Multivariate Analysis of Variance (MANOVA): I. Theory

Introduction

The purpose of a *t* test is to assess the likelihood that the means for two groups are sampled from the same sampling distribution of means. The purpose of an ANOVA is to test whether the means for two or more groups are taken from the same sampling distribution. The multivariate equivalent of the *t* test is Hotelling's T^2. Hotelling's T^2 tests whether the two *vectors* of means for the two groups are sampled from the same sampling distribution. MANOVA is the multivariate analogue to Hotelling's T^2. The purpose of MANOVA is to test whether the *vectors* of means for the two or more groups are sampled from the same sampling distribution. Just as Hotelling's T^2 will provide a measure of the likelihood of picking two random vectors of means out of the same hat, MANOVA gives a measure of the overall likelihood of picking *two or more* random vectors of means out of the same hat.

There are two major situations in which MANOVA is used. The first is when there are several correlated dependent variables, and the researcher desires a single, overall statistical test on this set of variables instead of performing multiple individual tests. The second, and in some cases, the more important purpose is to explore how independent variables influence some patterning of response on the dependent variables. Here, one literally uses an analogue of contrast codes on the dependent variables to test hypotheses about how the independent variables differentially predict the dependent variables.

MANOVA also has the same problems of multiple post hoc comparisons as ANOVA. An ANOVA gives one overall test of the equality of means for several groups for a single variable. The ANOVA will not tell you which groups differ from which other groups. (Of course, with the judicious use of a priori contrast coding, one can overcome this problem.) The MANOVA gives one overall test of the equality of mean *vectors* for several groups. But it cannot tell you which groups differ from which other groups on their mean *vectors*. (As with ANOVA, it is also possible to overcome this problem through the use of a priori contrast coding.) In addition, MANOVA will not tell you which variables are responsible for the differences in mean vectors. Again, it is possible to overcome this with proper contrast coding for the dependent variables

In this handout, we will first explore the nature of multivariate sampling and then explore the logic behind MANOVA.

1. MANOVA: Multivariate Sampling

To understand MANOVA and multivariate sampling, let us first examine a MANOVA design. Suppose a researcher in psychotherapy interested in the treatment efficacy of depression randomly assigned clinic patients into four conditions: (1) a placebo control group who received typical clinic psychotherapy and a placebo drug; (2) a placebo cognitive therapy group who received the placebo medication and systematic cognitive psychotherapy; (3) an active antidepressant medication group who received the

typical clinic psychotherapy; and (4) an active medication group who received cognitive therapy. This is a (2 x 2) factorial design with medication (placebo versus drug) as one factor and type of psychotherapy (clinic versus cognitive) as the second factor. Studies such as this one typically collect a variety of measures before treatment, during treatment, and after treatment. To keep the example simple, we will focus only on three outcome measures, say, Beck Depression Index scores (a self-rated depression inventory), Hamilton Rating Scale scores (a clinician rated depression inventory), and Symptom Checklist for Relatives (a rating scale that a relative completes on the patient--it was made up for this example). High scores on all these measures indicate more depression; low scores indicate normality. The data matrix would look like this:

Person	Drug	Psychotherapy	BDI	HRS	SCR
Sally	placebo	cognitive	12	9	6
Mortimer	drug	clinic	10	13	7
Miranda	placebo	clinic	16	12	4
.	.	.	.	.	.
Waldo	drug	cognitive	8	3	2

For simplicity, assume the design is balanced with equal numbers of patients in all four conditions. A univariate ANOVA on any single outcome measure would contain three effects, a main effect for psychotherapy, a mean effect for medication, and an interaction between psychotherapy and medication. The MANOVA will also contain the same three effects. The univariate ANOVA main effect for psychotherapy tells whether the clinic versus the cognitive therapy groups have different means, irrespective of their medication. The MANOVA main effect for psychotherapy tells whether the clinic versus the cognitive therapy group have different mean *vectors* irrespective of their medication; the vectors in this case are the (3 x 1) column vectors of (BDI, HRS, and SCR) means. The univariate ANOVA for medication tells whether the placebo group has a different mean from the drug group irrespective of psychotherapy. The MANOVA main effect for medication tells whether the placebo group has a different mean *vector* from the drug group irrespective of psychotherapy. The univariate ANOVA interaction tells whether the four means for a single variable differ from the value predicted from knowledge of the main effects of psychotherapy and drug. The MANOVA interaction term tells whether the four mean *vectors* differ from the vector predicted from knowledge of the main effects of psychotherapy and drug.

If you are coming to the impression that a MANOVA has all the properties as an ANOVA, you are correct. The only difference is that an ANOVA deals with a (1 x 1) mean vector for any group while a MANOVA deals with a (p x 1) vector for any group, p being the number of dependent variables, 3 in our example. Now let's think for a minute. What is the variance-covariance matrix for a single variable? It is a (1 x 1) matrix that has only one element, the variance of the variable. What is the variance-covariance matrix for p variables? It is now a (p x p) matrix with the variances on the diagonal and the covariances

on the off diagonals. The ANOVA partitions the (1 x 1) covariance matrix into a part due to error and a part due to hypotheses (the two main effects and the interaction term as described above). Or for our example, we can write

$$V_t = V_p + V_m + V_{(p*m)} + V_e \quad (1.1)$$

Equation (1.1) states that the total variability (V_t) is the sum of the variability due to psychotherapy (V_p), the variability due to medication (V_m), the variability due to the interaction of psychotherapy with medication ($V_{(p*m)}$), and error variability (V_e).

The MANOVA likewise partitions its (p x p) covariance matrix into a part due to error and a part due to hypotheses. Consequently, the MANOVA for our example will have a (3 x 3) covariance matrix for total variability, a (3 x 3) covariance matrix due to psychotherapy, a (3 x 3) covariance matrix due to medication, a (3 x 3) covariance matrix due to the interaction of psychotherapy with medication, and a (3 x 3) covariance matrix for error. Or we can now write

$$\mathbf{V}_t = \mathbf{V}_p + \mathbf{V}_m + \mathbf{V}_{(p*m)} + \mathbf{V}_e \quad (1.2)$$

where **V** now stands for the appropriate (3 x 3) matrix. Note how equation (1.2) equals (1.1) except that (1.2) is in matrix form. Actually, if we considered all the variances in a univariate ANOVA as (1 by 1) matrices and wrote equation (1.1) in matrix form, we would have equation (1.2).

Let's now interpret what these matrices in (1.2) mean. The $\mathbf{V}_e$ matrix will look like this

	BDI	HRS	CSR
BDI	V_{e1}	$cov(e_1,e_2)$	$cov(e_1,e_3)$
HRS	$cov(e_2,e_1)$	V_{e2}	$cov(e_2,e_3)$
CSR	$cov(e_3,e_1)$	$cov(e_3,e_2)$	V_{e3}

The diagonal elements are the error variances. They have the same meaning as the error variances in their univariate ANOVAs. In a univariate ANOVA the error variance is an average variability within the four groups. The error variance in MANOVA has the same meaning. That is, if we did a univariate ANOVA for the Beck Depression Inventory, the error variance would be the mean squares within groups. v_{e1} would be the same means squares within groups; it would literally be the same number as in the univariate analysis. The same would hold for the mean squares within groups for the HRS and the CSR.

The only difference then is in the off diagonals. The off diagonal covariances for the error matrix must all be interpreted as *within group* covariances. That is, $cov(e_1,e_2)$ tells us the extent to which individuals within a group who have high BDI scores also tend to have high HRS scores. Because this is a covariance matrix, the matrix can be scaled to correlations to ease inspection. For example, $corr(e_1,e_2) = cov(e_1,e_2)\,(v_{e1}v_{e2})^{-1/2}$.

I suggest that you always have this error covariance matrix and its correlation matrix printed and inspect the correlations. In theory, if you could measure a sufficient number of factors, covariates, etc. so that the only remaining variability is due to random noise, then these correlations should all go to 0. Inspection of these correlations is often a sobering experience because it demonstrates how far we have to go to be able to predict behavior.

What about the other matrices? In MANOVA, they will all have their analogues in the univariate ANOVA. For example, the variance in BDI due to psychotherapy calculated from a univariate ANOVA of the BDI would be the first diagonal element in the $\mathbf{V}_p$ matrix. The variance of HRS calculated from a univariate ANOVA is the second diagonal element in $\mathbf{V}_p$. The variance in CSR due to the interaction between psychotherapy and drug as calculated from a univariate ANOVA will be the third diagonal element in $\mathbf{V}_{(p*m)}$.

The off diagonal elements are all covariances and should be interpreted as *between group* covariances. That is, in $\mathbf{V}_p$, cov(1,2) = cov(BDI, HRS) tells us whether the psychotherapy group with the highest mean score on BDI also has the highest mean score on the HRS. If, for example, the cognitive therapy were more efficacious than the clinic therapy, then we should expect that all the covariances in $\mathbf{V}_p$ be large and positive. Again, cov(2,3) = cov(HRS, CSR) in the $\mathbf{V}_{(p*n)}$ matrix has the following interpretation: if we control for the main effects of psychotherapy and medication, then do groups with high average scores on the Hamilton also tend to have high average scores on the relative's checklist?

In theory, if there were no main effect for psychotherapy on any of the measures, then all the elements of $\mathbf{V}_p$ will be 0. If there were no main effect for medication, then $\mathbf{V}_m$ will be all 0's. And if there were no interaction, then all of $\mathbf{V}_{(p*m)}$ would be 0's. It makes sense to have these matrices calculated and printed out so that you can inspect them. However, just as most programs for univariate ANOVAs do not give you the variance components, most computer programs for MANOVA do not give you the variance component matrices. You can, however, calculate them by hand.

2. Understanding MANOVA

Understanding of MANOVA requires understanding of three basic principles. They are:

- An understanding of univariate ANOVA.
- Remembering that mathematicians work very hard to be lazy.
- A little bit of matrix algebra.

Each of these topics will now be convered in detail.

2.1 Understanding Univariate ANOVA.

Let us review univariate ANOVA by examining the expected results of a simple oneway ANOVA under the null hypothesis. The overall logic of ANOVA is to obtain two different estimates of the population variance; hence, the term analysis of *variance*.

The first estimate is based on the variance within groups. The second estimate of the variance is based on the variance of the means of the groups. Let us examine the logic behind this by referring to an example of a oneway ANOVA.

Suppose we select four groups of equal size and measure them on a single variable. Let *n* denote the sample size within each group. The null hypothesis assumes that the scores for inviduals in each of the four groups are all sampled from a single normal distribution with mean μ and variance σ^2. Thus, the expected value for the mean of the first group is μ and the expected value of the variance of the first group is σ^2, the expected value for the mean of the second group is μ and the expected value of the variance of the second group is σ^2, etc. The observed statistics and their expected values for this ANOVA are given in Table 1.

Table 1. Observed and expected means and variances of four groups in a oneway ANOVA under the null hypothesis.

		Group			
		1	2	3	4
Sample Size:		*n*	*n*	*n*	*n*
Means:	Observed	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
	Expected	μ	μ	μ	μ
Variances:	Observed	s_1^2	s_2^2	s_3^2	s_4^2
	Expected	σ^2	σ^2	σ^2	σ^2

Now concentrate on the rows for the variances in Table 1. There are four observed variances, one for each group, and they all have expected values of σ^2. Instead of having four different estimates of σ^2, we can obtain a better overall estimate of σ^2 by simply taking the average of the four estimates. That is,

$$\hat{\sigma}^2_{s^2} = \frac{s_1^2 + s_2^2 + s_3^2 + s_4^2}{4} \qquad (2.1)$$

The notation $\hat{\sigma}^2_{s^2}$ is used to denote that this is the estimate of σ^2 derived from the observed variances. (Note that the only reason we could add up the four variances is because the four groups are independent. Were they dependent in some way, this could not be done.)

Now examine the four means. A very famous theorem in statistics, the *central limit theorem*, states that when scores are normally distributed with mean μ and variance σ^2, then the the sampling distribution of means based on size *n* will be normally

distributed with an overall mean of μ and variance σ^2/n. Thus, the four means in Table 1 will be sampled from a single normal distribution with mean μ and variance σ^2/n. A second estimate of σ^2 may now be obtained by going through three steps: (1) treat the four means as raw scores, (2) calculate the variance of the four means, and (3) multiplying the result by n. Let $\bar{x}$ denote the overall mean of the four means, then

$$\hat{\sigma}^2_{\bar{x}} = n\frac{\sum_{i=1}^{4}(\bar{x}_i - \bar{x})^2}{4-1} \tag{2.2}$$

where $\hat{\sigma}^2_{\bar{x}}$ denotes the estimate of σ^2 based on the means and not the variance of the means.

We now have two separate estimates of σ^2, the first based on the within group variances ($\hat{\sigma}^2_{s^2}$) and the second based on the means ($\hat{\sigma}^2_{\bar{x}}$ $\sigma^2_{\bar{x}}$). If we take a ratio of the two estimates, we expect a value close to 1.0, or

$$E\left(\frac{\hat{\sigma}^2_{\bar{x}}}{\hat{\sigma}^2_{s^2}}\right) \approx 1. \tag{2.3}$$

This is the logic of the simple oneway ANOVA, although it is most often expressed in different terms. The estimate $\hat{\sigma}^2_{s^2}$ from equation (2.1) is the mean squares within groups. The estimate $\hat{\sigma}^2_{\bar{x}}$ from (2.2) is the mean squares between groups. And the ratio in (2.3) is the *F* ratio expected under the null hypothesis. In order to generalize to any number of groups, say *g* groups, we would perform the same step but substitute *g* in place of 4 in Equation (2.2).

Now the above derivations all pertain to the null hypothesis. The alternative hypothesis states that at least one of the four groups has been sampled from a different normal distribution than the other three means. Here, for the sake of exposition, it is assumed that scores in the four groups are sampled from different normal distributions with respective means of μ_1, μ_2, μ_3, and μ_4, but with the same variance, σ^2.

Because the variances do not differ, the estimate derived from the observed group variances in equation (2.1), or $\hat{\sigma}^2_{s^2}$, will remain a valid estimate of σ^2. However, the variance derived from the means using equation (2.2) is no longer an estimate of σ^2. If we performed the calculation on the right hand side of Equation (2.2), the expected results would be

$$E\left(n\frac{\sum_{i=1}^{4}(\bar{x}_i - \bar{x})^2}{4-1}\right) = n\sigma^2_{\mu} + \hat{\sigma}^2_{\bar{x}}. \tag{2.4}$$

Consequently, the expectation of the *F* ratio becomes

$$E(F) = \frac{n\sigma^2_{\mu} + \widehat{\sigma}^2_{\bar{x}}}{\widehat{\sigma}^2_{s^2}} = \frac{\widehat{\sigma}^2_{\bar{x}}}{\widehat{\sigma}^2_{s^2}} + \frac{n\sigma^2_{\mu}}{\widehat{\sigma}^2_{s^2}} \approx 1 + \frac{n\sigma^2_{\mu}}{\widehat{\sigma}^2_{s^2}}. \qquad (2.5)$$

Instead of having an expectation around 1.0 (which *F* has under the null hypothesis), the expectation under the alternative hypothesis will be sometihing greater than 1.0. Hence, the larger the *F* statistic, the more likely that the null hypothesis is false.

2.2 Mathematicians Are Lazy

The second step in understanding MANOVA is the recognition that mathematicians are lazy. Mathematical indolence, as applied to MANOVA, is apparent because mathematical statisticians do not bother to express the information in terms of the estimates of variance that were outlined above. Instead, all the information is expressed in terms of *sums of squares and cross products*. Although this may seem quite foreign to the student, it does save steps in calculations--something that was important in the past when computers were not available. We can explore this logic by once again returning to a simple oneway ANOVA.

Recall that the two estimates of the population variance in a oneway ANOVA are termed *mean squares*. Computationally, mean squares denote the quantity resulting from dividing the sum of squares by its associated degrees of freedom. The between group mean squares--which will now be called the *hypothesis* mean squares, is

$$MS_h = \frac{SS_h}{df_h}$$

and the within group mean squares--which will now be called the error mean squares, is

$$MS_e = \frac{SS_e}{df_e}.$$

The *F* statistic for any hypothesis is defined as the mean squares for the hypothesis divided by the mean squares for error. Using a little algebra, the *F* statistic can be shown to be the product of two ratios. The first ratio is the degrees of freedom for error divided by the degrees of freedom for the hypothesis. The second ratio is the sum of squares for the hypothesis divided by the sum of squares for error. That is,

$$F = \frac{MS_h}{MS_e} = \frac{\frac{SS_h}{df_h}}{\frac{SS_e}{df_e}} = \frac{df_e}{df_h}\frac{SS_h}{SS_e} \; .$$

The role of mathematical laziness comes about by developing a new statistic--called the *A* statistic here--that simplifies matters by removing the step of calculating the mean squares. The *A* statistic is simply the *F* statistic multiplied by the degrees of freedom for the hypothesis and divided by the degrees of freedom for error, or

$$A = \frac{df_h}{df_e} F = \frac{df_h}{df_e}\left(\frac{df_e}{df_h}\frac{SS_h}{SS_e}\right) = \frac{SS_h}{SS_e} \; .$$

The chief advantage of using the *A* statistic is that one never has to go through the trouble of calculating mean squares. Hence, given the overriding principle of mathematical indolence, that, of course, must be adhered to at all costs in statistics, it is preferable to develop ANOVA tables using the *A* statistic instead of the *F* statistic and to develop tables for the critical values of the *A* statistic to replace tables of critical values for the *F* statistic. Accordingly, we can refer to traditional ANOVA tables as "stupid," and to ANOVA tables using the *F* statistic as "enlightened."

The following tables demonstrate the difference between a traditional, "stupid" ANOVA and a modern, "enlightened" ANOVA.

Stupid ANOVA Table

Source	Sums of Squares	Degrees of Freedom	Mean Squares	F	p
Hypothesis	SS_h	df_h	MS_h		
Error	SS_e	df_e	MS_e		
Total	SS_t	df_t			

Critical Values of F (α = .05) (Stupid)

	df_e (numerator)				
df_h	**1**	**2**	**3**	**4**	**5**
1	161.45	199.50	215.71	224.58	230.16
2	18.51	19.00	19.16	19.25	19.30
3	10.13	9.55	9.28	9.12	9.01

Enlightened ANOVA Table				
Source	**Sums of Squares**	**Degrees of Freedom**	**A**	**p**
Hypothesis	SS_h	df_h		
Error	SS_e	df_e		
Total	SS_t	df_t		

Critical Values of A (α = .05)(Enlightened)					
	df_e (numerator)				
df_h	**1**	**2**	**3**	**4**	**5**
1	161.45	399.00	647.13	898.32	1150.80
2	9.26	19.00	28.74	38.50	48.25
3	3.38	6.37	9.28	12.16	15.02

2.3. Vectors and Matrices

The final step in understanding MANOVA is to express the information in terms of vectors and matrices. Assume that instead of a single dependent variance in the oneway ANOVA, there are three dependent variables. Under the null hypothesis, it is assumed that scores on the three variables for each of the four groups are sampled from a trivariate normal distribution mean vector

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

and variance-covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}.$$

(Recall that the quantity $\rho_{12}\sigma_1\sigma_2$ equals the covariance between variables 1 and 2.)

Under the null hypothesis, the scores for all those in group 1 will be sampled from this distribution, as will the scores for all those individuals in groups 2, 3, and 4. The observed and expected values for the four groups are given in Table 2.

Table 2. Observed and expected statistics for the mean vectors and the variance-covariance matrices of four groups in a oneway MANOVA under the null hypothesis.

		Group			
		1	2	3	4
Sample Size:		*n*	*n*	*n*	*n*
Mean Vector:	Observed	$\bar{\mathbf{x}}_1$	$\bar{\mathbf{x}}_2$	$\bar{\mathbf{x}}_3$	$\bar{\mathbf{x}}_4$
	Expected	μ	μ	μ	μ
Covariance Matrix:	Observed	$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$
	Expected	Σ	Σ	Σ	Σ

Note the resemblance between Tables 1 and 2. The only difference is that Table 2 is written in matrix notation. Indeed, if we consider the elements in Table 1 as (1 by 1) vectors or (1 by 1) matrices and then rewrite Table 1, we would get Table 2!

Once again, how can we obtain different estimates of Σ? Again, concentrate on the rows marked variances in Table 2. The easiest way to estimate Σ is to add up the covariance matrices for the four groups and divide by 4--or, in other words, take the average observed covariance matrix:

$$\hat{\Sigma}_w = \frac{\mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_3 + \mathbf{S}_4}{4}. \qquad (2.6)$$

Note how (2.6) is identical to (2.1) except that (2.6) is expressed in matrix notation.

What about the means? Under the null hypothesis, the means will be sampled from a trivariate normal distribution with mean vector μ and covariance matrix Σ/n. Consequently, to obtain an estimate of Σ based on the mean vectors for the four groups, we proceed with the same logic as that in the oneway ANOVAiven in section 2.1, but now apply it to the *vectors* of means. That is, treat the means as if they were raw scores and calculate the covariance matrix for the three "variables;" then multiply this result by *n*. Let $\bar{X}_{ij}$ denote the mean for the ith group on the jth variable. The data would look like this:

$$\begin{matrix} \bar{X}_{11} & \bar{X}_{12} & \bar{X}_{13} \\ \bar{X}_{21} & \bar{X}_{22} & \bar{X}_{23} \\ \bar{X}_{31} & \bar{X}_{32} & \bar{X}_{33} \\ \bar{X}_{41} & \bar{X}_{42} & \bar{X}_{43} \end{matrix}$$

This is identical to a data matrix with four observations and three variables. In this case, the observations are the four groups and the "scores" or numbers are the means. Let $\mathbf{SSCP}_{\bar{X}}$ denote the sums of squares and cross products matrix used to calculate the covariance matrix for these three variables. Then the estimate of Σ based on the means of the four groups will be

$$\hat{\Sigma}_b = n\frac{\mathbf{SSCP}_{\bar{X}}}{4-1}. \tag{2.7}$$

Note how (2.7) is identical in form to (2.2) except that (2.7) is expressed in matrix notation.

We now have two different estimates of Σ. The first, $\hat{\Sigma}_w$, is derived from the average covariance matrix within the groups, and the second, $\hat{\Sigma}_b$, is derived from the covariance matrix for the group mean vectors. Because both of them measure the same quantity, we expect

$$E(\hat{\Sigma}_b\hat{\Sigma}_w^{-1}) = \mathbf{I}. \tag{2.8}$$

or an identity matrix. Note that (2.8) is identical to (2.3) except that (2.8) is written in matrix notation. If the matrices in (2.8) were simply (1 by 1) matrices then their expectation would be a (1 by 1) identity matrix or simply 1.0. Just what we got in (2.3)!

Now examine the alternative hypothesis. Under the alternative hypothesis it is assumed that each group is sampled from a multivariate normal distribution that has the same covariance matrix, say Σ. Consequently, the average of the covariance matrices for the four groups in equation (2.6) remains an estimate of Σ. However, under the alternate hypothesis, the mean vector for at least one group is sampled from a multivariate normal with a different mean vector than the other groups. Consequently, the covariance matrix for the observed mean vectors will reflect the covariance due to true mean differences, say Σ_μ, plus the covariance matrix for sampling error, Σ/n. Thus, multiplying the covariance matrix of the means by n now gives

$$n\Sigma_\mu + \Sigma. \tag{2.9}$$

Once again, (2.9) is the same as (2.4) except for its expression in matrix notation. What is the expectation of the estimate from the observed means postmultiplied by the estimate from the observed covariance matrices under the alternative hypothesis? Substituting the expression in (2.9) for $\hat{\Sigma}_b$ in Equation (2.8) gives

$$E(\hat{\Sigma}_b\hat{\Sigma}_w^{-1}) = (n\Sigma_\mu + \Sigma)\Sigma^{-1} = n\Sigma_\mu\Sigma + \Sigma\Sigma^{-1} = n\Sigma_\mu\Sigma + \mathbf{I}. \tag{2.10}$$

Although it is a bit difficult to see how, (2.10) is the same as (2.5) except that (2.10) is expressed in matrix notation. Note that (2.10) is the sum of an identity matrix and another term. The identity matrix will have 1's on the diagonal, so the diagonals of the result will always be "1 + something." That "something" will always be a positive number [for technical reasons in matrix algebra]. Consequently, the diagonals in (2.10) will always be greater than 1.0. Again, if we consider the matrices in (2.10) as (1 by 1) matrices, we can verify that the expectation will always be greater than 1.0, just as we found in (2.5).

This exercise demonstrates how MANOVA is a natural extension of ANOVA. The only remaining point is to add the laziness of mathematical statisticians. In ANOVA, we saw how this laziness saved a compuational step by avoiding calculations of the means squares (which is simply another term for the variance) and expressing all the information in terms of the sums of squares. The same applies to MANOVA, instead of calculating the mean squares and mean products matrix (which is simply another term for a covariance matrix), MANOVA avoids this step by expressing all the information in terms of the sums of squares and cross products matrices.

The *A* statistic developed in section 2.2 was simply the ratio of the sum of squares for an hypothesis and the sum of squares for error. Let **H** denote the hypothesis sums of squares and cross products matrix, and let **E** denote the error sums of squares and cross products matrix. The multivariate equivalent of the *A* statistic is the matrix **A** which is

$$\mathbf{A} = \mathbf{H}\mathbf{E}^{-1} \tag{2.11}$$

Verify how Equation 2.11 is the matrix analogue of the *A* statistic given in section 2.2. Notice how mean squares (or, in other terms, covariance matrices) disappear from MANOVA just as they did for ANOVA. All hypothesis tests may be performed on matrix **A**. Parenthetically, note that because both **H** and **E** are symmetric, $\mathbf{H}\mathbf{E}^{-1} = \mathbf{E}^{-1}\mathbf{H}$. This is one special case where the order of matrix multiplication does not matter.

3.0. Hypothesis Testing in MANOVA

All current MANOVA tests are made on $\mathbf{A} = \mathbf{E}^{-1}\mathbf{H}$. That's the good news. The bad news is that there are four different multivariate tests that are made on $\mathbf{E}^{-1}\mathbf{H}$. Each of the four test statistics has its own associated *F* ratio. In some cases the four tests give an exact *F* ratio for testing the null hypothesis and in other cases the *F* ratio is approximated. The reason for four different statistics and for approximations is that the mathematics of MANOVA get so complicated in some cases that no one has ever been able to solve them. (Technically, the math folks can't figure out the sampling distribution of the *F* statistic in some multivariate cases.)

To understand MANOVA, it is not necessary to understand the derivation of the statistics. Here, all that is mentioned is their names and some properties. In terms of notation, assume that there are *q* dependent variables in the MANOVA, and let λ_i denote the ith eigenvalue of matrix **A** which, of course, equals $\mathbf{H}\mathbf{E}^{-1}$.

The first statistic is *Pillai's trace*. Some statisticians consider it to be the most powerful and most robust of the four statistics. The formula is

$$\text{Pillai's trace} = \text{trace}[\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}] = \sum_{i=1}^{q} \frac{\lambda_i}{1+\lambda_i}. \tag{3.1}$$

The second test statistic is *Hotelling-Lawley's trace*.

$$\text{Hotelling-Lawley's trace} = \text{trace}(\mathbf{A}) = \text{trace}(\mathbf{HE}^{-1}) = \sum_{i=1}^{q} \lambda_i . \tag{3.2}$$

The third is *Wilk's lambda* (Λ). (Here, the upper case, Greek Λ is used for Wilk's lambda to avoid confusion with the lower case, Greek λ often used to denote an eigenvalue. However, many texts use the lower case lambda as the notation for Wilk's lambda.) Wilk's Λ was the first MANOVA test statistic developed and is very important for several multivariate procedures in addition to MANOVA.

$$\text{Wilk's lambda} = \Lambda = \frac{|\mathbf{E}|}{|\mathbf{H}+\mathbf{E}|} = \prod_{i=1}^{q} \frac{1}{1+\lambda_i}. \tag{3.3}$$

The quantity (1 - Λ) is often interpreted as the proportion of variance in the dependent variables explained by the model effect. However, this quantity is not unbiased and can be quite misleading in small samples.

The fourth and last statistic is *Roy's largest root*. This gives an upper bound for the *F* statistic.

$$\text{Roy's largest root} = \max(\lambda_i). \tag{3.4}$$

or the maximum eigenvalue of $\mathbf{A} = \mathbf{HE}^{-1}$. (Recall that a "root" is another name for an eigenvalue.) Hence, this statistic could also be called Roy's largest eigenvalue. (In case you know where Roy's smallest root is, please let him know.)

Note how all the formula in equations (3.1) through (3.4) are based on the eigenvalues of $\mathbf{A} = \mathbf{HE}^{-1}$. This is the major reason why statstical programs such as SAS print out the eigenvalues and eigenvectors of $\mathbf{A} = \mathbf{HE}^{-1}$.

Once the statistics in (3.1) through (3.4) are obtained, they are translated into *F* statistics in order to test the null hypothesis. The reason for this translation is identical to the reason for converting Hotelling's T^2--the easy availability of published tables of the *F* distribution. The important issue to recognize is that in some cases, the *F* statistic is exact and in other cases it is approximate. Good statistical packages will inform you whether the *F* is exact or approximate.

In some cases, the four will generate identical *F* statistics and identical probabilities. In other's they will differ. When they differ, Pillai's trace is often used because it is the most powerful and robust. Because Roy's largest root is an upper bound

on *F*, it will give a lower bound estimate of the probability of *F*. Thus, Roy's largest root is generally disregarded when it is significant but the others are not significant.