

## 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.

## Tests for Two Groups Assuming a Two-Part Model

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

---

### Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

---

### Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

---

#### Solve For

##### Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power*, *Sample Size*, and *Mean Difference*. In most situations, you will likely select either *Power* or *Sample Size*.

The 'Solve For' parameter will be displayed on the vertical axis of any plots that are generated.

---

#### Power and Alpha

##### Power

Power is the probability of rejecting the null hypothesis when it is false. Power is equal to  $1 - \text{Beta}$ , so specifying power implicitly specifies beta. Beta is the probability obtaining a false negative with the statistical test. That is, it is the probability of accepting a false null hypothesis.

The valid range is 0 to 1. Different disciplines have different standards for setting power. The most common choice is 0.90, but 0.80 is also popular.

You can enter a single value, such as 0.90, or a series of values, such as .70 .80 .90, or .70 to .90 by .1.

When a series of values is entered, PASS will generate a separate calculation result for each value of the series.

##### Alpha

Alpha is the probability of obtaining a false positive with the statistical test. That is, it is the probability of rejecting a true null hypothesis. The null hypothesis is that the group proportions and means are equal.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.10.

Alpha is often set to 0.05 for two-sided tests such as this.

You can enter a single value, such as 0.05, or a series of values, such as .05 .10 .15, or .05 to .15 by .01.

When a series of values is entered, PASS will generate a separate calculation result for each value of the series.

## Tests for Two Groups Assuming a Two-Part Model

### Sample Size (When Solving for Sample Size)

#### Group Allocation

Select the option that describes the constraints on  $N1$  or  $N2$  or both.

The options are

- Equal ( $N1 = N2$ )**  
 This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.
- Enter  $N1$ , solve for  $N2$**   
 Select this option when you wish to fix  $N1$  at some value (or values), and then solve only for  $N2$ . Please note that for some values of  $N1$ , there may not be a value of  $N2$  that is large enough to obtain the desired power.
- Enter  $N2$ , solve for  $N1$**   
 Select this option when you wish to fix  $N2$  at some value (or values), and then solve only for  $N1$ . Please note that for some values of  $N2$ , there may not be a value of  $N1$  that is large enough to obtain the desired power.
- Enter  $R = N2/N1$ , solve for  $N1$  and  $N2$**   
 For this choice, you set a value for the ratio of  $N2$  to  $N1$ , and then **PASS** determines the needed  $N1$  and  $N2$ , with this ratio, to obtain the desired power. An equivalent representation of the ratio,  $R$ , is
 
$$N2 = R * N1.$$
- Enter percentage in Group 1, solve for  $N1$  and  $N2$**   
 For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then **PASS** determines the needed  $N1$  and  $N2$  with this percentage to obtain the desired power.

#### **$N1$ (Sample Size, Group 1)**

*This option is displayed if Group Allocation = "Enter  $N1$ , solve for  $N2$ "*

$N1$  is the number of items or individuals sampled from the Group 1 population.

$N1$  must be  $\geq 2$ . You can enter a single value or a series of values.

#### **$N2$ (Sample Size, Group 2)**

*This option is displayed if Group Allocation = "Enter  $N2$ , solve for  $N1$ "*

$N2$  is the number of items or individuals sampled from the Group 2 population.

$N2$  must be  $\geq 2$ . You can enter a single value or a series of values.

#### **$R$ (Group Sample Size Ratio)**

*This option is displayed only if Group Allocation = "Enter  $R = N2/N1$ , solve for  $N1$  and  $N2$ ."*

$R$  is the ratio of  $N2$  to  $N1$ . That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of  $N2$  to  $N1$  while solving for  $N1$  and  $N2$ . Only sample size combinations with this ratio are considered.

$N2$  is related to  $N1$  by the formula:

$$N2 = [R \times N1],$$

where the value  $[Y]$  is the next integer  $\geq Y$ .

## Tests for Two Groups Assuming a Two-Part Model

For example, setting  $R = 2.0$  results in a Group 2 sample size that is double the sample size in Group 1 (e.g.,  $N1 = 10$  and  $N2 = 20$ , or  $N1 = 50$  and  $N2 = 100$ ).

$R$  must be greater than 0. If  $R < 1$ , then  $N2$  will be less than  $N1$ ; if  $R > 1$ , then  $N2$  will be greater than  $N1$ . You can enter a single or a series of values.

### Percent in Group 1

*This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for  $N1$  and  $N2$ ."*

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for  $N1$  and  $N2$ . Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

---

## Sample Size (When Not Solving for Sample Size)

### Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal ( $N1 = N2$ )**  
This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.
- **Enter  $N1$  and  $N2$  individually**  
This choice permits you to enter different values for  $N1$  and  $N2$ .
- **Enter  $N1$  and  $R$ , where  $N2 = R * N1$**   
Choose this option to specify a value (or values) for  $N1$ , and obtain  $N2$  as a ratio (multiple) of  $N1$ .
- **Enter total sample size and percentage in Group 1**  
Choose this option to specify a value (or values) for the total sample size ( $N$ ), obtain  $N1$  as a percentage of  $N$ , and then  $N2$  as  $N - N1$ .

### Sample Size Per Group

*This option is displayed only if Group Allocation = "Equal ( $N1 = N2$ )."*

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for  $N1$ , and also the value for  $N2$ .

The Sample Size Per Group must be  $\geq 2$ . You can enter a single value or a series of values.

### $N1$ (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter  $N1$  and  $N2$  individually" or "Enter  $N1$  and  $R$ , where  $N2 = R * N1$ ."*

$N1$  is the number of items or individuals sampled from the Group 1 population.

$N1$  must be  $\geq 2$ . You can enter a single value or a series of values.

## Tests for Two Groups Assuming a Two-Part Model

### N2 (Sample Size, Group 2)

*This option is displayed only if Group Allocation = "Enter N1 and N2 individually."*

$N2$  is the number of items or individuals sampled from the Group 2 population.

$N2$  must be  $\geq 2$ . You can enter a single value or a series of values.

### R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter N1 and R, where  $N2 = R * N1$ ."*

$R$  is the ratio of  $N2$  to  $N1$ . That is,

$$R = N2/N1$$

Use this value to obtain  $N2$  as a multiple (or proportion) of  $N1$ .

$N2$  is calculated from  $N1$  using the formula:

$$N2 = [R \times N1],$$

where the value  $[Y]$  is the next integer  $\geq Y$ .

For example, setting  $R = 2.0$  results in a Group 2 sample size that is double the sample size in Group 1.

$R$  must be greater than 0. If  $R < 1$ , then  $N2$  will be less than  $N1$ ; if  $R > 1$ , then  $N2$  will be greater than  $N1$ . You can enter a single value or a series of values.

### Total Sample Size (N)

*This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."*

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines  $N1$  and  $N2$ .

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

### Percent in Group 1

*This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."*

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

## Tests for Two Groups Assuming a Two-Part Model

---

### Effect Size - Binomial Part ( $X = 0$ )

#### P1 (Prob $X = 0$ in Group 1)

Enter one or more values for the probability that a data value is zero in group 1. This value is compared with P2 in the chi-squared test.

You can enter a single value such as 0.1 or a series of values such as 0.1 0.2 0.3 or 0.1 to 0.9 by 0.1.

The range is given by  $0 \leq P1 < 1$ .

#### P2 (Prob $X = 0$ in Group 2)

Enter one or more values for the probability that a data value is zero in group 2. This value is compared with P1 in the chi-squared test.

You can enter a single value such as 0.1 or a series of values such as 0.1 0.2 0.3 or 0.1 to 0.9 by 0.1.

The range is given by  $0 \leq P2 < 1$ .

---

### Effect Size – Normal Part ( $X > 0$ )

#### $\delta$ (Mean Difference = $\mu_1 - \mu_2 | X > 0$ )

Enter one or more values for the difference between the group means at which the design is powered. Note that this is the mean of the continuous part of the data. It does not include the binomial (0) part.

The permissible range of values is  $0 \leq \delta$ . Note that  $\delta^2$  is used in the test statistic, so the sign does not matter.

You can enter a single value such as 4 or a series of values such as 1 2 3 or 1 to 10 by 1.

When a series of values is entered, PASS will generate a separate calculation result for each value of the series.

---

### Standard Deviation

#### $\sigma$ (Standard Deviation of $X | X > 0$ )

Enter one or more values for the standard deviation of the non-zero portion of the data. This is the standard deviation used for both groups.

$\sigma$  must be a positive number.

You can enter a single value, such as 5, or a series of values, such as 1 3 5 7 9, or 1 to 10 by 1.

When a series of values is entered, PASS will generate a separate calculation result for each value of the series.

Press the small ' $\sigma$ ' button to the right to obtain options for estimating the standard deviation.

## Tests for Two Groups Assuming a Two-Part Model

## 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  $\delta$  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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Groups Assuming a Two-Part Model** procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
P1 (Prob X = 0 in Group 1) .....	<b>0.80</b>
P2 (Prob X = 0 in Group 2) .....	<b>0.80</b>
$\delta$ (Mean Difference = $\mu_1 - \mu_2   X > 0$ ) .....	<b>100 to 350 by 50</b>
$\sigma$ (Standard Deviation of $X   X > 0$ ) .....	<b>85 to 100 by 5</b>

### Annotated Output

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

### Numeric Results

#### Numeric Results

H0:  $P1 - P2 = 0$  and  $\delta = \mu_1 - \mu_2 = 0$   
 H1:  $P1 - P2 \neq 0$  and/or  $\delta = \mu_1 - \mu_2 \neq 0$

Target Power	Actual Power	N1	N2	N	P1	P2	$\delta$	$\sigma$	Alpha
0.90	0.9019	92	92	184	0.80	0.80	100.0	85.0	0.050
0.90	0.9015	103	103	206	0.80	0.80	100.0	90.0	0.050
0.90	0.9021	115	115	230	0.80	0.80	100.0	95.0	0.050
0.90	0.9011	127	127	254	0.80	0.80	100.0	100.0	0.050
0.90	0.9027	41	41	82	0.80	0.80	150.0	85.0	0.050
0.90	0.9029	46	46	92	0.80	0.80	150.0	90.0	0.050
0.90	0.9014	51	51	102	0.80	0.80	150.0	95.0	0.050
0.90	0.9040	57	57	114	0.80	0.80	150.0	100.0	0.050
0.90	0.9019	23	23	46	0.80	0.80	200.0	85.0	0.050
0.90	0.9044	26	26	52	0.80	0.80	200.0	90.0	0.050
0.90	0.9047	29	29	58	0.80	0.80	200.0	95.0	0.050
0.90	0.9035	32	32	64	0.80	0.80	200.0	100.0	0.050
0.90	0.9075	15	15	30	0.80	0.80	250.0	85.0	0.050
0.90	0.9106	17	17	34	0.80	0.80	250.0	90.0	0.050
0.90	0.9115	19	19	38	0.80	0.80	250.0	95.0	0.050
0.90	0.9108	21	21	42	0.80	0.80	250.0	100.0	0.050

(Report Continues)

#### References

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

## Tests for Two Groups Assuming a Two-Part Model

### Report Definitions

Target Power is the desired power value entered in the procedure. Power is the probability of rejecting a false null hypothesis.

Actual Power is the actual power achieved. Because N1 and N2 are discrete, this value is usually slightly larger than the target power.

N1 is the sample size from group 1.

N2 is the sample size from group 2.

N is the total sample size from both groups.  $N = N1 + N2$ .

P1 is the probability that  $X = 0$  in group 1.

P2 is the probability that  $X = 0$  in group 2.

$\delta = \mu1 - \mu2$  is the difference between group means at which power and sample size calculations are made.

$\sigma$  is the assumed population standard deviation for each of the two groups.

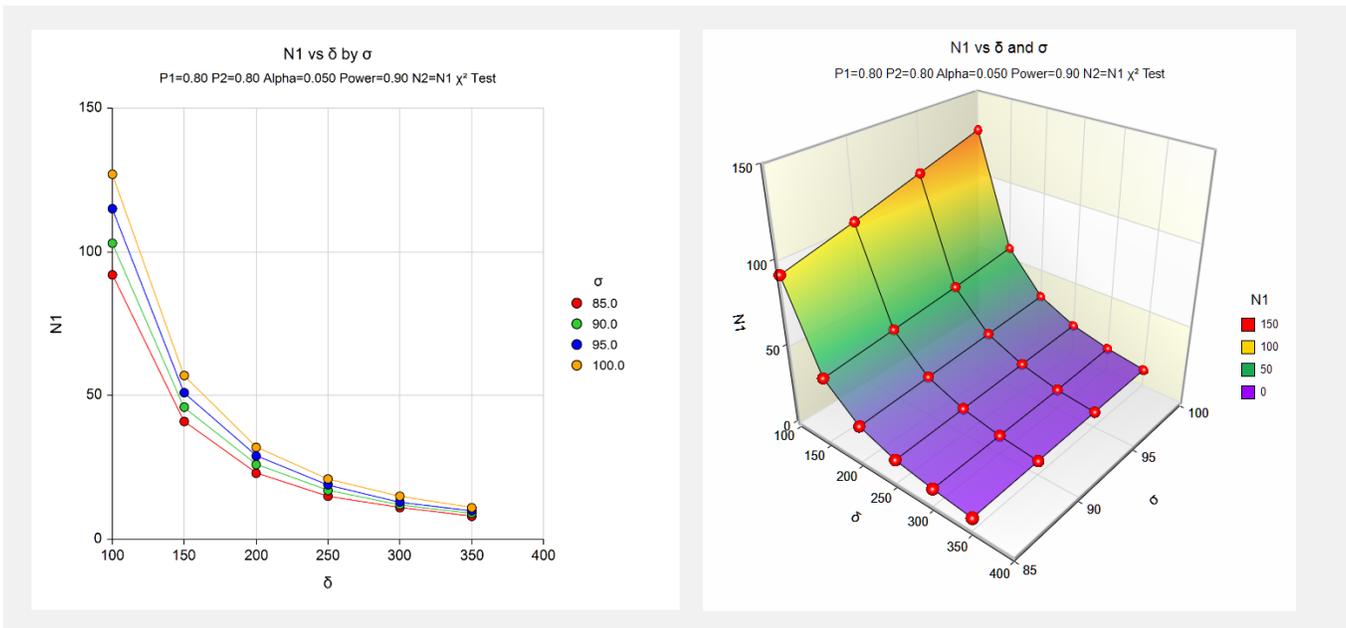
Alpha is the probability of rejecting a true null hypothesis, H0.

### Summary Statements

A test comparing two groups will use a two-part model which is a combination of a discrete distribution and a continuous distribution. The discrete part is a 'probability mass' at zero. The continuous part gives the distribution of values greater than zero. A chi-squared test with 2 degrees of freedom combines a test of the difference in proportions of 0's in the two groups with a test of the difference in group means of the continuous part. Group sample sizes of 92 and 92 achieve 90% power to reject the null hypothesis of equality of both parts at a significance level of 0.050. The proportion of 0's in group 1 is 0.80 and in group 2 is 0.80. The difference in the means of the continuous part is 100.0. The standard deviation of the continuous values is 85.0..

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

### Plots Section



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

## Tests for Two Groups Assuming a Two-Part Model

**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  $P1 = 0.1$ ,  $P2 = 0.2$ ,  $\delta = 0.3$ ,  $\sigma = 1$ ,  $N1 = N2 = 100$ , and  $\alpha = 0.05$ . He reports a power of 0.702

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Groups Assuming a Two-Part Model** procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>100</b>
P1 (Prob X = 0 in Group 1) .....	<b>0.1</b>
P2 (Prob X = 0 in Group 2) .....	<b>0.2</b>
$\delta$ (Mean Difference = $\mu_1 - \mu_2   X > 0$ ) .....	<b>0.3</b>
$\sigma$ (Standard Deviation of $X   X > 0$ ).....	<b>1</b>

**Output**

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

**Numeric Results**

<b>Numeric Results</b>									
H0: $P1 - P2 = 0$ and $\delta = \mu_1 - \mu_2 = 0$									
H1: $P1 - P2 \neq 0$ and/or $\delta = \mu_1 - \mu_2 \neq 0$									
<b>Power</b>	<b>N1</b>	<b>N2</b>	<b>N</b>	<b>P1</b>	<b>P2</b>	<b><math>\delta</math></b>	<b><math>\sigma</math></b>	<b>Alpha</b>	
<b>0.7019</b>	100	100	200	0.10	0.20	0.3	1.0	0.050	

You can see that **PASS** also calculates the power to be 0.702 (within rounding).