

One-Way Multiple Analysis of Variance[©]

When one wishes to determine whether two or more groups differ significantly on one or more optimally weighted linear combinations (**canonical variates** or discriminant functions) of two or more normally distributed dependent variables, a one-way multiple analysis of variance is appropriate. We have already studied Discriminant Function Analysis, which is mathematically equivalent to a one-way MANOVA, so we are just shifting perspective here.

A Manipulation Check for Michelle Plaster's Thesis

Download Plaster.dat from my [StatData page](#) and MANOVA1.sas from my [SAS Programs page](#). Run the program. The data are from Michelle Plaster's thesis, which you may wish to look over (in Joyner or in the Psychology Department). The analyses she did are very similar to, but not identical to, those we shall do for instructional purposes. Male participants were shown a picture of one of three young women. Pilot work had indicated that the one woman was beautiful, another of average physical attractiveness, and the third unattractive. Participants rated the woman they saw on each of twelve attributes. Those ratings are our dependent variables. To simplify the analysis, we shall deal with only the ratings of physical attractiveness (PHYATTR), happiness (HAPPY), INDEPENDence, and SOPHISTication. The purpose of the research was to investigate the effect of the defendant's physical attractiveness (and some other variables which we shall ignore here) upon the sentence recommended in a simulated jury trial. The MANOVA on the ratings served as a check on the effectiveness of our manipulation of the physical attractiveness (PA) independent variable.

Screening the Data

The overall means and standard deviations on the dependent variables were obtained so that I could standardize them and then compute scores on the canonical variates. Although you can not see it in this output, there is **heterogeneity of variance** on the physical attractiveness variable. The F_{max} is 4.62. With n 's approximately equal, I'm not going to worry about it. If F_{max} were larger and sample sizes less nearly equal, I might try data transformations to stabilize the variance or I might randomly discard scores to produce equal n 's and thus greater robustness.

One should look at the distributions of the dependent variables within each cell. I have done so, (using SAS), but have not provided you with the statistical output. Such output tends to fill many pages, and I generally do not print it, I just take notes from the screen. Tests of significance employed in MANOVA do assume that the sampling distributions of the means of the dependent variables and linear combinations

[©] Copyright 2006 Karl L. Wuensch - All rights reserved.

of the dependent variables are **normal** within each cell. With large sample sizes (or small samples with approximately equal n 's), the tests may be fairly robust to violations of this assumption. I must confess that there are some problems with the normality assumption for these data, especially the phyattr variable. It is negatively skewed in one cell, positively skewed in another, due to ceiling and floor effects. I was unable to find anything in my bag of tricks (nonlinear transformations and deletion of outliers) that ameliorated this problem to my satisfaction, so I decided to stick with the data as they are, with the caution that they are not exactly normal.

The MANOVA with Canonical Analysis

PROC ANOVA is employed to do the multiple analysis of variance. Note that there are multiple dependent variables listed in the model statement. The **NOUNI** option is used to suppress output of the univariate analyses of variance.

MANOVA H = PA / CANONICAL indicates that we want to do a MANOVA test of the null hypothesis that there is no effect of our physical attractiveness manipulation and that we want statistics about the canonical variates.

An effect will have as many **roots (discriminant functions, canonical variates)** as it has degrees of freedom, or it will have as many roots as there are DVs, whichever is fewer. Our manipulation has 3 levels, 2 df , so we have two roots. The first root is

that weighted combination of the ratings that maximizes the ratio $\frac{SS_{among_groups}}{SS_{within_groups}}$. These

sums of squares are those that would be obtained were we to do a univariate ANOVA using our manipulation as the IV and the weighted linear combination of our four ratings variables as the single DV. This ratio is called the **eigenvalue**. This is a very context-dependent definition of an eigenvalue, but, given our reliance on the computer and our usually avoiding matrix algebra, I think it appropriate to leave it at that. For our data, it is 1.767. The second root, orthogonal to the first, is that weighted linear combination of the dependent variables that maximizes the eigenvalue using variance in the variables not already captured by the earlier root(s). The Proportion .9133 for the first root is simply the first root's eigenvalue expressed as a percentage of the simple sum of all the roots' eigenvalues.

SAS gives us, along with the eigenvalues, tests of the significance of the roots. The first row gives us the test of the null hypothesis that all of the canonical variates have zero correlations with the independent variable. The second row tests the null hypothesis that all of the canonical variates from the second on have zero correlations with the independent variable, etc. As you can see, both of our canonical variates are significantly affected by our physical attractiveness manipulation.

Were we to do a univariate ANOVA on the first canonical variate (weighted linear combination of the dependent variables), $\frac{SS_{among_groups}}{SS_{total}}$ would be the first

squared canonical correlation. You should recognize this ratio as being η^2 . For our

data, η^2 , the squared canonical correlation, for the first root is .64, that is, 64% of the variance in the first canonical variate is accounted for by our manipulation. For the second canonical variate it is 14%.

We found that our groups differ significantly on two dimensions. Let us now look at the standardized discriminant function coefficients and the correlations between DVs and canonical variates to interpret those dimensions.

The **standardized canonical coefficients** may be treated as Beta weights in a multiple regression predicting C from Z-scores on the X's, where C is the canonical variate, the weighted linear combination of the variables which maximizes the eigenvalue.

$$C_i = B_1 Z_1 + B_2 Z_2 + \dots + B_p Z_p.$$

SAS has reported here "**total-sample standardized canonical correlation coefficients**" -- these are the coefficients obtained if the Z scores above are computed by using, for each variable, its mean and standard deviation ignoring groups (rather than within groups). Of course, one must realize that these coefficients reflect the contribution of one variable in the context of the other variables in the model. A low standardized coefficient might mean that the groups do not differ much on that variable, or it might just mean that that variable's correlation with the grouping variable is redundant with that of another variable in the model. Suppressor effects can also occur (they do in these data, on the second function, with sophistication having a β which is higher than is its simple correlation with the second canonical variate). Raw (unstandardized) coefficients are presented as well, but are rarely of any use.

Loadings (correlations between the dependent variables and the canonical variates) may be more helpful in terms of understanding the canonical variates we have created. These are available under the heading "Canonical Structure." I am accustomed to using the "Within" loadings, which are computed by computing the correlations within each group and then averaging them across groups. Generally, any variate with a loading of .40 or more must be considered to be important. One might apply a "meaningfulness criterion" with loadings between .30 and .40 -- that is, if they make sense, interpret them, if not, declare them too small to be worthy of your attention. ☺

Our loadings and our standardized coefficients indicate that canonical variate 1 is largely a matter of physical attractiveness and canonical variate 2 is largely Happiness and Independence. The role of the Sophistication rating is less clear. Neither of its loadings is high and it is involved in some sort of suppressor effect on the second canonical variate. It might be interesting to drop this variable and redo the analysis.

Following the canonical statistics, SAS gives us four different tests of the null hypothesis that none of the canonical variates is correlated with our treatment.

Hotelling's trace is the simple sum of all the roots' eigenvalues.

If we were ambitious (or foolish) enough to work out our MANOVA by hand (using matrix algebra, which is very arithmetic intensive, yuk), we would calculate treatment (among groups) and error (within groups) *SSCP* (sums of squares and cross products) matrices (like variance-covariance matrices, but with sums of squares on the main diagonal and cross products elsewhere -- dividing the elements by degrees of freedom would produce variances and covariances). Since we have 4 DVs, each would be a 4 x 4 matrix. In univariate ANOVA we test an effect by computing F , the ratio of a treatment variance (mean square) to error variance. In MANOVA we may test our treatment by finding the ratio of the determinant (generalized variance) of the error *SSCP* matrix to the determinant of the sum of the treatment and error *SSCP* matrices.

This ratio is called **Wilks' Lambda (Λ)**. Since the ratio is $\frac{\text{error}}{\text{error} + \text{treatment}}$, one may interpret the Wilks' Lambda as the proportion of variance in the DVs that is not accounted for by the IV. Unlike F , where large values cast doubt on the null hypothesis that the treatment groups do not differ, small values of Wilks' Lambda cast doubt on the null hypothesis. An **eta-squared** statistic can be computed as $\eta^2 = 1 - \Lambda$, .69 for our data. We may interpret this statistic as being the proportion of variance in the dependent variables that is accounted for by the physical attractiveness manipulation. When there are only two treatment groups, this η^2 will equal the squared canonical correlation.

Lambda (Λ) is usually used to stand for an eigenvalue. If we define theta as $\theta = \frac{\Lambda}{\Lambda + 1}$, then Wilks' Λ can be computed from these θ s. For our data, θ for the first root = $1.767/2.767 = 0.6386$, and, for the second root, $.168/1.168 = 0.1438$. **Wilks' Λ** is the product of all the $(1 - \theta)$ s, $(1 - .6386)(1 - .1438) = .309$.

Pillai's Trace is the sum of all the θ s, $.6386 + .1438 = .782$.

Roy's greatest characteristic root is simply the largest θ . Roy's gcr should be the most powerful test when the first root is much larger than the other roots.

Each of these statistics' significance levels is approximated using F .

Univariate ANOVAs on the Canonical Variates and the Original Variables

Now, look back at the data step of our program. I used the total-sample means and standard deviations to standardize the variables into Z _scores. I then computed, for each subject, canonical variate scores, using the total-sample standardized canonical correlation coefficients. **CV1** is canonical variate 1, **CV2** is canonical variate 2. I computed these canonical variate scores so I could use them as dependent variables in univariate analyses of variance. I used PROC ANOVA to do univariate ANOVAs with Fisher's LSD contrasts on these two canonical variates and also, for the benefit of those who just cannot understand canonical variates, on each rating variable.

Please note that for each canonical variate:

- If you take the $SS_{\text{among groups}}$ and divide by the $SS_{\text{within groups}}$, you get the **eigenvalue** reported earlier for that root.
- If you take the ratio of $SS_{\text{among groups}}$ to SS_{total} , you get the root's **squared canonical correlation**.
- The group means are the group **centroids** that would be reported with a discriminant function analysis.

I have never seen anyone else compute canonical scores like this and then do pairwise comparisons on them, but it seems sensible to me, and such analysis has been accepted in manuscripts I have published. As an example of this procedure, see my summary of an article on [Mock Jurors' Insanity Defense Verdict Selections](#).

Note that on the "beauty" canonical variate, the beautiful group's mean is significantly higher than the other two means, which are not significantly different from one another. On the "happily independent" variate, the average group is significantly higher than the other two groups.

The univariate ANOVAs reveal that our manipulation significantly affected every one of the ratings variables, with the effect on the physical attractiveness ratings being very large ($\eta^2 = .63$). Compared to the other groups, the beautiful woman was rated significantly more attractive and sophisticated, and the unattractive woman was rated significantly less independent and happy.

Unsophisticated Use of MANOVA

Unsophisticated users of MANOVA usually rely on the **Univariate F-tests** to try to interpret a significant MANOVA. These are also the same users who believe that the purpose of a MANOVA is to guard against **inflation of alpha across several DVs**. They promise themselves that they will not even look at the univariate ANOVAs for any effect which is not significant in the MANOVA. The logic is essentially the same as that in the Fisher's LSD protected test for making pairwise comparisons between means following a significant effect in ANOVA (this procedure has a lousy reputation -- not conservative enough -- these days, but is actually a good procedure when only three means are being compared; the procedure can also be generalized to other sort of analyses, most appropriately those involving effects with only two degrees of freedom). In fact, this "protection" is afforded only if the overall null hypothesis is true (none of the DVs is affected by the IV), not if some DVs are affected by the IV and others are not. If one DV is very strongly affected by the IV and another very weakly (for practical purposes, zero effect), the probability of finding the very weak effect to be "significant" is unacceptably high. Were we so unsophisticated as to take such a "MANOVA-protected univariate tests" approach, we would note that the MANOVA was significant, that univariate tests on Physically Attractive, Happy, Independent, and Sophisticated were significant, and then we might do some pairwise comparisons between groups on each of these "significant" dependent variables. I have included such univariate tests in my second invocation of PROC ANOVA. Persons who are

thinking of using MANOVA because they have an obsession about inflated familywise alpha and they think MANOVA somehow protects against this should consider another approach -- application of a **Bonferroni adjustment** to multiple univariate ANOVAs. That is, dispense with the MANOVA, do the several univariate ANOVAs, and evaluate each ANOVA using an adjusted criterion of significance equal to the familywise alpha you can live with divided by the number of tests in the family. For example, if I am willing to accept only a .05 probably of rejecting one or more of four true null hypotheses (four dependent variables, four univariate ANOVAs), I used an adjusted criterion of $.05/4 = .0125$. For each of the ANOVAs I declare the effect to be significant only if its $p \leq .0125$. This will, of course, increase the probability of making a Type II error (which is already the most likely sort of error), so I am not fond of making the Bonferroni adjustment of the per-comparison criterion of statistical significance. Read more on this in my document [Familywise Alpha and the Boogey Men](#).

While I have somewhat disparaged the “protected test” test use of MANOVA, I must confess that I sometimes employ it, especially in complex factorial designs where, frankly, interpreting the canonical variates for higher order effects is a bit too challenging for me. With a complex factorial analysis (usually from one of my colleagues, I usually having too much sense to embark on such ambitious projects) I may simply note which of the effects are significant in the MANOVA and then use univariate analyses to investigate those (and only those) effects in each of the dependent variables. Aside from any protection against inflating familywise alpha, the main advantage of this approach is that it may impress some reviewers of your manuscript, multivariate analyses being popular these days.

This unsophisticated approach just described ignores the correlations among the dependent variables and ignores the dimensions (canonical variates, discriminant functions) upon which MANOVA found the groups to differ. This unsophisticated user may miss what is really going on -- it is quite possible for none of the univariate tests to be significant, but the MANOVA to be significant.

Relationship between MANOVA and DFA

I also conducted a discriminant function analysis on these data just to show you the equivalence of MANOVA and DFA. Please note that the eigenvalues, canonical correlations, loadings, and canonical coefficients are identical to those obtained with the MANOVA.

SPSS MANOVA

SPSS has two routines for doing multiple analysis of variance, the GLM routine and the MANOVA routine. Let us first consider the GLM routine. Go to my [SPSS Data Page](#) and download the SPSS data file, Plaster.sav. Bring it into SPSS and click Analyze, General Linear Model, Multivariate. Move the dependents (phyattr, happy, indepen, and sophist) into the dependent variables box and the grouping variable (PA, manipulated physical attractiveness) into the fixed factor box. Under Options, select

descriptive statistics and homogeneity tests. Now, look at the output. Note that you do get Box's M , which is not available in SAS. You also get the tests of the significance of all roots simultaneously tested and Roy's test of only the first root, as in SAS. Levene's test is used to test the significance of group differences in variance on the original variables. Univariate ANOVAs are presented, as are basic descriptive statistics. Missing, regrettably, are any canonical statistics, and these are not available even if you select every optional statistic available with this routine. What a bummer. Apparently the folks at SPSS have decided that people that can only point and click will never do a truly multivariate analysis of variance, that they only use what I have called the unsophisticated approach, so they have made the canonical statistics unavailable.

Fortunately, the canonical statistics are available from the SPSS MANOVA routine. While this routine was once available on a point-and-click basis, it is now available only by syntax -- that is, you must enter the program statements in plain text, just like you would in SAS, and then submit those statements via the SPSS Syntax Editor. Point your browser to my [SPSS programs page](#) and download the file MANOVA1.sps to your hard drive or diskette. From the Windows Explorer or My Computer, double click on the file. The SPSS Syntax Editor opens with the program displayed. Click on Run, All, and the output will appear. Look at the output. If you find a downwards pointing red triangle at the bottom of the displayed output, there is more than what you can see. To see all of the output, click in the output box, so the borders of the box appear. Click on and drag downwards the resizing icon in the middle of the bottom border.

I did not ask SPSS to print the variance/covariance matrices for each cell, but I did get the determinants of these matrices, which are used in Box's M . You may think of the determinants as being indices of the generalized variance within a variance/covariance matrix. For each cell, for our 4 dependent variable design, that matrix is a 4 x 4 matrix with variances on the main diagonal and covariances (between each pair of DVs) off the main diagonal. **Box's M** tests the null hypothesis that the variance/covariance matrices in the population are identical across cells. If this null hypothesis is false, the pooled variance/covariance matrix used by SPSS is inappropriate. Box's M is notoriously powerful, and one generally doesn't worry unless $p < .001$ and sample sizes unequal. Using Pillai's trace (rather than Wilks' Λ) may improve the robustness of the test in these circumstances. One can always randomly discard scores to produce equal n 's. Since our n 's are nearly equal, I'll just use Pillai's trace and not discard any scores.

Look at the pooled within cells (thus eliminating any influence of the grouping variable) correlation matrix – it is now referred to as the WITHIN+RESIDUAL correlation matrix. The dependent variables are generally correlated with one another. **Bartlett's test of sphericity** tests the null hypothesis that in the population the correlation matrix for the DVs is an identity matrix (each $r_{jj} = 0$). That is clearly not the case with our DVs. Bartlett's test is based on the determinant of the within-cells correlation matrix. If the determinant is small, the null hypothesis is rejected. If the determinant is very small (zero to several decimal places), then at least one of the

variables is nearly a perfect linear combination of the other variables. This creates a problem known as **multicollinearity**. With multicollinearity your solution is suspect -- another random sample from the same population would be likely to produce quite different results. When this problem arises, I recommend that you delete one or more of the variables. If one of your variables is a perfect linear combination of the others (for example, were you to include as variables SAT_{Total} , SAT_{Math} and SAT_{Verbal}), the analysis would crash, due to the **singularity** of a matrix which needs to be inverted as part of the solution. If you have a multicollinearity problem but just must retain all of the DVs, you can replace your DVs with **principal component scores**. I recently had this problem with a data set on which I was doing a discriminant function analysis. I used SAS' PROC FACTOR to extract 8 orthogonal (independent) components from the set of 8 variables, outputting standardized component scores for each subject. The discriminant analysis was then done on these component scores. I used a little transformation magic so I could interpret the results with respect to the original variables rather than the component scores. I computed for each subject a score on each of the two discriminant functions (by multiplying the standardized discriminant function coefficients by the standardized component scores). Then I obtained loadings (correlations between the dependent and the canonical/discriminant variables) by correlating subjects' scores on the two discriminant functions with their scores on the original variables.

For our data, SPSS tells us that the **determinant of the within cells correlation matrix** has a log of $-.37725$. Using my calculator, $\text{natural log } (.689) = -.37725$ -- that is, the determinant is $.689$, not near zero, apparently no problem with multicollinearity. Another way to check on multicollinearity is to compute the R^2 between each variable and all the others (or $1-R^2$, the tolerance). This will help you identify which variable you might need to delete to avoid the multicollinearity problem. I used SAS' PROC FACTOR to get the R^2 's, which are:

Prior Communality Estimates: SMC

PHYATTR	HAPPY	INDEPEN	SOPHIST
0.136803	0.270761	0.133796	0.289598

Note that the MANOVA routine has given us all of the canonical statistics that we are likely to need to interpret our multivariate results. If you compare the canonical coefficients and loadings to those obtained with SAS, you will find that each SPSS coefficient equals minus one times the SAS coefficient. While SAS constructed canonical variates measuring physical attractiveness (CV1) and happiness/independence (CV2), SPSS constructed canonical variates measuring physical unattractiveness and unhappiness/dependence. The standardized coefficients presented by SPSS are computed using within group statistics.