

Chapter 425

Tests for Two Groups Assuming a Two-Part Model with Detection Limits

Introduction

This procedure provides sample size and power calculations for comparing two groups when the data come from a two-part model with left censoring because the data is below a detection limit. Two-part models assume that the data distribution is a mixture of a probability mass at zero (or some other constant) and a continuous distribution for data values greater than zero. Often, the continuous distribution is assumed to be log-normal. The power formula is based on the work of Chu, Nie, and Cole (2006).

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. This model can be extended by also assuming that data values below a detection limit are too small to be observed. The density of a variable X from this model is written as

$$f_i(x) = [P_i + (1 - P_i)\Phi(\varepsilon_{iL})]^\theta \times \left[\frac{(1 - P_i)}{\sigma} \varphi\left(\frac{x_i - \mu_i}{\sigma}\right) \right]^{1-\theta}$$

where $\varepsilon_{iL} = \frac{x_L - \mu_i}{\sigma}$ and P_i is the probability that $X = 0$.

Note that the means and standard deviation are in the scale of the transformed data values.

The test statistic for this model is the usual test for the difference in means constructed on the observed (non-zero) values.

Tests for Two Groups Assuming a Two-Part Model with Detection Limits

Power

The formula given by Chu *et al.* (2006) assumes that the group variances are equal. The formula given by Chu *et al.* (2006) is

$$\text{Power} = \Phi \left[\Phi^{-1}(\alpha/c) + \frac{|\mu_{1L} - \mu_{2L}|}{\sqrt{\frac{\sigma_{1L}^2}{N_1(1 - P_1)[1 - \Phi(\varepsilon_{1L})]} + \frac{\sigma_{2L}^2}{N_2(1 - P_2)[1 - \Phi(\varepsilon_{2L})]}} \right]$$

where c is 1 for one-sided tests and 2 for two-sided tests, $\Phi(x)$ is the CDF of the standard normal distribution, and $\varphi(x)$ is the standard normal density.

$$\begin{aligned} \mu_{iL} &= \mu_i + \sigma\varphi(\varepsilon_{iL})[1 - \Phi(\varepsilon_{iL})]^{-1} \\ \sigma_{iL} &= \sigma\{1 + \varepsilon_{iL}\varphi(\varepsilon_{iL})\}[1 - \Phi(\varepsilon_{iL})]^{-1} - \varphi^2(\varepsilon_{iL})[1 - \Phi(\varepsilon_{iL})]^{-2} \end{aligned}$$

The sample size can be determined using a simple search algorithm

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* and *Sample Size*.

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

Test Direction

Alternative Hypothesis

Specify whether the alternative hypothesis of the test is one-sided or two-sided. If a one-sided test is chosen, the hypothesis test direction is chosen based on whether the difference ($\delta = \mu_1 - \mu_2$) is greater than or less than zero.

Note that this parameter impacts the value of alpha. The value of alpha is used directly for one-sided tests. For two-sided tests, alpha is replaced by alpha/2.

Two-Sided Hypothesis Test

H0: $\mu_1 - \mu_2 = 0$ vs. H1: $\mu_1 - \mu_2 \neq 0$

One-Sided Hypothesis Tests

Upper: H0: $\mu_1 - \mu_2 \leq 0$ vs. H1: $\mu_1 - \mu_2 > 0$

Lower: H0: $\mu_1 - \mu_2 \geq 0$ vs. H1: $\mu_1 - \mu_2 < 0$.

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 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 and 0.025 for one-sided tests.

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.

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.

Tests for Two Groups Assuming a Two-Part Model with Detection Limits

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 ≥ 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 ≥ 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 $\geq Y$.

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 $N2$.

Tests for Two Groups Assuming a Two-Part Model with Detection Limits

- **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 ≥ 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 ≥ 2 . You can enter a single value or a series of values.

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 ≥ 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.

Tests for Two Groups Assuming a Two-Part Model with Detection Limits

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.

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 should not account for values that are below the detection limit. 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 should not account for values that are below the detection limit. 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 ($X > 0$)

$\mu 1$ (Mean of Group 1)

Enter one or more values for the mean of group 1 at which the design is powered. If a transformation (such as logarithm) has been applied, this is the mean of the transformed data.

Note that this is the mean of the continuous part of the data.

$\mu 1$ is unbounded. The only restriction is that $|\mu 1 - \mu 2| > 0$, that is, $\mu 1 \neq \mu 2$.

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.

$\mu 2$ (Mean of Group 2)

Enter one or more values for the mean of group 2 at which the design is powered. If a transformation (such as logarithm) has been applied, this is the mean of the transformed data.

Note that this is the mean of the continuous part of the data.

$\mu 2$ is unbounded. The only restriction is that $|\mu 1 - \mu 2| > 0$, that is, $\mu 1 \neq \mu 2$.

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

σ (Standard Deviation of X)

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

σ 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 ' σ ' button to the right to obtain options for estimating the standard deviation.

Example 1 – Finding the Sample Size

Researchers wish to compare the average medical expenditure by patients in two well-defined groups. Group 1 will represent the treatment group and group 2 will represent the control group. Since medical expenses are often zero, a two-part model is used. The logarithm base 10 of the expense values will be used to normalize the data values and to make the two group variances nearly equal. Subjects with expenses less than 50 will not be recorded as such. Hence, the detection limit is set to $\log(50) = 1.7$.

The researchers want to determine how many participants are needed if μ_1 is set to 3.45, μ_2 to 2, the standard deviation to 2 or 2.5, $P_1 = P_2 = 0.80$, the significance level of a two-sided test to 0.05, and the desired power to 0.8.

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 with Detection Limits** 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>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided
Power	0.80
Alpha	0.05
Group Allocation	Equal (N1 = N2)
P1 (Prob X = 0 in Group 1)	0.8
P2 (Prob X = 0 in Group 2)	0.8
μ_1 (Mean of Group 1)	3.45
μ_2 (Mean of Group 2)	2
σ (Standard Deviation of X)	2.25
X_L (Detection Limit)	1.7

Annotated Output

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

Numeric Results

Numeric Results

Two-sided alternative hypothesis

Actual Power	N1	N2	N	P1	P2	μ_1	μ_2	X_L	σ	Alpha
0.8001	1052	1052	2104	0.80	0.80	3.00	2.00	1.70	2.00	0.050
0.8001	1798	1798	3596	0.80	0.80	3.00	2.00	1.70	2.50	0.050
0.8010	207	207	414	0.80	0.80	4.00	2.00	1.70	2.00	0.050
0.8003	365	365	730	0.80	0.80	4.00	2.00	1.70	2.50	0.050
0.8037	76	76	152	0.80	0.80	5.00	2.00	1.70	2.00	0.050
0.8011	136	136	272	0.80	0.80	5.00	2.00	1.70	2.50	0.050

References

Chu, H., Nie, L., Cole, S.R. 2006. 'Sample size and statistical power assessing the effect of interventions in the context of mixture distributions with detection limits'. *Statistics in Medicine*, Vol. 25, Pages 2647-2657.

Tests for Two Groups Assuming a Two-Part Model with Detection Limits

Report Definitions

Actual Power is the actual power achieved. Because N_1 and N_2 are discrete, this value is usually slightly larger than the target power.

N_1 is the sample size from group 1.

N_2 is the sample size from group 2.

N is the total sample size from both groups. $N = N_1 + N_2$.

P_1 is the probability that $X = 0$ in group 1.

P_2 is the probability that $X = 0$ in group 2.

μ_1 is the mean of the measured data in group 1.

μ_2 is the mean of the measured data in group 2.

X_L is lower detection limit. Values of X below this amount cannot be detected by the measuring device.

σ is the assumed population standard deviation for each of the two groups.

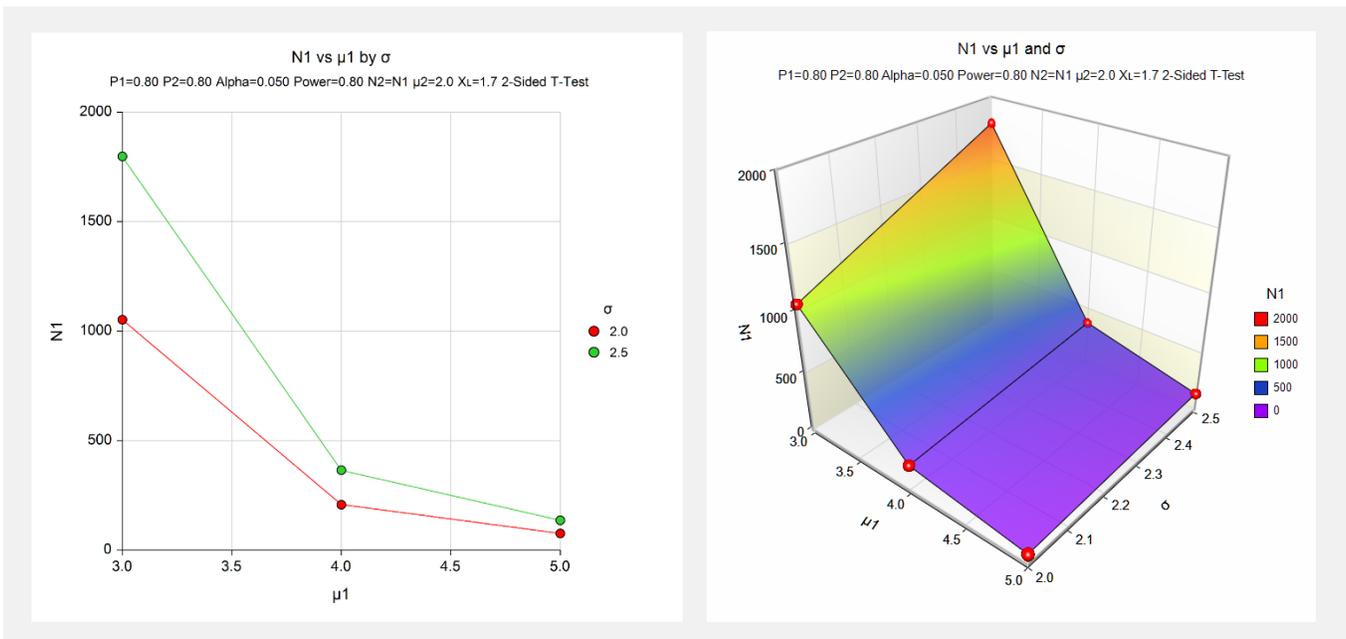
Alpha is the probability of rejecting a true null hypothesis, H_0 .

Summary Statements

A study is being conducted to compare the means of two normally distributed populations. Besides the continuous data, there are two additional types of data that may be found. First, there is a probability mass at a certain constant (usually zero). Second, some data values may be left-censored (unobserved) because they are below a detection limit. The test statistic is the difference between the means of the detected data. Group sample sizes of 1052 and 1052 achieve 80% power to reject the null hypothesis of mean equality at a two-sided significance level of 0.050. The proportion of 0's in group 1 is 0.80 and in group 2 is 0.80. The mean of group 1 of the observed data is 3.00. The mean of group 2 of the observed data is 2.00. The standard deviation of the observed values is 2.00. The detection limit, below which the data values are not observed, is 1.70.

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

Plots Section



These plots show the relationship of the mean, the standard deviation, and the sample size.

Example 2 – Validation using Chu *et al.* (2006)

Chu (2006) gives an example of calculated values on page 2652 of their article. The researchers want to determine how many participants are needed if μ_1 is set to 1.8, μ_2 to 2.4, the standard deviation to 1, $P_1 = P_2 = 0.20$, the significance level of a two-sided test to 0.05, and the desired power to 0.8. If X_L is set to 0, they concluded $N_1 = 61$ and $N_2 = 63$. If X_L is set to $\log(0.5) = -0.69$, they concluded that $N_1 = 56$ and $N_2 = 57$.

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 with Detection Limits** 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>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided
Power	0.80
Alpha	0.05
Group Allocation	Equal (N1 = N2)
P1 (Prob X = 0 in Group 1)	0.2
P2 (Prob X = 0 in Group 2)	0.2
μ_1 (Mean of Group 1)	1.8
μ_2 (Mean of Group 2)	2.4
σ (Standard Deviation of X)	1
X_L (Detection Limit)	0 -0.69

Output

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

Numeric Results

Numeric Results										
Two-sided alternative hypothesis										
Actual Power	N1	N2	N	P1	P2	μ_1	μ_2	X_L	σ	Alpha
0.8008	56	56	112	0.20	0.20	1.80	2.40	-0.69	1.00	0.050
0.8036	62	62	124	0.20	0.20	1.80	2.40	0.00	1.00	0.050

In both cases, **PASS** has obtained a sample size within 1 subject of the article. Note that the sample sizes determined in the article were based on slightly different search algorithms.