

Sample Size Calculations

Practical Methods for Engineers and Scientists

(Software solutions to selected example problems, Version: 17 August 2010)

Sample Size Calculations
Practical Methods for Engineers and Scientists
(Software solutions to selected example problems, Version: 17 August 2010)

Paul Mathews

Sample Size Calculations: Practical Methods for Engineers and Scientists
Paul Mathews
paul@mmbstatistical.com

Copyright ©2010 Paul Mathews

All rights reserved. No part of this publication may be reproduced or stored in any form or by any means without the prior written permission of the publisher.

Published by:
Mathews Malnar and Bailey, Inc.
217 Third Street, Fairport Harbor, OH 44077
Phone: 440-350-0911
Fax: 440-350-7210
Web: www.mmbstatistical.com

ISBN 978-0-615-32461-6

Original release: 24 February 2010

Contents

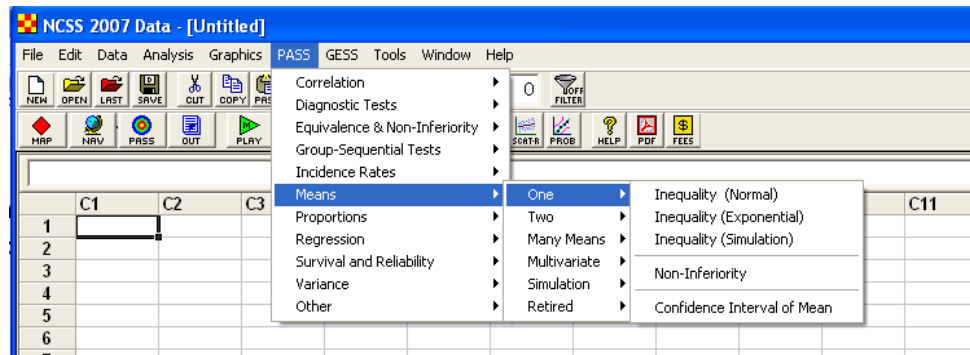
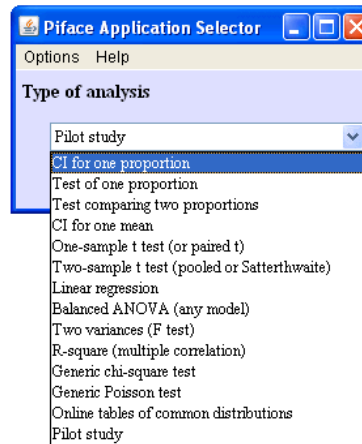
1	Fundamentals	7
1.1	Motivation for Sample Size Calculations	7
1.2	Rationale for Sample Size and Power Calculations	7
1.3	Rationale for Hypothesis Tests	12
1.4	Practical Considerations	16
1.5	Problems and Solutions	22
1.6	Software	22
2	Means	23
2.1	Assumptions	23
2.2	One Mean	23
2.3	Two Independent Means	35
2.4	Equivalence Tests	47
2.5	Contrasts	51
2.6	Multiple Comparisons Tests	52
3	Standard Deviations	57
3.1	One Standard Deviation	57
3.2	Two Standard Deviations	63
3.3	Coefficient of Variation	68
4	Proportions	71
4.1	One Proportion (Large Population)	71
4.2	One Proportion (Small Population)	85
4.3	Two Proportions	88
4.4	Equivalence Tests	98
4.5	Chi-square Tests	102

5	Poisson Counts	107
5.1	One Poisson Count	107
5.2	Two Poisson Counts	115
5.3	Tests for Many Poisson Counts	119
5.4	Correcting for Background Counts	120
6	Regression	121
6.1	Linear Regression	121
6.2	Logistic Regression	128
7	Correlation and Agreement	131
7.1	Pearson's Correlation	131
7.2	Intraclass Correlation	135
7.3	Cohen's Kappa	137
7.4	Receiver Operating Characteristic (ROC) Curves	139
7.5	Bland-Altman Plots	142
8	Designed Experiments	145
8.1	One-Way Fixed Effects ANOVA	145
8.2	Randomized Block Design	152
8.3	Balanced Full Factorial Design with Fixed Effects	153
8.4	Random and Mixed Models	159
8.5	Nested Designs	161
8.6	Two-Level Factorial Designs	162
8.7	Two-Level Factorial Designs with Centers	171
8.8	Response Surface Designs	172
9	Reliability and Survival	175
9.1	Reliability Parameter Estimation	175
9.2	Reliability Demonstration Tests	181
9.3	Two-Sample Reliability Tests	189
9.4	Interference	192
10	Statistical Quality Control	195
10.1	Statistical Process Control	195
10.2	Process Capability	197
10.3	Tolerance Intervals	199
10.4	Acceptance Sampling	200
10.5	Gage R&R Studies	211

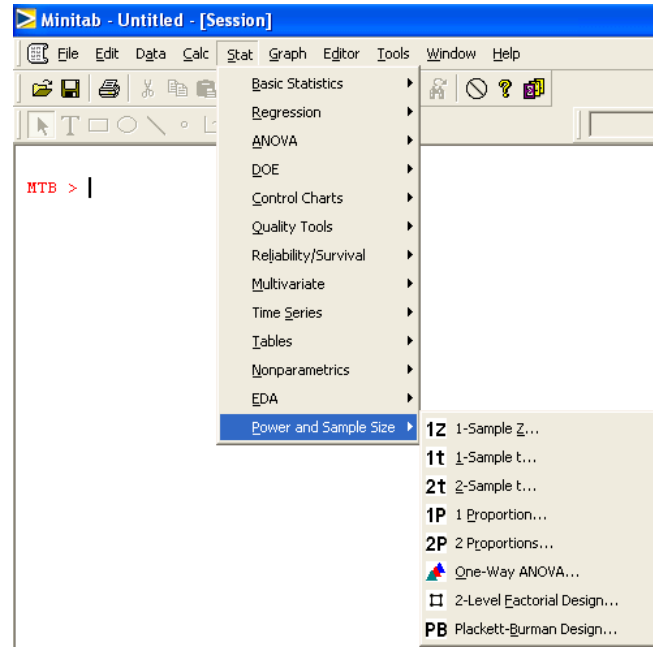
11 Resampling Methods	213
11.1 Software Requirements	213
11.2 Monte Carlo	213
11.3 Bootstrap	214

The purpose of this document is to present solutions to selected example problems from the book using PASS (2005), MINITAB (V15 and V16), Piface (V1.72), and R. (This version of the document, compiled on 17 August 2010, does not yet contain solutions using R.) All of the programs have more sample size and power calculation capabilities than what is included in the book. Some packages are particularly strong in certain areas. For example, PASS has the broadest scope, Piface offers an unmatched collection of ANOVA methods involving fixed, random, mixed, and nested designs and supports custom ANOVA models, and MINITAB has special methods for quality engineers including attribute and variables sampling plan design and reliability study design.

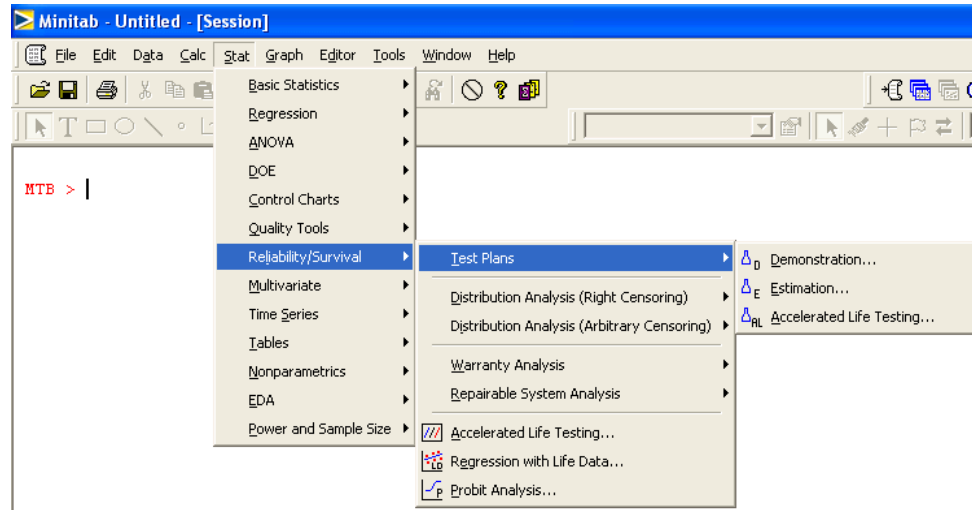
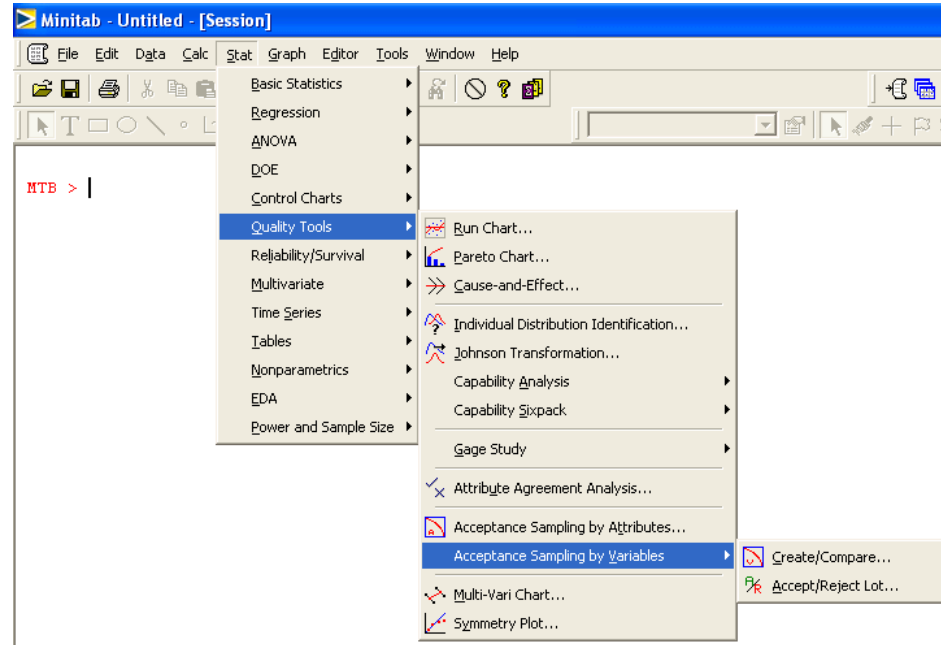
The following figures show screen captures of some of the methods available in Piface, PASS, and MINITAB.



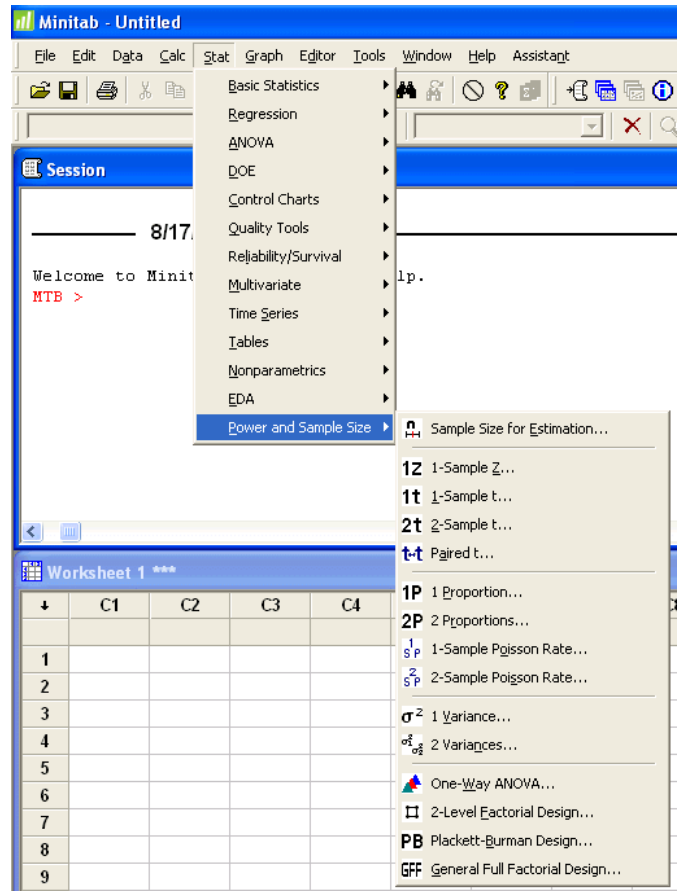
MINITAB V15:



MINITAB V16:



Most of the programs emphasize sample size and power calculations for hypothesis tests but they can be tricked into performing approximate sample size calculations for



confidence intervals by setting the hypothesis test power to $\pi = 0.50$. This trick is exact when the sampling distribution is normal because $z_{0.50} = 0$ and it is reasonably accurate when the sampling distribution is other than normal but symmetric. Be more careful when the sampling distribution is asymmetric.

The solutions in the book don't use the continuity correction when discrete distributions are approximated with continuous ones, however, the software solutions often include the continuity correction so answers to problems may differ slightly. If you have access to software that provides more accurate methods, then definitely use the software.

Some software provides several analysis methods for the same problem. For example, PASS offers six different methods for the significance test for one proportion expressed in terms of the proportion difference. The different methods usually give similar answers.

This document will be revised occasionally. The current version was compiled on 17 August 2010.

Chapter 1

Fundamentals

1.1 Motivation for Sample Size Calculations

1.2 Rationale for Sample Size and Power Calculations

Example 1.1 Express the confidence interval $P(3.1 < \mu < 3.7) = 0.95$ in words.

Solution: The confidence interval indicates that we can be 95% confident that the true but unknown value of the population mean μ falls between $\mu = 3.1$ and $\mu = 3.7$. Apparently, the mean of the sample used to construct the confidence interval is $\bar{x} = 3.4$ and the confidence interval half-width is $\delta = 0.3$.

Example 1.2 Data are to be collected for the purpose of estimating the mean of a mechanical measurement. Data from a similar process suggest that the standard deviation will be $\sigma_x = 0.003mm$. Determine the sample size required to estimate the value of the population mean with a 95% confidence interval of half-width $\delta = 0.002mm$.

Solution: With $z_{\alpha/2} = z_{0.025} = 1.96$ in Equation 1.4, the required sample size is

$$n = \left(\frac{1.96 \times 0.003}{0.002} \right)^2 = 8.64.$$

The sample size must be an integer; therefore, we round the calculated value of n up to $n = 9$.

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

```

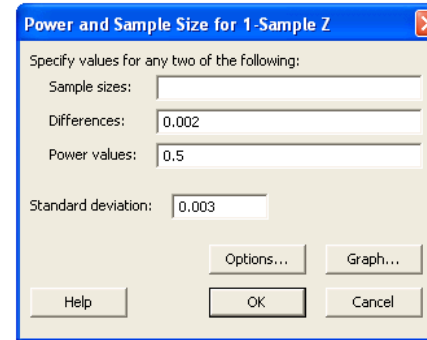
MTB > Power;
SUBC> ZOne;
SUBC> Difference 0.002;
SUBC> Power 0.5;
SUBC> Sigma 0.003;
SUBC> GPCurve.
    
```

Power and Sample Size

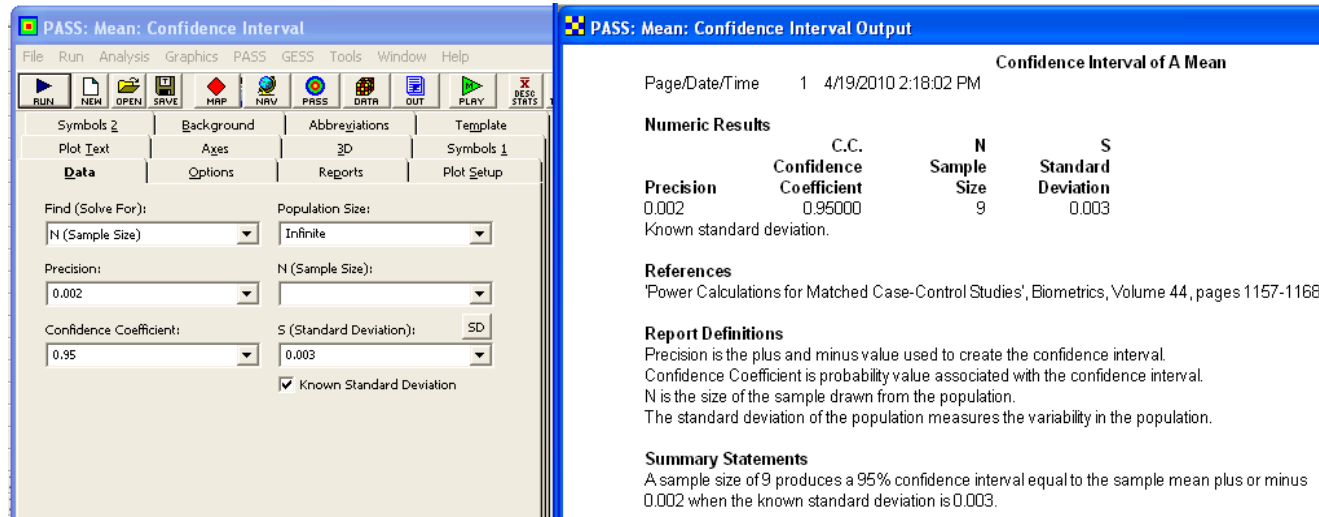
1-Sample Z Test

Testing mean = null (versus not = null)
 Calculating power for mean = null + difference
 Alpha = 0.05 Assumed standard deviation = 0.003

Difference	Sample Size	Target Power	Actual Power
0.002	9	0.5	0.516005



From **PASS> Means> One> Confidence Interval of Mean:**

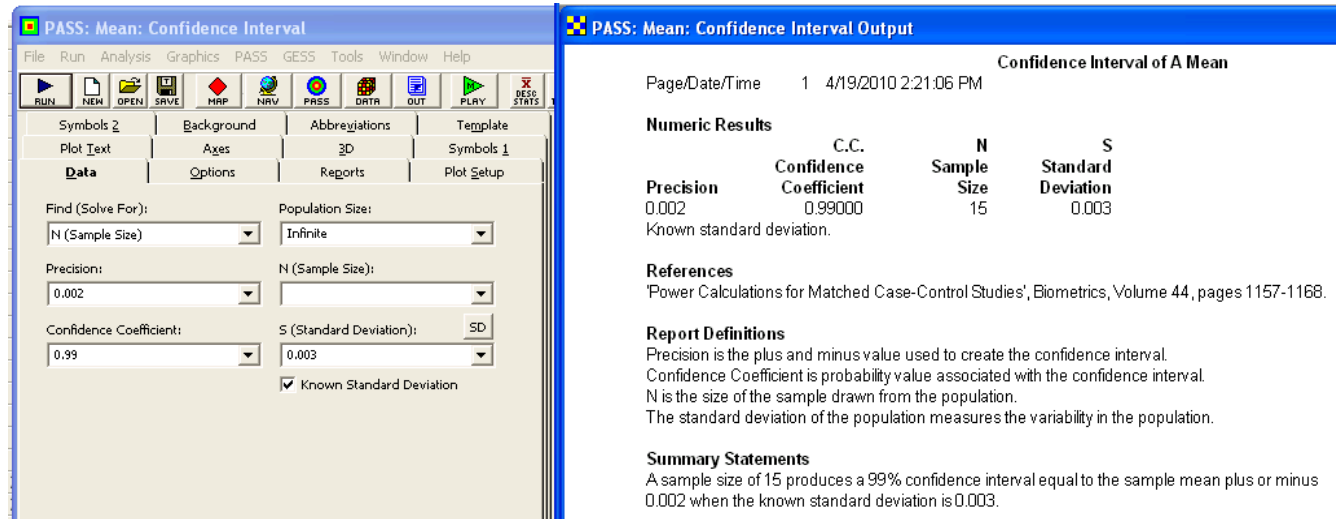


Example 1.3 What is the new sample size in Example 1.2 if the process owner prefers a 99% confidence interval?

Solution: With $z_{\alpha/2} = z_{0.005} = 2.575$ in Equation 1.4 the required sample size is

$$n = \left(\frac{2.575 \times 0.003}{0.002} \right)^2 = 15.$$

From **MINITAB> Stat> Power and Sample Size> 1-Sample Z:**



```

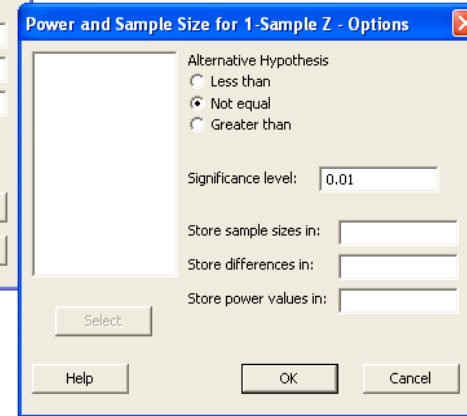
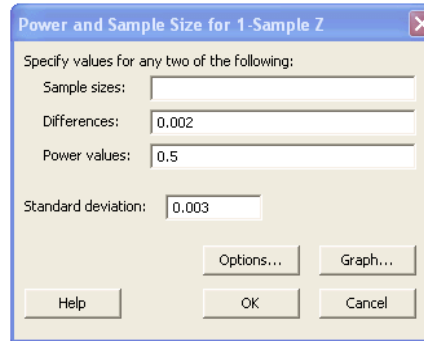
MTB > Power;
SUBC> ZOne;
SUBC> Difference 0.002;
SUBC> Power 0.5;
SUBC> Sigma 0.003;
SUBC> Alpha 0.01;
SUBC> GPCurve.
    
```

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus not = null)
 Calculating power for mean = null + difference
 Alpha = 0.01 Assumed standard deviation = 0.003

Difference	Sample Size	Target Power	Actual Power
0.002	15	0.5	0.502457



From PASS> Means> One> Confidence Interval of Mean:

Example 1.4 What is the new sample size in Example 1.2 if the process owner prefers a 95% confidence level with $\delta = 0.001mm$ half-width?

Solution: With $z_{0.025} = 1.96$ and $\delta = 0.001mm$ in Equation 1.4 the required sample size is

$$n = \left(\frac{1.96 \times 0.003}{0.001} \right)^2 = 35.$$

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

```
MTB > Power;
SUBC> ZOne;
SUBC> Difference 0.001;
SUBC> Power 0.5;
SUBC> Sigma 0.003;
SUBC> GPCurve.
```

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.003

Difference	Sample Size	Target Power	Actual Power
0.001	35	0.5	0.504854

From PASS> Means> One> Confidence Interval of Mean:

Confidence Interval of A Mean

Page/Date/Time 1 4/19/2010 2:22:52 PM

Numeric Results

	C.C. Confidence Coefficient	N Sample Size	S Standard Deviation
Precision	0.001	35	0.003

Known standard deviation.

References
"Power Calculations for Matched Case-Control Studies", Biometrics, Volume 44, pages 1157-1168.

Report Definitions
Precision is the plus and minus value used to create the confidence interval.
Confidence Coefficient is probability value associated with the confidence interval.
N is the size of the sample drawn from the population.
The standard deviation of the population measures the variability in the population.

Summary Statements
A sample size of 35 produces a 95% confidence interval equal to the sample mean plus or minus 0.001 when the known standard deviation is 0.003.

Example 1.5 Determine the sample size required to estimate the mean of a population when $\sigma_x = 30$ is known and the population mean must not exceed the sample mean by more than $\delta = 10$ with 95% confidence.

Solution: A one-sided upper 95% confidence interval is required of the form

$$P(\mu < \bar{x} + \delta) = 0.95.$$

1.2. Rationale for Sample Size and Power Calculations

With $z_{0.05} = 1.645$ in Equation 1.8 the necessary sample size is

$$n = \left(\frac{1.645 \times 30}{10} \right)^2 = 25.$$

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

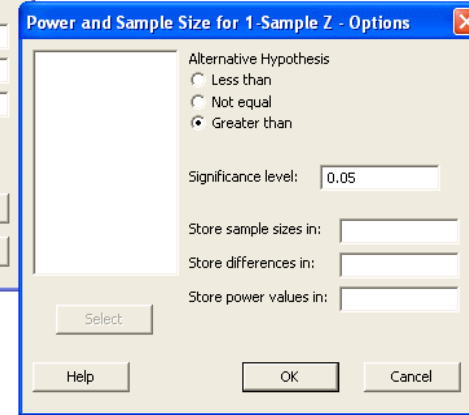
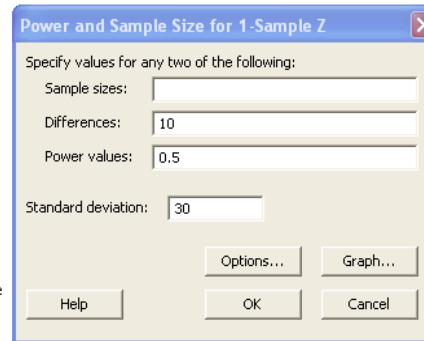
```
MTB > Power;
SUBC> ZOne;
SUBC> Difference 10;
SUBC> Power 0.5;
SUBC> Sigma 30;
SUBC> Alternative 1;
SUBC> GPCurve.
```

Power and Sample Size

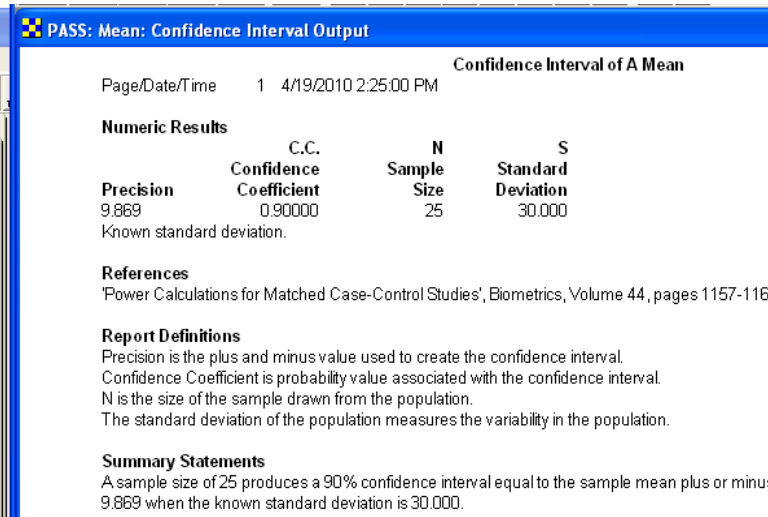
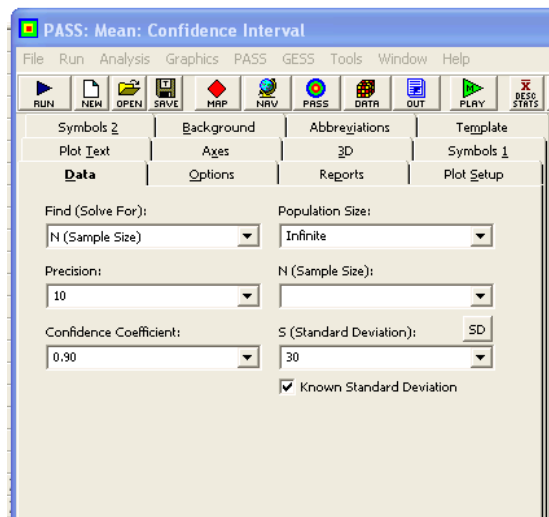
1-Sample Z Test

Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 30

Difference	Sample Size	Target Power	Actual Power
10	25	0.5	0.508701



From PASS> Means> One> Confidence Interval of Mean:



1.3 Rationale for Hypothesis Tests

Example 1.6 An experiment is planned to test the hypotheses $H_0 : \mu = 3200$ versus $H_A : \mu \neq 3200$. The process is known to be normally distributed with standard deviation $\sigma_x = 400$. What sample size is required to detect a practically significant shift in the process mean of $\delta = 300$ with power $\pi = 0.90$?

Solution: With $\beta = 1 - \pi = 0.10$ and assuming $\alpha = 0.05$ in Equation 1.12, the sample size required to detect a shift from $\mu = 3200$ to $\mu = 2900$ or $\mu = 3500$ with 90% power is

$$\begin{aligned} n &= \left(\frac{(z_{0.025} + z_{0.10}) \sigma_x}{\delta} \right)^2 \\ &= \left(\frac{(1.96 + 1.282) 400}{300} \right)^2 \\ &= 19 \end{aligned}$$

where the calculated value of n was rounded up to the nearest integer value.

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

```
MTB > Power;
SUBC> ZOne;
SUBC> Difference 300;
SUBC> Power 0.9;
SUBC> Sigma 400;
SUBC> GPCurve.
```

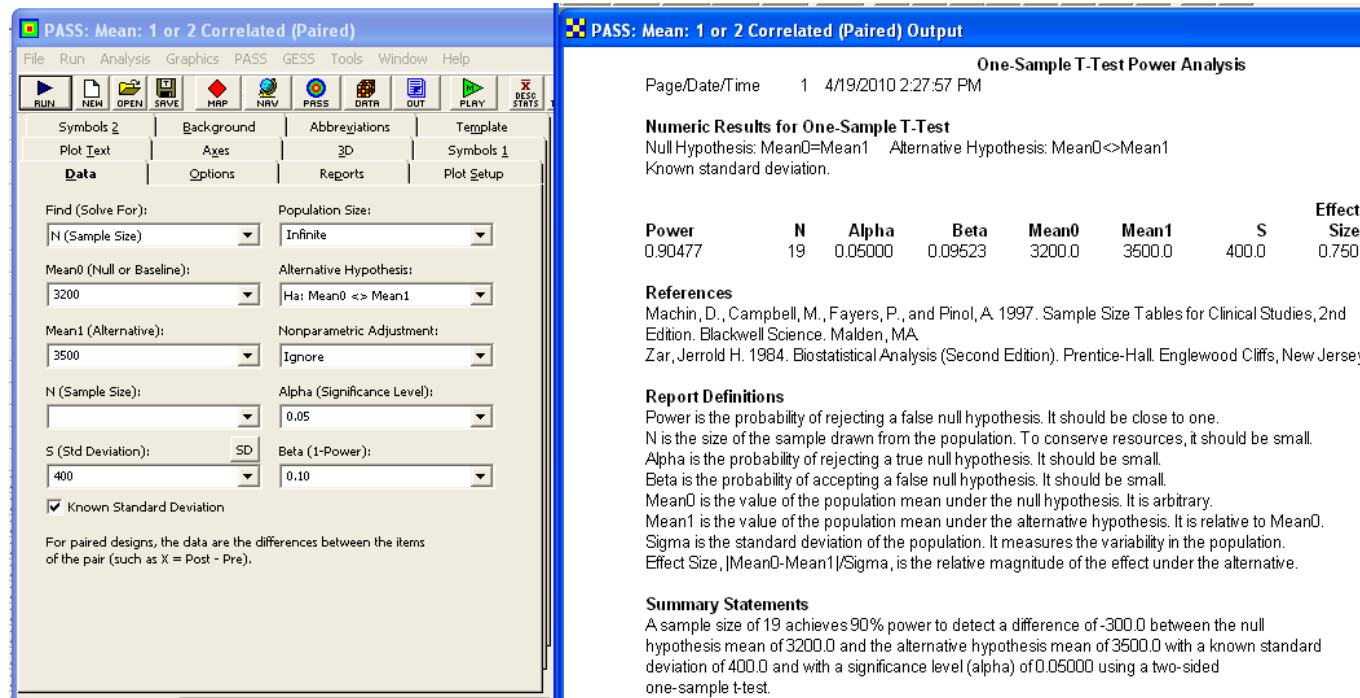
Power and Sample Size

1-Sample Z Test

Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 400

Difference	Sample Size	Target Power	Actual Power
300	19	0.9	0.904769

From PASS> Means> One> Inequality (Normal):



PASS: Mean: 1 or 2 Correlated (Paired)

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For): Population Size:
 N (Sample Size) Infinite

Mean0 (Null or Baseline): Alternative Hypothesis:
 3200 Ha: Mean0 < Mean1

Mean1 (Alternative): Nonparametric Adjustment:
 3500 Ignore

N (Sample Size): Alpha (Significance Level):
 0.05

S (Std Deviation): SD Beta (1-Power):
 400 0.10

Known Standard Deviation

For paired designs, the data are the differences between the items of the pair (such as X = Post - Pre).

PASS: Mean: 1 or 2 Correlated (Paired) Output

Page/Date/Time 1 4/19/2010 2:27:57 PM

One-Sample T-Test Power Analysis

Null Hypothesis: Mean0=Mean1 Alternative Hypothesis: Mean0<Mean1
 Known standard deviation.

Power	N	Alpha	Beta	Mean0	Mean1	S	Effect Size
0.90477	19	0.05000	0.09523	3200.0	3500.0	400.0	0.750

References
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science, Malden, MA.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population. To conserve resources, it should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean0 is the value of the population mean under the null hypothesis. It is arbitrary.
 Mean1 is the value of the population mean under the alternative hypothesis. It is relative to Mean0.
 Sigma is the standard deviation of the population. It measures the variability in the population.
 Effect Size, |Mean0-Mean1|/Sigma, is the relative magnitude of the effect under the alternative.

Summary Statements
 A sample size of 19 achieves 90% power to detect a difference of -300.0 between the null hypothesis mean of 3200.0 and the alternative hypothesis mean of 3500.0 with a known standard deviation of 400.0 and with a significance level (alpha) of 0.05000 using a two-sided one-sample t-test.

Example 1.7 An experiment will be performed to test $H_0 : \mu = 8.0$ versus $H_A : \mu > 8.0$. What sample size is required to reject H_0 with 90% power when $\mu = 8.2$? The process is known to be normally distributed with $\sigma_x = 0.2$.

Solution: For the one-tailed hypothesis test with $\alpha = 0.05$, $\delta = 0.2$ and $\beta = 1 - \pi = 0.10$ the required sample size is

$$\begin{aligned}
 n &= \left(\frac{(z_\alpha + z_\beta) \sigma_x}{\delta} \right)^2 \\
 &= \left(\frac{(z_{0.05} + z_{0.10}) \sigma_x}{\delta} \right)^2 \\
 &= \left(\frac{(1.645 + 1.282) 0.2}{0.2} \right)^2 \\
 &= 9.
 \end{aligned}$$

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

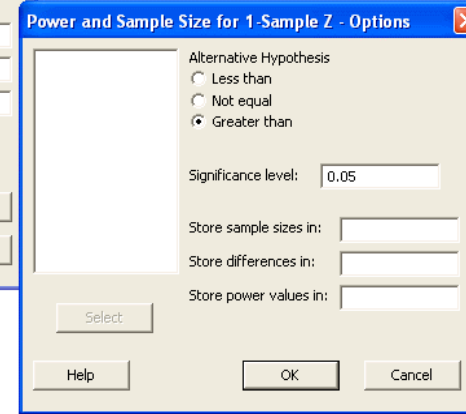
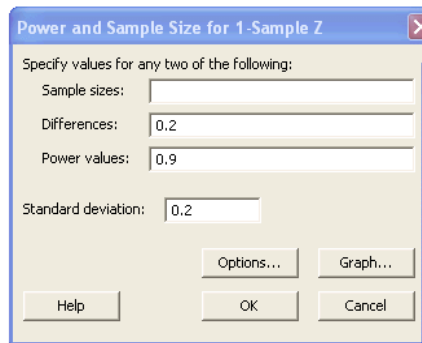
```
MTB > Power;
SUBC> ZOne;
SUBC> Difference 0.2;
SUBC> Power 0.9;
SUBC> Sigma 0.2;
SUBC> Alternative 1;
SUBC> GPCurve.
```

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus > null)
 Calculating power for mean = null + difference
 Alpha = 0.05 Assumed standard deviation = 0.2

Difference	Sample Size	Target Power	Actual Power
0.2	9	0.9	0.912315



From **PASS > Means > One > Inequality (Normal)**:

One-Sample T-Test Power Analysis

Page/Date/Time 1 4/19/2010 2:30:11 PM

Numeric Results for One-Sample T-Test
 Null Hypothesis: Mean0=Mean1 Alternative Hypothesis: Mean0<Mean1
 Known standard deviation.

Power	N	Alpha	Beta	Mean0	Mean1	S	Effect Size
0.91231	9	0.05000	0.08769	8.0	8.2	0.2	1.000

References
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population. To conserve resources, it should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean0 is the value of the population mean under the null hypothesis. It is arbitrary.
 Mean1 is the value of the population mean under the alternative hypothesis. It is relative to Mean0.
 Sigma is the standard deviation of the population. It measures the variability in the population.
 Effect Size, |(Mean0-Mean1)/Sigma, is the relative magnitude of the effect under the alternative.

Summary Statements
 A sample size of 9 achieves 91% power to detect a difference of -0.2 between the null hypothesis mean of 8.0 and the alternative hypothesis mean of 8.2 with a known standard deviation of 0.2 and with a significance level (alpha) of 0.05000 using a one-sided one-sample t-test.

Example 1.8 Calculate the p value for the test performed under the conditions of Example 1.6 if the sample mean was $\bar{x} = 3080$.

Solution: Figure 1.4 shows the contributions to the p value from the two tails of the \bar{x} distribution under H_0 . The z test statistic that corresponds to \bar{x} is

$$\begin{aligned} z &= \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \\ &= \frac{\bar{x} - \mu_0}{\sigma_x / \sqrt{n}} \\ &= \frac{3080 - 3200}{400 / \sqrt{19}} \\ &= -1.31, \end{aligned}$$

so the p value is

$$\begin{aligned} p &= 1 - \Phi(-1.31 < z < 1.31) \\ &= 0.19. \end{aligned}$$

Because $(p = 0.19) > (\alpha = 0.05)$, the observed sample mean is statistically consistent with $H_0 : \mu = 3200$, so we can not reject H_0 .

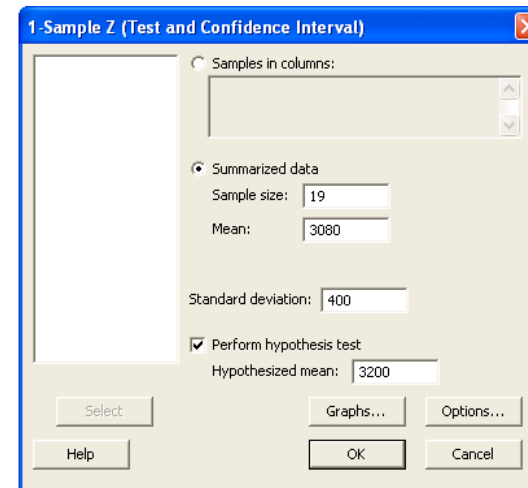
From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

```
MTB > OneZ 19 3080;
SUBC> Sigma 400;
SUBC> Test 3200.
```

One-Sample Z

Test of mu = 3200 vs not = 3200
The assumed standard deviation = 400

N	Mean	SE Mean	95% CI	Z	P
19	3080.0	91.8	(2900.1, 3259.9)	-1.31	0.191



Example 1.9 Calculate the p value for the test performed under the conditions of Example 1.7 if the sample mean was $\bar{x} = 8.39$.

Solution: Figure 1.5 shows the single contribution to the p value from the right tail of the \bar{x} distribution under H_0 . The z test statistic that corresponds to \bar{x} is

$$\begin{aligned} z &= \frac{8.39 - 8.2}{0.2 / \sqrt{9}} \\ &= 2.85, \end{aligned}$$

so the p value is

$$\begin{aligned} p &= \Phi(2.85 < z < \infty) \\ &= 0.0022. \end{aligned}$$

Because $(p = 0.0022) < (\alpha = 0.05)$, the observed sample mean is an improbable result under $H_0 : \mu = 8.2$, so we must reject H_0 .

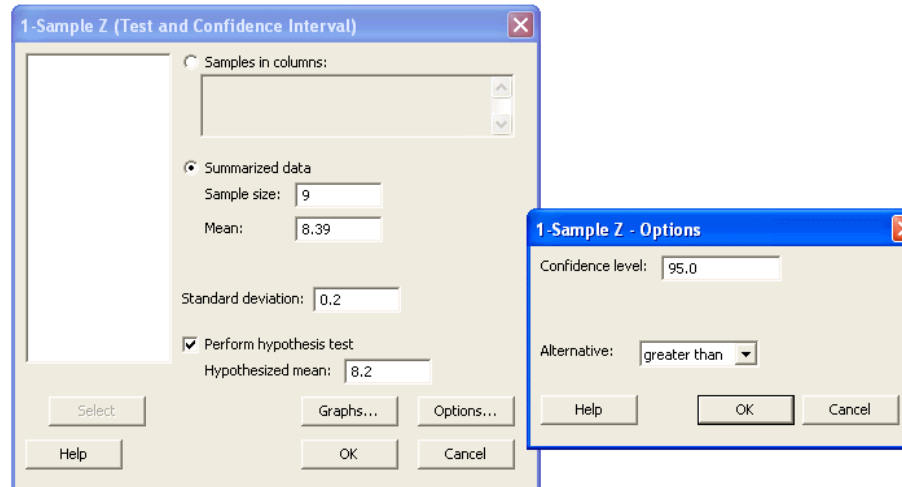
From MINITAB> **Stat> Power and Sample Size> 1-Sample Z:**

```
MTB > OneZ 9 8.39;
SUBC> Sigma 0.2;
SUBC> Test 8.2;
SUBC> Alternative 1.
```

One-Sample Z

Test of $\mu = 8.2$ vs > 8.2
The assumed standard deviation = 0.2

N	Mean	SE Mean	95% Lower Bound	Z	P
9	8.3900	0.0667	8.2803	2.85	0.002



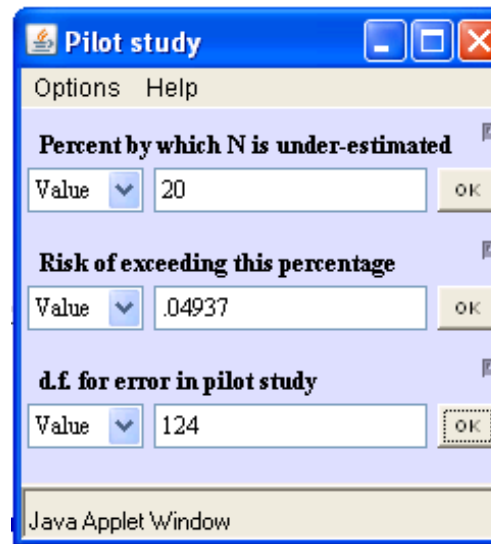
1.4 Practical Considerations

Example 1.10 What sample size is required for a pilot study to estimate the standard deviation to be used in the sample size calculation for a primary experiment if the sample size for the primary experiment should be within 20% of the correct value with 90% confidence?

Solution: With $\delta = 0.20$ and $\alpha = 0.10$ in Equation 1.20, the required sample size for the preliminary experiment to estimate the standard deviation is

$$\begin{aligned} n &\simeq 2 \left(\frac{1.645}{0.20} \right)^2 \\ &\simeq 136. \end{aligned}$$

From **Piface> Pilot Study:**



From PASS> Variance> Variance: 1 Group:

One Variance Power Analysis

Page/Date/Time 1 4/19/2010 3:19:51 PM

Numeric Results when H0: V0 = V1 versus Ha: V0<>V1

Power	N	V0	V1	Alpha	Beta
0.501763	155	1.0000	1.2000	0.100000	0.498237

References
 Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
 Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press, Ames, Iowa.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population.
 V0 is the value of the population variance under the null hypothesis.
 V1 is the value of the population variance under the alternative hypothesis.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements
 A sample size of 155 achieves 50% power to detect a difference of 0.2000 between the null hypothesis variance of 1.0000 and the alternative hypothesis variance of 1.2000 using a two-sided, Chi-square hypothesis test with a significance level (alpha) of 0.100000.

Example 1.11 An engineer must obtain approval from his manager to test a certain number of units to determine the mean response for a validation study. The standard deviation of the response is $\sigma_x = 600$ and the smallest practically significant shift in the mean that the experiment should detect is understood to be $\delta = 400$. What graph should the engineer use to present his case?

Solution: The value of the effect size of interest is firm at $\delta = 400$. The sample size is going to affect the power of the test, so an appropriate graph is power versus sample size. The sample size required to obtain a specified value of power for the test of $H_0 : \delta = 0$ versus $H_A : \delta \neq 0$ is given by Equation 1.12. Figure 1.6 shows the resulting power curve. The sample size required to obtain 80% power is $n = 18$ and the sample size required for 90% power is $n = 24$.

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

```
MTB > Power;
SUBC> ZOne;
SUBC> Difference 400;
SUBC> Power 0.80 0.90;
SUBC> Sigma 600;
SUBC> GPCurve;
SUBC> NSize 4:40.
```

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 600

Difference	Sample Size	Target Power	Actual Power
400	18	0.8	0.807430
400	24	0.9	0.904228

From PASS> Means> One> Inequality (Normal):

PASS: Mean: 1 or 2 Correlated (Paired)

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For): N (Sample Size) Population Size: Infinite
 Mean0 (Null or Baseline): 0 Alternative Hypothesis: Ha: Mean0 <> Mean1
 Mean1 (Alternative): 400 Nonparametric Adjustment: Ignore
 N (Sample Size): Alpha (Significance Level): 0.05
 S (Std Deviation): 600 SD Beta (1-Power): 0.10 0.20
 Known Standard Deviation

For paired designs, the data are the differences between the items of the pair (such as X = Post - Pre).

Opt 1 Template Id:

PASS: Mean: 1 or 2 Correlated (Paired) Output

Page/Date/Time 1 4/19/2010 3:29:01 PM

One-Sample T-Test Power Analysis

Null Hypothesis: Mean0=Mean1 Alternative Hypothesis: Mean0<>Mean1
 Known standard deviation.

Power	N	Alpha	Beta	Mean0	Mean1	S	Effect Size
0.90423	24	0.05000	0.09577	0.0	400.0	600.0	0.667
0.80743	18	0.05000	0.19257	0.0	400.0	600.0	0.667

References

Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA

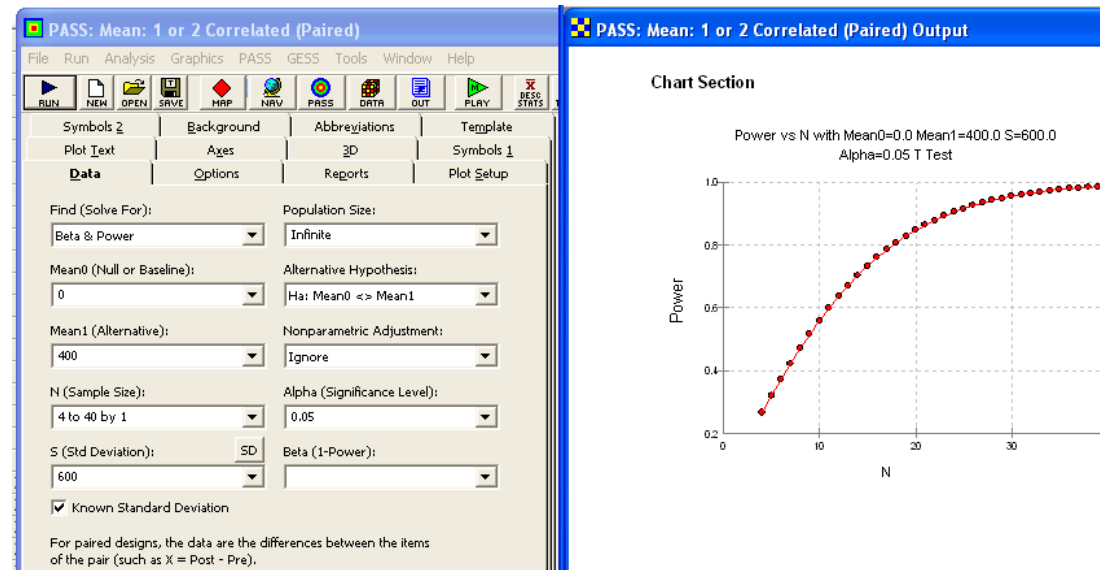
Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population. To conserve resources, it should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean0 is the value of the population mean under the null hypothesis. It is arbitrary.
 Mean1 is the value of the population mean under the alternative hypothesis. It is relative to Mean0.
 Sigma is the standard deviation of the population. It measures the variability in the population.
 Effect Size, |Mean0-Mean1|/Sigma, is the relative magnitude of the effect under the alternative.

Summary Statements

A sample size of 24 achieves 90% power to detect a difference of -400.0 between the null hypothesis mean of 0.0 and the alternative hypothesis mean of 400.0 with a known standard deviation of 600.0 and with a significance level (alpha) of 0.05000 using a two-sided one-sample t-test.



Example 1.12 Suppose that the manager in Example 1.11 approves the use of $n = 24$ units in the validation study. What power does the study have to reject H_0 when the effect size is $\delta = 200, 400,$ and 600 ?⁷

Solution: The power is given by

$$\pi = \Phi(-z_\beta < z < \infty)$$

where z_β is determined from Equation 1.12:

$$z_\beta = \sqrt{n} \frac{\delta}{\sigma_x} - z_{\alpha/2}$$

Figure 1.7 shows the power as a function of effect size. The power to reject H_0 when $\delta = 200$ is $\pi \simeq 0.37$, when $\delta = 400$ is $\pi \simeq 0.90$, and when $\delta = 600$ is $\pi \simeq 1$.

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

```
MTB > Power;
SUBC> ZOne;
SUBC> Sample 24;
SUBC> Difference 400;
SUBC> Sigma 600;
SUBC> GPCurve.
```

Power and Sample Size

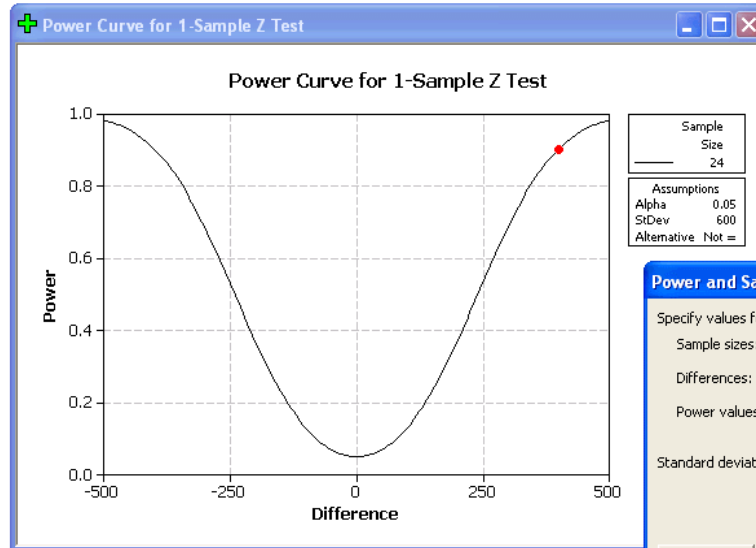
1-Sample Z Test

Testing mean = null (versus not = null)
 Calculating power for mean = null + difference
 Alpha = 0.05 Assumed standard deviation = 600

Difference	Sample Size	Power
400	24	0.904228

Power Curve for 1-Sample Z Test

MTB >



Power and Sample Size for 1-Sample Z

Specify values for any two of the following:

Sample sizes: 24

Differences: 400

Power values:

Standard deviation: 600

Options... Graph... Help OK Cancel

From PASS> Means> One> Inequality (Normal):

PASS: Mean: 1 or 2 Correlated (Paired)

File Run Analysis Graphics PASS GESS Tools Window Help

RUN NEW OPEN SAVE MAP NRV PASS DATA OUT PLAY REG STATS

Symbols Background Abbreviations Template
 Plot Text Axes 3D Symbols
 Data Options Reports Plot Setup

Find (Solve For): Beta & Power

Population Size: Infinite

Mean0 (Null or Baseline): 0

Alternative Hypothesis: Ha: Mean0 <=> Mean1

Mean1 (Alternative): 50 to 600 by 10

Nonparametric Adjustment: Ignore

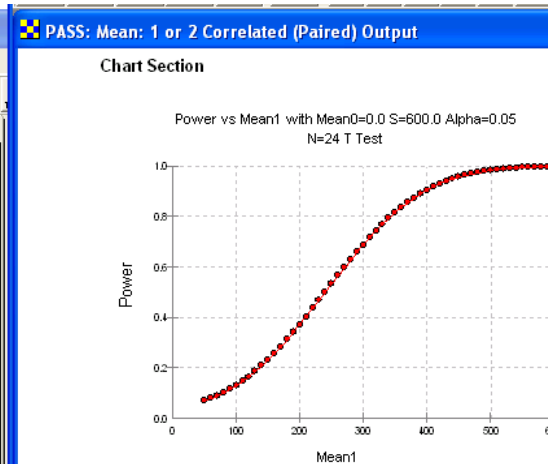
N (Sample Size): 24

Alpha (Significance Level): 0.05

S (Std Deviation): 600

Beta (1-Power):

Known Standard Deviation



1.5 Problems and Solutions**1.6 Software**

Chapter 2

Means

2.1 Assumptions

2.2 One Mean

Example 2.1 Find the sample size required to estimate the unknown mean of a population to within ± 3 with 95% confidence if the population standard deviation is known to be $\sigma = 5$.

Solution: With $\alpha = 0.05$, $z_{0.025} = 1.96$, and $\delta = 3$ in Equation ??, the required sample size is

$$\begin{aligned} n &\geq \left(\frac{1.96 \times 5}{3} \right)^2 \\ &\geq 11. \end{aligned}$$

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

```

MTB > Power;
SUBC> ZOne;
SUBC> Difference 3;
SUBC> Power 0.5;
SUBC> Sigma 5;
SUBC> GPCurve.

```

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus not = null)
 Calculating power for mean = null + difference
 Alpha = 0.05 Assumed standard deviation = 5

Difference	Sample Size	Target Power	Actual Power
3	11	0.5	0.512010

Example 2.2 Find the sample size required to estimate the unknown mean of a population to within $\delta = 3$ measurement units with 95% confidence if the estimated population standard deviation is $\hat{\sigma} = 5$.

Solution: From Equation 2.7 with $t_{0.025} \simeq (z_{0.025} = 1.96)$ in the first iteration,

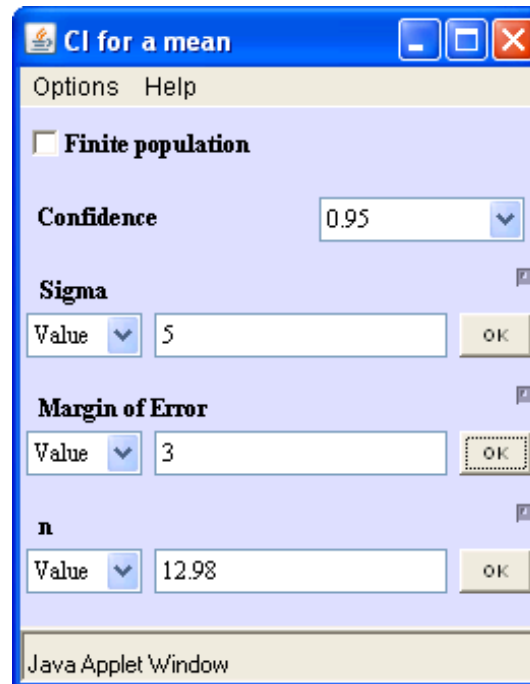
$$n \geq \left(\frac{1.96 \times 5}{3} \right)^2 = 11.$$

In the second iteration with $t_{0.025,10} = 2.228$,

$$n \geq \left(\frac{2.228 \times 5}{3} \right)^2 = 14.$$

Another iteration indicates that $n = 13$ is the smallest sample size that satisfies the sample size condition.

From **Piface> CI for one mean:**



From MINITAB> Stat> Power and Sample Size> 1-Sample t:

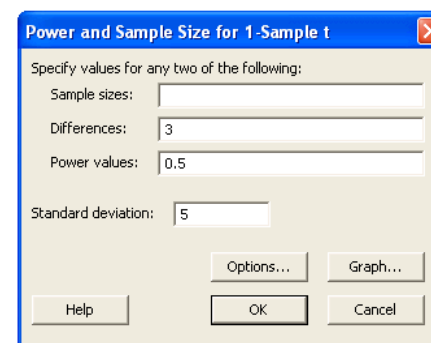
```
MTB > Power;
SUBC> TOne;
SUBC> Difference 3;
SUBC> Power 0.5;
SUBC> Sigma 5;
SUBC> GPCurve.
```

Power and Sample Size

1-Sample t Test

Testing mean = null (versus not = null)
 Calculating power for mean = null + difference
 Alpha = 0.05 Assumed standard deviation = 5

Difference	Sample Size	Target Power	Actual Power
3	13	0.5	0.511701



From MINITAB (V16)> Stat> Power and Sample Size> Sample Size for Estimation> Mean (Normal):

```
MTB > SSCI;
SUBC> NMean;
SUBC> Sigma 5;
SUBC> Confidence 95.0;
SUBC> IType 0;
SUBC> MError 3.
```

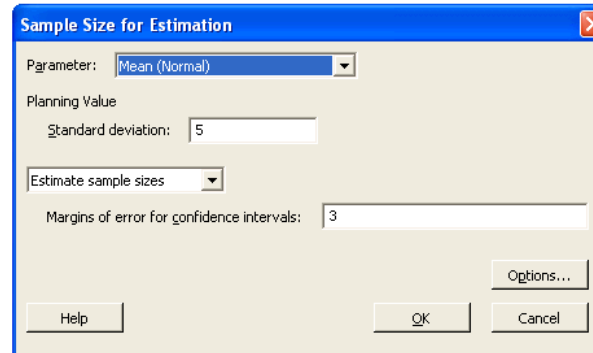
Sample Size for Estimation

Method

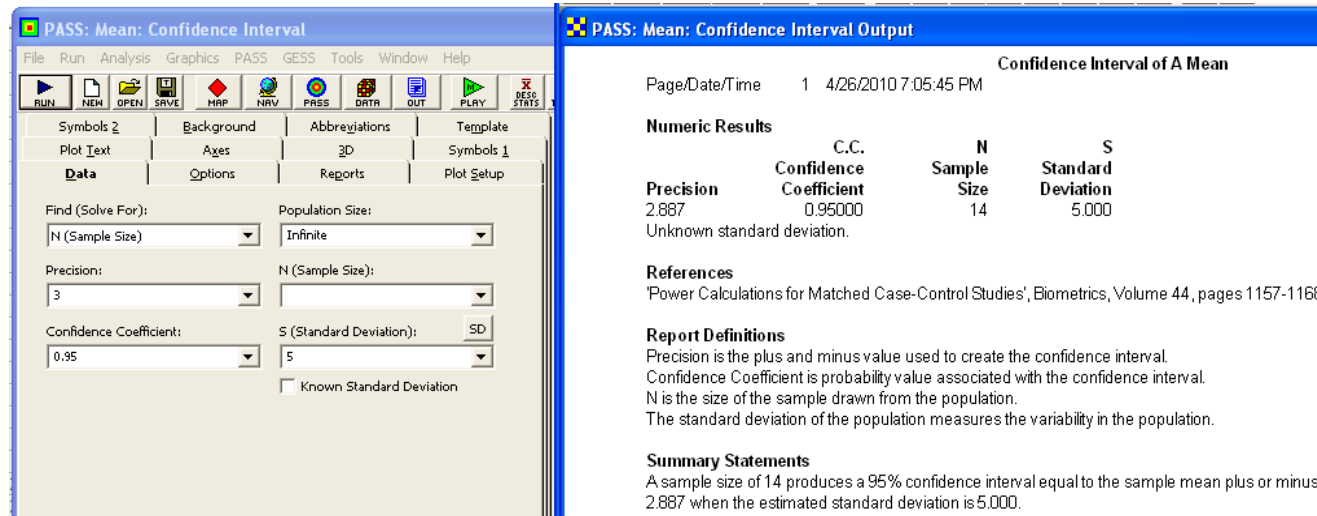
Parameter	Mean
Distribution	Normal
Standard deviation	5 (estimate)
Confidence level	95%
Confidence interval	Two-sided

Results

Margin of Error	Sample Size
3	14



From PASS> Means> One> Confidence Interval of Mean:



Example 2.3 For the one-sample test of $H_0 : \mu = 30$ versus $H_A : \mu \neq 30$ when the population is known to be normal with $\sigma = 3$, what sample size is required to detect a shift to $\mu = 32$ with 90% power?

Solution: By Equation 2.16 with $\delta = 2$, $z_{0.025} = 1.96$, and $z_{0.10} = 1.28$, the necessary sample size is

$$n \geq (1.96 + 1.28)^2 \left(\frac{3}{2}\right)^2 = 24.$$

From PASS> Means> One> Inequality (Normal):

PASS: Mean: 1 or 2 Correlated (Paired)

File Run Analysis Graphics PASS GESS Tools Window Help

RUN NEW OPEN SAVE MAP NAV PASS DATA OUT PLAY DESP STARTS

Symbols \downarrow Background Abbreviations Template
Plot Text Axes 3D Symbols \downarrow
Data Options Reports Plot Setup

Find (Solve For): Population Size:
N (Sample Size) Infinite

Mean0 (Null or Baseline): Alternative Hypothesis:
30 Ha: Mean0 <=> Mean1

Mean1 (Alternative): Nonparametric Adjustment:
32 Ignore

N (Sample Size): Alpha (Significance Level):
0.05

S (Std Deviation): SD Beta (1-Power):
3 0.10

Known Standard Deviation

For paired designs, the data are the differences between the items of the pair (such as X = Post - Pre).

PASS: Mean: 1 or 2 Correlated (Paired) Output

One-Sample T-Test Power Analysis

Numeric Results for One-Sample T-Test
Null Hypothesis: Mean0=Mean1 Alternative Hypothesis: Mean0<>Mean1
Known standard deviation.

Power	N	Alpha	Beta	Mean0	Mean1	S	Effect Size
0.90423	24	0.05000	0.09577	30.0	32.0	3.0	0.667

References
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science, Malden, MA
Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population. To conserve resources, it should be small.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
Mean0 is the value of the population mean under the null hypothesis. It is arbitrary.
Mean1 is the value of the population mean under the alternative hypothesis. It is relative to Mean0.
Sigma is the standard deviation of the population. It measures the variability in the population.
Effect Size, $|\text{Mean0} - \text{Mean1}| / \text{Sigma}$, is the relative magnitude of the effect under the alternative.

Summary Statements
A sample size of 24 achieves 90% power to detect a difference of -2.0 between the null hypothesis mean of 30.0 and the alternative hypothesis mean of 32.0 with a known standard deviation of 3.0 and with a significance level (alpha) of 0.05000 using a two-sided one-sample t-test.

Example 2.4 For the one-sample test of $H_0 : \mu = 30$ versus $H_A : \mu \neq 30$, what sample size is required to detect a shift to $\mu = 32$ with 90% power? The population standard deviation is unknown but expected to be $\sigma \simeq 1.5$.

Solution: The sample size condition given by Equation 2.21 is transcendental, so the correct value of n must be determined iteratively. With $t \simeq z$ as a first guess, $z_{0.025} = 1.96$, $z_{0.10} = 1.282$, and

$$n = (1.96 + 1.282)^2 \left(\frac{1.5}{2} \right)^2 = 6.$$

Then with $df_\epsilon = 5$, $t_{0.025,5} = 2.571$, and $t_{0.10,5} = 1.476$ the new sample size estimate is

$$n \geq (2.571 + 1.476)^2 \left(\frac{1.5}{2} \right)^2 = 9.21.$$

Further iterations are required because $(n = 6) \not\geq 9.21$. Another iteration indicates that $n = 9$ delivers the desired power.

From Piface> One-sample t test (or paired t):

One-sample (or paired) t test

Options Help

sigma

Value OK

True $|\mu - \mu_0|$

Value OK

n

Value OK

power

Value OK

Solve for

alpha Two-tailed

Java Applet Window

From MINITAB> Stat> Power and Sample Size> 1-Sample t:

```
MTB > Power;
SUBC> TOne;
SUBC> Difference 2;
SUBC> Power 0.90;
SUBC> Sigma 1.5;
SUBC> GPCurve.
```

Power and Sample Size

1-Sample t Test

Testing mean = null (versus not = null)
 Calculating power for mean = null + difference
 Alpha = 0.05 Assumed standard deviation = 1.5

Difference	Sample Size	Target Power	Actual Power
2	9	0.9	0.936743

Power and Sample Size for 1-Sample t

Specify values for any two of the following:

Sample sizes:

Differences:

Power values:

Standard deviation:

Options... Graph...

Help OK Cancel

From PASS> Means> One> Inequality (Normal):

PASS: Mean: 1 or 2 Correlated (Paired) Output

One-Sample T-Test Power Analysis

Numeric Results for One-Sample T-Test
 Null Hypothesis: Mean0=Mean1 Alternative Hypothesis: Mean0<>Mean1
 Unknown standard deviation.

Power	N	Alpha	Beta	Mean0	Mean1	S	Effect Size
0.93674	9	0.05000	0.06326	30.0	32.0	1.5	1.333

References
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population. To conserve resources, it should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean0 is the value of the population mean under the null hypothesis. It is arbitrary.
 Mean1 is the value of the population mean under the alternative hypothesis. It is relative to Mean0.
 Sigma is the standard deviation of the population. It measures the variability in the population.
 Effect Size, $|\text{Mean0}-\text{Mean1}|/\text{Sigma}$, is the relative magnitude of the effect under the alternative.

Summary Statements
 A sample size of 9 achieves 94% power to detect a difference of -2.0 between the null hypothesis mean of 30.0 and the alternative hypothesis mean of 32.0 with an estimated standard deviation of 1.5 and with a significance level (alpha) of 0.05000 using a two-sided one-sample t-test.

Example 2.5 Find the approximate and exact power for the solution obtained for Example 2.4.

Solution: With $n = 9$ and $t_{0.025,8} = 2.306$ the approximate power by Equation 2.19 is

$$\begin{aligned}
 \pi &= P\left(-\infty < t < \frac{\delta}{\hat{\sigma}/\sqrt{n}} - t_{\alpha/2}\right) \\
 &= P\left(-\infty < t < \frac{2}{1.5/\sqrt{9}} - 2.306\right) \\
 &= P(-\infty < t < 1.694) \\
 &= 0.9356.
 \end{aligned}$$

From Equation 2.23 the t distribution noncentrality parameter is

$$\phi = \frac{2}{1.5/\sqrt{9}} = 4.00,$$

so, from Equation 2.22,

$$t_{0.025} = 2.306 = t_{\beta,4.0},$$

which is satisfied by $\beta = 0.0633$ and power $\pi = 1 - \beta = 0.9367$. This value is in excellent agreement with the value obtained by the approximate method even though the sample size is relatively small.

From Piface> One-sample t test (or paired t):

Example 2.6 Compare the sample sizes for the two-independent-samples experiment and the paired-sample experiment if they must detect a bias between two treatments of $\Delta\mu = 2$ with 90% power when the standard deviation of individual units is $\hat{\sigma}_x = 2$ and the measurement precision error is $\hat{\sigma}_\epsilon = 0.5$.

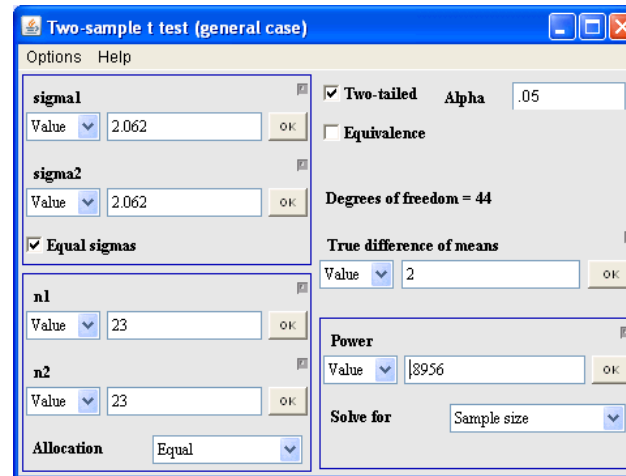
Solution: For the two-independent-sample t test the characteristic standard deviation for each treatment is (from Equation 2.27)

$$\hat{\sigma}_{independent} = \sqrt{2^2 + 0.5^2} = 2.062.$$

Then, from Equation 2.62, the required sample size for each treatment is

$$\begin{aligned}
 n &\geq 2(t_{0.025} + t_{0.10})^2 \left(\frac{\hat{\sigma}_{independent}}{\Delta\mu} \right)^2 \\
 &\geq 2(t_{0.025} + t_{0.10})^2 \left(\frac{2.062}{2} \right)^2 \\
 &\geq 24.
 \end{aligned}$$

From Piface> Two-sample t test (pooled or Satterthwaite):



From MINITAB> Stat> Power and Sample Size> 2-Sample t:

```

MTB > Power;
SUBC>   TTwo;
SUBC>   Difference 2;
SUBC>   Power 0.90;
SUBC>   Sigma 2.062;
SUBC>   GPCurve.
    
```

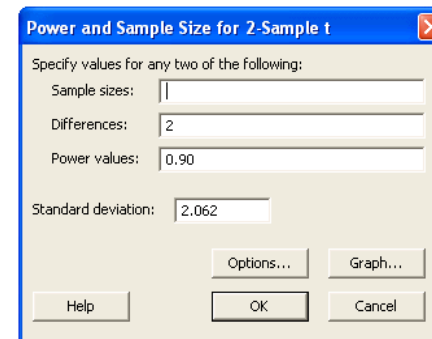
Power and Sample Size

2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
 Calculating power for mean 1 = mean 2 + difference
 Alpha = 0.05 Assumed standard deviation = 2.062

Difference	Sample Size	Target Power	Actual Power
2	24	0.9	0.908083

The sample size is for each group.



From PASS> Means> Two> Independent> Inequality (Normal) [Differences]:

PASS: Means: 2: Inequality [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
 Data | Options | Reports | Plot Setup

Find (Solve For): N1
 Mean1 (Mean of Group 1): 0
 Mean2 (Mean of Group 2): 2
 N1 (Sample Size Group 1):
 N2 (Sample Size Group 2): Use R
 R (Sample Allocation Ratio): 1.0

Alternative Hypothesis: Ha: Mean1 <> Mean2
 Nonparametric Adjustment: Ignore
 Alpha (Significance Level): 0.05
 Beta (1-Power): 0.10
 S1 (Std Deviation Group 1): 2.062
 S2 (Std Deviation Group 2): S1
 Known Std Deviation

PASS: Means: 2: Inequality [Differences] Output

Two-Sample T-Test Power Analysis

Numeric Results for Two-Sample T-Test
 Null Hypothesis: Mean1=Mean2. Alternative Hypothesis: Mean1 <> Mean2
 The standard deviations were assumed to be unknown and equal.

Power	Allocation		Alpha	Beta	Mean1	Mean2	S1	S2
	N1	N2						
0.90808	24	24	1.000	0.05000	0.09192	0.0	2.0	2.1

References
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science, Malden, MA.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. Power should be close to one.
 N1 and N2 are the number of items sampled from each population. To conserve resources, they should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean1 is the mean of populations 1 and 2 under the null hypothesis of equality.
 Mean2 is the mean of population 2 under the alternative hypothesis. The mean of population 1 is unchanged.
 S1 and S2 are the population standard deviations. They represent the variability in the populations.

Summary Statements
 Group sample sizes of 24 and 24 achieve 91% power to detect a difference of -2.0 between the null hypothesis that both group means are 0.0 and the alternative hypothesis that the mean of group 2 is 2.0 with estimated group standard deviations of 2.1 and 2.1 and with a significance level (alpha) of 0.05000 using a two-sided two-sample t-test.

For the paired-sample t test, the characteristic standard deviation for the Δx_i can be estimated from Equation 2.28:

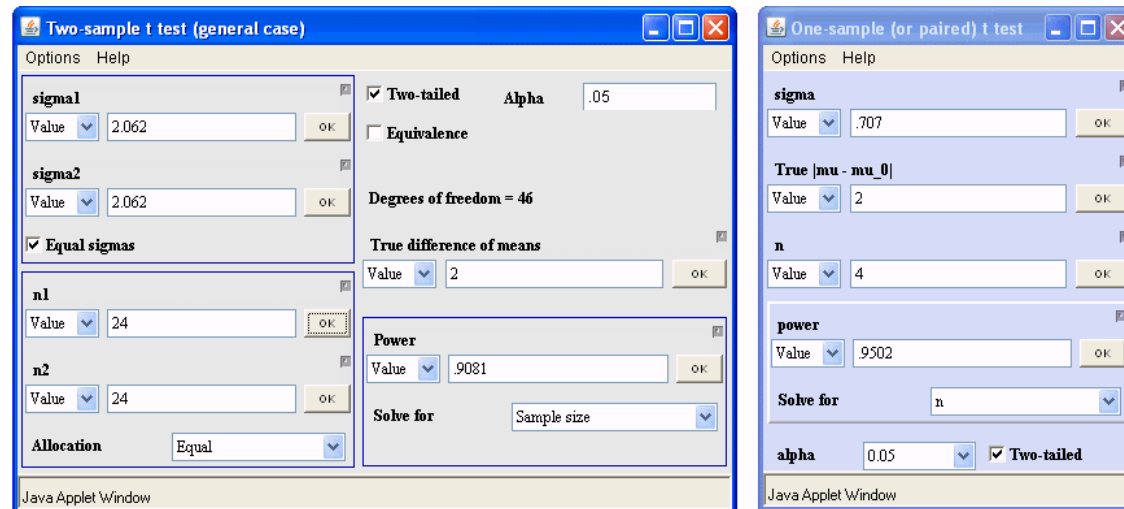
$$\hat{\sigma}_{\Delta x} = \sqrt{2}\hat{\sigma}_{\epsilon} = \sqrt{2} \times 0.5 = 0.707.$$

Then, from Equation 2.21, the required sample size is approximately

$$\begin{aligned}
 n &\geq (t_{0.025} + t_{0.10})^2 \left(\frac{\hat{\sigma}_{\Delta x}}{\Delta \mu} \right)^2 \\
 &\geq (t_{0.025} + t_{0.10})^2 \left(\frac{0.707}{2} \right)^2 \\
 &\geq 4
 \end{aligned}$$

and further iterations confirm that $n = 4$. When the independent-samples design requires two samples of size $n = 24$ units each, for a total of 48 measurements, the paired-sample design requires only $n = 4$ units for a total of 8 measurements!

From Piface> One-sample t test (or paired t):



From MINITAB (V16)> Stat> Power and Sample Size> Paired t:

```

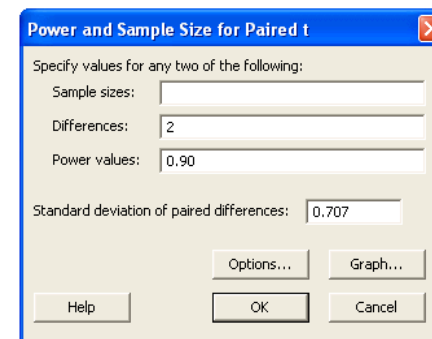
MTB > Power;
SUBC>   TPaired;
SUBC>   Difference 2;
SUBC>   Power 0.90;
SUBC>   Sigma 0.707;
SUBC>   GPCurve.
    
```

Power and Sample Size

Paired t Test

Testing mean paired difference = 0 (versus not = 0)
 Calculating power for mean paired difference = difference
 Alpha = 0.05 Assumed standard deviation of paired differences = 0.707

Difference	Sample Size	Target Power	Actual Power
2	4	0.9	0.950211



From MINITAB> Stat> Power and Sample Size> 1-Sample t:

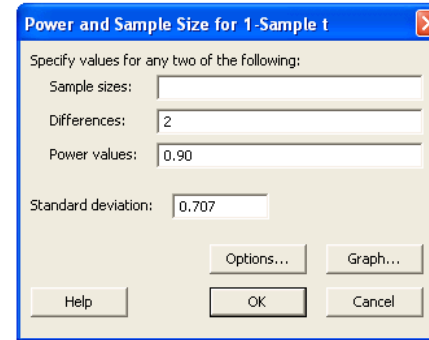
```
MTB > Power;
SUBC> TOne;
SUBC> Difference 2;
SUBC> Power 0.90;
SUBC> Sigma 0.707;
SUBC> GPCurve.
```

Power and Sample Size

1-Sample t Test

Testing mean = null (versus not = null)
 Calculating power for mean = null + difference
 Alpha = 0.05 Assumed standard deviation = 0.707

Difference	Sample Size	Target Power	Actual Power
2	4	0.9	0.950211



From PASS> Means> One> Inequality (Normal):

PASS: Mean: 1 or 2 Correlated (Paired)

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For):
 N (Sample Size) Infinite
 Mean0 (Null or Baseline): 0
 Alternative Hypothesis: Ha: Mean0 <=> Mean1
 Mean1 (Alternative): 2
 Nonparametric Adjustment: Ignore
 N (Sample Size):
 Alpha (Significance Level): 0.05
 S (Std Deviation): 0.707
 Beta (1-Power): 0.10
 Known Standard Deviation

For paired designs, the data are the differences between the items of the pair (such as X = Post - Pre).

PASS: Mean: 1 or 2 Correlated (Paired) Output

Page/Date/Time 1 4/26/2010 7:32:05 PM

One-Sample T-Test Power Analysis

Null Hypothesis: Mean0=Mean1 Alternative Hypothesis: Mean0<>Mean1
 Unknown standard deviation.

Power	N	Alpha	Beta	Mean0	Mean1	S	Effect Size
0.95021	4	0.05000	0.04979	0.0	2.0	0.7	2.829

References
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science, Malden, MA.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population. To conserve resources, it should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean0 is the value of the population mean under the null hypothesis. It is arbitrary.
 Mean1 is the value of the population mean under the alternative hypothesis. It is relative to Mean0.
 Sigma is the standard deviation of the population. It measures the variability in the population.
 Effect Size, |Mean0-Mean1|/Sigma, is the relative magnitude of the effect under the alternative.

Summary Statements
 A sample size of 4 achieves 95% power to detect a difference of -2.0 between the null hypothesis mean of 0.0 and the alternative hypothesis mean of 2.0 with an estimated standard deviation of 0.7 and with a significance level (alpha) of 0.05000 using a two-sided one-sample t-test.

2.3 Two Independent Means

Example 2.7 Find the sample sizes required for the a) equal-allocation and b) optimal-allocation conditions if the 95% two-sided confidence interval for $\Delta\mu$ must have half-width $\delta = 0.003$ when $\sigma_1 = 0.003$ and $\sigma_2 = 0.006$. Compare the total sample sizes required by the two methods.

Solution:

a) By Equation 2.32 the sample size required for equal allocation is

$$n = (1.96)^2 \frac{(0.003)^2 + (0.006)^2}{(0.003)^2} = 20.$$

From Piface> Two-sample t test (pooled or Satterthwaite):

The screenshot shows a dialog box titled "Two-sample t test (general case)". It contains several input fields and checkboxes:

- signal**: Value OK
- sigma2**: Value OK
- Equal sigmas
- n1**: Value OK
- n2**: Value OK
- Allocation**: Equal
- Two-tailed Alpha
- Equivalence
- Degrees of freedom = 29.41
- True difference of means**: Value OK
- Power**: Value OK
- Solve for**: Sample size

From PASS> Means> Two> Independent> Inequality (Normal) [Differences]:

PASS: Means: 2: Inequality [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For): N1
 Alternative Hypothesis: Ha: Mean1 <> Mean2
 Mean1 (Mean of Group 1): 0
 Nonparametric Adjustment: Ignore
 Mean2 (Mean of Group 2): 0.003
 Alpha (Significance Level): 0.05
 N1 (Sample Size Group 1):
 Beta (1-Power): 0.5
 N2 (Sample Size Group 2): Use R
 S1 (Std Deviation Group 1): SD 0.003
 R (Sample Allocation Ratio): 1.0
 S2 (Std Deviation Group 2): 0.006
 Known Std Deviation

PASS: Means: 2: Inequality [Differences] Output

Two-Sample T-Test Power Analysis

Numeric Results for Two-Sample T-Test
 Null Hypothesis: Mean1=Mean2. Alternative Hypothesis: Mean1 <> Mean2
 The standard deviations were assumed to be known and unequal.

	Allocation								
Power	N1	N2	Ratio	Alpha	Beta	Mean1	Mean2	S1	S2
0.51601	20	20	1.000	0.05000	0.48399	0.0	0.0	0.0	0.0

References
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science, Malden, MA.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. Power should be close to one.
 N1 and N2 are the number of items sampled from each population. To conserve resources, they should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean1 is the mean of populations 1 and 2 under the null hypothesis of equality.
 Mean2 is the mean of population 2 under the alternative hypothesis. The mean of population 1 is unchanged.
 S1 and S2 are the population standard deviations. They represent the variability in the populations.

Summary Statements
 Group sample sizes of 20 and 20 achieve 52% power to detect a difference of 0.0 between the null hypothesis that both group means are 0.0 and the alternative hypothesis that the mean of group 2 is 0.0 with known group standard deviations of 0.0 and 0.0 and with a significance level (alpha) of 0.05000 using a two-sided two-sample t-test.

b) By Equations 2.33a and b, the sample sizes required for optimal allocation are

$$n_1 = (1.96)^2 \frac{(0.003)(0.003 + 0.006)}{(0.003)^2} = 12$$

$$n_2 = 11.5 \left(\frac{0.006}{0.003} \right) = 24.$$

For the equal-allocation method, the total sample size is $2n = 40$, and for the optimal-allocation method, the total sample size is $n_1 + n_2 = 36$ - a 10% savings in sample size.

From Piface > Two-sample t test (pooled or Satterthwaite):

Example 2.8 Determine the sample size required to obtain a confidence interval half-width $\delta = 50$ when $\hat{\sigma}_1 = \hat{\sigma}_2 = 80$.

Solution: With $t_{0.025} \simeq z_{0.025}$ for the first iteration, the sample size is

$$n = 2 \left(\frac{1.96 \times 80}{50} \right)^2 = 20. \quad (2.1)$$

Another iteration with $t_{0.025,38} = 2.024$ gives

$$n = 2 \left(\frac{2.024 \times 80}{50} \right)^2 = 21. \quad (2.2)$$

A third iteration (not shown) confirms that $n = 21$ is the necessary sample size.

From **Piface**> **Two-sample t test (pooled or Satterthwaite)**:

Two-sample t test (general case)

Options Help

signal

Value σ_1 80 OK

sigma2

Value σ_2 80 OK

Equal sigmas

n1

Value n_1 21 OK

n2

Value n_2 21 OK

Allocation Equal

Two-tailed Alpha .05

Equivalence

Degrees of freedom = 40

True difference of means

Value $\mu_1 - \mu_2$ 50 OK

Power

Value $1 - \beta$.5066 OK

Solve for Sample size

Java Applet Window

From MINITAB> Stat> Power and Sample Size> 2-Sample t:

From PASS> Means> Two> Independent> Inequality (Normal) [Differences]:

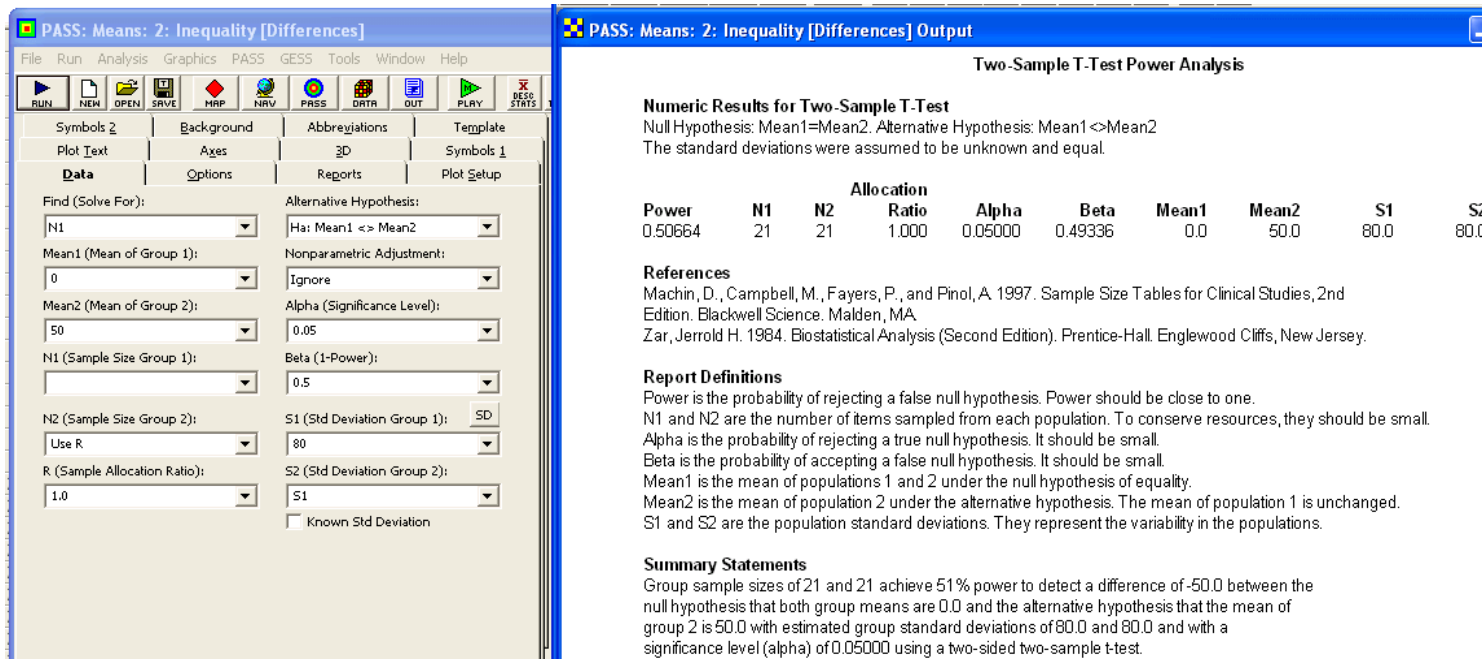
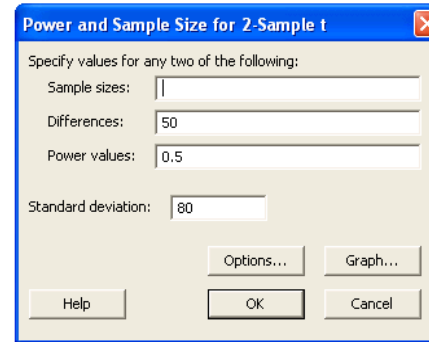
```
MTB > Power;
SUBC> TTwo;
SUBC> Difference 50;
SUBC> Power 0.5;
SUBC> Sigma 80;
SUBC> GPCurve.
```

Power and Sample Size

2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
 Calculating power for mean 1 = mean 2 + difference
 Alpha = 0.05 Assumed standard deviation = 80

Difference	Sample Size	Target Power	Actual Power
50	21	0.5	0.506639



Example 2.9 What optimal sample sizes are required to determine a confidence interval for the difference between two population means with confidence interval half-width $\delta = 15$ when $\hat{\sigma}_1 = 24$ and $\hat{\sigma}_2 = 8$?

Solution: From Equations 2.33a and b, initial guesses for the sample sizes are

$$n_1 \simeq (1.96)^2 \frac{24(24+8)}{15^2} \simeq 14$$

and

$$n_2 \simeq 14 \left(\frac{8}{24} \right) \simeq 5.$$

To obtain optimal sample size allocation n_1 and n_2 must be in the ratio

$$(n_1 : n_2) = (\hat{\sigma}_1 : \hat{\sigma}_2) = (24 : 8) = (3 : 1),$$

so reasonable choices for the sample sizes are $n_1 = 15$ and $n_2 = 5$. By Equation 2.41, the t distribution degrees of freedom will be

$$df_\epsilon = \frac{\left(\frac{24^2}{15} + \frac{8^2}{5} \right)^2}{\frac{24^4}{15^2(15+1)} + \frac{8^4}{5^2(5+1)}} - 2 = 20.$$

With $t_{0.025,20} = 2.086$ the next iteration on the sample sizes gives

$$n_1 \simeq (2.086)^2 \frac{24(24+8)}{15^2} \simeq 15$$

and

$$n_2 \simeq 15 \left(\frac{8}{24} \right) = 5$$

which must be the correct values.

From **Piface**> **Two-sample t test (pooled or Satterthwaite):**

The screenshot shows a Java Applet window titled "Two-sample t test (general case)". The window contains several input fields and options:

- signal**: Value 24, OK button.
- sigma2**: Value 8, OK button.
- Equal sigmas**
- n1**: Value 15, OK button.
- n2**: Value 5, OK button.
- Allocation**: Optimal (dropdown menu).
- Two-tailed** Alpha .05
- Equivalence**
- Degrees of freedom** = 17.92
- True difference of means**: Value 15, OK button.
- Power**: Value .5094, OK button.
- Solve for**: Sample size (dropdown menu).

At the bottom of the window, it says "Java Applet Window".

Example 2.10 Calculate the sample size for the two-sample t test to reject H_0 with 90% power when $|\mu_1 - \mu_2| = 5$. Assume that the sample sizes will be equal and that the two populations have equal standard deviations estimated to be $\hat{\sigma}_\epsilon = 3$. Compare the approximate and exact powers.

Solution: With $\Delta\mu = 5$ and $\hat{\sigma}_\epsilon = 3$ in Equation 2.62, the sample size predicted in the first iteration with $t \simeq z$ is

$$n = 2 \left(\frac{(1.96 + 1.282) 3}{5} \right)^2 = 8.$$

A second and third iteration indicate that the required sample size is $n = 9$.

With $n = 9$ for both samples, $df_\epsilon = 18 - 2 = 16$ and the approximate power is given by Equations 2.58 and 2.60:

$$\begin{aligned}\pi &= P\left(-\infty < t < \sqrt{\frac{n}{2}} \frac{\Delta\mu}{\hat{\sigma}_\epsilon} - t_{0.025,16}\right) \\ &= P\left(-\infty < t < \sqrt{\frac{9}{2}} \frac{5}{3} - 2.12\right) \\ &= P(-\infty < t < 1.416) \\ &= 0.912.\end{aligned}$$

The t distribution noncentrality parameter is given by Equation 2.64:

$$\phi = \sqrt{\frac{9}{2}} \frac{5}{3} = 3.536.$$

The exact power is determined by Equation 2.63 with $\alpha = 0.05$:

$$t_{0.025} = 2.120 = t_{\beta, 3.536},$$

which is satisfied by $\beta = 0.087$, so the exact power is $\pi = 0.913$. The exact power is in excellent agreement with the approximate power despite the somewhat small sample size.

From **Piface**> **Two-sample t test (pooled or Satterthwaite)**:

Two-sample t test (general case)

Options Help

sigma1 Value \downarrow 3 OK

sigma2 Value \downarrow 3 OK

Equal sigmas

n1 Value \downarrow 9 OK

n2 Value \downarrow 9 OK

Allocation Equal \downarrow

Two-tailed Alpha .05

Equivalence

Degrees of freedom = 16

True difference of means Value \downarrow 5 OK

Power Value \downarrow .9125 OK

Solve for Sample size \downarrow

Java Applet Window

From MINITAB> Stat> Power and Sample Size> 2-Sample t:

```
MTB > Power;
SUBC> TTwo;
SUBC> Difference 5;
SUBC> Power 0.90;
SUBC> Sigma 3;
SUBC> GPCurve.
```

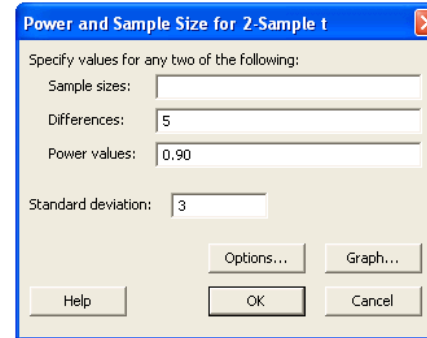
Power and Sample Size

2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
 Calculating power for mean 1 = mean 2 + difference
 Alpha = 0.05 Assumed standard deviation = 3

Difference	Sample Size	Target Power	Actual Power
5	9	0.9	0.912548

The sample size is for each group.



From PASS> Means> Two> Independent> Inequality (Normal) [Differences]:

Two-Sample T-Test Power Analysis

Numeric Results for Two-Sample T-Test
 Null Hypothesis: Mean1=Mean2. Alternative Hypothesis: Mean1 <> Mean2
 The standard deviations were assumed to be unknown and equal.

Power	N1	N2	Allocation Ratio	Alpha	Beta	Mean1	Mean2	S1	S2
0.91255	9	9	1.000	0.05000	0.08745	0.0	5.0	3.0	3.0

References
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. Power should be close to one.
 N1 and N2 are the number of items sampled from each population. To conserve resources, they should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean1 is the mean of populations 1 and 2 under the null hypothesis of equality.
 Mean2 is the mean of population 2 under the alternative hypothesis. The mean of population 1 is unchanged.
 S1 and S2 are the population standard deviations. They represent the variability in the populations.

Summary Statements
 Group sample sizes of 9 and 9 achieve 91% power to detect a difference of -5.0 between the null hypothesis that both group means are 0.0 and the alternative hypothesis that the mean of group 2 is 5.0 with estimated group standard deviations of 3.0 and 3.0 and with a significance level (alpha) of 0.05000 using a two-sided two-sample t-test.

Example 2.11 Determine the size of the second sample under the conditions described in Example 2.10 if the first sample size must be $n = 6$.

Solution: From Example 2.10, the optimal equal sample sizes are $n' = 9$. If $n_1 = 6$ is fixed, then, from Equation 2.68, the approximate value of the second sample size must be

$$n_2 = \frac{6 \times 9}{(2 \times 6) - 9} = 18.$$

In the equal- n solution, we had $n_1 + n_2 = 18$ and $df_\epsilon = 16$ with 91% power; therefore, we know that $n_1 + n_2 = 6 + 18 = 24$ and $df_\epsilon = 22$ will give a slightly larger power, so the next guess for n_2 can be a value less than $n_2 = 18$. By appropriate guesses and iterations, the required value of n_2 is determined to be $n_2 = 15$ with approximate power

$$\begin{aligned} \pi &= P\left(-\infty < t < \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \left(\frac{\Delta\mu}{\hat{\sigma}}\right) - t_{0.025,19}\right) \\ &= P\left(-\infty < t < \frac{1}{\sqrt{\frac{1}{6} + \frac{1}{15}}} \left(\frac{5}{3}\right) - 2.093\right) \\ &= P(-\infty < t < 1.357) \\ &= 0.905. \end{aligned}$$

From Piface > Two-sample t test (pooled or Satterthwaite):

Two-sample t test (general case)

Options Help

signal

Value σ_1 3 OK

sigma2

Value σ_2 3 OK

Equal sigmas

n1

Value n_1 6 OK

n2

Value n_2 15 OK

Allocation Independent

Two-tailed Alpha .05

Equivalence

Degrees of freedom = 19

True difference of means

Value $\mu_1 - \mu_2$ 5 OK

Power

Value $1 - \beta$.9051 OK

Solve for Sample size

Java Applet Window

From PASS> Means> Two> Independent> Inequality (Normal) [Differences]:

PASS: Means: 2: Inequality [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
 Data | Options | Reports | Plot Setup

Find (Solve For): N2
 Alternative Hypothesis: Ha: Mean1 <=> Mean2
 Mean1 (Mean of Group 1): 0
 Nonparametric Adjustment: Ignore
 Mean2 (Mean of Group 2): 5
 Alpha (Significance Level): 0.05
 N1 (Sample Size Group 1): 6
 Beta (1-Power): 0.10
 N2 (Sample Size Group 2):
 S1 (Std Deviation Group 1): SD
 3
 R (Sample Allocation Ratio): 1.0
 S2 (Std Deviation Group 2): S1
 3.0
 Known Std Deviation

PASS: Means: 2: Inequality [Differences] Output

Page/Date/Time 1 4/26/2010 8:40:47 PM

Two-Sample T-Test Power Analysis

Numeric Results for Two-Sample T-Test
 Null Hypothesis: Mean1=Mean2. Alternative Hypothesis: Mean1 <=> Mean2
 The standard deviations were assumed to be unknown and equal.

Power	Allocation		Alpha	Beta	Mean1	Mean2	S1	S2	
	N1	N2							Ratio
0.90514	6	15	2.500	0.05000	0.09486	0.0	5.0	3.0	3.0

References
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. Power should be close to one.
 N1 and N2 are the number of items sampled from each population. To conserve resources, they should be small.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Mean1 is the mean of populations 1 and 2 under the null hypothesis of equality.
 Mean2 is the mean of population 2 under the alternative hypothesis. The mean of population 1 is unchanged.
 S1 and S2 are the population standard deviations. They represent the variability in the populations.

Summary Statements
 Group sample sizes of 6 and 15 achieve 91% power to detect a difference of -5.0 between the null hypothesis that both group means are 0.0 and the alternative hypothesis that the mean of group 2 is 5.0 with estimated group standard deviations of 3.0 and 3.0 and with a significance level (alpha) of 0.05000 using a two-sided two-sample t-test.

2.4 Equivalence Tests

Example 2.12 Determine the sample size required for a one-sample equivalence test of the hypotheses $H_0 : \mu < 490$ or $\mu > 510$ versus $H_A : 490 < \mu < 510$ if the experiment must have 90% power to reject H_0 when $\mu = 505$ and $\sigma = 4$.

Solution: With $\mu_0 = 500$, $\mu = 505$, and $\delta = 10$, the sample size given by Equation 2.74 is

$$\begin{aligned}
 n &= \left(\frac{(z_{0.05} + z_{0.10}) \sigma}{\delta - \Delta\mu} \right)^2 \\
 &= \left(\frac{(1.645 + 1.282) 4}{10 - 5} \right)^2 \\
 &= 6.
 \end{aligned}$$

Example 2.13 Determine the power of the two independent-sample equivalence test where μ_1 and μ_2 are considered to be practically equivalent if $|\Delta\mu| < 2$ when $\Delta\mu = 0.2$, $\sigma_1 = \sigma_2 = 2$, and $n_1 = n_2 = 20$.

Solution: With $\delta = 2$ as the limit of practical equivalence, the hypotheses to be tested are

$$\begin{aligned}
 H_{01} &: \Delta\mu \leq -2 \text{ versus } H_{A1} : \Delta\mu > -2 \\
 H_{02} &: \Delta\mu \geq 2 \text{ versus } H_{A2} : \Delta\mu < 2.
 \end{aligned}$$

From Equation 2.79 with $\Delta\mu = 0.2$, the power of the equivalence test is

$$\begin{aligned}\pi &= \Phi\left(\frac{-2 - 0.2}{\sqrt{\frac{2}{20^2}}} + 1.645 < z < \frac{2 - 0.2}{\sqrt{\frac{2}{20^2}}} - 1.645\right) \\ &= \Phi(-1.83 < z < 1.20) \\ &= 0.85.\end{aligned}$$

PASS and Piface do the two-sample t equivalence test which gives power comparable to that of the z test for this example with relatively large error degrees of freedom ($df_\epsilon = 20 + 20 - 2 = 38$).

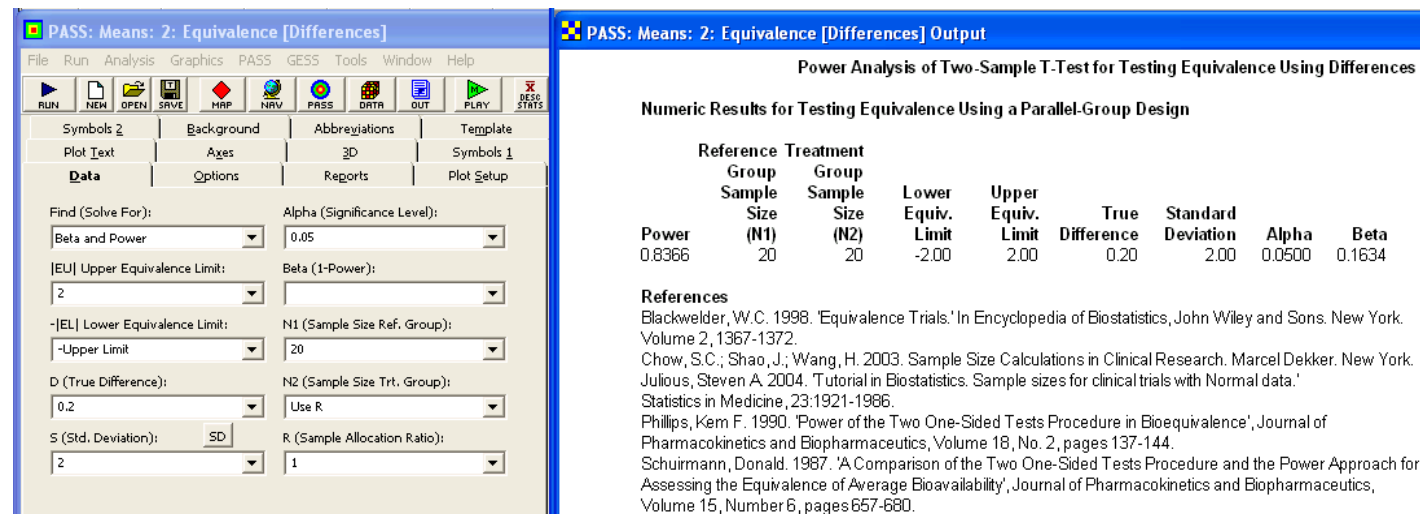
Piface> Two-sample t test:

The screenshot shows a Java Applet window titled "Two-sample t test (general case)". The window contains several input fields and checkboxes:

- sigma1:** Value 2
- sigma2:** Value 2
- Equal sigmas:** Checked
- n1:** Value 20
- n2:** Value 20
- Allocation:** Equal
- Two-tailed:** Checked
- Alpha:** .05
- Equivalence Threshold:** 2
- Degrees of freedom:** 38
- True difference of means:** Value 2
- Power:** Value .8366
- Solve for:** Sample size

At the bottom of the window, it says "Java Applet Window".

PASS> Means> Two> Independent> Equivalence [Difference]:



PASS: Means: 2: Equivalence [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
 Data | Options | Reports | Plot Setup

Find (Solve For): Alpha (Significance Level):

|E| Upper Equivalence Limit: Beta (1-Power):

-|E| Lower Equivalence Limit: N1 (Sample Size Ref. Group):

D (True Difference): N2 (Sample Size Trt. Group):

S (Std. Deviation): R (Sample Allocation Ratio):

PASS: Means: 2: Equivalence [Differences] Output

Power Analysis of Two-Sample T-Test for Testing Equivalence Using Differences

Numeric Results for Testing Equivalence Using a Parallel-Group Design

	Reference Group Sample Size (N1)	Treatment Group Sample Size (N2)	Lower Equiv. Limit	Upper Equiv. Limit	True Difference	Standard Deviation	Alpha	Beta
Power	20	20	-2.00	2.00	0.20	2.00	0.0500	0.1634

References

Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.

Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.

Julious, Steven A. 2004. 'Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data.' Statistics in Medicine, 23:1921-1986.

Phillips, Kem F. 1990. 'Power of the Two One-Sided Tests Procedure in Bioequivalence', Journal of Pharmacokinetics and Biopharmaceutics, Volume 18, No. 2, pages 137-144.

Schuurmann, Donald. 1987. 'A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability', Journal of Pharmacokinetics and Biopharmaceutics, Volume 15, Number 6, pages 657-680.

Example 2.14 What sample size is required in Example 2.13 to obtain 90% power?

Solution: From Equation 2.80 with $\beta = 0.10$, the sample size is

$$\begin{aligned}
 n &= 2 \left(\frac{(1.645 + 1.282) 2}{2 - 0.2} \right)^2 \\
 &= 22.
 \end{aligned}$$

PASS and Piface do the two-sample t equivalence test which gives comparable sample size to that of the z test for this example with relatively large error degrees of freedom.

From Piface> **Two-sample t test:**

Two-sample t test (general case)

Options Help

sigma1 Value OK

sigma2 Value OK

Equal sigmas

n1 Value OK

n2 Value OK

Allocation Equal

Two-tailed Alpha

Equivalence Threshold

Degrees of freedom = 46

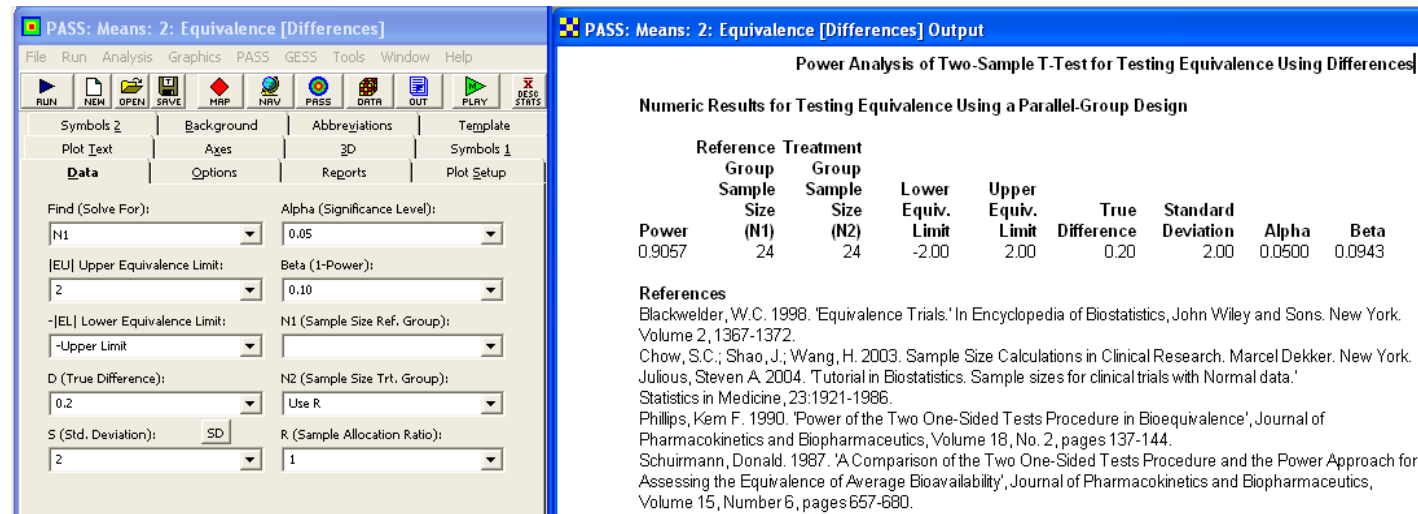
True difference of means Value OK

Power Value OK

Solve for Sample size

Java Applet Window

From PASS> Means> Two> Independent> Equivalence [Difference]:



PASS: Means: 2: Equivalence [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For): N1 Alpha (Significance Level): 0.05

|EU| Upper Equivalence Limit: 2 Beta (1-Power): 0.10

-|EL| Lower Equivalence Limit: -Upper Limit N1 (Sample Size Ref. Group):

D (True Difference): 0.2 N2 (Sample Size Trt. Group): Use R

S (Std. Deviation): SD R (Sample Allocation Ratio): 1

PASS: Means: 2: Equivalence [Differences] Output

Power Analysis of Two-Sample T-Test for Testing Equivalence Using Differences

Numeric Results for Testing Equivalence Using a Parallel-Group Design

	Reference Group Sample Size (N1)	Treatment Group Sample Size (N2)	Lower Equiv. Limit	Upper Equiv. Limit	True Difference	Standard Deviation	Alpha	Beta
Power	24	24	-2.00	2.00	0.20	2.00	0.0500	0.0943

References

Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.

Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.

Julious, Steven A. 2004. Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data. Statistics in Medicine, 23:1921-1986.

Phillips, Kern F. 1990. 'Power of the Two One-Sided Tests Procedure in Bioequivalence', Journal of Pharmacokinetics and Biopharmaceutics, Volume 18, No. 2, pages 137-144.

Schuurmann, Donald. 1987. 'A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability', Journal of Pharmacokinetics and Biopharmaceutics, Volume 15, Number 6, pages 657-680.

2.5 Contrasts

Example 2.15 How many observations per treatment group are required to estimate the contrast

$$\mu_c = \left(\frac{\mu_1 + \mu_2 + \mu_3}{3} \right) - \mu_4$$

to within $\delta = 80$ measurement units with 95% confidence if the one-way ANOVA standard error is $s_e = 200$?

Solution: The goal is to obtain a 95% confidence interval for the contrast of the form given in Equation 2.85 with a confidence interval half-width of $\delta = 80$. The contrast coefficients are $c_i = \left\{ \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -1 \right\}$. If there are sufficient error degrees of freedom so that $t \simeq z$, then, from Equation 2.87, the approximate sample size is

$$\begin{aligned}
 n &\simeq \left(\frac{1.96 \times 200}{80} \right)^2 \left(\left(\frac{1}{3} \right)^2 + \left(\frac{1}{3} \right)^2 + \left(\frac{1}{3} \right)^2 + (-1)^2 \right) \\
 &\simeq 33.
 \end{aligned}$$

With $df_e = k(n - 1) = 4(32) = 128$ error degrees of freedom the $t \simeq z$ approximation is satisfied, so the sample size is accurate.

Piface doesn't calculate the sample size, but it can be used to confirm the answer by showing that the sample size $n = 33$ produces about 50% power. From **Piface> Balanced ANOVA> One-way ANOVA> Differences/Contrasts:**



From PASS> Means> Many Means> ANOVA: One-Way:

One Way ANOVA Power Analysis

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.50605	33.00	4	132	0.05000	0.49395	34.64	200.00	0.1732

References

Desu, M. M. and Raghavarao, D. 1990. Sample Size Methodology. Academic Press. New York.
 Fleiss, Joseph L. 1986. The Design and Analysis of Clinical Experiments. John Wiley & Sons. New York.
 Kirk, Roger E. 1982. Experimental Design: Procedures for the Behavioral Sciences. Brooks/Cole. Pacific Grove, California.

Summary Statements

In a one-way ANOVA study, sample sizes of 33, 33, 33, and 33 are obtained from the 4 groups whose means are to be compared using a planned comparison (contrast). The total sample of 132 subjects achieves 51% power to detect a non-zero contrast of the means versus the alternative that the contrast is zero using an F test with a 0.05000 significance level. The value of the contrast of the means is -240.00. The common standard deviation within a group is assumed to be 200.00.

2.6 Multiple Comparisons Tests

Example 2.16 Determine the sample size required per treatment to detect a difference $\Delta\mu = 200$ between two treatment means using Bonferroni-corrected two-sample t tests for all possible pairs of five treatments with 90% power. Assume that the five populations are normal and homoscedastic with $\hat{\sigma}_\epsilon = 100$.

Solution: With $k = 5$ treatments there will be $K = \binom{5}{2} = 10$ two-sample t tests to perform. To restrict the family error rate to $\alpha_{family} = 0.05$, the Bonferroni-corrected error rate for individual tests is

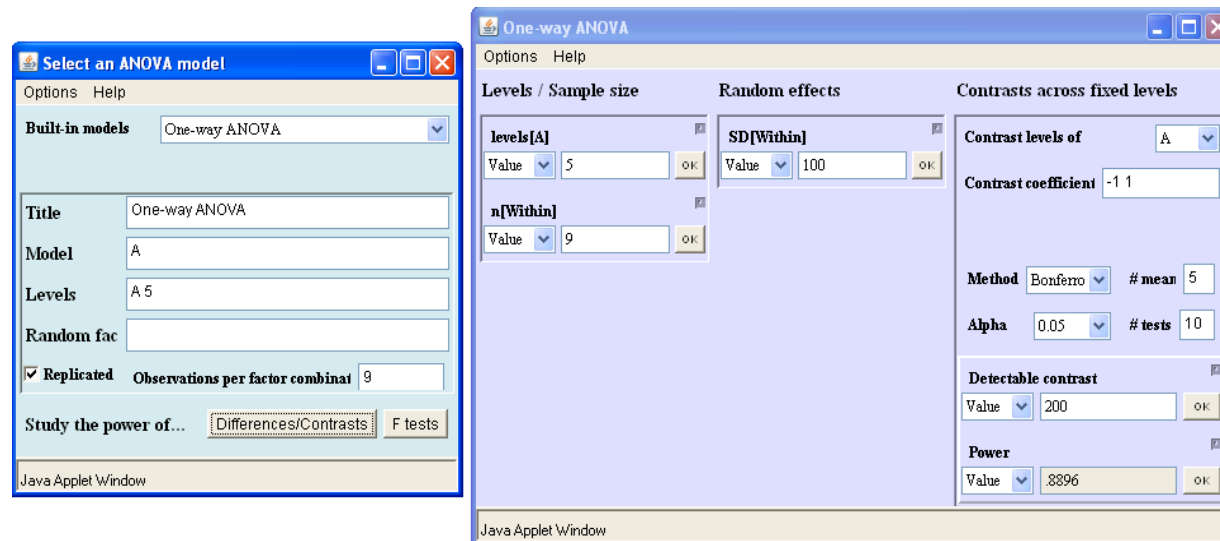
$$\alpha = \frac{0.05}{10} = 0.005.$$

By Equation 2.62 with $t \simeq z$, the sample size is

$$\begin{aligned} n &= 2 \left(\frac{(z_{0.0025} + z_{0.10}) \hat{\sigma}_\epsilon}{\delta} \right)^2 \\ &= 2 \left(\frac{(2.81 + 1.282) 100}{200} \right)^2 = 9. \end{aligned}$$

There will be $df_\epsilon = df_{total} - df_{model} = (5 \times 9 - 1) - (4) = 40$ degrees of freedom to estimate $\hat{\sigma}_\epsilon$ from the pooled treatment standard deviations, so the approximation $t \simeq z$ is justified.

From **Piface > Balanced ANOVA > One-way ANOVA > Differences/Contrasts:**



PASS uses a more conservative method for analyzing multiple comparisons which gives a larger sample size.

Example 2.17 Determine the approximate power for the sample size calculated in Example 2.16.

Solution: The approximate power for the test is given by Equations 2.58 and 2.60 with $\alpha = 0.005$:

$$\begin{aligned}
 \pi &= P(-\infty < t < t_\beta) \\
 &= P\left(-\infty < t < \left(\sqrt{\frac{n}{2}} \frac{\Delta\mu}{\hat{\sigma}} - t_{\alpha/2}\right)\right) \\
 &= P\left(-\infty < t < \left(\sqrt{\frac{9}{2}} \frac{200}{100} - t_{0.0025,40}\right)\right) \\
 &= P(-\infty < t < 1.273) \\
 &= 0.895.
 \end{aligned}$$

Example 2.18 Bonferroni's method becomes very conservative when the number of tests gets very large. A less conservative method for determining α for individual tests is given by Sidak's method:

$$\alpha = 1 - (1 - \alpha_{family})^{1/K}. \quad (2.93)$$

Compare the sample sizes determined using Bonferroni's and Sidak's methods for multiple comparisons between all possible pairs of fifteen treatments when the tests must detect a difference of $\Delta\mu = 8$ with 90% power when $\hat{\sigma}_\epsilon = 6$.

Solution: The number of multiple comparisons tests required is

$$\binom{15}{2} = \frac{15 \times 14}{2} = 105.$$

By Bonferroni's method with $\alpha_{family} = 0.05$, the α for individual tests is

$$\alpha = \frac{0.05}{105} = 0.000476,$$

so with $t \simeq z$ in Equation 2.62 the sample size is

$$\begin{aligned}
 n &= 2 \left(\frac{(z_{0.000476/2} + z_{0.10}) \hat{\sigma}_\epsilon}{\delta} \right)^2 \\
 &= 2 \left(\frac{(3.494 + 1.282) 6}{8} \right)^2 = 26.
 \end{aligned}$$

By Sidak's method (Equation 2.63), the α for individual tests is

$$\alpha = 1 - (1 - 0.05)^{1/105} = 0.000488,$$

so the sample size is

$$\begin{aligned}
 n &= 2 \left(\frac{(z_{0.000488/2} + z_{0.10}) \hat{\sigma}_\epsilon}{\delta} \right)^2 \\
 &= 2 \left(\frac{(3.487 + 1.282) 6}{8} \right)^2 = 26.
 \end{aligned}$$

Even with over 100 multiple comparisons, the sample sizes by the two calculation methods are still equal.

Example 2.19 An experiment will be performed to compare four treatment groups to a control group. Determine the sample size required to detect a difference $\delta = 200$ between the treatments and the control using Bonferroni-corrected two-sample t tests with 90% power. Use a balanced design with the same number of observations in each of the five groups and assume that the five populations are normal and homoscedastic with $\hat{\sigma}_\epsilon = 100$.

Solution: To restrict the family error rate to $\alpha_{family} = 0.05$ with $K = 4$ tests, the Bonferroni-corrected error rate for individual tests is

$$\alpha = \frac{0.05}{4} = 0.0125.$$

By Equation 2.62 with $t \simeq z$, the sample size is

$$\begin{aligned} n &= 2 \left(\frac{(z_{0.0125/2} + z_{0.10}) \hat{\sigma}_\epsilon}{\delta} \right)^2 \\ &= 2 \left(\frac{(2.50 + 1.282) 100}{200} \right)^2 = 8. \end{aligned}$$

Despite the small treatment-group sample size, the approximation $t \simeq z$ is justified because there will be $df_\epsilon = df_{total} - df_{model} = (5 \times 8 - 1) - (4) = 35$ degrees of freedom to estimate $\hat{\sigma}_\epsilon$ from the five pooled treatment standard deviations.

Piface offers Dunnett's test, but it uses the Bonferroni correction to approximate Dunnett's method so it gives the same result. From **Piface**> **Balanced ANOVA**> **One-way ANOVA**> **Differences/Contrasts**:

Example 2.20 Repeat Example 2.19 using the optimal allocation of units to treatments and controls.

Solution: From Equation 2.98 with $t \simeq z$ and $K = 4$,

$$n_i = \left(1 + \frac{1}{\sqrt{4}} \right) \left(\frac{(2.50 + 1.282) 100}{200} \right)^2 = 6$$

and

$$n_0 = n_i \sqrt{K} = 6\sqrt{4} = 12.$$

The approximation $t \simeq z$ is still justified because the error degrees of freedom will be $df_\epsilon = (4 \times 6 + 12) - 4 = 32$. The original experiment required $5 \times 8 = 40$ units, but the optimal experiment requires only $4 \times 6 + 12 = 36$ units to obtain the same power.

Chapter 3

Standard Deviations

3.1 One Standard Deviation

Example 3.1 Determine the sample size required to construct the 95% confidence interval for σ based on a random sample of size n drawn from a normal population if the confidence interval half-width must be about 10% of the sample standard deviation.

Solution: From Table 3.1 the sample size must be about $n = 200$. The lower and upper confidence limits will fall at about -9% and +11% relative to the sample standard deviation, so the asymmetry for this relatively large sample size is not too severe.

From **PASS**> **Variance**> **Variance: 1 Group** (Note that when the **Scale** text box is set to *Standard Deviation*, other text boxes on the form with labels that refer to variances are interpreted as standard deviations.):

The screenshot shows the PASS: Variance: 1 software interface. The left pane displays the input parameters for a one-variance power analysis:

- Find (Solve For): N
- Scale: Standard Deviation
- V0 (Baseline Variance): 1
- Alternative Hypothesis: Ha: V0 <> V1
- V1 (Alternative Variance): 1.1
- Alpha (Significance Level): 0.05
- N (Sample Size): [empty]
- Beta (1-Power): 0.5
- Known Mean

The right pane displays the output for the analysis:

One Variance Power Analysis

Numeric Results when H0: S0 = S1 versus Ha: S0 <> S1

Power	N	S0	S1	Alpha	Beta
0.500515	200	1.0000	1.1000	0.050000	0.499485

References

Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
 Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press. Ames, Iowa.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population.
 S0 is the value of the population standard deviation under the null hypothesis.
 S1 is the value of the population standard deviation under the alternative hypothesis.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

A sample size of 200 achieves 50% power to detect a difference of 0.1000 between the null hypothesis standard deviation of 1.0000 and the alternative hypothesis standard deviation of 1.1000 using a two-sided, Chi-square hypothesis test with a significance level (alpha) of 0.050000.

Example 3.2 Use the large sample approximation method to determine the sample size for the situation in Example 3.1.

Solution: The required confidence interval has the form

$$P(s(1 - 0.10) < \sigma < s(1 + 0.10)) = 0.95.$$

With $\alpha = 0.05$ and $\delta = 0.10$ in Equation 3.8, the sample size required to obtain a confidence interval of the desired half-width is

$$n = \frac{1}{2} \left(\frac{1.96}{0.10} \right)^2 = 193,$$

which is in excellent agreement with the original solution.

From MINITAB (V16) > Stat > Power and Sample Size > Sample Size for Estimation > Standard deviation (Normal):

```
MTB > SSCI;
SUBC> NStDev 100;
SUBC> Confidence 95.0;
SUBC> IType 0;
SUBC> HError 10.
```

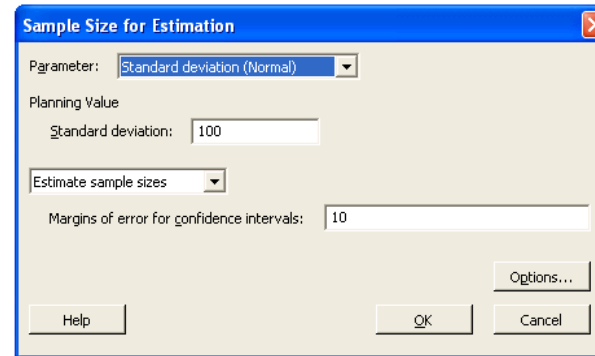
Sample Size for Estimation

Method

Parameter	Standard deviation
Distribution	Normal
Standard deviation	100
Confidence level	95%
Confidence interval	Two-sided

Results

Margin of Error	Sample Size
10	234



Example 3.3 For the test of $H_0 : \sigma^2 = 10$ versus $H_A : \sigma^2 > 10$, find the power associated with $\sigma^2 = 20$ when the sample size is $n = 20$ using $\alpha = 0.05$.

Solution: From Equation 3.14 the power is given by

$$\begin{aligned} \pi &= P\left(\chi_{0.95}^2 \left(\frac{10}{20}\right) < \chi^2 < \infty\right) \\ &= P(15.1 < \chi^2 < \infty) \\ &= 0.72. \end{aligned}$$

From MINITAB (V16)> Stat> Power and Sample Size> 1 Variance:

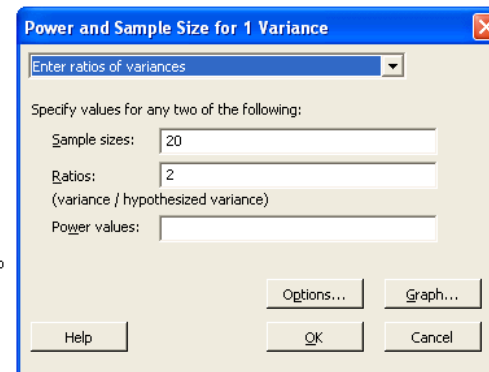
```
MTB > Power;
SUBC> OneVariance;
SUBC> Sample 20;
SUBC> Ratio 2;
SUBC> Alternative 1;
SUBC> Alpha 0.05;
SUBC> GPCurve.
```

Power and Sample Size

Test for One Variance

Testing variance = null (versus > null)
Calculating power for (variance / null) = ratio
Alpha = 0.05

Ratio	Sample Size	Power
2	20	0.718025



From PASS> Variance> Variance: 1 Group:

The image shows two windows from the PASS software. The left window, titled "PASS: Variance: 1", is the input interface. It has a menu bar (File, Run, Analysis, Graphics, PASS, GESS, Tools, Window, Help) and a toolbar with icons for RUN, NEW, OPEN, SAVE, HRP, NVV, PASS, DTR, OUT, and PLRY. Below the toolbar are several tabs: Symbols 2, Background, Abbreviations, Templates, Plot Text, Axes, 3D, Symbols, Data, Options, Reports, and Plot Setup. The main area contains input fields for "Find (Solve For):" (Beta and Power), "Scale:" (Variance), "V0 (Baseline Variance):" (10), "Alternative Hypothesis:" (Ha: V0 < V1), "V1 (Alternative Variance):" (20), "Alpha (Significance Level):" (0.05), "N (Sample Size):" (20), and "Beta (1-Power):" (0.1). There is also a checkbox for "Known Mean".

The right window, titled "PASS: Variance: 1 Output", displays the results of a "One Variance Power Analysis". It includes a table of "Numeric Results when H0: V0 = V1 versus Ha: V0 < V1":

Power	N	V0	V1	Alpha	Beta
0.718025	20	10.0000	20.0000	0.050000	0.281975

Below the table are "References" (Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York. Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press. Ames, Iowa. Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.) and "Summary Statements" (A sample size of 20 achieves 72% power to detect a difference of 10.0000 between the null hypothesis variance of 10.0000 and the alternative hypothesis variance of 20.0000 using a one-sided, Chi-square hypothesis test with a significance level (alpha) of 0.050000.)

Example 3.4 Find the sample size required to reject $H_0 : \sigma^2 = 40$ with 90% power when $\sigma^2 = 100$ using $H_A : \sigma^2 > 40$ with $\alpha = 0.05$.

Solution: From Equation 3.15 with $\sigma_0^2 = 40$ and $\sigma_1^2 = 100$, the necessary sample size is the smallest value of n that meets the requirement

$$\frac{\chi_{0.95}^2}{\chi_{0.10}^2} \leq \frac{100}{40} \leq 2.5.$$

By inspecting Table 3.2 and a table of χ^2 values, the required sample size is $n = 22$ for which

$$\left(\frac{\chi_{0.95}^2}{\chi_{0.10}^2} = \frac{32.67}{13.24} = 2.469 \right) \leq 2.5.$$

From MINITAB (V16)> Stat> Power and Sample Size> 1 Variance:

```
MTB > Power;
SUBC> OneVariance;
SUBC> Ratio 2.5;
SUBC> Power 0.90;
SUBC> Alternative 1;
SUBC> Alpha 0.05;
SUBC> GPCurve.
```

Power and Sample Size

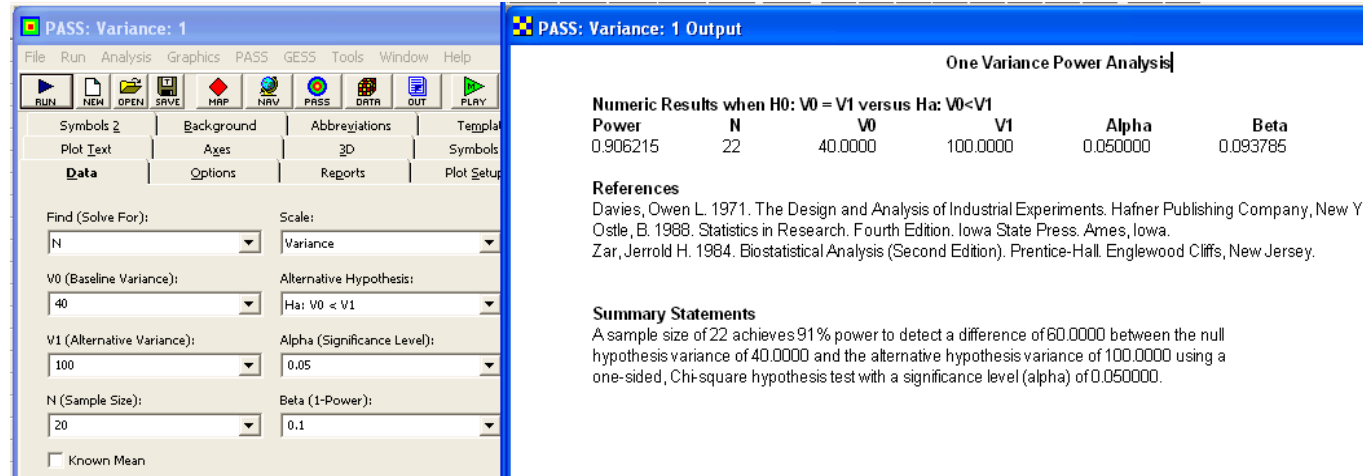
Test for One Variance

Testing variance = null (versus > null)
Calculating power for (variance / null) = ratio
Alpha = 0.05

Ratio	Sample Size	Target Power	Actual Power
2.5	22	0.9	0.906215

The image shows the "Power and Sample Size for 1 Variance" dialog box. It has a dropdown menu for "Enter ratios of variances" and a section titled "Specify values for any two of the following:". The "Ratios:" field is set to 2.5, and the "Power values:" field is set to 0.90. There are buttons for "Options...", "Graph...", "Help", "OK", and "Cancel".

From PASS> Variance> Variance: 1 Group:



Example 3.5 Find the sample size required to reject $H_0 : \sigma = 0.003$ in favor of $H_A : \sigma < 0.003$ with 90% power when in fact $\sigma = 0.001$.

Solution: With $\alpha = 0.05$ and $\beta = 1 - \pi = 0.10$, the sample size condition given by Equation 3.17 is

$$\frac{\chi_{0.90}^2}{\chi_{0.05}^2} \geq \left(\frac{0.003}{0.001} \right)^2$$

$$\geq 9.0,$$

which, from Table 3.2, is satisfied by $n = 5$.

From MINITAB (V16)> Stat> Power and Sample Size> 1 Variance:

From PASS> Variance> Variance: 1 Group:

```

MTE > Power;
SUBC> OneVariance;
SUBC> StDeviation;
SUBC> Ratio 0.3333;
SUBC> Power 0.90;
SUBC> Alternative -1;
SUBC> Alpha 0.05;
SUBC> GPCurve.

Power and Sample Size

Test for One Standard Deviation

Testing StDev = null (versus < null)
Calculating power for (StDev / null) = ratio
Alpha = 0.05

      Sample Target
      Ratio Size Power Actual Power
0.3333     6   0.9   0.933121

```

One Variance Power Analysis

Numeric Results when H0: S0 = S1 versus Ha: S0 > S1					
Power	N	S0	S1	Alpha	Beta
0.933069	6	0.0030	0.0010	0.050000	0.066931

References
Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press, Ames, Iowa.
Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.

Summary Statements
A sample size of 6 achieves 93% power to detect a difference of 0.0020 between the null hypothesis standard deviation of 0.0030 and the alternative hypothesis standard deviation of 0.0010 using a one-sided, Chi-square hypothesis test with a significance level (alpha) of 0.050000.

Example 3.6 Compare the power determined by the large-sample approximation method to the exact power determined in Example 3.3.

Solution: The null hypothesis may be written as $H_0 : \ln(\sigma) = \ln(\sqrt{10})$ and we wish to find the power to reject H_0 when $\ln(\sigma) = \ln(\sqrt{20})$ with $n = 20$. From Equation 3.20 we have

$$z_\beta = \sqrt{2 \times 20} \ln \left(\sqrt{\frac{20}{10}} \right) - z_{0.05} = 0.547$$

and by Equation 3.19 the approximate power is

$$\begin{aligned}\pi &= \Phi(-0.547 < z < \infty) \\ &= 0.71.\end{aligned}$$

This result is still in good agreement with the exact power of 72% despite the rather small sample size.

Example 3.7 Compare the sample size determined by the large-sample approximation method to the exact sample size determined in Example 3.4.

Solution: The problem is to find the sample size to reject $H_0 : \ln(\sigma) = \ln(\sqrt{40})$ with 90% power when $\ln(\sigma) = \ln(\sqrt{100})$. With $\alpha = 0.05$ and $\beta = 0.10$ in Equation ?? the approximate sample size required is

$$n = \frac{1}{2} \left(\frac{1.645 + 1.282}{\ln\left(\sqrt{\frac{100}{40}}\right)} \right)^2 = 21,$$

which is in good agreement with the exact sample size of $n = 22$.

3.2 Two Standard Deviations

Example 3.8 What equal- n sample size is required by an experiment to deliver a confidence interval for the ratio of two independent population standard deviations if the true ratio should fall within 20% of the experimental ratio with 95% confidence?

Solution: The goal of the experiment is to determine an interval of the form

$$P\left(\frac{s_1}{s_2}(1 - 0.2) < \frac{\sigma_1}{\sigma_2} < \frac{s_1}{s_2}(1 + 0.2)\right) = 1 - \alpha.$$

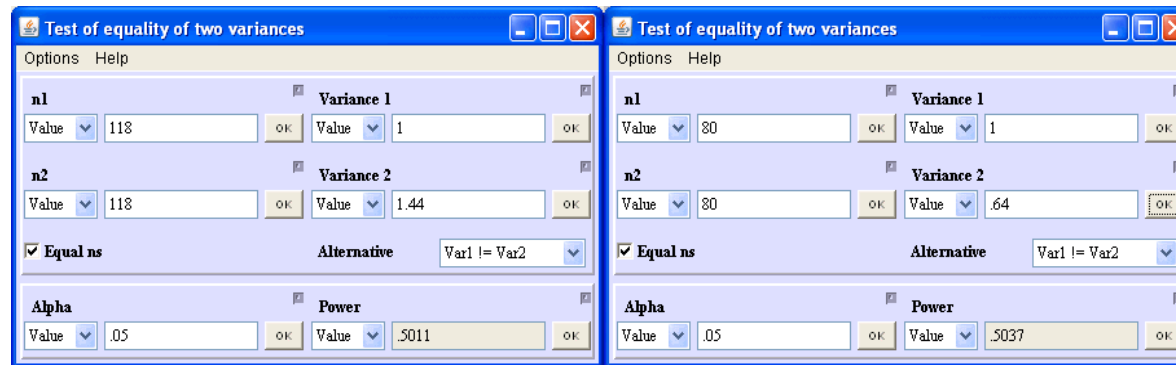
Then, from Equation ?? with $\delta = 0.2$, the required sample sizes are

$$n_1 = n_2 = \left(\frac{z_{0.025}}{\delta}\right)^2 = \left(\frac{1.96}{0.20}\right)^2 = 97.$$

Piface and PASS support the F test for two variances which can be tricked into confirming the sample size for the confidence interval. The confidence interval is asymmetric so the sample size is taken as the average of the sample sizes required for $\sigma_2/\sigma_1 = 1.2$ and $\sigma_2/\sigma_1 = 0.8$.

From Piface> **Two Variances (F Test):**

From PASS> **Variance> Variance: 2 Groups:**



Power Analysis of Two Variances

Numeric Results when H0: S1 = S2 versus Ha: S1<>S2							
Power	N1	N2	S1	S2	Alpha	Beta	
0.501070	118	118	1.0000	1.2000	0.050000	0.498930	

Numeric Results when H0: S1 = S2 versus Ha: S1<>S2							
Power	N1	N2	S1	S2	Alpha	Beta	
0.503697	80	80	1.0000	0.8000	0.050000	0.49630	

The confidence interval will be asymmetric so the sample size calculation was run for S2 20% greater and 20% less than S1. The average of the two resulting sample sizes is $n = (118 + 80)/2 = 99$.

References
 Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
 Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press, Ames, Iowa.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.]

Summary Statements
 Group sample sizes of 118 and 118 achieve 50% power to detect a ratio of 0.8333 between the group one standard deviation of 1.0000 and the group two standard deviation of 1.2000 using a two-sided F test with a significance level (alpha) of 0.050000.

The sample sizes are $n = 118$ and $n = 80$, respectively, so the average sample size is $n = (118 + 80) / 2 = 99$ which is in excellent agreement with the normal approximation.

Example 3.9 Find the power to reject $H_0 : \sigma_1^2 = \sigma_2^2$ in favor of $H_A : \sigma_1^2 > \sigma_2^2$ if $n_1 = n_2 = 26$, $\sigma_1^2 = 15$, and $\sigma_2^2 = 5$ using $\alpha = 0.05$.

Solution: From Equation 3.34 the power is

$$\begin{aligned}\pi &= P\left(\left(\frac{\sigma_2}{\sigma_1}\right)^2 F_{1-\alpha} < F < \infty\right) \\ &= P\left(\left(\frac{5}{15}\right) F_{0.95,25,25} < F < \infty\right) \\ &= P(0.652 < F < \infty) \\ &= 0.854.\end{aligned}$$

From MINITAB (V16)> Stat> Power and Sample Size> 2 Variances:

```
MTB > Power;
SUBC> TwoVariance;
SUBC> Sample 26;
SUBC> Ratio 3;
SUBC> Alternative 1;
SUBC> Alpha 0.05;
SUBC> FTest;
SUBC> GPCurve.
```

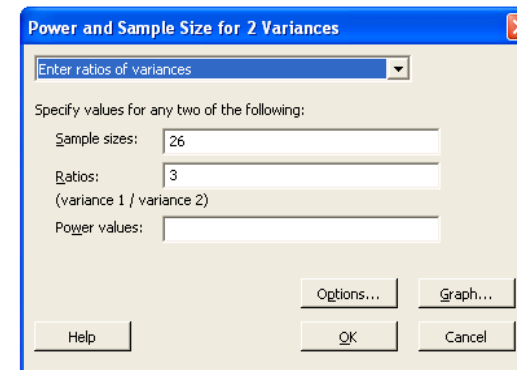
Power and Sample Size

Test for Two Variances

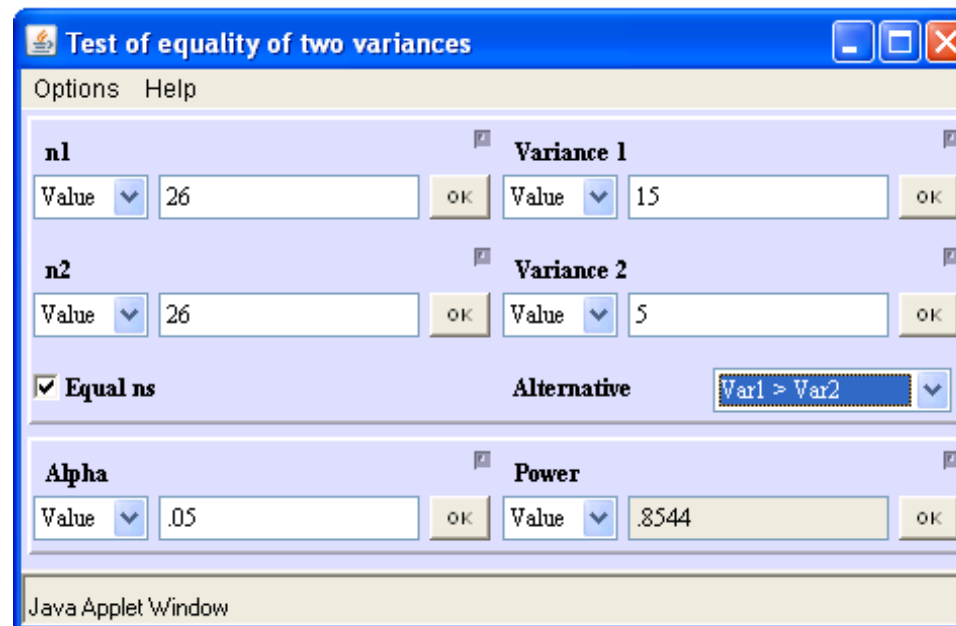
Testing (variance 1 / variance 2) = 1 (versus >)
Calculating power for (variance 1 / variance 2) = ratio
Alpha = 0.05
Method: F Test

Ratio	Sample Size	Power
3	26	0.854374

The sample size is for each group.



From Piface> Two Variances (F Test):



From PASS> Variance> Variance: 2 Groups:

Power Analysis of Two Variances

Numeric Results when H0: V1 = V2 versus Ha: V1 > V2

Power	N1	N2	V1	V2	Alpha	Beta
0.854374	26	26	15.0000	5.0000	0.050000	0.145626

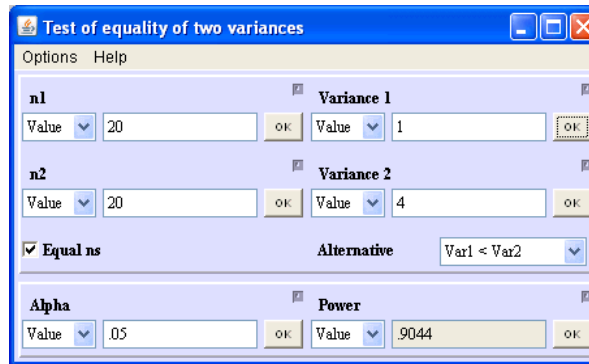
References

Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
 Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press. Ames, Iowa.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Summary Statements

Group sample sizes of 26 and 26 achieve 85% power to detect a ratio of 3.0000 between the group one variance of 15.0000 and the group two variance of 5.0000 using a one-sided F test with a significance level (alpha) of 0.050000.

Example 3.10 What equal sample size is required to detect a factor of two difference between two population standard deviations with 90% power and $\alpha = 0.05$?



Solution: A factor of two difference in population standard deviation corresponds to a factor of four difference in variance, so we need to determine the sample size such that

$$P\left(\frac{1}{4}F_{1-\alpha} < F < \infty\right) = 0.90.$$

By iterating through several values of sample size, we find that when $n_1 = n_2 = 20$, $F_{0.95,19,19} = 2.168$ and $F_{0.096} = 2.168/4 = 0.542$, which satisfies the problem statement.

From MINITAB (V16)> Stat> Power and Sample Size> 2 Variances:

```
MTB > Power;
SUBC> TwoVariance;
SUBC> StDeviation;
SUBC> Ratio 2;
SUBC> Power 0.90;
SUBC> Alternative 1;
SUBC> Alpha 0.05;
SUBC> FTest;
SUBC> GPcurve.
```

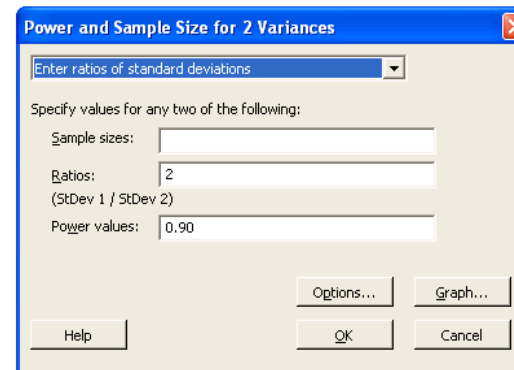
Power and Sample Size

Test for Two Standard Deviations

Testing (StDev 1 / StDev 2) = 1 (versus >)
 Calculating power for (StDev 1 / StDev 2) = ratio
 Alpha = 0.05
 Method: F Test

Ratio	Sample Size	Target Power	Actual Power
2	20	0.9	0.904437

The sample size is for each group.



From Piface> Two Variances (F Test):

From PASS> Variance> Variance: 2 Groups:

The screenshot shows the PASS software interface. The left window, titled "PASS: Variances: 2", contains the following input parameters:

- Find (Solve For): N1
- Scale: Variance
- V1 (Variance of Group 1): 1
- Alternative Hypothesis: Ha: V1 < V2
- V2 (Variance of Group 2): 4
- N1 (Sample Size Group 1): (blank)
- Alpha (Significance Level): .05
- N2 (Sample Size Group 2): Use R
- Beta (1-Power): 0.10
- R (Sample Allocation Ratio): 1.0

The right window, titled "PASS: Variances: 2 Output", displays the results of a "Power Analysis of Two Variances":

Numeric Results when H0: V1 = V2 versus Ha: V1 < V2

Power	N1	N2	V1	V2	Alpha	Beta
0.904437	20	20	1.0000	4.0000	0.050000	0.095563

References

Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
 Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press. Ames, Iowa.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Summary Statements

Group sample sizes of 20 and 20 achieve 90% power to detect a ratio of 0.2500 between the group one variance of 1.0000 and the group two variance of 4.0000 using a one-sided F test with a significance level (alpha) of 0.050000.

Example 3.11 Repeat Example 3.9 using the large-sample approximation method.

Solution: From the information given in the example problem statement

$$z_{\beta} = \frac{\ln\left(\sqrt{\frac{15}{5}}\right)}{\sqrt{\frac{1}{2}\left(\frac{1}{26} + \frac{1}{26}\right)}} - 1.645 = 1.16$$

so the power is

$$\pi = \Phi(-1.16 < z < \infty) = 0.877,$$

which is in good agreement with the exact solution of $\pi = 0.854$.

Example 3.12 Repeat Example 3.10 using the large-sample approximation method.

Solution: From the information given in the example problem statement

$$n_1 = n_2 = \left(\frac{1.645 + 1.282}{\ln(2)}\right)^2 = 18,$$

which slightly underestimates the exact solution $n = 20$.

3.3 Coefficient of Variation

Example 3.13 Determine the sample size required to estimate the population coefficient of variation to within $\pm 25\%$ with 95% confidence if the coefficient of variation is expected to be about 30%.

Solution: With $\alpha = 0.05$, $\delta = 0.25$, and $\widehat{CV} = 0.3$ in Equation 3.45, the sample size must be

$$n = \left(\frac{1.96}{0.25} \right)^2 \left((0.3)^2 + \frac{1}{2} \right) = 37.$$

Example 3.14 Determine the sample size required to reject $H_0 : CV = 0.5$ with 90% power when $CV = 0.8$.

Solution: With $CV_0 = 0.5$, $CV_1 = 0.8$, $\alpha = 0.05$, and $\beta = 0.10$ in Equation 3.49, the required sample size is

$$n = \left(\frac{1.96 \times 0.5 \sqrt{(0.5)^2 + \frac{1}{2}} + 1.282 \times 0.8 \sqrt{(0.8)^2 + \frac{1}{2}}}{0.8 - 0.5} \right)^2 = 42.$$

Example 3.15 Determine the sample size required to reject $H_0 : CV_1 = CV_2$ in favor of $H_A : CV_1 \neq CV_2$ with 90% power when $CV_1 = 0.3$ and $CV_2 = 0.5$.

Solution: With $CV_1 = 0.3$, $CV_2 = 0.5$, $\alpha = 0.05$, and $\beta = 0.10$ in Equation 3.57, the required sample size is

$$n = \left(\frac{1.96 \times 0.3 \sqrt{(0.3)^2 + \frac{1}{2}} + 1.282 \times 0.5 \sqrt{(0.5)^2 + \frac{1}{2}}}{0.3 - 0.5} \right)^2 = 26.$$

Chapter 4

Proportions

4.1 One Proportion (Large Population)

Example 4.1 How large a random sample is required to demonstrate that the fraction defective of a process is less than 1% with 95% confidence?

Solution: The required confidence interval has the form

$$P(0 < p < 0.01) = 0.95$$

so $p_U = 0.01$ and $\alpha = 0.05$. If we assume that the sample size is small compared to the lot size, then Equation 4.4 can be used to approximate the sample size. However, because the number of defectives allowed in the sample was not specified, we must consider the possibility of different X values. For $X = 0$, by the rule of three (Equation 4.5), the sample size is

$$\begin{aligned} n &\approx \frac{3}{0.01} \\ &\approx 300. \end{aligned}$$

For $X = 1$, by Equation 4.4

$$\begin{aligned} n &\approx \frac{\chi_{0.95,4}^2}{2(0.01)} \\ &\approx \frac{9.49}{2(0.01)} \\ &\approx 475. \end{aligned}$$

The values of n can be found for other choices of X in a similar manner.

Piface> CI for one proportion, PASS> Proportions> One Group> Confidence Interval - Proportion, and MINITAB> Stat> Power and Sample Size> 1 Proportion use the normal approximation to the binomial distribution to calculate the sample size for a symmetric two-tailed confidence interval but the normal approximation isn't valid for this problem.

Example 4.2 What fraction of a large population must be inspected and found to be free of defectives to be 95% confident that the population contains no more than ten defectives?

Solution: The goal of the experiment is to demonstrate that the population defective count satisfies the confidence interval $P(0 < S \leq 10) = 0.95$. With $X = 0$ and $\alpha = 0.05$ in Equation 4.7, the fraction of the population that will need to be inspected is

$$\frac{n}{N} \approx \frac{\chi_{0.95}^2}{2S_U} \quad (4.1)$$

$$\begin{aligned} &\approx \frac{3}{10} \\ &\approx 0.30. \end{aligned} \quad (4.2)$$

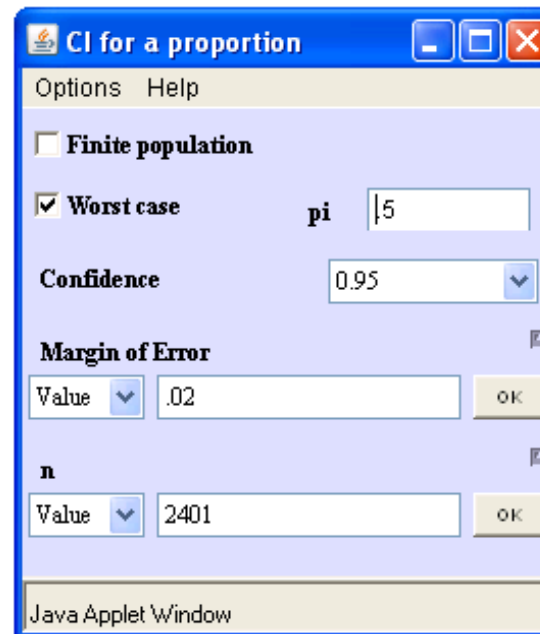
This result violates the small-sample approximation requirement that $n \ll N$, but it provides a good starting point for iterations toward a more accurate result. When n becomes a substantial fraction of N , use the method shown in Section 10.4.1.2 instead. (This example is re-solved using that method in Example 10.21.)

Example 4.3 How many people should be polled to estimate voter preference for two candidates in a close election if the poll result must be within 2% of the truth with 95% confidence?

Solution: From Equation 4.15 with confidence interval half-width $\delta = 0.02$ the required sample size is

$$n = \frac{1}{(0.02)^2} = 2500.$$

From Piface > CI for one proportion:



From MINITAB (V16) > Stat > Power and Sample Size > Sample Size for Estimation > Proportion (Binomial):

```

MTB > SSCI;
SUBC> BProportion 0.5;
SUBC> Confidence 95.0;
SUBC> IType 0;
SUBC> MError 0.02.
    
```

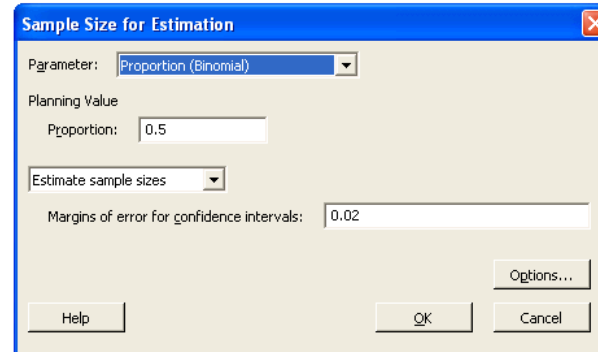
Sample Size for Estimation

Method

Parameter	Proportion
Distribution	Binomial
Proportion	0.5
Confidence level	95%
Confidence interval	Two-sided

Results

Margin of Error	Sample Size
0.02	2449



From PASS > Proportions > One Group > Confidence Interval - Proportion:

Confidence Interval of A Proportion

Numeric Results

	C.C.	N	P0
Precision	Confidence Coefficient	Sample Size	Baseline Proportion
0.02000	0.95108	2377	0.50000

References

Desu, M. M. and Raghavarao, D. 1990. Sample Size Methodology. Academic Press. New York.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.
Hahn, G. J. and Meeker, W.G. 1991. Statistical Intervals. John Wiley & Sons. New York.

Summary Statements

A sample size of 2377 produces a 95% confidence interval equal to the sample proportion plus or minus 0.02000 when the estimated proportion is 0.50000.

From MINITAB > Stat > Power and Sample Size > 1 Proportion

```

MTB > Power;
SUBC> POne;
SUBC> PAlternative 0.52;
SUBC> Power 0.5;
SUBC> PNull 0.50;
SUBC> GPCurve.

```

Power and Sample Size

Test for One Proportion

Testing proportion = 0.5 (versus not = 0.5)
Alpha = 0.05

Alternative Proportion	Sample Size	Target Power	Actual Power
0.52	2401	0.5	0.500058

Example 4.4 Find the power to reject $H_0 : p = 0.1$ when in fact $p = 0.2$ and the sample will be of size $n = 200$.

Solution: Under both H_0 and H_A the sample size is sufficiently large to justify the use of normal approximations to the binomial distributions. From Equation 4.21 with $\alpha = 0.05$ we have

$$z_\beta = \frac{\sqrt{200} |0.2 - 0.1| - z_{0.025} \sqrt{(0.1)(1 - 0.1)}}{\sqrt{(0.2)(1 - 0.2)}} = 2.066,$$

so the power is

$$\begin{aligned} \pi &= 1 - \Phi(-\infty < z < 2.066) \\ &= 0.981. \end{aligned}$$

From **Piface**> **Test of one proportion:**

Sample size for one proportion

Options Help

Null value (p_0)

Value OK

Actual value (p)

Value OK

Sample size

Value OK

Alternative **Alpha**

Method

Power

Value OK

Java Applet Window

From PASS> Proportions> One Group> Inequality [Differences]:

PASS: Proportion: Inequality [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 Background Abbreviations Template
 Plot Text Axes 3D Symbols 1
 Data Options Reports Plot Setup

Find (Solve For): Alternative Hypothesis:
 Alternative Difference (P1-P0): Null Proportion (P0):
 n (Sample Size): Alpha (Significance):
 N (Population Size): Beta (1-Power):
 Test Statistic in Reports:

If a search fails, increase the Maximum Iterations on the Options tab.

PASS: Proportion: Inequality [Differences] Output

Power Analysis of One Proportion

Numeric Results for testing H0: P = P0 versus H1: P <> P0
Test Statistic: Z Test using S(P0)

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Difference (P1 - P0)	Target Alpha	Actual Alpha	Beta	Reject H0 if Z > Then
0.9821	200	0.1000	0.2000	0.1000	0.0500	0.0439	0.0179	1.9600

References

Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Summary Statements

A sample size of 200 achieves 98% power to detect a difference (P1-P0) of 0.1000 using a two-sided Z test that uses S(P0) to estimate the standard deviation. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0439. These results assume that the population proportion under the null hypothesis is 0.1000.

From MINITAB> Stat> Power and Sample Size> 1 Proportion:

```

MTB > Power;
SUBC> POne;
SUBC> Sample 200;
SUBC> PAlternative 0.2;
SUBC> PNull 0.1;
SUBC> GPCurve.
  
```

Power and Sample Size

Test for One Proportion

Testing proportion = 0.1 (versus not = 0.1)
 Alpha = 0.05

Alternative Proportion	Sample Size	Power
0.2	200	0.980565

Power and Sample Size for 1 Proportion

Specify values for any two of the following:

Sample sizes:

Alternative values of p:

Power values:

Hypothesized p:

Example 4.5 What sample size is required to reject $H_0 : p = 0.05$ when in fact $p = 0.10$ using a two-sided test with 90% power?

Solution: Assuming that the sample size will be sufficiently large to justify the normal approximation method, from Equation 4.22 the required sample size is

$$\begin{aligned}
 n &= \left(\frac{1.96\sqrt{(0.05)(1-0.05)} + 1.282\sqrt{(0.10)(1-0.10)}}{0.10 - 0.05} \right)^2 \\
 &= 264.
 \end{aligned}$$

From Piface> Test of one proportion:

Sample size for one proportion

Options Help

Null value (p_0)

Value OK

Actual value (p)

Value OK

Sample size

Value OK

Alternative **Alpha**

Method

Power

Value OK

Java Applet Window

From PASS> Proportions> One Group> Inequality [Differences]:

PASS: Proportion: Inequality [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For): Alternative Hypothesis:

Alternative Difference (P1-P0): Null Proportion (P0):

n (Sample Size): Alpha (Significance):

N (Population Size): Beta (1-Power):

Test Statistic in Reports:

If a search fails, increase the Maximum Iterations on the Options tab.

PASS: Proportion: Inequality [Differences] Output

Power Analysis of One Proportion

Numeric Results for testing $H_0: P = P_0$ versus $H_1: P <> P_0$
Test Statistic: Z Test using S(P0)

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Difference (P1 - P0)	Target Alpha	Actual Alpha	Beta	Reject H0 if $ Z > \text{Then}$
0.9031	245	0.0500	0.1000	0.0500	0.0500	0.0555	0.0969	1.9600

References

Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Summary Statements

A sample size of 245 achieves 90% power to detect a difference (P1-P0) of 0.0500 using a two-sided Z test that uses S(P0) to estimate the standard deviation. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0555. These results assume that the population proportion under the null hypothesis is 0.0500.

From MINITAB> Power and Sample Size> 1 Proportion:

```

MTB > Power;
SUBC> POne;
SUBC> PAlternative 0.1;
SUBC> Power 0.9;
SUBC> PNull 0.05;
SUBC> GPCurve.

Power and Sample Size

Test for One Proportion

Testing proportion = 0.05 (versus not = 0.05)
Alpha = 0.05

Alternative Proportion   Sample Size   Target Power   Actual Power
0.1                      264          0.9           0.900470
    
```

Power and Sample Size for 1 Proportion

Specify values for any two of the following:

Sample sizes:

Alternative values of p:

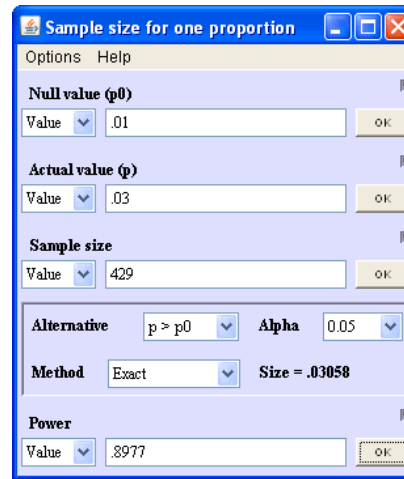
Power values:

Hypothesized p:

Example 4.6 What sample size is required to reject $H_0 : p = 0.01$ with 90% power when in fact $p = 0.03$?

Solution: The hypotheses to be tested are $H_0 : p = 0.01$ versus $H_A : p > 0.01$ and the two points on the OC curve are $(p_0, 1 - \alpha) = (0.01, 0.95)$ and $(p_1, \beta) = (0.03, 0.10)$. The exact simultaneous solution to Equations 4.24 and 4.25, obtained using Larson’s nomogram and then iterating to the exact solution using a binomial calculator, is $(n, c) = (390, 7)$. The distributions of the success counts under H_0 and H_A are shown in Figure 4.2.

From Piface> Test of one proportion:



From PASS > Proportions > One Group > Inequality [Differences]:

Power Analysis of One Proportion

Numeric Results for testing $H_0: P = P_0$ versus $H_1: P > P_0$
 Test Statistic: Exact Test

Power	N	Proportion Given H_0 (P_0)	Proportion Given H_1 (P_1)	Difference ($P_1 - P_0$)	Target Alpha	Actual Alpha	Beta	Reject H_0 if $R \geq$ This
0.9001	390	0.0100	0.0300	0.0200	0.0500	0.0445	0.0999	8

References
 Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Summary Statements
 A sample size of 390 achieves 90% power to detect a difference ($P_1 - P_0$) of 0.0200 using a one-sided binomial test. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0445. These results assume that the population proportion under the null hypothesis is 0.0100.

From MINITAB > Power and Sample Size > 1 Proportion:

```

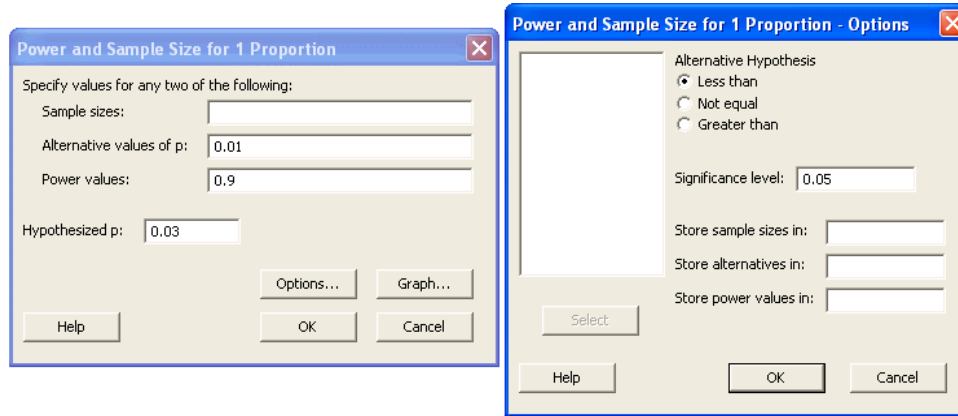
MTB > Power;
SUBC> POne;
SUBC> PAlternative 0.01;
SUBC> Power 0.9;
SUBC> PNull 0.03;
SUBC> Alternative -1;
SUBC> GPCurve.

Power and Sample Size

Test for One Proportion

Testing proportion = 0.03 (versus < 0.03)
Alpha = 0.05

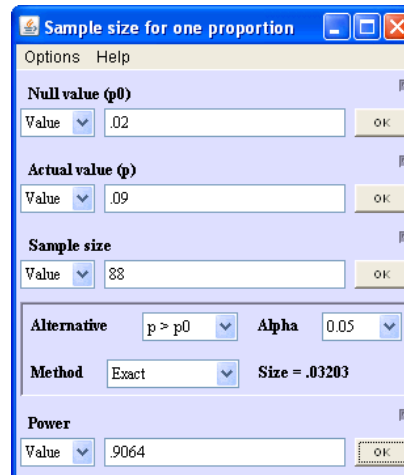
Alternative Sample Target
Proportion Size Power Actual Power
0.01 417 0.9 0.900542
    
```



Example 4.7 Use Larson’s nomogram to find n and c for the sampling plan for defectives that will accept 95% of lots with 2% defectives and 10% of lots with 8% defectives. Draw the OC curve.

Solution: Figure 4.3 shows the solution using Larson’s nomogram with the two specified points on the OC curve at $(p, P_A(H_0)) = (0.02, 0.95)$ and $(0.09, 0.10)$. The required sampling plan is $n = 100$ and $c = 4$. The OC curve is shown in Figure 4.4. Points on the OC were obtained by rocking a line about the point at $n = 100$ and $c = 4$ in the nomogram and reading off p and P_A values.

From Piface> **Test of one proportion:**



From PASS> Proportions> One Group> Inequality [Differences]:

PASS: Proportion: Inequality [Differences] Output

Power Analysis of One Proportion

Numeric Results for testing $H_0: P = P_0$ versus $H_1: P > P_0$
Test Statistic: Exact Test

Power	N	Proportion Given H_0 (P0)	Proportion Given H_1 (P1)	Difference (P1 - P0)	Target Alpha	Actual Alpha	Beta	Reject H_0 if $R \geq$ This
0.9012	87	0.0200	0.0900	0.0700	0.0500	0.0307	0.0988	5

References
 Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Summary Statements
 A sample size of 87 achieves 90% power to detect a difference (P1-P0) of 0.0700 using a one-sided binomial test. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0307. These results assume that the population proportion under the null hypothesis is 0.0200.

From MINITAB> Stat> Power and Sample Size> 1 Proportion:

```
MTB > Power;
SUBC> POne;
SUBC> PAlternative 0.09;
SUBC> Power 0.9;
SUBC> PNull 0.02;
SUBC> Alternative 1;
SUBC> GPCurve.
```

Power and Sample Size

Test for One Proportion

Testing proportion = 0.02 (versus > 0.02)
 Alpha = 0.05

Alternative Proportion	Sample Size	Target Power	Actual Power
0.09	73	0.9	0.900639

Power and Sample Size for 1 Proportion

Specify values for any two of the following:

Sample sizes:

Alternative values of p:

Power values:

Hypothesized p:

Options... Graph... Help OK Cancel

Power and Sample Size for 1 Proportion - Options

Alternative Hypothesis

Less than

Not equal

Greater than

Significance level:

Store sample sizes in:

Store alternatives in:

Store power values in:

Select Help OK Cancel

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Attributes:

Example 4.8 What sample size is required to reject $H_0 : p = 0.03$ with 90% power when in fact $p = 0.01$?

Solution: The hypotheses to be tested are $H_0 : p = p_0$ versus $H_A : p < p_0$ and the two points on the OC curve are $(p_0, 1 - \alpha) = (0.03, 0.95)$ and $(p_1, \beta) = (0.01, 0.10)$. The exact simultaneous solution to Equations 4.26 and 4.27, determined using Larson’s nomogram followed by manual iterations with a binomial calculator, is $(n, r) = (436, 7)$.

From Piface> Test of one proportion:

Acceptance Sampling by Attributes

Measurement type: Go/no go
 Lot quality in proportion defective
 Use binomial distribution to calculate probability of acceptance

Acceptable Quality Level (AQL) 0.02
 Producer's Risk (Alpha) 0.05

 Rejectable Quality Level (RQL or LTPD) 0.09
 Consumer's Risk (Beta) 0.1

Generated Plan(s)

Sample Size 87
 Acceptance Number 4

Accept lot if defective items in 87 sampled ≤ 4 ; Otherwise reject.

Proportion Defective	Probability Accepting	Probability Rejecting
0.02	0.969	0.031
0.09	0.099	0.901

Acceptance Sampling by Attributes ✖

Create a Sampling Plan Options...

Measurement type: Go / no go (defective) Graphs...

Units for quality levels: Proportion defective

Acceptable quality level (AQL): 0.02

Rejectable quality level (RQL or LTPD): 0.09

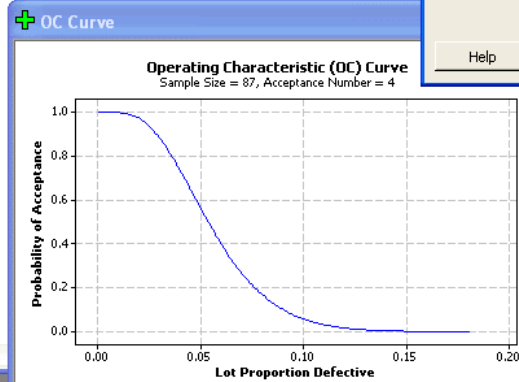
Producer's risk (Alpha): 0.05

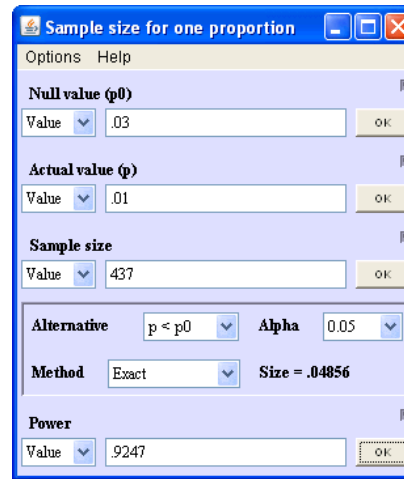
Consumer's risk (Beta): 0.10

Lot size:

OK
Cancel

Help





From PASS> Proportions> One Group> Inequality [Differences]:

Power Analysis of One Proportion

Numeric Results for testing H0: P = P0 versus H1: P < P0
Test Statistic: Exact Test

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Difference (P1 - P0)	Target Alpha	Actual Alpha	Beta	Reject H0 if R<=This
0.9255	436	0.0300	0.0100	-0.0200	0.0500	0.0493	0.0745	7

References
 Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Summary Statements
 A sample size of 436 achieves 93% power to detect a difference (P1-P0) of -0.0200 using a one-sided binomial test. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0493. These results assume that the population proportion under the null hypothesis is 0.0300.

From MINITAB> Stat> Power and Sample Size> 1 Proportion:

```

MTB > Power;
SUBC> POne;
SUBC> PAlternative 0.01;
SUBC> Power 0.9;
SUBC> PNull 0.03;
SUBC> Alternative -1;
SUBC> GPCurve.

```

Power and Sample Size

Test for One Proportion

Testing proportion = 0.03 (versus < 0.03)
Alpha = 0.05

Alternative Proportion	Sample Size	Target Power	Actual Power
0.01	417	0.9	0.900542

Power and Sample Size for 1 Proportion

Specify values for any two of the following:

Sample sizes:

Alternative values of p:

Power values:

Hypothesized p:

Options... Graph...

Help OK Cancel

Power and Sample Size for 1 Proportion - Options

Alternative Hypothesis

Less than

Not equal

Greater than

Significance level:

Store sample sizes in:

Store alternatives in:

Store power values in:

Select

Help OK Cancel

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Attributes:

Acceptance Sampling by Attributes

Measurement type: Go/no go
 Lot quality in proportion defective
 Use binomial distribution to calculate probability of acceptance

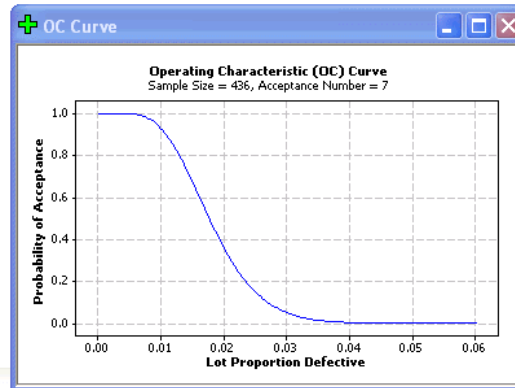
Acceptable Quality Level (AQL) 0.01
 Producer's Risk (Alpha) 0.1
 Rejectable Quality Level (RQL or LTPD) 0.03
 Consumer's Risk (Beta) 0.05

Generated Plan(s)

Sample Size 436
 Acceptance Number 7

Accept lot if defective items in 436 sampled <= 7; Otherwise reject.

Proportion Defective	Probability Accepting	Probability Rejecting
0.01	0.926	0.074
0.03	0.049	0.951



Acceptance Sampling by Attributes

Create a Sampling Plan

Measurement type: Go / no go (defective)

Units for quality levels: Proportion defective

Acceptable quality level (AQL): 0.01
 Rejectable quality level (RQL or LTPD): 0.03

Producer's risk (Alpha): 0.10
 Consumer's risk (Beta): 0.05

Lot size:

OK
 Cancel
 Help

4.2 One Proportion (Small Population)

Example 4.9 Suppose that a sample of size $n = 20$ drawn from a population of $N = 100$ units was found to have $X = 2$ defective units. Determine the one-sided upper confidence limit for the population fraction defective.

Solution: From the following hypergeometric probabilities:

$$h(0 \leq x \leq 2; S = 26, N = 100, n = 20) = 0.0555$$

$$h(0 \leq x \leq 2; S = 27, N = 100, n = 20) = 0.0448$$

the smallest value of S that satisfies the inequality in Equation 4.34 is $S = 27$, so the 95% one-sided upper confidence limit for S is $S_U = 27$ or

$$P(S \leq 27) \geq 0.95.$$

Example 4.10 A hospital is asked by an auditor to confirm that its billing error rate is less than 10% for a day chosen randomly by the auditor. However, it is impractical to inspect all 120 bills issued on that day. How many of the bills must be inspected to demonstrate, with 95% confidence, that the billing error rate is less than 10%?

Solution: The goal of the analysis is to demonstrate that the one-sided upper 95% confidence limit on the billing error rate p is 10% or

$$P(p \leq 0.10) = 0.95.$$

Under the assumption that the auditor will accept a zero defectives sampling plan, by the rule of three (Equation 4.5) the approximate sample size must be

$$n \simeq \frac{3}{p} = \frac{3}{0.10} = 30.$$

Because $n = 30$ is large compared to $N = 120$, the finite population correction factor (Equation 4.16) should be used and gives

$$\begin{aligned} n' &= \frac{30}{1 + \frac{30-1}{120}} \\ &= 25. \end{aligned}$$

Iterations with a hypergeometric probability calculator show that $n = 26$ is the smallest sample size that gives 95% confidence that the billing error rate is less than 10%.

Example 4.11 What sample size n must be drawn from a population of size $N = 200$ and found to be free of defectives if we need to demonstrate, with 95% confidence, that there are no more than four defectives in the population?

Solution: The goal of the experiment is to demonstrate the confidence interval

$$P(0 \leq S \leq 4) \geq 0.95$$

using a zero-successes ($X = 0$) sampling plan. By the small-sample binomial approximation with $S_U = 4$ and $\alpha = 0.05$, the required sample size by Equation 4.40 is given by

$$n = \frac{\ln(0.05)}{\ln\left(1 - \frac{4}{200}\right)} = 149,$$

which violates the small-sample assumption. By Equation 4.42, the rare-event binomial approximation gives

$$\begin{aligned} n &\geq N \left(1 - \alpha^{1/S_U}\right) \\ &\geq 200 \left(1 - 0.05^{1/4}\right) \\ &\geq 106. \end{aligned}$$

This solution meets the requirements of the rare-event approximation method, but just to check this result, the corresponding exact hypergeometric probability is $h(0; 4, 200, 106) = 0.047$ which is less than $\alpha = 0.05$ as required, however, because $h(0, 4, 200, 105) = 0.049$, the sample size $n = 105$ is the exact solution to the problem.

Example 4.12 A biologist needs to test the fraction of female frogs in a single brood, but the sex of the frog tadpoles is difficult to determine. The hypotheses to be tested are $H_0 : p = 0.5$ versus $H_A : p > 0.5$ where p is the fraction of the frogs that are female. If there are $N = 212$ viable frogs in the brood, how many of them must she sample to reject H_0 with 90% power when $p = 0.65$?

Solution: The exact sample size (n) and acceptance number (c) have to be determined by iteration. The approximate sample size given by the large-sample binomial approximation method in Equation 4.22 with $p_0 = 0.5$, $\alpha = 0.05$, $p_1 = 0.65$, and $\beta = 0.10$ is

$$n = \left(\frac{1.645\sqrt{0.5(1-0.5)} + 1.282\sqrt{0.65(1-0.65)}}{0.65 - 0.5} \right)^2 = 92.$$

However, this sample size is large compared to the population size, so the finite population correction factor (Equation 4.16) must be used, which gives

$$\begin{aligned} n' &= \frac{92}{1 + \frac{92-1}{212}} \\ &= 65. \end{aligned}$$

The exact values of n and c are determined from the simultaneous solution of Equations 4.43 and Equation 4.44 with $S_0 = Np_0 = 106$ and $S_1 = Np_1 = 138$, which gives

$$\begin{aligned} \sum_{x=0}^c h(x; S = 106, N = 212, n) &\geq 0.95 \\ \sum_{x=0}^c h(x; S = 138, N = 212, n) &\leq 0.10. \end{aligned}$$

Using a hypergeometric calculator with $n = 65$ we find

$$\begin{aligned} \sum_{x=0}^{38} h(x; S = 106, N = 212, n = 65) &= 0.963 \\ \sum_{x=0}^{38} h(x; S = 138, N = 212, n = 65) &= 0.117, \end{aligned}$$

which satisfies the $1 - \alpha \geq 0.95$ requirement but does not satisfy the $\beta \leq 0.10$ requirement. A few more iterations determine that $n = 69$ and $c = 40$ gives $\alpha = 0.039$ and $\pi = 0.912$ which meets both requirements. This means that the biologist must sample $n = 69$ frogs and can reject H_0 if $S > 40$.

The one proportion methods in Piface and MINITAB can be used to find the first step in the solution, $n = 92$, but they don't provide the opportunity to apply a small population correction. PASS does support finite populations using the binomial method instead of the normal approximation. From **PASS> Proportions> One Group> Inequality [Differences]:**

Power Analysis of One Proportion

Numeric Results for testing H0: P = P0 versus H1: P > P0. N = 212.
Test Statistic: Exact Test

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Difference (P1 - P0)	Target Alpha	Actual Alpha	Beta	Reject H0 if R >= This
0.9036	64	0.5000	0.6500	0.1500	0.0500	0.0497	0.0964	38

References
 Chow, S.C., Shao, J., Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 Fleiss, J.L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Summary Statements
 A sample size of 64 from a population of 212 achieves 90% power to detect a difference (P1-P0) of 0.1500 using a one-sided binomial test. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0497. These results assume that the population proportion under the null hypothesis is 0.5000.

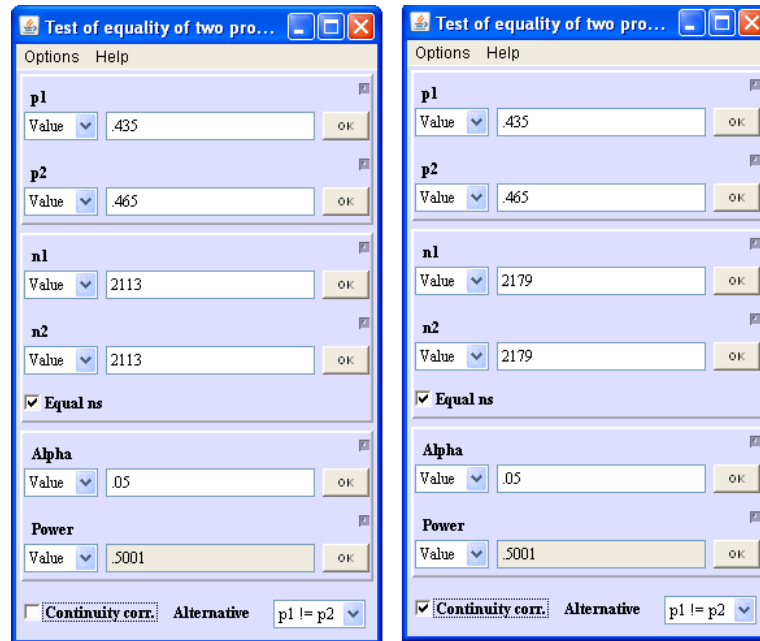
4.3 Two Proportions

Example 4.13 Determine the sample size required to estimate the difference between two proportions to within 0.03 with 95% confidence if both proportions are expected to be about 0.45. Assume that the two sample sizes will be equal.

Solution: From Equation 4.54 with $\delta = 0.03$, $\bar{p} = 0.45$, $n_1/n_2 = 1$, and $\alpha = 0.05$, the required sample size is

$$\begin{aligned} n_1 &= n_2 = \left(\frac{1.96}{0.03} \right)^2 (2 \times 0.45 \times (1 - 0.45)) \\ &= 2113. \end{aligned}$$

From **Piface** > **Test comparing two proportions** without and with the continuity correction:



From MINITAB> Stat> Power and Sample Size> Two Proportions:

```
MTB > Power;
SUBC> PTwo;
SUBC> PrOne 0.435;
SUBC> Power 0.5;
SUBC> PrTwo 0.465;
SUBC> GPCurve.
```

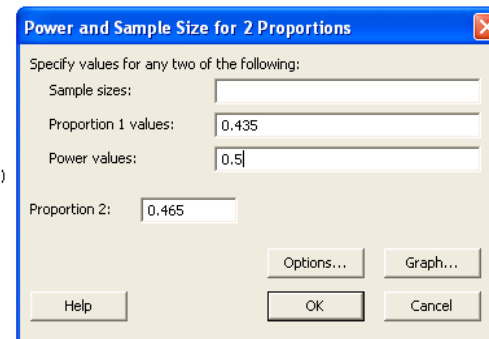
Power and Sample Size

Test for Two Proportions

Testing proportion 1 = proportion 2 (versus not =)
 Calculating power for proportion 2 = 0.465
 Alpha = 0.05

Proportion 1	Sample Size	Target Power	Actual Power
0.435	2113	0.5	0.500081

The sample size is for each group.



From PASS> Proportions> Two Groups: Independent> Inequality [Differences]:

PASS: Proportions: Inequality [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Abbreviations | Axes | 3D | Symbols 1 | Symbols 2 | Template

Data | Options | Reports | Plot Setup | Plot Text

Find (Solve For): N1

Alternative Hypothesis (H1): Two-Sided

Test Statistic: Z Test (Pooled)

D1 (Difference|H1 = P1 - P2): 0.03

P2 (Control Group Proportion): P

P1 is the rate in the Treatment group.
P2 is the rate in the Control group.
P1 = P2 + D1

N1 (Sample Size Group 1): 50 to 400 by 50

N2 (Sample Size Group 2): Use R

R (Sample Allocation Ratio): 1.0

Alpha (Significance Level): 0.05

Beta (1-Power): 0.5

PASS: Proportions: Inequality [Differences] Output

Two Independent Proportions (Null Case) Power Analysis

Numeric Results of Tests Based on the Difference: P1 - P2
H0: P1-P2=0. H1: P1-P2=D1<>0. Test Statistic: Z test with pooled variance

	Sample Size Grp 1	Sample Size Grp 2	Prop H1 Grp 1 or Trtmt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
Power	2113	2113	0.4650	0.4350	0.0000	0.0300	0.0500	0.4999	0.4999

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.

References

Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 D'Agostino, R.B., Chase, W., Belanger, A. 1988. 'The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations', The American Statistician, August 1988, Volume 42 Number 3, pages 198-202.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.

Example 4.14 An experiment is planned to estimate the risk ratio. The two proportions are expected to be $p_1 \simeq 0.2$ and $p_2 \simeq 0.05$. Determine the optimal allocation ratio and the sample size required to determine the risk ratio to within 20% of its true value with 95% confidence?

Solution: A 95% confidence interval for the risk ratio is required of the form in Equation 4.56. With $p_1 = 0.2$ and $p_2 = 0.05$, the anticipated value of the risk ratio is $RR \simeq 0.2/0.05 = 4$ and from Equation 4.62 the optimal sample size allocation ratio is

$$\frac{n_1}{n_2} = \sqrt{\frac{0.05/0.95}{0.2/0.8}} = 0.4588.$$

Then with $\delta = 0.2$ and $\alpha = 0.05$ in Equation 4.61, the required sample size n_1 is

$$\begin{aligned} n_1 &= \left(\frac{1.96}{0.2}\right)^2 \left(\frac{1-0.2}{0.2} + \frac{1-0.05}{0.05} (0.4588)\right) \\ &= 1222 \end{aligned}$$

and the sample size n_2 is

$$n_2 = \frac{n_1}{\left(\frac{n_1}{n_2}\right)} = \frac{1222}{0.4588} = 2664.$$

These sample sizes minimize the total number of samples required for the experiment.

Example 4.15 An experiment is planned to estimate the odds ratio. The two proportions are expected to be $p_1 \simeq 0.5$ and $p_2 \simeq 0.25$. Determine the optimal allocation ratio and the sample size required to determine, with 90% confidence, the odds ratio to within 20% of its true value?

Solution: The desired confidence interval has the form given by Equation 4.64 with $\delta = 0.2$. With $p_1 = 0.5$ and $p_2 = 0.25$, the anticipated value of the odds ratio is $OR = \frac{0.5/0.5}{0.25/0.75} = 3$ and from Equation 4.70 the optimal sample size allocation ratio is

$$\frac{n_1}{n_2} = \sqrt{\frac{0.25 \times 0.75}{0.5 \times 0.5}} = 0.866.$$

Then with $\delta = 0.2$ and $\alpha = 0.10$ in Equation 4.69, the required sample size n_1 is

$$\begin{aligned} n_1 &= \left(\frac{1.645}{0.2}\right)^2 \left(\frac{1}{0.5 \times 0.5} + \frac{1}{0.25 \times 0.75} (0.866)\right) \\ &= 584 \end{aligned}$$

and the sample size n_2 is

$$n_2 = \frac{n_1}{\left(\frac{n_1}{n_2}\right)} = \frac{584}{0.866} = 675.$$

Example 4.16 Determine the power for Fisher's test to reject $H_0 : p_1 = p_2$ in favor of $H_A : p_1 < p_2$ when $p_1 = 0.01$, $p_2 = 0.50$, and $n_1 = n_2 = 8$.

Solution: The Fisher's test p values for all possible combinations of x_1 and x_2 were calculated using Equation 4.71 and are shown in Table 4.3. The few cases that are statistically significant, where $p \leq 0.05$, are shown in a bold font in the upper right corner of the table. Table 4.4 shows the contributions to the power given by the product of the two binomial probabilities in Equation 4.74. The sum of the individual contributions, that is, the power of Fisher's test, is $\pi = 0.60$.

From **PASS> Proportions> Two Groups: Independent> Inequality [Differences]:**

PASS: Proportions: Inequality [Proportions]

File Run Analysis Graphics PASS GESS Tools Window Help

Abbreviations | 3D | Symbols 1 | Symbols 2 | Backgro

Data | Options | Reports | Plot Setup | Plot Tex

Find (Solve For): Alternative Hypothesis (H1):
 Beta and Power One-Sided (H1:P1<P2)
 Test Statistic: Fisher's Exact Test
 N1 (Sample Size Group 1): 8
 P1 (Group 1 Proportion |H1): 0.01
 N2 (Sample Size Group 2): Use R
 P2 (Control Group Proportion): P
 R (Sample Allocation Ratio): 1.0
 Alpha (Significance Level): 0.05
 Beta (1-Power):

PASS: Proportions: Inequality [Proportions] Output

Two Independent Proportions (Null Case) Power Analysis

Numeric Results of Tests Based on the Difference: P1 - P2
 H0: P1-P2>=0. H1: P1-P2=D1<0. Test Statistic: Fisher's Exact test

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop H1 Grp 1 or Trtmnt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.5984	8	8	0.0100	0.5000	0.0000	-0.4900	0.0500		0.4016

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.

References
 Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 D'Agostino, R.B., Chase, W., Belanger, A. 1988. The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations', The American Statistician, August 1988, Volume 42 Number 3, pages 198-202.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.

MINITAB's default sample size and power calculator uses the normal approximation but the Fisher's test power can be calculated using the custom MINITAB macro *fisherspower.mac* that is posted on www.mmbstatistical.com.

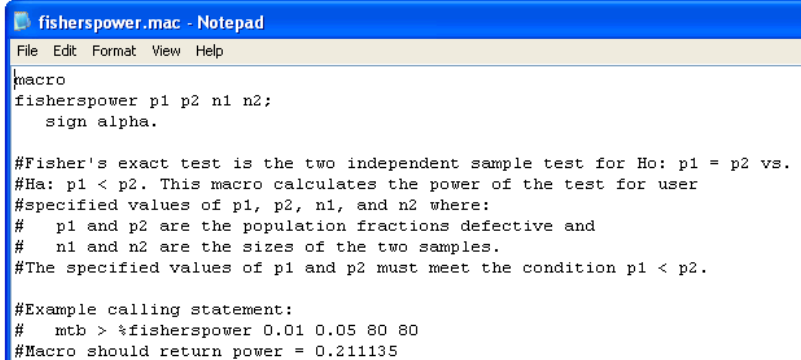
```

MTB > %fisherspower 0.01 0.50 8 8
Executing from file: C:\Program Files\Minitab 15\English\Macros\fisherspower.MAC

power    0.598399

MTB >

```



```

fisherspower.mac - Notepad
File Edit Format View Help
macro
fisherspower p1 p2 n1 n2;
    sign alpha.

#Fisher's exact test is the two independent sample test for Ho: p1 = p2 vs.
#Ha: p1 < p2. This macro calculates the power of the test for user
#specified values of p1, p2, n1, and n2 where:
# p1 and p2 are the population fractions defective and
# n1 and n2 are the sizes of the two samples.
#The specified values of p1 and p2 must meet the condition p1 < p2.

#Example calling statement:
# mtb > %fisherspower 0.01 0.05 80 80
#Macro should return power = 0.211135

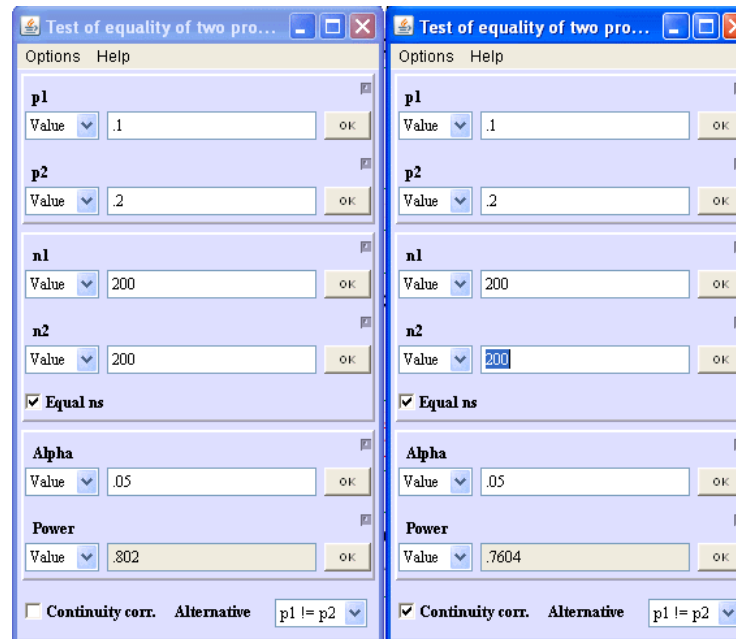
```

Example 4.17 Determine the power for the test of $H_0 : p_1 = p_2$ versus $H_A : p_1 \neq p_2$ when $n_1 = n_2 = 200$, $p_1 = 0.10$, and $p_2 = 0.20$. Use a two-tailed test with $\alpha = 0.05$

Solution: The normal approximation to the binomial distribution is justified for both samples, so with $\hat{p} = 0.15$ and $\Delta\hat{p} = 0.10$ in Equations 4.78 and 4.79, the power is

$$\begin{aligned}
 \pi &= \Phi\left(-\infty < z < \frac{0.10}{\sqrt{\frac{2(0.15)(1-0.15)}{200}}} - 1.96\right) \\
 &= \Phi(-\infty < z < 0.84) \\
 &= 0.80.
 \end{aligned}$$

From **Piface**> **Test comparing two proportions** without and with the continuity correction:



From MINITAB> Stat> Power and Sample Size> Two Proportions:

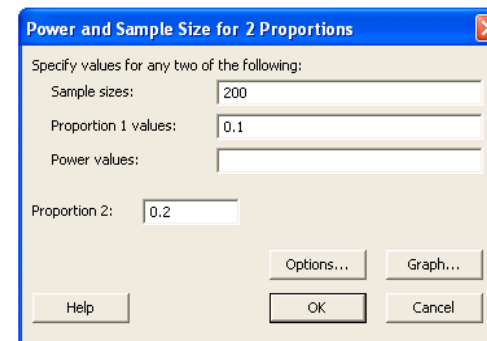
Power and Sample Size

Test for Two Proportions

Testing proportion 1 = proportion 2 (versus not =)
 Calculating power for proportion 2 = 0.2
 Alpha = 0.05

Proportion 1	Sample Size	Power
0.1	200	0.802049

The sample size is for each group.



From PASS> Proportions> Two Groups: Independent> Inequality [Differences]:

PASS: Proportions: Inequality [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Abbreviations | Axes | 3D | Symbols 1 | Symbols 2 | Background | Data | Options | Reports | Plot Setup | Plot Text

Find (Solve For): Beta and Power
 Alternative Hypothesis (H1): Two-Sided
 Test Statistic: Z Test (Pooled)
 N1 (Sample Size Group 1): 200
 D1 (Difference)|H1 = P1 - P2): -0.1
 N2 (Sample Size Group 2): Use R
 P2 (Control Group Proportion): P
 R (Sample Allocation Ratio): 1.0
 Alpha (Significance Level): 0.05
 Beta (1-Power):

P1 is the rate in the Treatment group.
 P2 is the rate in the Control group.
 P1 = P2 + D1

PASS: Proportions: Inequality [Differences] Output

Two Independent Proportions (Null Case) Power Analysis

Numeric Results of Tests Based on the Difference: P1 - P2
 H0: P1-P2=0. H1: P1-P2=D1<>0. Test Statistic: Z test with pooled variance

	Sample Size Grp 1	Sample Size Grp 2	PropH1 Grp 1 or Trtmnt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
Power	200	200	0.1000	0.2000	0.0000	-0.1000	0.0500		0.1980

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.

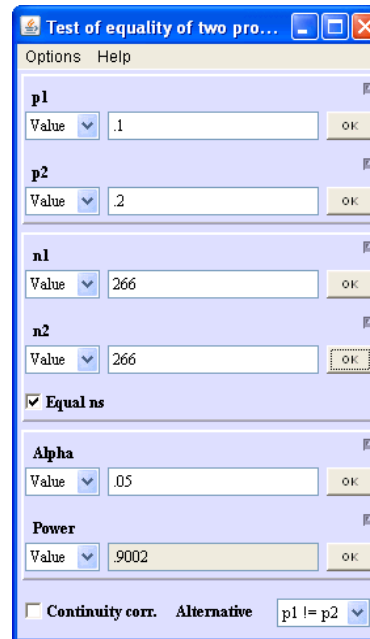
References
 Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 D'Agostino, R.B., Chase, W., Belanger, A. 1988. The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations', The American Statistician, August 1988, Volume 42 Number 3, pages 198-202.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
 Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.

Example 4.18 What common sample size is required to resolve the difference between two proportions with 90% power using a two-sided test when $p_1 = 0.10$ and $p_2 = 0.20$ is expected?

Solution: From Equation 4.80 with $\hat{p} = 0.15$ and $\Delta\hat{p} = 0.10$ the required sample size is

$$\begin{aligned} n &= \frac{2 \times 0.15 \times 0.85}{(0.10)^2} (1.28 + 1.96)^2 \\ &= 268. \end{aligned}$$

From Piface> Test comparing two proportions without the continuity correction:



From MINITAB> Stat> Power and Sample Size> Two Proportions:

```
MTB > Power;
SUBC> PTwo;
SUBC> PrOne 0.1;
SUBC> Power 0.90;
SUBC> PrTwo 0.2;
SUBC> GPCurve.
```

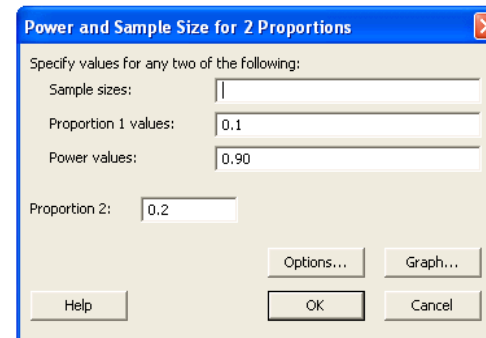
Power and Sample Size

Test for Two Proportions

Testing proportion 1 = proportion 2 (versus not =)
 Calculating power for proportion 2 = 0.2
 Alpha = 0.05

Proportion 1	Sample Size	Target Power	Actual Power
0.1	266	0.9	0.900155

The sample size is for each group.



From PASS> Proportions> Two Groups: Independent> Inequality [Differences]:

PASS: Proportions: Inequality [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Abbreviations | 3D | Symbols 1 | Symbols 2 | Temp | Backgro

Data | Options | Reports | Plot Setup | Plot Tex

Find (Solve For): N1
Alternative Hypothesis (H1): Two-Sided
Test Statistic: Z Test (Pooled)
D1 (Difference|H1 = P1 - P2): -0.1
N1 (Sample Size Group 1):
N2 (Sample Size Group 2): Use R
P2 (Control Group Proportion): P
R (Sample Allocation Ratio): 1.0
Alpha (Significance Level): 0.05
Beta (1-Power): 0.10
P1 is the rate in the Treatment group.
P2 is the rate in the Control group.
P1 = P2 + D1

PASS: Proportions: Inequality [Differences] Output

Two Independent Proportions (Null Case) Power Analysis

Numeric Results of Tests Based on the Difference: P1 - P2
H0: P1-P2=0. H1: P1-P2=D1<->0. Test Statistic: Z test with pooled variance

	Sample Size Grp 1	Sample Size Grp 2	Prop H1 or Trtmnt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
Power	N1	N2							
0.9002	266	266	0.1000	0.2000	0.0000	-0.1000	0.0500		0.0998

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.

References
Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
D'Agostino, R.B., Chase, W., Belanger, A. 1988. The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations', The American Statistician, August 1988, Volume 42 Number 3, pages 198-202.
Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.

Example 4.19 Repeat the calculation of the sample size for Example 4.18.

Solution: From Equation 4.86 the required sample size is

$$n = \frac{1}{2} \left(\frac{1.28 + 1.96}{\arcsin\sqrt{0.10} - \arcsin\sqrt{0.20}} \right)^2 = 261,$$

which is in excellent agreement with the sample size determined by the normal approximation method.

Example 4.20 Repeat the calculation of the sample size for Example 4.18 using the log risk ratio method.

Solution: With $RR = 0.1/0.2 = 0.5$, $\alpha = 0.05$, $\beta = 0.10$, and $n_1/n_2 = 1$ in Equation 4.90, the required sample size is

$$n_1 = n_2 = \left(\frac{1.96 + 1.282}{\ln(0.5)} \right)^2 \left(\frac{1 - 0.1}{0.1} + \frac{1 - 0.2}{0.2} \right) = 285,$$

which is in excellent agreement with the sample size obtained by the normal approximation method.

See **PASS > Proportions > Two Groups: Independent > Inequality [Ratios]**.

Example 4.21 Repeat the calculation of the sample size for Example 4.18 using the log odds ratio method.

Solution: With $n_1/n_2 = 1$ in Equation 4.97, the required common sample size is

$$n = \left(\frac{1.96 + 1.282}{\ln\left(\frac{0.10/0.90}{0.20/0.80}\right)} \right)^2 \left(\frac{1}{0.10(0.90)} + \frac{1}{0.20(0.80)} \right) = 278,$$

which is in excellent agreement with the sample size obtained by the normal approximation method.

See **PASS> Proportions> Two Groups: Independent> Inequality [Odds Ratios]**.

Example 4.22 Determine the number of subjects required for McNemar's test to reject $H_0 : RR = 1$ with 90% power when $RR = 2$ and the rate of discordant observations is estimated to be $p_D = 0.2$ from a preliminary study.

Solution: With $\beta = 0.10$ and $\alpha = 0.05$ in Equation 4.105, the approximate number of subjects required for the study is

$$\begin{aligned} \sum_i \sum_j \hat{f}_{ij} &\approx \frac{(1.282 + 1.96)^2}{0.20} \left(\frac{2 + 1}{2 - 1} \right)^2 \\ &\approx 473. \end{aligned}$$

From **PASS> Proportions> Two Groups: Paired or Correlated> Inequality (McNemar) [Odds Ratios]**:

The screenshot shows the PASS software interface for McNemar Test Power Analysis. The left window displays the input parameters, and the right window displays the output results.

McNemar Test Power Analysis								
Numeric Results for Two-Sided Test								
Power	N	P10	P01	Difference (P10-P01)	Proportion Discordant	Odds Ratio	Alpha	Beta
0.90051	489	0.133	0.067	0.067	0.200	2.000	0.05000	0.09949

References

Schork, M. and Williams, G. 1980. 'Number of Observations Required for the Comparison of Two Correlated Proportions.' Communications in Statistics-Simula. Computa., B9(4), 349-357.

Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.

Summary Statements

A sample size of 489 pairs achieves 90% power to detect an odds ratio of 2.000 using a two-sided McNemar test with a significance level of 0.05000. The odds ratio is equivalent to a difference between two paired proportions of 0.067 which occurs when the proportion in cell 1,2 is 0.133 and the proportion in cell 2,1 is 0.067. The proportion of discordant pairs is 0.200.

Example 4.23 Determine the McNemar's test power to reject $H_0 : RR = 1$ in favor of $H_A : RR \neq 1$ for a study with 200 subjects when in fact $RR = 3$ using $p_D = 0.3$.

Solution: With 200 subjects in the study, the expected number of discordant pairs is

$$\hat{f}_{12} + \hat{f}_{21} = p_D \sum_i \sum_j \hat{f}_{ij} = 0.3 \times 200 = 60.$$

Under H_A with $RR = 3$, we have $\hat{f}_{21} = 15$ and $\hat{f}_{12} = 45$, so the expected value of the McNemar's z_M statistic is

$$z_M = \frac{|45 - 15|}{\sqrt{45 + 15}} = 3.87$$

and the approximate power is

$$\begin{aligned} \pi &= P(-\infty < z < z_\beta) \\ &= P(-\infty < z < (z_M - z_{\alpha/2})) \\ &= P(-\infty < z < (3.87 - 1.96)) \\ &= P(-\infty < z < 1.91) \\ &= 0.972. \end{aligned}$$

From PASS> Proportions> Two Groups: Paired or Correlated> Inequality (McNemar) [Odds Ratios]:

The screenshot shows the PASS software interface for a McNemar Test Power Analysis. The left pane shows the input parameters, and the right pane shows the results and definitions.

Input Parameters (Left Pane):

- Find (Solve For): Beta and Power
- Alternative Hypothesis: Two-Sided
- Odds Ratio (P10/P01): 3
- N (Number of Pairs): 200
- Proportion Discordant (P10+P01): 0.3
- Alpha (Significance Level): 0.05
- Use Approximations if N >: 1500
- Beta (1 - Power):

Results (Right Pane):

McNemar Test Power Analysis

Numeric Results for Two-Sided Test

Power	N	P10	P01	Difference (P10-P01)	Proportion Discordant	Odds Ratio	Alpha	Beta
0.97407	200	0.225	0.075	0.150	0.300	3.000	0.05000	0.02593

Report Definitions:

- Power is the probability of rejecting a false null hypothesis. It should be close to one.
- N is the number of pairs in the sample.
- P10 is the proportion of pairs in cell 1,2 of the 2x2 table.
- P01 is the proportion of pairs in cell 2,1 of the 2x2 table.
- Difference is the difference between proportions parameter under the alternative hypothesis.
- Proportion Discordant is the total of P10 and P01.
- Odds Ratio is the value of this parameter under the alternative hypothesis.
- Alpha is the probability of rejecting a true null hypothesis. It should be small.
- Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements:

A sample size of 200 pairs achieves 97% power to detect an odds ratio of 3.000 using a two-sided McNemar test with a significance level of 0.05000. The odds ratio is equivalent to a difference between two paired proportions of 0.150 which occurs when the proportion in cell 1,2 is 0.225 and the proportion in cell 2,1 is 0.075. The proportion of discordant pairs is 0.300.

Table at the bottom of the left pane:

TREATMENT	STANDARD			OR = P10 / P01
	Yes	No	Total	
Yes	P11	P10	Pt	
No	P01	P00	1-Pt	
Total	Ps	1-Ps	1	

4.4 Equivalence Tests

Example 4.24 For the test of $H_0 : p < 0.45$ or $p > 0.55$ versus $H_A : 0.45 < p < 0.55$, calculate the exact and approximate power when $p = 0.5$ assuming that the sample size is $n = 800$ and $\alpha = 0.05$.

Solution: The value of x_1 , determined from Equation 4.106, is $x_1 = 384$ because

$$\sum_{x=0}^{383} b(x; n = 800, p_1 = 0.45) = 0.952.$$

The value of x_2 , determined from Equation 4.107, is $x_2 = 416$ because

$$\sum_{x=0}^{416} b(x; n = 800, p_2 = 0.55) = 0.048.$$

4.4. Equivalence Tests

Then the power when $p = 0.5$ is given by Equation 4.108:

$$\begin{aligned}\pi &= \sum_{x=383}^{416} b(x; n = 800, p_2 = 0.50) \\ &= 0.757.\end{aligned}$$

The approximate power by the normal approximation method, given by Equation 4.111, is

$$\begin{aligned}\pi &= \Phi\left(\frac{0.45 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{800}}} + 1.645 < z < \frac{0.55 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{800}}} - 1.645\right) \\ &= \Phi(-1.183 < z < 1.183) \\ &= 0.763,\end{aligned}$$

which is in good agreement with the exact solution.

From **PASS> Proportions> One Group: Equivalence [Differences]**:

PASS: Proportion: Equivalence [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For):
 Beta and Power

Baseline Proportion (PB):

Actual Difference (d1):

n (Sample Size):

Equivalence Difference (d0):

Alpha (Significance):

Beta (1-Power):

Test Statistic in Reports:

If a search fails, increase the Maximum Iterations on the Options tab.
 The test statistic used is specified under the Reports tab.

PASS: Proportion: Equivalence [Differences] Output

Power Analysis of One Proportion Equivalence

Numeric Results for testing H_0 : Non-Equivalence versus H_1 : Equivalence
 Test Statistic: Exact Test

Power	N	Equiv. Diff. (d0)	Lower Equiv. Prop. (POL)	Upper Equiv. Prop. (POU)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 if R1<=R<=R2 (R1 R2)
0.7567	800	0.0500	0.4500	0.5500	0.0000	0.5000	0.0500	0.0476	0.2433	384 416

Report Definitions

Power is the probability of concluding equivalence when the proportions are equivalent.
 N is the size of the sample drawn from the population.
 The equivalence difference is the maximum value of the difference that is still considered unimportant.
 The actual difference is the value of the difference under the alternative hypothesis.
 PB is the baseline or standard value of the proportion. This is the value under the current treatment.
 POL and POU are the limits between which an equivalent proportion must fall.
 d0 is the smallest absolute difference that is still considered equivalent.
 d1 is the value of the difference under the alternative hypothesis.
 Alpha is the probability of concluding equivalence when the proportions are non-equivalent.
 Beta is the probability concluding non-equivalence when the proportions are equivalent.

Summary Statements

A sample size of 800 achieves 76% power to detect a difference (PO-PB) of 0.0500 using a two-sided binomial test. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0476. These results assume a baseline proportion (PB) of 0.5000 and that the actual difference (P1-PB) is 0.0000.

Example 4.25 An experiment is to be performed to test the hypotheses $H_0 : p_1 \neq p_2$ versus $H_A : p_1 = p_2$. The two proportions are expected to be $p \simeq 0.12$ and the limit of practical equivalence is $\delta = 0.02$. What sample size is required to reject H_0 when $p_1 = p_2$ with 80% power?

Solution: With $\alpha = 0.05$ and $n_1/n_2 = 1$ in Equation 4.118, the sample size $n = n_1 = n_2$ is

$$\begin{aligned} n &= 2(z_{0.05} + z_{0.10})^2 \frac{0.12(1 - 0.12)}{(0.02)^2} \\ &= 2(1.645 + 1.282)^2 \frac{0.12(1 - 0.12)}{(0.02)^2} \\ &= 4524. \end{aligned}$$

From PASS> Proportions> Two Groups: Independent> Equivalence [Differences]:

PASS: Proportions: Equivalence [Differences]

File Run Analysis Graphics PASS GESS Tools Window Help

Abbreviations | 3D | Symbols 1 | Symbols 2 | Background | Template

AXES | Options | Reports | Plot Setup | Plot Text

Find (Solve For): N1

Test Statistic: Z Test (Pooled) N1 (Sample Size Group 1):

D0.U (Upper Equiv. Difference): 0.02 N2 (Sample Size Group 2): Use R

D0.L (Lower Equiv. Difference): -0.02 R (Sample Allocation Ratio): 1.0

D1 (Actual Difference): 0 Alpha (Significance Level): .05

P2 (Reference Group Proportion): 0.12 Beta (1-Power): .20

Group 1 is the Treatment or Experimental group.
Group 2 is the Control, Baseline, Standard, or Reference group.

PASS: Proportions: Equivalence [Differences] Output

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2<D0.U.
Test Statistic: Z test (pooled)

	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Target Alpha	Actual Alpha
Power	4522	4522	0.1200	0.1000	0.1400	-0.0200	0.0200	0.0000	0.0500	

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.

Report Definitions
 Power is the probability of concluding equivalence when equivalence is correct.
 Beta is the probability of accepting a false H0. Beta = 1 - Power.
 N1 and N2 are the sizes of the samples drawn from the corresponding groups.
 P2 is the response rate for group two which is the standard, reference, baseline, or control group.
 P1.0L is the smallest treatment-group response rate that still yields an equivalence conclusion.
 P1.0U is the largest treatment-group response rate that still yields an equivalence conclusion.
 D0.L is the lowest difference that still results in the conclusion of equivalence.
 D0.U is the highest difference that still results in the conclusion of equivalence.
 D1 is the actual difference, P1-P2, at which the power is calculated.
 'Target Alpha' is the probability of rejecting a true null hypothesis that was desired.
 'Actual Alpha' is the value of alpha that is actually achieved. Only available for exact results.
 'Grp 1' refers to Group 1 which is the treatment or experimental group.
 'Grp 2' refers to Group 2 which is the reference, standard, or control group.
 'Equiv.' refers to a small amount that is not of practical importance.
 Actual refers to the true value at which the power is computed.

Summary Statements
 Sample sizes of 4522 in the treatment group and 4522 in the reference group achieve 80% power to detect equivalence. The margin of equivalence, given in terms of the difference, extends from -0.0200 to 0.0200. The actual difference is 0.0000. The reference group proportion is 0.1200. The calculations assume that two, one-sided pooled Z tests are used. The significance level of the test is 0.0500.

Example 4.26 What sample size is required if the true difference between the two proportions in Example 4.254.25 is $\Delta p = 0.01$?

Solution: p_1 and p_2 are not specified, but they are both approximately $p = 0.12$, so from Equation 4.119 the sample size must be

$$\begin{aligned}
 n_1 &\simeq 2(z_\alpha + z_\beta)^2 \frac{p(1-p)}{(\delta - |\Delta p|)^2} \\
 &\simeq 2(1.645 + 0.842)^2 \frac{0.12(1-0.12)}{(0.02 - 0.01)^2} \\
 &\simeq 13063.
 \end{aligned}$$

From PASS> Proportions> Two Groups: Independent> Equivalence [Differences]:

The screenshot displays the PASS software interface. The main window is titled "PASS: Proportions: Equivalence [Differences]" and shows the "Find (Solve For):" field set to "N1". The "Test Statistic:" is "Z Test (Pooled)". The "D0.U (Upper Equiv. Difference):" is "0.02", "D0.L (Lower Equiv. Difference):" is "-0.02", "D1 (Actual Difference):" is "0.01", and "P2 (Reference Group Proportion):" is "0.12". The "N1 (Sample Size Group 1):" is "13524", "N2 (Sample Size Group 2):" is "13524", "R (Sample Allocation Ratio):" is "1.0", and "Alpha (Significance Level):" is ".05".

The output window is titled "PASS: Proportions: Equivalence [Differences] Output" and shows the "Power Analysis of Equivalence Tests of Two Independent Proportions". The "Numeric Results for Equivalence Tests Based on the Difference: P1 - P2" are as follows:

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Target Alpha	Actual Alpha
0.8000	13524	13524	0.1200	0.1000	0.1400	-0.0200	0.0200	0.0100	0.0500	0.0500

The "Report Definitions" section explains the terms used in the output table. "Power" is the probability of concluding equivalence when equivalence is correct. "Beta" is the probability of accepting a false H0. "N1" and "N2" are the sizes of the samples drawn from the corresponding groups. "P2" is the response rate for group two which is the standard, reference, baseline, or control group. "P1.0L" is the smallest treatment-group response rate that still yields an equivalence conclusion. "P1.0U" is the largest treatment-group response rate that still yields an equivalence conclusion. "D0.L" is the lowest difference that still results in the conclusion of equivalence. "D0.U" is the highest difference that still results in the conclusion of equivalence. "D1" is the actual difference, P1-P2, at which the power is calculated. "Target Alpha" is the probability of rejecting a true null hypothesis that was desired. "Actual Alpha" is the value of alpha that is actually achieved. Only available for exact results. "Grp 1" refers to Group 1 which is the treatment or experimental group. "Grp 2" refers to Group 2 which is the reference, standard, or control group. "Equiv." refers to a small amount that is not of practical importance. "Actual" refers to the true value at which the power is computed.

The "Summary Statements" section states: "Sample sizes of 13524 in the treatment group and 13524 in the reference group achieve 80% power to detect equivalence. The margin of equivalence, given in terms of the difference, extends from -0.0200 to 0.0200. The actual difference is 0.0100. The reference group proportion is 0.1200. The calculations assume that two, one-sided pooled Z tests are used. The significance level of the test is 0.0500."

4.5 Chi-square Tests

Example 4.27 Confirm the sample size for Example 4.18 using the χ^2 test method for a 2×2 table.

Solution: Under H_A with $p_1 = 0.10$ and $p_2 = 0.20$ the expected proportion of observations in each cell of the 2×2 table is

$$(p_{ij})_A = \frac{1}{2} \begin{Bmatrix} 0.1 & 0.9 \\ 0.2 & 0.8 \end{Bmatrix} = \begin{Bmatrix} 0.05 & 0.45 \\ 0.1 & 0.4 \end{Bmatrix}.$$

Under H_0 with $p_1 = p_2 = (0.1 + 0.2) / 2 = 0.15$ the expected distribution of observations is

$$(p_{ij})_0 = \frac{1}{2} \begin{Bmatrix} 0.15 & 0.85 \\ 0.15 & 0.85 \end{Bmatrix} = \begin{Bmatrix} 0.075 & 0.425 \\ 0.075 & 0.425 \end{Bmatrix}.$$

From Equation 4.121 with a total of $2 \times 268 = 536$ observations the noncentrality parameter is

$$\begin{aligned} \phi &= 536 \left(\frac{(0.05 - 0.075)^2}{0.075} + \frac{(0.1 - 0.075)^2}{0.075} + \frac{(0.45 - 0.425)^2}{0.425} \right. \\ &\quad \left. + \frac{(0.4 - 0.425)^2}{0.425} \right) \\ &= 10.51. \end{aligned}$$

Then, with $\alpha = 0.05$ and $df = 1$ degree of freedom in Equation 4.122,

$$\chi_{0.95}^2 = 3.8415 = \chi_{\beta, 10.51}^2$$

which is satisfied by $\beta = 0.10$, so the power is $\pi = 1 - \beta = 0.90$ and is consistent with the original example problem solution.

Example 4.28 A large school district intends to perform pass/fail testing of students from four large schools to test for performance differences among schools. If 50 students are chosen randomly from each school, what is the power of the χ^2 test to reject the null hypothesis of homogeneity when the student failure rates at the four schools are in fact 10%, 10%, 10%, and 30%?

Solution: To calculate the power of the χ^2 test we must specify the two 2×4 tables (result by school) associated with $(p_{ij})_0$ and $(p_{ij})_A$. From the problem statement, under H_A with $(p_{1j})_A = \{0.1, 0.1, 0.1, 0.3\}$, the table of $(p_{ij})_A$ is

$$\begin{aligned} (p_{ij})_A &= \frac{1}{4} \begin{Bmatrix} 0.1 & 0.1 & 0.1 & 0.3 \\ 0.9 & 0.9 & 0.9 & 0.7 \end{Bmatrix} \\ &= \begin{Bmatrix} 0.025 & 0.025 & 0.025 & 0.075 \\ 0.225 & 0.225 & 0.225 & 0.175 \end{Bmatrix}. \end{aligned}$$

The mean failure rate of all four schools is $(3(0.1) + 0.3) / 4 = 0.15$ under H_0 , so the corresponding table of $(p_{ij})_0$ is

$$\begin{aligned} (p_{ij})_0 &= \frac{1}{4} \begin{Bmatrix} 0.15 & 0.15 & 0.15 & 0.15 \\ 0.85 & 0.85 & 0.85 & 0.85 \end{Bmatrix} \\ &= \begin{Bmatrix} 0.0375 & 0.0375 & 0.0375 & 0.0375 \\ 0.2125 & 0.2125 & 0.2125 & 0.2125 \end{Bmatrix}. \end{aligned}$$

Under these definitions, the χ^2 distribution noncentrality parameter is

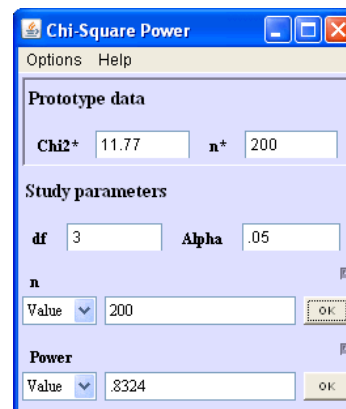
$$\begin{aligned}\phi &= 200 \left[3 \left(\frac{(0.025 - 0.0375)^2}{0.0375} \right) + \frac{(0.075 - 0.0375)^2}{0.0375} \right. \\ &\quad \left. + 3 \left(\frac{(0.225 - 0.2125)^2}{0.2125} \right) + \frac{(0.175 - 0.2125)^2}{0.2125} \right] \\ &= 11.77.\end{aligned}$$

The χ^2 test statistic will have $df = (2 - 1)(4 - 1) = 3$ degrees of freedom, so the critical value of the test statistic is $\chi_{0.95,3}^2 = 7.81$. The power of the test determined from the condition

$$\chi_{0.95}^2 = 7.81 = \chi_{1-\pi, 11.77}^2$$

is $\pi = 0.833$.

From **Piface** > **Generic chi-square test**:



From **PASS** > **Proportions** > **Multi-Group: Chi-Square Test** with effect size $W = \sqrt{\phi/N} = \sqrt{11.77/200} = 0.2425$:

The screenshot shows two windows from the PASS software. The left window, titled "PASS: Chi-Square Test", has a menu bar (File, Run, Analysis, Graphics, PASS, GESS, Tools, Window, Help) and a toolbar with icons for RUN, NEW, OPEN, SAVE, MAP, NRV, PASS, DTRA, OUT, and PLAY. Below the toolbar are tabs for Symbols, Background, Abbreviations, and Templates. The main area contains input fields for "Find (Solve For):" (Beta and Power), "W (Effect Size):" (0.2425), "DF (Degrees of Freedom):" (3), "Alpha (Significance Level):" (.05), and "N (Sample Size):" (200). The right window, titled "Chi-Square Effect Size Estimator", has tabs for "Contingency Table" and "Multinomial Test". It features a grid for entering values, a "Reset" button, and output fields for Chi-Square (11.764706), DF (3), Effect Size (W) (0.242536), Prob Level (0.008234), N (200), Rows (2), and Columns (4). A note at the bottom states: "This window calculates the Chi-Square probability level (PROB LEVEL) and the effect size (W) based on values entered into a two-way table. The values may be table counts or percentages."

Chi-Square Test Power Analysis

Numeric Results for Chi-Square Test

Power	N	W	Chi-Square	DF	Alpha	Beta
0.83210	200	0.2425	11.7613	3	0.05000	0.16790

References

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population. To conserve resources, it should be small.
 W is the effect size--a measure of the magnitude of the Chi-Square that is to be detected.
 DF is the degrees of freedom of the Chi-Square distribution.
 Alpha is the probability of rejecting a true null hypothesis.
 Beta is the probability of accepting a false null hypothesis.

Summary Statements

A sample size of 200 achieves 83% power to detect an effect size (W) of 0.2425 using a 3 degrees of freedom Chi-Square Test with a significance level (alpha) of 0.05000.

Example 4.29 What is the power to reject the claim that a die is balanced ($H_0 : \theta_i = \frac{1}{6}$ for $i = 1$ to 6) when it is in fact slightly biased toward one die face ($H_A : \theta_i = \{0.16, 0.16, 0.16, 0.16, 0.16, 0.20\}$) based on 100 rolls of the die?

Solution: The table of observations will have six cells and there will be no parameters estimated from the sample data, so the χ^2 test will have $df = 6 - 1 = 5$ degrees of freedom. From Equation 4.121 the noncentrality parameter will be

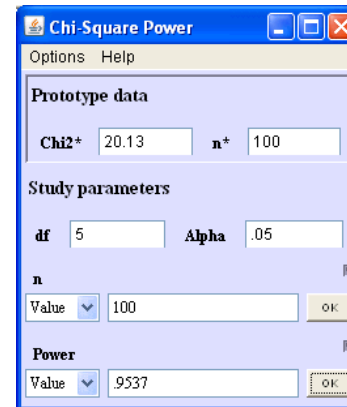
$$\phi = 100 \left[5 \left(\frac{(0.16 - \frac{1}{6})^2}{\frac{1}{6}} \right) + \frac{(0.20 - \frac{1}{6})^2}{\frac{1}{6}} \right] = 20.13.$$

With $\alpha = 0.05$ we have $\chi_{0.95}^2 = 11.07$, so the power to reject H_0 is determined from the condition

$$\chi_{0.95}^2 = 11.07 = \chi_{1-\pi, 20.13}^2,$$

which is satisfied by $\pi = 0.954$.

From **Piface**> **Generic chi-square test**:



From **PASS**> **Proportions**> **Multi-Group: Chi-Square Test** with effect size $W = \sqrt{\phi/N} = \sqrt{20.13/100} = 0.4487$:

PASS: Chi-Square Test Output

Chi-Square Test Power Analysis

Numeric Results for Chi-Square Test						
Power	N	W	Chi-Square	DF	Alpha	Beta
0.95371	100	0.4487	20.1332	5	0.05000	0.04629

References
Cohen, Jacob. 1988. Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population. To conserve resources, it should be small.
W is the effect size--a measure of the magnitude of the Chi-Square that is to be detected.
DF is the degrees of freedom of the Chi-Square distribution.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A sample size of 100 achieves 95% power to detect an effect size (W) of 0.4487 using a 5 degrees of freedom Chi-Square Test with a significance level (alpha) of 0.05000.

Chapter 5

Poisson Counts

5.1 One Poisson Count

Example 5.1 How many Poisson events must be observed if the relative error of the estimate for λ must be no larger than $\pm 10\%$ with 95% confidence?

Solution: The desired confidence interval for λ has the form

$$P\left(\frac{x}{n}(1 - 0.10) < \lambda < \frac{x}{n}(1 + 0.10)\right) = 0.95,$$

so $\delta = 0.10$ and from Equation 5.10

$$x = \left(\frac{1.96}{0.10}\right)^2 = 385.$$

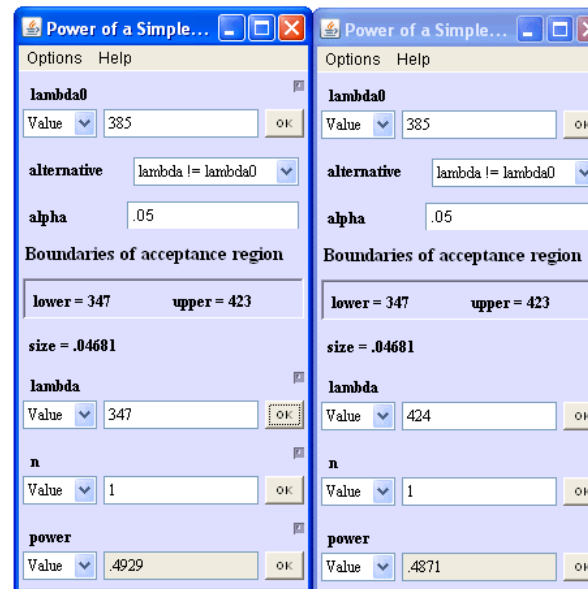
That is, if the Poisson process is sampled until $x = 385$ counts are obtained, then the 95% confidence limits for λ will be

$$UCL/LCL = \left(\frac{385}{n}\right)(1 \pm 0.10)$$

or

$$P\left(\frac{346}{n} < \lambda < \frac{424}{n}\right) = 0.95.$$

From Piface > Generic Poisson test:



From MINITAB (V16) > Stat > Power and Sample Size > Sample Size for Estimation > Mean (Poisson):

```

MTB > SSCI;
SUBC> PMean 10;
SUBC> Confidence 95.0;
SUBC> IType 0;
SUBC> MError 1.

```

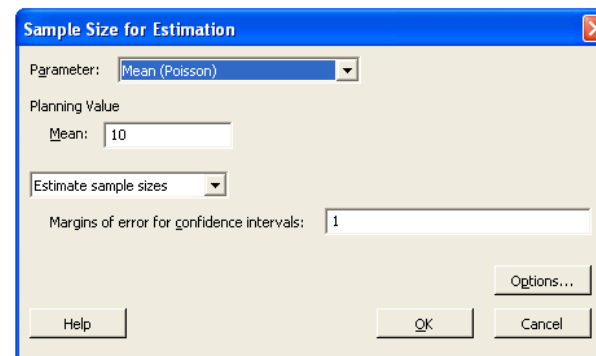
Sample Size for Estimation

Method

Parameter	Mean
Distribution	Poisson
Mean	10
Confidence level	95%
Confidence interval	Two-sided

Results

Margin of Error	Sample Size
1	43



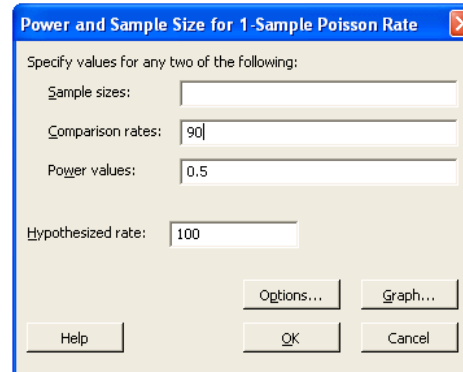
From MINITAB (V16) > Stat > Power and Sample Size > 1-Sample Poisson Rate:

```
MTB > Power;
SUBC> OneRate;
SUBC> RCompare 90;
SUBC> Power 0.5;
SUBC> RNull 100;
SUBC> Alternative 0;
SUBC> Alpha 0.05;
SUBC> Length 1.0;
SUBC> GPCurve.
```

Power and Sample Size

Test for 1-Sample Poisson Rate
 Testing rate = 100 (versus not = 100)
 Alpha = 0.05
 "Length" of observation = 1

Comparison Rate	Sample Size	Target Power	Actual Power
90	4	0.5	0.516846

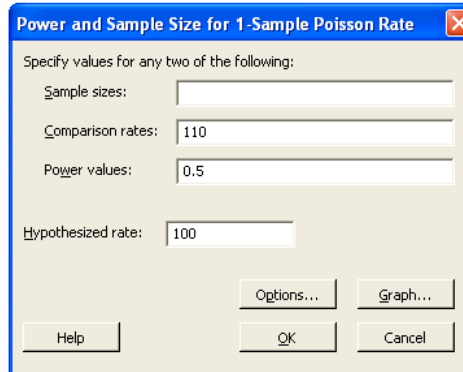


```
MTB > Power;
SUBC> OneRate;
SUBC> RCompare 110;
SUBC> Power 0.5;
SUBC> RNull 100;
SUBC> Alternative 0;
SUBC> Alpha 0.05;
SUBC> Length 1.0;
SUBC> GPCurve.
```

Power and Sample Size

Test for 1-Sample Poisson Rate
 Testing rate = 100 (versus not = 100)
 Alpha = 0.05
 "Length" of observation = 1

Comparison Rate	Sample Size	Target Power	Actual Power
110	4	0.5	0.515305



Example 5.2 For the hypothesis test of $H_0 : \lambda = 4$ versus $H_A : \lambda > 4$ based on a sample of size $n = 2$ units using $\alpha \leq 0.05$, determine the power to reject H_0 when $\lambda = 8$.
Solution: Under H_0 , the distribution of the observed number of counts x will be Poisson with $\mu_x = n\lambda_0 = 2 \times 4 = 8$. The acceptance interval for H_0 will be $0 \leq x \leq 13$ because

$$(1 - \text{Poisson}(0 \leq x \leq 12; 8) = 0.064) > 0.05$$

$$(1 - \text{Poisson}(0 \leq x \leq 13; 8) = 0.034) < 0.05,$$

so the exact significance level for the test will be $\alpha = 0.034$. With $\lambda = 8$, the power to reject H_0 is

$$\pi = 1 - \text{Poisson}(0 \leq x \leq 13; n\lambda = 16) = 0.725.$$

The count distributions under H_0 and H_A are shown in Figure 5.1.

From Piface> **Generic Poisson test**:

The screenshot shows the 'Power of a Simple Poisson Test' dialog box with the following settings:

- lambda0**: Value 4
- alternative**: lambda > lambda0
- alpha**: .05
- Boundaries of acceptance region**: upper = 12
- size**: .03418
- lambda**: Value 8
- n**: Value 2
- power**: Value .7255

At the bottom of the dialog, it says 'Java Applet Window'.

MINITAB V16 uses the normal approximation to the Poisson distribution so its answers are different from the exact answers. From **MINITAB (V16)> Stat> Power and Sample Size> 1-Sample Poisson Rate**:

```

MTE > Power;
SUBC> OneRate;
SUBC> Sample 2;
SUBC> RCompare 8;
SUBC> RNull 4;
SUBC> Alternative 1;
SUBC> Alpha 0.05;
SUBC> Length 1.0;
SUBC> GPCurve.

```

Power and Sample Size

Test for 1-Sample Poisson Rate

Testing rate = 4 (versus > 4)
 Alpha = 0.05
 "Length" of observation = 1

Comparison Rate	Sample Size	Power
8	2	0.798679

Example 5.3 For the hypothesis test of $H_0 : \lambda = 4$ versus $H_A : \lambda > 4$ based on a sample of size $n = 5$ units, determine the power to reject H_0 when $\lambda = 9$. Use the square root transformation method with $\alpha = 0.05$.

Solution: By Equation 5.16, the power to reject $H_0 : \lambda = 4$ when $\lambda = 9$ is

$$\begin{aligned}
 \pi &= \Phi\left(-2\sqrt{5}\left(\sqrt{9} - \sqrt{4}\right) + z_{0.05} < z < \infty\right) \\
 &= \Phi(-2.83 < z < \infty) \\
 &= 0.9977.
 \end{aligned}$$

From Piface> Generic Poisson test:

Power of a Simple Poisson Test

Options Help

lambda0

Value OK

alternative

alpha

Boundaries of acceptance region

upper = 27

size = .03433

lambda

Value OK

n

Value OK

power

Value OK

Java Applet Window

From MINITAB (V16) > Stat > Power and Sample Size > 1-Sample Poisson Rate:


```

MTB > Power;
SUBC> OneRate;
SUBC> Sample 5;
SUBC> RCompare 9;
SUBC> RNull 4;
SUBC> Alternative 1;
SUBC> Alpha 0.05;
SUBC> Length 1.0;
SUBC> GPCurve.

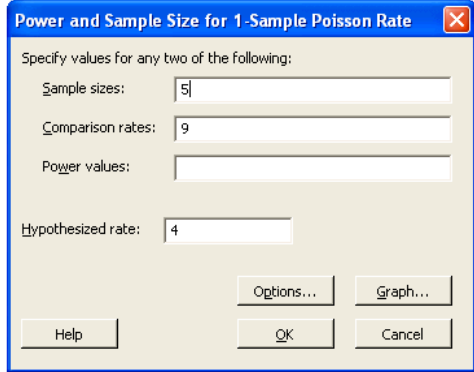
```

Power and Sample Size

Test for 1-Sample Poisson Rate

Testing rate = 4 (versus > 4)
Alpha = 0.05
"Length" of observation = 1

Comparison Rate	Sample Size	Power
9	5	0.995733



Example 5.4 How many sampling units must be inspected to reject $H_0 : \lambda = 10$ with 90% power in favor of $H_A : \lambda > 10$ when in fact $\lambda = 15$?

Solution: By Equation 5.17 the necessary sample size is

$$n = \frac{1}{4} \left(\frac{1.645 + 1.282}{\sqrt{15} - \sqrt{10}} \right)^2 = 4.2,$$

which rounds up to $n = 5$ sampling units.

From **Piface**> **Generic Poisson test**:

Power of a Simple Poisson Test

Options Help

lambda0
Value ▾ 10 OK

alternative lambda > lambda0 ▾

alpha .05

Boundaries of acceptance region

upper = 61

size = .04239

lambda
Value ▾ 15 OK

n
Value ▾ 5 OK

power
Value ▾ .9288 OK

Java Applet Window

From MINITAB (V16) > Stat > Power and Sample Size > 1-Sample Poisson Rate:

```

MTB > Power;
SUBC> OneRate;
SUBC> Sample 5;
SUBC> RCompare 9;
SUBC> RNull 4;
SUBC> Alternative 1;
SUBC> Alpha 0.05;
SUBC> Length 1.0;
SUBC> GPCurve.

```

Power and Sample Size

Test for 1-Sample Poisson Rate

Testing rate = 4 (versus > 4)
Alpha = 0.05
"Length" of observation = 1

Comparison Rate	Sample Size	Power
9	5	0.995733

5.2 Two Poisson Counts

Example 5.5 What optimal sample sizes are required to estimate the difference between two Poisson means with 30% precision if the means are expected to be $\lambda_1 = 25$ and $\lambda_2 = 16$?

Solution: The difference between the means is expected to be $\Delta\lambda = 9$, so the confidence interval half-width must be 30% of that, or

$$\delta = 0.3 \times 9 = 2.7.$$

From Equation 5.24, the optimal sample size ratio is

$$\frac{n_1}{n_2} = \sqrt{\frac{\lambda_1}{\lambda_2}} = \sqrt{\frac{25}{16}} = 1.25.$$

From Equation 5.22, with $\alpha = 0.05$, the sample size n_1 must be

$$n_1 = \left(\frac{1.96}{2.7}\right)^2 (25 + 1.25 \times 16) = 23.7$$

and the sample size n_2 must be

$$n_2 = \frac{n_1}{\left(\frac{n_1}{n_2}\right)} = \frac{23.7}{1.25} = 18.96,$$

which round up to $n_1 = 24$ and $n_2 = 19$.

Example 5.6 How many Poisson counts are required to estimate the ratio of the means of two independent Poisson distributions to within 20% of the true ratio with 95% confidence if the sample sizes will be the same and the ratio of the means is expected to be $\lambda_1/\lambda_2 \simeq 2$?

Solution: With $n_1/n_2 = 1$, $\lambda_1/\lambda_2 = 2$, $z_{0.025} = 1.96$, and $\delta = 0.02$ in Equation 5.30, the number of Poisson counts required in the first sample is

$$\begin{aligned} x_1 &= (1 + 1 \times 2) \left(\frac{1.96}{0.2} \right)^2 \\ &= 289. \end{aligned}$$

The corresponding required counts in the second sample are about half of the counts in the first: $x_2 = 289/2 = 145$.

Example 5.7 Determine the power to reject $H_0 : \lambda_1 = \lambda_2$ in favor of $H_A : \lambda_1 < \lambda_2$ when $\lambda_1 = 10$, $n_1 = 8$ and $\lambda_2 = 15$, $n_2 = 6$. Use the large-sample normal approximation, square root transform, and F test methods with $\alpha = 0.05$.

Solution: The expected number of counts from the first (x_1) and second (x_2) populations are both large enough to justify the large sample approximation method. By this method the power is

$$\begin{aligned} \pi &= \Phi \left(-\infty < z < \frac{15 - 10}{\sqrt{\frac{15}{6} + \frac{10}{8}}} - 1.645 \right) \\ &= \Phi(-\infty < z < 0.937) \\ &= 0.826. \end{aligned}$$

By the log-transformation method the power is

$$\begin{aligned} \pi &= \Phi \left(-\infty < z < \frac{\log(15/10)}{\sqrt{\frac{1}{6 \times 15} + \frac{1}{8 \times 10}}} - 1.645 \right) \\ &= \Phi(-\infty < z < 0.994) \\ &= 0.840. \end{aligned}$$

By the square-root transform method the power is

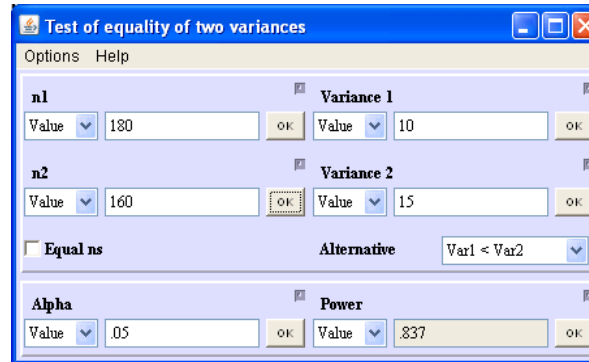
$$\begin{aligned} \pi &= \Phi \left(-\infty < z < \frac{\sqrt{15} - \sqrt{10}}{\frac{1}{2} \sqrt{\frac{1}{8} + \frac{1}{6}}} - 1.645 \right) \\ &= \Phi(-\infty < z < 1.01) \\ &= 0.838. \end{aligned}$$

By the F test method the power is

$$\begin{aligned} \pi &= P \left(\frac{10}{15} F_{0.95, 2(6)(15), 2(8)(10)} < F < \infty \right) \\ &= P(0.86 < F < \infty) \\ &= 0.837. \end{aligned}$$

MINITAB V16 supports the two-sample Poisson method but only for equal sample sizes.

By the F test method using **Piface**> **Two variances (F Test)** with $n_1 = 2 \times 6 \times 15 = 180$ and $n_2 = 2 \times 8 \times 10 = 160$:



By the F test method using **PASS**> **Variance**> **Variance: 2 Groups**:

PASS: Variances: 2

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols z | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For): Scale:

V1 (Variance of Group 1): Alternative Hypothesis:

V2 (Variance of Group 2): N1 (Sample Size Group 1):

Alpha (Significance Level): N2 (Sample Size Group 2):

Beta (1-Power): R (Sample Allocation Ratio):

PASS: Variances: 2 Output

Power Analysis of Two Variances

Numeric Results when H0: V1 = V2 versus Ha: V1 < V2

Power	N1	N2	V1	V2	Alpha	Beta
0.835662	180	160	10.0000	15.0000	0.050000	0.164338

References

Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
 Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press. Ames, Iowa.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N1 is the size of the sample drawn from the population 1.
 N2 is the size of the sample drawn from the population 2.
 V1 is the value of the population variance of group 1.
 V2 is the value of the population variance of group 2.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

Group sample sizes of 180 and 160 achieve 84% power to detect a ratio of 0.6667 between the group one variance of 10.0000 and the group two variance of 15.0000 using a one-sided F test with a significance level (alpha) of 0.050000.

Example 5.8 What minimum total counts are required for the two-sample counts test to detect a factor of two difference between the count rates with 90% power? Assume that the two sample sizes will be equal.

Solution: The hypotheses to be tested are $H_0 : \lambda_1/\lambda_2 = 1$ versus $H_A : \lambda_1/\lambda_2 > 1$. From Equation 5.45, which is expressed in terms of the ratio of the two means, the number of

count events x_1 required to reject H_0 when $\lambda_1/\lambda_2 = 2$ is

$$\begin{aligned} x_1 &= \left(1 + \frac{\lambda_1}{\lambda_2}\right) \left(\frac{z_\alpha + z_\beta}{\ln(\lambda_1/\lambda_2)}\right)^2 \\ &= (1 + 2) \left(\frac{1.645 + 1.282}{\ln(2)}\right)^2 \\ &= 54. \end{aligned}$$

Because $\lambda_2 = \lambda_1/2$, the corresponding number of x_2 counts is $x_2 = 54/2 = 27$.

From MINITAB (V16) > Stat > Power and Sample Size > 2-Sample Poisson Rate:

```
MTB > Power;
SUBC> TwoRate;
SUBC> RCompare 2;
SUBC> Power 0.90;
SUBC> RBaseline 1;
SUBC> Alternative 1;
SUBC> Alpha 0.05;
SUBC> Length 1.0;
SUBC> GPCurve.
```

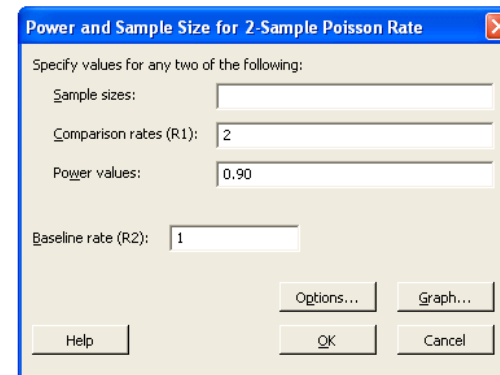
Power and Sample Size

Test for 2-Sample Poisson Rate

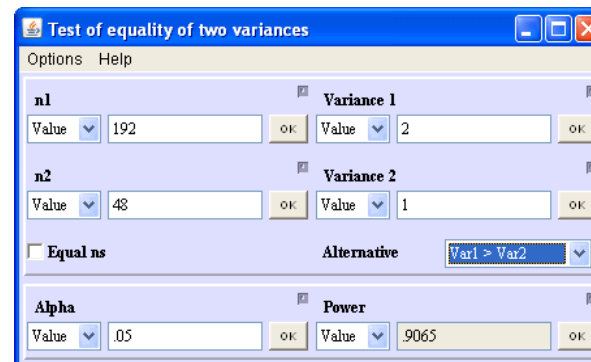
Testing comparison rate = baseline rate (versus >)
Calculating power for baseline rate = 1
Alpha = 0.05
"Lengths" of observation for sample 1, sample 2 = 1, 1

Comparison Rate	Sample Size	Target Power	Actual Power
2	26	0.9	0.903039

The sample size is for each group.



By trial and error using the F test method in Piface > Two variances (F Test), 90% power is obtained with $x_1 = 192/(2 \times 2) = 48$ and $x_2 = 48/(2 \times 2) = 24$:



5.3 Tests for Many Poisson Counts

Example 5.9 In a test for differences between mean counts from five different processes, determine the power to reject $H_0 : \lambda_i = \lambda_j$ for all i, j pairs when $\lambda_1 = \lambda_2 = \lambda_3 = 16$, $\lambda_4 = 9$, $\lambda_5 = 25$ and $n = 3$ units from each process are inspected. The number of counts will be reported for each unit. Assume that the test will be performed using one-way ANOVA applied to the square root transformed counts.

Solution: After the square root transform, the transformed treatment means are $\lambda'_1 = \lambda'_2 = \lambda'_3 = 4$, $\lambda'_4 = 3$, and $\lambda'_5 = 5$. The grand transformed mean is $\bar{\lambda}' = 4$, so the treatment biases relative to the grand mean are 0, 0, 0, -1, and 1, respectively. The ANOVA F test noncentrality parameter is then

$$\phi = \frac{E(SS_{Treatment})}{E(MS_\epsilon)} = \frac{3(0^2 + 0^2 + 0^2 + (-1)^2 + (1)^2)}{(\frac{1}{2})^2} = 24$$

where $E(MS_\epsilon) = (\sigma')^2 = (\frac{1}{2})^2$ is the error variance of the transformed counts. The ANOVA will have $df_{Treatment} = 4$ and $df_\epsilon = 15 - 1 - 4 = 10$, so the F test critical value will be $F_{0.95,4,24} = 3.48$. The power to reject H_0 is then given by Equation 8.1:

$$\begin{aligned} F_{1-\alpha} &= F_{1-\pi, \phi} \\ 3.48 &= F_{1-\pi, 24} \\ 3.48 &= F_{0.11, 24, } \end{aligned}$$

so the power is $\pi = 0.89$ to reject H_0 for the specified set of means.

Example 5.10 In a test for differences among the means of five Poisson populations, determine the probability of rejecting $H_0 : \lambda_i = \lambda_0$ for all i when $\lambda_i = \{16, 16, 16, 12, 20\}$. The number of units inspected is $n_i = 4$ for all i .

Solution: Given the Poisson means specified under H_A , the value of λ_0 under H_0 is given by

$$\lambda_0 = \frac{1}{5}(16 + 16 + 16 + 12 + 20) = 16.$$

With $n_i = n = 4$, the noncentrality parameter is given by

$$\begin{aligned} \phi &= n \sum_{i=1}^k \frac{(\lambda_{A,i} - \lambda_0)^2}{\lambda_0} \\ &= 4 \left(\frac{(0)^2}{16} + \frac{(0)^2}{16} + \frac{(0)^2}{16} + \frac{(-4)^2}{16} + \frac{(4)^2}{16} \right) = 8. \end{aligned}$$

The power is determined from Equation 5.61:

$$\chi_{0.95}^2 = 9.49 = \chi_{0.395, 8}^2$$

where the central and noncentral χ^2 distributions both have $\nu = 5 - 1 = 4$ degrees of freedom, so the power is $\pi = 0.605$.

5.4 Correcting for Background Counts

Example 5.11 In a two-sample test for counts, what common sample size $n = n_1 = n_2$ is required to distinguish $\lambda_1 = \lambda_2 = 6$ from $\lambda_1 = 6, \lambda_2 = 15$ with 90% power in the presence of a background count rate of $\lambda_0 = 10$?

Solution: From Equation 5.55, modified to account for the background count rate, the necessary sample size to reject $H_0 : \lambda_1 = \lambda_2$ in favor of $H_A : \lambda_1 < \lambda_2$ with 90% power and $\alpha = 0.05$ is given by

$$\begin{aligned} n &= \frac{1}{2} \left(\frac{z_\alpha + z_\beta}{(\sqrt{\lambda_2 + \lambda_0} - \sqrt{\lambda_1 + \lambda_0})} \right)^2 \\ &= \frac{1}{2} \left(\frac{1.645 + 1.282}{\sqrt{25} - \sqrt{16}} \right)^2 = 5. \end{aligned} \tag{5.1}$$

Chapter 6

Regression

6.1 Linear Regression

Example 6.1 Designed experiments frequently involve two or three equally weighted levels of x . Compare the sample sizes required for these two important special cases if they must both deliver a β_1 confidence interval half-width δ and the observations are taken over the same x range from x_{min} to x_{max} . For the three-level case, assume that the middle level will be midway between x_{min} and x_{max} .

Solution: The subscripts 2 and 3 will be used to indicate parameters from the two-level and three-level cases, respectively. For the two-level case, from Equation 6.12 with $k_2 = 2$ and $\Delta x_2 = x_{max} - x_{min}$, the sample size per x level will be

$$\begin{aligned} n_2 &\geq 2 \left(\frac{t_{\alpha/2} \hat{\sigma}_\epsilon}{\delta \Delta x_1} \right)^2 \\ &\geq 2 \left(\frac{t_{\alpha/2} \hat{\sigma}_\epsilon}{\delta (x_{max} - x_{min})} \right)^2. \end{aligned} \tag{6.1}$$

For the three-level case with $k_3 = 3$ and $\Delta x_3 = \frac{1}{2} (x_{max} - x_{min})$ the sample size per x level will be

$$\begin{aligned} n_3 &\geq \frac{1}{2} \left(\frac{t_{\alpha/2} \hat{\sigma}_\epsilon}{\delta \Delta x_2} \right)^2 \\ &\geq 2 \left(\frac{t_{\alpha/2} \hat{\sigma}_\epsilon}{\delta (x_{max} - x_{min})} \right)^2. \end{aligned} \tag{6.2}$$

Because $n_2 = n_3$, $N_2 = 2n_2$, and $N_3 = 3n_2$, the two experiments appear to have the same ability to resolve β_1 even though the three-level experiment requires 50% more observations! This means that the middle observations in the three-level experiment are effectively wasted for the purpose of estimating β_1 . This statement is not entirely true because the middle observations in the three-level experiment do add error degrees of freedom, which potentially decrease n_3 compared to n_2 for the same δ . In general, the purpose of using three levels of x in an experiment is not to improve the precision of the β_1 estimate; rather, three levels are used to allow a linear lack of fit test, which is not possible using just two levels of x .

Example 6.2 Compare the sample sizes required to estimate the slope parameter with equal precision for two experiments if x is uniformly distributed over the interval from x_{min} to x_{max} in the first experiment and if x has two levels, x_{min} and x_{max} , in the second experiment.

Solution: From Equations 6.9 and 6.13 the ratio of the total number of observations required by the two experiments is

$$\frac{N_{\text{uniform } x}}{N_{\text{two levels of } x}} \simeq \frac{12 \left(\frac{t_{\alpha/2} \hat{\sigma}_\epsilon}{\delta(x_{max} - x_{min})} \right)^2}{4 \left(\frac{t_{\alpha/2} \hat{\sigma}_\epsilon}{\delta \Delta x} \right)^2}$$

where the $t_{\alpha/2}$ values may differ a bit because of the difference in error degrees of freedom. Both experiments cover the same x range, so $\Delta x = x_{max} - x_{min}$ and the sample size ratio reduces to:

$$\frac{N_{\text{uniform } x}}{N_{\text{two levels of } x}} \simeq 3.$$

That is, three times as many observations are required in an experiment that uses uniformly distributed x values than if the x values are concentrated at the ends of the x range. Because we saw in Example 6.1 that the experiment with three evenly spaced, equally weighted levels of x requires 1.5 times as many observations as the two-level experiment, other methods of taking evenly spaced, equally weighted observations of x must give experiment sample size ratios between 1.5 and 3. Obviously, the two-level equally weighted method is the most efficient method for specifying x values for an experiment.

Example 6.3 For an experiment to be analyzed by linear regression with a single predictor, how many observations are required to reject $H_0 : \beta_1 = 0$ in favor of $H_A : \beta_1 \neq 0$ with 90% power for $\beta_1 = 10$ when a) the distribution of x values will be normal with $\mu_x \simeq 15$ and $\sigma_x \simeq 2$; b) an equal number of observations will be taken at $x = 10$ and $x = 20$; c) uniformly distributed values of x will be used over the interval $10 \leq x \leq 20$; and d) an equal number of observations will be taken at $x = 10, 15,$ and 20 . Experience with the process tells us the standard error of the model is expected to be $\sigma_\epsilon = 30$.

Solution:

a) With $t \simeq z$ in Equation 6.19, the first iteration to find N gives

$$N = (z_{0.025} + z_{0.10})^2 \left(\frac{30}{10 \times 2} \right)^2 = 24.$$

Further iterations indicate that the required sample size is $N = 26$.

From **Piface**> **Linear regression**:

The screenshot shows a Java Applet window titled "Linear Regression". The window contains several input fields and a "Solve for" dropdown menu. The "No. of predictors" field is set to 1, "Error SD" is 30, "SD of x[j]" is 2, "Detectable beta[j]" is 10, "Sample size" is 26, "Alpha" is .05, and "Power" is 9033. The "Two-tailed" checkbox is checked. The "Solve for" dropdown is set to "Sample size".

Parameter	Value
No. of predictors	1
Error SD	30
SD of x[j]	2
Detectable beta[j]	10
Sample size	26
Alpha	.05
Power	9033
Two-tailed	<input checked="" type="checkbox"/>
Solve for	Sample size

From PASS> Regression> Linear Regression:

The screenshot shows the PASS: Regression: Linear software interface. The left pane displays the input parameters for a Linear Regression Power Analysis:

- Find (Solve For): N
- Alternative Hypothesis: Two-Sided
- B0 (Slope|H0): 0.0
- Residual Variance Method: S (Std. Dev. of Residuals)
- B (Slope|H1): 10
- SY (Std Deviation of Y):
- Alpha (Significance Level): .05
- R (Correlation):
- Beta (1-Power): 0.10
- S (Std Dev of Residuals): 30
- N (Sample Size):
- SX (Std Deviation of X's): 2

The right pane displays the output results for the Linear Regression Power Analysis:

Linear Regression Power Analysis

Numeric Results for Two-Sided Testing of B = B0 where B0 = 0.00

Power	Sample Size (N)	Slope (B)	Standard Deviation of X (SX)	Standard Deviation of Residuals (S)	Alpha	Beta
0.90329	26	10.00	2.00	30.00	0.05000	0.09671

References
Neter, J., Wasserman, W., and Kutner, M. 1983. Applied Linear Regression Models. Richard D. Irwin, Inc. Chicago, Illinois.

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population. To conserve resources, it should be small.
B0 is the slope under the null hypothesis.
B is the slope at which the power is calculated.
SX is the standard deviation of the X values.
S is the standard deviation of the residuals.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements
A sample size of 26 achieves 90% power to detect a change in slope from 0.00 under the null hypothesis to 10.00 under the alternative hypothesis when the standard deviation of the X's is 2.00, the standard deviation of the residuals is 30.00, and the two-sided significance level is 0.05000.

b) The standard deviation of the x values will be

$$\sigma_x = \sqrt{\frac{SS_x}{N}} = \sqrt{\frac{1}{N} \frac{N}{2} \left((-5)^2 + (5)^2 \right)} = 5.$$

The first iteration to find N , with $t \simeq z$, gives

$$N = (z_{0.025} + z_{0.10})^2 \left(\frac{30}{10 \times 5} \right)^2 = 4.$$

Further iterations indicate that $N = 7$ observations are required.

From **Piface**> **Linear regression**:

Linear Regression

Options Help

No. of predictors
Value 1 OK

SD of x[j]
Value 2.89 OK

Alpha
Value .05 OK

Two-tailed

Error SD
Value 30 OK

Detectable beta[j]
Value 10 OK

Sample size
Value 14 OK

Power = .9105

Solve for Sample size

Linear Regression

Options Help

No. of predictors
Value 1 OK

SD of x[j]
Value 5 OK

Alpha
Value .05 OK

Two-tailed

Error SD
Value 30 OK

Detectable beta[j]
Value 10 OK

Sample size
Value 7 OK

Power = .9359

Solve for Sample size

c) For uniformly distributed x , the standard deviation of the x values is

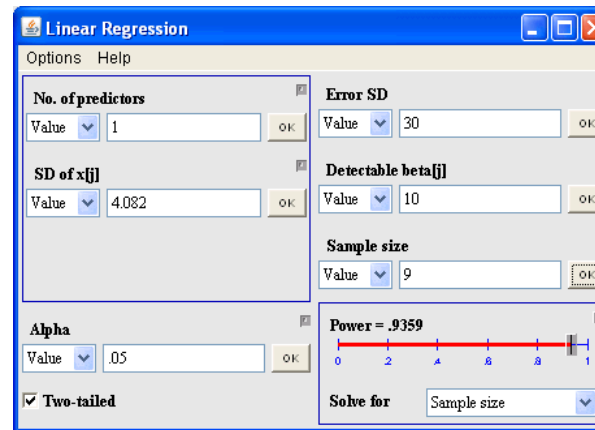
$$\sigma_x = \frac{x_{\max} - x_{\min}}{\sqrt{12}} = \frac{10}{\sqrt{12}} = 2.89.$$

With $t \simeq z$, the first iteration to find N gives

$$N = (z_{0.025} + z_{0.10})^2 \left(\frac{30}{10 \times 2.89} \right)^2 = 12.$$

Further iterations indicate that $N = 14$ observations are required.

From **Piface**> **Linear regression**:



d) The standard deviation of the x values will be

$$\sigma_x = \sqrt{\frac{SS_x}{N}} = \sqrt{\frac{1}{N} \frac{N}{3} \left((-5)^2 + (0)^2 + (5)^2 \right)} = 4.0825.$$

The first iteration to find N , with $t \simeq z$, gives

$$N = (z_{0.025} + z_{0.10})^2 \left(\frac{30}{10 \times 4.0825} \right)^2 = 6.$$

Further iterations indicate that $N = 9$ observations are required.

From **Piface** > **Linear regression**:

Example 6.4 What is the power to reject H_0 for the situation described in Example 6.3a if the sample size is $N = 20$?

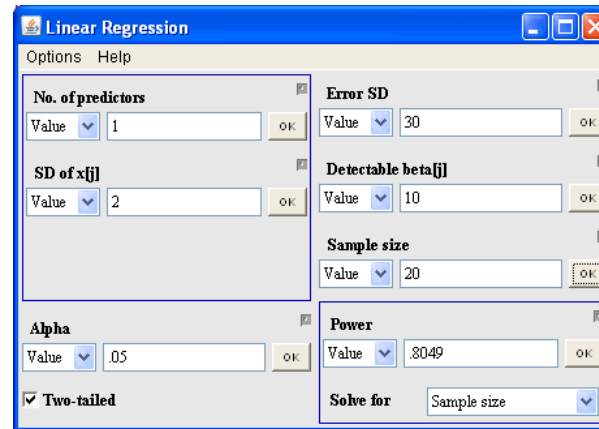
Solution: From Equation 6.18 with $SS_x = N\sigma_x^2$ and $df_\epsilon = 20 - 2 = 18$,

$$\begin{aligned} t_\beta &= \frac{|\beta_1| \sqrt{N} \sigma_x}{\sigma_\epsilon} - t_{0.025, 18} \\ &= \frac{10\sqrt{20} \cdot 4.0825}{30} - 2.10 \\ &= 0.881. \end{aligned}$$

The power, as given by Equation 6.17, is

$$\begin{aligned} \pi &= P(-\infty < t < 0.881) \\ &= 0.805. \end{aligned}$$

From **Piface** > **Linear regression**:



From PASS > Regression > Linear Regression:

PASS: Regression: Linear

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
 Data | Options | Reports | Plot Setup

Find (Solve For): Beta and Power
 Alternative Hypothesis: Two-Sided
 B0 (Slope|H0): 0.0
 Residual Variance Method: S (Std. Dev. of Residuals)
 B (Slope|H1): 10
 SY (Std Deviation of Y):
 Alpha (Significance Level): .05
 R (Correlation):
 Beta (1-Power): 0.10
 S (Std Dev of Residuals): 30
 N (Sample Size): 20
 SX (Std Deviation of X's): 2

PASS: Regression: Linear Output

Linear Regression Power Analysis

Numeric Results for Two-Sided Testing of $B = B_0$ where $B_0 = 0.00$

	Sample Size (N)	Slope (B)	Standard Deviation of X (SX)	Standard Deviation of Residuals (S)	Alpha	Beta
Power	20	10.00	2.00	30.00	0.05000	0.19509

References
 Neter, J., Wasserman, W., and Kutner, M. 1983. Applied Linear Regression Models. Richard D. Irwin, Inc. Chicago, Illinois.

Report Definitions
 Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population. To conserve resources, it should be small.
 B0 is the slope under the null hypothesis.
 B is the slope at which the power is calculated.
 SX is the standard deviation of the X values.
 S is the standard deviation of the residuals.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements
 A sample size of 20 achieves 80% power to detect a change in slope from 0.00 under the null hypothesis to 10.00 under the alternative hypothesis when the standard deviation of the X's is 2.00, the standard deviation of the residuals is 30.00, and the two-sided significance level is 0.05000.

6.2 Logistic Regression

Example 6.5 What sample size is required for an experiment to be analyzed by logistic regression if $H_0 : \beta_1 = 0$ should be rejected in favor of $H_A : \beta_1 \neq 0$ with 90% power when x is dichotomous with associated proportions $p_1 = 0.04$ and $p_2 = 0.08$?

Solution: The odds ratio for the given proportions is

$$OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{0.04/0.96}{0.08/0.92} = 0.479.$$

The required sample size is given by Equation 4.97:

$$n = \left(\frac{z_{0.025} + z_{0.10}}{\ln(0.479)} \right)^2 \left(\frac{1}{0.04(0.96)} + \frac{1}{0.08(0.92)} \right) = 770.$$

From PASS> Regression> Logistic Regression the total sample size is:

PASS: Regression: Logistic

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols
 Data | Options | Reports | Plot Setup

Find (Solve For): X (Independent Variable):

P0 (Baseline Prob Y=1): Hypothesis Test:

Use P1 or Odds Ratio: N (Sample Size):

P1 (Alt. Prob Y=1): OR % N with X=1 (Binary Only):

Odds Ratio (Odds1/Odds0): Alpha (Significance Level):

R-Squared Other X's: Beta (1-Power):

PASS: Regression: Logistic Output

Logistic Regression Power Analysis

Numeric Results

Power	N	Pcmt N X=1	P0	P1	Odds Ratio	R Squared	Alpha	Beta
0.89993	1477	50.000	0.040	0.080	2.087	0.000	0.05000	0.10007

References

Hsieh, F.Y., Block, D.A., and Larsen, M.D. 1998. 'A Simple Method of Sample Size Calculation for Linear and Logistic Regression', *Statistics in Medicine*, Volume 17, pages 1623-1634.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population.
 P0 is the response probability at the mean of X.
 P1 is the response probability when X is increased to one standard deviation above the mean.
 Odds Ratio is the odds ratio when P1 is on top. That is, it is $[P1/(1-P1)]/[P0/(1-P0)]$.
 R-Squared is the R2 achieved when X is regressed on the other independent variables in the regression.
 Alpha is the probability of rejecting a true null hypothesis.
 Beta is the probability of accepting a false null hypothesis.

Summary Statements

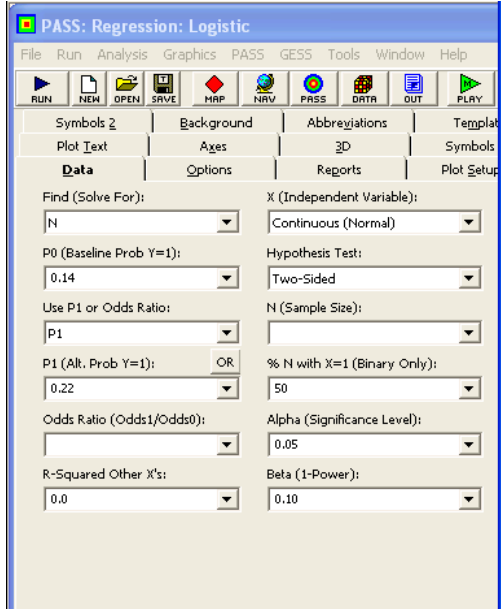
A logistic regression of a binary response variable (Y) on a binary independent variable (X) with a sample size of 1477 observations (of which 50% are in the group X=0 and 50% are in the group X=1) achieves 90% power at a 0.05000 significance level to detect a change in Prob(Y=1) from the baseline value of 0.040 to 0.080. This change corresponds to an odds ratio of 2.087.

Example 6.6 What sample size is required for an experiment to be analyzed by logistic regression if $H_0 : \beta_1 = 0$ should be rejected in favor of $H_A : \beta_1 \neq 0$ with 90% power when x is normally distributed with expected success proportions $p(x = \mu) = 0.14$ and $p(x = \mu + \sigma) = 0.22$.

Solution: From Equation 6.22 the required sample size is

$$n = \frac{(1.96 + 1.282)^2}{0.14(0.86) \left(\ln \left(\frac{0.14/0.86}{0.22/0.78} \right) \right)^2} = 289.$$

From PASS> Regression> Logistic Regression:



PASS: Regression: Logistic

File Run Analysis Graphics PASS GESS Tools Window Help

RUN NEW OPEN SAVE HNP NRV PASS DTRA OUT PLAY

Symbols 2 Background Abbreviations Templat
Plot Text Axes 3D Symbols
Data Options Reports Plot Setup

Find (Solve For): N
X (Independent Variable): Continuous (Normal)

P0 (Baseline Prob Y=1): 0.14
Hypothesis Test: Two-Sided

Use P1 or Odds Ratio: P1
N (Sample Size):

P1 (Alt. Prob Y=1): 0.22 OR % N with X=1 (Binary Only): 50

Odds Ratio (Odds1/Odds0):
Alpha (Significance Level): 0.05

R-Squared Other X's: 0.0
Beta (1-Power): 0.10

PASS: Regression: Logistic Output

Logistic Regression Power Analysis

Numeric Results

Power	N	P0	P1	Odds Ratio	R Squared	Alpha	Beta
0.89912	288	0.140	0.220	1.733	0.000	0.05000	0.10088

References
Hsieh, F. Y., Block, D. A., and Larsen, M. D. 1998. 'A Simple Method of Sample Size Calculation for Linear and Logistic Regression', *Statistics in Medicine*, Volume 17, pages 1623-1634.

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population.
P0 is the response probability at the mean of X.
P1 is the response probability when X is increased to one standard deviation above the mean.
Odds Ratio is the odds ratio when P1 is on top. That is, it is $[P1/(1-P1)]/[P0/(1-P0)]$.
R-Squared is the R2 achieved when X is regressed on the other independent variables in the regression.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A logistic regression of a binary response variable (Y) on a continuous, normally distributed variable (X) with a sample size of 288 observations achieves 90% power at a 0.05000 significance level to detect a change in Prob(Y=1) from the value of 0.140 at the mean of X to 0.220 when X is increased to one standard deviation above the mean. This change corresponds to an odds ratio of 1.733.

Chapter 7

Correlation and Agreement

7.1 Pearson's Correlation

Example 7.1 Determine the number of paired observations required to obtain the following confidence interval for the population correlation:

$$P(0.9 < \rho < 0.99) = 0.95.$$

Solution: The Fisher's Z -transformed confidence interval is

$$\begin{aligned} P(Z_{0.9} < Z_\rho < Z_{0.99}) &= 0.95 \\ P(1.472 < Z_\rho < 2.647) &= 0.95. \end{aligned}$$

Then with $\alpha = 0.05$ in Equation 7.10, the required sample size is

$$\begin{aligned} n &= 4 \left(\frac{1.96}{2.647 - 1.472} \right)^2 + 3 \\ &= 15. \end{aligned}$$

Example 7.2 An experiment is planned to test $H_0 : \rho = 0.9$ versus $H_A : \rho < 0.9$ on the basis of $n = 28$ paired observations. Determine the power of the test to reject H_0 when $\rho = 0.7$.

Solution: Under H_0 following Fisher's transform we have

$$(\mu_Z)_0 = \frac{1}{2} \ln \left(\frac{1 + 0.9}{1 - 0.9} \right) = 1.472$$

and by Equation 7.4

$$\sigma_Z = \frac{1}{\sqrt{28 - 3}} = 0.2.$$

For the one-sided left-tailed test, the critical value of Z that distinguishes the accept/reject regions is given by

$$\begin{aligned} Z_{A/R} &= (\mu_Z)_0 - z_{\alpha} \sigma_z \\ &= 1.472 - z_{0.05} (0.2) \\ &= 1.472 - 1.645 (0.2) \\ &= 1.143. \end{aligned}$$

The corresponding Z value under H_A when $\rho = 0.7$ is

$$(\mu_Z)_A = \frac{1}{2} \ln \left(\frac{1 + 0.7}{1 - 0.7} \right) = 0.867.$$

Then the power to reject H_0 when $\rho = 0.7$ is

$$\begin{aligned} \pi &= \Phi(-\infty < Z < Z_{A/R}; (\mu_Z)_A, \sigma_Z) \\ &= \Phi(-\infty < Z < 1.143; 0.867, 0.2) \\ &= \Phi(-\infty < z < 1.38) \\ &= 0.916. \end{aligned}$$

From PASS> Correlations> Correlations: One:

The screenshot shows the PASS software interface. The left window, titled 'PASS: Correlation: 1', displays the input parameters for a one-sided correlation power analysis. The right window, titled 'PASS: Correlation: 1 Output', displays the results and definitions.

Input Parameters (Left Window):

- Find (Solve For): Beta and Power
- R0 (Baseline Correlation): 0.9
- Alternative Hypothesis: $H_a: R_0 > R_1$
- R1 (Alternative Correlation): 0.7
- Alpha (Significance Level): .05
- N (Sample Size): 28
- Beta (1-Power): (empty)

Output Results (Right Window):

One Correlation Power Analysis

Numeric Results when $H_a: R_0 > R_1$

Power	N	Alpha	Beta	R0	R1
0.92319	28	0.05000	0.07681	0.90000	0.70000

References

- Graybill, Franklin. 1961. An Introduction to Linear Statistical Models. McGraw-Hill. New York, New York.
- Guenther, William C. 1977. 'Desk Calculation of Probabilities for the Distribution of the Sample Correlation Coefficient', The American Statistician, Volume 31, Number 1, pages 45-48.
- Zar, Jerrold H. 1984. Biostatistical Analysis. Second Edition. Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions

- Power is the probability of rejecting a false null hypothesis. It should be close to one.
- N is the size of the sample drawn from the population. To conserve resources, it should be small.
- Alpha is the probability of rejecting a true null hypothesis. It should be small.
- Beta is the probability of accepting a false null hypothesis. It should be small.
- R0 is the value of the population correlation under the null hypothesis.
- R1 is the value of the population correlation under the alternative hypothesis.

Summary Statements

A sample size of 28 achieves 92% power to detect a difference of 0.20000 between the null hypothesis correlation of 0.90000 and the alternative hypothesis correlation of 0.70000 using a one-sided hypothesis test with a significance level of 0.05000.

Example 7.3 Find the power to reject $H_0 : \rho_1 = \rho_2$ in favor of $H_A : \rho_1 \neq \rho_2$ when $\rho_1 = 0.99$, $\rho_2 = 0.95$, and $n_1 = n_2 = 30$.

PASS: Correlations: 2

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols
 Data | Options | Reports | Plot Setup

Find (Solve For): Beta and Power

Alternative Hypothesis: $H_A: R_1 \neq R_2$

R1 (Correlation Group 1): 0.99

N1 (Sample Size Group 1): 30

R2 (Correlation Group 2): 0.95

N2 (Sample Size Group 2): Use R

Alpha (Significance Level): .05

R (Sample Allocation Ratio): 1.0

Beta (1-Power):

PASS: Correlations: 2 Output

Two Correlations Power Analysis

Numeric Results when $H_A: R_1 \neq R_2$

Power	N1	N2	Allocation Ratio	R1	R2	Difference (R1-R2)	Alpha	Beta
0.84945	30	30	1.000	0.99000	0.95000	0.04000	0.05000	0.15055

References

Zar, Jerrold H. 1984. Biostatistical Analysis. Second Edition. Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one. N1 and N2 are the sizes of the samples drawn from the two populations. To conserve resources, it should be small.

Allocation Ratio is N_2/N_1 so that $N_2 = N_1 \times R$.

Alpha is the probability of rejecting a true null hypothesis. It should be small.

Beta is the probability of accepting a false null hypothesis. It should be small.

R1 is the value of both correlations under the null hypothesis.

R2 is the correlation in group two under the alternative hypothesis.

Summary Statements

Group sample sizes of 30 and 30 achieve 85% power to detect a difference of 0.04000 between the null hypothesis that both group correlations are 0.99000 and the alternative hypothesis that the correlation in group 2 is 0.95000 using a two-sided z test (which uses Fisher's z-transformation) with a significance level of 0.05000.

Solution: The Fisher-transformed difference between the two correlations under H_A is

$$\begin{aligned}
 \Delta Z &= Z_1 - Z_2 \\
 &= \frac{1}{2} \ln \left(\frac{1 + 0.99}{1 - 0.99} \right) - \frac{1}{2} \ln \left(\frac{1 + 0.95}{1 - 0.95} \right) \\
 &= 0.815.
 \end{aligned}$$

From Equations 7.11 and 7.14 with $\alpha = 0.05$ the power is

$$\begin{aligned}
 \pi &= \Phi \left(-\infty < z < \left(\frac{0.815}{\sqrt{\frac{2}{30-3}}} - 1.96 \right) \right) \\
 &= \Phi (-\infty < z < 1.03) \\
 &= 0.85.
 \end{aligned}$$

From **PASS > Correlations > Correlations: Two:**

Example 7.4 What sample size should be drawn from two populations to perform the two-sample test for correlation ($H_0 : \rho_1 = \rho_2$ versus $H_A : \rho_1 \neq \rho_2$) with 90% power to reject H_0 when $\rho_1 = 0.9$ and $\rho_2 = 0.8$?

Solution: The Fisher-transformed difference between the two correlations is

$$\begin{aligned}\Delta Z &= Z_1 - Z_2 \\ &= \frac{1}{2} \ln \left(\frac{1 + 0.9}{1 - 0.9} \right) - \frac{1}{2} \ln \left(\frac{1 + 0.8}{1 - 0.8} \right) \\ &= 0.374.\end{aligned}$$

From Equation 7.15 with $\alpha = 0.05$ and $\beta = 0.10$ the required common sample size is

$$\begin{aligned}n &= 2 \left(\frac{1.96 + 1.28}{0.374} \right)^2 + 3 \\ &= 154.\end{aligned}$$

From PASS> Correlations> Correlations: Two:

PASS: Correlations: 2

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols Background Abbreviations Templates
 Plot Text Axes 3D Symbols
 Data Options Reports Plot Setup

Find (Solve For): Alternative Hypothesis:

R1 (Correlation Group 1): N1 (Sample Size Group 1):

R2 (Correlation Group 2): N2 (Sample Size Group 2):

Alpha (Significance Level): R (Sample Allocation Ratio):

Beta (1-Power):

PASS: Correlations: 2 Output

Two Correlations Power Analysis

Numeric Results when Ha: R1<>R2

Power	N1	N2	Allocation Ratio	R1	R2	Difference (R1-R2)	Alpha	Beta
0.90084	154	154	1.000	0.90000	0.80000	0.10000	0.05000	0.09916

References
Zar, Jerrold H. 1984. Biostatistical Analysis. Second Edition. Prentice-Hall. Englewood Cliffs, New Jersey.

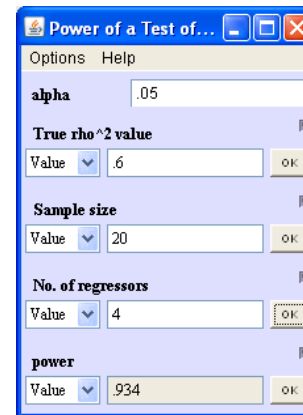
Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 and N2 are the sizes of the samples drawn from the two populations. To conserve resources, it should be small.
Allocation Ratio is N2/N1 so that N2 = N1 x R.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
R1 is the value of both correlations under the null hypothesis.
R2 is the correlation in group two under the alternative hypothesis.

Summary Statements
Group sample sizes of 154 and 154 achieve 90% power to detect a difference of 0.10000 between the null hypothesis that both group correlations are 0.90000 and the alternative hypothesis that the correlation in group 2 is 0.80000 using a two-sided z test (which uses Fisher's z-transformation) with a significance level of 0.05000.

Example 7.5 Determine the power to reject $H_0 : \rho^2 = 0$ when in fact $\rho^2 = 0.6$ based on a sample of $n = 20$ observations taken with four random covariates.

Solution: The regression model for y as a function of the four predictors will have $df_{model} = k = 4$ model degrees of freedom and $df_{\epsilon} = n - k - 1 = 20 - 4 - 1 = 15$ error degrees of freedom. The F distribution noncentrality parameter from Equation 7.19 with $\rho^2 = 0.6$ is

$$\phi = 20 \frac{0.6}{1 - 0.6} = 30.$$



From Equation 7.18 we have

$$\begin{aligned} F_{0.95} &= F_{1-\pi, 30} \\ 3.056 &= F_{0.024, 30}, \end{aligned}$$

so the power is $\pi = 1 - 0.024 = 0.976$.

From **Piface > R-square (multiple correlation)**: (I can't explain the discrepancy between my solution and the solution from Piface. There is a comment in Piface's **Help > This Dialog** that there is a discrepancy between it and the references.)

7.2 Intraclass Correlation

Example 7.6 An experiment will be performed to determine the single-rater intraclass correlation in a one-way design with $r = 2$ observations per subject. How many subjects must be sampled if the desired confidence interval for ICC is $P(0.7 < ICC < 0.9) = 0.95$?

Solution: By Equation 7.31, the desired confidence interval for ICC transforms into the following confidence interval for Z_{ICC} :

$$P(0.867 < Z_{ICC} < 1.472) = 0.95.$$

Then, from Equation 7.36 with $r = 2$ observations per subject and $\alpha = 0.05$, the number of subjects required is

$$\begin{aligned} n &= 4 \left(\frac{1.96}{1.472 - 0.867} \right)^2 + \frac{3}{2} \\ &= 44. \end{aligned}$$

Example 7.7 Confirm the answer to Example 7.6 using the method of Donner and Koval.

Solution: Assuming that $\widehat{ICC} = 0.8$, the sample size according to Donner and Koval is given by Equation 7.38:

$$\begin{aligned} n &= \frac{8}{2(2-1)} \left(\frac{1.96(1-0.8)(1+(2-1)0.8)}{0.9-0.7} \right)^2 \\ &= 50, \end{aligned}$$

which is in reasonable agreement with the sample size determined by the Fisher's transformation method.

Example 7.8 How many subjects are required in an experiment to reject $H_0 : ICC = 0.6$ with 80% power when $ICC = 0.8$ and two raters will rate each subject? Confirm the sample size by calculating the exact power.

Solution: From Equation 7.31 the Z_{ICC} values corresponding to $ICC = 0.6$ and $ICC = 0.8$ are $Z_0 = 0.693$ and $Z_1 = 1.099$, respectively. From Equation 7.42 with $r = 2$ and $\alpha = 0.05$, the approximate sample size is

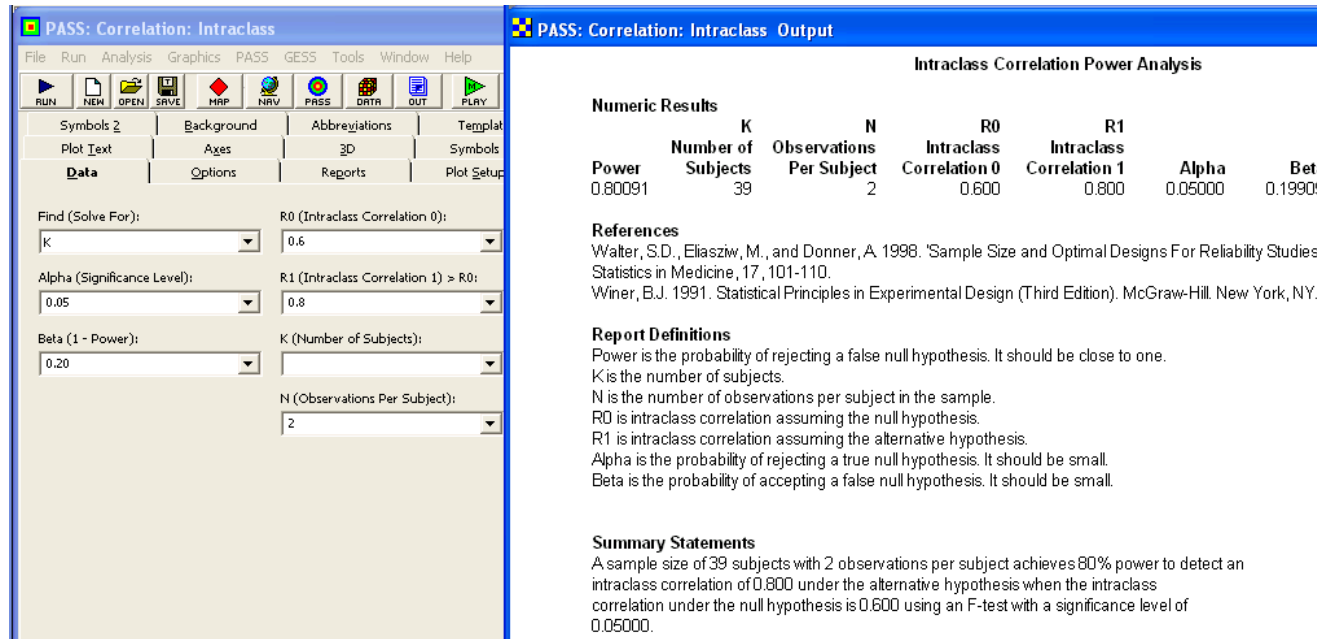
$$\begin{aligned} n &= \left(\frac{z_{0.05} + z_{0.20}}{Z_1 - Z_0} \right)^2 + \frac{3}{2} \\ &= \left(\frac{1.645 + 0.84}{1.099 - 0.693} \right)^2 + \frac{3}{2} \\ &= 39. \end{aligned}$$

The exact power is given by Equation 7.39 where the F distribution has $df_1 = 39 - 1 = 38$ numerator degrees of freedom and $df_2 = 39(2 - 1) = 39$ denominator degrees of freedom. The power is given by

$$\begin{aligned} \pi &= P \left(\frac{1 + 2 \left(\frac{0.6}{1-0.6} \right)}{1 + 2 \left(\frac{0.8}{1-0.8} \right)} F_{0.95} < F < \infty \right) \\ &= P(0.760 < F < \infty) \\ &= 0.80, \end{aligned}$$

which is in excellent agreement with the target power.

From **PASS> Correlation> Intraclass Correlation:**



7.3 Cohen's Kappa

Example 7.9 How many units should two operators evaluate in an attribute inspection agreement experiment to be analyzed using Cohen's κ if the true value of the unknown κ must be determined to within ± 0.10 with 95% confidence? A preliminary experiment indicated that $\kappa \simeq 0.85$ and $p_e \simeq 0.5$.

Solution: With $\alpha = 0.05$ and $\delta = 0.10$ in Equation 7.55, the required sample size is

$$n = \frac{0.85(1 - 0.85)}{1 - 0.5} \left(\frac{1.96}{0.10} \right)^2 = 98.$$

Example 7.10 Calculate the power to reject $H_0 : \kappa = 0.4$ in favor of $H_A : \kappa > 0.4$ when $\kappa = 0.7$ if a sample of size $n = 70$ is allocated to $k = 3$ categories in the ratio 0.4 : 0.5 : 0.1.

Solution: The expected chance agreement by Equation 7.46 is $p_e = 0.4^2 + 0.5^2 + 0.1^2 = 0.42$. The two κ values of interest have intermediate values not covered by the large- or small- κ approximations, so it is necessary to estimate $\sigma_{\hat{\kappa}}$ using Equation 7.47. Under H_0 with $\kappa = 0.4$ and $p_e = 0.42$ in Equation 7.44, we have

$$\begin{aligned}
 p_o &= 0.4(1 - 0.42) + 0.42 \\
 &= 0.652,
 \end{aligned}$$

so

$$\begin{aligned}
 \sigma_{\hat{\kappa}_0} &\simeq \frac{1}{1 - 0.42} \sqrt{\frac{0.652(1 - 0.652)}{70}} \\
 &\simeq 0.0982.
 \end{aligned}$$

Under H_A with $\kappa = 0.7$ we have

$$\begin{aligned} p_o &= 0.7(1 - 0.42) + 0.42 \\ &= 0.826, \end{aligned}$$

so

$$\begin{aligned} \sigma_{\hat{\kappa}_1} &\simeq \frac{1}{1 - 0.42} \sqrt{\frac{0.826(1 - 0.826)}{70}} \\ &\simeq 0.0781. \end{aligned}$$

Then with $\alpha = 0.05$, z_β is given by Equation 7.58:

$$\begin{aligned} z_\beta &= \frac{(0.7 - 0.4) - 1.645(0.0982)}{0.0781} \\ &= 1.77 \end{aligned}$$

and the power is given by Equation 7.57:

$$\begin{aligned} \pi &= \Phi(-\infty < z < 1.77) \\ &= 0.962. \end{aligned}$$

Example 7.11 An experiment is to be performed to test for agreement between two methods of categorizing a dichotomous response. The hypotheses to be tested are $H_0 : \kappa = 0$ versus $H_A : \kappa > 0$ where κ is Cohen's kappa. How many units must be inspected if the test should have 90% power to reject H_0 when $\kappa = 0.40$ and the total number of units to be inspected is evenly split between the two categories?

Solution: Because the units will be balanced between the two categories, the agreement expected by chance from Equation 7.51 is $p_e \simeq 0.5$. With $\alpha = 0.05$, $\pi = 0.90$, $\beta = 1 - \pi = 0.10$, and $\delta = 0.4 - 0 = 0.4$ in Equation 7.59, the required sample size is

$$n \simeq \frac{0.5}{1 - 0.5} \left(\frac{z_{0.05} + z_{0.10}}{\delta} \right)^2 = \left(\frac{1.645 + 1.282}{0.40} \right)^2 = 54.$$

Example 7.12 An experiment is to be performed to test for agreement between two raters using a categorical four-state response. The hypotheses to be tested are $H_0 : \kappa = 0.8$ versus $H_A : \kappa > 0.8$. How many units must be inspected if the test should have 90% power to reject H_0 when $\kappa = 0.9$? The units to be inspected are evenly balanced across the four categories.

Solution: From Equation 7.51 with $k = 4$ categories, $p_e \simeq 0.25$. With $\alpha = 0.05$, $\beta = 1 - \pi = 0.10$, and $\delta = 0.9 - 0.8 = 0.1$ in Equation 7.60, the required sample size is

$$\begin{aligned} n &\simeq \frac{1}{1 - 0.25} \left(\frac{1.645\sqrt{0.8 \times 0.2} + 1.282\sqrt{0.9 \times 0.1}}{0.9 - 0.8} \right)^2 \\ &\simeq 145. \end{aligned}$$

7.4 Receiver Operating Characteristic (ROC) Curves

Example 7.13 What sample size is required to estimate the value of an ROC curve's AUC to within ± 0.05 with 95% confidence if the AUC value is expected to be about 90%?

Solution: The desired confidence interval has the form

$$P\left(\widehat{AUC} - 0.05 < AUC < \widehat{AUC} + 0.05\right) = 0.95.$$

With $\alpha = 0.05$, $AUC = 0.90$, and $\delta = 0.05$ in Equation 7.66, the required sample size is

$$\begin{aligned} n &\approx \frac{1 - AUC}{2} \left(\frac{z_{0.025}}{\delta}\right)^2 \\ &\approx \frac{1 - 0.90}{2} \left(\frac{1.96}{0.05}\right)^2 \\ &\approx 77. \end{aligned}$$

That is, about 77 positives and 77 negatives are required. The large-sample and large AUC assumptions are reasonably satisfied, so this approximate sample size should be accurate.

From **PASS**> **Diagnostic Tests**> **ROC Curve - 1 Test**:

PASS: ROC Curve: 1

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 Background Abbreviations Template
 Plot Text Axes 3D Symbols 1
 Data Options Reports Plot Setup

Find (Solve For): Alternative Hypothesis:

AUC0 (Area Under Curve): Lower FPR: Upper FPR:

AUC1: B (SD Ratio):

Alpha (Significance Level): N+ (Size of Positive Group):

Beta (1-Power): N- (Size of Negative Group):

Type of Data: R (Sample Allocation Ratio):

Opt 118 Template Id:

31
32

Variable Info Sheet1

7 15 Search for a specified cell value.

PASS: ROC Curve: 1 Output

One ROC Curve Power Analysis

Numeric Results for Testing AUC0 = AUC1 with Continuous Data
 Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 1.0. B = 1.000. Allocation Ratio = 1.000.

Power	N+	N-	AUC0'	AUC1'	Diff'	AUC0	AUC1	Diff	Alpha	Beta
0.5065	79	79	0.9000	0.9500	0.0500	0.9000	0.9500	0.0500	0.0500	0.4935

References
 Hanley, J. A. and McNeil, B. J. 1983. 'A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases.' *Radiology*, 148, 839-843. September, 1983.
 Obuchowski, N. and McClish, D. 1997. 'Sample Size Determination for Diagnostic Accuracy Studies Involving Binormal ROC Curve Indices.' *Statistics in Medicine*, 16, pages 1529-1542.

Report Definitions
 Power is the probability of rejecting a false null hypothesis.
 N+ is the sample size from the positive (diseased) population.
 N- is the sample size from the negative (non-diseased) population.
 Alloc Ratio is the Sample Allocation Ratio (R = N- / N+).
 AUC0' is the adjusted area under the ROC curve under the null hypothesis.
 AUC1' is the adjusted area under the ROC curve under the alternative hypothesis.
 Diff' is AUC1' - AUC0'. This is the adjusted difference to be detected.
 AUC0 is the actual area under the ROC curve under the null hypothesis.
 AUC1 is the actual area under the ROC curve under the alternative hypothesis.
 Diff is AUC1 - AUC0. This is the difference to be detected.
 Alpha is the probability of rejecting a true null hypothesis.
 Beta is the probability of accepting a false null hypothesis.
 FPR1, FPR2 are the lower and upper bounds on the false positive rates.
 B is the ratio of the standard deviations of the negative and positive groups.

Summary Statements
 A sample of 79 from the positive group and 79 from the negative group achieve 51% power to detect a difference of 0.0500 between the area under the ROC curve (AUC) under the null hypothesis of 0.9000 and an AUC under the alternative hypothesis of 0.9500 using a two-sided z-test at a significance level of 0.0500. The data are continuous responses. The AUC is computed between false positive rates of 0.000 and 1.000. The ratio of the standard deviation of the responses in the negative group to the standard deviation of the responses in the positive group is 1.000.

Example 7.14 What sample size is required to reject $H_0 : AUC = 0.9$ in favor of $H_A : AUC \neq 0.9$ with 90% power when $AUC = 0.95$?

Solution: With $\alpha = 0.05$ in Equation 7.67, the required sample size is approximately

$$\begin{aligned}
 n &= \left(\frac{\sqrt{\frac{1-0.90}{2}} \approx 0.025 + \sqrt{\frac{1-0.95}{2}} \approx 0.10}{0.95 - 0.90} \right)^2 \\
 &= \left(\frac{\sqrt{\frac{1-0.90}{2}} 1.96 + \sqrt{\frac{1-0.95}{2}} 1.282}{0.95 - 0.90} \right)^2 \\
 &= 165.
 \end{aligned}$$

From PASS> Diagnostic Tests> ROC Curve - 1 Test:

PASS: ROC Curve: 1

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
Data | Options | Reports | Plot Setup

Find (Solve For): Alternative Hypothesis:

AUC0 (Area Under Curve): Lower FPR: Upper FPR:

AUC1: B (SD Ratio):

Alpha (Significance Level): N+ (Size of Positive Group):

Beta (1-Power): N- (Size of Negative Group):

Type of Data: R (Sample Allocation Ratio):

PASS: ROC Curve: 1 Output

One ROC Curve Power Analysis

Numeric Results for Testing AUC0 = AUC1 with Continuous Data
 Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 1.0. B = 1.000. Allocation Ratio = 1.000.

Power	N+	N-	AUC0'	AUC1'	Diff'	AUC0	AUC1	Diff	Alpha	Beta
0.9001	167	167	0.9000	0.9500	0.0500	0.9000	0.9500	0.0500	0.0500	0.0999

References
 Hanley, J. A. and McNeil, B. J. 1983. 'A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases.' *Radiology*, 148, 839-843. September, 1983.
 Obuchowski, N. and McClish, D. 1997. 'Sample Size Determination for Diagnostic Accuracy Studies Involving Binormal ROC Curve Indices.' *Statistics in Medicine*, 16, pages 1529-1542.

Report Definitions
 Power is the probability of rejecting a false null hypothesis.
 N+ is the sample size from the positive (diseased) population.
 N- is the sample size from the negative (non-diseased) population.
 Alloc Ratio is the Sample Allocation Ratio (R = N- / N+).
 AUC0' is the adjusted area under the ROC curve under the null hypothesis.
 AUC1' is the adjusted area under the ROC curve under the alternative hypothesis.
 Diff' is AUC1 - 'AUC0. This is the adjusted difference to be detected.
 AUC0 is the actual area under the ROC curve under the null hypothesis.
 AUC1 is the actual area under the ROC curve under the alternative hypothesis.
 Diff is AUC1 - AUC0. This is the difference to be detected.
 Alpha is the probability of rejecting a true null hypothesis.
 Beta is the probability of accepting a false null hypothesis.
 FPR1, FPR2 are the lower and upper bounds on the false positive rates.
 B is the ratio of the standard deviations of the negative and positive groups.

Summary Statements
 A sample of 167 from the positive group and 167 from the negative group achieve 90% power to detect a difference of 0.0500 between the area under the ROC curve (AUC) under the null hypothesis of 0.9000 and an AUC under the alternative hypothesis of 0.9500 using a two-sided z-test at a significance level of 0.0500. The data are continuous responses. The AUC is computed between false positive rates of 0.000 and 1.000. The ratio of the standard deviation of the responses in the negative group to the standard deviation of the responses in the positive group is 1.000.

Example 7.15 What sample size is required to reject $H_0 : AUC = 0.5$ versus $H_A : AUC > 0.5$ with 90% power when $AUC = 0.75$?

Solution: With $\beta = 0.10$ when $AUC = 0.75$ in Equation 7.67, the required sample size is approximately

$$\begin{aligned}
 n &= \left(\frac{\sqrt{\frac{1}{6}} z_{0.05} + \sqrt{\frac{1-0.75}{2}} z_{0.10}}{0.75 - 0.50} \right)^2 \\
 &= \left(\frac{\sqrt{\frac{1}{6}} 1.645 + \sqrt{\frac{1-0.75}{2}} 1.282}{0.75 - 0.50} \right)^2 \\
 &= 21.
 \end{aligned}$$

The large-sample assumption is only marginally satisfied, so this sample size may be somewhat inaccurate.

From PASS> Diagnostic Tests> ROC Curve - 1 Test:

PASS: ROC Curve: 1

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols Background Abbreviations Template
 Plot Text Axes 3D Symbols
 Data Options Reports Plot Setup

Find (Solve For): Alternative Hypothesis:
 AUC0 (Area Under Curve): Lower FPR: Upper FPR:
 AUC1: B (SD Ratio):
 Alpha (Significance Level): N+ (Size of Positive Group):
 Beta (1-Power): N- (Size of Negative Group):
 Type of Data: R (Sample Allocation Ratio):

Opt 1 Template Id:

31					
32					

Variable Info Sheet1

7 15 This is the spreadsheet that lets you enter and edit your data.

PASS: ROC Curve: 1 Output

One ROC Curve Power Analysis

Numeric Results for Testing AUC0 = AUC1 with Continuous Data
 Test Type = One-Sided. FPR1 = 0.0. FPR2 = 1.0. B = 1.000. Allocation Ratio = 1.000.

Power	N+	N-	AUC0'	AUC1'	Diff'	AUC0	AUC1	Diff	Alpha	Beta
0.9033	20	20	0.5000	0.7500	0.2500	0.5000	0.7500	0.2500	0.0500	0.0967

References
 Hanley, J. A. and McNeil, B. J. 1983. 'A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases.' *Radiology*, 148, 839-843. September, 1983.
 Obuchowski, N. and McClish, D. 1997. 'Sample Size Determination for Diagnostic Accuracy Studies Involving Binormal ROC Curve Indices.' *Statistics in Medicine*, 16, pages 1529-1542.

Report Definitions
 Power is the probability of rejecting a false null hypothesis.
 N+ is the sample size from the positive (diseased) population.
 N- is the sample size from the negative (non-diseased) population.
 Alloc Ratio is the Sample Allocation Ratio (R = N- / N+).
 AUC0' is the adjusted area under the ROC curve under the null hypothesis.
 AUC1' is the adjusted area under the ROC curve under the alternative hypothesis.
 Diff' is AUC1' - AUC0'. This is the adjusted difference to be detected.
 AUC0 is the actual area under the ROC curve under the null hypothesis.
 AUC1 is the actual area under the ROC curve under the alternative hypothesis.
 Diff is AUC1 - AUC0. This is the difference to be detected.
 Alpha is the probability of rejecting a true null hypothesis.
 Beta is the probability of accepting a false null hypothesis.
 FPR1, FPR2 are the lower and upper bounds on the false positive rates.
 B is the ratio of the standard deviations of the negative and positive groups.

Summary Statements
 A sample of 20 from the positive group and 20 from the negative group achieve 90% power to detect a difference of 0.2500 between the area under the ROC curve (AUC) under the null hypothesis of 0.5000 and an AUC under the alternative hypothesis of 0.7500 using a one-sided z-test at a significance level of 0.0500. The data are continuous responses. The AUC is computed between false positive rates of 0.000 and 1.000. The ratio of the standard deviation of the responses in the negative group to the standard deviation of the responses in the positive group is 1.000.

7.5 Bland-Altman Plots

Example 7.16 What minimum sample size is required to demonstrate the agreement between two methods to measure length by the Bland-Altman method if the limits of agreement are $LOA_{U/L} = \pm 3cm$ and the standard deviation of the differences had been estimated to be $\hat{\sigma}_d = 0.65cm$ from historical data? Assume that the limits of agreement must cover 99% of the samples with 95% confidence and that there is no bias between the two methods, i.e., $\mu_d = 0$.

Solution: The two-sided normal distribution tolerance interval factor k_2 is given by

$$\begin{aligned}
 k_2 &= \frac{LOA}{\hat{\sigma}_d} \\
 &= \frac{3cm}{0.65cm} \\
 &= 4.615.
 \end{aligned}$$

With $\alpha = 0.05$ and $Y = 0.99$ in Appendix E, Table E.7, the smallest sample size that gives $k_2 \leq 4.615$ is $n = 10$.

Chapter 8

Designed Experiments

8.1 One-Way Fixed Effects ANOVA

Example 8.1 In a one-way classification design with four treatments and five observations per treatment, determine the power of the ANOVA to reject H_0 if the treatment means are $\mu_i = \{40, 55, 55, 50\}$. The four populations are expected to be normal and homoscedastic with $\sigma_\epsilon = 8$.

Solution: The grand mean is $\bar{\mu} = 50$ so the treatment biases relative to the grand mean are $\tau_i = \{-10, 5, 5, 0\}$. From Equation 8.2 the F distribution noncentrality parameter is,

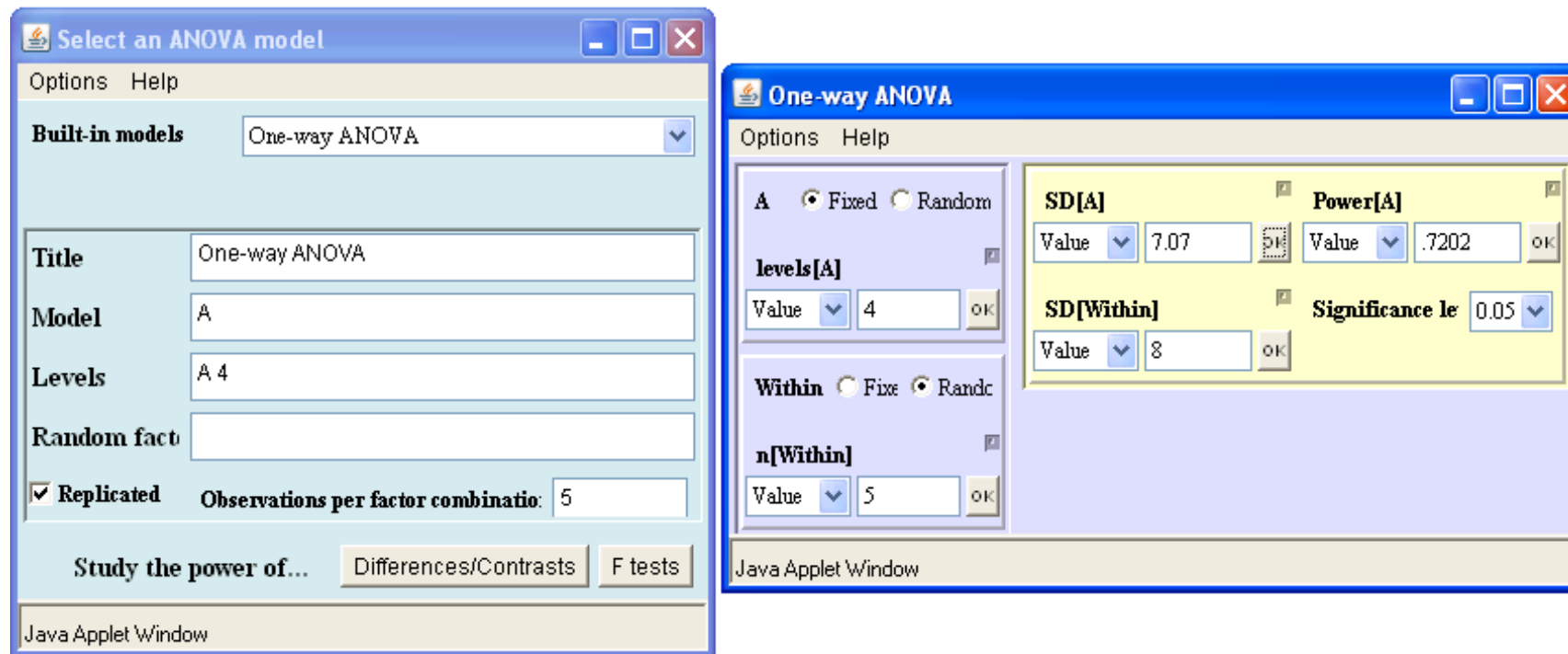
$$\phi = \frac{n \sum_{i=1}^k \tau_i^2}{\sigma_\epsilon^2} = \frac{5 \left((-10)^2 + (5)^2 + (5)^2 + (0)^2 \right)}{8^2} = 11.72.$$

The F statistic will have $df_{treatments} = 4 - 1 = 3$ and $df_\epsilon = 4(5 - 1) = 16$ degrees of freedom. The power is 72% as determined from Equation 8.1:

$$F_{0.95} = 3.239 = F_{0.280, 11.72}.$$

From `Piface > Balanced ANOVA (any model) > One-way ANOVA` with:

$$s_A = \sqrt{\frac{\left((-10)^2 + (5)^2 + (5)^2 + (0)^2 \right)}{4 - 1}} = 7.07$$



From PASS> Means> Many Means> ANOVA: One-Way:

The image shows two windows from the Minitab PASS software. The left window is the 'PASS: Means: ANOVA: One Way' dialog box, and the right window is the 'PASS: Means: ANOVA: One Way Output' results window.

Dialog Box Settings:

- Find (Solve For): Beta and Power
- k (Number of Groups): 4
- Hypothesized Means: 40 50 55 55
- Alpha (Significance Level): .05
- n (Sample Size Multiplier): 5
- Beta (1-Power): 0.1
- Group Sample Size Pattern: Equal
- S (Std Dev of Subjects): 8

Output Window: One Way ANOVA Power Analysis

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.72038	5.00	4	20	0.05000	0.27962	6.12	8.00	0.7655

Report Definitions

- Power is the probability of rejecting a false null hypothesis. It should be close to one.
- n is the average group sample size.
- k is the number of groups.
- Total N is the total sample size of all groups combined.
- Alpha is the probability of rejecting a true null hypothesis. It should be small.
- Beta is the probability of accepting a false null hypothesis. It should be small.
- Sm is the standard deviation of the group means under the alternative hypothesis.
- Standard deviation is the within group standard deviation.
- The Effect Size is the ratio of Sm to standard deviation.

Summary Statements

In a one-way ANOVA study, sample sizes of 5, 5, 5, and 5 are obtained from the 4 groups whose means are to be compared. The total sample of 20 subjects achieves 72% power to detect differences among the means versus the alternative of equal means using an F test with a 0.05000 significance level. The size of the variation in the means is represented by their standard deviation which is 6.12. The common standard deviation within a group is assumed to be 8.00.

MINITAB> Stat> Power and Sample Size> One-Way ANOVA cannot be used to solve this problem because it does not allow specification of the individual treatment means or the standard deviation of the treatment means. The steps required to calculate the power from the model and error degrees of freedom and the noncentrality parameter using MINITAB> Calc> Probability Distributions> F are:

```
MTB invcdf 0.95;
SUBC f 3 16.
```

F distribution with 3 DF in numerator and 16 DF in denominator

```
P(~X~==~x~)      x
0.95  3.23887
```

```
MTB cdf 3.23887;
SUBC f 3 16 11.72.
```

F distribution with 3 DF in numerator and 16 DF in denominator and noncentrality parameter 11.72

```
x  P(~X~==~x~)
3.23887  0.279568
```

```
Power = 1 - 0.280 = 0.720
```

Example 8.2 Determine the power of the ANOVA to reject H_0 in a one-way classification design with four treatments and five observations per treatment if the treatment biases from the grand mean are $\tau_i = \{-12, 12, 0, 0\}$. The four populations are expected to be normal and homoscedastic with $\sigma_\epsilon = 8$.

Solution: From Equation 8.5 with $\delta = 24$, the F distribution noncentrality parameter is

$$\phi = \frac{n}{2} \left(\frac{\delta}{\sigma_\epsilon} \right)^2 = \frac{5}{2} \left(\frac{24}{8} \right)^2 = 22.5.$$

The F statistic will have $df_{treatments} = 4 - 1 = 3$ and $df_\epsilon = 4(5 - 1) = 16$ degrees of freedom. The power is 95.4% as determined from Equation 8.1

$$F_{0.95} = 3.239 = F_{0.046, 22.5}.$$

From Piface > **Balanced ANOVA (any model)** > **One-way ANOVA** with:

$$s_A = \sqrt{\frac{((-12)^2 + (12)^2 + (0)^2 + (0)^2)}{4 - 1}} = 9.80$$

The image shows two screenshots of Minitab software windows. The left window is titled "Select an ANOVA model" and has a menu bar with "Options" and "Help". Under "Built-in models", "One-way ANOVA" is selected in a dropdown menu. Below this, there are input fields for "Title" (One-way ANOVA), "Model" (A), "Levels" (A 4), and "Random fact" (empty). A checkbox for "Replicated" is checked, and "Observations per factor combination" is set to 5. At the bottom, there are buttons for "Study the power of...", "Differences/Contrasts", and "F tests". The right window is titled "One-way ANOVA" and also has a menu bar with "Options" and "Help". It shows settings for factor A: "Fixed" is selected, "levels[A]" is 4, "Within" is "Random", and "n[Within]" is 5. On the right side, "SD[A]" is 9.8, "Power[A]" is .9537, "SD[Within]" is 8, and "Significance level" is 0.05. Both windows are identified as "Java Applet Window" at the bottom.

From MINITAB > **Stat** > **Power and Sample Size** > **One-Way ANOVA**:

```
MTB > Power;
SUBC> OneWay 4;
SUBC> Sample 5;
SUBC> MaxDifference 24;
SUBC> Sigma 8;
SUBC> GPCurve.
```

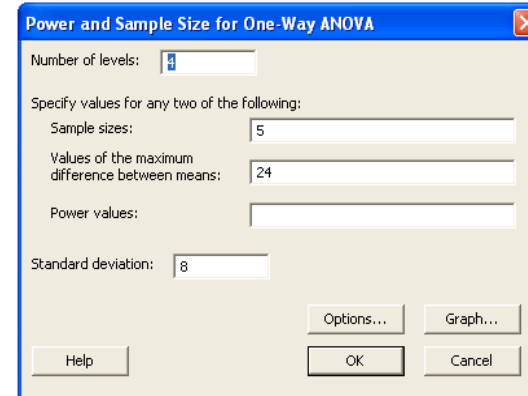
Power and Sample Size

One-way ANOVA

Alpha = 0.05 Assumed standard deviation = 8 Number of Levels = 4

SS	Sample	Maximum
Means	Size	Power Difference
288	5	0.953578 24

The sample size is for each level.



From PASS> Means> Many Means> ANOVA: One-Way:

One Way ANOVA Power Analysis

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.95358	5.00	4	20	0.05000	0.04642	8.49	8.00	1.0607

References

Desu, M. M. and Raghavarao, D. 1990. Sample Size Methodology. Academic Press. New York.
 Fleiss, Joseph L. 1986. The Design and Analysis of Clinical Experiments. John Wiley & Sons. New York.
 Kirk, Roger E. 1982. Experimental Design: Procedures for the Behavioral Sciences. Brooks/Cole. Pacific Grove, California.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
 n is the average group sample size.
 k is the number of groups.
 Total N is the total sample size of all groups combined.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.
 Sm is the standard deviation of the group means under the alternative hypothesis.
 Standard deviation is the within group standard deviation.
 The Effect Size is the ratio of Sm to standard deviation.

Summary Statements

In a one-way ANOVA study, sample sizes of 5, 5, 5, and 5 are obtained from the 4 groups whose means are to be compared. The total sample of 20 subjects achieves 95% power to detect differences among the means versus the alternative of equal means using an F test with a 0.05000 significance level. The size of the variation in the means is represented by their standard deviation which is 8.49. The common standard deviation within a group is assumed to be 8.00.

Example 8.3 In a one-way classification design with four treatments and five observations per treatment, determine the power of the ANOVA to reject H_0 if the treatment biases from the grand mean are $\tau_i = \{18, -6, -6, -6\}$. The four populations are expected to be normal and homoscedastic with $\sigma_\epsilon = 8$.

Solution: From Equation 8.6 with $\delta = 24$, the F distribution noncentrality parameter is

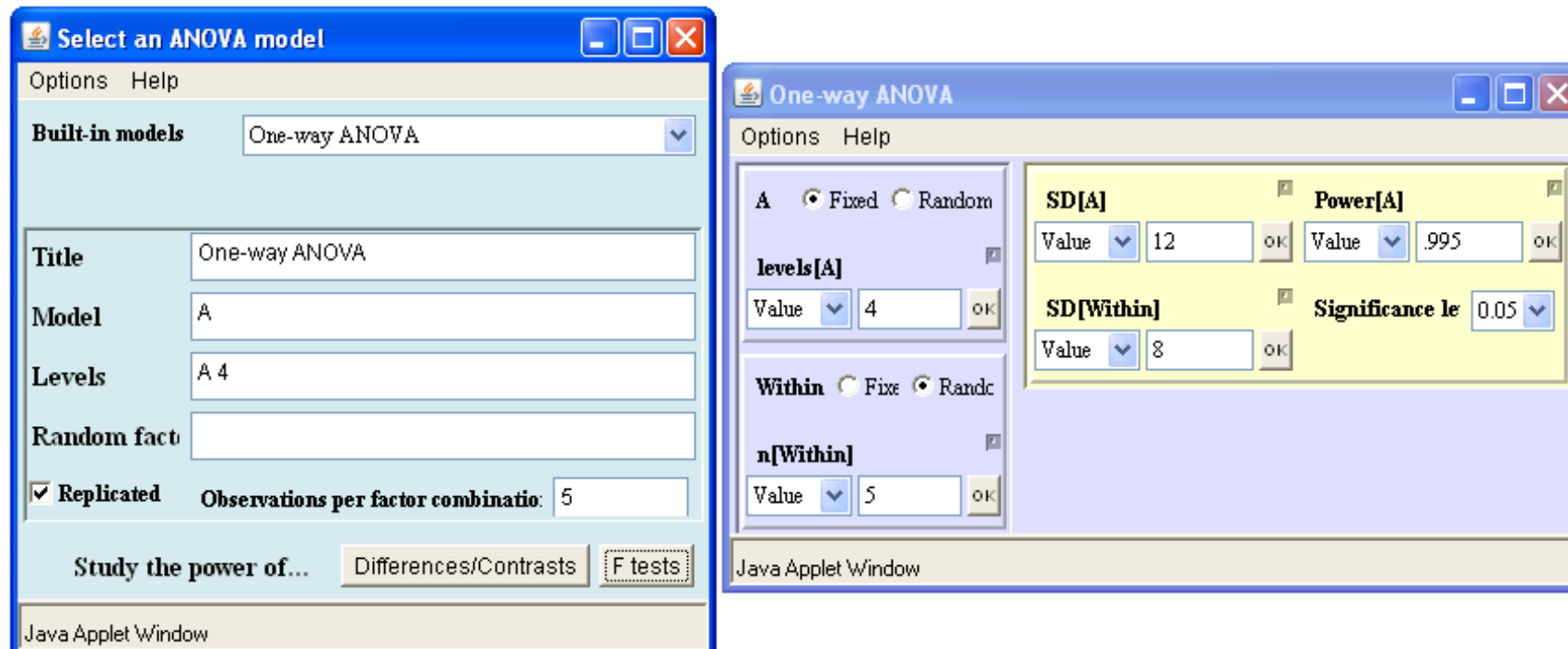
$$\phi = \frac{n(k-1)}{k} \left(\frac{\delta}{\sigma_\epsilon} \right)^2 = \frac{5 \times 3}{4} \left(\frac{24}{8} \right)^2 = 33.75.$$

The F statistic will have $df_{treatments} = 4 - 1 = 3$ and $df_\epsilon = 4(5 - 1) = 16$ degrees of freedom. The power is 99.5% as determined from Equation 8.1:

$$F_{0.95} = 3.239 = F_{0.005, 33.75}.$$

From Piface > **Balanced ANOVA (any model)** > **One-way ANOVA** with:

$$s_A = \sqrt{\frac{((18)^2 + (-6)^2 + (-6)^2 + (-6)^2)}{4 - 1}} = 12.0$$



From PASS > Means > Many Means > ANOVA: One-Way:

The screenshot shows the PASS software interface. The left pane displays the 'ANOVA: One Way' input settings, and the right pane displays the 'One Way ANOVA Power Analysis' output results.

ANOVA: One Way Input Settings:

- Find (Solve For): Beta and Power
- k (Number of Groups): 4
- Hypothesized Means: 18 -6 -6 -6
- Alpha (Significance Level): .05
- n (Sample Size Multiplier): 5
- Beta (1-Power): 0.1
- Group Sample Size Pattern: Equal
- S (Std Dev. of Subjects): 8

One Way ANOVA Power Analysis Output:

One Way ANOVA Power Analysis								
Numeric Results								
	Average		Total			Std Dev	Standard	
Power	n	k	N	Alpha	Beta	of Means	Deviation	Effect
						(Sm)	(S)	Size
0.99497	5.00	4	20	0.05000	0.00503	10.39	8.00	1.2990

References:

- Desu, M. M. and Raghavarao, D. 1990. Sample Size Methodology. Academic Press. New York.
- Fleiss, Joseph L. 1986. The Design and Analysis of Clinical Experiments. John Wiley & Sons. New York.
- Kirk, Roger E. 1982. Experimental Design: Procedures for the Behavioral Sciences. Brooks/Cole. Pacific Grove, California.

Report Definitions:

- Power is the probability of rejecting a false null hypothesis. It should be close to one.
- n is the average group sample size.
- k is the number of groups.
- Total N is the total sample size of all groups combined.
- Alpha is the probability of rejecting a true null hypothesis. It should be small.
- Beta is the probability of accepting a false null hypothesis. It should be small.
- Sm is the standard deviation of the group means under the alternative hypothesis.
- Standard deviation is the within group standard deviation.
- The Effect Size is the ratio of Sm to standard deviation.

Summary Statements:

In a one-way ANOVA study, sample sizes of 5, 5, 5, and 5 are obtained from the 4 groups whose means are to be compared. The total sample of 20 subjects achieves 99% power to detect differences among the means versus the alternative of equal means using an F test with a 0.05000 significance level. The size of the variation in the means is represented by their standard deviation which is 10.39. The common standard deviation within a group is assumed to be 8.00.

Example 8.4 Determine the power to reject H_0 by one-way ANOVA when the treatment means are $\mu_i = \{50, 30, 40, 40, 40\}$ and the sample sizes are $n_i = \{12, 12, 20, 20, 15\}$. The five populations are expected to be normal and homoscedastic with $\sigma_\epsilon = 13$.

Solution: The grand mean of the experimental data is expected to be

$$\frac{\sum n_i \mu_i}{\sum n_i} = \frac{(50 \times 12) + \dots + (40 \times 15)}{12 + \dots + 15} = 40.$$

The treatment biases relative to the grand mean are $\tau_i = \{10, -10, 0, 0, 0\}$ so the noncentrality parameter is

$$\phi = \frac{12(10)^2 + 12(-10)^2 + 20(0)^2 + 20(0)^2 + 15(0)^2}{13^2} = 14.2.$$

The ANOVA will have $df_{treatments} = 5 - 1 = 4$ and $df_\epsilon = \sum n_i - 1 - 4 = 74$ degrees of freedom. The power is 84.7% as determined from Equation 8.1:

$$F_{0.95} = 2.495 = F_{0.153, 14.2}.$$

8.2 Randomized Block Design

Example 8.5 Recalculate the power for Example 8.1 if the experiment is built as a randomized block design and the standard deviation of the population of block biases is $\sigma_{blocks} = 4$.

Solution: The F distribution noncentrality parameter ($\phi = 11.72$) and the treatment degrees of freedom ($df_{treatments} = 3$) will be unchanged from the original solution, but if the experiment is built in five blocks with one replicate in each block, the new error degrees of freedom for the RBD will be

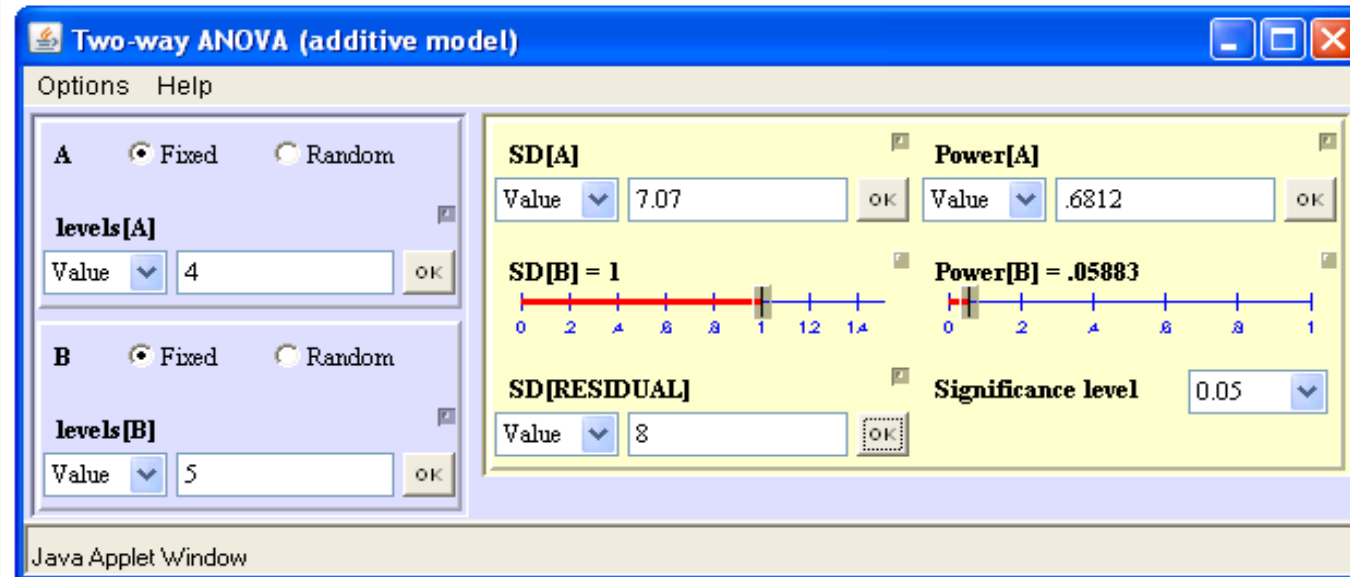
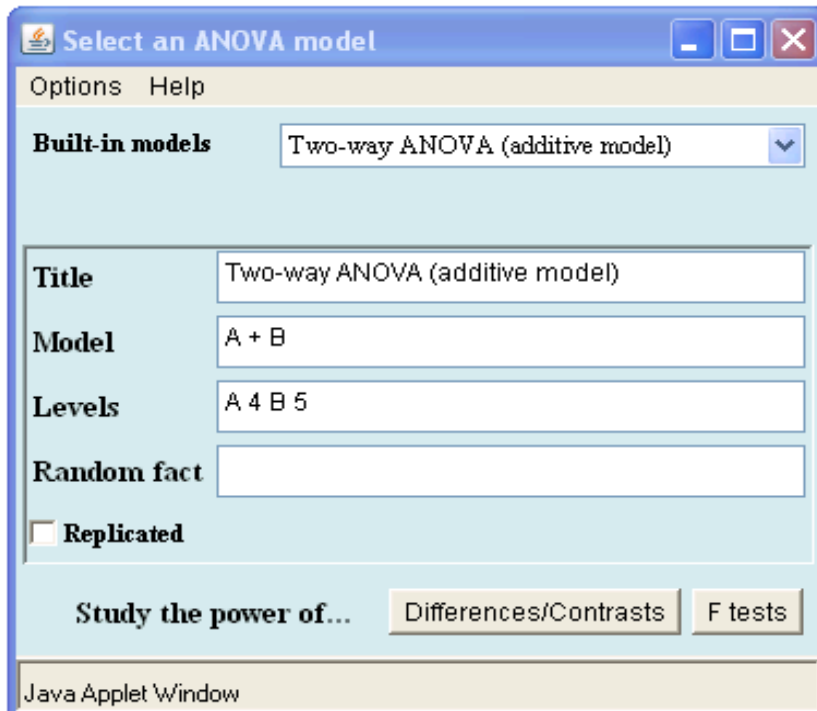
$$\begin{aligned} df_e &= df_{total} - df_{treatments} - df_{blocks} \\ &= 19 - 3 - 4 \\ &= 12. \end{aligned}$$

The power of the RBD is 68% as determined from

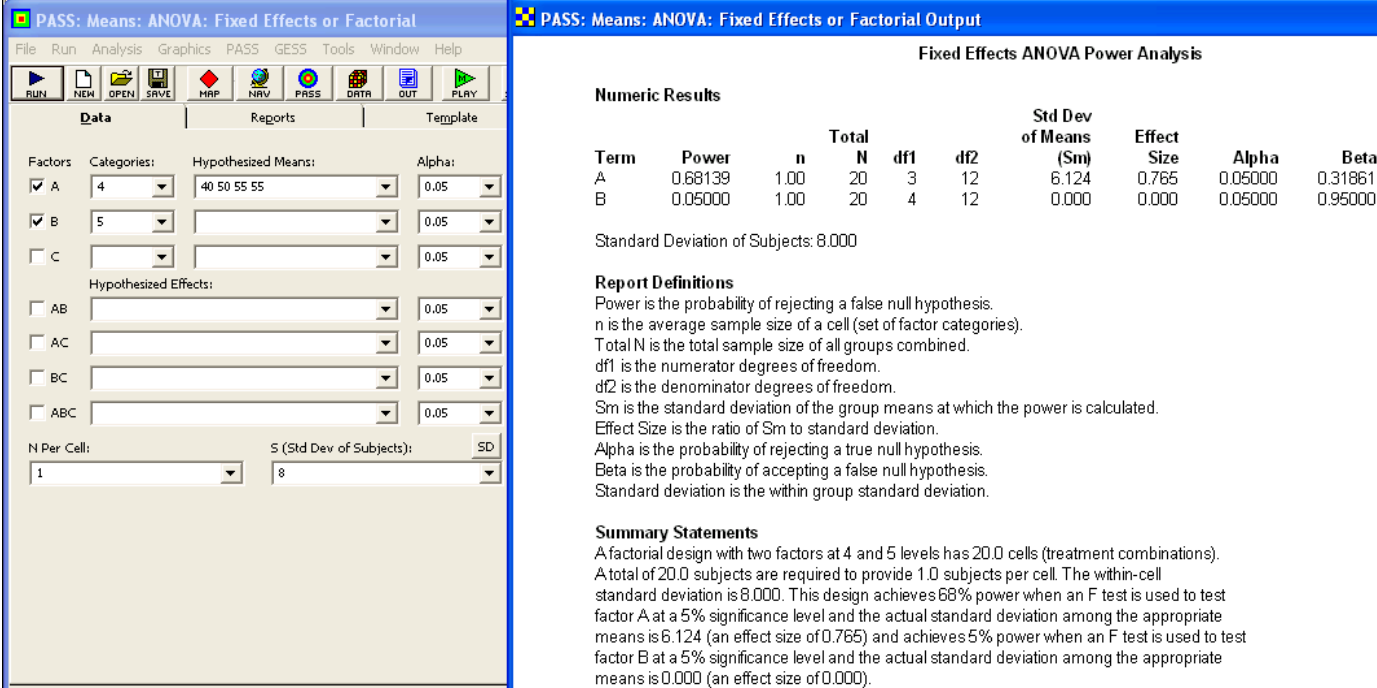
$$F_{0.95} = 3.490 = F_{0.32, 11.72}.$$

This is slightly lower than the original power (72%) because the RBD has fewer error degrees of freedom than the CRD. The RBD's power is not affected by the block variation because it separates that variation from the error variation that is used to determine the power.

From Piface > **Balanced ANOVA (any model)** > **Two-way ANOVA (additive model)** with:



From PASS> Means> Many Means> ANOVA: Fixed Effect:



The screenshot shows the PASS software interface. The left window is titled "PASS: Means: ANOVA: Fixed Effects or Factorial" and contains the following settings:

- Factors: A (checked), B (checked), C (unchecked)
- Categories: A (4), B (5), C ()
- Hypothesized Means: A (40 50 55 55), B (), C ()
- Alpha: 0.05 for all factors and interactions
- Hypothesized Effects: AB (), AC (), BC (), ABC ()
- N Per Cell: 1
- S (Std Dev of Subjects): 8

The right window is titled "PASS: Means: ANOVA: Fixed Effects or Factorial Output" and displays the "Fixed Effects ANOVA Power Analysis" results:

Numeric Results

Term	Power	n	Total N	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	0.68139	1.00	20	3	12	6.124	0.765	0.05000	0.31861
B	0.05000	1.00	20	4	12	0.000	0.000	0.05000	0.95000

Standard Deviation of Subjects: 8.000

Report Definitions

Power is the probability of rejecting a false null hypothesis.
n is the average sample size of a cell (set of factor categories).
Total N is the total sample size of all groups combined.
df1 is the numerator degrees of freedom.
df2 is the denominator degrees of freedom.
Sm is the standard deviation of the group means at which the power is calculated.
Effect Size is the ratio of Sm to standard deviation.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.
Standard deviation is the within group standard deviation.

Summary Statements

A factorial design with two factors at 4 and 5 levels has 20.0 cells (treatment combinations). A total of 20.0 subjects are required to provide 1.0 subjects per cell. The within-cell standard deviation is 8.000. This design achieves 68% power when an F test is used to test factor A at a 5% significance level and the actual standard deviation among the appropriate means is 6.124 (an effect size of 0.765) and achieves 5% power when an F test is used to test factor B at a 5% significance level and the actual standard deviation among the appropriate means is 0.000 (an effect size of 0.000).

Example 8.6 A 40-run experiment was performed using an RBD with $k = 5$ treatments and $r = 8$ blocks. The ANOVA table from the experiment is shown in Table 8.2. Calculate the blocking efficiency and the increase in the number of runs required to obtain the same estimation precision for treatment means using a CRD.

Solution: The blocking efficiency as determined from Equation 8.13 is

$$E = \frac{(7 \times 14) + (8 \times 4 \times 4)}{(5 \times 8 - 1) 4} = 1.45.$$

That is, the CRD will require about 45% more runs than the RBD because it ignores the variation associated with block effects. Because the number of runs in the RBD was $kr = 40$, the number of runs required for the CRD to obtain the same estimation precision for the treatment means would be $Ekr = 1.45 \times 40 = 58$. Apparently the blocking was beneficial and should be used in future studies.

8.3 Balanced Full Factorial Design with Fixed Effects

Example 8.7 A $2 \times 3 \times 5$ full factorial experiment with four replicates is planned. The experiment will be blocked on replicates and the ANOVA model will include main effects and two-factor interactions. Determine the power to detect a difference $\delta = 300$ units between two levels of each study variable if the standard error of the model is expected to

be $\sigma_\epsilon = 500$.

Solution: If the three study variables are given the names A , B , and C and have $a = 2$, $b = 3$, and $c = 5$ levels, respectively, then the degrees of freedom associated with the terms in the model will be $df_{blocks} = 3$, $df_A = 1$, $df_B = 2$, $df_C = 4$, $df_{AB} = 2$, $df_{AC} = 4$, $df_{BC} = 8$, and

$$\begin{aligned} df_\epsilon &= df_{total} - df_{model} \\ &= (2 \times 3 \times 5 \times 4 - 1) - (3 + 1 + 2 + 4 + 2 + 4 + 8) \\ &= 119 - 24 \\ &= 95. \end{aligned}$$

From Equation 8.2, the F distribution noncentrality parameter for variable A with treatment biases $\alpha_1 = -150$ and $\alpha_2 = 150$ is

$$\begin{aligned} \phi_A &= \frac{bcn \sum_{i=1}^2 \alpha_i^2}{\sigma_\epsilon^2} \\ &= \frac{3 \times 5 \times 4 \times \left((-150)^2 + 150^2 \right)}{500^2} \\ &= 10.8. \end{aligned}$$

The distribution of F_A will have $df_A = 1$ numerator and $df_\epsilon = 95$ denominator degrees of freedom, so the power associated with A is given by Equation 8.1:

$$F_{0.95} = 3.942 = F_{1-\pi_A, 10.8},$$

which is satisfied by $\pi_A = 0.908$ or 90.8%.

Similarly, the F distribution noncentrality parameter for B with biases $\beta_1 = -150$, $\beta_2 = 150$, and $\beta_3 = 0$ is

$$\begin{aligned} \phi_B &= \frac{acn \sum_{i=1}^3 \beta_i^2}{\sigma_\epsilon^2} \\ &= \frac{2 \times 5 \times 4 \times \left((-150)^2 + 150^2 + 0^2 \right)}{500^2} \\ &= 7.2. \end{aligned}$$

The distribution of F_B will have $df_B = 2$ numerator and $df_\epsilon = 95$ denominator degrees of freedom, so the power associated with B is given by

$$F_{0.95} = 3.093 = F_{1-\pi_B, 7.2},$$

which is satisfied by $\pi_B = 0.654$.

Finally, the F distribution noncentrality parameter for C with biases $\gamma_1 = -150$, $\gamma_2 = 150$, and $\gamma_3 = \gamma_4 = \gamma_5 = 0$ is

$$\begin{aligned}\phi_C &= \frac{abn \sum_{i=1}^5 \gamma_i^2}{\sigma_\epsilon^2} \\ &= \frac{2 \times 3 \times 4 \times \left((-150)^2 + 150^2 + 0^2 + 0^2 + 0^2 \right)}{500^2} \\ &= 4.32.\end{aligned}$$

The distribution of F_C will have $df_C = 4$ numerator and $df_\epsilon = 95$ denominator degrees of freedom, so the power associated with C is given by

$$F_{0.95} = 2.469 = F_{1-\pi_C, 4.32},$$

which is satisfied by $\pi_C = 0.328$. These three power calculations confirm by example that the power to detect a variable effect decreases as the number of variable levels increases.

From MINITAB> Stat> Power and Sample Size> General Full Factorial Design (MINITAB only reports the power for the variable with the most levels, which in this case is C):

```
MTB > Power;
SUBC> FDesign;
SUBC> NLevels 2 3 5;
SUBC> Reps 4;
SUBC> MaxDifference 300;
SUBC> Sigma 500;
SUBC> TOrder 2;
SUBC> FitB;
SUBC> Alpha 0.05;
SUBC> GPCurve.

Power and Sample Size

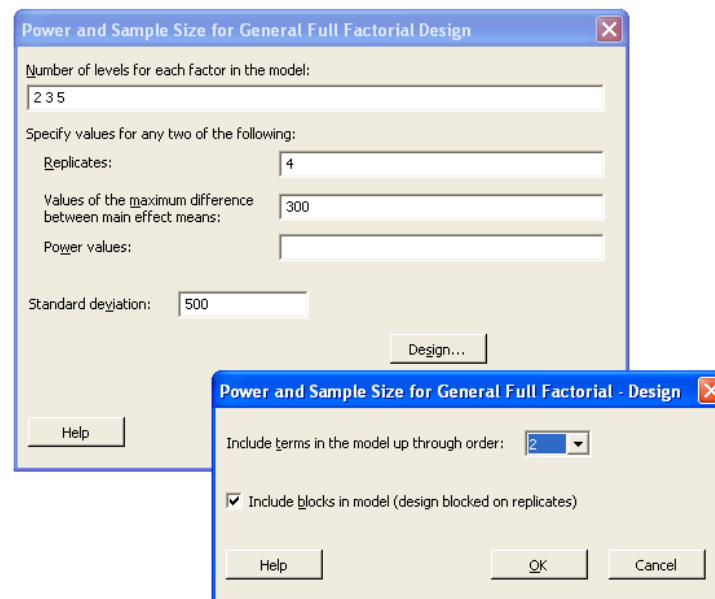
General Full Factorial Design

Alpha = 0.05 Assumed standard deviation = 500

Factors: 3 Number of levels: 2, 3, 5

Include terms in the model up through order: 2
Include blocks in model.

Maximum Total
Difference Reps Runs Power
300 4 120 0.328605
```



From Piface > Balanced ANOVA (any model) > Three-way ANOVA with

$$s_A = \sqrt{(2(150)^2) / (2 - 1)} = 212.1$$

$$s_B = \sqrt{(2(150)^2 + (0)^2) / (3 - 1)} = 150.0$$

$$s_C = \sqrt{(2(150)^2 + 3(0)^2) / (5 - 1)} = 106.1$$

Select an ANOVA model

Options Help

Built-in models: Three-way ANOVA

Title: Three-way ANOVA

Model: BI + A + B + C + A*B + A*C + B*C

Levels: BI 4 A 2 B 3 C 5

Random factors:

Replicated Observations per factor combination: 1

Study the power of... Differences/Contrasts F tests

Three-way ANOVA

Options Help

BI Fixed Random

levels[B] = 4

A Fixed Random

levels[A] = 2

B Fixed Random

levels[B] = 3

C Fixed Random

levels[C] = 5

SD[B] = 1 Power[B] = .05002

SD[A] Value: 212.1 OK Power[A] Value: .9019 OK

SD[B] Value: 150 OK Power[B] Value: .6538 OK

SD[C] Value: 106.1 OK Power[C] Value: .3288 OK

SD[A*B] = 1 Power[A*B] = .05001

SD[A*C] = 1 Power[A*C] = .05001

SD[B*C] = 1 Power[B*C] = .05001

SD[RESIDUAL] Value: 500 OK Significance level: 0.05

From PASS> Means> Many Means> ANOVA: Fixed Effect:

The screenshot shows the Minitab PASS interface for ANOVA Power Analysis. The left pane shows the configuration for a 3-factor ANOVA with factors A, B, and C. Factor A has 2 levels, B has 3 levels, and C has 5 levels. The standard deviation of subjects is set to 500.000. The right pane displays the 'Fixed Effects ANOVA Power Analysis' results.

Term	Power	n	Total N	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	0.90217	4.00	120	1	98	150.000	0.300	0.05000	0.09783
B	0.65427	4.00	120	2	98	122.474	0.245	0.05000	0.34573
C	0.32908	4.00	120	4	98	94.868	0.190	0.05000	0.67092
AB	0.05000	4.00	120	2	98	0.000	0.000	0.05000	0.95000
AC	0.05000	4.00	120	4	98	0.000	0.000	0.05000	0.95000
BC	0.05000	4.00	120	8	98	0.000	0.000	0.05000	0.95000

Standard Deviation of Subjects: 500.000

Summary Statements
 A factorial design with three factors at 2, 3, and 5 levels has 30.0 cells (treatment combinations). A total of 120.0 subjects are required to provide 4.0 subjects per cell. The within-cell standard deviation is 500.000. This design achieves 90% power when an F test is used to test factor A at a 5% significance level and the actual standard deviation among the appropriate means is 150.000 (an effect size of 0.300), achieves 65% power when an F test is used to test factor B at a 5% significance level and the actual standard deviation among the appropriate means is 122.474 (an effect size of 0.245), achieves 33% power when an F test is used to test factor C at a 5% significance level and the actual standard deviation among the appropriate means is 94.868 (an effect size of 0.190), achieves 5% power when an F test is used to test the AB interaction at a 5% significance level and the actual standard deviation among the appropriate means is 0.000 (an effect size of 0.000), achieves 5% power when an F test is used to test the AC interaction at a 5% significance level and the actual standard deviation among the appropriate means is 0.000 (an effect size of 0.000), and achieves 5% power when an F test is used to test the BC interaction at a 5% significance level and the actual standard deviation among the appropriate means is 0.000 (an effect size of 0.000).

MINITAB doesn't have a built in capability to do sample size and power calculations for multi-way ANOVA, however, the custom macro *power.mac* (posted at www.nmbstatistical.com/Samplesize.htm) can be used to calculate the power for one design variable at a time in a balanced multi-way ANOVA. For the first variable:

```
MTB %power c1
Executing from file: C:\Program Files\Minitab 15\English\Macros\power.MAC
```

```
Do you want to specify your design from the terminal or from a column?
```

```
If from terminal a column will be created in the column specified.
Otherwise the column specified will be the input.
```

```
(terminal=1, column=2)
DATA 1
```

```
How many runs are in one replicate?
DATA 30
```

```
How many replicates?
DATA 4
```

```
How many levels does the variable have?
DATA 2
```

How many model degrees of freedom are there?

DATA 24

What is the standard deviation?

DATA 500

What is the smallest difference that you want to detect between two levels?

DATA 300

N	120.000
Runs	30.0000
reps	4.00000
Levels	2.00000
dfmodel	24.0000
dferror	95.0000
Fcrit	3.94122
lambda	10.8000
sigma	500.000
delta	300.000
Power	0.901990

8.4 Random and Mixed Models

Piface can calculate sample size and power for random and mixed models but MINITAB and PASS can not.

Example 8.8 A balanced full factorial experiment is to be performed using $a = 3$ levels of a fixed variable A , $b = 5$ randomly selected levels of a random variable B , and $n = 4$ replicates. Determine the power to reject $H_0 : \alpha_i = 0$ for all i when the A -level biases are $\alpha_i = \{-20, 20, 0\}$ with $\sigma_B = 25$, $\sigma_{AB} = 0$, and $\sigma_\epsilon = 40$. Assume that the AB interaction term will be included in the ANOVA even though its expected variance component is 0.

Solution: The ANOVA table with the equations for the expected mean squares is shown in Table 8.3. From the ANOVA table, the error mean square used for testing the A effect (that is, the denominator of F_A) is

$$\begin{aligned} MS_{\epsilon(A)} &= MS_{AB} \\ &= \hat{\sigma}_\epsilon^2 + n\hat{\sigma}_{AB}^2. \end{aligned}$$

The noncentrality parameter for the test of the fixed effect A is given by Equation 8.15:

$$\begin{aligned} \phi_A &= \frac{N \sum_{i=1}^a \alpha_i^2}{a MS_{\epsilon(A)}} \\ &= \frac{3 \times 5 \times 4}{3} \frac{(-20)^2 + (20)^2 + (0)^2}{(40)^2 + 4(0)^2} \\ &= 10. \end{aligned}$$

With $df_A = 2$, $df_{AB} = 8$, and $\alpha = 0.05$ in Equation 8.1

$$F_{0.95} = 4.459 = F_{1-\pi, 10, 0},$$

which is satisfied by $\pi = 0.640$.

From Piface > **Balanced ANOVA (any model)** with:

$$s_A = \sqrt{\left((-20)^2 + (20)^2 + (0)^2 \right) / (3 - 1)} = 20$$

The image shows two screenshots of software interfaces for ANOVA model selection and parameter configuration.

The left screenshot, titled "Select an ANOVA model", shows the following settings:

- Built-in models:** (Define your own)
- Title:** Mixed Model
- Model:** A + B + A*B
- Levels:** A 3 B 5
- Random fac:** B
- Replicated** Observations per factor combination: 4
- Study the power of...** Differences/Contrasts F tests

The right screenshot, titled "Mixed Model", shows the following settings:

- A:** Fixed (selected), Random (unselected). levels[A] = 3. Slider range 0 to 4.
- B:** Fixed (unselected), Random (selected). n[B] = 5. Slider range 0 to 7.
- Within:** Fixed (unselected), Random (selected). n[Within] = 4. Slider range 0 to 6.
- SD[A]:** Value 20
- Power[A]:** Value .64
- SD[B]:** Value 25
- Power[B]:** Value .628
- SD[A*B]:** Value 0
- Power[A*B]:** Value .05
- SD[Within]:** Value 40
- Significance level:** 0.05

Example 8.9 Determine the power to reject $H_0 : \sigma_B^2 = 0$ when $\sigma_B = 25$, $\sigma_{AB} = 0$, and $\sigma_\epsilon = 40$ for Example 8.8. Retain the AB interaction term in the model even though its variance component is 0.

Solution: The ANOVA table with the equations for the expected mean squares is shown in Table 8.3. From Equation 8.19 under the specified conditions, the expected F_B value

is approximately

$$\begin{aligned}
 E(F_B) &\simeq \frac{E(MS_B)}{E(MS_{AB})} \\
 &\simeq \frac{\sigma_\epsilon^2 + n\sigma_{AB}^2 + an\sigma_B^2}{\sigma_\epsilon^2 + n\sigma_{AB}^2} \\
 &\simeq \frac{(40)^2 + 4(0)^2 + 3 \times 4 \times (25)^2}{(40)^2 + 4(0)^2} \\
 &\simeq 5.69.
 \end{aligned}$$

With $df_B = 4$, $df_{AB} = 8$, and $\alpha = 0.05$, the critical F value for the test for the B effect is $F_{0.95,4,8} = 3.838$, so from Equation 8.20 the power is approximately

$$\begin{aligned}
 \pi &\simeq P\left(\frac{3.838}{5.69} < F < \infty\right) \\
 &\simeq P(0.675 < F < \infty) \\
 &\simeq 0.618.
 \end{aligned}$$

See Piface solution to Example 8.8.

8.5 Nested Designs

Example 8.10 A is a fixed variable with three levels and B is a random variable with four levels nested within each level of A . The nested design is crossed with a five-level fixed variable C and one replicate of the experiment will be built. The model to be fitted is: $A + B(A) + C + AC + BC$. Find the power to detect a difference $\delta = 40$ between two levels of A assuming that the standard deviations for the random effects are $\sigma_B = 12$, $\sigma_{BC} = 4$, and $\sigma_\epsilon = 10$.

Solution: The ANOVA table with the equations for the expected mean squares is shown in Table 8.4 where the α_i are the A -level biases, the τ_i are the C -level biases, and the γ_i are the AC interaction biases. The error mean square used for testing the A effect is $MS_{\epsilon(A)} = MS_{B(A)}$. The noncentrality parameter for the test of the fixed effect A is given by Equation 8.15:

$$\begin{aligned}
 \phi_A &= \frac{a \times b \times c \times n}{a} \frac{\sum_{i=1}^a \alpha_i^2}{\sigma_\epsilon^2 + n\sigma_{BC}^2 + cn\sigma_{B(A)}^2} \\
 &= \frac{3 \times 4 \times 5 \times 1}{3} \frac{(-20)^2 + (20)^2 + (0)^2}{(10)^2 + (4)^2 + 5(12)^2} \\
 &= 19.14.
 \end{aligned}$$

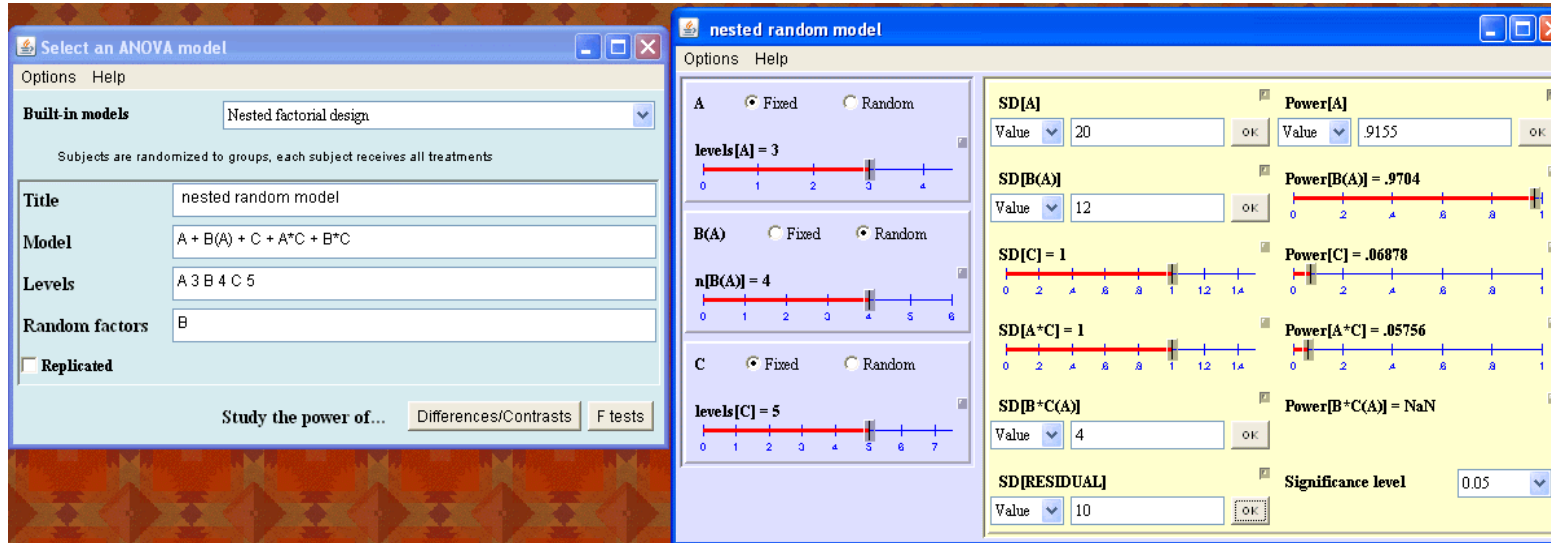
With $df_A = 2$, $df_{B(A)} = 3(4 - 1) = 9$, and $\alpha = 0.05$ in Equation 8.1:

$$F_{0.95} = 4.256 = F_{1-\pi, 19, 14}$$

which is satisfied by $\pi = 0.915$.

From Piface > **Balanced ANOVA (any model)** with:

$$s_A = \sqrt{\frac{2(20)^2 + (0)^2}{3-1}} = 20.0$$



8.6 Two-Level Factorial Designs

Example 8.11 Use the method of Equation 8.24 to determine the number of replicates required to detect an effect of size $\delta = 6$ with 90% power in a 2^4 experiment when $\sigma_\epsilon = 10$. Assume that the ANOVA model will include main effects and two-factor interactions.

Solution: With $t \simeq z$ in the first iteration of Equation 8.24, the number of replicates required to deliver 90% power to detect the difference $\delta = 6$ between two levels of a design variable is

$$\begin{aligned} n &\geq \frac{1}{2^{4-2}} (1.96 + 1.282)^2 \left(\frac{10}{6}\right)^2 \\ &\geq 8. \end{aligned}$$

Another iteration (not shown) confirms that $n = 8$ is the correct number of replicates.

See Example 8.12.

Example 8.12 Use the method of Equation 8.21 to confirm the solution to Example 8.11.

Solution: By Equation 8.22 the F distribution noncentrality parameter is

$$\phi = 8 \times 2^{4-2} \left(\frac{6}{10}\right)^2 = 11.5.$$

The central and noncentral F distributions will have $df_i = 1$ numerator and $df_\epsilon = df_{total} - df_{model} = (8 \times 2^4 - 1) - (4 + 6) = 117$ denominator degrees of freedom. The power, determined from the condition

$$\begin{aligned} F_{0.95} &= F_{1-\pi, 11.5} \\ 3.922 &= F_{0.080, 11.5} \end{aligned}$$

is $\pi = 0.920$ or 92.0%. This value is slightly larger than the 90% goal because the calculated value of n was fractional and was rounded up to the nearest integer. With $n = 7$ the power is slightly less than 90%.

From **Piface** > **Balanced ANOVA (any model)** with:

$$s_A = \sqrt{\frac{2(3)^2}{2-1}} = 4.243$$

The screenshot displays the MINITAB Power and Sample Size dialog for a 2^4 factorial design. The main window is titled "2^4" and includes an "Options" menu. The settings are as follows:

- Factor A:** Fixed, levels[A] = 2, SD[A] = 4.243, Power[A] = .9202
- Factor B:** Fixed, levels[B] = 2, SD[B] = 1, Power[B] = .1247
- Factor C:** Fixed, levels[C] = 2, SD[C] = 1, Power[C] = .1247
- Factor D:** Fixed, levels[D] = 2, SD[D] = 1, Power[D] = .1247
- Residual:** SD[Residual] = 10, Significance level = 0.05

The "Select an ANOVA model" dialog is open, showing the following configuration:

- Title:** 2^4
- Model:** $A + B + C + D + A*B + A*C + A*D + B*C + B*D + C*D$
- Levels:** A 2 B 2 C 2 D 2
- Random factors:** (empty)
- Replicated:** Observations per factor combination = 8
- Study the power of...:** Differences/Contrasts, F tests

From MINITAB> Stat> Power and Sample Size> 2-Level Factorial Design (with 5 terms removed from the model: four three-factor interactions and the one four-factor interaction):

```

MTB > Power;
SUBC> FFDesign 4 16;
SUBC> Effect 6;
SUBC> Power 0.90;
SUBC> CPBlock 0;
SUBC> Sigma 10;
SUBC> Omit 5;
SUBC> FitC;
SUBC> FitB;
SUBC> GPCurve.

Power and Sample Size

2-Level Factorial Design

Alpha = 0.05 Assumed standard deviation = 10

Factors: 4 Base Design: 4, 16
Blocks: none

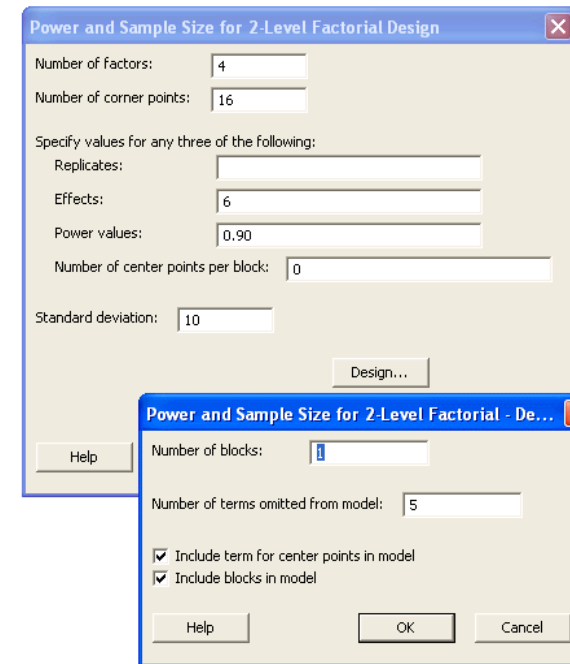
Number of terms omitted from model: 5

Center Points Effect Reps Total Runs Target Power Actual Power
0 6 8 128 0.9 0.920162

Power Curve for 2-Level Factorial Design

MTB >

```



Example 8.13 Suppose that two more two-level variables were added to the 2^4 experiment with $n = 8$ replicates from Example 8.11 without any increase in the total number of runs. Calculate the power for the resulting 2^6 experiment.

Solution: The 2^6 experiment must have $n = 2$ replicates to maintain the same number of runs as the original experiment. Because $8 \times 2^4 = 2 \times 2^6$, the F distribution noncentrality parameter will be unchanged. The new error degrees of freedom for the F distributions will be $df_e = (2 \times 2^6 - 1) - (6 + 15) = 106$. The power, determined from

$$\begin{aligned}
 F_{0.95} &= F_{1-\pi, 11, 5} \\
 3.931 &= F_{0.081, 11, 5}
 \end{aligned}$$

is $\pi = 0.919$ or 91.9%. This example confirms that adding variables to a 2^k design without increasing the total number of observations has little effect on the power provided that the error degrees of freedom remains large.

From **Piface**> **Balanced ANOVA (any model)**:

The screenshot displays the Minitab Power and Sample Size tool for a 2^6 factorial design. The main window shows settings for factors A through F, each with 2 levels and a standard deviation of 1. A central dialog box "Select an ANOVA model" is open, showing a full model with all 42 terms. The background displays a grid of power plots for various terms, all showing a power of 0.08671. The residual standard deviation is set to 10, and the significance level is 0.05.

Main Window Settings:

- Factor A: Fixed, levels[A] = 2, SD[A] = 4.243, Power[A] = .9198
- Factor B: Fixed, levels[B] = 2, SD[B] = 1, Power[B] = .08671
- Factor C: Fixed, levels[C] = 2, SD[C] = 1, Power[C] = .08671
- Factor D: Fixed, levels[D] = 2, SD[D] = 1, Power[D] = .08671
- Factor E: Fixed, levels[E] = 2, SD[E] = 1, Power[E] = .08671
- Factor F: Fixed, levels[F] = 2, SD[F] = 1, Power[F] = .08671
- Residual: Fixed, Replications = 2, SD[Residual] = 10
- Significance level: 0.05

Select an ANOVA model Dialog:

- Title: 2^6
- Model: $A+B+C+D+E+F+A*B+A*C+A*D+A*E+A*F+B*C+B*D+B*E+B*F+C*D+$
- Levels: A 2 B 2 C 2 D 2 E 2 F 2
- Random factors: (empty)
- Replicated: Observations per factor combination: 2
- Study the power of...: Differences/Contrasts, F tests

Power Plots (All show Power = 0.08671):

- SD[A*C] = 1, Power[A*C] = .08671
- SD[A*D] = 1, Power[A*D] = .08671
- SD[A*E] = 1, Power[A*E] = .08671
- SD[A*F] = 1, Power[A*F] = .08671
- SD[B*C] = 1, Power[B*C] = .08671
- SD[B*D] = 1, Power[B*D] = .08671
- SD[B*E] = 1, Power[B*E] = .08671
- SD[B*F] = 1, Power[B*F] = .08671
- SD[C*D] = 1, Power[C*D] = .08671
- SD[C*E] = 1, Power[C*E] = .08671
- SD[C*F] = 1, Power[C*F] = .08671
- SD[D*E] = 1, Power[D*E] = .08671
- SD[D*F] = 1, Power[D*F] = .08671
- SD[E*F] = 1, Power[E*F] = .08671

From MINITAB> Stat> Power and Sample Size> 2-Level Factorial Design (with 42 terms removed from the model: 20 three-factor interactions, 15 four-factor interactions, 6 five-factor interactions, and the one six-factor interaction):

```

MTB > Power;
SUBC> FFDesign 6 64;
SUBC> Reps 2;
SUBC> Effect 6;
SUBC> CPBlock 0;
SUBC> Sigma 10;
SUBC> Omit 42;
SUBC> FitC;
SUBC> FitB;
SUBC> GPCurve.

Power and Sample Size

2-Level Factorial Design

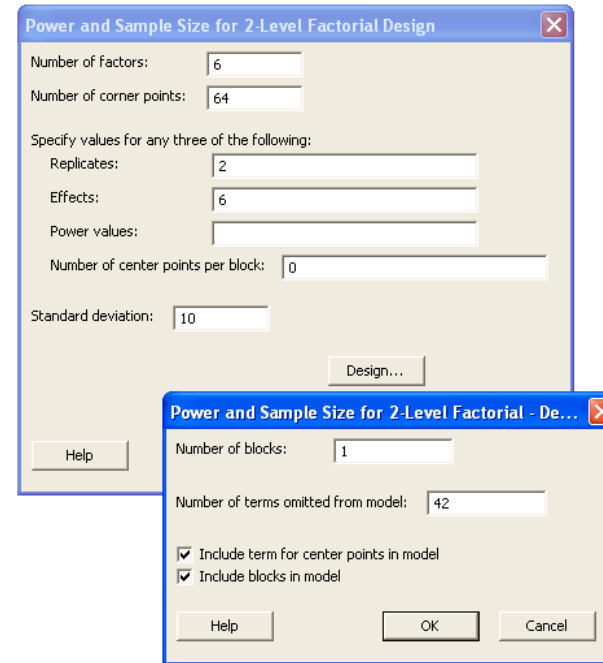
Alpha = 0.05 Assumed standard deviation = 10

Factors: 6 Base Design: 6, 64
Blocks: none

Number of terms omitted from model: 42

Center Points Effect Reps Total Runs Power
0 6 2 128 0.919727

```



Example 8.14 Derive a simplified expression for the total number of observations required for a 2^k experiment to detect a difference δ between two levels of a design variable assuming $\alpha = 0.05$ and $\beta = 0.10$. Under what conditions should this expression be valid?

Solution: From Equation 8.25 the total number of replicates required for a 2^k design to have 90% power to detect a difference δ between two levels of a design variable is approximately

$$\begin{aligned}
 n2^k &\geq 4(z_{0.025} + z_{0.10})^2 \left(\frac{\sigma_\epsilon}{\delta}\right)^2 \\
 &\geq 42 \left(\frac{\sigma_\epsilon}{\delta}\right)^2.
 \end{aligned} \tag{8.1}$$

This condition will be strictly valid when df_ϵ is large so that the $t \simeq z$ approximation is well satisfied.

Example 8.15 How many replicates of a 2^3 design are required to determine the regression coefficient for a main effect with precision $\delta = 300$ with 95% confidence when the standard error of the model is expected to be $\sigma_\epsilon = 600$?

Solution: If the error degrees of freedom are sufficiently large that $t_{0.025} \simeq z_{0.025}$ then

$$\begin{aligned}
 n &\geq \frac{1}{2^3} \left(\frac{1.96 \times 600}{300}\right)^2 \\
 &\geq 2.
 \end{aligned}$$

With only $2 \times 2^3 = 16$ total runs, the $t_{0.025} \simeq z_{0.025}$ assumption is not satisfied. Another iteration shows that the transcendental sample size condition is satisfied for $n = 3$ replicates of the 2^3 design.

Example 8.16 What is the power for the 2_{IV}^{4-1} design with two replicates to detect a difference of $\delta = 10$ between two levels of a design variable if $\sigma_\epsilon = 5$?

Solution: With two replicates the total number of experimental runs will be $2(2^{4-1}) = 16$. Because the experiment design is resolution IV, the model can include main effects and only three of the six possible two-factor interactions, so $df_{model} = 4 + 3 = 7$. Then, the error degrees of freedom will be $df_\epsilon = (16 - 1) - 7 = 8$. The F distribution noncentrality parameter associated with a difference of $\delta = 10$ between two levels of a design variable is given by a slightly modified form of Equation 8.22:

$$\begin{aligned}\phi &= n2^{(k-p)-2} \left(\frac{\delta}{\sigma_\epsilon} \right)^2 \\ &= 2 \times 2^{(4-1)-2} \left(\frac{10}{5} \right)^2 \\ &= 16.0\end{aligned}\tag{8.2}$$

where $p = 1$ accounts for the half-fractionation of the full factorial design. Then, by Equation 8.21

$$\begin{aligned}F_{0.95} &= F_{1-\pi, 16} \\ 5.318 &= F_{0.063, 16}.\end{aligned}$$

The power is $\pi = 1 - 0.063 = 0.937$.

From MINITAB> Stat> Power and Sample Size> 2-Level Factorial Design:

Power Curve for 2-Level Factorial Design

```
MTE > Power;
SUBC> FFDesign 4 8;
SUBC> Reps 2;
SUBC> Effect 10;
SUBC> CPBlock 0;
SUBC> Sigma 5;
SUBC> FitC;
SUBC> FitB;
SUBC> GPCurve.
```

Power and Sample Size

2-Level Factorial Design

Alpha = 0.05 Assumed standard deviation = 5

Factors: 4 Base Design: 4, 8
Blocks: none

Center Points	Effect	Reps	Total Runs	Power
0	10	2	16	0.936743

Power and Sample Size for 2-Level Factorial Design

Number of factors: 4

Number of corner points: 8

Specify values for any three of the following:

Replicates: 2

Effects: 10

Power values:

Number of center points per block: 0

Standard deviation: 5

Design... Options... Graph... Help OK Cancel

Example 8.17 How many replicates of a 2_V^{5-1} design are required to have 90% power to detect a difference $\delta = 0.4$ between two levels of a design variable? Assume that ten of the fifteen possible terms will drop out of the model and that the standard error will be $\sigma_\epsilon = 0.18$.

Solution: If a model with main effects and two factor interactions is fitted to one replicate of the 2_V^{5-1} design, there will not be any degrees of freedom left to estimate the error, so either the experiment must be replicated or some terms must be dropped from the model. Under the assumption that the number of replicates is large, so that we can take $t \simeq z$ in the first iteration of Equation 8.24, we have

$$\begin{aligned} n &\geq \frac{1}{2^{(5-1)-2}} (z_{0.025} + z_{0.10})^2 \left(\frac{0.18}{0.4} \right)^2 \\ &\geq 0.532. \end{aligned}$$

Obviously, the $t \simeq z$ approximation is not satisfied, so at least one more iteration is required. If only one replicate of the half-fractional factorial design is built and ten of the fifteen possible terms are dropped from the model, the error degrees of freedom will be $df_\epsilon = 15 - 10 = 5$. Then, for the second iteration of Equation 8.24, we have

$$\begin{aligned} n &\geq \frac{1}{2^{(5-1)-2}} (t_{0.025} + t_{0.10})^2 \left(\frac{0.18}{0.4} \right)^2 \\ &\geq \frac{1}{2^{(5-1)-2}} (2.228 + 1.372)^2 \left(\frac{0.18}{0.4} \right)^2 \\ &\geq 0.656, \end{aligned}$$

which rounds up to $n = 1$. Calculation of the power (not shown) confirms $\pi = 0.98$ for one replicate.

From MINITAB> Stat> Power and Sample Size> 2-Level Factorial Design:

Power Curve for 2-Level Factorial Design

```

MTB > Power;
SUBC> FFDesign 5 16;
SUBC> Reps 1;
SUBC> Effect 0.4;
SUBC> CPBlock 0;
SUBC> Sigma 0.18;
SUBC> Omit 10;
SUBC> FitC;
SUBC> FitB;
SUBC> GPCurve.

```

Power and Sample Size

2-Level Factorial Design

Alpha = 0.05 Assumed standard deviation = 0.18

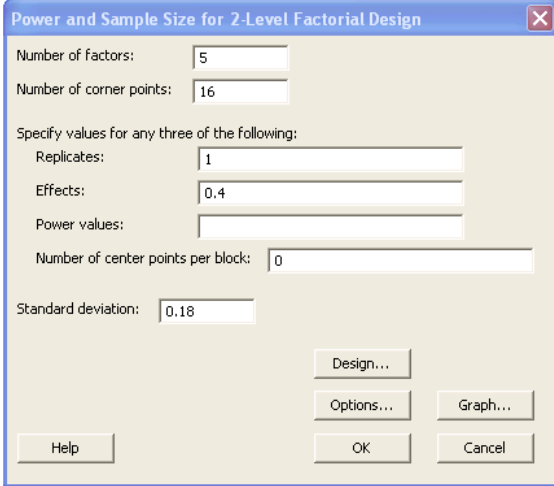
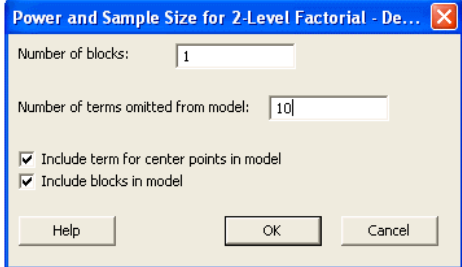
Factors: 5 Base Design: 5, 16
Blocks: none

Number of terms omitted from model: 10

Center Points	Effect	Reps	Total Runs	Power
0	0.4	1	16	0.978702

Power Curve for 2-Level Factorial Design

MTB >

Example 8.18 How many replicates of a 9-variable 12-run Plackett-Burman design are required to detect a difference $\delta = 7000$ between two levels of a variable with 90% power if the standard error is expected to be $\sigma_\epsilon = 4000$?

Solution: Plackett-Burman designs are resolution III, so their models may contain only main effects. If the experiment is run with only one replicate, then $df_\epsilon = (12 - 1) - 9 = 2$ and the large-sample approximation is obviously not satisfied. If enough replicates are run so that the large-sample approximation is satisfied, then with $\alpha = 0.05$, $\beta = 0.10$ and $t \simeq z$, the approximate number of replicates required is

$$\begin{aligned}
 n &\geq \frac{4}{12} (1.96 + 1.282)^2 \left(\frac{4000}{7000} \right)^2 \\
 &\geq 2.
 \end{aligned}$$

Another iteration confirms that two replicates are sufficient to achieve 90% power.

From MINITAB> Stat> Power and Sample Size> Plackett-Burman Design:

```

MTB > Power;
SUBC>   PBDesign 9 12;
SUBC>   Effect 7000;
SUBC>   Power 0.90;
SUBC>   CPBlock 0;
SUBC>   Sigma 4000;
SUBC>   FitC;
SUBC>   GPCurve.

```

Power and Sample Size

Plackett-Burman Design

Alpha = 0.05 Assumed standard deviation = 4000

Factors: 9 Design: 12
Center pts (total): 0

Center Points	Effect	Reps	Total Runs	Target Power	Actual Power
0	7000	2	24	0.9	0.975221

8.7 Two-Level Factorial Designs with Centers

Example 8.19 Calculate the power to detect a difference $\delta = 1400$ between two levels of a study variable in a 2^3 design with three replicates built in blocks with two center points per block. Include terms for main effects, two-factor interactions, lack of fit, and blocks in the model. The standard error is expected to be $\sigma_\epsilon = 1000$.

Solution: The experiment will have $3(2^3 + 2) = 30$ total observations, so the total degrees of freedom will be $df_{total} = 29$. The degrees of freedom for the model will be

$$\begin{aligned}
 df_{model} &= df_{blocks} + df_{main\ effects} + df_{interactions} + df_{LOF} \\
 &= 2 + 3 + 3 + 1 \\
 &= 9,
 \end{aligned}$$

so the error degrees of freedom will be

$$df_\epsilon = df_{total} - df_{model} = 29 - 9 = 20.$$

The power π to reject $H_0 : \delta = 0$ for the main effect of any one of the study variables is given by Equation 8.21 with one numerator and twenty denominator degrees of freedom where the F distribution noncentrality parameter, as given in Equation 8.22, is

$$\begin{aligned}
 \phi &= 3 \times 2^{3-2} \left(\frac{1400}{1000} \right)^2 \\
 &= 11.76.
 \end{aligned}$$

With $\alpha = 0.05$ and

$$\begin{aligned}
 F_{1-\alpha} &= F_{1-\pi, \phi} \\
 F_{0.95} &= 4.351 = F_{1-\pi, 8.17},
 \end{aligned}$$

we find the power to be $\pi = 0.903$.

From MINITAB> Stat> Power and Sample Size> 2-Level Factorial Design:

```

MTB > Power;
SUBC>   FFDesign 3 8;
SUBC>   Reps 3;
SUBC>   Effect 1400;
SUBC>   CPBlock 2;
SUBC>   Sigma 1000;
SUBC>   Blocks 3;
SUBC>   Omit 1;
SUBC>   FitC;
SUBC>   FitB;
SUBC>   GPCurve.

Power and Sample Size

2-Level Factorial Design

Alpha = 0.05   Assumed standard deviation = 1000

Factors: 3   Base Design: 3, 8
Blocks: 3

Number of terms omitted from model: 1
Including a term for center points in model.
Including blocks in model.

Center
Points
Per
Block Effect Reps Total Runs Power
2 1400 3 30 0.903298

Power Curve for 2-Level Factorial Design

MTB >

```

Example 8.20 Determine the ratio of the precisions of the estimates for the lack of fit and main effects in a 2^4 plus centers design when two center points are used per replicate.
Solution: From Equation 8.39 with $k = 4$ and $n_0 = 2$ the ratio of the lack of fit and main effect precision estimates will be

$$\begin{aligned} \frac{\delta_{**}}{\delta} &= \sqrt{1 + \frac{2^4}{2}} \\ &= 3. \end{aligned}$$

That is, the confidence interval for the lack of fit estimate will be three times wider than the confidence interval for the main effects.

8.8 Response Surface Designs

Example 8.21 How many replicates of a three-variable Box-Behnken design are required to estimate the regression coefficients associated with main effects, two-factor interactions, and quadratic terms to within $\delta = \pm 2$ with 95% confidence if the standard error is expected to be $\sigma_\epsilon = 5$?

Solution: A first estimate for the number of replicates required to estimate the regression coefficients associated with main effects is given by Equation 8.43 with $t_{0.025} \simeq 2$ and, from Table 8.5 for the $BB(3)$ design, $SS_{Main\ Effects} = 8$ is

$$\begin{aligned} n &\geq \frac{1}{8} \left(\frac{2 \times 5}{2} \right)^2 \\ &\geq 4. \end{aligned}$$

With $n = 4$ replicates, the error degrees of freedom will be

$$\begin{aligned} df_{\epsilon} &= df_{total} - df_{model} \\ &= (4 \times 15 - 1) - 9 \\ &= 50, \end{aligned}$$

so the approximation for $t_{0.025}$ is justified. Another iteration with $n = 3$ replicates indicates that the precision of the regression coefficient estimates would be slightly greater than $\delta = 2$, so $n = 4$ replicates are required.

From Table 8.5 for two-factor interactions, $SS_{Interaction} = 4$, so the number of replicates required to estimate the regression coefficients associated with two-factor interactions with confidence interval half-width $\delta = 2$ with 95% confidence is

$$\begin{aligned} n &\geq \frac{1}{4} \left(\frac{t_{0.025} \times 5}{2} \right)^2 \\ &\geq 6. \end{aligned}$$

From Table 8.5 for quadratic terms, $SS_{Quadratic} = 3.694$, so the number of replicates required to estimate the regression coefficients associated with quadratic terms is

$$\begin{aligned} n &\geq \frac{1}{3.694} \left(\frac{t_{0.025} \times 5}{2} \right)^2 \\ &\geq 7. \end{aligned}$$

Chapter 9

Reliability and Survival

9.1 Reliability Parameter Estimation

Example 9.1 How many units must be tested to failure to determine, with 20% precision and 95% confidence, the exponential mean life μ ?

Solution: From Equation 9.6 with $\alpha = 0.05$ and $\delta = 0.2$, the required number of failures is

$$r = \left(\frac{1.96}{0.20} \right)^2 = 97.$$

From PASS> Means> One> Inequality (Exponential):

The screenshot shows the PASS software interface. The left window, titled "PASS: Mean: Exponential: 1", displays the input parameters for the analysis. The right window, titled "PASS: Mean: Exponential: 1 Output", displays the results of the analysis.

Input Parameters (Left Window):

- Find (Solve For): Beta and Power
- Alternative Hypothesis: Ha: Theta0 <=> Theta1
- Theta0 (Baseline Mean Life): 100
- Termination Criterion: Fixed Time using Theta-hat
- Theta1 (Alternative Mean Life): 120
- Replacement Method: Without Replacement
- N (Sample Size): 97
- Alpha (Producer's Risk): 0.05
- t0 (Test Duration Time): 10000
- Beta (Consumer's Risk):
- E(t0) based on Theta1
- r (Number of Failures): 97

Output Results (Right Window):

Exponential Mean Power Analysis

Numeric Results
Test Based on Theta-hat with Fixed Running Time t0 and Without Replacement Sampling.
H0: Theta = Theta0. Ha: Theta = Theta1 <> Theta0. Reject H0 if Theta-hat <= Theta L or Theta-hat >= Theta U.

Power	N	t0	Theta0	Theta1	Target Alpha	Actual Alpha	Target Beta	Actual Beta	Theta L	Theta U
0.50379	9710000.000	100.0	120.0	0.05000	0.05000	0.49621	80.1	119.9		

Summary Statements
A sample size of 97 achieves 50% power to detect the difference between the null hypothesis mean lifetime of 100.0 and the alternative hypothesis mean lifetime of 120.0 at a 0.05000 significance level (alpha) using a two-sided test based on the elapsed time. Failing items are not replaced with new items. The study is terminated when it has run for 10000.000 time units.

From MINITAB> Stat> Reliability/Survival> Test Plans> Estimation:

```
MTB > Etestplan;
SUBC> EPtile 63.2;
SUBC> Dlower 20;
SUBC> Weibull;
SUBC> SetS 1;
SUBC> ScLocation 100;
SUBC> TwoSided.
```

Estimation Test Plans

Uncensored data

Estimated parameter: 63.2th percentile
 Calculated planning estimate = 99.9672
 Target Confidence Level = 95%
 Precision in terms of the lower bound of a two-sided confidence interval.

Planning distribution: Exponential
 Scale = 100

Precision	Sample Size	Actual Confidence Level
20	78	95.1330

```
MTB > Etestplan;
SUBC> EPtile 63.2;
SUBC> Dupper 20;
SUBC> Weibull;
SUBC> SetS 1;
SUBC> ScLocation 100;
SUBC> TwoSided.
```

Estimation Test Plans

Uncensored data

Estimated parameter: 63.2th percentile
 Calculated planning estimate = 99.9672
 Target Confidence Level = 95%
 Precision in terms of the upper bound of a two-sided confidence interval.

Planning distribution: Exponential
 Scale = 100

Precision	Sample Size	Actual Confidence Level
20	116	95.0499

The screenshot shows the 'Estimation Test Plans' dialog box. The 'Parameter to be Estimated' section has 'Percentile for percent' selected with a value of 63.2. The 'Precisions as distances from bound of CI to estimate' section has 'Lower bound' selected with a value of 20. The 'Assumed distribution' is set to 'Weibull'. The 'Specify planning values for two of the following' section has 'Shape (Weibull) or scale (other distributions)' set to 1 and 'Scale (Weibull or expo) or location (other dist)' set to 100. There are buttons for 'Right Cens...', 'Interval Cens...', 'Options...', 'OK', 'Cancel', and 'Help'.

The screenshot shows the 'Estimation Test Plans' dialog box. The 'Parameter to be Estimated' section has 'Percentile for percent' selected with a value of 63.2. The 'Precisions as distances from bound of CI to estimate' section has 'Upper bound' selected with a value of 20. The 'Assumed distribution' is set to 'Weibull'. The 'Specify planning values for two of the following' section has 'Shape (Weibull) or scale (other distributions)' set to 1 and 'Scale (Weibull or expo) or location (other dist)' set to 100. There are buttons for 'Right Cens...', 'Interval Cens...', 'Options...', 'OK', 'Cancel', and 'Help'.

Example 9.2 How many units must be tested to failure to determine, with 20% precision and 95% confidence, any failure percentile under the assumption that the reliability distribution is exponential?

Solution: The conditions required to estimate the failure percentiles are the same as those in Example 9.1, so the same number of failures required is $r = 97$.

From PASS> Means> One> Inequality (Exponential):

Exponential Mean Power Analysis

Numeric Results
 Test Based on Theta-hat with Fixed Running Time t0 and Without Replacement Sampling.
 H0: Theta = Theta0. Ha: Theta = Theta1 <> Theta0. Reject H0 if Theta-hat <= Theta L or Theta-hat >= Theta U.

Power	N	t0	Theta0	Theta1	Target Alpha	Actual Alpha	Target Beta	Actual Beta	Theta L	Theta U
0.50379	97	10000.000	100.0	120.0	0.05000	0.05000		0.49621	80.1	119.9

Summary Statements
 A sample size of 97 achieves 50% power to detect the difference between the null hypothesis mean lifetime of 100.0 and the alternative hypothesis mean lifetime of 120.0 at a 0.05000 significance level (alpha) using a two-sided test based on the elapsed time. Failing items are not replaced with new items. The study is terminated when it has run for 10000.000 time units.

Example 9.3 How many units must be tested to failure in an experiment to determine, with 95% confidence, the exponential reliability to within 10% of its true value if the expected reliability is 80%?

Solution: From Equation 9.12 with $\alpha = 0.05$, $\delta = 0.10$, and $\hat{R} = 0.80$, the required number of failures is

$$r = \left(\frac{1.96 \ln(0.80)}{0.10} \right)^2 = 20.$$

Example 9.4 How many units must be tested to failure to estimate, with 20% precision and 95% confidence, the Weibull scale factor if the shape factor is known to be $\beta = 2$?

Solution: The goal of the experiment is to obtain a confidence interval for the Weibull scale factor of the form given by Equation 9.17 with $\delta = 0.20$ and $\alpha = 0.05$. From Equation 9.19 the required number of failures is

$$r = \left(\frac{1.96}{2 \times 0.20} \right)^2 = 25.$$

The Weibull scale parameter is the 63.2% percentile. From MINITAB> Stat> Reliability/Survival> Test Plans> Estimation:

```

MTB > Etestplan;
SUBC> EPtile 63.2;
SUBC> Dlower 20;
SUBC> Weibull;
SUBC> SetS 2;
SUBC> ScLocation 100;
SUBC> TwoSided.

```

Estimation Test Plans

Uncensored data

Estimated parameter: 63.2th percentile
 Calculated planning estimate = 99.9836
 Target Confidence Level = 95%

Precision in terms of the lower bound of a two-sided confidence interval.

Planning distribution: Weibull
 Scale = 100, Shape (true value) = 2

Precision	Sample Size	Actual Confidence Level
20	20	95.4090

The MINITAB solution for the upper bound is $n = 29$, so the average of the two sample sizes is consistent with the approximate solution.

Example 9.5 How many units must be tested to failure to estimate, with 95% confidence, the Weibull shape parameter to within 20% of its true value?

Solution: The goal of the experiment is to produce a 95% confidence interval for β of the form given by Equation 9.20 with $\delta = 0.20$. From Equation 9.24 with $\alpha = 0.05$, the required number of failures is

$$r = 6 \left(\frac{1.96}{\pi \times 0.20} \right)^2 = 59.$$

Example 9.6 How many units must be tested to failure to estimate the Weibull reliability with 5% precision and 95% confidence when the expected reliability is 90%? Assume that the Weibull shape factor is known.

Solution: The desired confidence interval will have the form of Equation 9.31. From Equation 9.35 with $\delta = 0.05$, $\alpha = 0.05$, and $\hat{R} = 0.90$, the required number of failures is

$$r = \left(\frac{1.645}{0.05} \left(\frac{1 - 0.9}{0.9} \right) \right)^2 = 19.$$

Example 9.7 An experiment is planned to estimate, with 95% confidence, the time at which 10% of units will fail to within 1000 hours. The life distribution is expected to be normal with $\hat{\sigma}_t = 2000$ and all units will be tested to failure.

Solution: With $z_{\alpha/2} = z_{0.025} = 1.96$ and $z_f = z_{0.10} = 1.282$ in Equation 9.40, the sample size is

$$n = \left(\frac{1.96 \times 2000}{1000} \right)^2 \left(1 + \frac{(1.282)^2}{2} \right) = 28.$$

From MINITAB> Stat> Reliability/Survival> Test Plans> Estimation:

```
MTB > Etestplan;
SUBC> EPtile 10;
SUBC> Dupper 1000;
SUBC> Normal;
SUBC> ShScale 2000;
SUBC> ScLocation 0;
SUBC> TwoSided.
```

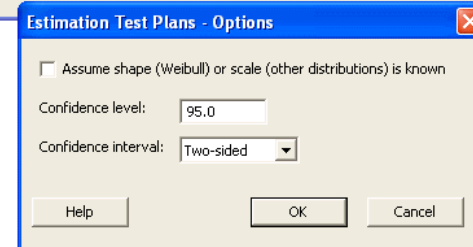
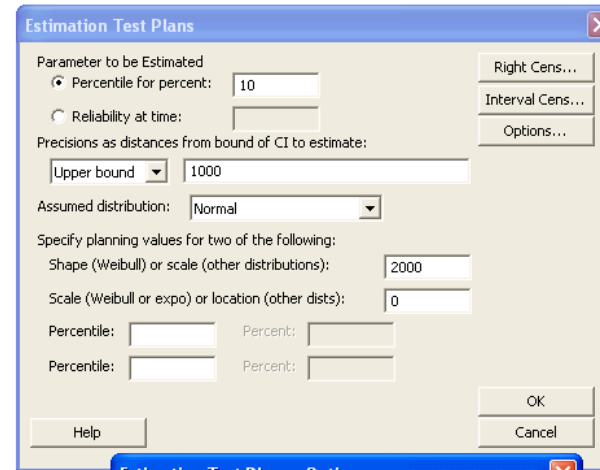
Estimation Test Plans

Uncensored data

Estimated parameter: 10th percentile
 Calculated planning estimate = -2563.10
 Target Confidence Level = 95%
 Precision in terms of the upper bound of a two-sided confidence interval.

Planning distribution: Normal
 Location = 0, Scale = 2000

Precision	Sample Size	Actual Confidence Level
1000	28	95.0065



Example 9.8 What sample size is required to estimate, with 95% confidence, the 24000 hour failure probability of a product to within 2% if the life distribution is expected to be normal with $\mu \simeq 20000$ and $\sigma \simeq 2000$?

Solution: With $x = 24000$ and $\hat{z} = (24000 - 20000)/2000 = 2$, the required confidence interval for the 24 hour failure probability has the form

$$P \left(\hat{\Phi}(2) - 0.02 < \Phi(x = 24000; \mu, \sigma) < \hat{\Phi}(2) + 0.02 \right) = 0.95.$$

From Equation 9.44 the required sample size to obtain this interval is

$$\begin{aligned} n &= \left(\frac{z_{0.025} \varphi(2)}{0.02} \right)^2 \left(1 + \frac{1}{2} 2^2 \right) \\ &= \left(\frac{1.96 \times 0.0540}{0.02} \right)^2 (3) \\ &= 85. \end{aligned}$$

From MINITAB> Stat> Reliability/Survival> Test Plans> Estimation the calculated sample sizes are $n = 20$ for the upper bound and $n = 147$ for the lower bound. Their average, $n = 84$, is in good agreement with the approximate solution.

The screenshot displays the Minitab interface with three overlapping windows related to Estimation Test Plans.

Top Window: Estimation Test Plans (Main)

Uncensored data

Estimated parameter: Reliability at time = 16000
 Calculated planning estimate = 0.977250
 Target Confidence Level = 95%
 Precision in terms of the upper bound of a two-sided confidence interval.

Planning distribution: Normal
 Location = 20000, Scale = 2000

Precision	Sample Size	Actual Confidence Level
0.02	20	95.4938

Bottom Window: Estimation Test Plans (Options)

Uncensored data

Estimated parameter: Reliability at time = 16000
 Calculated planning estimate = 0.977250
 Target Confidence Level = 95%
 Precision in terms of the lower bound of a two-sided confidence interval.

Planning distribution: Normal
 Location = 20000, Scale = 2000

Precision	Sample Size	Actual Confidence Level
0.02	147	95.0309

Right Window: Estimation Test Plans (Dialog)

Parameter to be Estimated

Percentile for percent: 63.2
 Reliability at time: 16000

Precisions as distances from bound of CI to estimate:

Lower bound: 0.02

Assumed distribution: Normal

Specify planning values for two of the following:

Shape (Weibull) or scale (other distributions): 2000
 Scale (Weibull or expo) or location (other dists): 20000

Percentile: Percent:
 Percentile: Percent:

Buttons: Right Cens..., Interval Cens..., Options..., OK, Cancel, Help

Bottom-Right Window: Estimation Test Plans - Options

Assume shape (Weibull) or scale (other distributions) is known

Confidence level: 95.0
 Confidence interval: Two-sided

Buttons: Help, OK, Cancel

9.2 Reliability Demonstration Tests

Example 9.9 How many units must be tested for 200 hours without any failures to show, with 95% confidence, that the *MTTF* of a system exceeds 400 hours. The life distribution is exponential and the test is time terminated.

Solution: We must determine the value of n with $r = 0$ failures in $t' = 200$ hours of testing such that

$$P(400 < \mu < \infty) = 0.95.$$

From the f' equation for the exponential distribution from Table 9.1 with $\mu_0 = 400$, the $t' = 200$ hour failure probability is

$$\begin{aligned} f' &= 1 - e^{-t'/\mu_0} \\ &= 1 - e^{-200/400} \\ &= 0.3935. \end{aligned}$$

With $r = 0$ and $\alpha = 0.05$ the smallest value of n that satisfies Equation 9.46 is $n = 6$ because

$$(b(0; 6, 0.3935) = 0.04977) < (\alpha = 0.05).$$

From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

```
MTB > DtestPlan 0;
SUBC> MTTF 400 ;
SUBC> TTime 200 ;
SUBC> Exponential;
SUBC> GPOGraph.
```

Demonstration Test Plans

```
Reliability Test Plan
Distribution: Exponential
MTTF Goal = 400, Target Confidence Level = 95%
```

Failure Test	Testing Time	Sample Size	Actual Confidence Level
0	200	6	95.0213

The screenshot shows the 'Demonstration Test Plans' dialog box in Minitab. The 'Minimum Value to be Demonstrated' section has three radio buttons: 'Scale (Weibull or expo) or location (other dist):', 'Percentile:', and 'Reliability:'. The 'MTTF:' radio button is selected, and the value '400' is entered in the adjacent text box. The 'Maximum number of failures allowed:' is set to '0'. The 'Testing times for each unit:' is set to '200'. The 'Distribution Assumptions' section has a dropdown menu for 'Distribution:' set to 'Exponential'. There are 'OK', 'Cancel', 'Help', 'Graphs...', and 'Options...' buttons.

Example 9.10 Determine how long ten units must be life tested with no more than one failure during the test period to demonstrate, with 90% confidence, that the mean life is greater than 1000 hours. Assume that the life distribution is exponential.

Solution: The goal of the experiment is to demonstrate that

$$P(1000 < \mu < \infty) = 0.90.$$

With $n = 10$ and $r = 1$, Equation 9.46 gives

$$\sum_{x=0}^1 b(x; n = 10, f') = 0.10 \quad (9.48)$$

which is satisfied by $f' = 0.337$. From the t' equation for the exponential distribution from Table 9.1, the required duration of the test in hours is

$$\begin{aligned} t' &= -\mu_0 \ln(1 - f') \\ &= -1000 \ln(1 - 0.337) \\ &= 411. \end{aligned}$$

From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

```
MTB > DtestPlan 1;
SUBC> MTTF 1000 ;
SUBC> Sample 10;
SUBC> Exponential;
SUBC> GPOPGraph;
SUBC> Confidence 90.
```

Demonstration Test Plans

Reliability Test Plan
Distribution: Exponential
MTTF Goal = 1000, Actual Confidence Level = 90%

Failure Test	Sample Size	Testing Time
1	10	410.751

Example 9.11 Determine how long ten units must be life tested with no more than one failure during the test period to demonstrate, with 90% confidence, that the Weibull scale factor is at least 1000 hours. Assume that the Weibull shape factor is known to be $\beta = 2$.

Solution: The design parameters of the RDT are the same as in Example 9.10, so Equation 9.48 still applies and the end-of-test failure probability is $f' = 0.337$. From the t' equation for the Weibull distribution from Table 9.1, the required duration of the test in hours is

$$\begin{aligned} t' &= \eta_0 (-\ln(1 - f'))^{\frac{1}{\beta}} \\ &= 1000 (-\ln(1 - 0.337))^{\frac{1}{2}} \\ &= 641. \end{aligned}$$

From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

```

MTB > DtestPlan 1;
SUBC> ScLocation 1000;
SUBC> Sample 10;
SUBC> Weibull;
SUBC> ShScale 2;
SUBC> GPOGraph;
SUBC> Confidence 90.

```

Demonstration Test Plans

```

Substantiation Test Plan
Distribution: Weibull, Shape = 2
Scale Goal = 1000, Actual Confidence Level = 90%

```

Failure Test	Sample Size	Testing Time
1	10	640.898

Example 9.12 Determine how long ten units must be life tested with no more than one failure during the test period to demonstrate, with 90% confidence, that the mean life is at least 1000 hours. Assume that the life distribution is normal with $\sigma = 100$.

Solution: The design parameters of the RDT are the same as in Example 9.10, so Equation 9.48 applies and the failure probability is $f' = 0.337$. From the equation for t' from Table 9.1, the required duration of the test in hours is

$$\begin{aligned}
 t' &= \mu + z_{f'}\sigma \\
 &= 10000 + z_{0.337}(100) \\
 &= 10000 + (-0.42 \times 100) \\
 &= 958.
 \end{aligned}$$

From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

```

MTB > DtestPlan 1;
SUBC> MTTF 1000 ;
SUBC> Sample 10;
SUBC> Normal;
SUBC> ShScale 100;
SUBC> GPOGraph;
SUBC> Confidence 90.

```

Demonstration Test Plans

```

Reliability Test Plan
Distribution: Normal, Scale = 100
MTTF Goal = 1000, Actual Confidence Level = 90%

```

Failure Test	Sample Size	Testing Time
1	10	957.892

Example 9.13 How many units must be tested for 400 hours without any failures to demonstrate 90% reliability at 600 hours, with 95% confidence? Assume that the reliability distribution is exponential.

Solution: In terms of the 600 hour failure probability, the goal of the experiment is to demonstrate

$$P(0 < f(600) < 0.10) = 0.95$$

based on a sample of size n tested to $t' = 400$ hours with $r = 0$ failures. From Table 9.2 the equation for f' for the exponential distribution gives

$$\begin{aligned}
 f' &= 1 - (1 - f_0)^{t'/t_0} \\
 &= 1 - (1 - 0.10)^{400/600} \\
 &= 0.0678.
 \end{aligned}$$

With $r = 0$, $f' = 0.0678$, and $\alpha = 0.05$, Equation 9.46 gives

$$b(0; n, 0.0678) \leq 0.05,$$

which is satisfied by $n = 43$.

From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

MTB > DtestPlan 0;
 SUBC> Reliability 0.90;
 SUBC> Time 600 ;
 SUBC> TTime 400 ;
 SUBC> Exponential;
 SUBC> GPOGraph.

Demonstration Test Plans

Reliability Test Plan
 Distribution: Exponential
 Reliability Goal = 0.9, Target Confidence Level = 95%

Failure Test	Testing Time	Sample Size	Actual Confidence Level
0	400	43	95.1215

Minimum Value to be Demonstrated
 Scale (Weibull or expo) or location (other dist):
 Percentile: Percent:
 Reliability: 0.90 Time: 600
 MTF:
 Maximum number of failures allowed: 0
 Sample sizes:
 Testing times for each unit: 400
 Distribution Assumptions
 Distribution: Exponential
 Shape (Weibull) or scale (other distributions):

Example 9.14 How long must ten units be life tested with no more than one failure during the test period to demonstrate, with 80% confidence, that the 3000-hour reliability is at least 90%. Assume that the life distribution is Weibull with $\beta = 1.8$.

Solution: The goal of the experiment is to demonstrate that

$$P(0.90 < R(3000) < 1) = 0.80$$

or in terms of the failure probability

$$P(0 < f(3000) < 0.10) = 0.80.$$

With $n = 10$, $r = 1$, and $\alpha = 0.20$, Equation 9.46 becomes

$$\sum_{x=0}^1 b(x; 10, f') \leq 0.20,$$

which is satisfied by $f' = 0.271$. From the equation for t' for the Weibull distribution from Table 9.2 with $f_0 = 0.10$ and $t_0 = 3000$, the test time is

$$\begin{aligned} t' &= t_0 \left(\frac{\ln(1 - f')}{\ln(1 - f_0)} \right)^{1/\beta} \\ &= 3000 \left(\frac{\ln(1 - 0.271)}{\ln(1 - 0.10)} \right)^{1/1.8} \\ &= 5523. \end{aligned}$$

From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

MTB > DtestPlan 1;
 SUBC> Reliability 0.90;
 SUBC> Time 3000 ;
 SUBC> Sample 10;
 SUBC> Weibull;
 SUBC> ShScale 1.8;
 SUBC> GPOGraph;
 SUBC> Confidence 80.

Demonstration Test Plans

Reliability Test Plan
 Distribution: Weibull, Shape = 1.8
 Reliability Goal = 0.9, Actual Confidence Level = 80%

Failure Test	Sample Size	Testing Time
1	10	5523.01

Demonstration Test Plans dialog box settings:
 Minimum Value to be Demonstrated: Scale (Weibull or expo) or location (other dist):
 Percentile: Percent:
 Reliability: 0.90 Time: 3000
 MTTF:
 Maximum number of failures allowed: 1
 Sample sizes: 10
 Testing times for each unit:
 Distribution Assumptions:
 Distribution: Weibull
 Shape (Weibull) or scale (other distributions): 1.8

Demonstration Test Plans - Options dialog box settings:
 Confidence level: 80

Example 9.15 How many units must be tested for 140 hours with no more than one failure to demonstrate that the 100 hour reliability is at least 95% with 90% confidence? Assume that the reliability distribution is normal with $\sigma = 20$.

Solution: The goal of the experiment is to demonstrate

$$P(0 < f(100) < 0.05) = 0.90$$

based on a sample of size n tested to $t' = 140$ hours with no more than $r = 1$ failures. From Table 9.2, the equation for $z_{f'}$ for the normal distribution gives

$$\begin{aligned} z_{f'} &= z_{f_0} + \left(\frac{t' - t_0}{\sigma} \right) \\ &= z_{0.05} + \left(\frac{140 - 100}{20} \right) \\ &= -1.645 + 2.0 \\ &= 0.355, \end{aligned}$$

which is satisfied by $f' = \Phi(-\infty < z < 0.355) = 0.639$. With $r = 1$, $f' = 0.639$, and $\alpha = 0.10$, Equation 9.46 becomes

$$\sum_{x=0}^1 b(x; n, 0.639) \leq 0.10,$$

which is satisfied by $n = 5$.

From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

Example 9.16 How many units must be tested without any failures to t_0 hours to demonstrate 90% reliability at t_0 hours with 95% confidence? Assume that the distribution is Weibull.

Solution: The goal of the experiment is to demonstrate

$$P(0 < f(t_0) < 0.10) = 0.95.$$

With $f_0 = 0.10$ and $t' = t_0$ in the Weibull equation for f' from Table 9.2

$$\begin{aligned} f' &= 1 - (1 - 0.10) \left(\frac{t_0}{t_0} \right)^\beta \\ &= 0.10. \end{aligned}$$

With $f' = 0.10$, $r = 0$, and $\alpha = 0.05$, Equation 9.46 is

$$b(0; n, 0.10) \leq 0.05,$$

which is satisfied by $n = 29$. This is just a case of the rule of three: $n = 3/f_0$.

Example 9.17 How long must 50 units be tested without any failures to demonstrate that the time at which the first 1% of the population fails exceeds 400 cycles? Assume that the life distribution is exponential and use the 95% confidence level.

Solution: The goal of the experiment is to demonstrate that

$$P(400 < t_{0.01} < \infty) = 0.95.$$

From Equation 9.46 with $r = 0$, $n = 50$, and $\alpha = 0.05$

$$b(0; 50, f') \leq 0.05,$$

which is satisfied by $f' = 0.058$. From the exponential form of t' from Table 9.2 with $t_0 = 400$, the required duration of the test in cycles is

$$\begin{aligned} t' &= t_0 \frac{\ln(1 - f')}{\ln(1 - f_0)} \\ &= 400 \frac{\ln(1 - 0.058)}{\ln(1 - 0.01)} \\ &= 2380. \end{aligned}$$

From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

```
MTB > DtestPlan 0;
SUBC> PTitle 1;
SUBC> Time 400;
SUBC> Sample 50;
SUBC> Exponential;
SUBC> GPOGraph.

Demonstration Test Plans

Reliability Test Plan
Distribution: Exponential
Percentile Goal = 400, Actual Confidence Level = 95%

Failure Test Sample Testing
          Size Time
          0    50 2384.58
```

Example 9.18 How many units must be tested to 30,000 cycles without any failures to demonstrate, with 95% confidence, that the 20,000 cycle reliability is at least 90%? The life distribution is known to be Weibull with $\beta = 3.1$.

Solution: The goal of the experiment is to demonstrate that

$$P(20000 < t_{0.10} < \infty) = 0.95.$$

The equation for f' for the Weibull distribution from Table 9.2 with $f_0 = 0.10$, $t_0 = 20000$, $t' = 30000$, and $\beta = 3.1$ gives

$$\begin{aligned} f' &= 1 - (1 - f_0)^{(t'/t_0)^\beta} \\ &= 1 - (1 - 0.10)^{(30000/20000)^{3.1}} \\ &= 0.3095. \end{aligned}$$

Then, with $r = 0$, $f' = 0.3095$, and $\alpha = 0.05$, Equation 9.46 gives

$$b(0; n, 0.3095) \leq 0.05,$$

which is satisfied by $n = 9$.

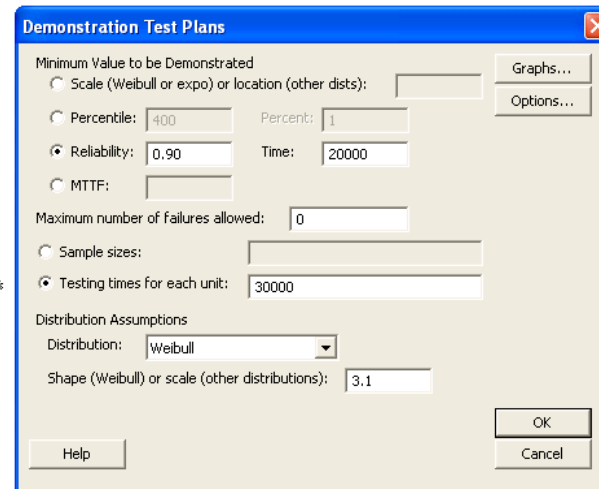
From MINITAB> Stat> Reliability/Survival> Test Plans> Demonstration:

```
MTB > DtestPlan 0;
SUBC> Reliability 0.90;
SUBC> Time 20000 ;
SUBC> TTime 30000 ;
SUBC> Weibull;
SUBC> ShScale 3.1;
SUBC> GPOGraph.
```

Demonstration Test Plans

Reliability Test Plan
 Distribution: Weibull, Shape = 3.1
 Reliability Goal = 0.9, Target Confidence Level = 95%

Failure Test	Testing Time	Sample Size	Actual Confidence Level
0	30000	9	96.4305



9.3 Two-Sample Reliability Tests

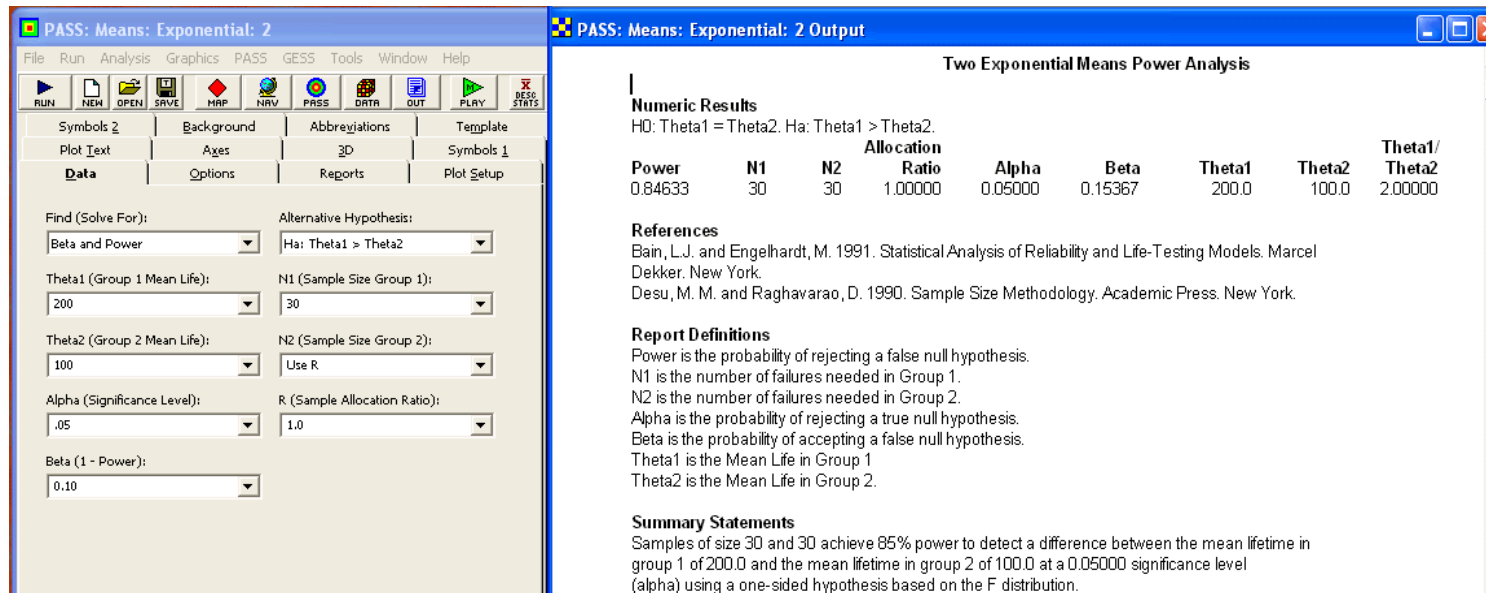
Example 9.19 A reliability experiment is to be performed to compare the mean life of two different product designs. Determine the power to reject $H_0 : \mu_1 = \mu_2$ in favor of $H_A : \mu_1 > \mu_2$ when $\mu_1 = 200$ hours and $\mu_2 = 100$ hours using two different strategies: a) $n_1 = n_2 = 30$ units, all tested to failure and b) $n_1 = 40, n_2 = 20$, and the test will be suspended when 90% of the units from one of the two designs have failed. Assume that both life distributions are exponential

Solution:

a) With $n_1 = n_2 = 30$ units tested to failure, the F test critical value will be $F_{0.95,60,60} = 1.534$ and by Equation 9.55 the power will be

$$\begin{aligned} \pi &= P\left(\left(\frac{100}{200} \times 1.534\right) < F < \infty\right) \\ &= P(0.767 < F < \infty) \\ &= 0.846 \end{aligned}$$

From PASS> Means> Two> Independent> Inequality (Exponential):



b) Under H_A in the second strategy, the second treatment group has fewer units with lower mean life so they should be exhausted first. The time at which 90% or 18 of these units will have failed is expected to be about $t = -100 \ln(0.1) = 230$ hours. At the same time about $40(1 - e^{-230/200}) = 27$ of the units from the first treatment group are expected to fail. If the test is suspended then, with $x_1 = 27$ and $x_2 = 18$, then the power will be

$$\begin{aligned}
 \pi &= P\left(\left(\frac{100}{200} \times F_{0.95,54,36}\right) < F < \infty\right) \\
 &= P(0.842 < F < \infty) \\
 &= 0.721.
 \end{aligned}$$

Under the second strategy, the test will end much earlier (i.e., when the 18th unit with 100 hour mean life fails versus when the 30th unit with 200 hour mean life fails); however, at the penalty of reduced experimental power.

Example 9.20 Determine how many units must be included in a study to compare the survival rates of two treatments using the log-rank test if the control treatment is expected to have about 20% survivors at the end of the study and the study should have 90% power to reject H_0 if the experimental treatment has 40% survivors at the end of the study. Assume that the hazard rates are proportional and that the sample sizes will be equal.

Solution: From the expected end-of-study conditions under H_A the log-hazard ratio is estimated to be

$$r_A \simeq \frac{\ln(0.40)}{\ln(0.20)} = 0.5693,$$

so the required sample size is

$$n_1 = n_2 = \left(\frac{z_{0.05} + z_{0.10}}{\ln(0.5693)}\right)^2 \left(\frac{1}{1-0.2} + \frac{1}{1-0.4}\right) = 79.$$

From PASS> Survival and Reliability> Log-Rank Survival (Simple):

PASS: Log Rank Survival: Simple Output

Log Rank Survival Power Analysis - Simple

Numeric Results						Hazard Ratio	One-Sided Alpha	Beta
Power	N	N1	N2	S1	S2			
0.9009	163	82	81	0.2000	0.4000	0.5693	0.0500	0.0991

Summary Statements
 A one-sided log rank test with an overall sample size of 163 subjects (of which 82 are in group 1 and 81 are in group 2) achieves 90% power at a 0.0500 significance level to detect a difference of 0.2000 between 0.2000 and 0.4000--the proportions surviving in groups 1 and 2, respectively. This corresponds to a hazard ratio of 0.5693. The proportion of patients lost during follow up was 0.0000. These results are based on the assumption that the hazard rates are proportional.

Example 9.21 Compare the power of the log-rank test to the power of the two-sample test for exponential mean life for Example 9.19b.

Solution: Because the hazard rate of an exponential distribution is constant, the proportional hazards assumption is satisfied. At 230 hours with $s_1(t') = 2/20 = 0.10$ and $s_2(t') = 13/40 = 0.325$ the hazard ratio under H_A will be

$$r_A = \frac{\ln(0.325)}{\ln(0.10)} = 0.488.$$

With $n_2/n_1 = 2$, $d_1(t') = 18$, and $d_2(t') = 27$, the z_β value from Equation 9.64 is

$$z_\beta = \frac{1 - 0.488}{1 + 2(0.488)} \sqrt{2(18 + 27)} - 1.96 = 0.50.$$

Then the power for the log-rank test is $\pi = \Phi(-\infty < z < 0.50) \simeq 0.69$, which is slightly less than the power for the two-sample exponential test for mean life, which was $\pi = 0.72$. The two-tailed test was used here to match the power obtained in Example 9.19.

From PASS> Survival and Reliability> Log-Rank Survival (Simple):

The screenshot shows the PASS software interface for a Log Rank Survival Power Analysis. The left pane displays the input parameters, and the right pane shows the resulting numeric results and summary statements.

Log Rank Survival Power Analysis - Simple

Numeric Results

Power	N	N1	N2	S1	S2	Hazard Ratio	One-Sided Alpha	Beta
0.6562	60	40	20	0.1000	0.3250	0.4881	0.0500	0.3438

Summary Statements
 A one-sided log rank test with an overall sample size of 60 subjects (of which 40 are in group 1 and 20 are in group 2) achieves 66% power at a 0.0500 significance level to detect a difference of 0.2250 between 0.1000 and 0.3250—the proportions surviving in groups 1 and 2, respectively. This corresponds to a hazard ratio of 0.4881. The proportion of patients lost during follow up was 0.0000. These results are based on the assumption that the hazard rates are proportional.

Example 9.22 Compare the sample size calculated by Lachin's method to that of Schoenfeld's method in Example 9.20.

Solution: From the information given in the problem statement and Equation 9.66 the sample size by Lachin's method must be

$$n_1 = n_2 = \frac{(1.645 + 1.282)^2}{2 - 0.2 - 0.4} \left(\frac{1 + 0.5693}{1 - 0.5693} \right)^2 = 82,$$

which is in good agreement with Schoenfeld's method, $n = 79$.

See the NCSS/PASS solution shown in Example 9.20. The manual calculation by Schoenfeld's method is in excellent agreement with PASS which uses the same method.

9.4 Interference

Example 9.23 A random sample of component strengths gave $n_S = 100$, $\hat{\mu}_S = 600$, and $\hat{\sigma}_S = 60$ and a random sample of loads gave $n_L = 36$, $\hat{\mu}_L = 450$, and $\hat{\sigma}_L = 40$. Both distributions are known to be normal. Determine the 90% upper confidence limit for the interference failure rate.

Solution: The point estimate for \hat{z}_f is given by Equation 9.73:

$$\hat{z}_f = \frac{450 - 600}{\sqrt{40^2 + 60^2}} = -2.08$$

and the corresponding point estimate for the interference failure rate is

$$\hat{f} = \Phi(-\infty < z < -2.08) = 0.0188.$$

The approximate standard deviation of the \hat{z}_f distribution is given by Equation 9.74:

$$\hat{\sigma}_{\hat{z}_f} = \sqrt{\frac{1}{40^2 + 60^2} \left(\frac{40^2}{36} + \frac{60^2}{100} + \frac{1}{2} \left(\frac{450 - 600}{40^2 + 60^2} \right)^2 \left(\frac{40^4}{36} + \frac{60^4}{100} \right) \right)} = 0.178.$$

Then, from Equation 9.76 with $z_{0.10} = 1.282$

$$\begin{aligned}\hat{z}_{f_U} &= -2.08 + 1.282 \times 0.178 \\ &= -1.85,\end{aligned}$$

so from Equation 9.75 the 90% upper confidence limit for the interference failure probability is

$$\begin{aligned}\hat{f}_U &= \Phi(-\infty < z < -1.85) \\ &= 0.032.\end{aligned}$$

That is, on the basis of the sample data, we can claim that the one-sided upper 90% confidence interval for the interference failure rate is

$$P(0 < f < 0.032) = 0.90.$$

Example 9.24 What sample size is required to demonstrate that the interference failure probability is less than 0.1% with 90% confidence if the strength distribution is known to be normal with $\mu_S = 20$ and $\sigma_S = 2$ and the load distribution is expected to be normal with $\hat{\mu}_L = 13$ and $\hat{\sigma}_L = 1$?

Solution: The point estimate for the interference failure probability determined from the S parameters and the L parameter estimates is

$$\begin{aligned}\hat{f} &= \Phi(-\infty < z < \hat{z}_f) \\ &= \Phi\left(-\infty < z < \frac{13 - 20}{\sqrt{1^2 + 2^2}}\right) \\ &= \Phi(-\infty < z < -3.13) \\ &= 0.000874.\end{aligned}$$

The sample size required to study the load distribution is given by swapping the relevant S and L subscripts in Equation 9.78:

$$\begin{aligned}n_L &= \left(\frac{z_\alpha}{\hat{z}_{f_U} - \hat{z}_f}\right)^2 \left(\frac{\hat{\sigma}_L^2}{\hat{\sigma}_L^2 + \sigma_S^2}\right) \left(1 + \frac{\hat{\sigma}_L^2}{2} \left(\frac{\hat{\mu}_L - \mu_S}{\hat{\sigma}_L^2 + \sigma_S^2}\right)^2\right) \\ &= \left(\frac{z_{0.10}}{z_{0.001} - z_{0.000874}}\right)^2 \left(\frac{1^2}{1^2 + 2^2}\right) \left(1 + \frac{1^2}{2} \left(\frac{13 - 20}{1^2 + 2^2}\right)^2\right) \\ &= \left(\frac{1.282}{-3.09 - (-3.13)}\right)^2 \left(\frac{1^2}{1^2 + 2^2}\right) \left(1 + \frac{1^2}{2} \left(\frac{13 - 20}{1^2 + 2^2}\right)^2\right) \\ &= 407.\end{aligned}\tag{9.1}$$

Example 9.25 What sample size is required to determine the 95% two-sided confidence interval for the exponential-exponential interference failure rate if the confidence limits must be within 50% of the predicted mean failure rate?

Solution: The goal of the experiment is to obtain a confidence interval for the exponential-exponential interference failure rate f of the form

$$\Phi(0.50\hat{f} < f < 1.50\hat{f}) = 0.95.$$

With $z_{0.025} = 1.96$ and $\delta = 0.50$ in Equation 9.90, the required equal sample sizes are

$$n_L = n_S = 2 \left(\frac{1.96}{0.50} \right)^2 = 31.$$

Example 9.26 How many measurements of mating components in a device must be taken to demonstrate, with 95% confidence, that their true interference failure rate does not exceed the observed failure rate by 20% if the two distributions are known to be Weibull with $\beta_S = 2.5$ and $\beta_L = 1.5$, respectively?

Solution: The goal of the experiment is to acquire sufficient information to demonstrate the following one-sided upper confidence interval for the interference failure rate f :

$$P\left(0 < f < \hat{f}(1 + 0.2)\right) = 0.95.$$

With $\delta = 0.2$ and $\alpha = 0.05$ in Equation 9.100, we obtain the sample size

$$\begin{aligned} n &= \left(\frac{1.645}{0.2 \times \Gamma\left(1 + \frac{2.5}{1.5}\right)} \right)^2 \left(1 + \frac{2.5^2}{1.5^2} \right) \\ &= 113. \end{aligned}$$

Chapter 10

Statistical Quality Control

10.1 Statistical Process Control

Example 10.1 Evaluate the following control chart run rule: A process is judged to be out of control if at least two of three consecutive observations falls beyond the same 2σ limit on the chart.

Solution: The rule is easy to identify on the chart, so it satisfies the first condition for a valid run rule. If the process is in control and the distribution of the statistic (call it w) is approximately normal, then the probability that any point on the chart falls above $\mu_w + 2\sigma_w$ is $p = \Phi(2 < z < \infty) = 0.023$. The probability that at least $x = 2$ of $n = 3$ consecutive points fall above that limit is given by the binomial probability

$$\sum_{x=2}^3 b(x; n = 3, p = 0.023) = 0.0016.$$

Because this pattern could also appear on the bottom half of the chart, the type I error rate for this rule is $\alpha = 2(0.0016) = 0.0032$, which is acceptably low, so the rule meets the second requirement for a valid control chart rule. If the process mean shifted to $\mu_w + 2\sigma_w$, then the probability that an observation would fall beyond $\mu_w + 2\sigma_w$ is $p = 0.5$ and the corresponding power of the rule is

$$\pi = \sum_{x=2}^3 b(x; n = 3, p = 0.5) = 0.5.$$

This meets the third requirement of a valid control chart rule. Because all three conditions are satisfied, that is: 1) the rule is easy to recognize, 2) it has a low type I error rate, and 3) it has good power to detect shifts in the process, then it is a valid control chart run rule.

Example 10.2 One of the weaknesses of defects charts when the sampling unit is small is that it is not possible to declare a process to be out of control on the lower side of the chart with a single observation. Evaluate the following special run rule for defects charts: If a defects chart's sampling unit size is sufficient to deliver $\lambda \geq 3$, then the process is out of control if two consecutive sampling units have 0 defects.

Solution: The chart obviously meets the first and third conditions for valid control chart run rules, but it is not clear if the second condition (low type I error rate) is satisfied. If the mean defect rate is $\lambda = 3$, then the probability of a sampling unit having 0 defects when the process is in control ($H_0 : \lambda = 3$) is $Poisson(x = 0; \lambda = 3) = 0.05$, a rather common occurrence. Under the same conditions, the probability of observing two consecutive zeros is $b(x = 2; n = 2, p = 0.05) = 0.0025$, but this is just the type I error rate for the rule. Because $\alpha = 0.0025$ is acceptably low, the rule meets all three conditions for a valid control chart run rule.

Example 10.3 When Walter Shewhart invented control charts, he expected that an operator would be using about four run rules to interpret at most three control charts. If each of Shewhart's run rules had $\alpha_i \simeq 0.004$, what is the expected overall type I error rate?

Solution: From Equation 10.4

$$\alpha_{FAMILY} \simeq 3 \times 4 \times 0.004 = 0.048.$$

That is, with three simultaneous charts and four run rules, Shewhart expected about 5% of the sampling intervals to result in type I errors.

Example 10.4 What is the minimum sample size required to have a positive lower control limit on a defectives chart if the process fraction defective is expected to be $p = 0.01$?

Solution: From Equation 10.6 the required sample size is

$$\begin{aligned} n &> \left(\frac{9(1 - 0.01)}{0.01} \right) \\ &> 891. \end{aligned}$$

Example 10.5 What is the smallest sampling unit size for a defects chart that will deliver no more than about 5% zero-defect observations when the process delivers 0.6 defects per unit?

Solution: From Equation 10.13 the mean number of defects per sampling unit is the value of λ that satisfies the condition

$$Poisson(x = 0; \lambda) = 0.05, \tag{10.1}$$

which is $\lambda = 3$. Consequently, the sampling unit size must be $3/0.6 = 5$ units.

Example 10.6 Calculate the power to reject $H_0 : \mu = 30$ when $\mu = 32$ if an \bar{x} chart is kept using $n = 4$ and $\sigma_x = 2$. Also determine the corresponding ARL .

Solution: The \bar{x} chart control limits will fall at

$$UCL/LCL = 30 \pm \frac{3 \times 2}{\sqrt{4}} = 33/27.$$

Assuming that the only out-of-control rule used is one point beyond three sigma limits, the power is given by Equation 10.18

$$\begin{aligned} \pi &= 1 - \Phi \left(27 < \bar{x} < 33; \mu_x = 32, \sigma_{\bar{x}} = \frac{2}{\sqrt{4}} = 1 \right) \\ &= 1 - \Phi(-5 < z < 1) \\ &= 0.16. \end{aligned}$$

Under the same conditions, the average number of subgroups that will have to be drawn after a shift from $\mu = 30$ to $\mu = 32$ to detect the shift is

$$ARL = \frac{1}{0.16} = 6.3.$$

From MINITAB> Stat> Power and Sample Size> 1-Sample Z:

MTB > Power;
 SUBC> ZOne;
 SUBC> Sample 4;
 SUBC> Difference 2;
 SUBC> Sigma 2;
 SUBC> Alpha 0.0026;
 SUBC> GPCurve.

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus not = null)
 Calculating power for mean = null + difference
 Alpha = 0.0026 Assumed standard deviation = 2

Difference	Sample Size	Power
2	4	0.155900

Power and Sample Size for 1-Sample Z

Specify values for any two of the following:

Sample sizes: 4
 Differences: 2
 Power values:
 Standard deviation: 2

Power and Sample Size for 1-Sample Z - Options

Alternative Hypothesis
 Less than
 Not equal
 Greater than

Significance level: 0.0026

Store sample sizes in:
 Store differences in:
 Store power values in:
 Select

10.2 Process Capability

Example 10.7 What sample size is required to determine c_p to within 10% of its true value with 90% confidence?

Solution: With $\delta = 0.10$ and $\alpha = 0.10$ in Equation 10.26 the required sample size is

$$n = \frac{1}{2} \left(\frac{1.645}{0.10} \right)^2 = 136.$$

The c_p value is inversely proportional to the standard deviation, so a standard deviation calculator can be used to determine the sample size required for a confidence interval for c_p . From **PASS> Variance> Variance: 1 Group**:

The screenshot shows the PASS: Variance: 1 software interface. The left pane displays input parameters for a one-variance power analysis:

- Find (Solve For): N
- Scale: Standard Deviation
- V0 (Baseline Variance): 100
- Alternative Hypothesis: Ha: V0 <> V1
- V1 (Alternative Variance): 110
- Alpha (Significance Level): 0.10
- N (Sample Size): [empty]
- Beta (1-Power): 0.5
- Known Mean

The right pane displays the output results for a One Variance Power Analysis:

Numeric Results when H0: S0 = S1 versus Ha: S0 <> S1

Power	N	S0	S1	Alpha	Beta
0.500628	141	100.0000	110.0000	0.100000	0.499372

References

Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
 Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press. Ames, Iowa.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population.
 S0 is the value of the population standard deviation under the null hypothesis.
 S1 is the value of the population standard deviation under the alternative hypothesis.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

A sample size of 141 achieves 50% power to detect a difference of 10.0000 between the null hypothesis standard deviation of 100.0000 and the alternative hypothesis standard deviation of 110.0000 using a two-sided, Chi-square hypothesis test with a significance level (alpha) of 0.100000.

Example 10.8 What sample size is required to estimate c_{pk} to within 5% of its true value with 90% confidence if $c_{pk} = 1.0$ is expected?

Solution: From Equation 10.30 with $\delta = 0.05$ and $\alpha = 0.05$ the required sample size is

$$n \simeq \left(\frac{1.645}{0.05} \right)^2 \left(\frac{1}{9(1.0)^2} + \frac{1}{2} \right) = 662.$$

Example 10.9 What sample size is required to estimate c_{pk} to within 5% of its true value with 90% confidence if c_{pk} is expected to be very large?

Solution: From Equation 10.31 with $\delta = 0.05$ and $\alpha = 0.05$ the required sample size is

$$n \simeq \frac{1}{2} \left(\frac{1.645}{0.05} \right)^2 = 541.$$

Example 10.10 Determine the sample size required to reject $H_0 : c_p = 1.33$ in favor of $H_A : c_p > 1.33$ with 90% power when $c_p = 1.5$.

Solution: With $(c_p)_0 = 1.33$, $(c_p)_1 = 1.5$, $\alpha = 0.05$, and $\beta = 0.10$ in Equation 10.33, the required sample size is

$$n \simeq \frac{1}{2} \left(\frac{1.645 + 1.282}{\ln \left(\frac{1.5}{1.33} \right)} \right)^2 = 297.$$

The hypothesis test for c_p can be performed using a sample size calculator for the standard deviation. By setting the standard deviations to the reciprocals of c_p in **PASS> Variance> Variance: 1 Group**:

PASS: Variance: 1

File Run Analysis Graphics PASS GESS Tools Window Help

Symbols 2 | Background | Abbreviations | Template
 Plot Text | Axes | 3D | Symbols 1
 Data | Options | Reports | Plot Setup

Find (Solve For): Scale:

V0 (Baseline Variance): Alternative Hypothesis:

V1 (Alternative Variance): Alpha (Significance Level):

N (Sample Size): Beta (1-Power):

Known Mean

PASS: Variance: 1 Output

One Variance Power Analysis

Numeric Results when H0: S0 = S1 versus Ha: S0>S1

Power	N	S0	S1	Alpha	Beta
0.900650	309	0.7500	0.6660	0.050000	0.099350

References

Davies, Owen L. 1971. The Design and Analysis of Industrial Experiments. Hafner Publishing Company, New York.
 Ostle, B. 1988. Statistics in Research. Fourth Edition. Iowa State Press. Ames, Iowa.
 Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
 N is the size of the sample drawn from the population.
 S0 is the value of the population standard deviation under the null hypothesis.
 S1 is the value of the population standard deviation under the alternative hypothesis.
 Alpha is the probability of rejecting a true null hypothesis. It should be small.
 Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

A sample size of 309 achieves 90% power to detect a difference of 0.0840 between the null hypothesis standard deviation of 0.7500 and the alternative hypothesis standard deviation of 0.6660 using a one-sided, Chi-square hypothesis test with a significance level (alpha) of 0.050000.

Example 10.11 Determine the sample size required to reject $H_0 : c_{pk} = 1.33$ in favor of $H_A : c_{pk} > 1.33$ with 90% power when $c_{pk} = 1.5$.

Solution: With $(c_{pk})_0 = 1.33$, $(c_{pk})_1 = 1.5$, $\alpha = 0.05$, and $\beta = 0.10$ in Equation 10.35, the sample size required to reject H_0 is

$$n = \left(\frac{1.645(1.33) \sqrt{\frac{1}{9 \times 1.33^2} + \frac{1}{2}} + 1.282(1.5) \sqrt{\frac{1}{9 \times 1.5^2} + \frac{1}{2}}}{1.5 - 1.33} \right)^2$$

$$= 326.$$

As expected, this value is comparable to the $n = 297$ sample size required for the test of c_p determined in Example 10.10 for similar conditions.

Example 10.12 Determine the sample size for Example 10.11 using the large sample approximation and compare the result to the original sample size.

Solution: From the information given in the original problem statement and Equation 10.36 the approximate sample size is

$$n \approx \frac{1}{2} \left(\frac{1.645(1.33) + 1.282(1.5)}{1.5 - 1.33} \right)^2$$

$$\approx 292.$$

This value is about 10% lower than the more accurate value calculated in Example 10.11.

10.3 Tolerance Intervals

Example 10.13 What sample size is required to be 95% confident that at least 99% of a population of continuous measurement values falls within the extreme values of the sample?

Solution: With $\alpha = 0.05$ and $p_U = 0.01$ the required sample size is approximately

$$\begin{aligned} n &\approx \frac{\chi_{0.95,4}^2}{2 \times 0.01} \\ &\approx 475. \end{aligned}$$

Further iterations indicate that the smallest value of n for which $\alpha \leq 0.05$ is $n = 473$, which leads to the following nonparametric tolerance interval for x :

$$P(0.99 < P(x_{min} \leq x \leq x_{max}) < 1) = 0.9502.$$

Example 10.14 What sample size n is required to be 95% confident that at least 99% of a population of continuous measurement values falls below the maximum value of the sample?

Solution: With $\alpha = 0.05$ and $p_U = 0.01$ the required sample size is

$$\begin{aligned} n &\approx \frac{\chi_{0.95,2}^2}{2 \times 0.01} \\ &\approx 300. \end{aligned}$$

Example 10.15 Determine the sample size required to obtain a 95% confidence two-sided 99% coverage normal distribution tolerance interval with tolerance limits $UTL/LTL = \bar{x} \pm 3.5s$.

Solution: The desired tolerance interval has the form

$$P(0.99 \leq \Phi(\bar{x} - 3.5s \leq x \leq \bar{x} + 3.5s) \leq 1) = 0.95.$$

From Appendix E.7, as sample of size $n = 25$ gives $k_2 = 3.46$. A spreadsheet (not shown) was set up to calculate k_2 as a function of n using Equation 10.47 with $p = 0.01$ and $\alpha = 0.05$. The spreadsheet indicated that the sample size $n = 24$ delivers $k_2 = 3.485$ and that $n = 23$ delivers $k_2 = 3.514$, so $n = 24$ should be used to be conservative. These approximate k_2 values differ from the exact values given in Appendix E.7 in the thousandths place.

Example 10.16 Determine the sample size required to obtain a 95% confidence 99% coverage normal distribution tolerance interval with one-sided upper tolerance limit $UTL = \bar{x} + 3s$.

Solution: The required interval has the form

$$P(0.99 < \Phi(-\infty < x \leq UTL) < 1) = 0.95 \tag{10.2}$$

where $UTL = \bar{x} + k_1s$ with $k_1 = 3$. From Table E.7 of Appendix E with $\alpha = 0.05$ and $Y = 0.99$, the required sample size is $n = 35$.

10.4 Acceptance Sampling

Example 10.17 Design the single sampling plan for attributes that will accept 95% of lots when the process fraction defective is 1% and accept only 10% of lots when the process fraction defective is 5%.

Solution: From the problem statement, the lots are coming from a continuous process, so the sampling plan will be Type B with points on the OC curve at $(AQL, 1 - \alpha) =$

$(0.01, 0.95)$ and $(RQL, \beta) = (0.05, 0.10)$. From Table 10.1 with $RQL/AQL = 0.05/0.01 = 5.0$, the acceptance number must be $c = 3$. Then, from the RQL condition, the required sample size is approximately

$$\begin{aligned} n &\approx \frac{\chi_{0.90,8}^2}{2(0.05)} \\ &\approx \frac{13.36}{2 \times 0.05} \\ &\approx 134. \end{aligned}$$

The exact sampling plan that meets the specifications in the problem statement is $n = 132$ and $c = 3$.

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Attributes:

```

MTB > AASAMPLING 1;
SUBC> AQL 0.01;
SUBC> RQL 0.05;
SUBC> CREATE;
SUBC> ALPHA 0.05;
SUBC> BETA 0.10;
SUBC> PROPORTION;
SUBC> GOC.

Acceptance Sampling by Attributes

Measurement type: Go/no go
Lot quality in proportion defective
Use binomial distribution to calculate probability of acceptance

Acceptable Quality Level (AQL)          0.01
Producer's Risk (Alpha)                 0.05

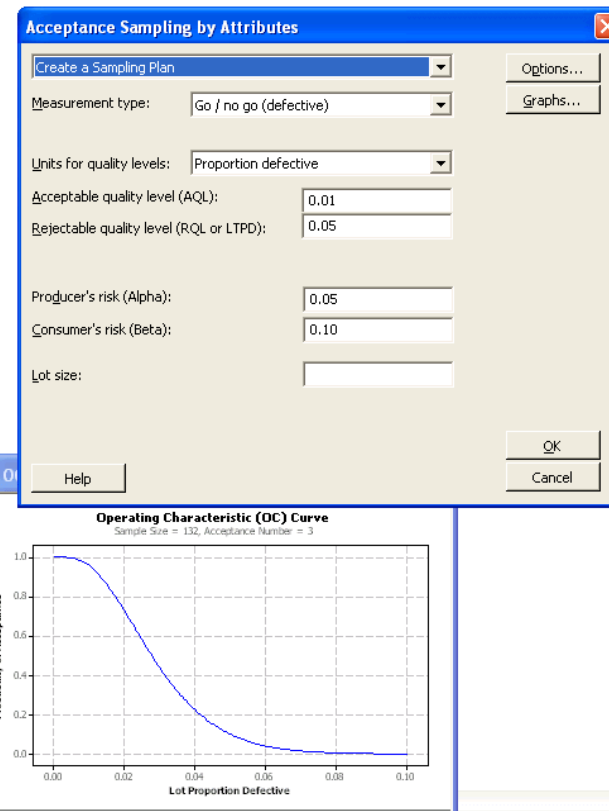
Rejectable Quality Level (RQL or LTPD)  0.05
Consumer's Risk (Beta)                  0.1

Generated Plan(s)

Sample Size      132
Acceptance Number  3

Accept lot if defective items in 132 sampled <= 3; Otherwise reject.

Proportion Defective  Probability Accepting  Probability Rejecting
0.01                   0.956           0.044
0.05                   0.099           0.901
    
```



Example 10.18 Find the $c = 0$ plans that meet a) the AQL requirement and b) the RQL requirement from Example 10.17. Plot the three OC curves on the same graph.

Solution:

a) The sample size for the $c = 0$ plan that meets the AQL requirement $(p, P_A) = (0.01, 0.95)$ is approximately

$$\begin{aligned} n &\approx \frac{\chi_{\alpha,2}^2}{2 \times AQL} \\ &\approx \frac{0.1026}{2 \times 0.01} \\ &\approx 6. \end{aligned}$$

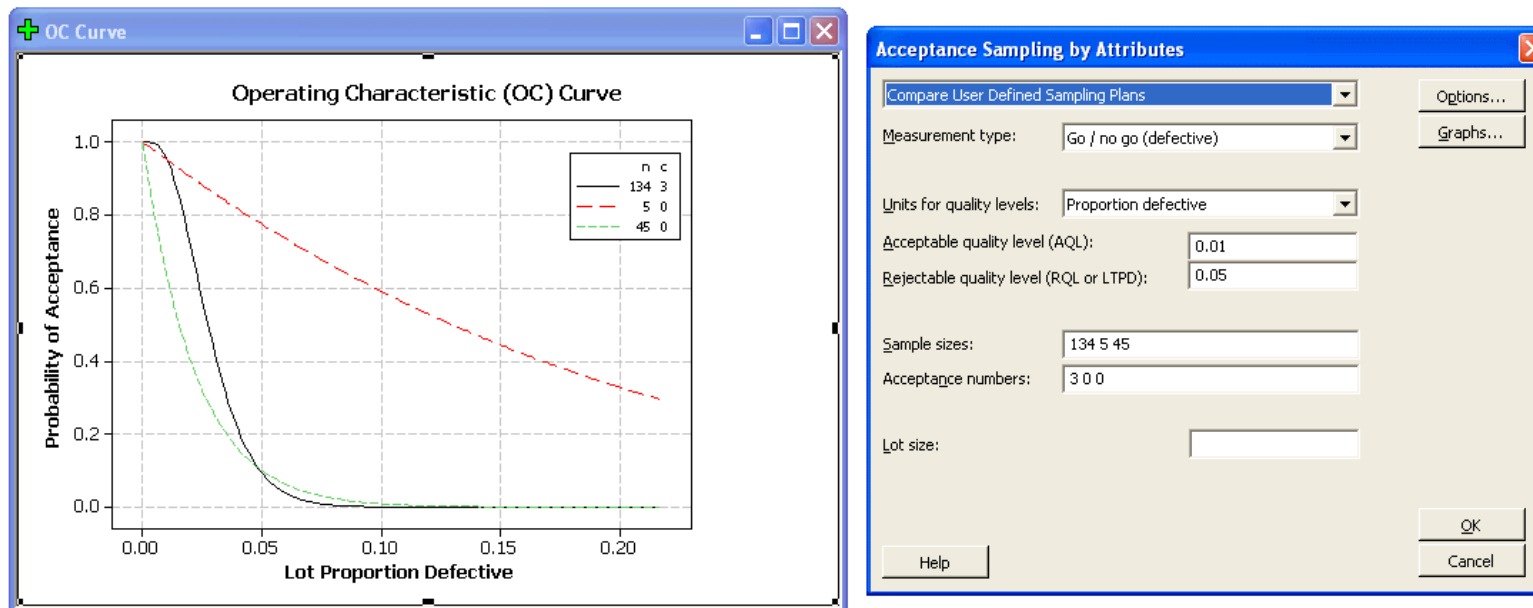
A few binomial calculations indicate that the exact sample size is $n = 5$ because $(b(0; 5, 0.01) = 0.951) > (1 - \alpha = 0.95)$.

b) The sample size for the $c = 0$ plan that meets the RQL requirement $(p, P_A) = (0.05, 0.10)$ is approximately

$$\begin{aligned} n &\approx \frac{\chi_{1-\beta,2}^2}{2 \times RQL} \\ &\approx \frac{4.61}{2 \times 0.05} \\ &\approx 46. \end{aligned}$$

The exact sample size is $n = 45$ because $(b(0; 45, 0.05) = 0.099) < (\beta = 0.10)$.

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Attributes:



Example 10.19 Determine the sampling plan for lots of size $N = 50$ that will accept 95% of the lots with $D \leq 1$ defectives and reject 90% of the lots with $D \geq 5$ defectives.

Solution: The sampling plan must meet the simultaneous conditions given by Equation 10.57 with $D_1 = 1$ and $\alpha \leq 0.05$:

$$\sum_{x=0}^c h(x; D_1 = 1, N = 50, n) \geq 0.95 \quad (10.3)$$

and Equation 10.58 with $D_2 = 5$ and $\beta \leq 0.10$:

$$\sum_{x=0}^c h(x; D_2 = 5, N = 50, n) < 0.10. \quad (10.4)$$

The acceptance number c is not specified, so different values of c must be considered. The approximate sample size for the $c = 0$ sampling plan to meet the condition in Equation 10.63 is given by Equation 10.61:

$$n \simeq 50 \left(1 - 0.10^{1/5} \right) = 19;$$

however, the condition in Equation 10.62 is not satisfied because

$$(h(x = 0; D_1 = 1, N = 50, n = 19) = 0.525) \not\geq 0.95.$$

Iterations with a hypergeometric calculator show that with $c = 1$ Equation 10.63 is satisfied when $n = 29$ because

$$\begin{aligned} \left(\sum_{x=0}^1 h(x; D_1 = 5, N = 50, n = 28) = 0.109 \right) &\not\leq 0.10 \\ \left(\sum_{x=0}^1 h(x; D_1 = 5, N = 50, n = 29) = 0.092 \right) &\leq 0.10 \end{aligned}$$

and Equation 10.62 is satisfied because

$$\left(\sum_{x=0}^1 h(x; D_1 = 1, N = 50, n = 29) = 1 \right) \geq 0.95.$$

The sampling plan that meets the requirements is $n = 29$ with $c = 1$.

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Attributes:

```

MTB > AASAMPLING 1;
SUBC> AQL 0.02;
SUBC> RQL 0.10;
SUBC> CREATE;
SUBC> ALPHA 0.05;
SUBC> BETA 0.10;
SUBC> LOTSIZE 50;
SUBC> PROPORTION;
SUBC> HYPERGEOMETRIC;
SUBC> GOC;
SUBC> GAOQ;
SUBC> GATI;
SUBC> ONEGRAPH.

```

Acceptance Sampling by Attributes

```

Measurement type: Go/no go
Lot quality in proportion defective
Lot size: 50
Use hypergeometric distribution to calculate probability of acceptance

```

Acceptable Quality Level (AQL)	0.02
Producer's Risk (Alpha)	0.05
Rejectable Quality Level (RQL or LTPD)	0.1
Consumer's Risk (Beta)	0.1

Generated Plan(s)

```

Sample Size      29
Acceptance Number 1

```

```

Accept lot if defective items in 29 sampled <= 1; Otherwise reject.

```

Example 10.20 A 100% inspection process for large lots is to be replaced with a $c = 0$ sampling plan. What fraction of each lot must be inspected if lots that contain five or more defectives must be rejected 90% of the time?

Solution: From Equation 10.61 with $D = 5$ and $P_A = 0.10$, the fraction of each lot that must be inspected is

$$\begin{aligned}\frac{n}{N} &\approx 1 - 0.10^{1/5} \\ &\approx 0.37.\end{aligned}$$

Example 10.21 The calculated sample size in Example 4.2 was quite large compared to the lot size, which violates the small-sample approximation assumption. Repeat that example using the small-lot-size method.

Solution: The solution in the example indicated that 30% of the lot needed to be inspected. From Equation 10.61, which takes the relatively large sample size into account, the fraction of the lot that has to be inspected is more accurately

$$\begin{aligned}\frac{n}{N} &\approx 1 - 0.05^{1/10} \\ &\approx 0.259.\end{aligned}$$

Example 10.22 Create the OC curves for normal, tightened, and reduced inspection under ANSI/ASQ Z1.4 using general inspection level II, single sampling, $N = 1000$, and $AQL = 1\%$.

Solution: The sampling plans determined for code letter J from the standard were normal ($n = 200, c = 5$), tightened ($n = 200, c = 3$), and reduced ($n = 80, c = 2$). The operating characteristic curves were calculated using Equation 10.54 and are shown in Figure 10.6. For reference, the figure also shows the OC curve for the sampling plan determined for the same conditions using the Squeglia zero acceptance number sampling standard, which is often used instead of ANSI/ASQ Z1.4.

Example 10.23 Determine the optimal rectifying inspection sampling plan for $LTPD = 0.04$ with $\beta = 0.10$ when the lot size is $N = 2500$ and the historical process fraction defective is $p = 0.01$.

Solution: A spreadsheet was used to solve Equations 10.66 and 10.67 as a function of acceptance number c as shown in Table 10.2. For the specified conditions, the sampling plan that minimizes ATI when $p = 0.01$ is $n = 232$ with $c = 5$. By comparison, the Dodge-Romig LTPD tables indicate a sampling plan with $n = 230$ and $c = 5$, which is in excellent agreement with the calculated plan.

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Attributes:

Acceptance Sampling by Attributes

Measurement type: Go/no go
 Lot quality in proportion defective
 Lot size: 2500
 Use hypergeometric distribution to calculate probability of acceptance

Acceptable Quality Level (AQL) 0.01
 Producer's Risk (Alpha) 0.05

Rejectable Quality Level (RQL or LTPD) 0.04
 Consumer's Risk (Beta) 0.1

Generated Plan(s)

Sample Size 194
 Acceptance Number 4

Accept lot if defective items in 194 sampled ≤ 4 ; Otherwise reject.

Proportion Defective	Probability Accepting	Probability Rejecting	AOQ	ATI
0.01	0.960	0.040	0.00886	285.2
0.04	0.100	0.900	0.00369	2269.5

Average outgoing quality limit (AOQL) = 0.01216 at 0.01840 proportion defective.

Acceptance Sampling by Attributes dialog box settings:

- Create a Sampling Plan
- Measurement type: Go / no go (defective)
- Units for quality levels: Proportion defective
- Acceptable quality level (AQL): 0.01
- Rejectable quality level (RQL or LTPD): 0.04
- Producer's risk (Alpha): 0.05
- Consumer's risk (Beta): 0.10
- Lot size: 2500

Acceptance Sampling by Attributes - Options dialog box settings:

- Use hypergeometric distribution for isolated lot
- Enter additional quality levels to calculate acceptance probabilities: (empty field)
- (Units: Proportion defective)
- You may increase Alpha and Beta slightly for alternative plans with smaller sample sizes
- Maximum Alpha allowed: (empty field)
- Maximum Beta allowed: (empty field)

Example 10.24 Find the rectifying inspection plan with $LTPD = 0.04$ and $\beta = 0.1$ for a lot size of $N = 50$.

Solution: For the given conditions, the spreadsheet method gives $n = 58$, which exceeds the lot size. From Equation 10.70 the sample size required for the $c = 0$ plan is

$$\begin{aligned} n &\simeq 50 \left(1 - 0.1^{\frac{1}{50 \times 0.04}} \right) \\ &\simeq 35. \end{aligned}$$

By comparison, the corresponding Dodge-Romig plan calls for $n = 34$ and is independent of the historical fraction defective.

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Attributes:

```

MTB > AASAMPLING 1;
SUBC> AQL 0.000001;
SUBC> RQL 0.04;
SUBC> CREATE;
SUBC> ALPHA 0.05;
SUBC> BETA 0.10;
SUBC> LOTSIZE 50;
SUBC> PROPORTION;
SUBC> HYPERGEOMETRIC;
SUBC> GOC;
SUBC> GAOQ;
SUBC> GATI;
SUBC> ONEGRAPH.

```

Acceptance Sampling by Attributes

Measurement type: Go/no go
 Lot quality in proportion defective
 Lot size: 50
 Use hypergeometric distribution to calculate probability of acceptance

Acceptable Quality Level (AQL)	0.000001
Producer's Risk (Alpha)	0.05
Rejectable Quality Level (RQL or LTPD)	0.04
Consumer's Risk (Beta)	0.1

Generated Plan(s)

Sample Size 34
 Acceptance Number 0

Accept lot if defective items in 34 sampled <= 0; Otherwise reject.

The image shows two overlapping dialog boxes from Minitab. The top dialog, titled "Acceptance Sampling by Attributes", has a dropdown menu set to "Create a Sampling Plan" and a "Measurement type" of "Go / no go (defective)". The "Units for quality levels" is set to "Proportion defective". The "Acceptable quality level (AQL)" is 0.000001, and the "Rejectable quality level (RQL or LTPD)" is 0.04. The "Producer's risk (Alpha)" is 0.05, and the "Consumer's risk (Beta)" is 0.10. The "Lot size" is 50. The bottom dialog, titled "Acceptance Sampling by Attributes - Options", has a checked checkbox for "Use hypergeometric distribution for isolated lot!". It includes a text box for "Enter additional quality levels to calculate acceptance probabilities:" and a note: "You may increase Alpha and Beta slightly for alternative plans with smaller sample sizes". There are input fields for "Maximum Alpha allowed:" and "Maximum Beta allowed:". Buttons for "Help", "OK", and "Cancel" are at the bottom.

Example 10.25 What sample size is required for a $c = 1$ rectifying inspection single-sampling plan to obtain 1% $AOQL$ if the lot size is $N = 300$? For what value of incoming fraction defective will AOQ be a maximum?

Solution: From Equation 10.76 with $A_1 = 0.839$ the required sample size is

$$n = \frac{1}{\frac{0.01}{0.839} + \frac{1}{300}} = 66.$$

AOQ will be at its maximum value, $AOQL$, when the incoming fraction defective is

$$p_c = \frac{\chi_1^2}{2n} = \frac{3.24}{2 \times 66} = 0.0245.$$

Example 10.26 Determine the sampling plan that minimizes the ATI for lots of size $N = 1000$ with $AOQL = 0.02$ when the historical defective rate is $p = 0.01$.

Solution: A spreadsheet was used to solve for n and ATI as a function of c using Equations 10.76 and 10.67. The results from the spreadsheet, shown in Table 10.4, indicate that the sampling plan that minimizes ATI is given by $n = 65$ and $c = 2$. By comparison, this is exactly the same plan indicated in the Dodge-Romig tables for these conditions.

Example 10.27 Find the single sampling plan for variables that will accept 95% of the lots with 1% defectives and reject 90% of the lots with 4% defectives when $\sigma = 30$ and the specification is one-sided with $USL = 700$.

Solution: The two specified points on the OC curve are $(p_0, 1 - \alpha) = (0.01, 0.95)$ and $(p_1, \beta) = (0.04, 0.10)$. From Equation 10.79 the required sample size is

$$\begin{aligned} n &= \left(\frac{z_{0.05} + z_{0.10}}{z_{0.01} - z_{0.04}} \right)^2 \\ &= \left(\frac{1.645 + 1.282}{2.33 - 1.75} \right)^2 \\ &= 26. \end{aligned}$$

The critical value of $\bar{x}_{A/R}$ is

$$\begin{aligned} \bar{x}_{A/R} &= \mu_0 + z_\alpha \sigma_{\bar{x}} \\ &= (USL - z_{p_0} \sigma_x) + z_\alpha \frac{\sigma_x}{\sqrt{n}} \\ &= (700 - 2.33 \times 30) + 1.645 \frac{30}{\sqrt{26}} \\ &= 640. \end{aligned}$$

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Variables> Create/Compare:


```
MTB > VASAMPLING;
SUBC> AQL 0.01;
SUBC> RQL 0.04;
SUBC> CREATE;
SUBC> ALPHA 0.05;
SUBC> BETA 0.10;
SUBC> PROPORTION;
SUBC> LSPEC 700;
SUBC> SIGMA 30;
SUBC> GOC.
```

Acceptance Sampling by Variables - Create/Compare

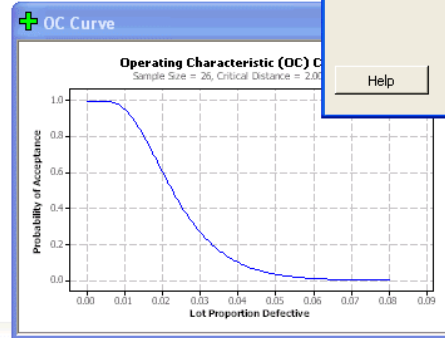
Lot quality in proportion defective

Lower Specification Limit (LSL)	700
Historical Standard Deviation	30
Acceptable Quality Level (AQL)	0.01
Producer's Risk (Alpha)	0.05
Rejectable Quality Level (RQL or LTPD)	0.04
Consumer's Risk (Beta)	0.1

Generated Plan(s)

Sample Size	26
Critical Distance (k Value)	2.00289

Z.LSL = (mean - lower spec)/historical standard deviation
 Accept lot if Z.LSL >= k; otherwise reject.



Acceptance Sampling by Variables (Create/Compare)

Create a Sampling Plan

Units for quality levels: Proportion defective

Acceptable quality level (AQL): 0.01

Rejectable quality level (RQL or LTPD): 0.04

Producer's risk (Alpha): 0.05

Consumer's risk (Beta): 0.10

Lower spec: 700

Upper spec:

Historical standard deviation: 30 (Optional)

Lot size:

Options...
Graphs...
OK
Cancel
Help

Example 10.28 Determine the sample size ratio for attributes and variables inspection plans that will accept 95% of the lots with 0.1% defectives and reject 95% of the lots with 0.4% defectives.

Solution: The two points on the OC curve are $(p_0 = 0.001, 1 - \alpha = 0.95)$ and $(p_1 = 0.004, \beta = 0.05)$. Because $\alpha = \beta = 0.05$ and both p_0 and p_1 are relatively small, from Equation 10.81 the ratio of the attributes- to variables-based sample sizes is approximately

$$\begin{aligned} \frac{n_{attributes}}{n_{variables}} &\approx \frac{1}{4} \left(\frac{z_{0.001} - z_{0.004}}{\sqrt{0.004} - \sqrt{0.001}} \right)^2 \\ &\approx \frac{1}{4} \left(\frac{3.090 - 2.652}{\sqrt{0.004} - \sqrt{0.001}} \right)^2 \\ &\approx 48. \end{aligned}$$

So, the attributes plan sample size will have to be about 48 times larger than the variables plan sample size to obtain the same performance for acceptable and rejectable quality levels!

Example 10.29 Find the single sampling plan for variables that will accept 95% of the lots with 1% defectives and reject 90% of the lots with 4% defectives. The specification is one-sided and σ is unknown.

Solution: The two specified points on the OC curve are $(p_0, 1 - \alpha) = (0.01, 0.95)$ and $(p_1, \beta) = (0.04, 0.10)$. From Equation 10.86, the condition that determines the sample size is

$$t_{0.95, n-1, -z_{0.01}\sqrt{n}} = t_{0.10, n-1, -z_{0.04}\sqrt{n}}$$

which is satisfied by $n = 78$ because

$$t_{0.95, 77, -20.58} = t_{0.10, 77, -15.46} = -17.75.$$

The accept/reject value of k for the test is

$$k = \frac{17.75}{\sqrt{78}} = 2.01.$$

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Variables> Create/Compare:

The screenshot displays the Minitab interface for creating an acceptance sampling plan. The main window shows the command sequence and the resulting plan parameters:

```

MTB > VASAMPLING;
SUBC> AQL 0.01;
SUBC> RQL 0.04;
SUBC> CREATE;
SUBC> ALPHA 0.05;
SUBC> BETA 0.10;
SUBC> PROPORTION;
SUBC> LSPEC 1000;
SUBC> GOC.
    
```

Acceptance Sampling by Variables - Create/Compare

Lot quality in proportion defective

Lower Specification Limit (LSL)	1000
Acceptable Quality Level (AQL)	0.01
Producer's Risk (Alpha)	0.05
Rejectable Quality Level (RQL or LTPD)	0.04
Consumer's Risk (Beta)	0.1

Generated Plan(s)

Sample Size	78
Critical Distance (k Value)	2.00278

Z.LSL = (mean - lower spec)/standard deviation
 Accept lot if Z.LSL >= k; otherwise reject.

Proportion Defective	Probability Accepting	Probability Rejecting
0.01	0.952	0.048

The **Operating Characteristic (OC) Curve** plot shows the Probability of Acceptance on the y-axis (ranging from 0.0 to 1.0) and Lot Proportion Defective on the x-axis (ranging from 0.00 to 0.09). The curve starts at a probability of 1.0 for 0% defectives and drops to approximately 0.1 at 4% defectives. The plot title indicates: "Operating Characteristic (OC) Curve, Sample Size = 78, Critical Distance = 2.00278".

Example 10.30 Plot the OC curves for the normal, tightened, and reduced sampling plans under ANSI/ASQ Z1.9 using a one-sided specification with Form 1, code letter F, and $AQL = 1\%$.

Solution: The sampling plans determined from the standard were normal ($n = 10, k = 1.72$), tightened ($n = 10, k = 1.84$), and reduced ($n = 4, k = 1.34$). The operating characteristic curves were calculated using Equations 10.84 and 10.85. For example, the OC curve for normal inspection is given by

$$t_{P_A,df,-z_p\sqrt{n}} = -k\sqrt{n}$$

$$t_{P_A,9,-z_p\sqrt{10}} = -1.72\sqrt{10}$$

$$t_{P_A,9,-z_p\sqrt{10}} = -5.439.$$

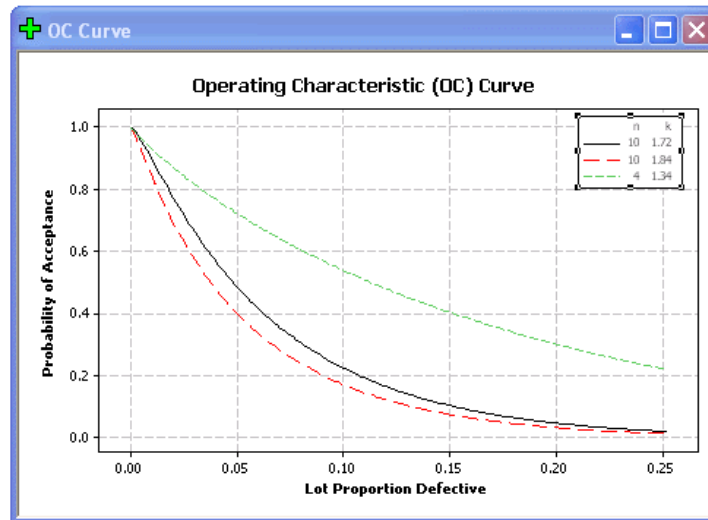
The OC curves are shown in Figure 10.9 and are in excellent agreement with the OC curves in the standard.

From MINITAB> Stat> Quality Tools> Acceptance Sampling by Variables> Create/Compare:

```
MTB > VASAMPLING;
SUBC> COMPARE;
SUBC> SIZE 10 10 4;
SUBC> CDISTANCE 1.72 1.84 1.34;
SUBC> PROPORTION;
SUBC> LSPEC 1000;
SUBC> GOC.
```

Acceptance Sampling by Variables - Create/Compare

Lot quality in proportion defective



The dialog box 'Acceptance Sampling by Variables (Create/Compare)' is shown. It has a dropdown menu set to 'Compare User Defined Sampling Plans'. There are buttons for 'Options...' and 'Graphs...'. The 'Units for quality levels' is set to 'Proportion defective'. There are input fields for 'Acceptable quality level (AQL):', 'Rejectable quality level (RQL or LTPD):', 'Lower spec:' (set to 1000), 'Upper spec:', and 'Historical standard deviation:' (Optional). The 'Sample sizes' field contains '10 10 4' and the 'Critical distances (k values)' field contains '1.72 1.84 1.34'. There is also a 'Lot size:' field. At the bottom, there are 'Help', 'OK', and 'Cancel' buttons.

10.5 Gage R&R Studies

Chapter 11

Resampling Methods

11.1 Software Requirements

11.2 Monte Carlo

Example 11.1 How many samples should be drawn from a Poisson population to estimate the mean with 20% precision and 95% confidence? The mean is expected to be $\lambda = 5$.

Solution: The sample size must be sufficient to deliver the following confidence interval for λ :

$$\begin{aligned}P\left(0.8\hat{\lambda} < \lambda < 1.2\hat{\lambda}\right) &= 0.95 \\P(4 < \lambda < 6) &= 0.95\end{aligned}$$

where $\hat{\lambda} = x/n$ and x is the number of counts observed in n sampling units. A MINITAB macro was used to draw 10000 random samples, all of size n , from a Poisson distribution with $\lambda = 5$, and calculate $\hat{\lambda}$ for each sample. The 2.5th and 97.5th $\hat{\lambda}$ percentiles were used to estimate the 95% confidence limits. Figure 11.1 shows the Monte Carlo confidence limits as a function of sample size. The sample size that delivers the desired confidence interval width is $n = 19$, which is in good agreement with the sample size given by Equation 5.11:

$$\begin{aligned}n &= \frac{1}{\lambda} \left(\frac{z_{\alpha/2}}{\delta}\right)^2 \\&= \frac{1}{5} \left(\frac{1.96}{0.2}\right)^2 \\&= 20.\end{aligned}$$

Example 11.2 Use the Monte Carlo method to confirm the answer to Example 9.25, that samples of size $n = 31$ are required to estimate, with 95% confidence, the exponential-exponential interference failure rate to within $\pm 50\%$.

Solution: A MINITAB macro was written that draws random samples of size $n = 31$ for exponential load and strength distributions, where the load distribution has mean μ_L ,

where μ_L comes from a uniform distribution with $1 \leq \mu_L \leq 10$, and the strength distribution has mean $\mu_S = k\mu_L$, where k comes from a uniform distribution with $3 \leq k \leq 10$. The macro loops 1000 times, collecting the ratio of the empirical interference failure rate \hat{f} to the known failure rate f from each pass. Figure 11.2 shows the histogram of \hat{f}/f . About 97% of the observed \hat{f} values fall within $\pm 50\%$ of their known f values, which is consistent with the intended 95% confidence level.

Example 11.3 An experiment is proposed to study the defective rate (p) of a process using a sample of size $n = 300$ units. The process is considered to be acceptable when $p \leq 0.01$, but $p \geq 0.03$ is unacceptable. Determine the acceptance number and power for the sampling plan.

Solution: The hypotheses to be tested are $H_0 : p \leq 0.01$ versus $H_A : p > 0.01$. The decision to reject H_0 or not is based on the observed number of defectives x in a sample of size $n = 300$. The critical value of x that distinguishes the accept and reject regions should be chosen such that $\alpha \leq 0.05$ for $p = 0.01$. To determine this critical value, MINITAB was used to select 10000 random samples from a binomial population with $n = 300$ and $p = 0.01$. The frequencies and cumulative percentages by x are shown in the H_0 columns of Table 11.1. From the *CumPct* column, the critical value of x must be set to $x_{A/R} = 6.5$ to provide good protection against type I errors ($\alpha = 1 - 0.9692 = 0.0308$). The frequency statistics under H_A in the Figure were created by selecting 10000 random samples from a binomial population with $n = 300$ and $p = 0.03$. From the *CumPct* column, the type II error probability relative to $x_{A/R} = 6.5$ is $\beta = 0.206$, so the power to reject H_0 when $p = 0.03$ is $\pi = 1 - \beta = 0.794$. This is in excellent agreement with the power calculated for the one-sample proportion test (see Section 4.1.2), which is $\pi = 0.797$.

Example 11.4 Tukey's quick test is a nonparametric two-sample test for location that is easy to perform using dotplots. The samples must be comparable in size and they must be slipped, that is, one sample must have the largest observation and the other sample must have the smallest observation. The test statistic, T , is the number of slipped or nonoverlapping points. The Tukey test rejects $H_0 : \mu_1 = \mu_2$ when $T \geq 7$. For example, in Figure 11.3 $T = 7 + 10 = 17$, so there is sufficient evidence to reject H_0 .

Use the Monte Carlo method to determine the power of Tukey's quick test as a function of effect size and sample size assuming that the two populations are normal and homoscedastic.

Solution: A MINITAB macro was written to draw 1000 random samples of equal sample size from two independent homoscedastic normal populations and count the number of times that H_0 was rejected by the Tukey quick test. Figure 11.4 shows that the power of the test is low for all sample sizes until the difference between the means is greater than 1.5 to 2.5 standard deviations. When the sample size is $n \geq 20$, Tukey's quick test has power $\pi \geq 0.90$ for differences between the means of 1.5 standard deviations or greater.

11.3 Bootstrap

Example 11.5 The following yield strength values (in thousands of *psi*) for a material were obtained in a pilot study: {56, 23, 25, 68, 35, 31, 13, 15, 48, 37, 57, 69, 50, 76, 50, 19, 88, 33, 10, 21}. What sample size is required to estimate, with 95% confidence, the mean yield strength to within ± 5000 psi?

Solution: The bootstrap percentile confidence interval width was studied as a function of sample size using 1000 resamples for sample sizes from $n = 30$ to 100. Figure 11.5 shows the confidence interval width, given by

$$2\delta = \hat{\theta}_{0.975}^* - \hat{\theta}_{0.025}^*$$

versus sample size. The sample size required to obtain the desired precision of the estimate is $n = 75$.

Example 11.6 The following data were obtained from a pilot study: {56, 48, 44, 62, 50, 47, 49, 57, 48, 55, 96, 47, 46, 47, 49, 72, 46, 61}. Determine the sample size required to obtain 90% power to reject $H_0 : \mu = 50$ when $\mu = 52$. The population is not normal, so assume that the experimental data will be analyzed using the bootstrap method.

Solution: If the population distribution is normal, the one-sample Student's t test would be an appropriate method of analysis. Because the normality assumption is not satisfied, the preferred method of analysis using the bootstrap method uses the analogous bootstrap- t distribution given by

$$t^* = \frac{\bar{x}^* - \mu_0}{s^*/\sqrt{n}} \quad (11.4)$$

where \bar{x}^* and s^* are the mean and standard deviation of bootstrap samples. Figure 11.6 shows the bootstrap distributions of t^* with samples of size $n = 18$ under $H_0 : \mu = 50$ and $H_A : \mu = 56$, where the sample data were shifted to the appropriate population means before bootstrapping using transformations of the form

$$x'_i = x_i - \bar{x} + \mu.$$

From the bootstrap- t distribution under H_0 , the acceptance interval for H_0 is $-4.29 \leq t^* \leq 1.62$. The acceptance interval is skewed because the original sample is skewed. From the bootstrap distribution under H_A , the power is $\pi = 0.802$, which does not meet the 90% power requirement.

Figure 11.7 shows the bootstrap- t test power versus sample size for samples from size $n = 16$ to 30. (Figure 11.6 was constructed from 1000 bootstrap samples. Each point in Figure 11.7 was constructed from 10000 bootstrap samples to reduce the noise in the power versus sample size plot.) The sample size required to obtain 90% power to reject H_0 when $\mu = 56$ is $n = 23$.