3

Hotelling's T^2

A Two-Group Multivariate Analysis

CHAPTER CONTENTS

Courtesy of the State Library of North Carolina

The Hotelling's T^2 was developed by Harold Hotelling (1895–1973) to extend the univariate *t* test with one dependent variable to a multivariate *t* test with two or more dependent variables (Hotelling, 1931). He attributes his interest in statistics to his professional relations with R. A. Fisher. He was an associate professor of mathematics at Stanford University from 1927 to 1931. He was a member of the Columbia University faculty from 1931 to 1946. While at Columbia University, he sponsored Henry Mann (nonparametric Mann–Whitney *U* statistic) and

(Continued)

(Continued)

Abraham Wald (decision theory, statistical sequential analysis) due to European anti-Semitism. Hotelling is well-known for his vision that universities should have a department of statistics. He spent much of his career as a professor of mathematical statistics at the University of North Carolina at Chapel Hill from 1946 until his death in 1973.

△ Overview

So why use more than one dependent variable? There are two main reasons. First, any treatment will affect participants in more than one way, so multiple measures on several criteria provide a more valid assessment of group differences (experimental vs. control). Second, the use of more criteria measures permits a better profile of group differences.

We can also examine the following question: "Why use a multivariate analysis rather than a univariate analysis?" There are several reasons from a statistical point of view.

First, the Type I error rate is inflated when using several univariate tests; for example, two univariate *t* tests would have a Type I error rate of (.95)(.95) = .90, so 1 − .90 = .10 (probability of falsely rejecting the null hypothesis; a Type I error rate), not the individual Type I error rate of .05. A researcher could test each univariate *t* test at the .025 level to avoid an inflated Type I error rate. This has been referred to as the Dunn–Bonferroni adjustment to the alpha level, where the alpha level is divided by the number of tests; for example, .05 divided by 2 = .025. The multivariate test could incorporate both the tests and keep the alpha level at the .05 level, thus maintaining the power for the test of group mean differences. The second reason is that the univariate test ignores covariance (correlation) among dependent variables. The separate univariate *t* tests would not include the relation among the dependent variables. Another good reason to conduct multivariate analyses is when a set of dependent variables have a theoretical basis or rationale for being together. The third reason is that a researcher may not find a single univariate mean difference between groups, but jointly, a mean difference may exist when considering the set of dependent variables. These three reasons for conducting a multivariate analysis provide a sound rationale to consider when analyzing data with multiple dependent variables.

Stevens (2009) pointed out that a researcher may not find a multivariate joint group mean difference for all dependent variables, so a researcher should check for subsets of dependent variables, which might be statistically significant. This situation may arise when a researcher uses subtest scores for the set of dependent variables, rather than using a total test score. Basically, one or more subtest mean differences may exist between the two groups, but the total test score mean is not statistically different. Similarly, two dependent variables might indicate multivariate statistical significance, but a third variable when included may suppress or negate the statistical significance of the other two variables.

△ Assumptions

When conducting the Hotelling T^2 statistic, it is important to consider the data assumptions that affect the statistical test. Four data assumptions are important to consider when computing the Hotelling T^2 test of group mean differences:

1. The data from population i are sampled from a population with mean vector $\boldsymbol{\mu}_i$.
   - This assumption implies that there are no subpopulations with different population means. A randomized experiment with subjects randomly assigned to experimental and control groups would meet this assumption.
2. The data from both populations have a common variance–covariance matrix—Σ.
   - We can test the null hypothesis that Σ_1 is equal to Σ_2 against the general alternative that they are not equal using a Box M test:

     H_0: $\Sigma_1 = \Sigma_2$

     H_A: $\Sigma_1 \neq \Sigma_2$

     Under the null hypothesis, H_0: $\Sigma_1 = \Sigma_2$, Bartlett's test statistic is approximately chi-square distributed with $P(P + 1)/2$ degrees of freedom; P = number of variables. If the Bartlett's test is statistically significant, then we reject the null hypothesis and assume that the variance–covariance matrices are different between the two groups.

3. The data values are independent.
    - The subjects from both populations are independently sampled.
    - Subjects from the two separate populations were independently randomly sampled. This does not mean that the variables are independent of one another.
    - The independence assumption is violated when using nonprobability, clustered, time series, and spatial sampled data. If data are dependent, then the results for some observations are going to be predictable from the results of other observations (linear dependency). The consequence of violating the assumption of independence is that the null hypothesis is rejected more often than if the independence assumption is met, and linear dependency results in a nonpositive definite matrix.
4. Both populations of data are multivariate normally distributed.

    We can check this using the following approaches:
    - Produce histograms for each variable to check for a symmetric distribution.
    - Produce scatter plots of variables to check for an elliptical display of points.
    - Run a Shapiro–Wilk test of multivariate normality.

Notes:

- The *central limit theorem* states that the dependent variable sample means are going to be approximately multivariate normally distributed regardless of the distribution of the original variables.
- Hotelling's T^2 test is robust to violations of assumptions of multivariate normality; however, the Box M test should not be used if data are not multivariate normally distributed.
- Hotelling's T^2 test is sensitive to violations of the assumption of equal variance–covariance matrices, especially when sample sizes are unequal, that is, $n_1 \neq n_2$. If the sample sizes are equal, the Hotelling's T^2 test is more robust.

△ Univariate Versus Multivariate Hypothesis

The expression of the univariate and multivariate hypotheses shows the extension of the univariate *t* test with a single dependent variable to the multivariate *t*-test case with multiple dependent variables. Instead of a single comparison of means between two groups, we express multiple dependent variable means for each group in a matrix vector. The univariate null hypothesis is expressed as follows:

$$\mathrm{H}_0\colon \mu_1 = \mu_2,$$

and the univariate t test is computed as follows:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}.$$

When the denominator of the formula is expressed as a pooled estimate of the common population variance for the two groups, squaring both sides reduces the formula to

$$t^2 = \frac{(\bar{y}_1 - \bar{y}_2)^2}{s^2_{\text{pooled}}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)},$$

which can be expressed as follows:

$$t^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{y}_1 - \bar{y}_2)(s^2_{\text{pooled}})^{-1}(\bar{y}_1 - \bar{y}_2).$$

The multivariate null hypothesis with P dependent variables is expressed in a matrix vector as follows:

$$\mathrm{H}_0 \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{P_1} \end{pmatrix} = \begin{pmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{P_2} \end{pmatrix},$$

and the Hotelling T^2 multivariate t test that replaces each variable with a vector of means ($\bar{Y}_1$ and $\bar{Y}_2$) for each group is computed as follows:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{Y}_1 - \bar{Y}_2)S^{-1}(\bar{Y}_1 - \bar{Y}_2).$$

Note: S^{-1} is an estimate of the common population covariance matrix of dependent variables for both groups, and capital Y letters are used to denote the matrix vectors of means.

We see from the univariate *t*-test formula that the two sample means for each group are replaced in the multivariate *t* test with a vector of means based on the number of dependent variables. Similarly, the common population covariance matrix in the univariate *t* test is expanded to include more than one dependent variable in the multivariate *t* test. The univariate and multivariate *t*-test formulas should look similar except for the inclusion of the matrix vector notation.

Statistical Significance

The univariate *t* test has a table of critical *t*-test values with varying degrees of freedom for checking statistical significance, while the Hotelling T^2 multivariate *t* test does not. However, statistical significance for both the univariate and multivariate *t* test can be tested using an *F* test.

The Hotelling T^2 statistic uses the sample size of each group, a vector of mean differences between groups, and the pooled sample estimate of the population variance–covariance matrix of the dependent variables. An assumption that the groups have equal variance–covariance matrix is required before testing for mean differences, which is generally computed as the Box's M test. The test of equal variance–covariance matrices between groups is an extension of the assumption in the univariate case, which is tested using the Levene's test of equal variance between two or more groups.

The Hotelling T^2 statistic is tested for significance using the *F* test. The *F*-test formula uses the sample sizes of each group, the number of dependent variables (*P*), and of course the T^2 value. The critical *F* value with numerator and denominator degrees of freedom (*df*) for $\alpha = .05$, .01, and .001 can be found in statistical tables for *F* values; however, software today reports the *F* test of statistical significance. Given the degrees of freedom as follows:

$$df_1 = P$$

$$df_2 = n_1 + n_2 - p - 1$$

The *F* value is computed as follows:

$$F = \left(\frac{df_1}{df_2}\right)T^2.$$

△ Practical Examples Using R

The multivariate *t* test(s) parallel the three types of group mean difference tests computed in the univariate case: (1) single sample, (2) independent sample, and (3) dependent (paired) sample (Hotelling T^2 R tutorial at http://www.uni-kiel.de/psychologie/rexrepos/posts/multHotelling.html). You will need to have the R software installed to conduct these mean difference tests, and, optionally, the *Rcommander* or *RStudio* software (see Preface). Once the software is installed, the R script commands can be entered and run for each type of group mean difference test.

Single Sample

The single-sample multivariate *t* test is computed when you have several dependent variables for a single sample and hypothesize that the vector of means are statistically different from zero (null hypothesis). Alternatively, the vector of dependent variable means could be tested for statistical significance from a specified population mean. An educator might conduct a single-sample multivariate *t* test when obtaining students' test scores on two or more tests, for example, midterm and final exams in a class. Alternatively, a teacher might test whether her students' SAT and ACT scores were statistically different from the population norms for the tests. The first step in conducting a multivariate single-sample *t* test is to install the R package(s) and load the functions. The second step is to read in or create the sample data frame for the number of dependent variables. A third step is to compute and print out the correlation between the dependent variable(s) and compute the means and standard deviations of the dependent variables. A fourth step could include a graph of the means for the dependent variables to visually show the magnitude of mean difference. Finally, a Hotelling T^2 test is computed. The Hotelling **T.2()** function reports the T^2 value, which is an *F* value, since $F = T^2$. The results of each step are output after running the R code for each example.

The following single-sample multivariate *t* test has two dependent variables, *Y*1 and *Y*2. The first dependent variable has scores that indicate the number of points subtracted from a pop quiz. The second dependent variable has scores that indicate the number of points awarded on a homework assignment. The teacher wants to test if the joint mean for these two dependent variables together are statistically significant for her 10 students. The R code for the necessary steps are highlighted, and the results are listed below each step.

R Code: Hotelling T^2 Single Sample

```
# Step 1
# Install R packages and load library of functions

> install.packages("ICSNP")
> install.packages("mvtnorm")
> library(ICSNP)
> library(mvtnorm)

# Step 2
# Enter data for two dependent variables in two separate
matrix vectors Y1 and Y2
# Y12 is a data frame that combines the two matrix vectors of
dependent variables
# The names() function assigns variable names to the
dependent variables
# The attach() function makes it possible to refer to
variables in data frame by their names
# Print out data in the Y12 data frame

> Y1 = c(-2,-4,-6,-3,-7,-2,-1,-8,-6,-9)
> Y2 = c(3,4,9,3,5,4,2,4,2,8)
> Y12 = data.frame(Y1,Y2)
> names(Y12) = c("Y1","Y2")
> attach(Y12)
> Y12

     Y1  Y2
1    -2   3
2    -4   4
3    -6   9
4    -3   3
5    -7   5
6    -2   4
7    -1   2
8    -8   4
9    -6   2
10   -9   8
```

Data for Y1 and Y2 dependent variables

```
# Step 3
# Print out correlation (cor) of dependent variables
# Print out means (mean) and standard deviations (sd) of
dependent variables
```

```
> cor(Y12)
> mean(Y1); sd(Y1)
> mean(Y2); sd(Y2)

            Y1         Y2
Y1  1.0000000 -0.5875697
Y2 -0.5875697  1.0000000
```

Correlation of dependent variables

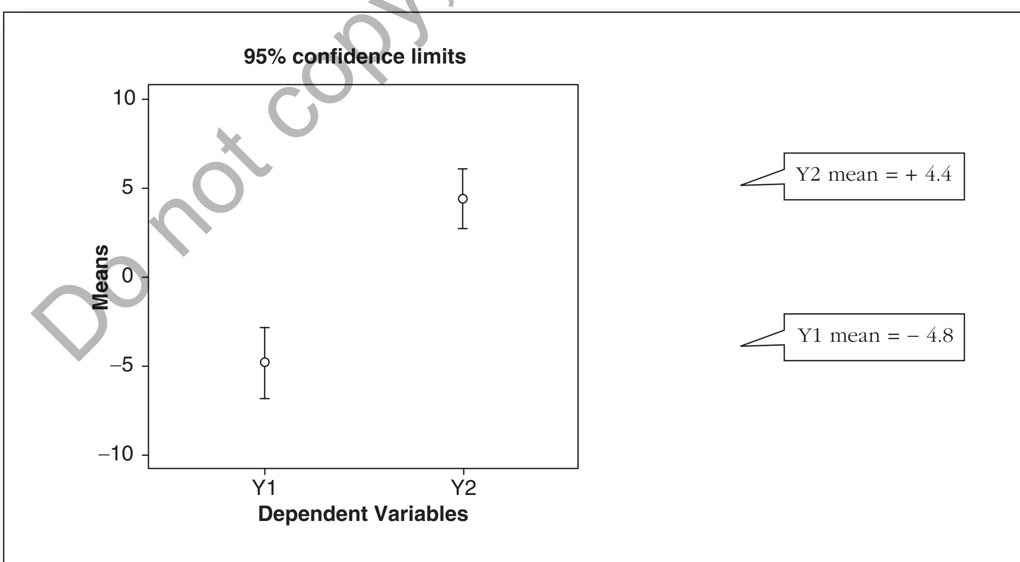
```
mean(Y1); sd(Y1)
[1] -4.8
[1] 2.780887
```

Y1 mean and standard deviation
Y2 mean and standard deviation

```
mean(Y2); sd(Y2)
[1] 4.4
[1] 2.366432

# Step 4
# Install and load R psych package
# Graph dependent variable means using error.bars() function

> install.packages("psych")
> library(psych)
> error.bars(Y12,bar=FALSE,ylab="Group Means",xlab="Dependent Variables",
+ ylim = c(-10,10),eyes=FALSE)
```



```
# Step 5
# Conduct a Hotelling T2 test of null hypothesis that
dependent means are different than zero
# muH0 assigns population means in matrix vector to equal zero

> muH0 = c(0, 0)
> HotellingsT2(Y12, mu=muH0)


Hotelling's one sample T2 test        Hotelling T² output

data: Y12
T.2 = 18.0899, df1 = 2, df2 = 8, p-value = 0.001075
alternative hypothesis: true location is not equal to c(0,0)
```

The results for the single-sample multivariate *t* test indicated that the two dependent variable means together are statistically significantly different from zero. The correlation matrix indicated that the two dependent variables were correlated, $r = -.587$. The Hotelling T^2 value was statistically significant: T.2 = 18.089 with 2 and 8 *df*, and $p = .001$. Therefore, the null hypothesis of no joint mean difference is rejected. The alternative hypothesis is accepted, which reflects a test of whether the joint sample means are different from zero [true location difference is not equal to *c*(0,0).

Two Independent Group Mean Difference

The two independent group multivariate *t* test is when you hypothesize that a set of dependent variable group means are different between two independent groups, for example, Rogerian and Adlerian counselors. The R code is highlighted for testing the null hypothesis of no mean difference, and the output is listed after the R code. I have placed comments before sets of R command lines to provide a brief explanation of what each set of commands are doing. There are three Rogerian counselors and six Adlerian counselors measured on two dependent variables by their clients. The first measure was counseling effectiveness and the second measure was counseling satisfaction based on a 10-point numerical scale.

R Code: Hotelling T^2 (Two Independent Samples)

```
# Step 1
# Install R packages and load library of functions
```

```
> install.packages("ICSNP")
> install.packages("mvtnorm")
> library(ICSNP)
> library (mvtnorm)

# Step 2
# Use data set from James Steven Book (2009, 5th Edition) P. 148
the independent variables Rogerian vs Adlerian
# Assign data to two matrices with different number of subjects
# Assign data to matrix for group membership variable, grp
# Print out the matrices

>  roger = matrix(c(1,3,2,3,7,2),3,2)
>  adler = matrix(c(4,6,6,5,5,4,6,8,8,10,10,6),6,2)
>  grp = matrix(c(1,1,1,2,2,2,2,2,2),9,1)
>  roger
>  adler
>  grp

roger
       [,1] [,2]
[1,]      1    3
[2,]      3    7
[3,]      2    2

adler
       [,1] [,2]
[1,]      4    6
[2,]      6    8
[3,]      6    8
[4,]      5   10
[5,]      5   10
[6,]      4    6

grp
       [,1]
[1,]      1
[2,]      1
[3,]      1
[4,]      2
[5,]      2
[6,]      2
[7,]      2
```

Matrix output for roger, adler, and grp

```
[8,]      2
[9,]      2

# Step 2 continued
# Combine the two dependent variable matrices
# Add variable names with names() function
# Use attach() function so variable names can be used
# Use factor() function to declare group variable as categorical

# Select p-value to print in decimal rather than
scientific notation

> Y = data.frame(rbind(roger,adler))
> names(Y) = c("effect","satis")
> attach(Y)
> factor(grp)
> options(scipen=999)
> Y

   effect  satis
1       1      3
2       3      7
3       2      2
4       4      6
5       6      8
6       6      8
7       5     10
8       5     10
9       4      6


# Step 3
# Print out correlation between effect and satis
# Print out means and standard deviations for two dependent variables

> cor(effect,satis)
> mean(effect);sd(effect)
> mean(satis);sd(satis)

> cor(effect,satis)
[1] 0.8295614
```

Correlation between effect and satis

```
> mean(effect);sd(effect)
[1] 4
[1] 1.732051
```

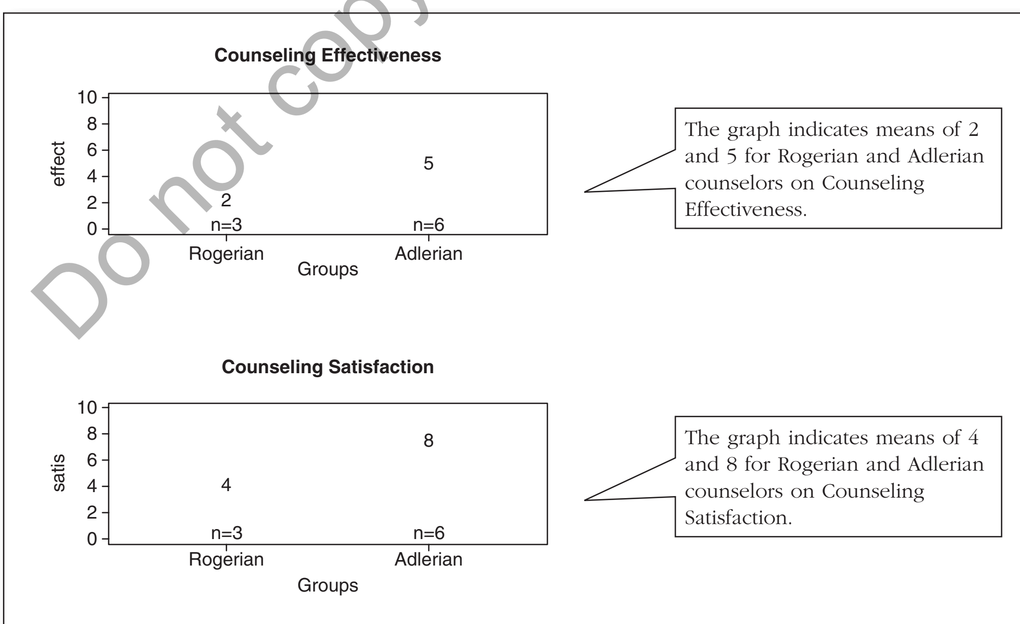
Means and standard deviations

```
> mean(satis);sd(satis)
[1] 6.666667
[1] 2.783882
```

```
# Step 4
# Install and load R gplots package
# Graph individual group dependent variable means

> install.packages("gplots")
> library(gplots)
> group = data.frame(grp)
> Yall = data.frame(group,Y)
> par(mfrow= (2, 1))
> plotmeans(effect ~ grp,data=Yall,ylim=c(0,10),xlab="Groups",
+   legends = c("Rogerian","Adlerian"),main="Counseling
+   Effectiveness",connect=FALSE,mean.labels=TRUE,col=NULL,p=1.0)
> plotmeans(satis ~ grp,data=Yall,ylim=c(0,10),xlab="Groups",
+  legends = c("Rogerian","Adlerian"),main ="Counseling
+ Satisfaction",connect=FALSE,mean.labels=TRUE,col=NULL,p=1.0)
```

Note: The + symbol indicates a line carryover due to page margins, which would not be entered in R. The entire single line must be entered.



```
# Step 5
# Install and load R biotools package
# Declare grp as factor variable
# Use Y data set and grp variable created before
# Compute Box M test of equal covariance matrices

> install.packages ("biotools")
> library(biotools)
> factor(grp)
> boxM(Y,grp)


Box's M-test for Homogeneity of Covariance Matrices

data: Y
Chi-Sq (approx.) = 0.1149, df = 3, p-value = 0.99
```

The Box M test results indicated that the variance–covariance matrices of the two groups were equal (χ^2 = .11, *df* = 3, *p* = .99).

Note: When group sizes are 20 or more and the number of dependent variables are 5 or more, the chi-square approximation is preferred, otherwise the *F* approximation is more accurate (Stevens, 2009).

```
# Step 6
# Compute Hotelling T2 based on two sample data matrices
> HotellingsT2(roger,adler)

Hotelling's two sample T2-test

data: roger and adler
T.2 = 9, df1 = 2, df2 = 6, p-value = 0.01562
alternative hypothesis: true location difference is not
equal to c(0,0)
```

T.2 = *F* value

or the alternative using a formula:

```
> factor(grp)
> HotellingsT2 (formula = Y ~ grp)


# Step 7
# Compute F test for Hotelling T-squared value
```

```
# Enter sample sizes, number of dependent variables, and
Hotelling T2 value

> n1 = 3
> n2 = 6
> p = 2
> T = 3

# Compute degrees of freedom for numerator and denominator

> df1 = p
> df2 = n1 + n2 - p - 1

# Compute F value and p-value

> Fval = (df2/df1) * T
> pval = round(1-pf(Fval,df1,df2),digits=3)


# print out T, F, df1, df2, and p-value

> cat("T = ",T,"F-value =",Fval,"df1 =",df1,"df2 =",df2,"
p-value=",pval,fill=FALSE,"\n")


T = 3   F-value = 9  df1 = 2
 df2 = 6  p-value = 0.016
```

The $T = 3$ and $F = 9$ computed here are the Hotelling T and F values reported in the SPSS and SAS output. The $p = .016$ alpha level is for the F test.

The results show that the two dependent variables were positively correlated, $r = .829$. The theoretical meaningfulness and correlation of the two dependent variables provided the rationale for conducting the multivariate *t* test. The first dependent variable had mean = 4 and standard deviation = 1.73, and the second dependent variable had mean = 6.67 and standard deviation = 2.78. The Box M test indicated that the covariance matrices were not statistically different, so we assumed them to be equal and proceeded with the multivariate *t* test. The results indicated that T.2 = 9, with 2 and 6 *df* and $p = .016$ (*Note:* The function reports T squared, which is equal to an F value—that is, $T^2 = (3)^2 = 9$. The null hypothesis of no group mean difference is rejected. The alternative hypothesis is accepted—true location difference is not equal to *c*(0,0)—which indicates that the two groups, Rogerian and Adlerian, had a statistically significant joint mean difference for counseling effectiveness and counseling satisfaction by clients. A graph of the individual group means for counseling effectiveness and counseling satisfaction shows that Adlerian counselors had higher client means than Rogerian counselors on both dependent variables.

Tip:

When covariance matrices are not homogeneous, a Wald test would be computed. The R code is as follows:

```
# Compute a Wald test when the covariance
matrices are not homogeneous.

> W = t(m1-m2)%*%solve(s1/n1+s2/n2)%*%(m1-m2)
> cat ("Wald test = ",W, fill=T)

Wald test = 20.15924.
```


Two Groups (Paired) Dependent Variable Mean Difference

The multivariate dependent *t* test is an extension of the univariate dependent *t* test with two or more dependent variables. The data entry is important because you will need to calculate the mean difference between the two groups on each dependent variable. The R code has been written to provide certain values prior to the actual Hotelling T^2 dependent *t* test. This includes printing out the difference scores, means, and standard deviations. The R code is described at each step in a text box. The R code shows two different approaches when conducting the multivariate dependent *t* test. The *first* approach is comparing the difference scores between two groups. The two groups are fifth-grade boys and girls. The dependent variable was the pop quiz test. The *second* approach is comparing all students on their difference scores. The pop quiz test was given twice, once after instruction and again 2 weeks later. The teacher wanted to test memory retention of the material taught in class. She hypothesized that students would not retain the information, and thus, they would score lower on the second administration of the pop quiz. The teacher not only wanted to see if there were differences between the boys and girls but also wanted to know if there was a difference overall for her students, hence the two different multivariate dependent *t*-test approaches.

R Code: Hotelling T^2 *(Two Paired Dependent Variables)*

```
# Step 1
# Install R package and load library of functions

> install.packages("ICSNP")
> install.packages("mvtnorm")
> library(ICSNP)
> library (mvtnorm)

# Step 2
# Enter data for two dependent variables (Pre and Post) by
the two groups (Boy and Girl)
# Place in a data frame and list variables

> PreBoy = c(12,16,18,12,10)
> PostBoy = c(14,16,18,10,12)
> PreGirl = c(5,2,7,4,15)
> PostGirl = c(8,2,16,10,14)
> mydata = data.frame(PreBoy,PostBoy,PreGirl,PostGirl)
> attach(mydata)
> mydata

    PreBoy  PostBoy  PreGirl  PostGirl

1       12       14        5        8
2       16       16        2        2
3       18       18        7       16
4       12       10        4       10
5       10       12       15       14

# Step 3
# Compute descriptive statistics on dependent variables for
each group

> mean(PreBoy);sd(PreBoy)
> mean(PostBoy);sd(PostBoy)
> mean(PreGirl);sd(PreGirl)
> mean(PostGirl);sd(PostGirl)

mean(PreBoy);sd(PreBoy)
[1] 13.6
[1] 3.286335
```

> Group 1: Boy
> Pop Quiz First Time = 13.6 average
> Pop Quiz Second Time = 14 average

```
mean(PostBoy);sd(PostBoy)
[1] 14
[1] 3.162278

mean(PreGirl);sd(PreGirl)
[1] 6.6
[1] 5.029911
```

Group 2: Girl
Pop Quiz First Time = 6.6 average
Pop Quiz Second Time = 10 average

```
mean(PostGirl);sd(PostGirl)
[1] 10
[1] 5.477226
```

Approach 1: Compare Boys and Girls Pop Quiz Difference Scores

In the first approach, we would first calculate the difference scores in each group. Then, we would calculate the mean difference for each group. The R commands are as follows.

```
# Calculate and print mean difference scores by group

> diffBoy = PostBoy - PreBoy
> diffGirl = PostGirl - PreGirl
> mean(diffBoy);sd(diffBoy)
> mean(diffGirl);sd(diffGirl)

mean(diffBoy);sd(diffBoy)
[1] 0.4
[1] 1.67332
```

Boy: Post − Pre = 0.4 mean difference

```
mean(diffGirl);sd(diffGirl)
[1] 3.4
[1] 4.159327
```

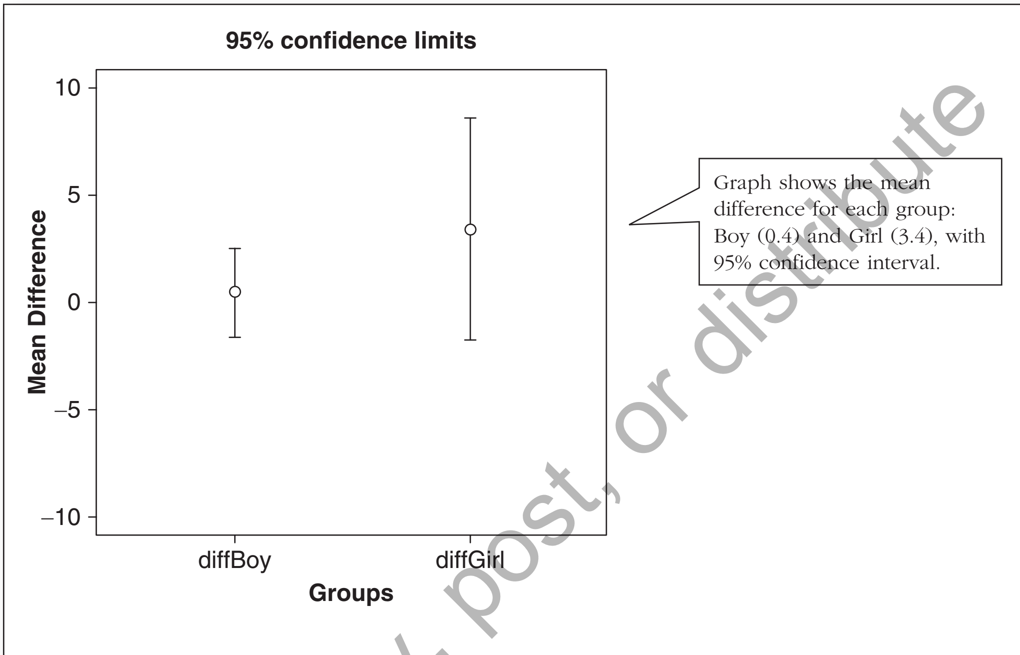
Girl: Post − Pre = 3.4 mean difference

We would then want to graph the dependent variable mean differences to visually inspect the magnitude of the mean difference. The R commands are as follows:

```
# Install and load psych package
# Create data frame of difference scores
# Graph dependent variable mean differences

> install.packages("psych")
> library(psych)
> YDiff = data.frame(diffBoy,diffGirl)
```

```
> YDiff
> error.bars(YDiff,bar=FALSE,ylab="Group Means",
xlab="Dependent Variables",ylim=c(-10,10),eyes=FALSE)
```



We can visually inspect the difference scores in each group with the following R command.

```
# Place difference scores in data frame

> YDiff


YDiff


    diffBoy diffGirl
1         2        3
2         0        0
3         0        9
4        -2        6
5         2       -1
```

Finally, we compute the Hotelling T^2 statistic separately on the difference scores for each group.

```
# Compute Hotelling T-square on difference scores for Boys

> muH0 = c(0)
> HotellingsT2(diffBoy, mu=muH0)


Hotelling's one sample T2-test

data: diffBoy
T.2 = 0.2857, df1 = 1, df2 = 4, p-value = 0.6213
alternative hypothesis: true location is not equal to c(0)
```

The mean difference (0.4) was not statistically significant for the Boys.

```
# Compute Hotelling T-square on difference scores for Girls

> muH0 = c(0)
> HotellingsT2(diffGirl, mu=muH0)


Hotelling's one sample T2-test

data: diffGirl
T.2 = 3.341, df1 = 1, df2 = 4, p-value = 0.1416
alternative hypothesis: true location is not equal to c(0)
```

The mean difference (3.4) was not statistically significant for the Girls.

Approach 2: Compare All Students in Class on Pre and Post Scores

The Hotelling T2 test can be computed for omnibus difference scores for all subjects in the data set. We first create the data set with the following R commands.

```
# Step 1
# Create data frame with dependent variables and group
membership variable

> g = 2 # number of groups
> N = 5 # number of subjects per group
> Group = matrix(rep(1:g, each=N))
> Pre = c(PreBoy,PreGirl)
```

```
> Post = c(PostBoy,PostGirl)
> All = data.frame(Pre,Post,Group)
> names(All) = c("Pre","Post","Group")
> factor(Group)
> All
```

```
   Pre Post  Group
1   12   14      1
2   16   16      1
3   18   18      1
4   12   10      1
5   10   12      1
6    5    8      2
7    2    2      2
8    7   16      2
9    4   10      2
10  15   14      2
```

```
# Step 2
# Compute Descriptive statistics on Pre and Post scores

> mean(Pre)
> mean(Post)

mean(Pre)

[1] 10.1

mean(Post)

[1] 12
```

The fifth-grade class had a 10.1 average Pop Quiz first time (Pre) and a 12 average Pop Quiz second time (Post).

```
# Step 3
# Create data matrix of difference scores for all students
# Print out difference scores
# Compute average difference score

> Diff = cbind(Post-Pre)
> Diff
> mean(Diff)
```

```
Diff
      [,1]
 [1,]    2
 [2,]    0
 [3,]    0
 [4,]   -2
 [5,]    2
 [6,]    3
 [7,]    0
 [8,]    9
 [9,]    6
[10,]   -1

mean(Diff)

[1] 1.9
```

Pop Quiz mean difference = (12 − 10.1) = 1.9.

```
# Compute Hotelling T-square on difference scores for all
students

> muH0 = c(0)
> HotellingsT2(Diff, mu=muH0)

Hotelling's one sample T2-test

data: Diff
T.2 = 3.1574, df1 = 1, df2 = 9, p-value = 0.1093
alternative hypothesis: true location is not equal to c(0)
```

The Pop Quiz mean difference is not statistically significant. $T^2 = F = 3.157$, $p = .109$.

The first approach conducted a multivariate dependent t test to test whether the fifth-grade boys differed on their Pop Quiz difference scores compared with the girls Pop Quiz difference scores. The boys had a 0.4 mean difference, while the girls had a 3.4 mean difference. For the boys, Hotelling $T^2 = 0.2857$, $df_1 = 1$, $df_2 = 4$, and p value = .6213, so we would retain the null hypothesis of no difference in Pop Quiz scores. For the girls, Hotelling $T^2 = 3.341$, $df_1 = 1$, $df_2 = 4$, p value = .1416, so we would retain the null hypothesis of no difference in Pop Quiz scores. The teacher was pleased that there was no statistical difference between the boys' and girls' Pop Quiz scores.

The second approach conducted a multivariate dependent t test to test whether all fifth-grade students in her class differed in their Pop Quiz scores. The data frame shows the Pre and Post scores for the dependent variables side by side. This helps our understanding that the mean difference is what is being tested for statistical significance. For example, the Pop Quiz mean was 10.1 the first time it was administered (Pre), and the Pop Quiz mean was 12 the second time it was administered (Post). So the mean difference is 12 − 10.1 = 1.9. The Hotelling $T^2 = 3.1574$, $df_1 = 1$, $df_2 = 9$, and p value = .1093, so we would retain the null hypothesis of no difference in Pop Quiz scores for all students. The teacher gave the same Pop Quiz both Pre and Post, so her interest was in whether students retained the information she taught. Therefore, the teacher was pleased that the students did retain the information; thus, no difference on average between the first and second administration of the Pop Quiz was a good finding. In contrast, researchers often design a study with a pretest, followed by a treatment, and then a posttest. In this type of research design, the researcher would expect a statistically significant difference if the treatment was effective and changed students' scores.

△ Power and Effect Size

There are several factors that affect the power of a statistical test to detect a mean difference. The factors that influence the power to detect a mean difference are as follows:

1. Type I error rate (alpha level)
2. Sample size
3. Effect size (difference in groups on the dependent variable)
4. Population standard deviation (homogeneous or heterogeneous)
5. Directionality of hypothesis (one-tail test vs. two-tail test)

When planning a research study, we would select values for these five criteria to compute power (http://www.cedu.nniu.edu/~walker/calculators/). Alternatively, we could determine sample size by selecting power and the other four criteria to compute the sample size needed for the study.

Their impact on power for each of these factors is briefly described as follows:

- *Type I error:* Probability of rejecting the null hypothesis when it is true (hypothesize that groups differ but really don't)
- *Sample size:* The larger the sample, the more representative of the population

- *Effect size:* The smaller the difference wanting to detect, the larger scaled difference needed
- *Population standard deviation:* Homogeneous (smaller sample size); heterogeneous (larger sample size)
- Directionality of hypothesis: Test for mean difference in one direction has more power over testing for mean differences in both tails.

We should also be concerned with Type II error rate, which is the counterintuitive testing of the Type I error rate, which is defined as follows:

- Type II error: Probability of accepting the null hypothesis when it is false (stated that groups don't differ but really do)

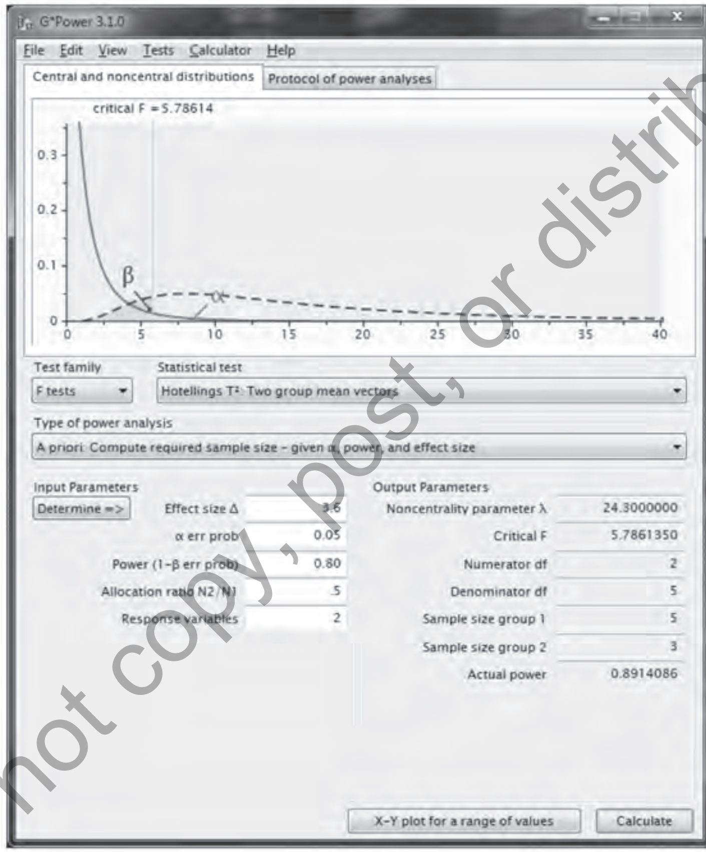
A Priori Power Estimation

A researcher can determine power when planning a study, which is an a priori determination, by selecting values listed above. Power is a statement of how probable you want to be in detecting a mean difference, for example, 80% probability of rejecting a null hypothesis when false. A popular free software, G*Power 3, determines the a priori power for different statistical tests (http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/).

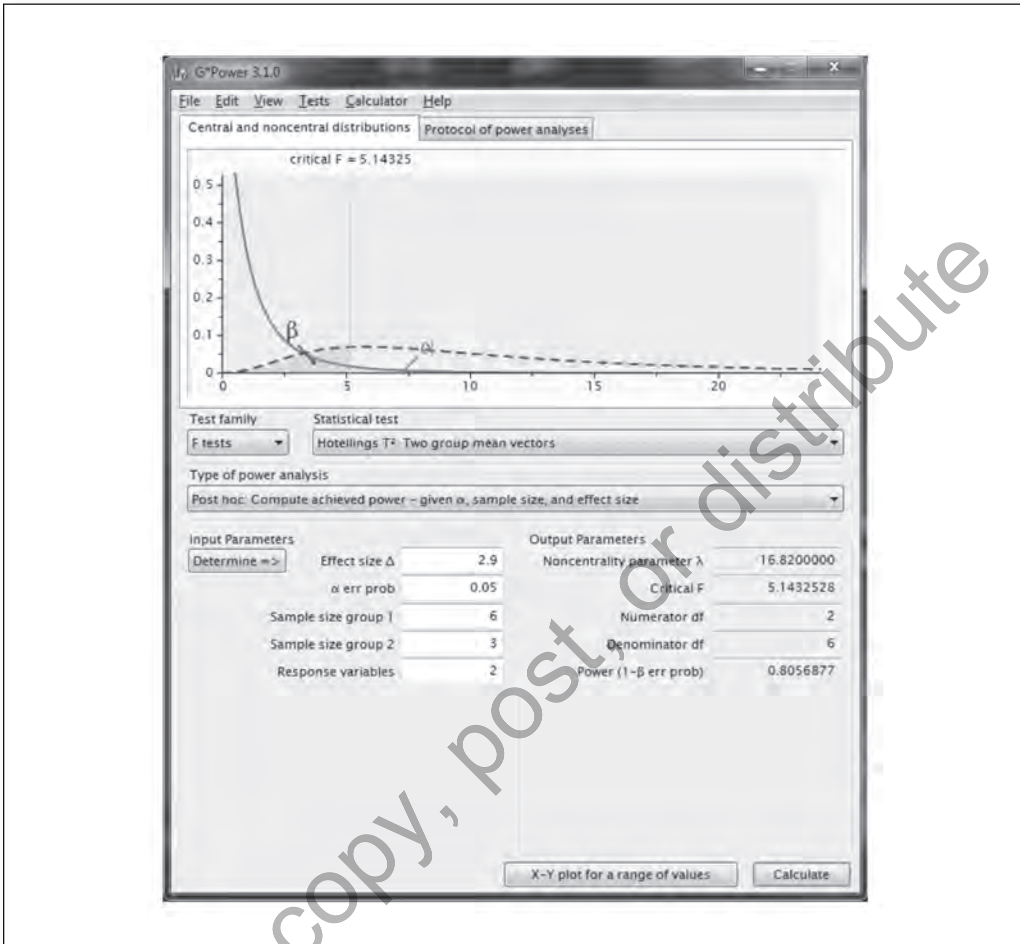
G*Power 3 has options for the Hotelling T^2 one group. We would enter the following values to determine the sample size: effect size (1.2), Type I alpha (.05), power (.80), and number of response variables (2). Sample size was 10, which is the number of subjects in the single-sample multivariate *t* test. We could detect a mean difference of 1.2 (effect size); our results indicated $Y1 = -4.8$ and $Y2 = 4.4$, which was greater than the specified effect size (1.2).

Note: Criteria in the dialog boxes can be varied to achieve different results for effect size, power, and number of response variables.

G*Power 3 also has options for the Hotelling T^2 two independent group. We would enter the following values to determine the sample size: effect size (3.6), Type I alpha (.05), power (.80), sample size ratio ($n_1 = 3/n_2 = 6$), and number of response variables (2). Effect size was selected to be 3.6 based on *Y*1 mean difference of 3.0 between Rogerian (mean = 2) and Adlerian (mean = 5) counselors on counseling effectiveness, and *Y*2 mean difference of 4.0 between Rogerian (mean = 4) and Adlerian (mean = 8) counselors on counseling satisfaction. The other criteria was selected to be Type I error rate or α = .05, power = .80, number of response variables = 2, and ratio of sample sizes (3/6) = .5. Sample size was given as Group 1 = 5 and Group 2 = 3 for power = .89, so we had sufficient sample size and power given our criteria.



G*Power also has other types of analysis options, which are shown in the pull-down menu. The dialog box below, for example, computes power based on alpha, sample size, and effect size. I input the values for the sample sizes of the two groups, number of response variables, and effect size, which yielded power = .805.



Effect Size Measures

The univariate effect size measures are generally given when reporting the general linear model results. These popular univariate effect size measures (how many standard deviation units the group means are separated by) are as follows:

1. Cohen's d

$Cohen's\ d = \frac{(\mu_1 - \mu_2)}{\sigma}$, where σ is the common population standard deviation.

2. Partial eta-squared

$$\eta_P^2 = \frac{(df \times F)}{\left(df_{\text{h}} \times F + df_{\text{e}} \right)}.$$

Note: df_{h} is degrees of freedom for hypothesis, and df_{e} is degrees of freedom for error term. A partial eta-squared = .01 (small), .06 (medium), and .14 (large) effect sizes.

The Mahalanobis D^2 measure is commonly reported as a multivariate effect size measure. It uses the vector of mean differences and the common population covariance matrix. The Mahalanobis D^2 measure is calculated as follows:

3. Mahalanobis D^2 (two-group means)

$$D^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2),$$

where the multivariate vector of means is used with the variance–covariance matrix.

$$\hat{D}^2 = (\bar{Y}_1 - \bar{Y}_2)'S^{-1}(\bar{Y}_1 - \bar{Y}_2).$$

The Mahalanobis D^2 is a measure of the separation of the independent group means without using the sample sizes of the groups (Hotelling T^2 without sample size). It yields a value that indicates the distance in space between the dependent variable means.

You can obtain the F and T^2 values from the R code and then calculate the D^2 effect size measure. The calculations using the R output from the multivariate independent two-group results would be as follows:

$$F = \left(\frac{df_1}{df_2}\right)T^2 = \left(\frac{6}{2}\right)3 = 9$$

$$T^2 = \left(\frac{df_1}{df_2}\right)F = \left(\frac{2}{6}\right)9 = 3$$

$$D^2 = \frac{NT^2}{n_1 n_2} = \frac{9(3)}{3(6)} = 1.5$$

The D^2 effect size = 1.5 is considered a large effect size.

△ Reporting and Interpreting

A researcher should provide the descriptive statistics for the Hotelling T^2 test of mean differences (means, standard deviations, and correlations). In addition, the Box M test of equal covariance matrices should be reported. This is followed by reporting the Hotelling T^2, degrees of freedom, and p value. The power and effect size information should also be given when possible. It is important to report these values along with the hypothesis or research question. An examination of published journal articles in your field will guide what information to report when conducting a Hotelling T^2 analysis. A basic write-up is provided to help with that understanding.

Rogerian and Adlerian counselors were compared on two dependent measures: counseling effectiveness and counseling satisfaction. The means for Adlerian counselors were higher than Rogerian counselors on the two dependent variables. A Hotelling T^2 two independent group analysis was conducted which indicated a statistically significant mean difference between the two groups ($T^2 = 3$, $df = 2, 6$, $p = .016$) for the two dependent variables. Adlerian counselors had higher mean scores on counseling effectiveness and counseling satisfaction (5 and 8) than Rogerian counselors (2 and 4). The multivariate results indicated a significant dependent variable joint effect. The multivariate effect size = 1.5 and power = .80.

SUMMARY

This chapter presented a two-group multivariate test of mean differences on two or more dependent variables. The Hotelling T^2 test can be conducted on a single sample, mean difference between two independent groups, or mean difference of a paired group. It is considered an extension of the univariate t-test method. The assumptions and practical examples demonstrated how R functions can be used to test the mean differences.

An important concept was also presented in the chapter, namely, power and effect size. The factors that affect power were illustrated using G*Power 3 software. The software permits the determination of sample size and/or power for the different multivariate tests. Additionally, the discussion of effect size measures relates the importance of looking beyond statistical significance to the practical importance and meaningfulness of interpretation given by an effect size measure. The relation and formula to convert F and T^2 into a D^2 effect size is important, especially when the statistical output does not readily provide an effect size measure.

EXERCISES

1. Create two data vectors and merge them into one using R code.
2. Create a single membership vector for two groups.
3. Create an R code for data analysis in the Hotelling T^2 two independent group example and show results.

WEB RESOURCES

Box's M test
http://en.wikiversity.org/wiki/Box's_M

Dunn–Bonferroni
http://en.wikipedia.org/wiki/Bonferroni_correction

G*Power 3
http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/

Hotelling Biography
http://en.wikipedia.org/wiki/Harold_Hotelling

Hotelling T^2 R tutorial
http://www.uni-kiel.de/psychologie/rexrepos/posts/multHotelling.html

Levene's test
http://en.wikipedia.org/wiki/Levene's_test

Power and effect size
http://www.cedu.niu.edu/~walker/calculators/

REFERENCES

Hotelling, H. (1931). The generalization of student's ratio. *Annals of Mathematical Statistics, 2*(3), 360–378.
Schumacker, R. (2014). *Learning statistics using R*. Thousand Oaks, CA: Sage.
Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York, NY: Routledge.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.