Chapter 423

Tests for Two Groups Assuming a Two-Part Model

Introduction

This procedure provides sample size and power calculations for comparing two groups when the data come from a two-part model. Two-part models assume that the data distribution is a mixture of a probability mass at zero and a continuous distribution for data values greater than zero. Often, the continuous distribution is assumed to be normal. The power formula is based on the work of Lachenbruch (2001).

Technical Details

Two-part models assume that the data distribution is a mixture of a probability mass at zero and a continuous distribution for data values greater than zero. The density of a variable *X* from this model is written as

$$f(x) = p_i^{\theta}[(1 - p_i)h_i(x)]^{1-\theta}$$

where *p* is the probability that *X* = 0 and $h_i(x)$ is a continuous density.

Test Statistic

The test statistic for the two-part model is

$$\chi^2_{(2)} = \left[\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right]^2 + U^2$$

where *U* is a statistic that compares the non-zero parts of the model. It can be assumed to be a *z*, *t*, or Wilcoxon test statistic computed on the non-zero data values.

Power Analysis

The power analysis is based on non-central chi-squared distribution. The non-centrality parameter of this distribution is given by

$$\lambda = \frac{(p_1 - p_2)^2 n_1 n_2}{\bar{p}(1 - \bar{p})(n_1 + n_2)} + \frac{\delta^2}{\sigma^2 \left[\frac{1}{(1 - p_1)n_1} + \frac{1}{(1 - p_2)n_2}\right]}$$

where $\delta = \mu_1 - \mu_2$, the difference in the means of the continuous portion of the two groups.

The power is the probability a two degree of freedom chi-squared random variable with non-centrality λ exceeds the critical value computed from a central chi-squared distribution.

Example 1 – Finding the Sample Size

Researchers wish to compare the average medical expenditure by patients in two well-defined groups. Since medical expenses are often zero, a two-part model is used. The researchers set δ to a range of values from 100 to 350. They set the standard deviation to the range 85 to 100 based on previous studies. They set P1 = P2 = 0.80. They want to determine how many participants are needed to achieve 90% power at a significance level of 0.05.

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**
Power.. **0.90**
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
P1 (Prob X = 0 in Group 1) **0.80**
P2 (Prob X = 0 in Group 2) **0.80**
δ (Mean Difference = μ1 - μ2|X > 0) .... **100 to 350 by 50**
σ (Standard Deviation of X|X > 0)................. **85 to 100 by 5**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Hypotheses: H0: P1 - P2 = 0 and δ = μ1 - μ2 = 0
H1: P1 - P2 ≠ 0 and/or δ = μ1 - μ2 ≠ 0

Power		Sample Size			Binomial Part (X = 0)		Normal Part (X > 0)		
							Mean Difference	Standard Deviation	
Target	Actual	N1	N2	N	P1	P2	δ	σ	Alpha
0.9	0.9019	92	92	184	0.8	0.8	100	85	0.05
0.9	0.9015	103	103	206	0.8	0.8	100	90	0.05
0.9	0.9021	115	115	230	0.8	0.8	100	95	0.05
0.9	0.9011	127	127	254	0.8	0.8	100	100	0.05
0.9	0.9027	41	41	82	0.8	0.8	150	85	0.05
0.9	0.9029	46	46	92	0.8	0.8	150	90	0.05
0.9	0.9014	51	51	102	0.8	0.8	150	95	0.05
0.9	0.9040	57	57	114	0.8	0.8	150	100	0.05
0.9	0.9019	23	23	46	0.8	0.8	200	85	0.05
0.9	0.9044	26	26	52	0.8	0.8	200	90	0.05
0.9	0.9047	29	29	58	0.8	0.8	200	95	0.05
0.9	0.9035	32	32	64	0.8	0.8	200	100	0.05
0.9	0.9075	15	15	30	0.8	0.8	250	85	0.05
0.9	0.9106	17	17	34	0.8	0.8	250	90	0.05
0.9	0.9115	19	19	38	0.8	0.8	250	95	0.05
0.9	0.9108	21	21	42	0.8	0.8	250	100	0.05
0.9	0.9225	11	11	22	0.8	0.8	300	85	0.05
0.9	0.9152	12	12	24	0.8	0.8	300	90	0.05
0.9	0.9072	13	13	26	0.8	0.8	300	95	0.05
0.9	0.9186	15	15	30	0.8	0.8	300	100	0.05
0.9	0.9198	8	8	16	0.8	0.8	350	85	0.05
0.9	0.9208	9	9	18	0.8	0.8	350	90	0.05
0.9	0.9200	10	10	20	0.8	0.8	350	95	0.05
0.9	0.9181	11	11	22	0.8	0.8	350	100	0.05

Target Power The desired power value entered in the procedure. Power is the probability of rejecting a false null hypothesis.
Actual Power The actual power achieved. Because N1 and N2 are discrete, this value is usually slightly larger than the target power.
N1 The sample size from group 1.
N2 The sample size from group 2.
N The total sample size from both groups. N = N1 + N2.
P1 The probability that X = 0 in group 1.
P2 The probability that X = 0 in group 2.
δ The difference between group means at which power and sample size calculations are made. δ = μ1 - μ2.
σ The assumed population standard deviation for each of the two groups.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A parallel two-group design will be used to compare two groups for the scenario where the data follow a two-part model. Two-part models assume that the data distribution for each group is a mixture of a probability mass at zero and a continuous distribution for data values greater than zero. The comparison of groups will be made using a two-sided, two-sample Chi-square test (with 2 degrees of freedom), with a Type I error rate (α) of 0.05. The test combines two comparisons: the difference in the proportion of 0's and the difference in means (H0: P1 - P2 = 0 and μ1 - μ2 = 0 versus H1: P1 - P2 ≠ 0 and/or H1: μ1 - μ2 ≠ 0). The proportion of 0's in Group 1 is assumed to be 0.8 and the proportion of 0's in Group 2 is assumed to be 0.8. The common standard deviation for the continuous values in both groups is assumed to be 85. To detect a difference in means of 100 with 90% power (for the combined test), the number of needed subjects will be 92 in Group 1 and 92 in Group 2.

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%	92	92	184	115	115	230	23	23	46
20%	103	103	206	129	129	258	26	26	52
20%	115	115	230	144	144	288	29	29	58
20%	127	127	254	159	159	318	32	32	64
20%	41	41	82	52	52	104	11	11	22
20%	46	46	92	58	58	116	12	12	24
20%	51	51	102	64	64	128	13	13	26
20%	57	57	114	72	72	144	15	15	30
20%	23	23	46	29	29	58	6	6	12
20%	26	26	52	33	33	66	7	7	14
20%	29	29	58	37	37	74	8	8	16
20%	32	32	64	40	40	80	8	8	16
20%	15	15	30	19	19	38	4	4	8
20%	17	17	34	22	22	44	5	5	10
20%	19	19	38	24	24	48	5	5	10
20%	21	21	42	27	27	54	6	6	12
20%	11	11	22	14	14	28	3	3	6
20%	12	12	24	15	15	30	3	3	6
20%	13	13	26	17	17	34	4	4	8
20%	15	15	30	19	19	38	4	4	8
20%	8	8	16	10	10	20	2	2	4
20%	9	9	18	12	12	24	3	3	6
20%	10	10	20	13	13	26	3	3	6
20%	11	11	22	14	14	28	3	3	6

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. 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. After solving for N1 and N2, 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, 115 subjects should be enrolled in Group 1, and 115 in Group 2, to obtain final group sample sizes of 92 and 92, respectively.

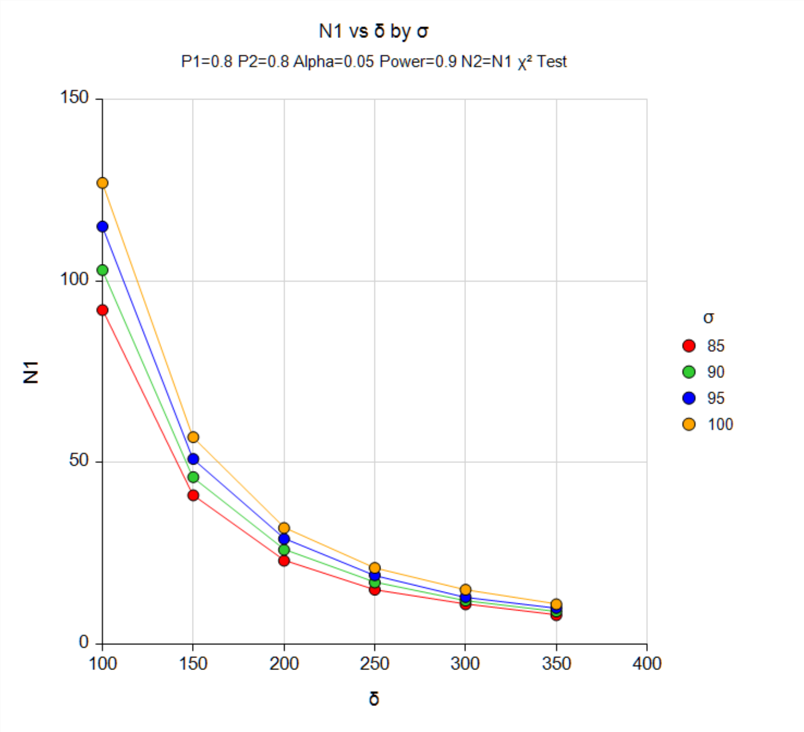
References

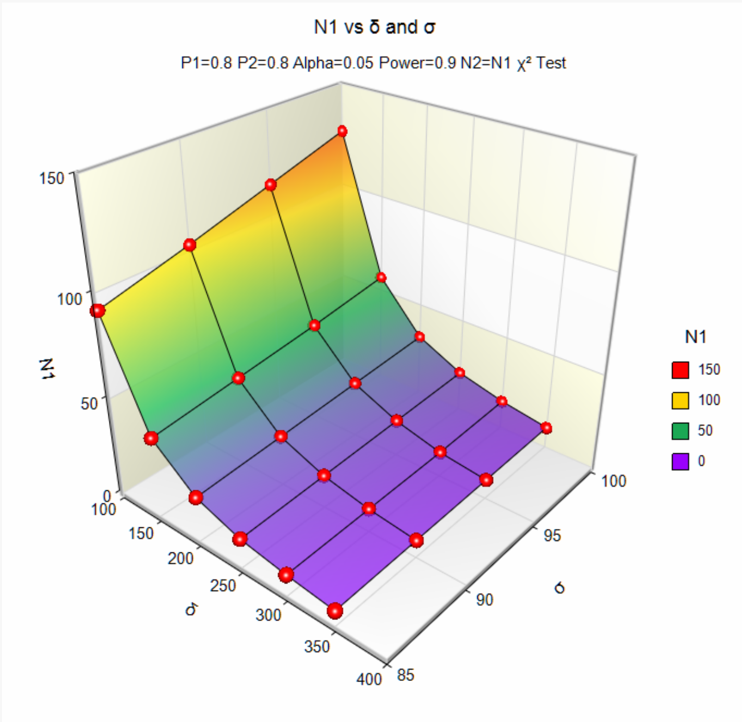
Lachenbruch, P.A. 2001. 'Power and sample size requirements for two-part models'. Statistics in Medicine, Vol. 20, Pages 1235-1238.

These reports show the values of each of the parameters, one scenario per row.

Plots Section

Plots





These plots show the relationship between the standard deviation and sample size for the two alpha levels.

Example 2 – Validation using Lachenbruch (2001)

Lachenbruch (2001) gives an example of calculated values on page 1237 of his article. The sixth line of Table I on this page gives the power when P1 = 0.1, P2 = 0.2, δ = 0.3, σ = 1, N1 = N2 = 100, and alpha = 0.05. He reports a power of 0.702.

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 .... **Power**
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **100**
P1 (Prob X = 0 in Group 1) **0.1**
P2 (Prob X = 0 in Group 2) **0.2**
δ (Mean Difference = μ1 - μ2|X > 0) **0.3**
σ (Standard Deviation of X|X > 0)................. **1**

Output

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

Numeric Results

Solve For: Power
Hypotheses: H0: P1 - P2 = 0 and δ = μ1 - μ2 = 0
H1: P1 - P2 ≠ 0 and/or δ = μ1 - μ2 ≠ 0

	Sample Size			Binomial Part (X = 0)		Normal Part (X > 0)		
						Mean Difference	Standard Deviation	
Power	N1	N2	N	P1	P2	δ	σ	Alpha
0.7019	100	100	200	0.1	0.2	0.3	1	0.05

The value for power calculated by **PASS** matches the expected result.