



## Statistics for Business and Economics

By Stuart Strother and Orlando Griego

Included in this preview:

- Copyright Page
- Table of Contents
- Excerpt of Chapter 1

For additional information on adopting this book for your class, please contact us at 800.200.3908 x501 or via e-mail at [info@cognella.com](mailto:info@cognella.com)

# Sneak Preview

# **STATISTICS FOR BUSINESS AND ECONOMICS**

By Stuart C. Strother & Orlando Griego



Copyright © 2011 University Readers Inc. All rights reserved. No part of this publication may be reprinted, reproduced, transmitted, or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information retrieval system without the written permission of University Readers, Inc.

First published in the United States of America in 2011 by Cognella, a division of University Readers, Inc.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

15 14 13 12 11                      1 2 3 4 5

Printed in the United States of America

ISBN: 978-1-60927-731-4

# Contents

---

## Part One: Descriptive Statistics

---

Chapter 1: Introduction	3
Chapter 2: Collecting Data	13
Chapter 3: Descriptive Statistics	31
Chapter 4: Descriptive Statistics: Graphs	51

---

## Part Two: Probability

---

Chapter 5: Probability	79
Chapter 6: Discrete Probability	97
Chapter 7: Continuous Probability: The Bell Curve	109
Chapter 8: Estimation and Confidence Intervals	123

---

## Part Three: Hypothesis Testing

---

Chapter 9: Basic Hypothesis Testing	139
Chapter 10: Small Sample Hypothesis Tests	155
Chapter 11: The ANOVA Test	169
Chapter 12: Chi-Squared Tests	183

---

## Part Four: Correlation and Regression

---

Chapter 13: Correlation and Regression	197
Chapter 14: Time Series Analysis	217
Chapter 15: Multivariate Regression	233
Chapter 16: Statistical Process Control	255
References	269
Appendices	271
Answers to Chapter Exercises	291
Index	309

# Part One

## Descriptive Statistics



An employee at the Yanjing Brewery in Beijing inspects bottles of beer. Photo: Barbara L. Strother

**T**he Yanjing Brewery employs a team of quality managers who inspect bottles of beer at their factory in Beijing, China. Each bottle should have a predetermined amount of liquid and air. Too much liquid and the bottle will burst from the pressure of the carbonation. Too much air and the beer will be flat. The job of the inspector is to pull bottles from the assembly line that have incorrect ratios of liquid and air. If the company can reduce the number of defective bottles that end up in shops and restaurants, the company will be more profitable. The inspectors report their findings to company managers using descriptive statistics.

In Part One of this textbook we focus on descriptive statistics. We will describe large sets of data using numeric statistics such as mean and range, and graphs such as pie charts and bar charts. The first chapter begins with introductory concepts that are important to the study of statistics. Because it is important to know where data come from before analyzing those data, our second chapter describes data collection techniques. The third chapter covers numeric statistics and the fourth chapter describes graphs.

# Chapter 1

## Introduction

### 1.1 Chapter introduction

---

**S**tatistics is the science of collecting data, describing data, and interpreting data for effective decision making. Virtually every field uses some form of statistical analysis, which is why most college majors include a statistics course. The ability to collect, describe, and interpret data is an essential part of a successful career in today's marketplace. If you learn the basic statistical techniques presented in this textbook, you will be a producer of statistical analysis, which will help you make decisions; you will also learn to be a good consumer of information, which will help you understand the analyses and decisions of others.

This textbook is written for students of business and economics, so the examples and applications come from the world of business and economics. Most of the statistical techniques described in this book, however, are applied techniques that are used in other fields including medicine, sociology, psychology, and others. Professionals in all these fields use statistical analysis in their decision-making process.

To succeed in mastering the applied statistical techniques presented in this book, we have learned that a typical student needs multiple exposures to the material. He or she should read the text before attending class, practice the end-of-chapter exercises, keep a running list of the formulas, and learn the key vocabulary words. Be careful though! The study of statistics has never been easy. We have made every effort to make the material as simple as possible, and we are confident that if you apply yourself, you will succeed.

### 1.2 How this book is organized

---

This book is organized in four sections. The first section describes data collection and *descriptive* statistical techniques, which are ways of organizing and describing raw data. *Raw data* is information that has been collected, but not organized in any way, such as these shoe prices at the Azusa, California Ross store:

\$29.95      \$19.95      \$44.00      \$31.95      \$29.95

With only five pairs of shoes, we can get some ideas about the Ross store's pricing policies just by reviewing the raw data (about \$30 for a pair of shoes), but this would be more difficult with 500 pairs of shoes. Instead of reporting raw data, we can use descriptive statistical techniques to better represent these data. For instance we can use numbers such as mean, median, or mode, or we can use graphs such as a column chart or a pie chart.

The second section of this book describes *probability* techniques, which are methods of determining the likelihood of an event happening in the future. This is quite useful in business, as managers make many decisions based on forecasts. If consumer demand will be high, a company needs to increase production in advance. If the economy is heading into recession, firms will cut back on production and hiring. Probability analysis is a critical part of business decision making.

The third section of the book covers *hypothesis testing*, which are techniques to determine whether two or more things are the same or different. In the recession of 2008–2009, many heavy equipment manufacturers such as Caterpillar and Kobelco are not planning on selling as much of their product as in previous years. How about companies in other industries? Should video game manufacturers such as Sony, Activision, and Nintendo follow the same strategy? That depends on whether or not consumer demand for video games is the same as for other products. Some people think video game sales are recession-proof. We can test this hypothesis (and give Mario and Sonic a hand) using the techniques described later in the book.



Photo: Stuart C. Strother

Each year, companies such as Kobelco use statistical techniques to forecast demand for their products.

The fourth and final section of the book describes *correlation* and *regression*, which are techniques for measuring the association between two or more variables. For example we will analyze the relationship between the number of bedrooms in a house and its price. Of course we expect a house

with more bedrooms to cost more, but how much more? And what about a garage or a pool? These questions can be answered using correlation and regression techniques.

### 1.3 Statistics and the scientific method

---

Most research follows what has become known as the *scientific method*, which is a process of trying to discover some truth (or perhaps disprove some fallacy). Steps of the scientific method are 1) define the question, 2) gather information and resources, 3) form hypotheses, 4) perform experiments and collect data, 5) analyze data, 6) interpret data, and 7) publish results.

The first few steps are more theoretical and require some knowledge of the field of study and the existing body of knowledge in that field. Before doing analysis of a topic, such as the effectiveness of local economic development policies, it is important to conduct a literature review on that topic to ensure that your research will in fact add something new and important to the existing state of knowledge. As scientists say, new work must “advance knowledge.”

Most of the statistical techniques in this textbook fit within the fifth step of the scientific method. That is, we will analyze data that have already been collected. To successfully analyze data, it is important to understand how data are collected, thus we start in Chapter 2 by introducing some frequently used data-collection techniques. To successfully interpret data (step 6 of the scientific method), researchers need to rely on their existing knowledge of the field, which should have been sharpened in the first few steps. Notice also the final step in the scientific method, which is publishing the results. In this textbook we focus on accurate analysis, but do not overlook the importance of publishing your results with brevity, clarity, and professionalism.

### 1.4 Types of data

---

Data are either quantitative or qualitative. *Quantitative* data are numeric data, such as the net worth of Britney Spears (\$100 million, as reported by Forbes magazine). We could also report the same data *qualitatively* by saying the pop queen is “rich.” Clearly the quantitative measure is more precise than the qualitative. Qualitative data are sometimes referred to as *attribute* data, especially by engineers and quality control managers working in manufacturing firms.

Data can also be either cross-sectional or longitudinal. *Cross-sectional* data have been measured at the same time, such as real estate data for a neighborhood that include the number of bedrooms in each home, number of bathrooms, square footage, and the market value of each home. *Longitudinal* data, also called *time-series* data, are measures of the same variable over time, such as the median home value of the one hundred fifty houses in a neighborhood, as measured each January from 1998 to 2008.

Because the market value of a home over time is a number that varies, we call this measure a *variable*. If something does not vary, we call it a *constant*, such as the number of bedrooms in a home. Most statistical analysis focuses on variables and the analysis attempts to figure out the amount of variation in that which is being measured.

Quantitative data are either discrete or continuous. *Discrete* data can only assume specific values, usually whole numbers, such as the number of bedrooms in house. Most houses have either one,





Every ten years the US government conducts a census, counting every American.

two, three, or four bedrooms. Because the data for the bedroom variable can only assume specific values, this is a discrete variable. *Continuous* data can assume any value within a reasonable range; for example, the number of square feet in a home such as a small condo with 901.5 square feet or a large single-family home with 2,695.2 square feet.

## 1.5 Sample and population

---

The term *population* refers to all cases of interest to the researcher. A quality control manager at the Gallo winery is interested in the taste and alcohol content of their wines, thus the population is each bottle of wine produced by the company. But if the manager took a drink from each bottle of wine, there would be none left to sell to their customers. For this reason, researchers usually rely on a *sample*, which is a subset of the population. Other reasons to use samples include: contacting the population may be too costly, and data from samples are usually reliable enough.

A *census* occurs when researchers attempt to collect data from the entire population, such as the US population census that occurs every ten years and the economic census that occurs every five years.

## 1.6 Levels of measurement

---

*Levels of measurement* are a way of analyzing how accurately a variable measures something, like the earlier example when we compared “rich” to “\$100 million.” Clearly the “100 million” is a more accurate measure. There are four different levels of measurement for variables.

The *nominal* level of measurement is the weakest measure. Nominal means “in name only,” and when working with nominal data, all we can do is name it, or classify it. For example Curren runs the Stinkdog Surf School and has customers as illustrated in Table 1.1. When planning the surfing lessons, Curren may consider the customers’ gender, skill level, ideal outdoor temperature, and the number of lessons completed. The first example, gender is the nominal level of measurement. That is, all we can do is *name* whether the student is male or female. Note that nominal data are always qualitative, that is non-numeric, data.

**Table 1.1 Customer data at Stinkdog Surf School**

Student	Gender	Surfing Skill	Ideal Temperature	Lessons
Bradley	male	intermediate	95 degrees	4
Jenna	female	expert	80 degrees	21
Jeremy	male	novice	90 degrees	0
Tiffany	female	intermediate	80 degrees	17

*Nominal* level data are mutually exclusive, that is they can only be in one category. For example, a person's gender is mutually exclusive. If a person is a male, the chance to be a female is "excluded." It is impossible to be both a male and a female.

Nominal-level data are also collectively exhaustive, meaning they must appear in some category. Gender is collectively exhaustive. When a surveyor asks a person, "What is your gender?" the person must choose one of the possible choices and answer either "male" or "female." The person cannot respond, "I'm neither male nor female, what are the other choices?"

The *ordinal* level of measurement is present when the data can be put in some order, such as lowest to highest. Ordinal data may be quantitative, such as first, second, or third place in a beauty contest; or it may be qualitative, such as the example in Table 1.1. We know that surfing skill can be ordered from lowest to highest: novice, intermediate, expert. What we don't know however, is the degree of difference between the ranks. In a beauty contest we may have two beautiful women and one ugly. The two beauties will place first and second, but the difference between first and second is not the same as the difference between second and third.

The *interval* level of measurement is like the ordinal level of measurement in that we can order the values from lowest to highest. We know that 90 degrees is higher than 89 degrees, and both are



Photo: Barbara L. Strother.

The number of fashion magazines sold is the ratio level of measurement.

warm. The difference is that interval level data have equal increments between each measure. That is, the difference between 89 degrees and 90 degrees is exactly 1 degree, which is the same as the difference between 90 and 91. These equal intervals characterize the interval level of measurement. Another aspect of the interval level of measurement is that the zero point is meaningless, in that it does not indicate the absence of the characteristic. A temperature of 0 degrees does not indicate the absence of temperature; in fact, it is a very cold temperature. Besides temperature, there are few examples of this level of measurement, but others are shoe and clothing sizes that may also be 0, especially for petite women's clothes.

*Ratio* is the fourth and strongest level of measurement. Ratio-level data have all the characteristics of the other three levels of measurement (can be named, ranked in order; have equal intervals between values) but the zero point is meaningful and indicates the absence of the characteristic. A person with zero dollars is sadly experiencing the absence of dollars. Similarly, if a surfing student has zero lessons, surfing lessons are absent. Table 1.2 shows a comparison of the four levels of measurement.

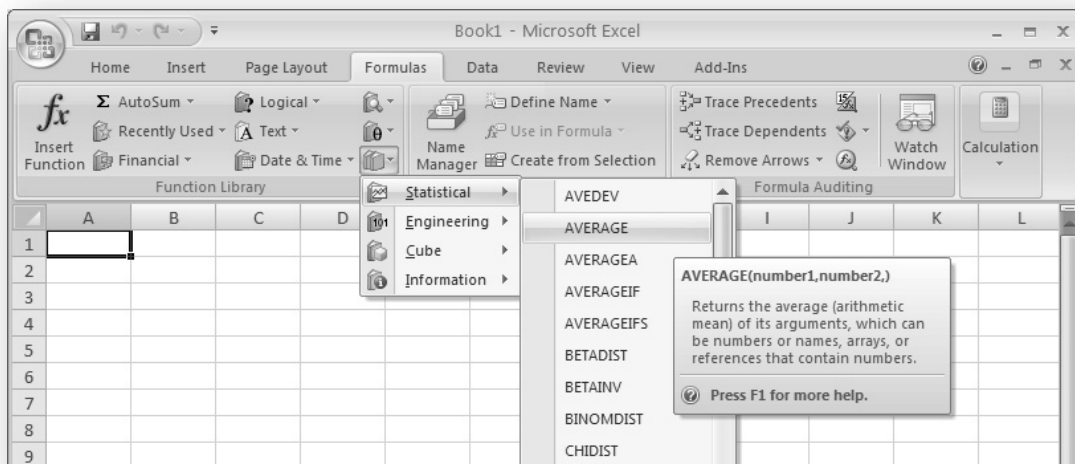
**Table 1.2 Levels of measurement**

Level of Measurement	Can be Classified	Can be Ordered	Equal Intervals	Non-Arbitrary Zero
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

## 1.7 Math and statistics software

In this textbook, you only need a basic level of mathematical skill. College algebra and a hand calculator will do. The calculations are not necessarily complex, but they can become tedious, especially when dealing with large data sets. In these cases we will give examples of how to use software in our analysis.

**Figure 1.1 Statistics functions in MS Excel**



There are a number of statistical software programs in the marketplace today. Minitab is common in business, especially among manufacturing firms. Those doing market research tend to prefer the PASW program (Predictive Analytics SoftWare, formerly known as SPSS, Statistical Program for Social Scientists). Both of these programs are quite sophisticated, but also expensive. All of the techniques shown in this book can be performed using the statistical functions in MS Excel, as shown in Figure 1.1.

## 1.7 Chapter summary

---

This chapter describes the study of statistics and introduces some important statistical jargon. To succeed in your study, read carefully, take notes, and have good class participation. Statistics can be an intimidating subject to study, but with hard work, you'll find success.

## Chapter terms

---

attribute: qualitative data

census: collecting data from the entire population

constant: something that stays the same

continuous: variables that can assume any values within a reasonable range

correlation: a measure of the association between two variables

cross-sectional: data collected at a single point in time, usually for multiple variables

descriptive statistics: techniques for organizing and describing raw data

discrete: variables that can only assume certain values, usually whole numbers

hypothesis testing: techniques to determine whether two or more things are the same or different

interval: level of measurement in which data can be classified, put in rank order, and the space between measures is equal

levels of measurement: ways of analyzing how accurately a variable measures something

longitudinal: data collected over a period of time, usually for a single variable

nominal: weak level of measurement in which the data can only be classified, but not ranked in any numeric way

ordinal: level of measurement in which data can be classified and ranked in some order

population: all cases of interest

probability: a number between 0 and 1 indicating the likelihood of an event occurring

qualitative: non-numeric data

quantitative: numeric data

ratio: strongest level of measurement in which data can be classified and put in rank order, space between measures is of equal intervals, and the zero point is non-arbitrary

raw data: information that has been collected but not organized

regression: using correlation to predict the value of one variable dependent upon another variable

sample: a portion of the population

sampling: collecting data from a portion of the population

scientific method: a process of discovering truth or disproving fallacy

statistics: the science of collecting, describing, and interpreting data for effective decision making

time series: data collected over a period of time

variable: something that varies

## Chapter exercises

---

1. Determine whether the following variables are quantitative or qualitative:
  - a. salary of professional baseball players
  - b. the number of runs scored in a baseball game
  - c. the numbers on the jerseys of baseball players
  - d. whether the mascot is a Lion, Tiger, or Yankee
2. Determine whether the following variables are cross-sectional or longitudinal:
  - a. the daily temperature in Hangzhou each day in August
  - b. the population of Hangzhou each year from 1900 to 2000
  - c. the age, salary, and marital status of the Hangzhou city council
3. Determine whether these variables are discrete or continuous:
  - a. the number of extras used in the *Benjamin Buttons* movie
  - b. the amount of money spent marketing the movie
  - c. the number of scenes of the movie
  - d. the average length of the scenes in the movie
4. Determine the level of measurement (nominal, ordinal, interval, or ratio) of the following variables:
  - a. the temperature in Shanghai on a given day
  - b. whether or not the day is considered “blue sky” or not (i.e., overcast)
  - c. the wind speed in Shanghai in kilometers per hour
  - d. the amount of rainfall in Shanghai in centimeters
  - e. whether the air quality is consider clear, polluted, or highly polluted
5. Review your local newspaper and find examples of numeric data that are
  - a. qualitative
  - b. quantitative
  - c. discrete
  - d. continuous
  - e. nominal level of measurement
  - f. ordinal
  - g. interval
  - h. ratio
6. Using MS Excel, determine how many formulas are built in for the following calculations:
  - a. average
  - b. standard deviation
  - c. variance
7. Select five of your friends and collect the following data for each of these variables:
  - a. name
  - b. gender
  - c. class in school (freshman, sophomore, etc.)
  - d. temperature at their favorite vacation spot
  - e. expected income after graduating from college

8. Determine whether the variables in the previous exercise are quantitative or qualitative.
9. Determine the level of measurement of each variable in the previous exercise.
10. Present the results of exercises 7–9 in a table that follows APA format, similar to this one:

**Table 1.3 Five friends**

<b>Name</b>	<b>Gender</b>	<b>Class</b>	<b>Ideal Vacation Temp. (F. degrees)</b>	<b>Post-College Income (x \$1,000)</b>
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
Qual. or Quant.	?	?	?	?
Level of Measurement	?	?	?	?