SPECIFICITY The percentage of cases in the "other" group correctly classified by the model, also known as the *correct identification of true negatives*. A measure of classification accuracy.

SPECIFICITY ASSUMPTION Requirement that the logistic regression model contain all relevant predictors and no irrelevant predictors. If the model is incorrectly specified, parameter estimates will be inaccurate.

STANDARD ERROR (*SE*) Estimates the variability from sample to sample in a model coefficient. Used in computing z scores and confidence intervals.

z TEST Tests whether a model coefficient differs from zero in the population. Large z values (in absolute value) mean that the population coefficient probably differs from 0. To compute z, divide a parameter estimate by its standard error.

References

Agresti, A. (1990). *Categorical data analysis*. Wiley: New York.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit, and probit models*. Beverly Hills, CA: Sage.

Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences*. New York: Oxford Press.

Fleiss, J. L., Williams, J. B. W., & Dubro, A. F. (1986). Logistic regression of psychiatric data. *Journal of Psychiatric Research, 20*, 145–209.

Hauck, W. W., & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association, 72*, 851–853.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.

Klecka, W. R. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. New York: Chapman and Hall.

Neter, J., & Wasserman, W. (1974). *Applied linear statistical models*. Homewood, IL: Irwin.

Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association, 73*, 699–705.

Shott, S. (1991). Logistic regression and discriminant analysis. *Journal of the American Veterinary Medical Association, 198*, 1902–1905.

Spiegel, D., Frischholz, E. J., Fleiss, J. L., & Spiegel, H. (1993). Predictors of smoking abstinence following a single-session restructuring intervention with self-hynosis. *American Journal of Psychiatry, 150*, 1090-1097.

Stablein, D. M., Miller, J. D., Choi, S. C., & Becker, D. P. (1980). Statistical methods for determining prognosis in severe head injury. *Neurosurgery, 6*, 243–246.

Walsh, A. (1987). Teaching understanding and interpretation of logit regression. *Teaching Sociology, 15*, 178–183.

Wright, R. (1992). *Semantic evaluation criteria and the expository writing skills of grade school children*. Unpublished doctoral dissertation, University of Illinois at Chicago.

8 Multivariate Analysis of Variance

Kevin P. Weinfurt

Maybe you have had this dream before: You are earnestly reading a research article you need to digest in order to work on a paper, grant application, or thesis. You successfully wade through the jargon found in the Introduction and Method sections and reach the climax of the article, the Results section, where you are confronted with something like this[1]:

> A 4 (Group) × 2 (Time) repeated measures multivariate analysis of variance revealed a significant multivariate main effect for Time [Wilks's $\Lambda = .406$, F (4, 113) = 41.28, $p < .001$, $\eta^2 = .59$], but no significant effect for Group [Wilks's $\Lambda = .865$, F (12, 229) = 1.41, $p > .05$] or the Group × Time interaction [Wilks's $\Lambda = .856$, F (12, 299) = 1.51, $p > .05$].

Although it may seem intimidating, such text need not be the stuff of nightmares. In short, multivariate analysis of variance (MANOVA) is used to assess the statistical significance of the effect of one or more independent variables on a set of two or more dependent variables.

I begin this chapter with an example of a research situation in which a MANOVA is used, followed by a discussion of some basic statistical concepts and the general purpose of a MANOVA. The assumptions underlying a MANOVA, as well as the consequences of violating those assumptions, are discussed next. Following this is a brief, nontechnical

1. This tongue-in-cheek paragraph was based on the Results section of an actual article published in a legitimate psychology journal, with very few modifications made. Unfortunately, the difference between exaggeration and reality is small in this case.

explanation of how a MANOVA is performed and how it relates to a traditional analysis of variance (ANOVA). Next, the ways in which MANOVA results are presented in journal articles and various analytical methods that often follow multivariate significance are reviewed. The latter sections of the chapter are devoted to multivariate analysis of covariance (MANCOVA), repeated measures MANOVA, and power analysis. I also provide a list of recommended readings and a glossary of terms and symbols used throughout the chapter.

A Hypothetical MANOVA Design

Imagine that a clinical psychologist interested in panic disorders (see Antony, Brown, & Barlow, 1992, for a recent review) designed and published the following study. One hundred subjects suffering from panic disorder were solicited to participate in a program to improve their condition. The author of the study wanted to determine the effectiveness of two different forms of therapy: relaxation training and cognitive–behavioral therapy. Subjects were randomly assigned to one of four groups: relaxation training only (relaxation), cognitive–behavioral therapy only (cog–behav), relaxation and cognitive–behavioral therapies combined (combined), and a control group with no therapy (control). Thus, the study is a 2 (relaxation) × 2 (cog–behav) factorial design (Campbell & Stanley, 1963). On the basis of a structured interview, the clinician assigned each subject a rating from 1 to 10 to index the severity of each subject's disorder at the outset of the study, with higher scores reflecting greater severity of the condition. Subjects participated in the program for 8 weeks, after which they were tested using three subscales of a panic disorder questionnaire. The subscales tapped different theoretical components of panic disorder: cognitive, emotional, and physiological. Each scale was scored on a 1–20 scale and then converted to standard scores, with higher scores indicating greater intensity of panic disorder. Table 1 displays each group's mean and standard deviation for the three panic subscales. This example is referred to throughout the chapter to demonstrate various aspects of a MANOVA.

Some Preliminary Statistical Concepts

Knowledge of a few key statistical concepts is required to understand MANOVA and issued related to its interpretation. They are Type I error, the Bonferroni inequality, effect size, and statistical power.

Table 1

Mean Hypothetical Standard Scores on the Cognitive, Emotional, and Physiological Subscales for the Four Experimental Groups

	Cognitive		Emotional		Physiological	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control	1.75	1.10	1.63	1.11	1.44	0.90
Relaxation	−0.19	0.66	−0.07	1.07	0.56	1.20
Cog–behav	0.03	0.92	0.33	0.90	0.20	1.02
Combined	0.45	0.96	−0.18	0.89	0.08	0.97

Note. There were 25 participants in each group. Cog–behav = cognitive–behavioral.

Type I Error

The Type I error rate, expressed as alpha (α), is the probability of rejecting the null hypothesis when it is true. In other words, it is the probability of detecting a significant effect when there is no real effect in nature (Kleinbaum, Kupper, & Muller, 1988). Statisticians speak of two types of alphas: actual and nominal. The actual alpha level is the actual (true) probability of making a Type I error. The nominal alpha level is the Type I error rate that the researcher desires (e.g., .05). The nominal alpha level is based on assumptions, and if these assumptions are tenable, the nominal alpha is equal to the actual alpha. However, if the assumptions are violated to some degree, the actual alpha may be different from the nominal alpha. Hence, it is important to evaluate a MANOVA's assumptions to better determine what the actual alpha might be in a given situation.

One can also speak of alphas at three different levels of analysis. Most researchers are familiar with the alpha for a single statistical test (e.g., a *t* test), indicating the probability of falsely rejecting the null hypothesis for that particular test of significance. There is also a *familywise alpha* (or comparisonwise alpha; Ryan, 1959) in situations in which multiple levels of an independent variable are being compared with respect to a dependent variable, as in a one-way ANOVA. The familywise alpha is the probability of falsely rejecting the null hypothesis for at least one of the statistical comparisons being made. The comparisons are considered a "family" of tests. There are many multiple-comparison procedures available for this situation, including the Tukey, Sheffé, the least significant difference, Newman-Keuls, and Duncan's multiple range (Keselman, Keselman, & Games, 1991; Kleinbaum et al., 1988; Seaman, Levin, & Serlin, 1991).

Finally, there is the *experimentwise alpha*, which is the probability of falsely rejecting the null hypothesis for at least one statistical test when several tests are used in the same study. Hence, in a study with two ANOVAs, the experimentwise alpha would be the probability of obtaining erroneously significant results for any test performed in the study, including the ANOVA omnibus tests, as well as any analysis of main and interaction effects done for each ANOVA (see the Glossary for definitions of main and interaction effects).

The Bonferroni Inequality

The Bonferroni inequality (Stevens, 1986) is an extremely important concept for understanding the familywise and experimentwise alphas, because it defines the maximum value of alpha for a given set of statistical tests. Basically, the Bonferroni inequality states that the overall alpha for a set of tests will be less than or equal to the sum of the alpha levels associated with each individual test. Therefore, if six t tests are performed using .05 as the criterion for rejecting the null hypothesis for each, then the overall alpha level is approximately $6(0.5) = .30$. (Actually, the overall alpha is .265; see Maxwell & Delaney, 1990.) Hence, there is a 30% chance of committing at least one Type I error across the six t tests.

When performing MANOVAs, there are often many statistical tests involved. If the null hypothesis is correct in every case, then the more tests one performs, the greater the possibility of committing a Type I error. In other words, the experimentwise alpha increases. When reading a research article, be aware of how many tests are being performed, the alpha level used for each test, and what the maximum experimentwise alpha could be using the Bonferroni inequality. It is surprising how many studies are published that base their conclusions on "significant" statistical results, when in actuality the experimentwise alpha level is so high that erroneous results are almost guaranteed.

Effect Size

In addition to determining whether a finding could have occurred by chance, it is also useful to know the magnitude of a finding. For example, consider a simple pretest–posttest study with some intervening treatment between testing periods. A paired t test could be used to ascertain whether the mean pretest score differed from the mean posttest score at a statistically significant level. One might also ask how large the difference between the two means is. In other words, what is the magnitude of the intervening treatment's effect? This is precisely the issue addressed by the notion of *effect size*.

Typically, an effect size is expressed as a number between 0 and 1, with higher values reflecting a larger effect. As discussed later, one measure of effect size in MANOVA is eta-square, which is roughly equivalent to the R^2 used in multiple regression (see chapter 2 in this book; also see Pedhazur, 1982). Cohen's (1977) classification of effect sizes has become somewhat of a standard in social research. The most often quoted standard refers to effect sizes measured as mean differences (as in the pretest–posttest example given earlier): Effect sizes around 0.20 are small, those around 0.50 are medium, and those larger than 0.80 are large. Although this classification scheme is often used to judge the magnitude of eta-squares, the more proper standard should be the one Cohen suggested for effect sizes measured via R^2 or other such indices: 0.01 is small, 0.09 is medium, and 0.25 or greater is large. The majority of social research produces small to medium effect sizes.

Statistical Power

Power is the probability of detecting a significant effect when the effect truly does exist in nature. A simple metaphor for understanding power is that of a flashlight. If a statistical test is a flashlight shining in the dark, then power is the brightness of the beam. One is able to see more of what is really there with a more powerful beam. Power is a function of sample size, effect size, and the nominal alpha level set by the researcher (Cohen, 1977). The larger the sample, the more power there is. The bigger the effect as it exists in nature, the greater the power to detect it. Power is expressed as a number ranging from zero to one, with zero indicating no power and one indicating perfect power. Power is an important notion when a MANOVA is concerned, because it is helpful to know how able a MANOVA is to reject the null hypothesis for a given sample size and effect size. In a research situation, if there is low power and no significant effects are obtained, then there is reason to believe that effects might truly exist in nature, but they are undetectable because of the research design or number of subjects. I say more about the power of a MANOVA in a later section.

MANOVA Basics

The Purpose of a MANOVA

Perhaps the best way to begin a discussion about a MANOVA is to consider its more familiar relative, the univariate ANOVA. In the ANOVA there is a continuous dependent variable (e.g., IQ) and one or more categorical independent variables (e.g., social class, gender). The purpose of the ANOVA is to determine whether the means of the dependent variable for each level of an independent variable are significantly different from each other. Therefore, if the independent variable social class has three levels (lower, middle, and upper), an ANOVA could tell whether one social class had a higher mean IQ than another class. This relationship between the independent and dependent variables is sometimes referred to as the "influence" or the "effect" of the independent variable on the dependent variable, even though the direction of the causal chain may not necessarily flow from independent to dependent variable. If a second independent variable was added to the model—gender, for instance—a 3 (social class) × 2 (gender) ANOVA could be performed to determine the influence of social class on IQ (as before), the influence of gender on IQ (e.g., Is there a difference between males and females with regard to IQ?), and the joint influence of social class and gender (e.g., Does the difference between the mean male and mean female IQ depend on the level of social class?). This last variable describes a statistical interaction. An interaction addresses whether the influence of one independent variable is altered by the level of another independent variable.

The MANOVA allows one to examine the effects of the independent variables in much the same way as the univariate ANOVA, including main effects, interaction effects, contrast analyses, covariates, and repeated measure effects. The meanings of these concepts are dealt with shortly. For now, note that the ANOVA is confined to situations in which there is only one dependent variable, hence the term *univariate analysis of variance*. *Multivariate analysis of variance* is a technique used for situations in which there is more than one dependent variable. In addition to having more than one dependent variable, the MANOVA design also requires that the dependent measures be correlated. If the variables being used are statistically correlated with one another, then there is an empirical relation between them, and they could be analyzed using a MANOVA.

Table 2

Correlations Between the Dependent Measures for the Hypothetical Panic Disorder Study

Subscale	Cognitive	Emotional	Physiological
Cognitive	—		
Emotional	.36	—	
Physiological	.25	.21	—

Note. $N = 100$. All $ps < .05$.

Ideally, the dependent variables should be theoretically correlated as well as empirically correlated. In the panic disorder example, there are three dependent variables (i.e., the three subscales of the panic disorder questionnaire). They are theoretically related in that they all assess panic disorder. Also, the variables are intercorrelated, as Table 2 illustrates, suggesting that this variable set may be a candidate for a MANOVA.

Why Multivariate Analyses?

At this point, a question may arise as to why a researcher would want to analyze two or more dependent variables at once. Why not use separate univariate tests for each dependent variable and avoid confusing people? Researchers in the social and biological sciences usually have one of two reasons for using a multivariate approach: controlling Type I error and providing a multivariate analysis of effects by taking into account the correlations between dependent measures.

Controlling Against Type I Error

As Huberty and Morris's (1989) survey demonstrates, MANOVAs are most often done with the intent of keeping the Type I error rate at the nominal alpha level. One school of thought, supported by Hummel and Sligo's (1971) research, maintains that a MANOVA should be conducted when there are multiple dependent variables, and if the multivariate test is significant, then univariate ANOVAs are conducted for each of the dependent measures. The theory is that by performing an overall omnibus test of significance first—the MANOVA—one is guarding against the chance of committing a Type I error that might occur as a result of unwarranted multiple ANOVAs. In recent years, however, this approach has come under attack (Huberty & Morris, 1989; Wilkinson, 1975), and these criticisms are reviewed later when discussing follow-up procedures.

Providing a Multivariate Analysis of Effects

If there are multiple dependent measures and they are intercorrelated, then the intercorrelations can be taken into account to provide a much richer multivariate analysis of the data. This is true for two reasons. First, intercorrelations between outcome measures suggest that the measures may be partially redundant. That is, there may a degree of conceptual overlapping. As an example, suppose that a psychologist performed a study on affect intensity (Larsen & Diener, 1987) that used the Affect Intensity Measure (AIM; Larsen, 1984) and galvanic skin response (GSR) as two of several dependent measures. Because scores on the AIM are positively correlated with GSR to emotional stimuli, it is reasonable to suppose that the two measures are tapping the same concept. Hence, if separate ANOVAs for each measure were performed and each produced significant results, could it really be said that significance was obtained for two completely separate dependent variables? Although there certainly were two separate measures, it is questionable whether the two tapped conceptually separate constructs. A MANOVA avoids this question by taking the correlations between the dependent measures into consideration. As long as the effects being tested are multivariate effects (i.e., taking all dependent variables at once), there will be no redundant information in the results of the MANOVA.

A second reason why a multivariate approach can offer a richer analysis of the data is that a MANOVA can detect when groups differ on a system of variables (Huberty & Morris, 1989). Taken individually, the dependent variables may not show significant group differences, but taken as a whole—as a system defining one or more theoretical constructs—differences caused by the independent variables may be revealed. This is accomplished by finding a *linear composite* of the dependent measures that maximizes the separation between the groups defined by the independent variable, resulting in the most statistically significant value of the MANOVA test statistic. A linear composite refers to some combination of the dependent measures (e.g., .46AIM + .24GSR − .10gender, etc.). This demonstrates how univariate tests using the AIM and GSR might not yield significance, but a linear combination of the dependent variable would (see Stevens, 1986). By examining such dependent variable systems, a researcher is sometimes better able to discern the effects of independent variables.

To summarize, MANOVA is a procedure used to test the significance of the effects of one or more categorical independent variables on two or more continuous dependent variables. In the panic disorder study, the researcher wants to test the effect of relaxation training and cognitive–behavioral therapy (two categorical independent variables) on the cognitive, emotional, and physiological subscales of the panic questionnaire (three continuous dependent variables). Having defined the MANOVA and some of the reasons it is used, I review the assumptions inherent in the MANOVA.

Assumptions of the MANOVA

All parametric statistical procedures are inferential procedures (i.e., they make inferences about populations). Mathematics and logic dictate that inferences be based on assumptions, and so like any other parametric statistical technique, the MANOVA has assumptions with which scientists must be concerned. These assumptions are not esoteric mathematic theory but conditions of the data that must be assessed (and hopefully satisfied) before trying to interpret the results of a MANOVA. The three necessary conditions are (a) multivariate normality, (b) homogeneity of the covariance matrices, and (c) independence of observations. For each of the three assumptions, I present the ANOVA analog to that assumption, a definition of the assumption, how it is tested, and what effect violating the assumption will have on the Type I error rate and the statistical power of the MANOVA.

Multivariate Normality

In the case of univariate ANOVA, the statistical tests are based on the assumption that the observations on the dependent variable be normally distributed for each group defined by the independent variables. Multivariate normality is a much harder criterion to satisfy than univariate normality. This is clearly seen when one considers some properties of data that are multivariate normal (Stevens, 1986). First, all of the individual dependent variables must be distributed normally. Second, any linear combination of the dependent variables must also be normally distributed. Finally, all subsets of the variables must have a multivariate normal distribution. Unfortunately, most researchers fail to report how well their data approximate a multivariate normal distribution, and so one will seldom find anything in an article using MANOVA that mentions this assumption. Still, there are procedures for assessing multivariate normality, as discussed by Stevens (1986, chapter 6). Because space limitations prevent a detailed discussion regarding these techniques, the interested reader should consult Stevens's thorough and systematic treatment of this topic.

What if the distribution of the dependent measures is not multivariate normal? In terms of Type I error rate, the MANOVA appears to be fairly robust. That is, violation of the multivariate normality assumption has a small effect on the actual alpha level with which the researcher is working (Stevens, 1986). With regard to statistical power, Olson (1974) found that extremely platykurtic distributions (i.e., those that are flat compared with the more peaked normal distribution) can reduce the power of the analysis. In practice, MANOVAs tend to be performed on data regardless of whether the data violate this assumption, because the general consensus is that the MANOVA is a robust procedure. How correct this belief is can be determined only by more studies on the MANOVA's ability to withstand violations of the multivariate normality condition.

Homogeneity of the Covariance Matrices

Univariate ANOVA requires that the variance of the dependent variable be the same for all groups defined by the independent variables. In the multivariate context, the variances for all of the dependent variables must be equal across the experimental groups defined by the independent variable. Additionally, MANOVA demands that the covariance—the variance shared between two variables—for all unique pairs of dependent measures be equal for all experimental groups. Table 3 shows the components of a covariance matrix for an experimental group in a study with three dependent measures (A, B, and C). Observe that each variable's variance is on the diagonal, and the covariance for each pair of variables make up the rest of the matrix. Each of these numbers is referred to as an element of the matrix; therefore, this matrix has nine elements.

Note also that this is a covariance matrix for only one experimental group. In order to say the covariance matrices are homogeneous across experimental groups, the covariance matrices for each group are examined and checked to see that each matrix element (e.g., Covariance AB, Variance A, Variance B, etc.) is equal for all of the groups. To put this into more concrete terms, consider the hypothetical panic disorder data. There are three dependent measures, and so the covariance matrix for each group will be a 3 × 3 table of variances and covariances. Table 4 shows the covariance matrices for the four experimental groups that are formed when the two independent variables (relax and cog–behav) are crossed.

In examining Table 4, one might ask whether the variance for the

Table 3

Components of a 3 × 3 Covariance Matrix (Three Dependent Variables) for One Experimental Group

Variable	A	B	C
A	Variance A	Covariance AB	Covariance AC
B	Covariance BA	Variance B	Covariance BC
C	Covariance CA	Covariance CB	Variance C

Table 4

Within-Groups Covariance Matrices for the Hypothetical Panic Disorder Study

Group	Cognitive	Emotional	Physiological
Control			
Cognitive	1.22	0.10	0.24
Emotional	0.10	1.25	−0.07
Physiological	0.24	−0.07	0.82
Relax			
Cognitive	0.44	0.05	−0.02
Emotional	0.05	1.13	0.24
Physiological	−0.02	0.24	1.44
Cog–behav			
Cognitive	0.85	0.02	−0.34
Emotional	0.02	0.80	−0.29
Physiological	−0.34	−0.29	1.05
Combined			
Cognitive	0.92	−0.01	0.16
Emotional	−0.01	0.78	−0.09
Physiological	0.16	−0.09	0.94

Note. Cog–behav = cognitive–behavioral.

cognitive subscale is significantly different across the four groups. (Observe that these variances are 1.22, 0.44, 0.85, and 0.92.) Moreover, it would be desirable to know whether the matrices of variances and covariances are equal for all four groups. Two indices used to test the homogeneity of covariance matrices are Box's M (Norusis, 1988) and Bartlett's chi-square (Green, 1978). Both tests are highly sensitive to violations of the multivariate normality assumption, and it is therefore recommended that the tenability of that assumption be thoroughly investigated. For illustration, assume that the panic disorder data conform reasonably well to the normality requirement, and so the results of the chi-square are presented here so that the reader might recognize such a test. The null hypothesis is that the groups have equal covariance matrices. Therefore, if the test yields statistical significance, the groups are not homogeneous with respect to covariance matrices. For the panic disorder study, Bartlett's test produces the following: χ^2 (18, N = 100) = 18.83, $p > .05$. Hence, it can be concluded that the panic data satisfy the homogeneity of covariance matrices assumption.

Now consider what happens when this assumption is not tenable. Summarizing the studies done on the subject, Stevens (1986) concluded that violation of the homogeneity of covariance matrices assumption when the number of subjects in each experimental group is approximately equal will lead to a slight reduction in statistical power. For cases in which the number of subjects in each group is markedly unequal, the Type I error rate will be either inflated or deflated depending on which matrices are the most different (see Stevens, 1986, p. 227, for more detail).

Independence of Observations

The last and most important assumption is exactly the same for a MANOVA as it is for a univariate ANOVA. Both procedures assume that observations are independent of one another. This means that a subject's scores on the dependent measures are not influenced by the other subjects in his or her experimental group. Hence, if a study involved having subjects interact in an experimental condition, it is possible that the subjects are affecting each other's scores. As an example, consider what might happen if schoolchildren were asked to answer survey questions aloud in a small group setting. In almost every group of children, there will be those who simply follow the lead of others; therefore, one would expect that these more passive children's responses to the survey questions would be influenced by a more dominant child's answers.

Such dependence among observations can be assessed by an *intraclass correlation* (Fleiss, 1986; Guilford, 1965). Stevens (1986) noted that even a small intraclass correlation—indicating a small degree of dependence among observations—can inflate the actual alpha to seven times the nominal alpha level the experimenter observes. Hence, although it is possible to speak of a MANOVA's relative robustness with regard to the first two assumptions, there is little room for violation of this last assumption. The critical journal reader can determine whether the tenability of the assumption is suspect by looking at the research methods. Were the experimental conditions administered on an individual basis or to a group? Could other subjects have affected a person's responses at the time the dependent variables were measured? In the panic disorder study, there is little chance of dependence among observations, because the therapies were individual therapies. If the relax and cog–behav programs were delivered in a group therapy context, then it would be wise to investigate the assumed independence of observations with an intraclass correlation. If dependence was found, then the group means would be the logical units of analysis, not the individual scores. Techniques for analyzing dependent data are being developed. (For data of dyads, e.g., see Mendoza and Graziano, 1982.)

The MANOVA Procedure

Having reviewed the conditions necessary for a MANOVA, I now present a conceptual discussion of how a MANOVA is performed, beginning with the null hypothesis that is tested in MANOVA and how it relates to the null hypothesis tested in univariate ANOVA. Then I show how matrix algebra is used in a MANOVA to carry out the same basic procedure that takes place in a univariate ANOVA. The various multivariate test statistics are then introduced, as well as the measure of explained variance, eta-square.

Null Hypothesis Testing

To begin, consider the simplest version of a univariate ANOVA, the t test for independent samples (Kleinbaum et al., 1988). The t test is used when two groups are compared on a single continuous dependent variable. As an example, imagine that 40 males and 40 females took the revised Wechsler Adult Intelligence Scale (WAIS–R; Kaplan & Saccuzzo,

1989) and that a verbal IQ score was obtained for each subject. A t test could be performed to test the null hypothesis that the mean IQ for the male population is equal to the mean IQ for the female population. If the t value exceeds the appropriate critical value, then the means for the two populations are not equal. Note that two single means (average male IQ and average female IQ) are compared.

What if a researcher were interested in comparing males and females on the six verbal subscales of the WAIS–R and wanted to do so in a multivariate analysis to account for the correlations between the six scales? Then one is no longer comparing two means. Rather, two sets of six means are being compared. The formal name for such a set is a *vector*. In this example, the male and female groups each have a vector of means consisting of the mean scores for each of the six WAIS–R subscales. Thus, the null hypothesis being tested is that the vector of means for the male population is equal to the vector of means for the female population. MANOVA will produce a test statistic (in this case the multivariate version of a t test, Hotelling's T^2) that will be compared with a critical value to obtain a significance level. If that probability is below the predetermined criterion for significance (e.g., $p < .05$), then it is concluded that the male and female populations have unequal vectors of means. Therefore, the important difference to remember between a univariate null hypothesis and a multivariate null hypothesis is that the univariate hypothesis considers single means, whereas the multivariate hypothesis considers vectors of means.

Just as a univariate ANOVA can test hypotheses concerning the effects of multiple independent variables, so too can a MANOVA. The null hypothesis being tested in a MANOVA is the same as it would be for an ANOVA, except that individual means are replaced with vectors of means. In the 2 (relax) × 2 (cog–behav) panic disorder study, there are two main effects and one interaction effect that can be tested. The null hypothesis for the first main effect is that the vector of the three subscale means (cognitive, emotional, and physiological) for the people in relaxation training is equal to the vector of means for those who are not in relaxation training. The null hypothesis for the cog–behav main effect is that the vector of subscale means for the people in the cognitive–behavioral therapy group will be same as the vector of means for those not in the cognitive–behavioral group. Finally, the null hypothesis for the interaction effect is that the four groups defined by crossing relax and cog–behav will have equal vectors.

Calculating MANOVA Test Statistics

To explain how a MANOVA test statistic is computed, it would be necessary to cover complex matrix operations and countless equations. Such an exposition is beyond the scope of this chapter, and I therefore limit this section to explaining how, at a conceptual level, the derivation of a MANOVA test statistic is the same as the derivation of an ANOVA test statistic. A natural way to start is to review how a univariate ANOVA is performed.

An ANOVA is an examination of means and mean differences, and the logic of the analysis is fairly straightforward. To begin, it is clear that not everyone has the exact same score on a dependent measure. If everyone did have the same score, say a 6.8 out of 10, then the overall (or *grand*) mean would be 6.8 with variance of zero. When not everyone scores exactly the same, then it is possible to express each person's score as a *deviation* from the grand mean. Statistical variance is a function of the sum of the squares of these deviations from a mean for an entire group. This is referred to as *sum of squares* (SS).

The ANOVA seeks to determine how much total variation can be explained by the variables in an experiment by dividing the total variance into two parts: (a) the variance attributable to the variables in the study and (b) the variance attributable to variables not included in the study, or "error." These variance components are called the *between sums of squares* ($SS_{between}$) and the *within sums of squares* (SS_{within}), respectively. Recall that sums of squares refer to the sum of the squared deviations about a mean, which is the primary component that is used to compute variance. One can readily see that the elements of an ANOVA are expressed in terms of differences, differences between an individual's score and the grand mean (SS_{total}), between an individual's score and his or her group's mean (SS_{within}), and between a group mean and the grand mean ($SS_{between}$). Hence, ANOVA tests whether the amount of variance explained by the independent variable ($SS_{between}$) is a significant proportion relative to the variance that has not been explained (SS_{within}).

In MANOVA, the same approach is taken, but the sum of squares is replaced with sum of the square and cross-product (SSCP) matrices. An SSCP matrix is similar to the matrix discussed when reviewing the homogeneity of covariance matrices assumption. An SSCP matrix consists of the sums of the squares (representing variances) for every dependent variable along the diagonal and cross-products (representing covariances) taking up the off-diagonal elements (recall that covariance is the amount

Table 5

Relationships Between Some Primary MANOVA Components

Effect size	B	W	Λ	*p*	η^2
No effect	Smaller	Larger	Larger	Larger	Smaller
Effect present	Larger	Smaller	Smaller	Smaller	Larger

Note. B = between-groups sum of squares cross-product (SSCP) matrix; W = within-groups SSCP matrix; *p* value = the probability of obtaining a particular value of Wilks's lambda; η^2 = the proportion of variance explained. MANOVA = multivariate analysis of variance.

of common variance shared by two variables). Just as a univariate ANOVA posits a total sum of squares and partitions it into between and within sums of squares, the MANOVA posits similar components, but in matrix form. Thus, there is a total (T) SSCP matrix that is divided into a within-groups (W) SSCP matrix and a between-groups (B) SSCP matrix. Matrix algebra allows one to derive a single number that expresses the amount of generalized variance (the variability present in a set of variables) for a particular matrix, call the *determinant.* Hence, one is able to compare the generalized variance of one matrix with another.

What kind of relationship should exist between T, B, and W if there is a significant effect? Because B addresses the amount of variance and covariance the effects can explain, and W is the variance and covariance remaining, then one would expect B to have a larger generalized variance than W. By the same token, one would expect W to be smaller when an effect is present. It is from the latter point that the oldest and most popular multivariate test statistic, Wilks's lambda (Λ), is derived (Tatsuoka, 1971). Basically, Wilks's lambda is a ratio of W to T. Hence, when lambda is small, the variance not explained by the independent variables is small. In order to determine how statistically significant lambda is, it is transformed into either an *F* or a chi-square statistic (Pedhazur, 1982).

If lambda is the proportion of variance not explained, then it follows that $1 - \Lambda$ is the proportion of variance explained by the independent variable's effect. This index is called eta-square (η^2; Huberty & Smith, 1982) and is analogous to other measures of explained variance such as R^2 in multiple regression (Pedhazur, 1982). Unlike some formulations of R^2, eta-squares for several variables are not additive (i.e., they will not add up to one), because different linear composites of the dependent variables are used to determine the effect of each independent variable. Table 5 demonstrates the relationship between the size of an independent

Table 6

2 × 2 MANOVA Results From the Hypothetical Panic Disorder Data

Effect	Λ	F^a	*df*	*p*	η^2
Relaxation	0.65	17.21	3, 94	.0001	.35
Cog–behav	0.71	12.82	3, 94	.0001	.29
Interaction	0.65	16.91	3, 94	.0001	.35

Note. Cog–behav = cognitive–behavioral.
[a] Rao's *F* transformation of Wilks's lambda.

variable's effect, the generalized variance of W and B, Wilks's lambda, the probability associated with the *F* or chi-square transformation of lambda, and eta-square.

There are other test statistics besides Wilks's lambda, all of which are some function of the T, W, and B matrices. Hotelling's T^2 is used when there are only two groups being compared on a set of dependent measures, and it can be transformed to an *F* statistic just as Wilks's lambda can in order to determine its statistical significance. Other test statistics for cases in which there are more than two levels of the independent variable or more than one independent variable include the Hotelling–Lawley trace, Roy's largest root, and the Phillai–Bartlett trace (Tatsuoka, 1971). Regarding which of these latter three statistics should be used, Stevens (1986) reviewed the literature and concluded that "as a general rule, it won't make that much of a difference which of the statistics is used" (p. 187).

As a way of further summarizing, Table 6 shows the preliminary results of the 2 (relaxation) × 2 (cog–behav) MANOVA performed in the panic disorder example. These are preliminary results, because no follow-up procedures were used to interpret the significant group differences observed. Such procedures are discussed in the next section. For now, the reader can observe three tests of significance, each addressing one of the three multivariate effects in the 2 × 2 MANOVA. The first significance test evaluates whether the vector of means for those subjects in relaxation training is equal to the vector of means for those not in relaxation training. The second significance test determines whether the group receiving cognitive–behavior therapy is equal to the group not receiving that therapy with respect to the vector of dependent variable means. Finally, the last test examines the interaction between relaxation and cog-behav to determine whether the four groups formed (relaxation, cog–behav, combined, and control) differ on the vector of dependent means.

For each effect shown in Table 6, there are two indices one should use to interpret the significance of an effect (Huberty & Smith, 1982). The first is the p value, which indicates the probability of obtaining a given effect if there was indeed no real effect in nature. In the panic disorder study, each multivariate effect was statistically significant at the .0001 level. The other criterion that should be used for assessing the significance of a multivariate effect is eta-squared, shown in the last column of Table 6. As I mentioned earlier, the eta-squares for each variable cannot be added together, even though the ones in Table 6 almost add to one. The proportions of explained variance shown in the table—.35, .29, and .35—would be extremely large according to Cohen's (1977) social science standards. Thus, it can be concluded that the three multivariate effects from the 2×2 panic disorder MANOVA are all significant because (a) the small p values suggest that the chance of the results being attributable to Type I error is small and (b) the eta-squares indicate that the effects are explaining nontrivial portions of the variance in the dependent measures.

Follow-Up Analyses for Significant Multivariate Effects

The most difficult part of performing and interpreting a MANOVA is determining what to do if a significant multivariate effect has been obtained. In the panic disorder example, that fact that there was a multivariate main effect for relaxation is interesting, but what does it mean in terms of that treatment's effect on the subjects' panic disorders? There are several procedures available for following up multivariate significance. Five of these procedures are reviewed here: multiple univariate ANOVAs, stepdown analysis, discriminant analysis, dependent variable contribution, and multivariate contrasts.

Multiple Univariate ANOVAs

By far the most popular way of proceeding from a significant effect in MANOVA is to perform univariate ANOVAs for each of the dependent variables (Bray & Maxwell, 1982). The reasoning behind this approach is that the preliminary MANOVA will control for Type I error. If the MANOVA yields significance, then it is considered acceptable to carry out multiple ANOVAs without undue inflation of the experimentwise alpha. Hummel and Sligo (1971) performed a simulation study to compare various methods for analyzing data in a MANOVA context and found that a MANOVA followed by univariate ANOVAs kept the experimentwise error rate the lowest. This often cited study serves as the major justification for using this technique.

This popular approach, however, has been criticized for three major reasons. The first is that Hummel and Sligo's (1971) simulation used data that do not resemble data found in real research situations. "Their use of equicorrelation matrices, in which all off-diagonal elements are equal, made their results applicable to almost no real data" (Wilkinson, 1975, p. 409). Hence, the findings from their study are highly suspect.

More important is the fact that a preliminary MANOVA protects the experimentwise alpha level only when the null hypothesis is true. Bray and Maxwell (1982) described the serious problem that occurs when the multivariate null hypothesis is rejected:

> In many cases the null hypothesis of no group differences is in fact false for one or more but not all variates. Hence, the multivariate test will produce significant results with a high probability if power is sufficient. However, univariate tests will then be performed for those variates for which there is no true differences as well as for those variates for which there is a difference. Because the individual alpha levels are not adjusted despite performing multiple significance tests, the overall multivariate test does not provide 'protection' for each of the p univariate tests. Consequently, in such cases the experimentwise error rate for the set of p univariate F ratios may be inflated above the nominal alpha level, even if the initial MANOVA test was significant. (p. 343)

A third reason for not following a significant MANOVA with univariate ANOVAs is that the separate ANOVAs ignore the correlations between the dependent variables, and as Bray and Maxwell (1982) noted, this means that valuable information concerning redundancies and conceptual relationships is left out. On the same note, Huberty and Morris (1989) distinguished between univariate questions and multivariate questions. Briefly defined, univariate research questions pertain to individual outcome variables, whereas multivariate questions are concerned with experimental effects for multiple dependent variables taken as a set. Huberty and Morris showed how separate univariate ANOVAs are inadequate for addressing multivariate questions.

Of course, some researchers may reply that they were never interested in the multivariate aspect of their data (i.e., in looking at the dependent variables as a set) but were only concerned with minimizing Type

I error. As pointed out earlier, the MANOVA–multiple ANOVA technique does not protect the experimentwise alpha when there is a significant multivariate effect present. If protection against Type I error is the concern, there are more appropriate techniques available, such as the various versions of the Bonferroni correction (see de Cani, 1984; Holland & Copenhaver, 1987, 1988; Holm, 1979; Maxwell & Delaney, 1990).

Stepdown Analysis

Recall that a MANOVA can be used when there is a theoretical relationship between dependent variables. If there is a logical a priori causal ordering of these variables, then a stepdown analysis may be appropriate (Bray & Maxwell, 1982; Stevens, 1986). Consider a study with three dependent variables, A, B, and C, wherein it is theorized that A causes B and B causes C. For such an analysis, a stepdown F is calculated for each dependent variable. For the variable that comes first in the causal ordering (A), the stepdown F is identical to the univariate ANOVA F for that variable. The stepdown F for the second variable (B) is calculated by performing an analysis of covariance (ANCOVA), whereby the influence of A is factored out, so that the unique contribution of B to group separation can be identified. (A more detailed discussion of ANCOVA is presented later when multivariate analysis of covariance [MANCOVA] is considered.) The stepdown F for the last variable, C, is derived from an ANCOVA that removes the effects of A and B, leaving only the variance uniquely associated with C. In this way, each dependent variable's contribution to the multivariate effect can be ascertained. Keep in mind when reading the results of a stepdown analysis that the ordering of the variables is extremely important for interpreting the results correctly.

As an example, consider the three dependent variables for the panic disorder study. Imagine that the experimenter has a theory which posits a biological cause of panic disorder, which produces a physiological state. This state initiates an emotional reaction, which is then conceptualized cognitively. Hence, the logical ordering of the dependent variables would be physiological, emotional, and cognitive. The stepdown F for each variable would give the relative contribution of each variable under the assumption of the causal ordering that was specified.

Discriminant Analysis

Another way of examining significant multivariate effects is to perform a discriminant analysis (DA) for each of the significant effects (Tatsuoka, 1971). A DA is a procedure that maximizes the separation between groups on some categorical variable by finding the optimal linear combination of several continuous variables. Recall that in a MANOVA, the same basic task is accomplished. In a MANOVA, the linear combination of the dependent variables that best separates the levels comprising the independent variables is determined. The DA allows one to discern subsets of the dependent variables that might constitute some underlying dimensions or constructs on which the experimental groups differ. A full discussion of DA is not possible here because of space limitations, and so the reader is referred to chapter 6 of this book, as well as other excellent treatments of the topic in relation to MANOVA (e.g., Huberty, 1984; Huberty & Smith, 1982; Pedhazur, 1982; Stevens, 1986).

Dependent Variable Contribution

Two lesser known procedures attempt to determine how much each dependent variable contributes to a multivariate effect. These techniques differ from stepdown analysis in that they look at the decrease in the multivariate effect as a function of removing a particular variable. Stepdown analysis tests one dependent variable at a time, controlling for the variables that precede it in the a priori causal ordering.

The first of these two techniques that evaluate the contribution of each dependent variable is from Wilkinson (1975). He suggested performing successive MANOVAs in which one dependent variable is left out in each analysis. The change in the multivariate F is examined to determine which variables are contributing the most to the multivariate effect and which are not. In other words, the procedure attempts to reveal which variable a researcher cannot do without. Of course, such a decision is not purely statistical. It might be the case that a particular variable does not contribute as strongly to the multivariate effect as others do, but the variable in question is essential to the model from a theoretical standpoint. A researcher might decide that a variable's theoretical necessity may supersede its statistical necessity.

The other approach for determining variable contribution is from Huberty and his associates (Huberty & Morris, 1989; Huberty & Smith, 1982). For a given set of dependent variables, Huberty recommended using an F-to-remove statistic for each dependent variable: "An F-to-remove tests the significance of the decrease in group separation if variable i is removed from the entire set of variables" (Huberty & Smith, 1982, p. 421). The F-to-remove values for the dependent variables are then rank ordered to determine the relative importance of each. There are strong similarities between Wilkinson's (1975) and Huberty's methods, both in mathematical form and practical function.

Although neither of these two methods are yet frequently seen in the journals, I discuss them here because they are viable approaches. Hence they might appear in future articles with increasing frequency.

Multivariate Contrasts

Whereas the previous procedures focused on the dependent variables (e.g., which contributes to the overall effect, what dimension underlies the variates, etc.), contrast analyses focus on the groups defined by the independent variables. Multivariate contrasts compare groups over a set of dependent variables simultaneously (Huberty & Morris, 1989). In other words, multivariate contrasts compare vectors of means. As Stevens (1986) observed, a contrast analysis is essentially a comparison between two groups. For simple (or pairwise) contrasts, the two groups being compared are two levels of some independent variable (e.g., male vs. female for the variable of gender). For complex contrasts, the two groups being compared can be combinations of levels of several dependent variables. Consider a study with four different treatments: T1 through T4. A simple contrast might be to compare T1 and T3. A complex analysis might compare T4 with a group formed by combining T1, T2, and T3. The latter contrast is complex because a group has been formed from the preexisting groups T1, T2, and T3.

The multivariate test statistic for comparing two groups in a multivariate contrast is Hotelling's T^2, which is transformed to an F to obtain a probability level. There are two basic ways of following up a significant multivariate comparison. The first is to use discriminant function analysis or obtain F-to-remove statistics as in the follow-up procedures for a significant MANOVA effect (Huberty & Smith, 1982). This method has already been discussed earlier. Note that this first approach retains the "multivariate" aspect of the data because it considers all of the dependent variables at once.

A second contrast procedure examines each dependent variable separately. Hence, a significant T^2 is followed by one of several methods for univariate pairwise comparisons such as multiple t tests, Tukey confidence intervals, Roy–Bose simultaneous confidence intervals, and modified Bonferroni procedures (Bray & Maxwell, 1982; Stevens, 1986). Each of these techniques have specific pros and cons regarding the control of

Type I error and power that are beyond the limits of this chapter (the interested reader can consult Keselman et al., 1991; Kleinbaum et al., 1988; Seaman et al., 1991; Stevens, 1986).

MANCOVA, Repeated Measures MANOVA, and Power Analysis

MANCOVA

In an ANOVA design, it is often desirable to eliminate as much error variance as possible so that a larger portion of the remaining variance is attributable to treatment effects or group differences (Kleinbaum et al., 1988). One way of reducing error variance is to measure a *covariate*—a continuous variable known to affect the dependent measures—whose effect on the dependent variable can then be factored out of the total variance. Hence, the levels of the independent variables are being compared using means on the dependent variable that have been adjusted using the covariate. In a multivariate analysis, covariates can be used in the same manner, changing a MANOVA to a MANCOVA—a multivariate analysis of covariance. Instead of comparing vectors of means, the MANCOVA compares vectors of adjusted means.

Consider the panic disorder study. The researcher assigned each subject a severity rating before the treatment period. It would be expected that the more severe cases of panic disorder would have generally higher scores on the cognitive, emotional, and physiological subscales, and so this severity rating might be an effective covariate. A MANCOVA could be performed to determine whether the vectors of adjusted means differed across the experimental groups.

Before conducting the analysis, however, the covariate needs to be carefully examined. A covariate should only be used if (a) there is a statistically significant linear relationship between the covariate and the dependent measures and (b) an additional assumption—the homogeneity of the regressions—is satisfied. The first condition can be tested with a simple correlation between the covariate and each dependent measure. For the panic disorder study, the severity rating correlated significantly with each of the dependent measures at the .05 level. The second condition, the homogeneity of regression hyperplanes assumption, requires the experimental groups to have equal regression slopes for the covariate. Another way of saying this is that the relationship between the covariate and the dependent measures must be equal for all of the groups. If this

Table 7

2 × 2 MANCOVA Results From the Hypothetical Panic Disorder Data

Effect	Λ	F	df	p	η^2
Relaxation	0.78	8.96	3, 93	.0001	.22
Cog–behav	0.87	4.74	3, 93	.005	.13
Interaction	0.74	10.65	3, 93	.0001	.26

Note. MANCOVA = multivariate analysis of variance; Cog–behav = cognitive–behavioral.

assumption was not satisfied, there would be a Group × Covariate interaction that would be overlooked. In the panic disorder MANCOVA, there was no interaction for the Severity × Relaxation or Severity × Cog-Behav conditions, and so the assumption is tenable.

Tests of the multivariate effects of relaxation, cog–behav, and the interaction between the two are performed. These tests, along the eta-squares, are shown in Table 7. Reading the results of a MANCOVA is exactly the same as reading the results of a MANOVA, except that in the former, variance attributable to individual differences on some covariate is partitioned out prior to examining the effects. Note that the F values for each effect are smaller than the corresponding F values for the 2 × 2 MANOVA presented earlier. Introducing a covariate can influence the F statistic in two ways. To the extent that the covariate accounts for variance in the dependent measures that is not explained by the independent variables, the F statistic will be larger. This is because the within-groups SSCP matrix W (reflecting unexplained variance) will be smaller, thereby making Wilks's lambda smaller. To the extent that the covariate accounts for variance in the dependent measures that is shared by an independent variable, the F statistic will decrease. This happens because the mean vectors of the dependent variables for the groups defined by the independent variable are brought closer together after adjusting for the covariate. In this example, the latter influence was evidently more predominant.

That the results of the MANCOVA are the same as they are for the regular MANOVA will not always be the case. In many instances, when the effect of a covariate on the set of dependent measures is removed, the effect of the remaining independent variables may be negligible. Such a finding may suggest one of the following: (a) The independent variable has an indirect effect on the dependent measures by working through the covariate; (b) the covariate has an indirect effect on the dependent measures via the independent variable, such that the independent variable does not explain variation in the dependent variables absent the covariate's effect on the independent variable; and (c) the covariate and independent measures are measures of the same variable, and so the presence of one will render the other's effect redundant.

Repeated Measures MANOVA

One of the most powerful and efficient research designs is the repeated measures design (Shaughnessy & Zechmeister, 1990). The term *repeated measures design* refers to a situation in which subjects are measured on more than one occasion. This design is powerful because error variance is reduced substantially and is efficient because fewer subjects are needed than in nonrepeated measures experimental designs. This can happen when subjects are assessed before and after a treatment (pretest–posttest) or when each subject participates in more than one experimental condition. To facilitate discussion, imagine that the author of the panic disorder study measured each subject on the cognitive, emotional, and physiological subscales before the intervention, as well as after. This would transform the study into a repeated measures design.

Researchers use various terms when describing repeated measures designs, and so it is helpful to review some of them here. Independent variables such as time (T1, T2, T3, etc.) or trial (Trial1, Trial2, etc.) are called *within* or *within-subjects* variables. For a within-subjects variable, each subject is measured on the dependent variables for every level of the within-subjects variable. Hence, a within-subjects variable in one study might be time, consisting of four points during the study when subjects were assessed on the dependent variables. A *between* or *between-subjects* variable is a grouping variable such as age or gender. A study can involve both within- and between-subjects independent variables. Hence, the modified panic disorder study constitutes a 2 (time) × 2 (relaxation) × 2 (cog–behav) repeated measure design, with repeated measures on the first variable (time). The other two variables—relaxation and cog–behav—are between-subjects variables.

In the standard MANOVA, vectors of means across levels of the independent variables are compared. For the repeated measures MANOVA, vectors of *mean differences* are compared across levels of the independent variables. These mean differences refer to differences in the value of the dependent measures between levels of the within-subjects variable. Thus, if GSR was measured four different times (T1 through T4) for each subject, the three difference scores for each subject would be T1–T2, T2–T3, and T3–T4. The mean difference variables would comprise the dependent measure vector, not the original scores.

A within-subjects effect such as time or trial can be analyzed in a manner analogous to a between-subjects effect. Returning to the GSR example, suppose that a researcher wanted to compare GSR with stressful stimuli over four time periods for male subjects and female subjects. This would constitute a 4 (time) × 2 (gender) repeated measures MANOVA. The MANOVA would produce three significance tests corresponding to three effects that can be tested: (a) the main effect for time (i.e., Do the subjects' GSRs differ over time?); (b) the main effect for gender (i.e., Do males and females differ with respect to GSR, ignoring the effect of time?); and (c) the interaction effect (i.e., Does the GSR of males change over time at a different rate than females?).

When conducting a repeated measures MANOVA, an additional assumption must be met. This *sphericity assumption* (Huynh & Mandeville, 1979) concerns the difference variables that are created from the original dependent variables. In the GSR and gender example, there were three such transformed variables, with each reflecting the difference between one pair of the original four dependent measures. The sphericity condition requires that the covariance matrix for the transformed variables be a diagonal matrix in which the values along the diagonal are equal and all the off-diagonal elements must be zero. In other words, the sphericity assumption requires that the variances of the transformed variables will be equal and the transformed variables are not intercorrelated. Failure to meet the sphericity condition will result in an inflated Type I error rate (Stevens, 1986).

One of the more sophisticated within-subjects analyses is the so-called *doubly multivariate repeated measures MANOVA* (Norusis, 1988, p. 283; SAS Institute, 1990, p. 988). A typical repeated measures MANOVA has only one dependent measure, such as GSR, which is measured multiple times. In the doubly multivariate repeated measure design, two or more dependent variables are measured at two or more points in time. Hence, the modified version of the panic disorder study constitutes such a design. There are three dependent measures (the three subscales) assessed twice for each subject (pretest and posttest). Recall that in the standard repeated measures MANOVA, a vector of mean differences for each group is compared. In the doubly multivariate design, a matrix of difference variables is used, because there is more than one dependent measure being analyzed.

Table 8

2 × 2 × 2 Doubly Multivariate Repeated Measures MANOVA Results From the Hypothetical Panic Disorder Data

Multivariate effect	Λ	*F*	*df*	*p*	η^2
Time	0.35	58.90	3, 94	.0001	.65
Relaxation	0.77	9.30	3, 94	.0001	.23
Cog–behav	0.81	7.13	3, 94	.0005	.19
Relaxation × Cog–Behav	0.74	10.99	3, 94	.0001	.26
Time × Relaxation	0.75	10.34	3, 94	.0001	.25
Time × Cog–Behav	0.80	7.94	3, 94	.0001	.20
Time × Relaxation × Cog–Behav	0.83	6.49	3, 94	.001	.17

Note. MANOVA = multivariate analysis of variance, Cog–Behav = cognitive–behavioral.

Table 8 shows the results of the doubly multivariate repeated measures MANOVA for the panic disorder study. Six different multivariate effects can be tested. The first is the main effect for time, which addresses the question of whether the dependent measures, taken as a set, change over time. The next three tests for the relaxation, cog–behav, and relaxation × cog–behav effects are somewhat superfluous, because they test for group differences collapsing across the two testing periods (although the reader may observe that each of these effects were significant). Of primary interest is the question of which groups did better on the posttest compared with the pretest. Hence, the Time × Relaxation, Time × Cog–Behav, and Time × Relaxation × Cog–Behav interactions should be examined. As Table 8 shows, all three interactions are highly significant, indicating that the change in pretest–posttest scores for the three subscales is different for the four experimental groups. Remember that the eta-squares cannot be combined.

Power Analysis in MANOVA

Earlier discussion concerned the notion of statistical power. In this section, the issue of power is addressed once again, focusing specifically on how power can be assessed for a MANOVA. It is desirable to know the power of a MANOVA for two reasons. First, one may wish to know the power a priori to determine the sample size needed to detect a given effect. This would be the case when one is designing a study or reviewing a proposal for a study. Second, one may be interested in a post hoc estimation of power to ascertain how well a MANOVA could detect an effect in a given study that was already done.

Recall that power is a function of sample size, the nominal alpha level specified, and effect size. Sample size and the nominal alpha level can easily be obtained from either a univariate or multivariate analysis, but little work has been done on measuring effect sizes for a MANOVA. The best work in this area has come from Stevens (1980, 1986). The effect size for a two-group univariate analysis is typically expressed as d and is a function of the difference between the two group means. Stevens showed that for the two-group multivariate case, the effect size can be measured as D^2, a function of the difference between the two vectors of means (see Stevens, 1986, p. 140). Determining the effect size when there are more than two experimental groups is more complicated. Those interested in determining the power of a MANOVA are strongly advised to consult Stevens (1980, 1986) for explicit information on how it is done, including easy-to-understand tables showing the power of a MANOVA for various numbers of dependent variables, groups, and subjects.

Conclusion

Recall the Results section paragraph that appeared on the first page of this chapter. What might have seemed intimidating and confusing earlier should now be more clear. The 4 (group) × 2 (time) MANOVA describes a repeated measures MANOVA with one between-subjects variable (group) and one within-subjects variable (time). The group variable has four levels corresponding to four groups, and the time variable has two levels corresponding to two testing periods. The values of Wilks's lambda, their F transformations, and their respective p values indicate that the only statistically significant independent variable was time. This denotes that the vector of dependent variable means for the first time period was significantly different from the vector of means for the second time period. For instance, one might say that scores on the dependent measures were higher at Time 2 than at Time 1. The eta-square listed for the time effect (.59) suggests that the effect explained over half of the variance in the dependent measures. The absence of a significant group effect means that the mean vectors for each of the four groups were equivalent. The absence of a significant Group × Time interaction indicates that the effect of one variable did not depend on the level of the other variable.

Suggestions for Further Reading

For introductory-level treatments of research designs and univariate ANOVA, see Campbell and Stanley (1963), Grimm (1993), Guilford (1965), Kerlinger (1986), Kleinbaum, Kupper, and Muller (1988), Pedhazur (1982), Shaughnessy and Zechmeister (1990), and Stevens (1990).

More advanced discussions of power analysis can be found in Cohen (1977) and Stevens (1980, 1986). Advanced discussions of Type I error rates and multiple comparison procedures can be found in de Cani (1984), Holland and Copenhaver (1987, 1988), Holm (1979), Keselman, Keselman, and Games (1991), Maxwell and Delaney (1990), and Seaman, Levin, and Serlin (1991).

Detailed treatments of matrix algebra and calculation of MANOVA test statistics can be found in Green (1978), Pedhazur (1982), Stevens (1986), and Tatsuoka (1971). For comprehensive review of MANOVA follow-up techniques, see Bray and Maxwell (1982), Maxwell and Delaney (1990), and Stevens (1986). For detailed discussions of particular follow-up procedures, see Huberty and Morris (1989), Huberty and Smith (1982), Tatsuoka (1971), and Wilkinson (1975).

Glossary

When appropriate, the symbol associated with a particular term is included in parentheses after the term.

ALPHA (α) The probability of rejecting the null hypothesis when it is true.

BETWEEN GROUPS The variation accounted for by the independent variables.

BETWEEN SUBJECTS A variable on which each subject can be found on only one level of the variable, such as age or gender.

CHI-SQUARE (χ^2) An inferential test statistic that multivariate statistics can be transformed to in order to derive a probability level.

COVARIANCE The variation in one variable that is shared by another variable.

COVARIATE A continuous variable that has a significant linear relationship with some dependent variable. Covariates are used to reduce error variance in an ANCOVA and MANCOVA.

EFFECT SIZE The magnitude of an independent variable's effect, usually expressed as a proportion of explained variance in the dependent variables.

ETA-SQUARE (η^2) An index of the proportion of explained variance; can vary from 0 to 1.

F STATISTIC A test statistic to which multivariate indices can be transformed, to derive a probability level.

HOTELLING'S T^2 A multivariate test statistic used when there is one independent variable with only two levels.

INTERACTION EFFECT An interaction occurs when the effect of an independent variable on some dependent variable depends on the level of another independent variable.

MAIN EFFECT The effect of a single independent variable on one or more dependent variables.

NULL HYPOTHESIS (H_0) The hypothesis that states that there is no difference in the mean values of one or more dependent variables across levels of one or more independent variables.

OMNIBUS TEST A test of the null hypothesis that none of the independent variables has an effect on any of the dependent variables.

POWER The probability of detecting a significant effect when the effect truly exists in nature.

PROBABILITY (p) The probability of falsely rejecting the null hypothesis.

REPEATED MEASURES An experimental design and corresponding analysis in which each subject is measured on the dependent variable for more than one level of an independent variable (e.g., Time 1 and Time 2).

SSCP A matrix with sums of squares for each dependent variable on the diagonal and sums of cross-products for every pair of variables filling the rest of the elements.

SUM OF CROSS-PRODUCTS For a set of observations, it is the sum of the products of each subjects' squared deviations from the mean on one variable and the squared deviation from the mean of another variable. It is an index of covariance.

SUM OF SQUARES The sum of the squared deviations about a mean for a set of observations. It is an index of variance used in analysis of variance and covariance procedures.

TYPE I ERROR Rejecting the null hypothesis when it is really true.

TYPE II ERROR Accepting the null hypothesis when it is really false.

VARIANCE The average of the squared deviations from a mean for a set of observations. It reflects the dispersion of values around a mean.

WILKS'S LAMBDA (Λ) A multivariate test statistic that expresses the proportion of unexplained variance in the dependent measures.

WITHIN GROUP The variation in the dependent measure that is not explained by the independent variables.

WITHIN SUBJECT An independent variable used in such a way that dependent variable values are obtained for every level of the within-subject variable.

References

Antony, M. M., Brown, T. A., & Barlow, D. H. (1992). Current perspectives on panic and panic disorder. *Current Directions in Psychologial Science, 1*, 79–82.

Bray, J. H., & Maxwell, S. E. (1982). Analyzing and interpreting significant MANOVAs. *Review of Educational Research, 52*, 340–367.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* San Diego, CA: Academic Press.

de Cani, J. S. (1984). Balancing Type 1 risk and loss of power in ordered Bonferroni procedures. *Journal of Educational Psychology, 76*, 1035–1037.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments.* New York: Wiley.

Green, P. E. (1978). *Analyzing multivariate data.* Hinsdale, IL: Dryden Press.

Grimm, L. (1993). *Statistical applications for the behavioral sciences.* New York: Wiley.

Guilford, J. P. (1965). *Fundamental statistics in psychology and education* (4th ed.). New York: McGraw-Hill.

Holland, B. S., & Copenhaver, M. D. (1987). An improved sequentially rejective Bonferroni procedure. *Biometrics, 43*, 417–423.

Holland, B. S., & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. *Psychological Bulletin, 104*, 145–149.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandanavian Journal of Statistics, 6*, 65–70.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. *Psychological Bulletin, 95*, 156–171.

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin, 105*, 302–308.

Huberty, C. J., & Smith, J. D. (1982). The study of effects in MANOVA. *Multivariate Behavioral Research, 17*, 417–482.

Hummel, T. J., & Sligo, J. R. (1971). Empirical comparison of univariate and multivariate analysis of variance procedures. *Psychological Bulletin, 76*, 49–57.
Huynh, H., & Mandeville, G. K. (1979). Validity conditions in repeated measures designs. *Psychological Bulletin, 86*, 964–973.
Kaplan, R. M., & Saccuzzo, D. P. (1989). *Psychological testing: Principles, applications, and applications* (2nd ed.). Pacific Grove, CA: Brooks/Cole.
Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart & Winston.
Keselman, H. J., Keselman, J. C., & Games, P. A. (1991). Maximum familywise Type I error rate: The least significant difference, Newman-Keuls, and other multiple comparison procedures. *Psychological Bulletin, 110*, 155–161.
Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariable methods* (2nd ed.). Boston: PWS-KENT.
Larsen, R. J. (1984). Theory and measurement of affect intensity as an individual difference characteristic. *Dissertation Abstracts International, 85*, 2297B. (University Microfilms No. 84-22112)
Larsen, R. J., & Diener, E. (1987). Affect intensity as an individual differences characteristic: A review. *Journal of Research in Personality, 21*, 1–39.
Maxwell, S. E., & Delaney, M. D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.
Mendoza, J. L., & Graziano, W. G. (1982). The statistical analysis of dyadic social behavior: A multivariate approach. *Psychological Bulletin, 92*, 532–540.
Norusis, M. J. (1988). *SPSS advanced statistics guide* (2nd ed.). Chicago: SPSS.
Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association, 69*, 894–908.
Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). Fort Worth, TX: Holt, Rinehart & Winston.
Ryan, T. A. (1959). Multiple comparisons in psychological research. *Psychological Bulletin, 56*, 26–47.
SAS Institute. (1990). *SAS/STAT user's guide* (4th ed.). Cary, NC: SAS Institute.
Seaman, M. A., Levin, J. R., & Serlin, R. (1991). New developments in pairwise multiple comparisons: Some powerful and practicable procedures. *Psychological Bulletin, 110*, 577–586.
Shaughnessy, J. J., & Zechmeister, E. B. (1990). *Research methods in psychology* (2nd ed.). New York: McGraw-Hill.
Stevens, J. P. (1980). Power of the multivariate analysis of variance tests. *Psychological Bulletin, 88*, 728–737.
Stevens, J. (1986). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Erlbaum.
Stevens, J. (1990). *Intermediate statistics: A modern approach.* Hillsdale, NJ: Erlbaum.
Tatsuoka, M. M. (1971). *Multivariate analysis: Techniques for educational and psychological research.* New York: Wiley.
Wilkinson, L. (1975). Response variable hypothesis in the multivariate analysis of variance. *Psychological Bulletin, 82*, 408–412.

9 Discriminant Analysis

A. Pedro Duarte Silva and Antonie Stam

Descriptive discriminant analysis (DA) is a statistical technique that allows one to identify *variables* (also known as *attributes*) that best discriminate members of two or more groups from one another. *Predictive* or *prescriptive* DA allows one to predict the group membership status of *subjects* (also known as *observations*, *cases*, or *entities*) of which the group status is unknown. For example, imagine that a researcher had a sample of patients (subjects) who had undergone a heart transplant operation (this sample of subjects is known as the *training* or *development* sample). For each patient, information was available regarding blood pressure, age, number of white blood cells, and percentage of normal weight. In addition, it was known which patients had survived for at least a year after the operation and which patients had died before 1 year (thus, the groups were alive or dead). Descriptive DA would be used to determine how well the variables allowed one to discriminate between the two groups. In this instance, the investigator could use a combination of the blood pressure, age, number of white blood cells, and percentage of normal weight variables to discriminate among those patients who had died and those who had survived. Some variables might be found to be very important, whereas others might turn out to be irrelevant. Now suppose that a new patient was considered for a transplant operation. In predictive DA, the classification rule derived using the training sample would be used to combine information concerning this new patient's blood pressure, age, and so forth, to predict whether the patient would survive the operation for at least 1 year.

In this chapter, we introduce a class of techniques for discriminating