleted is y = 2040.88 − 3.809x. (c) The slopes of the least-squares lines are noticeably affected by the outliers. They ranged from −2.861 to −5.236. 17. ln W = β0 + β1 ln L + ε,where β0 = ln a and β1 = b. 19. (a) A physical law. (b) It would be better to redo the experiment. If the results of an experiment violate a physical law, then something was wrong with the experiment, and you can't fix it by transforming variables. Supplementary Exercises for Chapter 7 1. (a) (b) 0 (c) The molar absorption coefficient M. (d) Testing H0: β0 = 0 vs. H1: β0 ≠ 0, t3 = −0.428 and 0.50 < P < 0.80 (P = 0.697), so the data are consistent with the Beer–Lambert law. Page 882 3. (a) (b) Ti+1 = 120.18 − 0.696Ti (c) (−0.888, −0.503) (d) 71.48 minutes (e) (68.40, 74.56) 5. (f) (45.00, 97.95) (a) (b) No. t9 = 1.274, 0.20 < P < 0.50 (P = 0.235). (c) Yes. t9 = −5.358, P < 0.001 (P = 0.000457). (d) Yes, since we can conclude that β1 ≠ 1, we can conclude that the machine is out of calibration. (e) (18.58, 20.73) (f) (75.09, 77.23) (g) No, when the true value is 20, the result of part 7. (e) shows that a 95% confidence interval for the mean of the measured values is (18.58, 20.73). Therefore it is plausible that the mean measurement will be 20, so that the machine is in calibration. (a) y = −2.6903 + 0.0391x (b) For β0:(−6.0171, 6.6366), for β1:(0.0030, 0.0752). (c) 0.9465 (d) (0.7848,1.1082) 9. (e) (0.1118, 1.7812) (a) ln y = β0 + β1 ln x,where β0 = ln k and β1 = r. (b) The least-squares line is ln y = −1.7058 + 0.65033 ln x. Therefore . (c) t3 = 4.660, P = 0.019. No, it is not plausible. and 11. (a) y = 337.13 + 0.098006x. (b) Page 883 (c) ln y = −0.46584 + 0.81975 ln x. (d) (e) (38.75, 5103.01) 13. (a) −0.87015 (b) y = 814.89 − 143.33x (c) The residuals corresponding to x ≥ 0.5 are more spread out than those corresponding to x < 0.5. (d) y = 805.82 − 111.13x (e) y = 764.52 − 60.534x (f) 0.5689 (g) 0.1162 (h) less than or equal to 0.5 15. (ii) 17. (a) 145.63 (b) . Note that r is negative because the slope of the leastsquares line is negative. (c) 145.68. 19. (a) We need to minimize the sum of squares respect to and set it equal to 0 , so . (b) Let obtaining . Then .We take the derivative with . Then . Page 884 21. From the answer to Exercise 20, we know that . Now , and 23. Section 8.1 1. (a) 49.617 kg/mm2 (b) 33.201 kg/mm2 (c) 2.1245 kg/mm2 3. There is no obvious pattern to the residual plot, so the linear model appears to fit well. Page 885 5. (a) 25.465 (b) No, the predicted change depends on the values of the other independent variables, because of the interaction terms. (c) 0.9691 (d) F9,17 = 59.204. Yes, the null hypothesis can be rejected. 7. (a) 2.3411 L (b) 0.06768 L 9. (c) Nothing is wrong. In theory, the constant estimates FEV1 for an individual whose values for the other variables are all equal to zero. Since these values are outside the range of the data (e.g., no one has zero height), the constant need not represent a realistic value for an actual person. (a) 3.572 (b) 0.098184 (c) Nothing is wrong. The constant estimates the pH for a pulp whose values for the other variables are all equal to zero. Since these values are outside the range of the data (e.g., no pulp has zero density), the constant need not represent a realistic value for an actual pulp. (d) (3.4207, 4.0496) (e) (2.2333, 3.9416) (f) Pulp B. The standard deviation of its predicted pH (SE Fit) is smaller than that of pulp A (0.1351 vs. 0.2510). 11. (a) −2.05 (b) 0.3512 (c) −0.2445 (d) 4.72 (e) 13.92 (f) 18.316 (g) 4.54 (h) 9 13. (a) 135.92° F (b) No. The change in the predicted flash point due to a change in acetic acid concentration depends on the butyric acid concentration as well, because of the interaction between these two variables. (c) Yes. The predicted flash point will change by −13.897°F. 15. (a) 0.2286, −0.5743, 0.3514, 0.1057, −0.1114, 0.0000 (b) SSE = 0.5291, SST = 16.7083 (c) s2 = 0.1764 (d) R2 = 0.9683 (e) F = 45.864. There are 2 and 3 degrees of freedom. (f) Yes, the P-value corresponding to the F statistic with 2 and 3 degrees of freedom is between 0.001 and 0.01, so it is less than 0.05. 17. (a) 2.0711 (b) 0.17918 (c) PP is more useful, because its P-value is small, while the P-value of CP is fairly large. (d) The percent change in GDP would be expected to be larger in Sweden, because the coefficient of PP is negative. 19. (a) y = −0.012167 + 0.043258t + 2.9205t2 (b) (2.830, 3.011) (c) (5.660, 6.022) (d) , P = 0.278, , P = 0.350, , P = 0.000. (e) No, the P-value of 0.278 is not small enough to reject the null hypothesis that β0 = 0. (f) No, the P-value of 0.350 is not small enough to reject the null hypothesis that β1 = 0. Section 8.2 1. (a) Predictor Constant x1 Coef 6.3347 1.2915 StDev 2.1740 0.1392 T 2.9138 9.2776 β0 differs from 0 (P = 0.009), ft differs from 0 (P = 0.000). P 0.009 0.000 (b) Predictor Constant x2 Coef 53.964 −0.9192 StDev 8.7737 0.2821 T 6.1506 −3.2580 P 0.000 0.004 β0 differs from 0 (P = 0.000), ft differs from 0 (P = 0.004). (c) Predictor Constant x1 Coef 12.844 1.2029 StDev 7.5139 0.1707 T 1.7094 7.0479 P 0.104 0.000 x2 −0.1682 0.1858 −0.90537 0.377 β0 may not differ from 0 (P = 0.104), β1 differs from 0 (P = 0.000), β2 may not differ from 0 (P = 0.377). Page 886 (d) The model in part (a) is the best. When both x1 and x2 are in the model, only the coefficient of x1 is significantly different from 0. In addition, the value of R2 is only slightly greater (0.819 vs. 0.811) for the model containing both x1 and x2 than for the model containing x1 alone. 3. 5. (a) Plot (i) came from engineer B, and plot (ii) came from engineer A. We know this because the variables x1 and x2 are both significantly different from 0 for engineer A but not for engineer B. Therefore engineer B is the one who designed the experiment to have the dependent variables nearly collinear. (b) Engineer A's experiment produced the more reliable results. In engineer B's experiment, the two dependent variables are nearly collinear. (a) For R1 < 4, the least-squares line is R2 = 1.23 + 0.264R1. For R1 ≥ 4, the least-squares line is R2 = −0.190 + 0.710R1. (b) The relationship is clearly nonlinear when R1 < 4. Predictor Constant R1 Predictor Constant R1 Predictor Coef 1.2840 0.21661 StDev 0.26454 0.23558 T 4.8536 0.91947 P 0.000 0.368 0.0090189 0.044984 0.20049 0.843 Coef −1.8396 4.4987 StDev 0.56292 0.75218 T −3.2680 5.9809 P 0.004 0.000 −1.7709 0.22904 0.30789 0.039454 −5.7518 5.8053 0.000 0.000 Coef StDev T P Constant R1 −2.6714 6.0208 2.0117 3.6106 −1.3279 1.6675 0.200 0.112 −2.7520 0.49423 −0.02558 2.2957 0.61599 0.05930 −1.1988 0.80234 −0.43143 0.245 0.432 0.671 (c) Page 887 (d) The correlation coefficient between (e) (f) and and is 0.997. are nearly collinear. The cubic model is best. The quadratic is inappropriate because the residual plot exhibits a pattern. The residual plots for both the cubic and quartic models look good; however, there is no reason to include in the model since it merely confounds the effect of . Section 8.3 1. (a) False (b) True (c) False 3. 5. 7. 9. (d) True v. iv. The four-variable model with the highest value of R2 has a lower R2 than the three-variable model with the highest value of R2. This is impossible. (a) 0.2803 (b) Three degrees of freedom in the numerator and 157 in the denominator. (c) P > 0.10. The reduced model is plausible. (d) This is not correct. It is possible for a group of variables to be fairly strongly related to an independent variable, even though none of the variables individually is strongly related. (e) No mistake. If y is the dependent variable, then the total sum of squares is This quantity does not involve the independent variables. 11. No, F3,17 = 10.96, P < 0.001. 13. (a) Predictor Constant x (b) Coef 37.989 1.0774 StDev 53.502 0.041608 T 0.71004 25.894 Predictor Coef StDev T Constant −253.45 132.93 −1.9067 x 1.592 0.22215 7.1665 2 −0.00020052 0.000085328 −2.3499 x P 0.487 0.000 P 0.074 0.000 0.031 Page 888 . (c) (d) (e) The quadratic model seems more appropriate. The P-value for the quadratic term is fairly small (0.031), and the residual plot for the quadratic model exhibits less of a pattern. (There are a couple of points somewhat detached from the rest of the plot, however.) (f) 1683.5 (g) (1634.7, 1732.2) 15. (a) Predictor Constant x1 x2 (b) Predictor Coef 25.613 0.18387 StDev 10.424 0.12353 T 2.4572 1.4885 P 0.044 0.180 −0.015878 0.0040542 −3.9164 0.006 Coef StDev T P Constant x 1 14.444 0.17334 16.754 0.20637 0.86215 0.83993 0.414 0.425 Page 889 (c) Predictor Constant x2 Coef 40.370 −0.015747 StDev 3.4545 0.0043503 T 11.686 −3.6197 P 0.000 0.007 (d) The model containing x2 as the only independent variable is best. There is no evidence that the coefficient of x1 differs from 0. 17. The model y = β0 + β1x2 + ε is a good one. One way to see this is to compare the fit of this model to the full quadratic model. The ANOVA table for the full model is Source Regression Residual error Total DF 5 9 14 SS 4.1007 3.9241 8.0248 MS 0.82013 0.43601 F 1.881 P 0.193 F 6.8285 P 0.021 The ANOVA table for the model y = β0 + β1x2 + ε is Source Regression Residual error Total DF 1 13 14 SS 2.7636 5.2612 8.0248 MS 2.7636 0.40471 From these two tables, the F statistic for testing the plausibility of the reduced model is . The null distribution is F4,9, P > 0.10. The large P-value indicates that the reduced model is plausible. Supplementary Exercises for Chapter 8 1. (a) 24.6% (b) 5.43% 3. (c) No, we need to know the oxygen content. (a) 0.207 (b) 0.8015 (c) 3.82 (d) 1.200 (e) 2 (f) 86.81 (g) 43.405 (h) 30.14 5. (i) 14 (a) Predictor Constant Speed Pause Speed2 Pause2 Speed · Pause Coef StDev T 10.84 2.2749 39.432 −0.073851 0.023379 −3.1589 −0.12743 0.013934 −9.1456 0.0011098 0.00048887 2.2702 0.0016736 0.00024304 6.8861 −0.00024272 0.00027719 −0.87563 Analysis of Variance Source DF Regression 5 Residual error 24 Total 29 SS 31.304 2.6462 33.95 MS 6.2608 0.11026 F 56.783 P 0.000 0.004 0.000 0.032 0.000 0.390 P 0.000 Page 890 (b) We drop the interaction term Speed · Pause. Predictor Constant Speed Pause Speed2 Pause2 Coef 10.967 −0.079919 −0.13253 0.0011098 0.0016736 Analysis of Variance Source DF Regression 4 Residual error 25 Total 29 StDev 0.23213 0.022223 0.01260 0.00048658 0.0002419 SS 31.22 2.7307 33.95 MS 7.8049 0.10923 T 47.246 −3.5961 −10.518 2.2809 6.9185 P 0.000 0.001 0.000 0.031 0.000 F 71.454 P 0.000 Comparing this model with the one in part (a), F1,24 = 0.77, P > 0.10. (c) There is some suggestion of heteroscedasticity, but it is hard to be sure without more data. (d) No, compared with the full model containing Speed, Pause, Speed2, and Pause2, and Speed · Pause, the F statistic is F3,24 = 15.70, and P < 0.001. Page 891 (e) (f) 7. The model containing the dependent variables Speed, Pause, Speed2, and Pause2 has both the lowest value of Cp and the largest value of adjusted R2. The residual plot shows an obvious curved pattern, so the linear model is not appropriate. Page 892 There is no obvious pattern to the residual plot, so the quadratic model appears to fit well. 9. There is no obvious pattern to the residual plot, so the cubic model appears to fit well. (a) 182.52, 166.55, 187.56 (b) 236.39, 234.18, 237.02 (c) 176.80, 163.89, 180.78 (d) (iv). The output does not provide much to choose from between the two-variable models. In the three-variable model, none of the coefficients are significantly different from 0, even though they were significant in the two-variable models. This suggests collinearity. 11. (a) Following are the values of SSE and their degrees of freedom for models of degrees 1, 2, 3, and 4. Linear Quadratic Cubic Quartic 18 17 16 15 2726.55 481.90 115.23 111.78 To compare quadratic vs. linear, . To compare cubic vs. quadratic, . To compare quartic vs. cubic, . The cubic model is selected by this procedure. Page 893 (b) The cubic model is y = 27.937 + 0.48749x + 0.85104x2 − 0.057254x3 The estimate y is maximized when dy/dx = 0. dy/dx = 0.48749 + 1.70208x − 0.171762x2. Therefore x = 10.188. 13. (a) Let y1 represent the lifetime of the sponsor's paint, y2 represent the lifetime of the competitor's paint, x1 represent January temperature, x2 represent July temperature, and x3 represent precipitation. Then one good model for y1 is .A good model for y2 is . (b) 15. (a) . Predictor Constant x Coef 0.25317 −0.041561 StDev 0.0065217 0.040281 T 38.819 −1.0318 P 0.000 0.320 (b) Predictor Constant x x2 Coef 0.21995 0.58931 −2.2679 StDev 0.0038434 0.06146 0.2155 T 57.23 9.5886 −10.524 P 0.000 0.000 0.000 (c) Predictor Constant x x2 x3 Coef 0.22514 0.41105 −0.74651 −3.6728 StDev 0.0068959 0.20576 1.6887 4.043 T 32.648 1.9977 −0.44206 −0.90843 P 0.000 0.069 0.666 0.382 (d) Predictor Constant x x2 x3 x4 Coef 0.23152 0.10911 3.4544 −26.022 40.157 StDev 0.013498 0.58342 7.7602 40.45 72.293 T 17.152 0.18702 0.44515 −0.64333 0.55548 P 0.000 0.855 0.665 0.533 0.590 (e) The quadratic model. The coefficient of x3 in the cubic model is not significantly different from 0. Neither is the coefficient of x4 in the quartic model. (f) 0.258 17. (a) Predictor Coef StDev T Constant −0.093765 0.092621 −1.0123 x1 0.63318 2.2088 0.28666 P 0.335 0.780 x2 2.5095 0.30151 8.3233 0.000 x1x2 5.318 −0.3214 0.15209 8.2231 0.17396 1.5778 0.64672 0.532 −1.8475 0.094 0.09639 0.925 Page 894 Analysis of Variance Source DF SS Regression 5 20.349 Residual error 10 0.045513 Total 15 20.394 MS 4.0698 0.0045513 F P 894.19 0.000 (b) The model containing the variables x1, x2, and is a good one. Here are the coefficients along with their standard deviations, followed by the analysis of variance table. Predictor Constant x1 x2 Coef −0.088618 2.1282 StDev 0.068181 0.30057 T −1.2997 7.0805 P 0.218 0.000 2.4079 0.13985 17.218 0.000 −0.27994 0.059211 −4.7279 0.000 Analysis of Variance Source DF SS MS F P Regression 3 20.346 6.782 1683.9 0.000 Residual error 12 0.048329 0.0040275 Total 15 20.394 The F statistic for comparing this model to the full quadratic model is , so it is reasonable to drop and x1x2 from the full quadratic model. All the remaining coefficients are significantly different from 0, so it would not be reasonable to reduce the model further. (c) The model with the best adjusted R2 (0.99716) contains the variables x2, , and . This model is also the model with the smallest value of Mallows' Cp (2.2). This is not the best model, since it contains but not x1. The model containing x1, x2, and , suggested in the answer to part (b), is better. Note that the adjusted R2 for the model in part (b) is 0.99704, which differs negligibly from that of the model with the largest adjusted R2 value. 19. (a) Predictor Constant t t2 Coef 1.1623 0.059718 −0.00027482 StDev 0.17042 0.0088901 0.000069662 T P 6.8201 0.006 6.7174 0.007 −3.9450 0.029 (b) 17.68 minutes (c) (0.0314, 0.0880) (d) The reaction rate is decreasing with time if β2 < 0. We therefore test H0:β2 ≥ 0vs. H1: β2 < 0. The test statistic is t3 = 3.945, P = 0.029/2 = 0.0145. It is reasonable to conclude that the reaction rate decreases with time. 21. y = β0 + β1x1 + β2x2 + β3x1x2 + ε. Page 895 23. (a) The 17-variable model containing the independent variables x21, x22,and x23 has adjusted R2 equal to 0.98446. The fitted model is (b) The eight-variable model containing the independent variables x1, x2, x5, x8, x10, x11, x14,and x21 has Mallows' Cp equal to 1.7. The fitted model is (c) Using a value of 0.15 for both α-to-enter and α-to-remove, the equation chosen by stepwise regression is y = −927.72 + 142.40x5 + 0.081701x7 + 21.698x10 + 0.41270x16 + 45.672x21. (d) The following 13-variable model has adjusted R2 equal to 0.95402. (There are also two 12-variable models whose adjusted R2 is only very slightly lower.) (e) The two-variable model z = −1660.9 + 0.67152x7 + 134.28x10 has Mallows' Cp equal to −4.0. (f) Using a value of 0.15 for both α-to-enter and α-to-remove, the equation chosen by stepwise regression is z = −1660.9 + 0.67152x7 + 134.28x10 (g) The following 17-variable model has adjusted R2 equal to 0.97783. (h) The following 13-variable model has Mallows' Cp equal to 8.0. (i) Using a value of 0.15 for both α-to-enter and α-to-remove, the equation chosen by stepwise regression is w = 130.92 − 28.085x2 + 113.49x5 + 0.16802x9 − 0.20216x16 + 11.417x19 + 12.068x21 − 78.371x23. Section 9.1 1. (a) Source Temperature Error Total DF 3 16 19 SS 202.44 18.076 220.52 MS 67.481 1.1297 F 59.731 P 0.000 (b) Yes. F3,16 = 59.731, P < 0.001 (P ≈ 0). 3. (a) Source Treatment Error Total DF 4 11 15 SS 19.009 22.147 41.155 MS 4.7522 2.0133 F 2.3604 P 0.117 (b) No. F4,11 = 2.3604, P > 0.10 (P = 0.117). Page 896 5. (a) Source Age Error Total DF 5 73 78 SS 3.8081 7.0274 10.835 MS 0.76161 0.096266 F 7.9115 P 0.000 MS 0.064062 0.034085 F 1.8795 P 0.142 (b) Yes, F5,73 = 7.9115, P < 0.01 7. (a) Source Group Error Total DF 3 62 65 SS 0.19218 2.1133 2.3055 (b) No. F3,62 = 1.8795, P > 0.10 (P = 0.142). 9. (a) Source Plant Error Total DF 2 25 27 SS 481.45 690.26 1171.7 MS 240.73 27.611 F 8.7186 P 0.001 (b) Yes, F2,25 = 8.7186, 0.001 < P < 0.01 (P = 0.001341). 11. No, F3,16 = 15.8255, P < 0.001 (P ≈ 4.8 × 10−5). 13. (a) Source Temperature Error Total DF 3 16 19 SS 58.650 36.837 95.487 MS 19.550 2.3023 F 8.4914 P 0.001 (b) Yes, F3,16 = 8.4914, 0.001 < P < 0.01 (P = 0.0013). 15. (a) s = 1.517 (b) 18 (c) 38 17. (a) Source Machine Error Total DF 4 30 34 SS 6862 6529.1 13391 MS 1715.5 217.64 F 7.8825 P 0.000 (b) Yes, F4,30 = 7.8825, P ≈ 0 19. (a) Source Soil Error Total DF 2 23 25 SS 2.1615 4.4309 6.5924 MS 1.0808 0.19265 F 5.6099 P 0.0104 (b) Yes, F2,23 = 5.6099, 0.01 < P < 0.05 (P = 0.0104). Page 897 Section 9.2 1. (a) Yes, F5,6 = 46.64, P ≈ 0. (b) q6,6,.05 = 5.63. The value of MSE is 0.00508. The 5% critical value is therefore . Any pair that differs by more than 0.284 can be concluded to be different. The following pairs meet this criterion: A and B, A and C, A and D, A and E, B and C, B and D, B and E, B and F, D and F. (c) t6,.025/15 = 4.698 (the value obtained by interpolating is 4.958). The value of MSE is 0.00508. The 5% critical value is therefore . Any pair that differs by more than 0.335 may be concluded to be different. The following pairs meet this criterion: A and B, A and C, A and D, A and E, B and C, B and D, B and E, B and F, D and F. (d) The Tukey–Kramer method is more powerful, since its critical value is smaller (0.284 vs. 0.335). 3. (e) Either the Bonferroni or the Tukey–Kramer method can be used. (a) MSE = 2.9659, Ji = 12 for all i. There are seven comparisons to be made. Now t88,.025/7 = 2.754, so the 5% critical value is . All the sample means of the noncontrol formulations differ from the sample mean of the control formulation by more than this amount. Therefore we conclude at the 5% level that all the noncontrol formulations differ from the control formulation. (b) There are seven comparisons to be made. We should use the Studentized range value q7,88,.05 This value is not in the table, so we will use q7,60,.05 = 4.31, which is only slightly larger. The 5% critical value is . All the noncontrol formulations differ from the sample mean of the control formulation by more than this amount. Therefore we conclude at the 5% level that all the noncontrol formulations differ from the control formulation. 5. (c) The Bonferroni method is more powerful, because it is based on the actual number of comparisons being made, which is 7. The Tukey–Kramer method is based on the largest number of comparisons that could be made, which is (7)(8)/2 = 28. (a) t16,.025/6 = 3.0083 (the value obtained by interpolating is 3.080). The value of MSE is 2.3023. The 5% critical value is therefore . We may conclude that the mean for 750°Cdiffers from the means for 850°C and 900°C, and that the mean for 800° differs from the mean for 900°C. (b) q4,16,.05 = 4.05. The value of MSE is 2.3023. The 5% critical value is therefore . We may conclude that the mean for 750°C differs from the means for 850°C and 900°C, and that the mean for 800° differs from the mean for 900°C. 7. (c) The Tukey–Kramer method is more powerful, because its critical value is smaller. (a) t16,.025/3 = 2.6730 (the value obtained by interpolating is 2.696). The value of MSE is 2.3023. The 5% critical value is therefore . We may conclude that the mean for 900°C differs from the means for 750°C and 800°C. (b) q4,16,.05 = 4.05. The value of MSE is 2.3023. The 5% critical value is therefore . We may conclude that the mean for 900°C differs from the means for 750°C and 800°C. 9. (c) The Bonferroni method is more powerful, because its critical value is smaller. (a) t73,.025 = 1.993, MSE = 0.096266, the sample sizes are 12 and 15. The sample means are , . The 95% confidence interval is , or (0.1133, 0.5923). (b) The sample sizes are J1 = 12, J2 = 12, J3 = 13, J4 = 12, J5 = 15, J6 = 15. MSE = 0.096266. We should use the Studentized range value q6,73,.05. This value is not in the table, so we will use q6,60,.05 = 4.16, which is only slightly larger. The values of q6,60,.05 and the values of the differences are presented in the following two tables. Page 898 1 1 − 2 0.37260 2 3 4 5 6 0.37260 0.365360.37260 0.353480.35348 − 0.365360.37260 0.353480.35348 3 0.365360.36536 − 4 0.372600.37260 0.36536 0.36536 0.345840.34584 − 5 0.353480.35348 0.345840.35348 0.353480.35348 − 0.33326 6 0.353480.35348 0.345840.35348 0.33326 1 1 0 2 0.0075 2 3 5 6 0.0075 0.49904 0.15083 0.5475 0.35283 0 3 0.499040.49154 0.49154 0.14333 0 4 0.150830.14333 0.34821 5 0.5475 4 − 0.54 0.34533 0.34821 0.0484620.14621 0 0.54 0.0484620.39667 0.39667 0.202 0 6 0.352830.34533 0.14621 0.202 0.19467 0.19467 0 The differences that are significant at the 5% level are: mean 1 differs from means 3 and 5; mean 2 differs from means 3 and 5; and mean 4 differs from mean 5. 11. (a) t8,.025 = 2.306, MSE = 1.3718. The sample means are and . The sample sizes are J1 = 5 and J3 = 3. The 95% confidence interval is therefore , or (1.330, 5.274). (b) The sample means are , , . The sample sizes are J1 = 5, J2 = J3 = 3. The upper 5% point of the Studentized range is q3,8,.05 = 4.04. The 5% critical value for and for is , and the 5% critical value for is . Therefore means 1 and 3 differ at the 5% level. 13. (a) MSTr = 19.554 so F = 19.554/3.85 = 5.08. There are 3 and 16 degrees of freedom, so 0.01 < P < 0.05. The null hypothesis of no difference is rejected at the 5% level. (b) q4,16,.05 = 4.05, so catalysts whose means differ by more than are significantly different at the 5% level. Catalysts 1 and 2 both differ significantly from catalyst 4. 15. Any value of MSE satisfying 5.099 < MSE < 6.035. Section 9.3 1. (a) 3 (b) 2 (c) 6 (d) 24 (e) (f) Source Oil Ring Interaction Error Total DF 3 2 6 24 35 SS 1.0926 0.9340 0.2485 1.7034 3.9785 MS 0.36420 0.46700 0.041417 0.070975 F 5.1314 6.5798 0.58354 P 0.007 0.005 0.740 Yes. F6,24 = 0.58354, P > 0.10 (P = 0.740). (g) No, some of the main effects of oil type are nonzero. F3,24 = 5.1314, 0.001 < P < 0.01 (P = 0.007). (h) No, some of the main effects of piston ring type are nonzero. F2,24 = 6.5798, 0.001 < P < 0.01 (P = 0.005). Page 899 3. (a) Source Mold Temp. Alloy Interaction Error Total DF 4 2 8 45 59 SS 69,738 8958 7275 115,845 201,816 MS 17,434.5 4479.0 909.38 2574.3 F 6.7724 1.7399 0.35325 P 0.000 0.187 0.939 (b) Yes. F8,45 = 0.35325, P > 0.10 (P = 0.939). (c) No, some of the main effects of mold temperature are nonzero. F4,45 = 6.7724, P < 0.001 (P ≈ 0). (d) Yes. F3,45 = 1.7399, P > 0.10, (P = 0.187). 5. (a) NaCl Na2HPO4 25°C 37°C Main Effects of Solution −9.1148 9.1148 Main Effects of Temperature 1.8101 −1.8101 Interactions Temperature 25°C −0.49983 0.49983 Solution NaCl Na2HPO4 (b) Source Solution Temperature Interaction Error Total DF 1 1 1 20 23 SS 1993.9 78.634 5.9960 7671.4 9750.0 MS 1993.9 78.634 5.9960 383.57 37°C 0.49983 −0.49983 F 5.1983 0.20500 0.015632 P 0.034 0.656 0.902 (c) Yes, F1,20 = 0.015632, P > 0.10 (P = 0.902). (d) Yes, since the additive model is plausible. The mean yield stress differs between Na2HPO4 and NaCl: F1,20 = 5.1983, 0.01 < P < 0.05 (P = 0.034). 7. (e) There is no evidence that the temperature affects yield stress: F1,20 = 0.20500, P > 0.10 (P = 0.656). (a) Source Adhesive Pressure Interaction Error Total DF 1 2 2 12 17 SS 17.014 35.663 4.3544 20.173 77.205 MS 17.014 17.832 2.1772 1.6811 4.5415 F 10.121 10.607 1.2951 P 0.008 0.002 0.310 (b) Yes. F2,12 = 1.2951, P > 0.10 (P = 0.310). (c) Yes, since the additive model is plausible. The mean strength differs between the two adhesives: F1,12 = 10.121, P < 0.01 (P = 0.008). 9. (d) Yes, since the additive model is plausible. The mean strength differs among the pressure: F2,12 = 10.607, P < 0.01 (P = 0.002). Page 900 (a) 80 120 150 Main Effects of Speed −13.074 −5.7593 19.463 Main Effects of Time −8.4259 −0.2037 8.6296 5 10 15 Interactions Speed 80 120 150 (b) 5 5.6481 2.3704 −8.0185 Source Speed Time Interaction Error Total DF 2 2 4 45 53 Time 10 0.75926 −0.018519 −0.74074 SS 10796 2619.1 1357.5 3816.3 18589 MS 5397.9 1309.6 339.38 84.807 15 −6.4074 −2.3519 8.7593 F 63.649 15.442 4.0018 P 0.000 0.000 0.007 (c) No, F4,45 = 4.0018, P < 0.01 (P = 0.007) (d) No, because the additive model is rejected. (e) No, because the additive model is rejected. 11. (a) CPTi-ZrO2 Main Effects of Material 0.044367 TiAlloy-ZrO2 −0.044367 Main Effects of Neck Length Short −0.018533 Medium −0.024833 Long 0.043367 Interactions CPTi-ZrO2 Short 0.0063333 Neck Length Medium −0.023767 Long 0.017433 TiAlloy-ZrO2 0.0063333 0.023767 −0.017433 (b) Source DF SS MS F P Taper Material 1 0.059052 0.059052 23.630 0.000 Neck Length 2 0.028408 0.014204 5.6840 0.010 Interaction 2 0.0090089 0.0045444 1.8185 0.184 Error 24 0.059976 0.002499 Total 29 0.15652 (c) Yes, the interactions may plausibly be equal to 0. The value of the test statistic is 1.8185, its null distribution is F224,and P > 0.10 (P = 0.184). (d) Yes, since the additive model is plausible. The mean coefficient of friction differs between CPTi-ZrO2 and TiAlloy-ZrO2: F1,24 = 23.630, P < 0.001. (e) Yes, since the additive model is plausible. The mean coefficient of friction is not the same for all neck lengths: F2,24 = 5.6840, P ≈ 0.01. To determine which pairs of effects differ, we use q3,24,.05 = 3.53. We compute . We conclude that the effect of long neck length differs from both short and medium lengths, but we cannot conclude that the effects of short and medium lengths differ from each other. Page 901 13. (a) 15 40 100 1:1 1:5 1:10 Main Effects of Concentration 0.16667 −0.067778 −0.098889 Main Effects of Delivery Ratio 0.73333 −0.30000 −0.43333 Interactions Concentration 15 40 100 (b) Source Concentration 1:1 0.66778 −0.20111 −0.46667 Delivery Ratio 1:5 −0.30222 −0.064444 0.36667 DF SS 2 0.37936 MS 0.18968 1:10 −0.36556 0.26556 0.10000 F P 3.8736 0.040 Delivery Ratio Interaction Error Total 2 4 18 26 7.34 3.4447 0.8814 12.045 3.67 74.949 0.000 0.86118 17.587 0.000 0.048967 (c) No. The value of the test statistic is 17.587, its null distribution is F4,18, and P ≈ 0. (d) The slopes of the line segments are quite different from one another, indicating a high degree of interaction. 15. (a) Nail Adhesive Main Effects of Attachment −1.3832 1.3832 Main Effects of Length −7.1165 −2.5665 9.683 Quarter Half Full Interactions Attachment Nail Adhesive (b) Source Length Half 0.33167 −0.33167 Quarter 0.48317 −0.48317 DF SS MS Full −0.51633 0.51633 F P Attachment Length Interaction Error Total 1 2 2 54 59 114.79 3019.8 10.023 107.29 3251.9 114.79 1509.9 5.0115 1.9869 57.773 759.94 2.5223 0.000 0.000 0.090 (c) The additive model is barely plausible: F2,54 = 2.5223, 0.05 < P < 0.10 (P = 0.090). Page 902 (d) Yes, the attachment method does affect the critical buckling load: F1,54 = 57.773, P ≈ 0. (e) Yes, the side member length does affect the critical buckling load: F2,54 = 759.94, P ≈ 0. To determine which effects differ at the 5% level, we should use q3,54,.05. This value is not found in Table A.9, so we approximate it with q3,40,.05 = 3.44. We compute . We conclude that the effects of quarter, half, and full all differ from each other. 17. (a) Source Wafer Operator Interaction Error Total DF 2 2 4 9 17 SS 114,661.4 136.78 6.5556 45.500 114,850.3 MS 57,330.7 68.389 1.6389 5.0556 F 11,340.1 13.53 0.32 P 0.000 0.002 0.855 (b) There are differences among the operators. F2,9 = 13.53, 0.01 < P < 0.001 (P = 0.002). 19. (a) Source PVAL DCM Interaction Error Total DF 2 2 4 18 26 SS 125.41 1647.9 159.96 136.94 2070.2 MS 62.704 823.94 39.990 7.6075 F 8.2424 108.31 5.2567 P 0.003 0.000 0.006 (b) Since the interaction terms are not equal to 0 (F4,18 = 5.2567, P = 0.006), we cannot interpret the main effects. Therefore we compute the cell means. These are PVAL 0.5 50 97.8 DCM (mL) 40 92.7 30 74.2 1.0 2.0 93.5 94.2 80.8 88.6 75.4 78.8 We conclude that a DCM level of 50 mL produces greater encapsulation efficiency than either of the other levels. If DCM = 50, the PVAL concentration does not have much effect. Note that for DCM = 50, encapsulation efficiency is maximized at the lowest PVAL concentration, but for DCM = 30 it is maximized at the highest PVAL concentration. This is the source of the significant interaction. Section 9.4 1. (a) NaOH concentration is the blocking factor, age is the treatment factor. (b) Source Treatment Blocks Error Total DF 3 4 12 19 SS 386.33 13.953 7.3190 407.60 MS 128.78 3.4882 0.6099 F 211.14 5.7192 P 0.000 0.008 Page 903 (c) Yes, F3,12 = 211.14, P ≈ 0 (d) q4, = 4.20, MSAB = 0.6099, and J = 5. The 5% critical value is therefore . The sample means are , , , and . We therefore conclude that age 0 differs from ages 4, 8, and 12, and that age 8 differs from age 12. 3. 12.05 (a) Source Lighting Block Interaction Error Total DF 3 2 6 24 35 SS 9943 11,432 6135 23,866 51,376 MS 3314.3 5716.0 1022.5 994.42 F 3.3329 5.7481 1.0282 P 0.036 0.009 0.431 F 2.5677 25.367 P 0.018 0.000 (b) Yes. The P-value for interactions is large (0.431). 5. (c) Yes. The P-value for lighting is small (0.036). (a) Source Variety Block DF 9 5 SS 339,032 1,860,838 MS 37,670 372,168 Error Total 45 59 660,198 2,860,069 14,671 (b) Yes, F9,45 = 2.5677, P = 0.018. 7. (a) Source Waterworks Block Error Total DF 3 14 42 59 SS 1253.5 1006.1 3585.0 5844.6 MS 417.84 71.864 85.356 F 4.8953 0.84193 P 0.005 0.622 (b) Yes, F3,42 = 4.8953, P = 0.005. (c) To determine which effects differ at the 5% level, we should use q4,42,.05. This value is not found in Table A.9, so we approximate it with q4,40,.05 = 3.79. The 5% critical value is . The sample means are , , , . We can conclude that A differs from both B and C. 9. (d) The P-value for the blocking factor is large (0.622), suggesting that the blocking factor (time) has only a small effect on the outcome. It might therefore be reasonable to ignore the blocking factor and perform a one-way ANOVA. (a) One motor of each type should be tested on each day. The order in which the motors are tested on any given day should be chosen at random. This is a randomized block design, in which the days are the blocks. It is not a completely randomized design, since randomization occurs only within blocks. (b) The test statistic is . Page 904 Section 9.5 1. 1 ad bd ab cd ac bc abcd A − + − + − + − + B − − + + − − + + C − − − − + + + + D − + + − + − − + The alias pairs are {A, BCD}, {B, ACD}, {C, ABD}, {D, ABC}, {AB, CD}, {AC, BD}, and {AD, BC} 3. (a) Term A B C AB AC BC ABC Error Total Effect 6.75 9.50 1.00 2.50 0.50 0.75 −2.75 DF 1 1 1 1 1 1 1 8 15 SS 182.25 361.00 4.00 25.00 1.00 2.25 30.25 122.00 727.75 MS 182.25 361.00 4.00 25.00 1.00 2.25 30.25 15.25 F 11.9508 23.6721 0.2623 1.6393 0.0656 0.1475 1.9836 P 0.009 0.001 0.622 0.236 0.804 0.711 0.197 (b) Factors A and B (temperature and concentration) seem to have an effect on yield. There is no evidence that pH has an effect. None of the interactions appear to be significant. Their P-values are all greater than 0.19. 5. (c) Since the effect of temperature is positive and statistically significant, we can conclude that the mean yield is higher when temperature is high. (a) Term A B C AB AC BC ABC Effect 3.3750 23.625 1.1250 −2.8750 −1.3750 −1.6250 1.8750 (b) No, since the design is unreplicated, there is no error sum of squares. (c) No, none of the interaction terms are nearly as large as the main effect of factor B. (d) If the additive model is known to hold, then the following ANOVA table shows that the main effect of B is not equal to 0, while the main effects of A and C may be equal to 0. Term A B C Error Effect DF 3.3750 1 23.625 1 1.1250 1 4 Sum of Squares 22.781 1116.3 2.5312 32.625 Mean Square 22.781 1116.3 2.5312 8.1562 F P 2.7931 0.170 136.86 0.000 0.31034 0.607 Total 7 1174.2 Page 905 7. (a) Term A B C AB AC BC ABC Effect −119.25 259.25 −82.75 101.75 −6.25 −52.75 −2.25 (b) No, because there is only one replication. (c) The AB interaction is among the larger effects. 9. (d) No, because the AB interaction is large. (a) Term A B C D AB AC AD BC BD CD ABC ABD ACD BCD ABCD Effect 1.2 3.25 −16.05 −2.55 2.0 2.9 −1.2 1.05 −1.45 −1.6 −0.8 −1.9 −0.15 0.8 0.65 (b) Factor C is the only one that really stands out. 11. (a) Term Effect DF Sum of Squares Mean Square F P A 14.245 1 B 8.0275 1 C −6.385 1 AB −1.68 1 AC −1.1175 1 BC −0.535 1 ABC −1.2175 1 Error 8 Total 15 811.68 257.76 163.07 11.29 4.9952 1.1449 5.9292 9.3944 1265.3 811.68 257.76 163.07 11.29 4.9952 1.1449 5.9292 1.1743 691.2 219.5 138.87 9.6139 4.2538 0.97496 5.0492 0.000 0.000 0.000 0.015 0.073 0.352 0.055 (b) All main effects are significant, as is the AB interaction. Only the BC interaction has a P-value that is reasonably large. All three factors appear to be important, and they seem to interact considerably with each other. 13. ii. Page 906 Supplementary Exercises for Chapter 9 1. Source Gypsum Error Total DF 3 8 11 SS 0.013092 0.12073 0.13383 MS 0.0043639 0.015092 F 0.28916 P 0.832 The value of the test statistic is F3,8 = 0.28916; P > 0.10 (P = 0.832). There is no evidence that the pH differs with the amount of gypsum added. 3. Source Day Error Total DF 2 36 38 SS 1.0908 0.87846 1.9692 MS 0.54538 0.024402 F 22.35 P 0.000 We conclude that the mean sugar content differs among the three days (F2,36 = 22.35, 5. P ≈ 0). (a) No. The variances are not constant across groups. In particular, there is an outlier in group 1. (b) No, for the same reasons as in part (a). (c) Source Group Error Total DF 4 35 39 SS 5.2029 5.1080 10.311 MS 1.3007 0.14594 F 8.9126 P 0.000 We conclude that the mean dissolve time differs among the groups (F4,35 = 8.9126, P ≈ 7. 9. 0). The recommendation is not a good one. The engineer is trying to interpret the main effects without looking at the interactions. The small P-value for the interactions indicates that they must be taken into account. Looking at the cell means, it is clear that if design 2 is used, then the less expensive material performs just as well as the more expensive material. The best recommendation, therefore, is to use design 2 with the less expensive material. (a) Source Base Instrument Interaction Error Total DF 3 2 6 708 719 SS 13,495 90,990 12,050 422,912 539,447 MS 4498.3 45,495 2008.3 597.33 F 7.5307 76.164 3.3622 P 0.000 0.000 0.003 (b) No, it is not appropriate because there are interactions between the row and column effects (F6,708 = 3.3622, P = 0.003). 11. (a) Yes. F4,15 = 8.7139, P = 0.001. (b) q5,20 = 4.23, MSE = 29.026, J = 4. The 5% critical value is therefore . The sample means for the five channels are , , , , . We can therefore conclude that channels 3 and 4 differ from channels 1, 2, and 5. 13. No. F4,289 = 1.5974, P > 0.10 (P = 0.175). 15. (a) s = 5.388 (b) 10 (c) 22 Page 907 17. (a) Term A B C D AB AC AD BC BD CD ABC ABD ACD BCD ABCD Effect 3.9875 2.0375 1.7125 3.7125 −0.1125 0.0125 −0.9375 0.7125 −0.0875 CD ABC ABD ACD BCD ABCD 0.6375 −0.2375 0.5125 0.4875 −0.3125 −0.7125 (b) The main effects are noticeably larger than the interactions, and the main effects for A and D are noticeably larger than those for B and C. (c) Term A B C D AB AC AD BC BD CD Interaction Total Sum of Mean Effect DF Squares Square F P 3.9875 1 63.601 63.601 68.415 0.000 2.0375 1 16.606 16.606 17.863 0.008 1.7125 1 11.731 11.731 12.619 0.016 3.7125 1 55.131 55.131 59.304 0.001 −0.1125 1 0.050625 0.050625 0.054457 0.825 0.0125 1 0.000625 0.000625 0.00067231 0.980 −0.9375 1 3.5156 3.5156 3.7818 0.109 0.7125 1 2.0306 2.0306 2.1843 0.199 −0.0875 1 0.030625 0.030625 0.032943 0.863 0.6375 1 1.6256 1.6256 1.7487 0.243 5 4.6481 0.92963 15 158.97 We can conclude that each of the factors A, B, C, and D has an effect on the outcome. (d) The F statistics are computed by dividing the mean square for each effect (equal to its sum of squares) by the error mean square 1.04. The degrees of freedom for each F statistic are 1 and 4. The results are summarized in the following table. Sum of Mean Term Effect DF Squares Square F P A 3.9875 1 63.601 63.601 61.154 0.001 B 2.0375 1 16.606 16.606 15.967 0.016 C 1.7125 1 11.731 11.731 11.279 0.028 D 3.7125 1 55.131 55.131 53.01 0.002 AB −0.1125 1 0.050625 0.050625

[illegible]