

Chapter 600

Hotelling's Two-Sample T^2

Introduction

This module calculates power for the Hotelling's two-group, T-squared (T^2) test statistic. Hotelling's T^2 is an extension of the univariate two-sample t-test to the case where the number of response variables is greater than one. These results may also be obtained using PASS's MANOVA test.

Assumptions

The following assumptions are made when using Hotelling's T^2 to analyze two groups of data.

1. The response variables are continuous.
2. The residuals follow the multivariate normal probability distribution with mean zero and constant variance-covariance matrix.
3. The subjects are independent.

Technical Details

The formulas used to perform a Hotelling's T^2 power analysis provide exact answers if the above assumptions are met. These formulas can be found in many places. We use the results in Rencher (1998). We refer you to that reference for more details.

Two-Group Technical Details

In the two-group case, sets of N_1 observations from group 1 and N_2 observations from group 2 are available on p response variables. We assume that all observations have the multivariate normal distribution with common variance covariance matrix Σ . The mean vectors of the two groups are assumed to be μ_1 and μ_2 under the alternative hypothesis. Under the null hypothesis, these mean vectors are assumed to be equal.

The value of T^2 is computed using the formula

$$T_{p, N_1+N_2-2}^2 = \frac{N_1 N_2}{N_1 + N_2} (\bar{y}_1 - \bar{y}_2)' S_{sp}^{-1} (\bar{y}_1 - \bar{y}_2)$$

where \bar{y}_1 and \bar{y}_2 are the vectors sample mean vectors of the two groups and S_{sp} is the pooled sample variance-covariance matrix.

To calculate power we need the non-centrality parameter for this distribution. This non-centrality parameter is defined as follows

Hotelling's Two-Sample T²

$$\begin{aligned}\lambda &= \frac{N1N2}{N1 + N2} (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) \\ &= \frac{N1N2}{N1 + N2} \Delta^2\end{aligned}$$

where

$$\Delta = \sqrt{(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)}$$

We define Δ as *effect size* because it provides an expression for the magnitude of the standardized difference between the null and alternative means.

Using this non-centrality parameter, the power of the Hotelling's T² may be calculated for any value of the means and standard deviations. Since there is a simple relationship between the non-central T² and the non-central F , calculations are actually based on the non-central F using the formula

$$\beta = \Pr(F' < F'_{\alpha, df1, df2, \lambda})$$

where

$$df1 = p$$

$$df2 = N1 + N2 - p - 1$$

Procedure Options

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

Design Tab

The Design tab contains many of the options that you will be primarily concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for.

When you choose to solve for *Sample Size*, the program searches for the lowest sample size that meets the alpha and power criterion you have specified.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Hotelling's Two-Sample T²**Alpha**

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

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 $N2$, 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.

 $N2$ (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 $\geq Y$.

Hotelling's Two-Sample T²

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 ≥ 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.

Hotelling's Two-Sample T²

N_2 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 N_1 and R , where $N_2 = R * N_1$."*

R is the ratio of N_2 to N_1 . That is,

$$R = N_2/N_1$$

Use this value to obtain N_2 as a multiple (or proportion) of N_1 .

N_2 is calculated from N_1 using the formula:

$$N_2 = [R \times N_1],$$

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 N_2 will be less than N_1 ; if $R > 1$, then N_2 will be greater than N_1 . 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 N_1 and N_2 .

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.

Effect Size – Response Variables

Number of Response Variables

Enter the number of response (dependent or Y) variables. For a true multivariate test, this value will be greater than one.

The number of mean differences entered in the Mean Differences box or in the Means column must equal this value. If you read-in the covariance matrix from the spreadsheet, the number of columns specified must equal this value.

Effect Size – Mean Differences

Mean Differences (= # of Response Vars)

Enter a list of values representing the mean differences under the alternative hypothesis. Under the null hypothesis, these values are all zero. The values entered here represent the differences that you want the experiment (study) to be able to detect.

Note that the number of values must match the number of Response Variables.

Hotelling's Two-Sample T^2

If you like, you can enter these values in a column on the spreadsheet. This column is specified using the 'Means Column' option. When that option is specified, any values entered here are ignored.

Means Differences Column

Use this option to specify the spreadsheet column containing the hypothesized mean differences. The response variables are represented down the rows. The number of rows with data must equal the number of response variables. When this option is used, the 'Mean Differences' box is ignored.

You can obtain the spreadsheet by selecting 'Window', then 'Data', from the menus.

Effect Size – Mean Multiplier

K (Means Multipliers)

These values are multiplied times the mean differences to give you various effect sizes. A separate power calculation is generated for each value of K. If you want to ignore this setting, enter '1'.

Covariance Tab

This tab specifies the covariance matrix.

Covariance Matrix Specification

Specify Which Covariance Matrix Input Method to Use

This option specifies which method will be used to define the covariance matrix.

- **Standard Deviation and Correlation**

This option generates a covariance matrix based on the settings for the standard deviation (SD) and the pattern of correlations as specified in the Correlation Pattern and R options.

- **Covariance Matrix Variables**

When this option is selected, the covariance matrix is read in from the columns of the spreadsheet. This is the most flexible method, but specifying a covariance matrix is tedious. You will usually only use this method when a specific covariance is given to you.

Note that the spreadsheet is shown by selecting the menus: 'Window' and then 'Data'.

Covariance Matrix Specification- Input Method = 'Standard Deviation and Correlation'

The parameters in this section provide a flexible way to specify Σ , the covariance matrix. Because the covariance matrix is symmetric, it can be represented as

Hotelling's Two-Sample T^2

$$\begin{aligned} \Sigma &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_p\rho_{1p} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \cdots & \sigma_2\sigma_p\rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_p\rho_{1p} & \sigma_2\sigma_p\rho_{2p} & \cdots & \sigma_p^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix} \end{aligned}$$

where p is the number of response variables.

Thus, the covariance matrix can be represented with complete generality by specifying the standard deviations $\sigma_1, \sigma_2, \dots, \sigma_p$ and the correlation matrix

$$R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}.$$

SD (Common Standard Deviation)

This value is used to generate the covariance matrix. This option specifies a single standard deviation to be used for all response variables. The square of this value becomes the diagonal elements of the covariance matrix. Since this is a standard deviation, it must be greater than zero.

This option is only used when the first Covariance Matrix Input Method is selected.

R (Correlation)

Specify a correlation to be used in calculating the off-diagonal elements of the covariance matrix. Since this is a correlation, it must be between -1 and 1. This option is only used when the first Covariance Matrix Input Method is selected.

Specify Correlation Pattern

This option specifies the pattern of the correlations in the variance-covariance matrix. Two options are available:

Hotelling's Two-Sample T^2

- **Constant**

The value of R is used as the constant correlation. For example, if $R = 0.6$ and $p = 6$, the correlation matrix would appear as

$$R = \begin{bmatrix} 1 & 0.600 & 0.600 & 0.600 & 0.600 & 0.600 \\ 0.600 & 1 & 0.600 & 0.600 & 0.600 & 0.600 \\ 0.600 & 0.600 & 1 & 0.600 & 0.600 & 0.600 \\ 0.600 & 0.600 & 0.600 & 1 & 0.600 & 0.600 \\ 0.600 & 0.600 & 0.600 & 0.600 & 1 & 0.600 \\ 0.600 & 0.600 & 0.600 & 0.600 & 0.600 & 1 \end{bmatrix}$$

- **1st-Order Autocorrelation**

The value of R is used as the base autocorrelation in a first-order, serial correlation pattern. For example, $R = 0.6$ and $p = 6$, the correlation matrix would appear as

$$R = \begin{bmatrix} 1 & 0.600 & 0.360 & 0.216 & 0.130 & 0.078 \\ 0.600 & 1 & 0.600 & 0.360 & 0.216 & 0.130 \\ 0.360 & 0.600 & 1 & 0.600 & 0.360 & 0.216 \\ 0.216 & 0.360 & 0.600 & 1 & 0.600 & 0.360 \\ 0.130 & 0.216 & 0.360 & 0.600 & 1 & 0.600 \\ 0.078 & 0.130 & 0.216 & 0.360 & 0.600 & 1 \end{bmatrix}$$

This pattern is often chosen as the most realistic when little is known about the correlation pattern and the responses variables are measured across time.

Covariance Matrix Specification- Input Method = 'Covariance Matrix Variables'

This option instructs the program to read the covariance matrix from the spreadsheet.

Spreadsheet Columns Containing the Covariance Matrix

This option designates the columns on the current spreadsheet holding the covariance matrix. It is used when the 'Specify Which Covariance Matrix Input Method to Use' option is set to *Covariance Matrix Variables*. The number of columns and number of rows must match the number of response variable at which the subjects are measured.

Hotelling's Two-Sample T²**Example 1 – Power and Validation**

Rencher (1998) pages 107-108 presents an example of power calculations for the two-group case in which the mean differences and covariance matrix are

$$\mu_1 - \mu_2 = \begin{bmatrix} 3 \\ -2 \\ 3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 6 & -3 & 3 \\ -3 & 5 & -6 \\ 3 & -6 & 9 \end{bmatrix}$$

When $N_1 = N_2 = 10, 12, 14, 16$ and the significance level is 0.05, Rencher calculated the power to be 0.6438, 0.7520, 0.8329, 0.8936, respectively.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Hotelling's Two-Sample T2** procedure window by expanding **Means**, then clicking on **Multivariate Means**, and then clicking on **Hotelling's Two-Sample T2**. 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**. You can see that the values have been loaded into the spreadsheet by clicking on the spreadsheet button.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
Sample Size Per Group.....	10 12 14 16
Number of Response Variables.....	3
Mean Differences.....	<i>blank</i>
Mean Differences Column	Differences
K (Means Multiplier).....	1.0

Covariance Tab

Specify Covariance Method.....	Covariance Matrix Columns
Spreadsheet Columns	VC_1-VC_3

Press the Spreadsheet button to enter the following values into the spreadsheet for columns VC_1 through VC_3

Row 1.....	6	-3	3
Row 2.....	-3	5	-6
Row 3.....	3	-6	9

Reports Tab

Show Numeric Results	Checked
Show Means Matrix	Checked
Show Covariance Matrix.....	Checked

Hotelling's Two-Sample T^2 **Output**

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

Numeric Report

Power	N1	N2	N	Multiply Means By (K)	Alpha	Effect Size	# of Y's (DF1)	DF2
0.64423	10	10	20	1.0000	0.050	1.41	3	16
0.75459	12	12	24	1.0000	0.050	1.41	3	20
0.83613	14	14	28	1.0000	0.050	1.41	3	24
0.89360	16	16	32	1.0000	0.050	1.41	3	28

Note that the power values obtained here are very close to those obtained by Rencher. We feel that our results are more accurate since Rencher's results were obtained by interpolation from Tang's tables.

Means Section

Means Section	
Name	Mean
Y1	3.0000
Y2	-2.0000
Y3	3.0000

This report shows the mean differences that were read in.

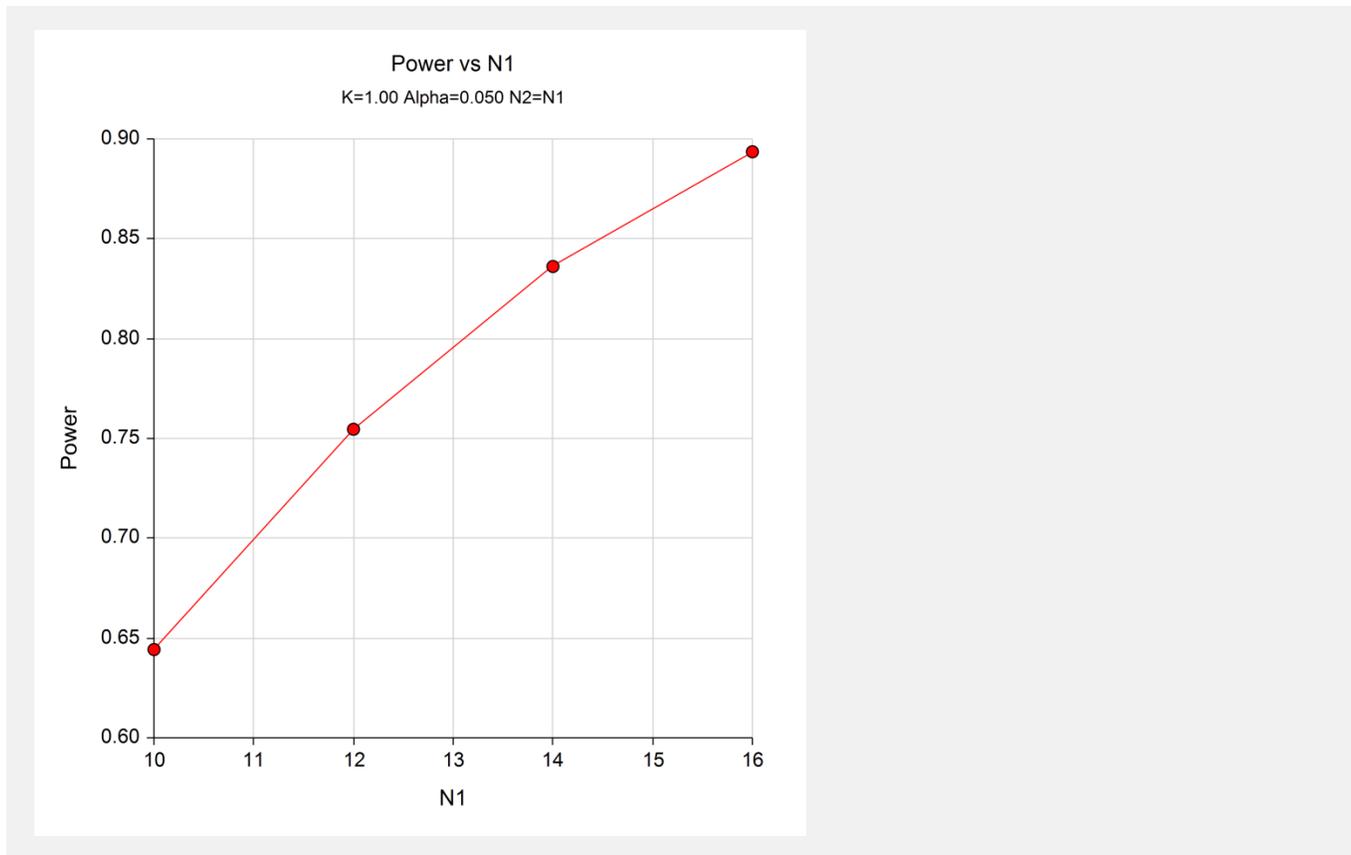
Variance-Covariance Matrix Section

Variance-Covariance Matrix Section			
Response	Y1	Y2	Y3
Y1	2.4495	-0.5477	0.4082
Y2	-0.5477	2.2361	-0.8944
Y3	0.4082	-0.8944	3.0000

SD's on diagonal. Correlations off diagonal.

This report shows the variance-covariance matrix that was read in from the spreadsheet or generated by the settings of on the Covariance tab. The standard deviations are given on the diagonal and the correlations are given off the diagonal.

Chart Section



This chart shows the relationship between power and $N1$.