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.

Test Statistic

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 null hypothesis, these mean vectors are assumed to be equal.

The value of T^2 is computed using the formula

$$T^2_{p,N_1+N_2-2} = \frac{N_1 N_2}{N_1+N_2}(\bar{y}_1 - \bar{y}_2)' S_{pool}^{-1}(\bar{y}_1 - \bar{y}_2)$$

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

Power Calculation

To calculate power, we need the non-centrality parameter for this distribution. With $\delta = \mu_1 - \mu_2$ as a vector of actual mean differences, this non-centrality parameter is defined as follows

$$\lambda = \frac{N_1 N_2}{N_1 + N_2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$$

$$= \frac{N_1 N_2}{N_1 + N_2}\delta'\Sigma^{-1}\delta$$

$$= \frac{N_1 N_2}{N_1 + N_2}\Delta^2$$

where

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

$$= \sqrt{\delta'\Sigma^{-1}\delta}$$

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

Using this non-centrality parameter, the power of the Hotelling's T^2 may be calculated for any values of the means and standard deviations. Since there is a simple relationship between the non-central T^2 and the non-central *F*, calculations are actually based on the non-central *F* using the formula

$$\text{Power} = 1 - \Pr\left(F' < F'_{\alpha, df_1, df_2, \lambda}\right)$$

where

$$df_1 = p$$

$$df_2 = N_1 + N_2 - p - 1$$

Variance-Covariance Matrix – Standard Deviation

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

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_p\rho_{1p} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \cdots & \sigma_2\sigma_p\rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_p\sigma_1\rho_{p1} & \sigma_p\sigma_2\rho_{p2} & \cdots & \sigma_p^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}.$$

where p is the number of response variables.

Thus, the covariance matrix can be represented with complete generality by specifying the standard deviations σ_1, σ_2, ..., σ_p and the correlation matrix

$$R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}.$$

Example 1 – Power and Validation using Rencher (1998)

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

$$\mu_1 - \mu_2 = \begin{bmatrix} 3 \\ -2 \\ 3 \end{bmatrix}$$

and the variance-covariance matrix is

$$\Sigma = \begin{bmatrix} 6 & -3 & 3 \\ -3 & 5 & -6 \\ 3 & -6 & 9 \end{bmatrix}$$

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

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 .... **Power**
Alpha.. **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **10 12 14 16**
Number of Response Variables **3**
Mean Difference Input Type.......................... **List of Mean Differences**
List of Mean Differences .... **3 -2 3**
K (Means Multiplier)....................................... **1**

Covariance Tab

Variance-Covariance Matrix Input Type......... **Variance-Covariance Matrix in Spreadsheet**
Columns Containing the V-C Matrix.............. **VC1-VC3**

Input Spreadsheet Data

Row	Diffs	VC1	VC2	VC3
1	3	6	-3	3
2	-2	-3	5	-6
3	3	3	-6	9

Output

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

Numeric Reports

Numeric Results

Solve For:	Power
Number of Response Variables:	3
Mean Differences:	3, -2, 3
Variance-Covariance Matrix:	Input Columns = VC1-VC3

	Sample Size			Means Multiplier	Degrees of Freedom			
Power	N1	N2	N	K	Number of Y's DF1	DF2	Effect Size	Alpha
0.64423	10	10	20	1	3	16	1.414	0.05
0.75459	12	12	24	1	3	20	1.414	0.05
0.83613	14	14	28	1	3	24	1.414	0.05
0.89360	16	16	32	1	3	28	1.414	0.05

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N1 and N2	The number of items sampled from each population.
N	The total sample size. N = N1 + N2.
K	A constant by which all means are multiplied.
DF1	The first degrees of freedom for T^2. It is the number of response variables.
DF2	The second degrees of freedom for T^2.
Effect Size	A standardized version for T^2 under the alternative hypothesis.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A two-group design with 3 response variables will be used to test whether there is a difference between the two multivariate means. The comparison will be made using a two-sample Hotelling's T^2 test with a Type I error rate (α) of 0.05. The variance-covariance matrix is assumed to be defined as 'Input Columns = VC1-VC3'. To detect an effect size of 1.414 (derived from the mean differences '3, -2, 3' and the variance-covariance matrix), with sample sizes of 10 in Group 1 and 10 in Group 2, the power is 0.64423.

.
.
.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N1	N2	N	Dropout-Inflated Enrollment Sample Size N1'	N2'	N'	Expected Number of Dropouts D1	D2	D
20%	10	10	20	13	13	26	3	3	6
20%	12	12	24	15	15	30	3	3	6
20%	14	14	28	18	18	36	4	4	8
20%	16	16	32	20	20	40	4	4	8

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 (as entered by the user). 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. 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, 13 subjects should be enrolled in Group 1, and 13 in Group 2, to obtain final group sample sizes of 10 and 10, respectively.

References

Rencher, Alvin C. 1998. Multivariate Statistical Inference and Applications. John Wiley. New York, New York.

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 Matrix

Means Matrix

Name	Mean
Y1	3
Y2	-2
Y3	3

This report shows the mean differences that were read in.

Variance-Covariance Matrix

Variance-Covariance Matrix

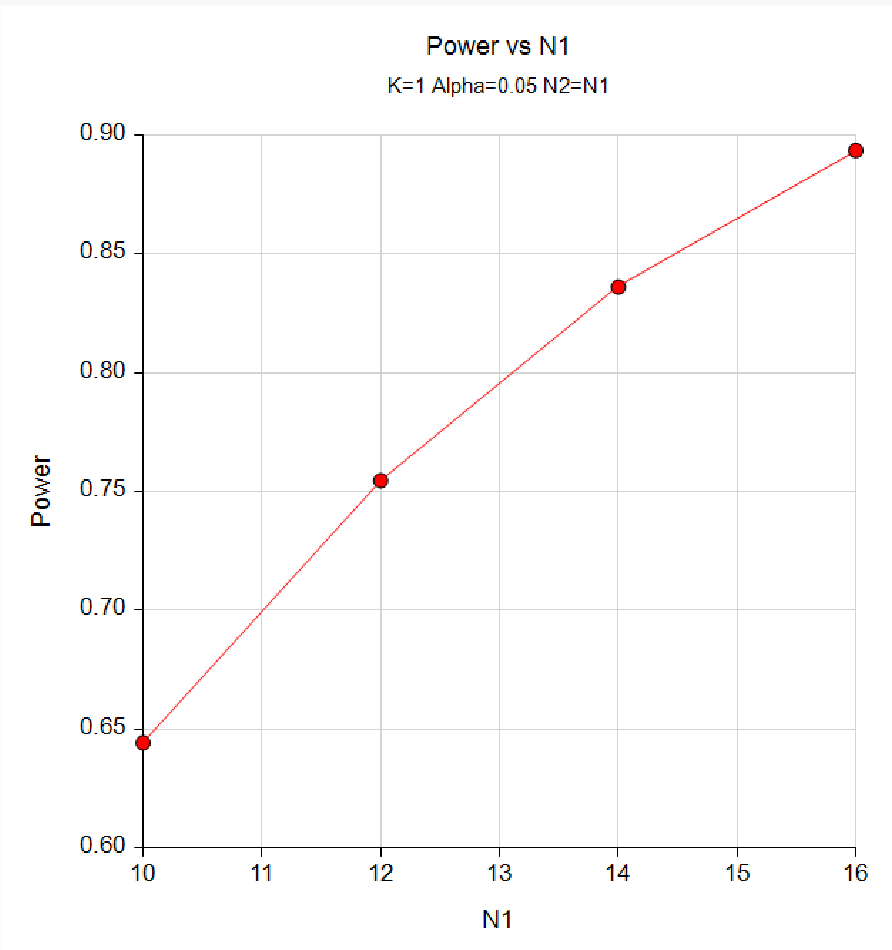
Response	Y1	Y2	Y3
Y1	2.449	-0.548	0.408
Y2	-0.548	2.236	-0.894
Y3	0.408	-0.894	3.000

σ's on the diagonal, ρ's on the off diagonal(s)

This report shows a modified variance-covariance matrix with σ's on the diagonal, ρ's on the off diagonal(s). This matrix is created from the variance-covariance matrix that was read in from the spreadsheet or generated by the σ and ρ settings on the Covariance tab.

Plots Section

Plots



This plot shows the relationship between *Power* and *N1*.

The very same report can be generated by inputting the mean differences from the spreadsheet (**Example 1b**):

Design Tab

Mean Difference Input Type.......................... **Mean Differences in Spreadsheet**
Column Containing the Mean Differences ..... **DIFFS**

Input Spreadsheet Data

Row	Diffs	VC1	VC2	VC3
1	3	6	-3	3
2	-2	-3	5	-6
3	3	3	-6	9

Numeric Reports

Numeric Results

Solve For: Power
Number of Response Variables: 3
Mean Differences: 3, -2, 3
Variance-Covariance Matrix: Input Columns = VC1-VC3

	Sample Size				Degrees of Freedom			
Power	N1	N2	N	Means Multiplier K	Number of Y's DF1	DF2	Effect Size	Alpha
0.64423	10	10	20	1	3	16	1.414	0.05
0.75459	12	12	24	1	3	20	1.414	0.05
0.83613	14	14	28	1	3	24	1.414	0.05
0.89360	16	16	32	1	3	28	1.414	0.05