

Chapter 410

Hotelling's Two-Sample T^2

Introduction

The two-sample Hotelling's T^2 is the multivariate extension of the common two-group Student's t -test. In a t -test, differences in the mean response between two populations are studied. T^2 is used when the number of response variables are two or more, although it can be used when there is only one response variable. The null hypothesis is that the group means for all response variables are equal.

T^2 makes the usual assumptions of equal variances and normally distributed residuals. Preliminary tests are provided that allow these assumptions to be evaluated. *Randomization tests* are provided that do not rely on these assumptions. These randomization tests should be used whenever you want exact results that do not rely on several assumptions.

Technical Details

Equal Covariance Case

The two-sample T^2 is used to test the equality of the mean vectors of two populations. Specifically, suppose a set of p response variables Y_1, Y_2, \dots, Y_p is measured for each of two groups. Assume that population 1 is distributed as $N_p(\mu_1, \Sigma_1)$ and population 2 is distributed as $N_p(\mu_2, \Sigma_2)$, where $N_p(\mu, \Sigma)$ is the p -variable multivariate normal distribution with mean vector μ and covariance matrix Σ . The null hypothesis that $\mu_1 = \mu_2$ can be tested using the test statistic

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_1 - \bar{y}_2)' S_{pl}^{-1} (\bar{y}_1 - \bar{y}_2)$$

where \bar{y}_1 and \bar{y}_2 are the two sample mean vectors, n_1 and n_2 are the two sample sizes, and S_{pl}^{-1} is the inverse of the pooled covariance matrix which is calculated using

$$S_{pl} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

Here, S_1 and S_2 are the estimated covariance matrices calculated from the two samples.

If we make the additional assumption that $\Sigma_1 = \Sigma_2$, T^2 follows Hotelling's T -squared distribution when the null hypothesis is true. That is, $T^2 \sim T_{p, n_1 + n_2 - 2}^2$. Reject the null hypothesis if $T^2 \geq T_{1-\alpha, p, n_1 + n_2 - 2}^2$. Note that rejecting the null hypothesis concludes that at least one pair of the p sets of group response means are unequal.

Unequal Covariance Case

When the experimental setting or a preliminary test such as Box's M test leads us to conclude that $\Sigma_1 \neq \Sigma_2$, an alternative to T^2 must be used. Several such multivariate Behrens-Fisher tests have been suggested in the statistical literature. Following the suggestions of Rencher (1998) derived from a large simulation study, we use the procedure suggested by Nel and van der Merwe (1986) since it was shown to have near optimal power while maintaining reasonable type-I error rates. The test statistic is computed using the formula

$$T^{2*} = (\bar{y}_1 - \bar{y}_2)' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2} \right)^{-1} (\bar{y}_1 - \bar{y}_2)$$

T^{2*} is approximately distributed as $T_{p,v}^2$ where v is given by

$$v = \frac{\text{tr} \left[\left(\frac{S_1}{n_1} + \frac{S_2}{n_2} \right) \left(\frac{S_1}{n_1} + \frac{S_2}{n_2} \right) \right] + \left[\text{tr} \left(\frac{S_1}{n_1} + \frac{S_2}{n_2} \right) \right]^2}{\frac{\text{tr} \left[\left(\frac{S_1}{n_1} \right) \left(\frac{S_1}{n_1} \right) \right] + \left[\text{tr} \left(\frac{S_1}{n_1} \right) \right]^2}{n_1 - 1} + \frac{\text{tr} \left[\left(\frac{S_2}{n_2} \right) \left(\frac{S_2}{n_2} \right) \right] + \left[\text{tr} \left(\frac{S_2}{n_2} \right) \right]^2}{n_2 - 1}}$$

Randomization Test

Because of the stringent assumptions that must be made when using this procedure, **NCSS** also includes a randomization test as outlined by Edgington (1987). Randomization tests are becoming more and more popular as the speed of computers allows them to be computed in seconds rather than hours.

A randomization test is conducted by enumerating all possible permutations of the sample data, calculating the test statistic for each permutation, and counting the number of permutations that result in a T^2 value greater than or equal to the actual T^2 value. Dividing this count by the number of permutations tried gives the significance level of the test.

For even moderate sample sizes, the total number of permutations is in the trillions, so a Monte Carlo approach is used in which the permutations are found by random selection rather than complete enumeration. Edgington suggests that at least 1,000 permutations be selected. We suggest that this be increased to 10,000.

Permutation results are provided for the equal covariance case, the unequal covariance case, and for all individual t tests.

Assumptions

The following assumptions are made when using T^2 .

1. Each population follows the multivariate normal distribution.
2. The two samples are independent.
3. The two covariance matrices are equal.

These are a set of restrictive assumptions that must be evaluated for each set of data. Box's M test may be used to test whether two covariances matrices are equal. Unfortunately, the accuracy of Box's M test is very sensitive to departures from multivariate normality (assumption 1).

Data Structure

The data must be entered in a format that places the response variables and values for the group side by side. An example of the data structure for a Hotelling's T^2 design is shown below. In this example, *WRATR* and *WRATA* are the two response variables. *Treatment* is the group variable. Note that this database has a fourth variable, *Disability*, which is ignored in this analysis. This database is stored in the file MANOVA1.

MANOVA1 Dataset

WRATR	WRATA	Treatment	Disability
115	108	1	1
98	105	1	1
107	98	1	1
90	92	2	1
85	95	2	1
80	81	2	1
100	105	1	2
105	95	1	2
95	98	1	2
70	80	2	2
85	68	2	2
78	82	2	2
89	78	1	3
100	85	1	3
90	95	1	3
65	62	2	3
80	70	2	3

Example 1 – Hotelling's Two-Sample T^2 Test

This section presents an example of how to run an analysis of the MANOVA1 data shown earlier. In this analysis, two groups are to be compared on two variables: WRATR and WRATA.

Setup

To run this example, complete the following steps:

1 Open the MANOVA1 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **MANOVA1** and click **OK**.

2 Specify the Hotelling's Two-Sample T^2 procedure options

- Find and open the **Hotelling's Two-Sample T^2** procedure using the menus or the Procedure Navigator.
- The settings 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.

Variables Tab

Response Variables.....	WRATR-WRATA
Group Variable.....	Treatment
Run Randomization Tests	Checked
Monte Carlo Samples	1000
Random Seed.....	3183150 (for reproducibility)

Reports Tab

All Reports	Checked
-------------------	----------------

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Descriptive Statistics

Descriptive Statistics

Variable	Means		Standard Deviations	
	1	2	1	2
WRATR	99.8889	78.3333	8.2832	8.0467
WRATA	96.3333	78.1111	9.7468	10.9481
Count	9	9	9	9

This report provides the means, standard deviations, and counts of each group. Look the values over to be certain that the right variables were selected.

Hotelling's T^2 Tests

Hotelling's T^2 Tests

Covariance Assumption	T^2	Degrees of Freedom		P-Value	
		DF1	DF2	Parametric Test	Randomization Test
Equal	31.64	2	16.0	0.0003	0.0020
Unequal	31.64	2	15.9	0.0003	0.0020

The randomization test results are based on 1000 Monte Carlo samples with a random seed of 3183150.

This report gives the results of the two T^2 tests: first for equal, and then for unequal, covariance matrices.

Covariance Assumption

Indicates the type of T^2 test displayed on this line. Either *Equal* or *Unequal* are possible. Even though you actually use only one of these two tests, both are reported here.

T^2

The values of the two test statistics are shown in this column. The top value is T^2 . The bottom value is T^{2*} . Note that these values will be equal if the two sample sizes are equal.

DF1

This is the number of response variables.

DF2

For the top line, this is $n_1 + n_2 - 2$. For the bottom line, this is the value of v .

Parametric Test P-Value

This is the p -value of the T^2 test. If this value is less than 0.05, we say that the test was significant at the 0.05 level. If the value is less than 0.01, we say that the test was significant at the 0.01 level. This result is accurate if all the assumptions of the corresponding test are met.

Hotelling's Two-Sample T^2

Randomization Test P-Value

This is the p -value of the randomization test. If this value is less than 0.05, we say that the test was significant at the 0.05 level. If the value is less than 0.01, we say that the test was significant at the 0.01 level.

This result is an exact result even if the data were not obtained by random sampling. The result is accurate even if the response variables were binary.

Individual Variable Influence Tests

Individual Variable Influence Tests

Variable Omitted	Others		Change		Alone	
	T^2	P-Value	T^2	P-Value	T^2	P-Value
WRATR	13.909	0.0018	17.731	0.0093	31.357	0.0000
WRATA	31.357	0.0000	0.283	0.7688	13.909	0.0018

This report provides information about the influence of each of the individual response variables on the overall T^2 value. This is accomplished by calculating the change in T^2 when a response variable is omitted.

Variable Omitted

This is the name of the variable shown on this line of the report.

 T^2 Others

This is the value of T^2 calculated with all response variables except the variable listed to the left.

P-Value Others

This is the p -value of the associated T^2 test.

 T^2 Change

This is the amount that T^2 is reduced when the response variable shown on this line is omitted.

P-Value Change

This is the p -value of the T^2 change test. It is computed using the fact that the change in T^2 is related to an F distribution using the formula

$$F_{\alpha,1,n_1+n_2-2-(p-1)} = \frac{T_p^2 - T_{p-1}^2}{n_1 + n_2 - 2 + T_{p-1}^2}$$

Note that this quantity tests the drop in T^2 when a variable is removed, conditional on the other response variables that are included. Another way of looking at this quantity is that it tests whether the variable omitted significantly increases the distance between the two populations. See Rencher (1998) page 109 for further details.

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

This is the value of T^2 calculated when only this response variable is used. It is the square of the common t test. It is the two-sided test of the null hypothesis that the means for this variable are equal, ignoring all other variables.

P-Value Alone

This is the p -value of the associated T^2 test.

Student's T-Tests**Student's T-Tests**

Variable	T^2 or Student's T	P-Value	
		Parametric Test	Randomization Test
All (T^2)	31.640	0.0003	0.0020
WRATR	5.600	0.0000	0.0010
WRATA	3.729	0.0018	0.0030

The randomization test results are based on 1000 Monte Carlo samples with a random seed of 3183150.
These individual t -test significance levels should only be used when the overall T^2 value is significant.

This report provides the results of individually conducting a two-sided t -test on each of the response variables. You might think that since there are a series of p t -tests being employed, a Bonferroni adjustment should be applied to the significance levels. However, if these individual tests are only considered when the overall T^2 is significant at the same level, such as 0.05, then their significance levels are "protected" by the T^2 test and the unadjusted significance levels given here can be used.

Variable

The variable whose results are presented on this line. Note that the first line gives the overall results for T^2 .

 T^2 or |Student's T|

The first line is the value of T^2 . The other lines are the two-sided Student's t -test values.

Parametric Test P-Value

These are the p -values of the test statistics given to the left. Note that if the individual tests are only used when the overall test is significant, these p -values are accurate even though several individual tests are made. The multivariate T^2 test is said to "protect" the significance levels of the individual tests.

Randomization Test P-Value

These are the results of randomization tests that are run on each of the variables. These tests are exact when the Monte Carlo sample size is large, say over 5000. These tests should be used when there is even a hint that the regular assumptions of the t -tests are not valid. For example, this p -value is accurate even when the response variable takes on binary values (the t -test assumes a continuous, normal response variable).

Hotelling's Two-Sample T²

Note that these values will change from run to run. As you increase the number of Monte Carlo samples, these values will become more and more stable. You may have to go as large as 100,000 before the results remain the same from run to run. This instability is due to the use of a random sample of all the trillions of permutations that are possible. As you increase the Monte Carlo sample size, you reduce the sampling error (and greatly increase the time it takes to generate the results).

Confidence Intervals for the Mean Differences

Confidence Intervals for the Mean Differences

Variable	Group Mean Difference	95% Bonferroni Confidence Interval Limits		95% Simultaneous Confidence Interval Limits	
		Lower	Upper	Lower	Upper
WRATR	21.5556	12.0365	31.0747	10.7665	32.3446
WRATA	18.2222	6.1396	30.3048	4.5277	31.9168

This report provides confidence intervals for the differences between the group means for each response variable. Two intervals are provided: Bonferroni and simultaneous.

Variable

The variable whose results are presented on this line.

Group Mean Difference

The actual difference between the group means for the corresponding response variable.

Bonferroni Confidence Interval Limits

Bonferroni confidence intervals are based on the formula

$$\bar{y}_{1j} - \bar{y}_{2j} \pm t_{\alpha/2p, n_1+n_2-2} \sqrt{\left(\frac{n_1 + n_2}{n_1 n_2} s_{pl,jj}\right)}$$

This formula is derived by applying a Bonferroni adjustment to the regular univariate confidence interval. This adjustment is made by dividing the alpha level by p , the number of such intervals to be created. These intervals are usually not as wide as the simultaneous intervals, yet still have an appropriate adjustment because of the multiple intervals that are being created.

Hotelling's Two-Sample T^2

Simultaneous Confidence Interval Limits

Simultaneous confidence intervals are based on the formula

$$\bar{y}_{1j} - \bar{y}_{2j} \pm \sqrt{T_{1-\alpha, p, n_1+n_2-2}^2 \left(\frac{n_1 + n_2}{n_1 n_2} s_{pl,jj} \right)}$$

This formula is derived from a formula for confidence intervals for *any* linear combination of the mean differences, including those that are generated after looking at the data. Because of this, these confidence intervals are extra wide and may not be of must use.

Bartlett-Box Homogeneity of Variance Tests

Bartlett-Box Homogeneity of Variance Tests

Variable(s) Tested	Test Value	Degrees of Freedom		Approximate F-Test		Approximate Chi-Square Test	
		DF1	DF2	F	P-Value	Chi²	P-Value
Box's M Test							
ALL	0.153	3	46080	0.044	0.9877	0.133	0.9877
Bartlett Individual Variable Tests							
WRATR	0.007	1	768	0.006	0.9367	0.006	0.9368
WRATA	0.108	1	768	0.101	0.7503	0.101	0.7505

This report provides a preliminary test of the assumption of equality of covariance matrices. If the data fails this test, you should use the unequal variance version of the T^2 test or the randomization test. The calculation of these tests is documented in the Technical Details section of the Equality of Covariance Matrices chapter and will not be repeated here.

Box's M test is very sensitive to non-normality. A significant value indicates either unequal covariance matrices, non-normality, or both. Hence, it is important to establish multivariate normality before concluding unequal covariance matrices using Box's M test.

The Chi-square approximation should be used when all group sample sizes are greater than 20 and p is less than 6. Otherwise, the F approximation is more accurate. **NCSS** supplies both the multivariate Box's M test and the individual Bartlett's tests so that when Box's M test is significant, you can determine which variables contribute to the variance.

Covariance and Correlation Matrix Determinants

Covariance and Correlation Matrix Determinants

Treatment	Log of Covariance Matrix Determinant	Correlation Matrix Determinant
1	8.3863	0.6730
2	8.4949	0.6301
All	8.4502	0.6528

This report gives the logarithm (base e) of the determinant of each of the covariance matrices and the determinant of each of the correlation matrices. The assumption of equality of covariance matrices forces us to also assume that these values are equal. Box's M test compares these values.

Eigenvalues of Covariance Matrices

Eigenvalues of Covariance Matrices

Number	Treatment		
	1	2	All
1	129.8207	152.5590	140.9319
2	33.7904	32.0521	33.1793

This report gives the eigenvalues of each of the individual covariance matrices as well as the pooled covariance matrix. Each column gives a set of eigenvalues. These eigenvalues summarize the covariance matrix into a few values. By comparing the largest eigenvalues across both groups, you can determine if the groups are different. Also, eigenvalues near zero indicate singularities in your data.

Eigenvalues of Correlation Matrices

Eigenvalues of Correlation Matrices

Number	Treatment		
	1	2	All
1	1.5718	1.6082	1.5893
2	0.4282	0.3918	0.4107

This report gives the eigenvalues of each of the individual correlation matrices followed by the eigenvalues of the within-group correlation matrix. Each column gives a set of eigenvalues.

This report is useful because the eigenvalues summarize the correlation matrix. By comparing the largest eigenvalues across all groups, you can determine if the groups are different. Also, eigenvalues near zero indicate singularities in your data.

Within Group Correlations\Covariances

Within-Group Correlation\Covariance for Treatment = All

Variable	Variable	
	WRATR	WRATA
WRATR	66.6806	49.8750
WRATA	0.5893	107.4306

Within-Group Correlation\Covariance for Treatment = 1

Variable	Variable	
	WRATR	WRATA
WRATR	68.6111	46.1667
WRATA	0.5718	95.0000

Within-Group Correlation\Covariance for Treatment = 2

Variable	Variable	
	WRATR	WRATA
WRATR	64.7500	53.5833
WRATA	0.6082	119.8611

This report displays correlations and covariances. The covariance matrices were labeled S_{sp} , S_1 , and S_2 in the formulas given earlier in this chapter. The correlations are shown in the lower-left half of the matrix and the covariances are shown on the diagonal and in the upper-right half of the matrix.