Example: Multivariate Analysis of Variance

Multivariate analyses of variance (MANOVA) differs from univariate analyses of variance (ANOVA) in the number of dependent variables utilized. The major purpose of univariate analyses of variance is to determine the effects of one or more independent variables upon ONE dependent variable. However, there arise situations in which measurements are available on MORE THAN ONE dependent variable.

For example, say you are interested in designing an experiment to test the effects of success of street kids exiting street life. “Success” can be measured in more than one way. It can be measured by both ‘the number of days a kid “sleeps” at home’ or ‘the number of days a kid returns to school’. In such a case, there are TWO dependent variables involved in the analysis.

In analyzing such data, one obvious (and popular) approach is to perform an ANOVA on each dependent variable separately. In our example above, two ANOVA’s could be carried out (i.e., one ANOVA with “home” scores
and one ANOVA with “school” scores. However, it is quite apparent that both the home and school variables are highly correlated. Thus, the two analyses are not independent. One’s ‘return to school’ is most likely positively correlated with ‘one’s return home’. The independent ANOVA ignores the interrelation between variables. Consequently, substantial information may be lost (SPSS, Release 9.0), and the resultant p values for tests of hypothesis for the two independent ANOVA’s are incorrect.

The extension of univariate analysis of variance to the case of multiple dependent variables is known as Multivariate Analysis of Variance (MANOVA). MANOVA allows for a direct test of the null hypothesis with respect to ALL the dependent variables in an experiment. In MANOVA, a linear function \( y \) of the dependent variables in the analysis is constructed, so that “inter-group differences” on \( y \) are maximized. The composite variable \( y \) is then treated in a manner somewhat similar to the dependent variable in a univariate ANOVA, with the null hypothesis being accepted or rejected accordingly.
Essentially, the multivariate technique known as **Discriminant analysis** is quite similar to **MANOVA**. In **discriminant analysis** we consider the problem of finding the best linear combination of variables for distinguishing several groups. The coefficients (Betas) are chosen in order that the *ratio* of the ‘between-groups sums of squares’ to the ‘total sums of squares’ is as large as possible. **MANOVA** can be viewed as a problem of first finding *linear combinations* of the dependent variables that ‘best separate the groups’ and then testing ‘whether these new variables are significantly different for the groups’. In the **MANOVA** situation, we already know which categories the cases belong to, and thus are not interested in classification **BUT** in identifying a composite variable (y) which illuminates the differences among the groups.

In both procedures the composite variable (y) is known as a **discriminant function**. In a **MANOVA**, discriminant function is a weighted sum of the dependent variables, with the weightings chosen such that the distributions of y for the various groups are separated to the greatest possible extent.
In a DISCRIMINANT analysis, the discriminant function is used to predict category membership to the greatest extent possible. Thus, both techniques have very much in common and this will be illustrated in the last section of this module. Remember that a MANOVA is used when the dependent variables are CORRELATED.

Example

Let’s see multivariate analysis of variance in action. We borrowed data from Moore and McCabe (1998:Data Appendix 4) in order to illustrate this procedure. The study is described by the researchers as follows:

Jim Baumann and Leah Jones of the Purdue University School of Education conducted a study to compare three methods of teaching reading comprehension. The 66 students who participated in the study were randomly assigned to the methods (22 to each). The standard practice of comparing new methods with a traditional one was used in this study. The traditional method is called Basal and the two innovative methods are called DRTA and Strat.

(Moore and McCabe, 1998, Data Appendix 4)
The 66 subjects in this study are divided into three groups (B, D, S) which we have coded as B=1, D=2, S=3. There are five independent variables which represent either pretest or posttest scores (i.e., pre1, pre2, post1, post2, and post3). Let’s imagine that we are interested in discovering whether these variables are significantly different for the three groups. Remember that we are using MANOVA since we suspect that the dependent variables are correlated.

To perform a multivariate analysis of variance in SPSS for this reading study, we need to complete the following steps:

1. Pull up the “Reading” data set.

2. Click ‘Statistics’ on the main menu bar.

3. Click ‘General Linear Model’ on the Statistics menu.

4. Select ‘GLM-Multivariate’ in the General Linear Model submenu. This will open a GLM-Multivariate box similar to that shown on the next page.
The next step is to specify the dependent variables and the fixed factors. In this example, the five reading scores (i.e. ‘pre1’, ‘pre2’, ‘post1’, ‘post2’, ‘post3’) are the Dependent variables. ‘Group’ is the Fixed Factor in this MANOVA. To specify the variable as a dependent or fixed factor, simply select the variable in the box on the left hand side, and click the arrow button to the left of the ‘Dependent Variables’ or ‘Fixed Factor(s)’ respectively. The GLM-Multivariate dialogue box should now appear as seen on the next page:
The commands for the basic MANOVA have been defined through the aforementioned four steps. We now want to include specific criteria for our MANOVA. In order to do this we must complete the following steps:

1. Click the ‘Model’ pushbutton at the top-right corner of the GLM-Multivariate dialogue box. This will open up the submenu and allow us to further specify our
analysis as follows:

- Click the ‘Custom’ button.
- Select ‘group(F)’ in the ‘Factors & Covariates’ box. Then move ‘group’ over to the ‘Model box’ by clicking the right arrow under ‘build term.’
- Select ‘Main effects’ in the drop-down window below the ‘Build Term(s)’ arrow.

Click ‘Continue’ to return to the GLM-Multivariate box.
2. Click the ‘Contrasts’ pushbutton. This will allow us to further specify our analysis as follows:

- Select ‘Simple’ in the ‘Contrast’ drop-down window.
- Click the ‘First’ pushbutton for the ‘Reference Category’.
- Click the ‘Change’ pushbutton. The ‘Factor’ submenu should now read ‘Group[Simple[first]]’ as seen below.

Click ‘Continue’ to return to the GLM-Multivariate box.
3. Click the ‘Plots’ pushbutton. This will open up this submenu and allow us to make the following selections:

- Select ‘group’ in the ‘Factors’ box and click the right arrow to the ‘Horizontal Axis:’. This will move ‘group’ into the ‘Horizontal Axis:’.
- The ‘Add’ button on the ‘Plots’ box is now an available option. Click this ‘Add’ button. This will now add ‘group’ to this box, and submenu should appear as follows:

![GLM - Multivariate: Profile Plots](image)

Click ‘Continue’ to return to the GLM-Multivariate box.
4. Click the ‘Post Hoc’ pushbutton. This will allow us to further specify our analysis as follows:

- Select ‘group’ in the ‘Factor(s):’ box and click the right arrow to the ‘Post Hoc Tests for:’ box. This will move ‘group’ into the ‘Post Hoc Tests for:’ box.

- Click the ‘Bonferroni’ box in the ‘Equal Variances Assumed’ section, and the submenu will appear as follows:

![Image of GLM - Multivariate: Post Hoc Multiple Comparisons for Observed Means]

Click ‘Continue’ to return to the GLM-Multivariate box.
5. Click the ‘Save’ pushbutton. This will allow us to further specify our analysis as follows:

- In the ‘Predicted Values’ section, select both the “Unstandardized” and “Standard error” boxes.
- In the ‘Diagnostics’ section, select both the “Cook’s distance” and ‘Leverage values’ boxes.
- In the ‘Residuals’ section, select both the ‘Unstandardized’ and ‘Standardized’ boxes.

Click ‘Continue’ to return to the GLM-Multivariate box.
6. Click the ‘Options’ pushbutton. This will allow us to further specify our analysis as follows:

- In the ‘Estimated Marginal Means’ section, select ‘group’. Then click the right arrow button to move ‘group’ into the ‘Display Means for:’ box.
- In the ‘Display’ section select the following boxes: ‘Descriptive statistics’; ‘Estimates of effect size’; ‘Observed power’; ‘Residual SSCP matrix’; ‘Homogeneity tests’; and ‘Residual plots’. Your submenu should now appear as below:
Click “Continue’ to return to the GLM-Multivariate box.

Click ‘OK’ to run this SPSS analysis.

Congratulations!
You have just completed the SPSS MANOVA for this module.
Now let’s see how to interpret our SPSS Data Output.
Multivariate Analysis of Variance

SPSS Output Explanation

Descriptives

SPSS provides us with a general idea of the distribution of scores in each group in the ‘Descriptive Statistics’ output.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GROUP</strong> Basal</td>
<td>10.50</td>
<td>2.97</td>
<td>22</td>
</tr>
<tr>
<td>DRTA</td>
<td>9.73</td>
<td>2.69</td>
<td>22</td>
</tr>
<tr>
<td>Strategies</td>
<td>9.14</td>
<td>3.34</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>9.79</td>
<td>3.02</td>
<td>66</td>
</tr>
<tr>
<td><strong>GROUP</strong> Basal</td>
<td>5.27</td>
<td>2.76</td>
<td>22</td>
</tr>
<tr>
<td>DRTA</td>
<td>5.09</td>
<td>2.00</td>
<td>22</td>
</tr>
<tr>
<td>Strategies</td>
<td>4.95</td>
<td>1.86</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>5.11</td>
<td>2.21</td>
<td>66</td>
</tr>
<tr>
<td><strong>GROUP</strong> Basal</td>
<td>6.68</td>
<td>2.77</td>
<td>22</td>
</tr>
<tr>
<td>DRTA</td>
<td>9.77</td>
<td>2.72</td>
<td>22</td>
</tr>
<tr>
<td>Strategies</td>
<td>7.77</td>
<td>3.93</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>8.08</td>
<td>3.39</td>
<td>66</td>
</tr>
<tr>
<td><strong>GROUP</strong> Basal</td>
<td>5.55</td>
<td>2.04</td>
<td>22</td>
</tr>
<tr>
<td>DRTA</td>
<td>6.23</td>
<td>2.09</td>
<td>22</td>
</tr>
<tr>
<td>Strategies</td>
<td>8.23</td>
<td>2.94</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>6.67</td>
<td>2.62</td>
<td>66</td>
</tr>
<tr>
<td><strong>GROUP</strong> Basal</td>
<td>41.05</td>
<td>5.64</td>
<td>22</td>
</tr>
<tr>
<td>DRTA</td>
<td>46.73</td>
<td>7.39</td>
<td>22</td>
</tr>
<tr>
<td>Strategies</td>
<td>44.27</td>
<td>5.77</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>44.02</td>
<td>6.64</td>
<td>66</td>
</tr>
</tbody>
</table>
A boxplot of the multiple dependent variables of our MANOVA can be easily created utilizing SPSS. The output is as follows:
It is obvious that there is an even spread within groups for each dependent variable. We can also perform a test of Homogeneity in order to know whether the variances for each of the groups are the same:

### Univariate Test Results

<table>
<thead>
<tr>
<th>Condition</th>
<th>Dependent Variable</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Eta Square</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PRE1</td>
<td>0.576</td>
<td>2</td>
<td>.209</td>
<td>.132</td>
<td>.329</td>
<td>.035</td>
<td>.264</td>
</tr>
<tr>
<td></td>
<td>PRE2</td>
<td>1.121</td>
<td>2</td>
<td>.561</td>
<td>.111</td>
<td>.895</td>
<td>.004</td>
<td>.223</td>
</tr>
<tr>
<td></td>
<td>POST1</td>
<td>3.121</td>
<td>2</td>
<td>1.061</td>
<td>.317</td>
<td>.007</td>
<td>.144</td>
<td>.635</td>
</tr>
<tr>
<td></td>
<td>POST2</td>
<td>5.485</td>
<td>2</td>
<td>1.274</td>
<td>.455</td>
<td>.001</td>
<td>.191</td>
<td>4.911</td>
</tr>
<tr>
<td></td>
<td>POST3</td>
<td>7.303</td>
<td>2</td>
<td>3.652</td>
<td>.481</td>
<td>.015</td>
<td>.125</td>
<td>3.962</td>
</tr>
<tr>
<td>Error</td>
<td>PRE1</td>
<td>2.455</td>
<td>63</td>
<td>9.087</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PRE2</td>
<td>7.136</td>
<td>63</td>
<td>5.034</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST1</td>
<td>0.500</td>
<td>63</td>
<td>0.167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST2</td>
<td>1.182</td>
<td>63</td>
<td>5.733</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST3</td>
<td>1.682</td>
<td>63</td>
<td>9.868</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Computed using alpha = .05
Univariate Homogeneity of Variance Tests

Variable .. POST1

Cochrans C(21,3) =  .50564, P =  .072 (approx.)
Bartlett-Box F(2,8930) =  1.86780, P =  .155

Variable .. POST3

Cochrans C(21,3) =  .45641, P =  .227 (approx.)
Bartlett-Box F(2,8930) =  .97142, P =  .379

Variable .. POST2

Cochrans C(21,3) =  .49688, P =  .090 (approx.)
Bartlett-Box F(2,8930) =  1.68980, P =  .185

Variable .. PRE2

Cochrans C(21,3) =  .50566, P =  .072 (approx.)
Bartlett-Box F(2,8930) =  1.91527, P =  .148

Variable .. PRE1

Cochrans C(21,3) =  .40980, P =  .538 (approx.)
Bartlett-Box F(2,8930) =  .48111, P =  .618
As seen through the Cochran’s C and the Bartlett Box F tests, the significance levels indicate that there is no reason to reject the hypotheses that the variances in the three groups are equal (all values are greater than .05).

These tests are UNIVARIATE and are a convenient starting point for examining homogeneity (covariance); however, we also need to simultaneously consider both the variances and the covariances. The “Box’s M” test is such a test. Box’s M is based on the determinants of the variance-covariance matrices in each cell, as well as the pooled variance-covariance matrix. Thus Box’s M provides us with a multivariate test for homogeneity (SPSS 9.0).

Multivariate test for Homogeneity of Dispersion matrices

<table>
<thead>
<tr>
<th>Box’s Test of Equality of Covariance Matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box's M</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>df1</td>
</tr>
<tr>
<td>df2</td>
</tr>
<tr>
<td>Sig.</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

1. Design: Intercept+GROUP

Dr. Robert Gebotys
February 2000
Chi-Square with 30 DF = 29.55742, P = .488 (Approx.)

Given the results of Box’s M test (either .493 or .488 depending on the F or Chi-square), we have no reason to suspect that homogeneity has been violated (values greater than .05).

**Bartlett’s Test of Sphericity**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>.000</td>
</tr>
<tr>
<td>Approx. Chi-Square</td>
<td>171.444</td>
</tr>
<tr>
<td>df</td>
<td>14</td>
</tr>
<tr>
<td>Sig.</td>
<td>.000</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the residual covariance proportional to an identity matrix.

1. Design: Intercept+GROUP

Multivariate analysis of variance is a procedure used when the dependent variables are correlated. The Bartlett’s Test of Sphericity is an option that is used to test the hypothesis that the population correlation matrix is an identity matrix (all diagonal terms are 1 and all off-diagonal terms are 0). The test is based
on the determinant of the error correlation matrix: a determinant that is close to 0 indicates that one or more variables are correlated. Thus, if the determinant is small, independence of the variables are rejected. As seen above, our determinant value is small (-1.17) with a significance value of .000. Therefore, we can reject the hypothesis of independence and conclude that our variables are correlated, hence the need for MANOVA.

Statistics for WITHIN+RESIDUAL correlations

Log(Determinant) = -1.17510
Bartlett test of sphericity = 71.09348 with 10 D. F.
Significance = .000

Another way to visualize variable independence is through the Half-normal plot of the correlation coefficients (see below). This plot is very similar to the normal plot except for the fact that positive and negative values are treated identically. This plot is used to test graphically the hypothesis that the correlation matrix is an identity matrix.

The Fischer’s Z-transform is the way in which the observed
correlation coefficients are transformed. If the correlation matrix is an identity matrix, then the plot should be linear and the line should pass through the origin. Our plot shows deviation from linearity, and thus suggests that the variables are not independent (as indicated by the Bartlett’s test).
### Multivariate Tests

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Eta Square</th>
<th>Concordant Observations</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.988</td>
<td>.990</td>
<td>5.000</td>
<td>9.000</td>
<td>.000</td>
<td>.988</td>
<td>4.952</td>
<td>1.000</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.012</td>
<td>.990</td>
<td>5.000</td>
<td>9.000</td>
<td>.000</td>
<td>.988</td>
<td>4.952</td>
<td>1.000</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>4.830</td>
<td>.990</td>
<td>5.000</td>
<td>9.000</td>
<td>.000</td>
<td>.988</td>
<td>4.952</td>
<td>1.000</td>
</tr>
<tr>
<td>Roy's Lambda</td>
<td>4.830</td>
<td>.990</td>
<td>5.000</td>
<td>9.000</td>
<td>.000</td>
<td>.988</td>
<td>4.952</td>
<td>1.000</td>
</tr>
<tr>
<td>GRO</td>
<td>.556</td>
<td>4.617</td>
<td>0.000</td>
<td>0.000</td>
<td>.000</td>
<td>.278</td>
<td>6.174</td>
<td>.999</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.510</td>
<td>4.725</td>
<td>0.000</td>
<td>8.000</td>
<td>.000</td>
<td>.286</td>
<td>7.248</td>
<td>.999</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>.832</td>
<td>4.828</td>
<td>0.000</td>
<td>6.000</td>
<td>.000</td>
<td>.294</td>
<td>8.280</td>
<td>.999</td>
</tr>
<tr>
<td>Roy's Lambda</td>
<td>.627</td>
<td>7.526</td>
<td>5.000</td>
<td>0.000</td>
<td>.000</td>
<td>.385</td>
<td>7.628</td>
<td>.999</td>
</tr>
</tbody>
</table>

1. Computed using alpha = .05
2. Exact statistic
3. The statistic is an upper bound on F that yields a lower bound on the significance of the effect.
4. Design: Intercept+GROUP

All four tests explore whether the means for each of the groups are the same. The first line contains the values of the parameters (S, M, N) used to discover significant levels in tables of the
exact distributions of the statistics. For the first three tests, the value of the test statistic is given, followed by its transformation to a statistic that has approximately an F distribution. The next two columns contain the numerator (hypothesis) and denominator (Error) degrees of freedom for the F statistic. The next column gives us the observed significance levels which are translated as the probability of observing a difference at least as large as the one found in the sample when there is no difference in the populations (SPSS 9.0). Note that the Roy’s test value cannot be transformed into a statistic and thus only the value is given. In our case, due to the significance values of .000 (sig column) we can conclude that the null hypothesis - that there is no difference is rejected. Therefore, we know that there are significant differences between the three groups on the means of the five variables.
Levene's Test of Equality of Error Variances

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE1</td>
<td>1.680</td>
<td>2</td>
<td>63</td>
<td>.195</td>
</tr>
<tr>
<td>PRE2</td>
<td>1.081</td>
<td>2</td>
<td>63</td>
<td>.345</td>
</tr>
<tr>
<td>POST1</td>
<td>2.748</td>
<td>2</td>
<td>63</td>
<td>.072</td>
</tr>
<tr>
<td>POST2</td>
<td>2.600</td>
<td>2</td>
<td>63</td>
<td>.082</td>
</tr>
<tr>
<td>POST3</td>
<td>.790</td>
<td>2</td>
<td>63</td>
<td>.458</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

1. Design: Intercept+GROUP
### Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Corrected Mean of Dependent Variable</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Eta Squared</th>
<th>Noncent Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE1</td>
<td>20.576^2</td>
<td>2</td>
<td>10.288</td>
<td>1.132</td>
<td>.329</td>
<td>.035</td>
<td>2.264</td>
<td>.241</td>
<td></td>
</tr>
<tr>
<td>PRE2</td>
<td>1.121^3</td>
<td>2</td>
<td>.561</td>
<td>.111</td>
<td>.895</td>
<td>.004</td>
<td>.223</td>
<td>.066</td>
<td></td>
</tr>
<tr>
<td>POST1</td>
<td>108.121^4</td>
<td>2</td>
<td>54.061</td>
<td>5.317</td>
<td>.007</td>
<td>.144</td>
<td>10.635</td>
<td>.821</td>
<td></td>
</tr>
<tr>
<td>POST2</td>
<td>85.485^5</td>
<td>2</td>
<td>42.742</td>
<td>7.455</td>
<td>.001</td>
<td>.191</td>
<td>14.911</td>
<td>.932</td>
<td></td>
</tr>
<tr>
<td>POST3</td>
<td>357.303^6</td>
<td>2</td>
<td>178.652</td>
<td>4.481</td>
<td>.015</td>
<td>.125</td>
<td>8.962</td>
<td>.747</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intercept</th>
<th>Dependent Variable</th>
<th>TYPE III</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Eta Squared</th>
<th>Noncent Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE1</td>
<td>322.970</td>
<td>2</td>
<td>322.970</td>
<td>1</td>
<td>395.858</td>
<td>.000</td>
<td>.917</td>
<td>695.858</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>PRE2</td>
<td>720.742</td>
<td>2</td>
<td>720.742</td>
<td>1</td>
<td>341.830</td>
<td>.000</td>
<td>.844</td>
<td>431.830</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>POST1</td>
<td>304.379</td>
<td>2</td>
<td>304.379</td>
<td>1</td>
<td>423.382</td>
<td>.000</td>
<td>.870</td>
<td>423.382</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>POST2</td>
<td>933.333</td>
<td>2</td>
<td>933.333</td>
<td>1</td>
<td>511.654</td>
<td>.000</td>
<td>.890</td>
<td>511.654</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>POST3</td>
<td>864.015</td>
<td>2</td>
<td>864.015</td>
<td>1</td>
<td>207.187</td>
<td>.000</td>
<td>.981</td>
<td>320.7187</td>
<td>.000</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GROUP</th>
<th>Dependent Variable</th>
<th>TYPE III</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Eta Squared</th>
<th>Noncent Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE1</td>
<td>20.576</td>
<td>2</td>
<td>10.288</td>
<td>1</td>
<td>1.132</td>
<td>.329</td>
<td>.035</td>
<td>2.264</td>
<td>.241</td>
<td></td>
</tr>
<tr>
<td>PRE2</td>
<td>1.121</td>
<td>2</td>
<td>.561</td>
<td>1</td>
<td>.111</td>
<td>.895</td>
<td>.004</td>
<td>.223</td>
<td>.066</td>
<td></td>
</tr>
<tr>
<td>POST1</td>
<td>108.121</td>
<td>2</td>
<td>54.061</td>
<td>1</td>
<td>5.317</td>
<td>.007</td>
<td>.144</td>
<td>10.635</td>
<td>.821</td>
<td></td>
</tr>
<tr>
<td>POST2</td>
<td>85.485</td>
<td>2</td>
<td>42.742</td>
<td>1</td>
<td>7.455</td>
<td>.001</td>
<td>.191</td>
<td>14.911</td>
<td>.932</td>
<td></td>
</tr>
<tr>
<td>POST3</td>
<td>357.303</td>
<td>2</td>
<td>178.652</td>
<td>1</td>
<td>4.481</td>
<td>.015</td>
<td>.125</td>
<td>8.962</td>
<td>.747</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error</th>
<th>Dependent Variable</th>
<th>TYPE III</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Eta Squared</th>
<th>Noncent Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE1</td>
<td>572.455</td>
<td>63</td>
<td>9.087</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRE2</td>
<td>317.136</td>
<td>63</td>
<td>5.034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST1</td>
<td>640.500</td>
<td>63</td>
<td>10.167</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST2</td>
<td>361.182</td>
<td>63</td>
<td>5.733</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST3</td>
<td>511.682</td>
<td>63</td>
<td>39.868</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th>Dependent Variable</th>
<th>TYPE III</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Eta Squared</th>
<th>Noncent Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE1</td>
<td>916.000</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRE2</td>
<td>2039.000</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST1</td>
<td>5053.000</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST2</td>
<td>380.000</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST3</td>
<td>733.000</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corrected Total</th>
<th>Dependent Variable</th>
<th>TYPE III</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Eta Squared</th>
<th>Noncent Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE1</td>
<td>593.030</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRE2</td>
<td>318.258</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST1</td>
<td>748.621</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST2</td>
<td>446.667</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POST3</td>
<td>868.985</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Computed using alpha = .05
2. R Squared = .035 (Adjusted R Squared = .004)
3. R Squared = .004 (Adjusted R Squared = -.028)
4. R Squared = .144 (Adjusted R Squared = .117)
5. R Squared = .191 (Adjusted R Squared = .166)
6. R Squared = .125 (Adjusted R Squared = .097)
The table below contains several sets of statistics for the discriminant functions. The eigenvalue represents the ratio of the Between-Groups sum of squares to the Within-Groups sum of squares. We see that our examples consists of two functions that contribute to group differences. The first function has an eigenvalue of .630 and explains 73% of the differences in the three groups (column: cumulative percentage). The final column gives the Canonical Correlation, and the square of this value is the ratio of the Between-groups sum of squares to the Total sum of squares. Therefore about 38.6% (.622^2) of the variability in the first discriminant score is attributable to between-group differences.

Eigenvalues and Canonical Correlations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.630</td>
<td>73.853</td>
<td>73.853</td>
<td>.622</td>
</tr>
<tr>
<td>2</td>
<td>.223</td>
<td>26.147</td>
<td>100.000</td>
<td>.427</td>
</tr>
</tbody>
</table>

Wilks’ Lambda can be seen as a measure of the proportion
of total variability not explained by group differences. In our case (see below) both functions are significant (last column-significance values less than .05). The second and third columns represent the numerator and denominator degrees of freedom for the F statistic. In sum, we can conclude that both functions are distinct and significant in discriminating amongst the groups.

Dimension Reduction Analysis

<table>
<thead>
<tr>
<th>Roots</th>
<th>Wilks L.</th>
<th>F</th>
<th>Hypoth. DF</th>
<th>Error DF</th>
<th>Sig. of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 TO 2</td>
<td>.50177</td>
<td>4.85822</td>
<td>10.00</td>
<td>118.00</td>
<td>.000</td>
</tr>
<tr>
<td>2 TO 2</td>
<td>.81771</td>
<td>3.34388</td>
<td>4.00</td>
<td>60.00</td>
<td>.015</td>
</tr>
</tbody>
</table>

-------------------------------------------
## DATA OUTPUT – Custom Hypothesis Tests

### Contrast Results (K Matrix)

<table>
<thead>
<tr>
<th>GROUP Simple Contrast</th>
<th>Level 2 vs. Level 1</th>
<th>Contrast Estimate</th>
<th>Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PRE1</td>
<td>PRE2</td>
</tr>
<tr>
<td></td>
<td>Contrast Estimate</td>
<td>-.773</td>
<td>-.182</td>
</tr>
<tr>
<td></td>
<td>Hypothesized Value</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Difference (Estimate - Hypothesized)</td>
<td>-.773</td>
<td>-.182</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.909</td>
<td>.676</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.398</td>
<td>.789</td>
</tr>
<tr>
<td></td>
<td>95% Confidence Interval for Difference</td>
<td>Lower Bound</td>
<td>-.258</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Upper Bound</td>
<td>1.044</td>
</tr>
<tr>
<td></td>
<td>Level 3 vs. Level 1</td>
<td>Contrast Estimate</td>
<td>-.1364</td>
</tr>
<tr>
<td></td>
<td>Hypothesized Value</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Difference (Estimate - Hypothesized)</td>
<td>-.1364</td>
<td>-.318</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.909</td>
<td>.676</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.139</td>
<td>.640</td>
</tr>
<tr>
<td></td>
<td>95% Confidence Interval for Difference</td>
<td>Lower Bound</td>
<td>-3.180</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Upper Bound</td>
<td>.453</td>
</tr>
</tbody>
</table>

1: Reference category = 1

### Multivariate Test Results

<table>
<thead>
<tr>
<th>Pillai's trace</th>
<th>Wilks' lam</th>
<th>Hotelling's</th>
<th>Roy's larg</th>
</tr>
</thead>
<tbody>
<tr>
<td>.556</td>
<td>.510</td>
<td>.832</td>
<td>.627</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value</th>
<th>F</th>
<th>df</th>
<th>df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Eta Squared</th>
<th>Noncentrality Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>.556</td>
<td>4.617</td>
<td>10.000</td>
<td>0.000</td>
<td>.000</td>
<td>.278</td>
<td>.46.174</td>
<td>.999</td>
<td></td>
</tr>
<tr>
<td>.510</td>
<td>4.725</td>
<td>10.000</td>
<td>8.000</td>
<td>.000</td>
<td>.286</td>
<td>.47.248</td>
<td>.999</td>
<td></td>
</tr>
<tr>
<td>.832</td>
<td>4.828</td>
<td>10.000</td>
<td>6.000</td>
<td>.000</td>
<td>.294</td>
<td>.48.280</td>
<td>.999</td>
<td></td>
</tr>
<tr>
<td>.627</td>
<td>7.526</td>
<td>5.000</td>
<td>0.000</td>
<td>.000</td>
<td>.385</td>
<td>.37.628</td>
<td>.999</td>
<td></td>
</tr>
</tbody>
</table>

1: Computed using alpha = .05
2: Exact statistic
3: The statistic is an upper bound on F that yields a lower bound on t.
Since our multivariate results were significant (remember the Pillais, Hotellings, Wilks and Roys tests above), we can now examine our univariate results shown below:

**EFFECT .. GROUP (Cont.)**

Univariate F-tests with (2,63) D. F.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypoth. SS</th>
<th>Error SS</th>
<th>Hypoth. MS</th>
<th>Error MS</th>
<th>F</th>
<th>Sig. of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST1</td>
<td>108.1212</td>
<td>640.5000</td>
<td>54.0606</td>
<td>10.1667</td>
<td>5.31744</td>
<td>.007</td>
</tr>
<tr>
<td>POST3</td>
<td>357.3030</td>
<td>2511.682</td>
<td>178.6515</td>
<td>39.86797</td>
<td>4.48108</td>
<td>.015</td>
</tr>
<tr>
<td>POST2</td>
<td>95.1212</td>
<td>356.4091</td>
<td>47.5606</td>
<td>5.65729</td>
<td>8.40696</td>
<td>.001</td>
</tr>
<tr>
<td>PRE2</td>
<td>1.1212</td>
<td>317.136</td>
<td>5.5606</td>
<td>5.03391</td>
<td>.11137</td>
<td>.895</td>
</tr>
<tr>
<td>PRE1</td>
<td>20.5776</td>
<td>572.455</td>
<td>10.288</td>
<td>9.08658</td>
<td>1.13221</td>
<td>.329</td>
</tr>
</tbody>
</table>

This table helps determine which variables contribute to the overall differences. In our case, we can see that the first three variables: Post1, Post 3, and Post2 all yield significant values (p
less than .05). It also makes sense that both Pretest 1 and Pretest 2 have no influence upon the three groups, since pretests are in fact supposed to NOT influence respondents.

Lastly, the table below contains the standardized discriminant function coefficients for the group term. The magnitude of the coefficients give us some idea of the variables contributing most to group differences. In our case, we can see that Post1 (.963), Post2 (.709), and Pre1 (-.867) have large coefficients and thus could play an important role in group differences within the first function. In the second function, Post1 and Post 2 both yield large coefficients.

### Standardized discriminant function coefficients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Function No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>POST1</td>
<td>.963</td>
</tr>
<tr>
<td>POST3</td>
<td>.194</td>
</tr>
<tr>
<td>POST2</td>
<td>.709</td>
</tr>
<tr>
<td>PRE2</td>
<td>-.379</td>
</tr>
<tr>
<td>PRE1</td>
<td>-.867</td>
</tr>
</tbody>
</table>
In addition, we can get a sense of the spread of group means of each discriminant function (in our case, we are interested in both functions) which was derived from the discriminant analysis procedure:

Canonical discriminant functions evaluated at group means (group centroids)

<table>
<thead>
<tr>
<th>Group</th>
<th>Func 1</th>
<th>Func 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.09294</td>
<td>.05163</td>
</tr>
<tr>
<td>2</td>
<td>.47132</td>
<td>-.58901</td>
</tr>
<tr>
<td>3</td>
<td>.62162</td>
<td>.53738</td>
</tr>
</tbody>
</table>

***

In summary, let’s return to our initial discussion around the similarities between MANOVA and discriminant analysis. As noted earlier, these two procedures are much alike, and here is
the test: I ran the same data through a discriminant analysis and look at the results. Do they look familiar at all?

### Discriminant Results on Same Data:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Wilks' Lambda</th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST1</td>
<td>.85557</td>
<td>5.3174</td>
<td>2</td>
<td>63</td>
<td>.0073</td>
</tr>
<tr>
<td>POST2</td>
<td>.78934</td>
<td>8.4070</td>
<td>2</td>
<td>63</td>
<td>.0006</td>
</tr>
<tr>
<td>POST3</td>
<td>.87546</td>
<td>4.481</td>
<td>2</td>
<td>63</td>
<td>.0152</td>
</tr>
<tr>
<td>PRE1</td>
<td>.96530</td>
<td>1.1322</td>
<td>2</td>
<td>63</td>
<td>.3288</td>
</tr>
<tr>
<td>PRE2</td>
<td>.99648</td>
<td>.1114</td>
<td>2</td>
<td>63</td>
<td>.8948</td>
</tr>
</tbody>
</table>

Wilks' Lambda (U-statistic) and univariate F-ratio with 2 and 63 degrees of freedom

Variable | Wilks' Lambda | F       | Significance |
----------|---------------|---------|--------------|
POST1     | .85557        | 5.3174  | .0073        |
POST2     | .78934        | 8.4070  | .0006        |
POST3     | .87546        | 4.481   | .0152        |
PRE1      | .96530        | 1.1322  | .3288        |
PRE2      | .99648        | .1114   | .8948        |

see page 8 for MANOVA output.

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Eigenvalue</th>
<th>% of Variance</th>
<th>Cumulative %</th>
<th>Canonical Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.6271</td>
<td>75.3</td>
<td>75.0</td>
<td>.621</td>
</tr>
<tr>
<td>Function 1</td>
<td>.2051</td>
<td>24.7</td>
<td>100.0</td>
<td>.413</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test of Function(s)</th>
<th>Wilks' Lambda</th>
<th>Chi-square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 through 2</td>
<td>.510</td>
<td>41.085</td>
<td>10</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>.830</td>
<td>11.389</td>
<td>4</td>
<td>.023</td>
</tr>
</tbody>
</table>
Canonical Discriminant Functions

<table>
<thead>
<tr>
<th>Pct of</th>
<th>Cum Canonical After Wilks'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fcn</td>
<td>Eigenvalue</td>
</tr>
<tr>
<td>1*</td>
<td>.6296</td>
</tr>
<tr>
<td>2*</td>
<td>.2229</td>
</tr>
</tbody>
</table>

* Marks the 2 canonical discriminant functions remaining in the analysis.

See page 7 for MANOVA output.

<table>
<thead>
<tr>
<th>standardized Canonical Discriminant Function Coefficien</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>PRE1</td>
</tr>
<tr>
<td>PRE2</td>
</tr>
<tr>
<td>POST1</td>
</tr>
<tr>
<td>POST2</td>
</tr>
<tr>
<td>POST3</td>
</tr>
</tbody>
</table>
Standardized canonical discriminant function coefficients

<table>
<thead>
<tr>
<th></th>
<th>Func 1</th>
<th>Func 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST1</td>
<td>.96336</td>
<td>-.81633</td>
</tr>
<tr>
<td>POST2</td>
<td>.70901</td>
<td>.73923</td>
</tr>
<tr>
<td>POST3</td>
<td>.19358</td>
<td>.00261</td>
</tr>
<tr>
<td>PRE1</td>
<td>-.86697</td>
<td>.25010</td>
</tr>
<tr>
<td>PRE2</td>
<td>-.37896</td>
<td>-.00682</td>
</tr>
</tbody>
</table>

see page 9 for MANOVA output.

We see that there are two functions, both are significant in examining group differences. With our univariate F-tests, we identify the same significant variables. And with our standardized canonical discriminate function coefficients, we see the same variables having strength (or significance) with each particular function. This is the identical test to MANOVA.

In conclusion, MANOVA is a procedure utilized when our dependent variables are correlated. Multivariate analysis helps control the error rate, or the chance of getting something significant when in actuality it is not. Very similar in process to discriminant analysis, MANOVA permits a direct test of the
null hypothesis with respect to ALL of the analysis’ dependent variables.