Chapter 256

Multi-Arm Tests for Treatment and Control Proportions

Introduction

This module computes power and sample size for multiple comparisons of treatment proportions versus a control proportion based on the results in Machin, Campbell, Tan, and Tan (2018). In this design, there are *k* treatment groups and one control group. A proportion is measured in each group. A total of *k* hypothesis tests are anticipated each comparing a treatment group with the common control group using a simple z-test of the difference between two proportions.

The Bonferroni adjustment of the type I error rate may be optionally made because several comparisons are being tested using the same data. Making a multiplicity adjustment is usually recommended, but not always. In fact, Saville (1990) advocates not applying it and Machin, Campbell, Tan, and Tan (2018) include omitting it as a possibility.

Background

Whether you want to test several doses of a single treatment or several types of treatments, good research practice requires that each treatment be compared with a control. For example, a popular three-arm design consists of three groups: control, treatment A, and treatment B. Two tests are run: treatment A versus control and treatment B versus the same control. This avoids having to obtain a second control group for treatment B. Besides the obvious efficiency in subjects, it may be easier to recruit subjects if their chances of receiving the new treatment are better than 50-50.

Technical Details

Suppose you need *k* treatment groups with response probabilities P_i that have samples of size N_i and one control group with response probability P_C that has a sample of size N_C. The total sample size is $N = N_1 + N_2 + \ldots + N_k + N_C$.

The hypotheses for two-sided tests are

$$H_{0i}: P_i = P_C \quad \text{versus} \quad H_{1i}: P_i \neq P_C$$

and for one-sided tests the hypotheses are

$$H_{0i}: P_i \leq P_C \quad \text{versus} \quad H_{1i}: P_i > P_C$$

or

$$H_{0i}: P_i \geq P_C \quad \text{versus} \quad H_{1i}: P_i < P_C$$

Test Statistics

Seven test statistics are available in this routine. These are

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 present here is summarized from Section 26.3 (pages 866 – 870) of the StatXact-5 documentation. In this case, the test statistic is

$$T = -\ln\left[\frac{\binom{n_i}{x_i}\binom{n_C}{x_C}}{\binom{N}{m}}\right]$$

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

$$\Pr(T \geq t | m, H_0) = \sum_{A(m)}\left[\frac{\binom{n_i}{x_i}\binom{n_C}{x_C}}{\binom{N}{m}}\right]$$

where

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

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

$$\Pr(T \geq t_\alpha | m, H_1) = \sum_{A(m, T \geq t_\alpha)}\left[\frac{b(x_i, n_i, p_i)b(x_C, n_C, p_C)}{\sum_{A(m)} b(x_i, n_i, p_i)b(x_C, n_C, p_C)}\right]$$

$$\begin{aligned} P(m) &= \Pr(x_i + x_C = m | H_1) \\ &= b(x_i, n_i, p_i)b(x_C, n_C, p_C) \end{aligned}$$

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

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) (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test is usually 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 formula for the test statistic is

$$z_t = \frac{\hat{p}_i - \hat{p}_C}{\hat{\sigma}_D}$$

Pooled Version

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

$$\hat{p} = \frac{n_i\hat{p}_i + n_C\hat{p}_C}{n_i + n_C}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n_i} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_C}}$$

Power

The following normal approximation to the binomial is used as presented in Chow et al. (2008).

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value is that value of *z* that leaves exactly the target value of alpha in the tail.
2. Use the normal approximation to binomial distribution to compute binomial probabilities, compute the power for the pooled and unpooled tests, respectively, using

Pooled: $1-\beta = \Pr\left(Z < \frac{z_\alpha\sigma_{D,p} + (p_i - p_C)}{\sigma_{D,u}}\right)$ **Unpooled:** $1-\beta = \Pr\left(Z < \frac{z_\alpha\sigma_{D,u} + (p_i - p_C)}{\sigma_{D,u}}\right)$

where

$$\sigma_{D,u} = \sqrt{\frac{p_i q_i}{n_i} + \frac{p_C q_C}{n_C}} \qquad \text{(unpooled standard error)}$$

$$\sigma_{D,p} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_i} + \frac{1}{n_C}\right)} \qquad \text{(pooled standard error)}$$

with $\bar{p} = \dfrac{n_i p_i + n_C p_C}{n_i + n_C}$ and $\bar{q} = 1 - \bar{p}$

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

$$z = \frac{(\hat{p}_i - \hat{p}_C) + \dfrac{F}{2}\left(\dfrac{1}{n_i} + \dfrac{1}{n_C}\right)}{\hat{\sigma}_D}$$

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

Pooled Version

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

$$\hat{p} = \frac{n_i \hat{p}_i + n_C \hat{p}_C}{n_i + n_C}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n_i} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_C}}$$

Power

The power of this test is computed using the enumeration procedure described for the *z*-test above. For large samples, approximate results based on the normal approximation to the binomial are used.

Conditional Mantel-Haenszel Test

The conditional Mantel-Haenszel test, see Lachin (2000) page 40, is based on the *index frequency*, x_{11}, from a 2x2 table. The formula for the z-statistic is

$$z = \frac{x_{11} - E(x_{11})}{\sqrt{V_c(x_{11})}}$$

where

$$E(x_{11}) = \frac{n_1 m_1}{N}$$

$$V_c(x_{11}) = \frac{n_1 n_2 m_1 m_2}{N^2(N-1)}$$

Power

The power of this test is computed using the normal approximation to the binomial described above.

Likelihood Ratio Test

In 1935, Wilks showed that the following quantity has a chi-square distribution with one degree of freedom. Using this test statistic to compare proportions is presented, among other places, in Upton (1982). The likelihood ratio test statistic is computed as

$$LR = 2\begin{bmatrix} a\ln(a) + b\ln(b) + c\ln(c) + d\ln(d) + \\ N\ln(N) - s\ln(s) - f\ln(f) - m\ln(m) - n\ln(n) \end{bmatrix}$$

Power

The power of this test is computed using the normal approximation to the binomial described above.

Multiplicity Adjustment

Because *k* z-tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that the Bonferroni adjustment be applied. This adjustment is easy to apply: the value of alpha that is used in the test is found by dividing the original alpha by the number of tests. For example, if the original alpha is set at 0.05 and the number of treatment (not including the control) groups is five, the individual tests will be conducted using an alpha of 0.01.

The main criticism of this procedure is that if there are many tests, the value of alpha becomes very small. To mitigate against this complaint, some statisticians recommend separating the treatment groups into those that are of primary interest and those that are of secondary interest. The Bonferroni adjustment is made by the using the number of primary treatments rather than the total number of treatments.

There are some who advocate ignoring the adjustment entirely in the case of randomized clinical trials. See for example Saville (1990) and the discussion in chapter 14 of Machin, Campbell, Tan, and Tan (2018).

Size of the Control Group

Because the control group is used over and over, some advocate increasing the number of subjects in this group. The standard adjustment is to include $\sqrt{k}$ subjects in the control group for each subject in one of the treatment groups. See Machin, Campbell, Tan, and Tan (2018, pages 231-232). Note that often, the treatment groups all have the same size.

Example 1 – Finding the Sample Size

A parallel-group clinical trial is being designed to compare three doses of a test compound against the standard therapy in patients with a specific type of disease. Suppose the standard therapy has a response rate of 60%. The investigators would like a sample size large enough to find statistical significance at the 0.05 level if the actual response rate is at least 70% and the power is 0.80 in each test. They also want to see the impact on sample size of response rates of 75% and 80%. The tests will be two-sided.

Following standard procedure, the control group multiplier will be set to $\sqrt{k} = \sqrt{3} = 1.732$ since the control group is used for three comparisons in this design.

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	**Sample Size**
Alternative Hypothesis	**Two-Sided (H1: Pi ≠ Pc)**
Test Type	**Z-Test (Unpooled)**
Power of Each Test	**0.80**
Overall Alpha	**0.05**
Bonferroni Adjustment	**None**
Group Allocation	**Enter Group Allocation Pattern, solve for group sample sizes**
Control Proportion	**0.6**
Control Sample Size Allocation	**1.73**
Set A Number of Groups	**3**
Set A Proportion	**0.7 0.75 0.8**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**0**
Set C Number of Groups	**0**
Set D Number of Groups	**0**
More	**Unchecked**

Output

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

Numeric Reports

Numeric Results

Solve For:	Sample Size
Group Allocation:	Enter Group Allocation Pattern, solve for group sample sizes
Test Type:	Z-Test with unpooled variance
Hypotheses:	H0: Pi = Pc vs. H1: Pi ≠ Pc
Number of Groups:	4
Bonferroni Adjustment:	None (Divisor = 1)

	Power		Sample Size			Comparison Statistics			
Comparison	Target	Actual	Ni	Allocation	Proportion Pi	Difference δi	Ratio Ri	Odds Ratio ORi	Alpha
Control			474	1.73	0.60				
vs A1	0.8	0.80041	274	1.00	0.70	0.10	1.16667	1.55556	0.05
vs A2	0.8	0.80041	274	1.00	0.70	0.10	1.16667	1.55556	0.05
vs A3	0.8	0.80041	274	1.00	0.70	0.10	1.16667	1.55556	0.05
Total			1296						
Control			197	1.73	0.60				
vs A1	0.8	0.80050	114	1.00	0.75	0.15	1.25000	2.00000	0.05
vs A2	0.8	0.80050	114	1.00	0.75	0.15	1.25000	2.00000	0.05
vs A3	0.8	0.80050	114	1.00	0.75	0.15	1.25000	2.00000	0.05
Total			539						
Control			102	1.73	0.60				
vs A1	0.8	0.80242	59	1.00	0.80	0.20	1.33333	2.66667	0.05
vs A2	0.8	0.80242	59	1.00	0.80	0.20	1.33333	2.66667	0.05
vs A3	0.8	0.80242	59	1.00	0.80	0.20	1.33333	2.66667	0.05
Total			279						

Comparison	The group that is involved in the comparison between the treatment and control displayed on this report line. The comparison may be made using the difference, ratio, or odds ratio--they all yield the same results.
Target Power	The power desired. Power is probability of rejecting a false null hypothesis for this comparison. This power is of the comparison shown on this line only.
Actual Power	The power actually achieved.
Ni	Sample Size. The number of subjects in the ith group. The total sample size, N, is shown as the last row of the column.
Allocation	The group sample size allocation pattern. The value on each row represents the relative number of subjects assigned to the group.
Pi	The response proportion in the ith group at which the power is calculated.
δi	The difference between the ith group proportion (Pi) and the control group proportion (Pc) at which the power is calculated. The formula is δi = Pi - Pc.
Ri	The ratio of the ith group proportion (Pi) and the control group proportion (Pc) at which the power is calculated. The formula is Ri = Pi / Pc.
ORi	The odds ratio of the ith treatment group proportion (Pi) and the control group proportion (Pc) at which the power is calculated. The formula is ORi = [Pi / (1 - Pi)] / [Pc / (1 - Pc)].
Alpha	The probability of rejecting the null hypothesis that the control mean is equal to the treatment mean described on this line.

Summary Statements

A parallel, 4-group design (with one control group and 3 treatment groups) will be used to test whether the proportion for each treatment group is different from the control group proportion (H0: Pi = Pc versus H1: Pi ≠ Pc). The hypotheses will be evaluated using 3 two-sided, two-sample, Bonferroni-adjusted Z-tests with unpooled variance, with an overall (experiment-wise) Type I error rate (α) of 0.05. The control group proportion is assumed to be 0.6. To detect the treatment proportions 0.7, 0.7, and 0.7 with at least 80% power for each test, the control group sample size needed will be 474 and the number of needed subjects for the treatment groups will be 274, 274, and 274 (totaling 1296 subjects overall).

Dropout-Inflated Sample Size

Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
1	20%	474	593	119
2	20%	274	343	69
3	20%	274	343	69
4	20%	274	343	69
Total		1296	1622	326
1	20%	197	247	50
2	20%	114	143	29
3	20%	114	143	29
4	20%	114	143	29
Total		539	676	137
1	20%	102	128	26
2	20%	59	74	15
3	20%	59	74	15
4	20%	59	74	15
Total		279	350	71

Group — Lists the group numbers.

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.

Ni — The evaluable sample size for each group at which power is computed (as entered by the user). If Ni subjects are evaluated out of the Ni' subjects that are enrolled in the study, the design will achieve the stated power.

Ni' — The number of subjects that should be enrolled in each group in order to obtain Ni evaluable subjects, based on the assumed dropout rate. Ni' is calculated by inflating Ni using the formula Ni' = Ni / (1 - DR), with Ni' 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.)

Di — The expected number of dropouts in each group. Di = Ni' - Ni.

Dropout Summary Statements

Anticipating a 20% dropout rate, group sizes of 593, 343, 343, and 343 subjects should be enrolled to obtain final group sample sizes of 474, 274, 274, and 274 subjects.

References

Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, 3rd Edition. Chapman & Hall/CRC. Boca Raton, FL. Pages 86-88.

D'Agostino, R.B., Chase, W., and 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., and Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.

Lachin, J.M. 2000. Biostatistical Methods. John Wiley & Sons. New York.

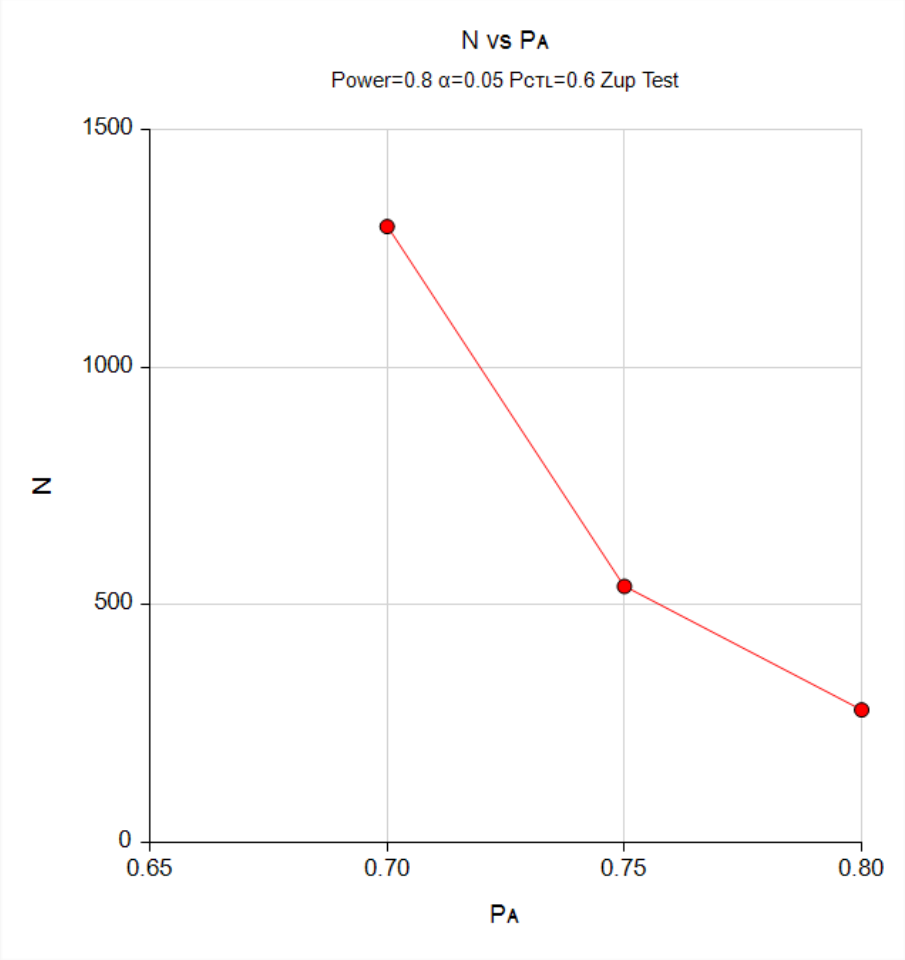
Machin, D., Campbell, M.J., Tan, S.B, and Tan, S.H. 2018. Sample Sizes for Clinical, Laboratory, and Epidemiology Studies, 4th Edition. Wiley Blackwell.

Ryan, T.P. 2013. Sample Size Determination and Power. John Wiley & Sons. Hoboken, New Jersey.

This report shows the numeric results of this power study. Notice that the results are shown in blocks of three rows at a time. Each block represents a single design.

Plots Section

Plots



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the sample size of decreasing the difference between the treatment and control proportions.

Example 2 – Validation using Chow, Shao, and Wang (2008)

Chow, Shao, and Wang (2008) page 100 gives an example in which the control response rate is 20%, treatment 1 response rate is 40%, and treatment 2 response rate is 50%. They use two-sided tests and set the overall significance level at 0.05. They set the target power at 0.80. They calculate the sample sizes for treatment 1 as about 95 and treatment 2 as about 45. Thus, the required sample size is 95.

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**
Alternative Hypothesis	**Two-Sided (H1: Pi ≠ Pc)**
Test Type	**Z-Test (Unpooled)**
Power of Each Test	**0.80**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
Control Proportion	**0.2**
Set A Number of Groups	**1**
Set A Proportion	**0.4**
Set B Number of Groups	**1**
Set B Proportion	**0.5**
Set C Number of Groups	**0**
Set D Number of Groups	**0**
More	**Unchecked**

Output

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

Numeric Results

Solve For:	Sample Size
Group Allocation:	Equal (Nc = N1 = N2 = ...)
Test Type:	Z-Test with unpooled variance
Hypotheses:	H0: Pi = Pc vs. H1: Pi ≠ Pc
Number of Groups:	3
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 2)

	Power		Sample Size		Comparison Statistics			Alpha	
Comparison	Target	Actual	Ni	Proportion Pi	Difference δi	Ratio Ri	Odds Ratio ORi	Overall	Bonferroni-Adjusted
Control			96	0.2					
vs A	0.8	0.80427	96	0.4	0.2	2.0	2.66667	0.05	0.025
vs B	0.8	0.99059	96	0.5	0.3	2.5	4.00000	0.05	0.025
Total			288						

PASS obtained 96 which matches Chow, Shao, and Wang's results to within rounding. The final design would be 96 subjects in each of the three groups.