Chapter 194

Fisher's Exact Test for Two Proportions

Introduction

This module computes power and sample size for hypothesis tests of the difference, ratio, or odds ratio of two independent proportions using Fisher's exact test. This procedure assumes that the difference between the two proportions is zero or their ratio is one under the null hypothesis.

The power calculations assume that random samples are drawn from two separate populations.

Technical Details

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability, p_i, is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of *m* and *n* individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	*a*	*c*	*m*
Control	*b*	*d*	*n*
Total	*s*	*f*	*N*

The following alternative notation is also used.

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Total	m_1	m_2	N

The binomial proportions p_1 and p_2 are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Comparing Two Proportions

When analyzing studies such as this, one usually wants to compare the two binomial probabilities, p_1 and p_2. Common measures for comparing these quantities are the difference and the ratio. If the binomial probabilities are expressed in terms of odds rather than probabilities, another common measure is the odds ratio. Mathematically, these comparison parameters are

Parameter	**Computation**
Difference	$\delta = p_1 - p_2$
Risk Ratio	$\phi = p_1 \,/\, p_2$
Odds Ratio	$\psi = \frac{p_1 \,/\, (1-p_1)}{p_2 \,/\, (1-p_2)} = \frac{p_1 q_2}{p_2 q_1}$

Tests analyzed by this routine are for the *null case*. This refers to the values of the above parameters under the null hypothesis. In the *null case*, the difference is zero and the ratios are one under the null hypothesis.

Hypothesis Tests

Several statistical tests have been developed for testing the inequality of two proportions. For large samples, the powers of the various tests are about the same. However, for small samples, the differences in the powers can be quite large. Hence, it is important to base the power analysis on the test statistic that will be used to analyze the data. If you have not selected a test statistic, you may wish to determine which one offers the best power in your situation. No single test is the champion in every situation, so you must compare the powers of the various tests to determine which to use.

Difference

The (risk) difference, $\delta = p_1 - p_2$, is perhaps the most direct measure for comparing two proportions. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 - p_2 = 0$ versus $H_1: p_1 - p_2 \neq 0$; this is often called the *two-tailed test*.
2. $H_0: p_1 - p_2 \leq 0$ versus $H_1: p_1 - p_2 > 0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 - p_2 \geq 0$ versus $H_1: p_1 - p_2 < 0$; this is often called the *lower-tailed test*.

The traditional approach for testing these hypotheses has been to use the Pearson chi-square test for large samples, the Yates chi-square for intermediate sample sizes, and the Fisher Exact test for small samples. Recently, some authors have begun questioning this solution. For example, based on exact enumeration, Upton (1982) and D'Agostino (1988) conclude that the Fisher Exact test and Yates test should never be used.

Ratio

The (risk) ratio, $\phi = p_1 \,/\, p_2$, is often preferred to the difference when the baseline proportion is small (less than 0.1) or large (greater than 0.9) because it expresses the difference as a percentage rather than an amount. In this null case, the null hypothesized ratio of proportions, ϕ_0, is one. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 \,/\, p_2 = \phi_0$ versus $H_1: p_1 \,/\, p_2 \neq \phi_0$; this is often called the *two-tailed test*.
2. $H_0: p_1 \,/\, p_2 \leq \phi_0$ versus $H_1: p_1 \,/\, p_2 > \phi_0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 \,/\, p_2 \geq \phi_0$ versus $H_1: p_1 \,/\, p_2 < \phi_0$; this is often called the *lower-tailed test*.

Odds Ratio

The odds ratio, $\psi = \frac{o_1}{o_2} = \frac{p_1 \,/\, (1-p_1)}{p_2 \,/\, (1-p_2)} = \frac{p_1 q_2}{p_2 q_1}$, is sometimes used to compare the two proportions because of its statistical properties and because some experimental designs require its use. In this null case, the null hypothesized odds ratio, ψ_0 ψ_0, is one. Three sets of statistical hypotheses can be formulated:

1. $H_0: \psi = \psi_0$ versus $H_1: \psi \neq \psi_0$; this is often called the *two-tailed test*.
2. $H_0: \psi \leq \psi_0$ versus $H_1: \psi > \psi_0$; this is often called the *upper-tailed test*.
3. $H_0: \psi \geq \psi_0$ versus $H_1: \psi < \psi_0$; this is often called the *lower-tailed test*.

Power Calculation

The power for a test statistic that is based on the normal approximation can be computed exactly using two binomial distributions. The following steps are taken to compute the power of such a test.

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value, $z_{critical}$, is that value of *z* that leaves exactly the target value of alpha in the appropriate tail of the normal distribution. For example, for an upper-tailed test with a target alpha of 0.05, the critical value is 1.645.
2. Compute the value of the test statistic, z_t, for every combination of x_{11} and x_{21}. Note that x_{11} ranges from 0 to n_1, and x_{21} ranges from 0 to n_2. A small value (around 0.0001) can be added to the zero cell counts to avoid numerical problems that occur when the cell value is zero.
3. If $z_t > z_{critical}$, the combination is in the rejection region. Call all combinations of x_{11} and x_{21} that lead to a rejection the set *A*.
4. Compute the power for given values of p_1 and p_2 as

$$1 - \beta = \sum_A \binom{n_1}{x_{11}} p_1^{x_{11}} q_1^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}.$$

5. Compute the actual value of alpha achieved by the design by substituting p_2 for p_1 to obtain

$$\alpha^* = \sum_A \binom{n_1}{x_{11}} p_2^{x_{11}} q_2^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}$$

$$= \sum_A \binom{n_1}{x_{11}} \binom{n_2}{x_{21}} p_2^{x_{11}+x_{21}} q_2^{n_1+n_2-x_{11}-x_{21}}.$$

When the values of n_1 and n_2 are large (say over 200), these formulas may take a little time to evaluate. In this case, a large sample approximation may be used.

Fisher's Exact Test

The most useful reference we found for power analysis of Fisher's Exact test was in the StatXact 5 (2001) documentation. The material presented here is summarized from Section 26.3 (pages 866 – 870) of the StatXact-5 documentation. In this case, the test statistic is

$$T = -\ln \left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}} \right]$$

The null distribution of T is based on the hypergeometric distribution. It is given by

$$\Pr(T \geq t | m, H_0) = \sum_{A(m)} \left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}} \right]$$

where

$$A(m) = \{\text{all pairs } x_1, x_2 \text{ such that } x_1 + x_2 = m, \text{ given } T \geq t\}$$

Conditional on *m*, the critical value, t_α, is the smallest value of *t* such that

$$Pr(T \geq t_\alpha | m, H_0) \leq \alpha$$

The power is defined as

$$1 - \beta = \sum_{m=0}^{N} P(m) Pr(T \geq t_\alpha | m, H_1)$$

where

$$\Pr(T \geq t_\alpha | m, H_1) = \sum_{A(m, T \geq t_\alpha)} \left[\frac{b(x_1, n_1, p_1) b(x_2, n_2, p_2)}{\sum_{A(m)} b(x_1, n_1, p_1) b(x_2, n_2, p_2)} \right]$$

$$\begin{aligned} P(m) &= Pr(x_1 + x_2 = m | H_1) \\ &= b(x_1, n_1, p_1) b(x_2, n_2, p_2) \end{aligned}$$

$$b(x, n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

When the normal approximation is used to compute power, the result is based on the pooled, continuity corrected Z test.

Z Test (or Chi-Square Test) with Continuity Correction (Pooled and Unpooled)

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922. Although this test is often expressed directly as a Chi-Square statistic, it is expressed here as a *z* statistic so that it can be more easily used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The continuity corrected z-test is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_D}$$

where *F* is -1 for lower-tailed, 1 for upper-tailed, and both -1 and 1 for two-sided hypotheses.

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

Example 1 – Finding Power

A study is being designed to study the effectiveness of a new treatment. Historically, the standard treatment has enjoyed a 60% cure rate. Researchers want to compute the power of the two-sided *z*-test at group sample sizes ranging from 50 to 650 for detecting differences of 0.05 and 0.10 in the cure rate at the 0.05 significance level.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Power**
Power Calculation Method	**Binomial Enumeration**
Maximum N1 or N2 for Binomial Enumeration	**1000**
Zero Count Adjustment Method	**Add to zero cells only**
Zero Count Adjustment Value	**0.0001**
Alternative Hypothesis	**Two-Sided**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**50 to 650 by 100**
Input Type	**Differences**
D1 (Difference|H1 = P1-P2)	**0.05 0.1**
P2 (Group 2 Proportion)	**0.6**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Reports

Numeric Results

Solve For: Power
Test Type: Fisher's Exact Test
Groups: 1 = Treatment, 2 = Control
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

	Sample Size			Proportion		Difference	Alpha	
Power*	N1	N2	N	P1	P2	D1	Target	Actual*
0.05398	50	50	100	0.65	0.6	0.05	0.05	0.03207
0.11908	150	150	300	0.65	0.6	0.05	0.05	0.03909
0.18341	250	250	500	0.65	0.6	0.05	0.05	0.04011
0.24952	350	350	700	0.65	0.6	0.05	0.05	0.04112
0.31619	450	450	900	0.65	0.6	0.05	0.05	0.04381
0.37874	550	550	1100	0.65	0.6	0.05	0.05	0.04418
0.43689	650	650	1300	0.65	0.6	0.05	0.05	0.04438
0.13196	50	50	100	0.70	0.6	0.10	0.05	0.03207
0.39398	150	150	300	0.70	0.6	0.10	0.05	0.03909
0.61766	250	250	500	0.70	0.6	0.10	0.05	0.04011
0.77218	350	350	700	0.70	0.6	0.10	0.05	0.04112
0.86945	450	450	900	0.70	0.6	0.10	0.05	0.04381
0.92824	550	550	1100	0.70	0.6	0.10	0.05	0.04418
0.96215	650	650	1300	0.70	0.6	0.10	0.05	0.04438

* Power was computed using binomial enumeration of all possible outcomes. Actual alpha is only computed for two-sided tests.

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N1 and N2 The number of items sampled from each population.
N The total sample size. N = N1 + N2.
P1 The proportion for Group 1 at which power and sample size calculations are made. This is the treatment or experimental group.
P2 The proportion for Group 2. This is the standard, reference, or control group.
D1 The difference assumed for power and sample size calculations. D1 = P1 - P2.
Target Alpha The input probability of rejecting a true null hypothesis.
Actual Alpha The value of alpha that is actually achieved.

Summary Statements

A parallel two-group design will be used to test whether the Group 1 (treatment) proportion (P1) is different from the Group 2 (control) proportion (P2) (H0: P1 - P2 = 0 versus H1: P1 - P2 ≠ 0). The comparison will be made using a two-sided, two-sample Fisher's Exact Test with a Type I error rate (α) of 0.05. The control group proportion is assumed to be 0.6. To detect a proportion difference (P1 - P2) of 0.05 (or P1 of 0.65) with sample sizes of 50 for the treatment group and 50 for the control group, the power is 0.05398.Group sample sizes of 50 in group 1 and 50 in group 2 achieve 0.05398% power to detect a proportion difference (P1 - P2) of 0.05. The proportion in group 1 (the treatment group) is assumed to be 0.6 under the null hypothesis and 0.65 under the alternative hypothesis. The proportion in group 2 (the control group) is 0.6. The test statistic used is the two-sided Fisher's Exact Test. The significance level of the test is targeted at 0.05. The significance level actually achieved by this design is 0.03207.

Power Detail Report

Test Type: Fisher's Exact Test
Groups: 1 = Treatment, 2 = Control
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Sample Size			Proportion		Difference	Normal Approximation		Binomial Enumeration	
N1	N2	N	P1	P2	D1	Power	Alpha	Power	Alpha
50	50	100	0.65	0.6	0.05	0.05284	0.05	0.05398	0.03207
150	150	300	0.65	0.6	0.05	0.11919	0.05	0.11908	0.03909
250	250	500	0.65	0.6	0.05	0.18503	0.05	0.18341	0.04011
350	350	700	0.65	0.6	0.05	0.25090	0.05	0.24952	0.04112
450	450	900	0.65	0.6	0.05	0.31569	0.05	0.31619	0.04381
550	550	1100	0.65	0.6	0.05	0.37839	0.05	0.37874	0.04418
650	650	1300	0.65	0.6	0.05	0.43824	0.05	0.43689	0.04438
50	50	100	0.70	0.6	0.10	0.13036	0.05	0.13196	0.03207
150	150	300	0.70	0.6	0.10	0.39486	0.05	0.39398	0.03909
250	250	500	0.70	0.6	0.10	0.61483	0.05	0.61766	0.04011
350	350	700	0.70	0.6	0.10	0.76985	0.05	0.77218	0.04112
450	450	900	0.70	0.6	0.10	0.86889	0.05	0.86945	0.04381
550	550	1100	0.70	0.6	0.10	0.92808	0.05	0.92824	0.04418
650	650	1300	0.70	0.6	0.10	0.96174	0.05	0.96215	0.04438

Dropout-Inflated Sample Size

	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
Dropout Rate	N1	N2	N	N1'	N2'	N'	D1	D2	D
20%	50	50	100	63	63	126	13	13	26
20%	150	150	300	188	188	376	38	38	76
20%	250	250	500	313	313	626	63	63	126
20%	350	350	700	438	438	876	88	88	176
20%	450	450	900	563	563	1126	113	113	226
20%	550	550	1100	688	688	1376	138	138	276
20%	650	650	1300	813	813	1626	163	163	326

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N1, N2, and N The evaluable sample sizes at which power is computed (as entered by the user). If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
N1', N2', and N' The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D1, D2, and D The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2.

Dropout Summary Statements

Anticipating a 20% dropout rate, 63 subjects should be enrolled in Group 1, and 63 in Group 2, to obtain final group sample sizes of 50 and 50, respectively.

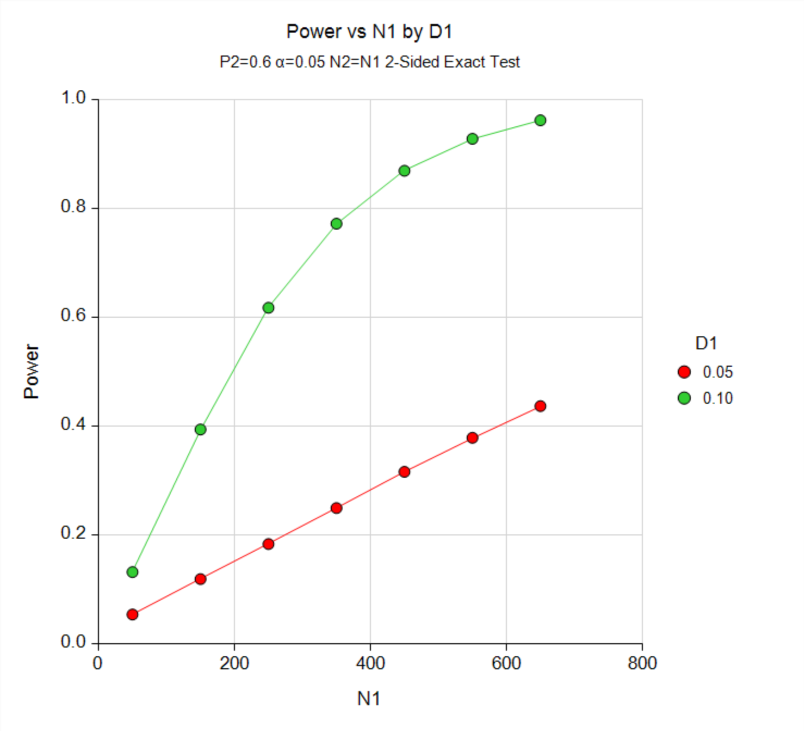
References

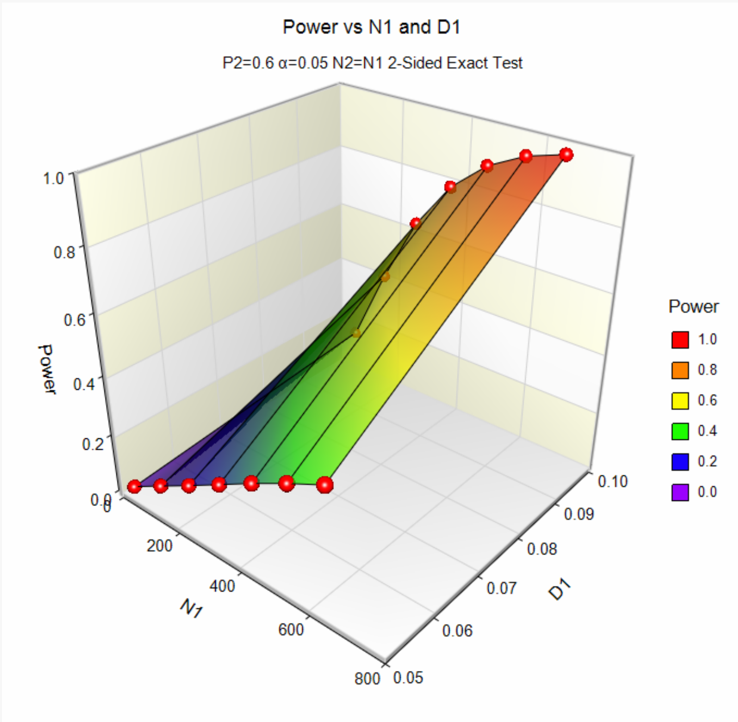
Bennett, B.M, and Hsu, P. 1960. 'On the power function of the exact test for the 2x2 contingency table', Biometrika, Volume 47, page 363-398.
Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Chapman & Hall/CRC. Boca Raton, Florida.
Julious, Steven A. 2010. Sample Sizes for Clinical Trials. CRC Press. New York.
Machin, D., Campbell, M., Tan, S.B., and Tan, S.H. 2018. Sample Size Tables for Clinical, Laboratory and Epidemiology Studies, 4th Edition. John Wiley & Sons. Hoboken, NJ.
Ryan, Thomas P. 2013. Sample Size Determination and Power. John Wiley & Sons. Hoboken, New Jersey.

This report shows the values of each of the parameters, one scenario per row. Notice that the approximate power values are close to the binomial enumeration values for almost all sample sizes.

Plots Section

Plots





The values from the table are displayed on the above plots.

Example 2 – Finding the Sample Size

A clinical trial is being designed to test effectiveness of new drug in reducing mortality. Suppose the current cure rate during the first year is 0.44. The sample size should be large enough to detect a difference in the cure rate of 0.10. Assuming the test statistic is a two-sided Fisher's Exact test with a significance level of 0.05, what sample size will be necessary to achieve 90% power?

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Sample Size**
Power Calculation Method	**Binomial Enumeration**
Maximum N1 or N2 for Binomial Enumeration	**1000**
Zero Count Adjustment Method	**Add to zero cells only**
Zero Count Adjustment Value	**0.0001**
Alternative Hypothesis	**Two-Sided**
Power	**0.90**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Input Type	**Proportions**
P1 (Group 1 Proportion|H1)	**0.54**
P2 (Group 2 Proportion)	**0.44**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For: Sample Size
Test Type: Fisher's Exact Test
Groups: 1 = Treatment, 2 = Control
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Power		Sample Size			Proportion		Difference	Alpha	
Target	Actual*	N1	N2	N	P1	P2	D1	Target	Actual*†
0.9	0.90028	546	546	1092	0.54	0.44	0.1	0.05	0.04207

* Power was computed using binomial enumeration of all possible outcomes. Actual alpha is only computed for two-sided tests.
† Warning: When solving for sample size with power computed using binomial enumeration, the target alpha level is not guaranteed. Actual alpha may be greater than target alpha in some cases. We suggest that you investigate sample sizes near the solution to find designs with an actual alpha you are willing to tolerate.

The required sample size is 546 per group.

As an exercise, change the Power Calculation Method to "Normal Approximation". When this is done, the sample size is 543—not much of a difference from the 546 that was found by exact power calculation. The actual alpha is 0.04207 which is close to the target of 0.05.

Example 3 – Validation using Bennett and Hsu (1960)

Bennett and Hsu (1960), page 396, present an example using Fisher's Exact test in which P1 = 0.8, P2 = 0.2, N1 = 10, N2 = 10, and alpha = 0.05. Assuming a one-sided test and equal sample allocation, they calculate the power to be 0.8054.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Power**
Power Calculation Method	**Binomial Enumeration**
Maximum N1 or N2 for Binomial Enumeration	**1000**
Zero Count Adjustment Method	**Add to zero cells only**
Zero Count Adjustment Value	**0.0**
Alternative Hypothesis	**One-Sided**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**10**
Input Type	**Proportions**
P1 (Group 1 Proportion|H1)	**0.8**
P2 (Group 2 Proportion)	**0.2**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For: Power
Test Type: Fisher's Exact Test
Groups: 1 = Treatment, 2 = Control
Hypotheses: H0: P1 - P2 ≤ 0 vs. H1: P1 - P2 > 0

	Sample Size			Proportion		Difference	Alpha	
Power*	N1	N2	N	P1	P2	D1	Target	Actual*
0.80539	10	10	20	0.8	0.2	0.6	0.05	

* Power was computed using binomial enumeration of all possible outcomes. Actual alpha is only computed for two-sided tests.

PASS found the power to be 0.80539, which matches the result in the journal.