CHAPTER

7 Multivariate Analysis of Variance and Covariance

7.1 General Purpose and Description

Multivariate analysis of variance (MANOVA) is a generalization of ANOVA to a situation in which there are several DVs. For example, suppose a researcher is interested in the effect of different types of treatments on several types of anxieties: test anxiety, anxiety in reaction to minor life stresses, and so-called free-floating anxiety. The IV is different treatment with three levels (desensitization, relaxation training, and a waiting-list control). After random assignment of subjects to treatments and a subsequent period of treatment, subjects are measured for test anxiety, stress anxiety, and free-floating anxiety. Scores on all three measures for each subject serve as DVs. MANOVA is used to ask whether a combination of the three anxiety measures varies as a function of treatment. MANOVA is statistically identical to discriminant analysis, the subject of Chapter 9. The difference between the techniques is one of emphasis only. MANOVA emphasizes the mean differences and statistical significance of differences among groups. Discriminant analysis emphasizes prediction of group membership and the dimensions on which groups differ.

ANOVA tests whether mean differences among groups on a single DV are likely to have occurred by chance. MANOVA tests whether mean differences among groups on a combination of DVs are likely to have occurred by chance. In MANOVA, a new DV that maximizes group differences is created from the set of DVs. The new DV is a linear combination of measured DVs, combined so as to separate the groups as much as possible. ANOVA is then performed on the newly created DV. As in ANOVA, hypotheses about means in MANOVA are tested by comparing variances—hence multivariate analysis of variance.

In factorial or more complicated MANOVA, a different linear combination of DVs is formed for each main effect and interaction. If gender of subject is added to the example as a second IV, one combination of the three DVs maximizes the separation of the three treatment groups, a second combination maximizes separation of women and men, and a third combination maximizes separation of the cells of the interaction. Further, if the treatment IV has more than two levels, the DVs can be recombined in yet other ways to maximize the separation of groups formed by comparisons.[1]

MANOVA has a number of advantages over ANOVA. First, by measuring several DVs instead of only one, the researcher improves the chance of discovering what it is that changes as a result of different treatments and their interactions. For instance, desensitization may have an advantage over

[1]The linear combinations themselves are of interest in discriminant analysis (Chapter 9).

relaxation training or waiting-list control, but only on test anxiety; the effect is missing if test anxiety isn't one of your DVs. A second advantage of MANOVA over a series of ANOVAs when there are several DVs is protection against inflated Type I error due to multiple tests of (likely) correlated DVs.

Another advantage of MANOVA is that, under certain, probably rare conditions, it may reveal differences not shown in separate ANOVAs. Such a situation is shown in Figure 7.1 for a one-way design with two levels. In this figure, the axes represent frequency distributions for each of two DVs, Y_1 and Y_2. Notice that from the point of view of either axis, the distributions are sufficiently overlapping that a mean difference might not be found in ANOVA. The ellipses in the quadrant, however, represent the distributions of Y_1 and Y_2 for each group separately. When responses to two DVs are considered in combination, group differences become apparent. Thus, MANOVA, which considers DVs in combination, may occasionally be more powerful than separate ANOVAs.

But there are no free lunches in statistics, either. MANOVA is a substantially more complicated analysis than ANOVA. There are several important assumptions to consider, and there is often some ambiguity in interpretation of the effects of IVs on any single DV. Further, the situations in which MANOVA is more powerful than ANOVA are quite limited; often MANOVA is considerably less powerful than ANOVA, particularly in finding significant group differences for a particular DV. Thus, our recommendation is to think very carefully about the need for more than one DV in light of the added complexity and ambiguity of analysis and the likelihood that multiple DVs may be redundant (see also Section 7.5.3). Even moderately correlated DVs diminish the power of MANOVA. Figure 7.2 shows a set of hypothetical relationships between a single IV and four DVs. DV1 is highly related to the IV and shares some variance with DV2 and DV3. DV2 is related to both DV1 and DV3 and shares very little unique variance with the IV, although by itself in a univariate ANOVA might be related to the IV. DV3 is somewhat related to the IV, but also to all of the other DVs. DV4 is highly

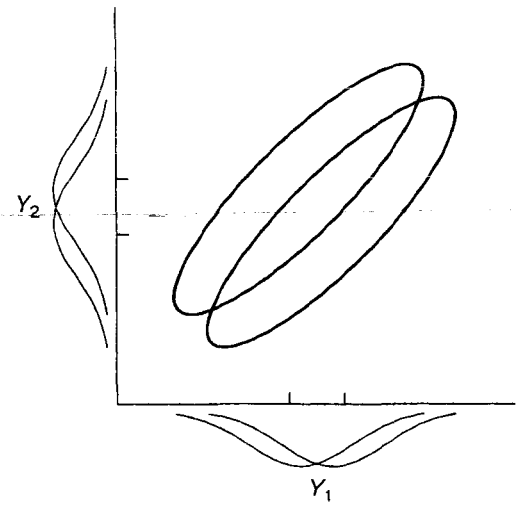


FIGURE 7.1 Advantage of MANOVA, which combines DVs, over ANOVA. Each axis represents a DV; frequency distributions projected to axes show considerable overlap, while ellipses, showing DVs in combination, do not.

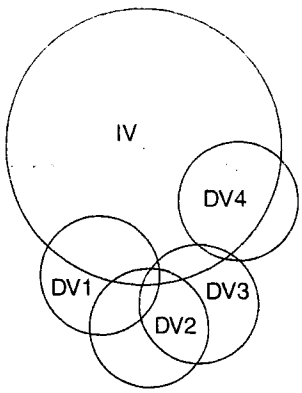


FIGURE 7.2 Hypothetical relationships among a single IV and four DVs.

related to the IV and shares only a little bit of variance with DV3. Thus, DV2 is completely redundant with the other DVs, and DV3 adds only a bit of unique variance to the set. However, DV2 would be useful as a CV if that use made sense conceptually. DV2 reduces the total variance in DV1 and DV2, and most of the variance reduced is *not* related to the IV. Therefore, DV2 reduces the error variance in DV1 and DV3 (the variance that is not overlapping with the IV).

Multivariate analysis of covariance (MANCOVA) is the multivariate extension of ANCOVA (Chapter 6). MANCOVA asks if there are statistically significant mean differences among groups after adjusting the newly created DV for differences on one or more covariates. For the example, suppose that before treatment subjects are pretested on test anxiety, minor stress anxiety, and free-floating anxiety. When pretest scores are used as covariates, MANCOVA asks if mean anxiety on the composite score differs in the three treatment groups, after adjusting for preexisting differences in the three types of anxieties.

MANCOVA is useful in the same ways as ANCOVA. First, in experimental work, it serves as a noise-reducing device where variance associated with the covariate(s) is removed from error variance; smaller error variance provides a more powerful test of mean differences among groups. Second, in nonexperimental work, MANCOVA provides statistical matching of groups when random assignment to groups is not possible. Prior differences among groups are accounted for by adjusting DVs as if all subjects scored the same on the covariate(s). (But review Chapter 6 for a discussion of the logical difficulties of using covariates this way.)

ANCOVA is used after MANOVA (or MANCOVA) in Roy-Bargmann stepdown analysis where the goal is to assess the contributions of the various DVs to a significant effect. One asks whether, after adjusting for differences on higher-priority DVs serving as covariates, there is any significant mean difference among groups on a lower-priority DV. That is, does a lower-priority DV provide additional separation of groups beyond that of the DVs already used? In this sense, ANCOVA is used as a tool in interpreting MANOVA results.

Although computing procedures and programs for MANOVA and MANCOVA are not as well developed as for ANOVA and ANCOVA, there is in theory no limit to the generalization of the model, despite complications that arise. There is no reason why all types of designs—one-way, factorial, repeated measures, nonorthogonal, and so on—cannot be extended to research with several DVs. Questions of effect size, specific comparisons, and trend analysis are equally interesting with

MANOVA. In addition, there is the question of importance of DVs—that is, which DVs are affected by the IVs and which are not.

MANOVA developed in the tradition of ANOVA. Traditionally, MANOVA was applied to experimental situations where all, or at least some, IVs are manipulated and subjects are randomly assigned to groups, usually with equal cell sizes. Discriminant analysis (Chapter 9) developed in the context of nonexperimental research where groups are formed naturally and are not usually the same size. MANOVA asks if mean differences among groups on the combined DV are larger than expected by chance; discriminant analysis asks if there is some combination of variables that reliably separates groups. But there is no mathematical distinction between MANOVA and discriminant analysis. At a practical level, computer programs for discriminant analysis are more informative but are also, for the most part, limited to one-way designs. Therefore, analysis of one-way MANOVA is deferred to Chapter 9 and the present chapter covers factorial MANOVA and MANCOVA.

Mason (2003) used a 2 × 5 between-subjects MANOVA to investigate male and female high school students' beliefs about math. The six scales serving as DVs were in agreement with items concerning ability to solve difficult math problems, need for complex procedures for word problems, importance of understanding concepts, importance of word problems, effect of effort, and usefulness of math in everyday life. Multivariate tests of both main effects were statistically significant, but the interaction was not. Post hoc Tukey HSD tests were used to investigate the individual DVs. Belief in usefulness of math and need for complex procedures increased over the grades; belief in ability to solve difficult problems increased from the first to second year and then decreased. Girls were found to be more likely to believe in the importance of understanding concepts than boys.

A more complex MANOVA design was employed by Pisula (2003) who studied responses to novelty in high- and low-avoidance rats. IVs were sex of rat, subline (high vs. low avoidance), and 8 time intervals. Thus, this was a 2 × 2 × 8 mixed between-between-within MANOVA. DVs were four durations spent inside various zones, duration of object contact, duration of floor sniffing, and number of walking onsets. Multivariate results were not reported, but the table of (presumably) univariate F tests suggests significant results for all effects except the sex by subline interaction. All DVs showed significant differences over trials. All DVs associated with time spent inside various zones also showed significant differences between high- and low-avoidance sublines, as did number of walking onsets. Duration of object contact and number of walking onsets showed sex differences. All DVs except walking onsets also showed significant two-way interactions, and duration of object contact showed a significant three-way interaction.

A MANCOVA approach was taken by Hay (2003) to investigate quality of life variables in bulimic eating disorders. Two types of disorders were identified: regular binge eating and extreme weight control. These each were compared with a non–eating-disordered group in separate MANCOVAs. It is not clear why (or if) these were not combined into a single three-group one-way MANCOVA with planned comparisons between each eating-disorder group and the comparison group. Covariates were age, gender, income level, and BMI (body mass index). Three sets of DVs (physical and mental health components of SF-36 scores, eight SF-36 subscale scores, and six utility AqoL scores) were entered into three separate MANCOVAs for each of the comparisons, resulting in a total of 6 MANCOVAs. The emphasis in interpretation was on variance explained (η^2) for each analysis. For example, 23% of the variance in mental and physical scores was associated with regular binge eating after adjusting for CVs, but only 5% of the variance was associated with extreme weight control behaviors. Similarly, binge eating was associated with greater variance in SF-36 subscale scores and in AqoL scores than were extreme weight control behavior.

7.2 Kinds of Research Questions

The goal of research using MANOVA is to discover whether behavior, as reflected by the DVs, is changed by manipulation (or other action) of the IVs. Statistical techniques are currently available for answering the types of questions posed in Sections 7.2.1 through 7.2.8.

7.2.1 Main Effects of IVs

Holding all else constant, are mean differences in the composite DV among groups at different levels of an IV larger than expected by chance? The statistical procedures described in Sections 7.4.1 and 7.4.3 are designed to answer this question, by testing the null hypothesis that the IV has no systematic effect on the optimal linear combination of DVs.

As in ANOVA, "holding all else constant" refers to a variety of procedures: (1) controlling the effects of other IVs by "crossing over" them in a factorial arrangement, (2) controlling extraneous variables by holding them constant (e.g., running only women as subjects), counterbalancing their effects, or randomizing their effects, or (3) using covariates to produce an "as if constant" state by statistically adjusting for differences on covariates.

In the anxiety-reduction example, the test of main effect asks: Are there mean differences in anxiety—measured by test anxiety, stress anxiety, and free-floating anxiety—associated with differences in treatment? With addition of covariates, the question is: Are there differences in anxiety associated with treatment, after adjustment for individual differences in anxiety prior to treatment?

When there are two or more IVs, separate tests are made for each IV. Further, when sample sizes are equal in all cells, the separate tests are independent of one another (except for use of a common error term) so that the test of one IV in no way predicts the outcome of the test of another IV. If the example is extended to include gender of subject as an IV, and if there are equal numbers of subjects in all cells, the design produces tests of the main effect of treatment and of gender of subject, the two tests independent of each other.

7.2.2 Interactions among IVs

Holding all else constant, does change in the DV over levels of one IV depend on the level of another IV? The test of interaction is similar to the test of main effect, but interpreted differently, as discussed more fully in Chapter 3 and in Sections 7.4.1 and 7.4.3. In the example, the test of interaction asks: Is the pattern of response to the three types of treatments the same for men as it is for women? If the interaction is significant, it indicates that one type of treatment "works better" for women while another type "works better" for men.

With more than two IVs, there are multiple interactions. Each interaction is tested separately from tests of other main effects and interactions, and these tests (but for a common error term) are independent when sample sizes in all cells are equal.

7.2.3 Importance of DVs

If there are significant differences for one or more of the main effects or interactions, the researcher usually asks which of the DVs are changed and which are unaffected by the IVs. If the main effect of treatment is significant, it may be that only test anxiety is changed while stress anxiety and free-floating

anxiety do not differ with treatment. As mentioned in Section 7.1, Roy-Bargmann stepdown analysis is often used where each DV is assessed in ANCOVA with higher-priority DVs serving as covariates. Stepdown analysis and other procedures for assessing importance of DVs appear in Section 7.5.3.

7.2.4 Parameter Estimates

Ordinarily, marginal means are the best estimates of population parameters for main effects and cell means are the best estimates of population parameters for interactions. But when Roy-Bargmann stepdown analysis is used to test the importance of the DVs, the means that are tested are adjusted means rather than sample means. In the example, suppose free-floating anxiety is given first, stress anxiety second, and test anxiety third priority. Now suppose that a stepdown analysis shows that only test anxiety is affected by differential treatment. The means that are tested for test anxiety are not sample means, but sample means adjusted for stress anxiety and free-floating anxiety. In MANCOVA, additional adjustment is made for covariates. Interpretation and reporting of results are based on both adjusted and sample means, as illustrated in Section 7.6. In any event, means are accompanied by some measure of variability: standard deviations, standard errors, and/or confidence intervals.

7.2.5 Specific Comparisons and Trend Analysis

If an interaction or a main effect for an IV with more than two levels is significant, you probably want to ask which levels of main effect or cells of interaction are different from which others. If, in the example, treatment with three levels is significant, the researcher would be likely to want to ask if the pooled average for the two treated groups is different from the average for the waiting-list control, and if the average for relaxation training is different from the average for desensitization. Indeed, the researcher may have planned to ask these questions instead of the omnibus F questions about treatment. Similarly, if the interaction of gender of subject and treatment is significant, you may want to ask if there is a significant difference in the average response of women and men to, for instance, desensitization.

Specific comparisons and trend analysis are discussed more fully in Sections 7.5.4, 3.2.6, 6.5.4.3, and 8.5.2.

7.2.6 Effect Size

If a main effect or interaction reliably affects behavior, the next logical question is: How much? What proportion of variance of the linear combination of DV scores is attributable to the effect? You can determine, for instance, the proportion of the variance in the linear combination of anxiety scores that is associated with differences in treatment. These procedures are described in Section 7.4.1. Procedures are also available for finding the effect sizes for individually significant DVs as demonstrated in Section 7.6, along with confidence intervals for effect sizes.

7.2.7 Effects of Covariates

When covariates are used, the researcher normally wants to assess their utility. Do the covariates provide statistically significant adjustment and what is the nature of the DV-covariate relationship? For

example, when pretests of test, stress, and free-floating anxiety are used as covariates, to what degree does each covariate adjust the composite DV? Assessment of covariates is demonstrated in Section 7.6.3.1.

7.2.8 Repeated-Measures Analysis of Variance

MANOVA is an alternative to repeated-measures ANOVA in which responses to the levels of the within-subjects IV are simply viewed as separate DVs. Suppose, in the example, that measures of test anxiety are taken three times (instead of measuring three different kinds of anxiety once), before, immediately after, and 6 months after treatment. Results could be analyzed as a two-way ANOVA, with treatment as a between-subjects IV and tests as a within-subject IV, or as a one-way MANOVA, with treatment as a between-subjects IV and the three testing occasions as three DVs.

As discussed in Sections 3.2.3 and 8.5.1, repeated measures ANOVA has the often-violated assumption of sphericity. When the assumption is violated, significance tests are too liberal and some alternative to ANOVA is necessary. Other alternatives are adjusted tests of the significance of the within-subjects IV (e.g., Huynh-Feldt), decomposition of the repeated-measures IV into an orthogonal series of single degree of freedom tests (e.g., trend analysis), and profile analysis of repeated measures (Chapter 8).

7.3 Limitations to Multivariate Analysis of Variance and Covariance

7.3.1 Theoretical Issues

As with all other procedures, attribution of causality to IVs is in no way assured by the statistical test. This caution is especially relevant because MANOVA, as an extension of ANOVA, stems from experimental research where IVs are typically manipulated by the experimenter and desire for causal inference provides the reason behind elaborate controls. But the statistical test is available whether or not IVs are manipulated, subjects randomly assigned, and controls implemented. Therefore, the inference that significant changes in the DVs are caused by concomitant changes in the IVs is a logical exercise, not a statistical one.

Choice of variables is also a question of logic and research design rather than of statistics. Skill is required in choosing IVs and levels of IVs, as well as DVs that have some chance of showing effects of the IVs. A further consideration in choice of DVs is the extent of likely correlation among them. The best choice is a set of DVs that are uncorrelated with each other because they each measure a separate aspect of the influence of the IVs. When DVs are correlated, they measure the same or similar facets of behavior in slightly different ways. What is gained by inclusion of several measures of the same thing? Might there be some way of combining DVs or deleting some of them so that the analysis is simpler?

In addition to choice of number and type of DVs is choice of the order in which DVs enter a stepdown analysis if Roy-Bargmann stepdown F is the method chosen to assess the importance of DVs (see Section 7.5.3.2). Priority is usually given to more important DVs or to DVs that are considered causally prior to others in theory. The choice is not trivial because the significance of a DV

may well depend on how high a priority it is given, just as in sequential multiple regression the significance of an IV is likely to depend on its position in the sequence.

When MANCOVA is used, the same limitations apply as in ANCOVA. Consult Sections 6.3.1 and 6.5 for a review of some of the hazards associated with interpretation of designs that include covariates.

Finally, the usual limits to generalizability apply. The results of MANOVA and MANCOVA generalize only to those populations from which the researcher has randomly sampled. And although MANCOVA may, in some very limited situations, adjust for failure to randomly assign subjects to groups, MANCOVA does not adjust for failure to sample from segments of the population to which one wishes to generalize.

7.3.2 Practical Issues

In addition to the theoretical and logical issues discussed above, the statistical procedure demands consideration of some practical matters.

7.3.2.1 Unequal Sample Sizes, Missing Data, and Power

Problems associated with unequal cell sizes are discussed in Section 6.5.4.2. Problems caused by incomplete data (and solutions to them) are discussed in Chapters 4 and 6 (particularly Section 6.3.2.1). The discussion applies to MANOVA and, in fact, may be even more relevant because, as experiments are complicated by numerous DVs and, perhaps, covariates, the probability of missing data increases.

In addition, when using MANOVA, it is necessary to have more cases than DVs in every cell. With numerous DVs this requirement can become burdensome, especially when the design is complicated and there are numerous cells. There are two reasons for the requirement. The first is associated with the assumption of homogeneity of variance-covariance matrices (see Section 7.3.2.4). If a cell has more DVs than cases, the cell becomes singular and the assumption is untestable. If the cell has only one or two more cases than DVs, the assumption is likely to be rejected. Thus MANOVA as an analytic strategy may be discarded because of a failed assumption when the assumption failed because the cases-to-DVs ratio is too low.

Second, the power of the analysis is lowered unless there are more cases than DVs in every cell because of reduced degrees of freedom for error. One likely outcome of reduced power is a nonsignificant multivariate F, but one or more significant univariate Fs (and a very unhappy researcher). Sample sizes in each cell must be sufficient in any event to ensure adequate power. There are many software programs available to calculate required sample sizes depending on desired power and anticipated means and standard deviations in an ANOVA. An Internet search for "statistical power" reveals a number of them, some of which are free. One quick-and-dirty way to apply these is to pick the DV with the smallest expected difference that you want to show statistical significance—your minimum significant DV. One program specifically designed to assess power in MANOVA is GANOVA (Woodward, Bonett, & Brecht, 1990). Another is NCSS PASS (2002), which now includes power analysis for between-subjects MANOVA. Required sample size also may be estimated through SPSS MANOVA by a process of successive approximation. For post hoc estimates of power at a given sample size, you compute a constant weighting variable, weight cases by that variable, and rerun the analysis until desired power is achieved (David P. Nichols, SPSS, personal com-

munication, April 19, 2005). Matrix input is useful for a priori estimates of sample size using SPSS MANOVA (D'Amico, Neilands, & Zambarano, 2001).

Power in MANOVA also depends on the relationships among DVs. Power for the multivariate test is highest when the pooled within-cell correlation among two DVs is high and negative. The multivariate test has much less power when the correlation is positive, zero, or moderately negative. An interesting thing happens, however, when one of two DVs is affected by the treatment and the other is not. The higher the absolute value of the correlation between the two DVs, the greater the power of the multivariate test (Woodward et al., 1990).

7.3.2.2 Multivariate Normality

Significance tests for MANOVA, MANCOVA, and other multivariate techniques are based on the multivariate normal distribution. Multivariate normality implies that the sampling distributions of means of the various DVs in each cell and all linear combinations of them are normally distributed. With univariate F and large samples, the central limit theorem suggests that the sampling distribution of means approaches normality even when raw scores do not. Univariate F is robust to modest violations of normality as long as there are at least 20 degrees of freedom for error in a univariate ANOVA and the violations are not due to outliers (Section 4.1.5). Even with unequal n and only a few DVs, a sample size of about 20 in the smallest cell should ensure robustness (Mardia, 1971). In Monte Carlo studies, Seo, Kanda, and Fujikoshi (1995) have shown robustness to nonnormality in MANOVA with overall $N = 40$ ($n = 10$ per group).

With small, unequal samples, normality of DVs is assessed by reliance on judgment. Are the individual DVs expected to be fairly normally distributed in the population? If not, is some transformation likely to produce normality? With a nonnormally distributed covariate consider transformation or deletion. Covariates are often included as a convenience in reducing error, but it is hardly a convenience if it reduces power.

7.3.2.3 Absence of Outliers

One of the more serious limitations of MANOVA (and ANOVA) is its sensitivity to outliers. Especially worrisome is that an outlier can produce either a Type I or a Type II error, with no clue in the analysis as to which is occurring. Therefore, it is highly recommended that a test for outliers accompany any use of MANOVA.

Several programs are available for screening for univariate and multivariate outliers (cf. Chapter 4). *Run tests for univariate and multivariate outliers for each cell of the design separately and change, transform, or eliminate them.* Report the change, transformation, or deletion of outlying cases. Screening runs for within-cell univariate and multivariate outliers are shown in Sections 6.6.1.4 and 7.6.1.4.

7.3.2.4 Homogeneity of Variance-Covariance Matrices

The multivariate generalization of homogeneity of variance for individual DVs is homogeneity of variance-covariance matrices as discussed in Section 4.1.5.3.[2] The assumption is that variance-covariance matrices within each cell of the design are sampled from the same population variance-

[2]In MANOVA, homogeneity of variance for each of the DVs is also assumed. See Section 8.3.2.4 for discussion and recommendations.

covariance matrix and can reasonably be pooled to create a single estimate of error.[3] If the within-cell error matrices are heterogeneous, the pooled matrix is misleading as an estimate of error variance.

The following guidelines for testing this assumption in MANOVA are based on a generalization of a Monte Carlo test of robustness for T^2 (Hakstian, Roed, & Lind, 1979). If sample sizes are equal, robustness of significance tests is expected; disregard the outcome of Box's M test, a notoriously sensitive test of homogeneity of variance-covariance matrices available through SPSS MANOVA.

However, if sample sizes are unequal and Box's M test is significant at $p < .001$, then robustness is not guaranteed. The more numerous the DVs and the greater the discrepancy in cell sample sizes, the greater the potential distortion of alpha levels. Look at both sample sizes and the sizes of the variances and covariances for the cells. If cells with larger samples produce larger variances and covariances, the alpha level is conservative so that null hypotheses can be rejected with confidence. If, however, cells with smaller samples produce larger variances and covariances, the significance test is too liberal. Null hypotheses are retained with confidence but indications of mean differences are suspect. Use Pillai's criterion instead of Wilks' lambda (see Section 7.5.2) to evaluate multivariate significance (Olson, 1979); or equalize sample sizes by random deletion of cases, if power can be maintained at reasonable levels.

7.3.2.5 Linearity

MANOVA and MANCOVA assume linear relationships among all pairs of DVs, all pairs of covariates, and all DV–covariate pairs in each cell. Deviations from linearity reduce the power of the statistical tests because (1) the linear combinations of DVs do not maximize the separation of groups for the IVs, and (2) covariates do not maximize adjustment for error. Section 4.1.5.2 provides guidelines for checking for and dealing with nonlinearity. If serious curvilinearity is found with a covariate, consider deletion; if curvilinearity is found with a DV, consider transformation—provided, of course, that increased difficulty in interpretation of a transformed DV is worth the increase in power.

7.3.2.6 Homogeneity of Regression

In Roy-Bargmann stepdown analysis (Section 7.5.3.2) and in MANCOVA (Section 7.4.3) it is assumed that the regression between covariates and DVs in one group is the same as the regression in other groups so that using the average regression to adjust for covariates in all groups is reasonable.

In both MANOVA and MANCOVA, if Roy-Bargmann stepdown analysis is used, the importance of a DV in a hierarchy of DVs is assessed in ANCOVA with higher-priority DVs serving as covariates. Homogeneity of regression is required for each step of the analysis, as each DV, in turn, joins the list of covariates. If heterogeneity of regression is found at a step, the rest of the stepdown analysis is uninterpretable. Once violation occurs, the IV-"covariate" interaction is itself interpreted and the DV causing violation is eliminated from further steps.

In MANCOVA (like ANCOVA) heterogeneity of regression implies that there is interaction between the IV(s) and the covariates and that a different adjustment of DVs for covariates is needed in different groups. If interaction between IVs and covariates is suspected, MANCOVA is an inappropriate analytic strategy, both statistically and logically. Consult Sections 6.3.2.7 and 6.5.5 for alternatives to MANCOVA where heterogeneity of regression is found.

[3]Don't confuse this assumption with the assumption of sphericity that is relevant to repeated-measures ANOVA or MANOVA, as discussed in Section 6.5.4.1 and 8.5.1.

For MANOVA, test for stepdown homogeneity of regression, and for MANCOVA, test for overall and stepdown homogeneity of regression. These procedures are demonstrated in Section 7.6.1.6.

7.3.2.7 *Reliability of Covariates*

In MANCOVA as in ANCOVA, the *F* test for mean differences is more powerful if covariates are reliable. If covariates are not reliable, either increased Type I or Type II errors can occur. Reliability of covariates is discussed more fully in Section 6.3.2.8.

In Roy-Bargmann stepdown analysis where all but the lowest-priority DV act as covariates in assessing other DVs, unreliability of any of the DVs (say, $r_{yy} < .8$) raises questions about stepdown analysis as well as about the rest of the research effort. When DVs are unreliable, use another method for assessing the importance of DVs (Section 7.5.3) and report known or suspected unreliability of covariates and high-priority DVs in your Results section.

7.3.2.8 *Absence of Multicollinearity and Singularity*

When correlations among DVs are high, one DV is a near-linear combination of other DVs; the DV provides information that is redundant to the information available in one or more of the other DVs. It is both statistically and logically suspect to include all the DVs in analysis and *the usual solution is deletion of the redundant DV.* However, if there is some compelling theoretical reason to retain all DVs, a principal components analysis (cf. Chapter 13) is done on the pooled within-cell correlation matrix, and component scores are entered as an alternative set of DVs.

SAS and SPSS GLM protect against multicollinearity and singularity through computation of pooled within-cell tolerance (1 − SMC) for each DV; DVs with insufficient tolerance are deleted from analysis. In SPSS MANOVA, singularity or multicollinearity may be present when the determinant of the within-cell correlation matrix is near zero (say, less than .0001). Section 4.1.7 discusses multicollinearity and singularity and has suggestions for identifying the redundant variable(s).

7.4 Fundamental Equations for Multivariate Analysis of Variance and Covariance

7.4.1 Multivariate Analysis of Variance

A minimum data set for MANOVA has one or more IVs, each with two or more levels, and two or more DVs for each subject within each combination of IVs. A fictitious small sample with two DVs and two IVs is illustrated in Table 7.1. The first IV is degree of disability with three levels—mild, moderate, and severe—and the second is treatment with two levels—treatment and no treatment. These two IVs in factorial arrangement produce six cells; three children are assigned to each cell so there are 3 × 6 or 18 children in the study. Each child produces two DVs: score on the reading subtest of the Wide Range Achievement Test (WRAT-R) and score on the arithmetic subtest (WRAT-A). In addition an IQ score is given in parentheses for each child to be used as a covariate in Section 7.4.3.

The test of the main effect of treatment asks: Disregarding degree of disability, does treatment affect the composite score created from the two subtests of the WRAT? The test of interaction asks: Does the effect of treatment on a difference composite score from the two subtests differ as a function of degree of disability?

TABLE 7.1 Small-Sample Data for Illustration of Multivariate Analysis of Variance

	Mild			Moderate			Severe		
	WRAT-R	*WRAT-A*	*(IQ)*	*WRAT-R*	*WRAT-A*	*(IQ)*	*WRAT-R*	*WRAT-A*	*(IQ)*
Treatment	115	108	(110)	100	105	(115)	89	78	(99)
	98	105	(102)	105	95	(98)	100	85	(102)
	107	98	(100)	95	98	(100)	90	95	(100)
Control	90	92	(108)	70	80	(100)	65	62	(101)
	85	95	(115)	85	68	(99)	80	70	(95)
	80	81	(95)	78	82	(105)	72	73	(102)

The test of the main effect of disability is automatically provided in the analysis but is trivial in this example. The question is: Are scores on the WRAT affected by degree of disability? Because degree of disability is at least partially defined by difficulty in reading and/or arithmetic, a significant effect provides no useful information. On the other hand, the absence of this effect would lead us to question the adequacy of classification.

The sample size of three children per cell is highly inadequate for a realistic test but serves to illustrate the techniques of MANOVA. Additionally, if causal inference is intended, the researcher should randomly assign children to the levels of treatment. The reader is encouraged to analyze these data by hand and by computer. Syntax and selected output for this example appear in Section 7.4.2 for several appropriate programs.

MANOVA follows the model of ANOVA where variance in scores is partitioned into variance attributable to difference among scores within groups and to differences among groups. Squared differences between scores and various means are summed (see Chapter 3); these sums of squares, when divided by appropriate degrees of freedom, provide estimates of variance attributable to different sources (main effects of IVs, interactions among IVs, and error). Ratios of variances provide tests of hypotheses about the effects of IVs on the DV.

In MANOVA, however, each subject has a score on each of several DVs. When several DVs for each subject are measured, there is a matrix of scores (subjects by DVs) rather than a simple set of DVs within each group. Matrices of difference scores are formed by subtracting from each score an appropriate mean; then the matrix of differences is squared. When the squared differences are summed, a sum-of-squares-and-cross-products matrix, an **S** matrix, is formed, analogous to a sum of squares in ANOVA (Section 16.4). Determinants[4] of the various **S** matrices are found, and ratios between them provide tests of hypotheses about the effects of the IVs on the linear combination of DVs. In MANCOVA, the sums of squares and cross products in the **S** matrix are adjusted for covariates, just as sums of squares are adjusted in ANCOVA (Chapter 6).

The MANOVA equation for equal n is developed below through extension of ANOVA. The simplest partition apportions variance to systematic sources (variance attributable to differences

[4]A determinant, as described in Appendix A, can be viewed as a measure of generalized variance for a matrix.

between groups) and to unknown sources of error (variance attributable to differences in scores within groups). To do this, differences between scores and various means are squared and summed.

$$\sum_i \sum_j (Y_{ij} - \text{GM})^2 = n \sum_j (\bar{Y}_j - \text{GM})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_j)^2 \tag{7.1}$$

The total sum of squared differences between scores on Y (the DV) and the grand mean (GM) is partitioned into sum of squared differences between group means ($\bar{Y}_j$) and the grand mean (i.e., systematic or between-groups variability), and sum of squared difference between individual scores (Y_{ij}) and their respective group means.

or

$$\text{SS}_{total} = \text{SS}_{bg} - \text{SS}_{wg}$$

For designs with more than one IV, SS_{bg} is further partitioned into variance associated with the first IV (e.g., degree of disability, abbreviated D), variance associated with the second IV (treatment, or T), and variance associated with the interaction between degree of disability and treatment (or DT).

$$\begin{aligned} n_{km} \sum_k \sum_m (DT_{km} - \text{GM}_{km})^2 &= n_k \sum_k (D_k - \text{GM})^2 + n_m \sum_m (T_m - \text{GM})^2 \\ &+ \left[n_{km} \sum_k \sum_m (DT_{km} - \text{GM})^2 - n_k \sum_k (D_k - \text{GM})^2 - n_m \sum_m (T_m - \text{GM})^2 \right] \end{aligned} \tag{7.2}$$

The sum of squared differences between cell (DT_{km}) means and the grand mean is partitioned into (1) sum of squared differences between means associated with different levels of disability (D_k) and the grand mean; (2) sum of squared differences between means associated with different levels of treatment (T_m) and the grand mean; and (3) sum of squared differences associated with combinations of treatment and disability (DT_{km}) and the grand mean, from which differences associated with D_k and T_m are subtracted. Each n is the number of scores composing the relevant marginal or cell mean.

or

$$\text{SS}_{bg} = \text{SS}_D + \text{SS}_T + \text{SS}_{DT}$$

The full partition for this factorial between-subjects design is

$$\begin{aligned} \sum_i \sum_k \sum_m (Y_{ikm} - \text{GM})^2 &= n_k \sum_k (D_k - \text{GM})^2 + n_m \sum_m (T_m - \text{GM})^2 \\ &+ \left[n_{km} \sum_k \sum_m (DT_{km} - \text{GM})^2 - n_k \sum_k (D_k - \text{GM})^2 - n_m \sum_m (T_m - \text{GM})^2 \right] \\ &+ \sum_i \sum_k \sum_m (Y_{ikm} - DT_{km})^2 \end{aligned} \tag{7.3}$$

For MANOVA, there is no single DV but rather a column matrix (or vector) of Y_{ikm} values of scores on each DV. For the example in Table 7.1, column matrices of Y scores for the three children in the first cell of the design (mild disability with treatment) are

$$Y_{i11} = \begin{bmatrix} 115 \\ 108 \end{bmatrix} \begin{bmatrix} 98 \\ 105 \end{bmatrix} \begin{bmatrix} 107 \\ 98 \end{bmatrix}$$

Similarly, there is a column matrix of disability—D_k—means for mild, moderate, and severe levels of D, with one mean in each matrix for each DV.

$$\mathbf{D_1} = \begin{bmatrix} 95.83 \\ 96.50 \end{bmatrix} \quad \mathbf{D_2} = \begin{bmatrix} 88.83 \\ 88.00 \end{bmatrix} \quad \mathbf{D_3} = \begin{bmatrix} 82.67 \\ 77.17 \end{bmatrix}$$

where 95.83 is the mean on WRAT-R and 96.50 is the mean on WRAT-A for children with mild disability, averaged over treatment and control groups.

Matrices for treatment—T_m—means, averaged over children with all levels of disability are

$$\mathbf{T_1} = \begin{bmatrix} 99.89 \\ 96.33 \end{bmatrix} \quad \mathbf{T_2} = \begin{bmatrix} 78.33 \\ 78.11 \end{bmatrix}$$

Similarly, there are six matrices of cell means (DT_{km}) averaged over the three children in each group.

Finally, there is a single matrix of grand means (**GM**), one for each DV, averaged over all children in the experiment.

$$\mathbf{GM} = \begin{bmatrix} 89.11 \\ 87.22 \end{bmatrix}$$

As illustrated in Appendix A, differences are found by simply subtracting one matrix from another, to produce difference matrices. The matrix counterpart of a difference score, then, is a difference matrix. To produce the error term for this example, the matrix of grand means (**GM**) is subtracted from each of the matrixes of individual scores ($\mathbf{Y}_{ikm}$). Thus for the first child in the example:

$$(\mathbf{Y_{111}} - \mathbf{GM}) = \begin{bmatrix} 115 \\ 108 \end{bmatrix} - \begin{bmatrix} 89.11 \\ 87.22 \end{bmatrix} = \begin{bmatrix} 25.89 \\ 20.75 \end{bmatrix}$$

In ANOVA, difference scores are squared. The matrix counterpart of squaring is multiplication by a transpose. That is, each column matrix is multiplied by its corresponding row matrix (see Appendix A for matrix transposition and multiplication) to produce a sum-of-squares and cross-products matrix. For example, for the first child in the first group of the design:

$$(\mathbf{Y_{111}} - \mathbf{GM})(\mathbf{Y_{111}} - \mathbf{GM})' = \begin{bmatrix} 25.89 \\ 20.78 \end{bmatrix} [25.89 \quad 20.78] = \begin{bmatrix} 670.29 & 537.99 \\ 537.99 & 431.81 \end{bmatrix}$$

These matrices are then summed over subjects and over groups, just as squared differences are summed in univariate ANOVA.[5] The order of summing and squaring is the same in MANOVA as in ANOVA for a comparable design. The resulting matrix (**S**) is called by various names: sum-of-squares and cross-products, cross-products, or sum-of-products. The MANOVA partition of sums-of-squares and cross-products for our factorial example is represented below in a matrix form of Equation 7.3:

$$\sum_i \sum_k \sum_m (\mathbf{Y}_{ikm} - \mathbf{GM})(\mathbf{Y}_{ikm} - \mathbf{GM})'$$

$$= n_k \sum_k (\mathbf{D}_k - \mathbf{GM})(\mathbf{D}_k - \mathbf{GM})' + n_m \sum_m (\mathbf{T}_m - \mathbf{GM})(\mathbf{T}_m - \mathbf{GM})'$$

$$+ \left[n_{km} \sum_k \sum_m (\mathbf{DT}_{km} - \mathbf{GM})(\mathbf{DT}_{km} - \mathbf{GM})' - n_k \sum_k (\mathbf{D}_k - \mathbf{GM})(\mathbf{D}_k - \mathbf{GM})' \right.$$

$$\left. - n_m \sum_m (\mathbf{T}_m - \mathbf{GM})(\mathbf{T}_m - \mathbf{GM})' \right] + \sum_i \sum_k \sum_m (\mathbf{Y}_{ikm} - \mathbf{DT}_{km})(\mathbf{Y}_{ikm} - \mathbf{DT}_{km})'$$

or

$$\mathbf{S}_{\text{total}} = \mathbf{S}_D + \mathbf{S}_T + \mathbf{S}_{DT} + \mathbf{S}_{S(DT)}$$

The total cross-products matrix ($\mathbf{S}_{\text{total}}$) is partitioned into cross-products matrices for differences associated with degree of disability, with treatment, with the interaction between disability and treatment, and for error–subjects within groups ($\mathbf{S}_{S(DT)}$).

For the example in Table 7.1, the four resulting cross-products matrices[6] are

$$\mathbf{S}_D = \begin{bmatrix} 570.29 & 761.72 \\ 761.72 & 1126.78 \end{bmatrix} \qquad \mathbf{S}_T = \begin{bmatrix} 2090.89 & 1767.56 \\ 1767.56 & 1494.22 \end{bmatrix}$$

$$\mathbf{S}_{DT} = \begin{bmatrix} 2.11 & 5.28 \\ 5.28 & 52.78 \end{bmatrix} \qquad \mathbf{S}_{S(DT)} = \begin{bmatrix} 544.00 & 31.00 \\ 31.00 & 539.33 \end{bmatrix}$$

Notice that all these matrices are symmetrical, with the elements top left to bottom right diagonal representing sums of squares (that, when divided by degrees of freedom, produce variances), and with the off-diagonal elements representing sums of cross products (that, when divided by degrees of freedom, produce covariances). In this example, the first element in the major diagonal (top left to bottom right) is the sum of squares for the first DV, WRAT-R, and the second element is the sum of

[5]We highly recommend using a matrix algebra program, such as a spreadsheet or SPSS MATRIX, MATLAB, or SAS IML, to follow the more complex matrix equations to come.

[6]Numbers producing these matrices were carried to 8 digits before rounding.

squares for the second DV, WRAT-A. The off-diagonal elements are the sums of cross-products between WRAT-R and WRAT-A.

In ANOVA, sums of squares are divided by degrees of freedom to produce variances, or mean squares. In MANOVA, the matrix analog of variance is a determinant (see Appendix A); the determinant is found for each cross-products matrix. In ANOVA, ratios of variances are formed to test main effects and interactions. In MANOVA, ratios of determinants are formed to test main effects and interactions when using Wilks' lambda (see Section 7.5.2 for additional criteria). These ratios follow the general form

$$\Lambda = \frac{|\mathbf{S}_{\text{error}}|}{|\mathbf{S}_{\text{effect}} + \mathbf{S}_{\text{error}}|} \tag{7.4}$$

Wilks' lambda (Λ) is the ratio of the determinant of the error cross-products matrix to the determinant of the sum of the error and effect cross-products matrices.

To find Wilks' lambda, the within-groups matrix is added to matrices corresponding to main effects and interactions before determinants are found. For the example, the matrix produced by adding the $\mathbf{S}_{DT}$ matrix for interaction to the $\mathbf{S}_{S(DT)}$ matrix for subjects within groups (error) is

$$\mathbf{S}_{DT} + \mathbf{S}_{S(DT)} = \begin{bmatrix} 2.11 & 5.28 \\ 5.28 & 52.78 \end{bmatrix} + \begin{bmatrix} 544.00 & 31.00 \\ 31.00 & 539.33 \end{bmatrix} = \begin{bmatrix} 546.11 & 36.28 \\ 36.28 & 592.11 \end{bmatrix}$$

For the four matrices needed to test main effect of disability, main effect of treatment, and the treatment-disability interaction, the determinants are

$$|\mathbf{S}_{S(DT)}| = 292436.52$$
$$|\mathbf{S}_{D} + \mathbf{S}_{S(DT)}| = 1228124.71$$
$$|\mathbf{S}_{T} + \mathbf{S}_{S(DT)}| = 2123362.49$$
$$|\mathbf{S}_{DT} + \mathbf{S}_{S(DT)}| = 322040.95$$

At this point a source table, similar to the source table for ANOVA, is useful, as presented in Table 7.2. The first column lists sources of variance; in this case the two main effects and the interaction. The error term does not appear. The second column contains the value of Wilks' lambda.

Wilks' lambda is a ratio of determinants, as described in Equation 7.4. For example, for the interaction between disability and treatment, Wilks' lambda is

$$\Lambda = \frac{|\mathbf{S}_{S(DT)}|}{|\mathbf{S}_{DT} + \mathbf{S}_{S(DT)}|} = \frac{292436.52}{322040.95} = .908068$$

Tables for evaluating Wilks' lambda directly are rare, however, an approximation to F has been derived that closely fits Λ. The last three columns of Table 7.2, then, represent the approximate F values and their associated degrees of freedom.

TABLE 7.2 Multivariate Analysis of Variance of WRAT-R and WRAT-A Scores

Source of Variance	Wilks' Lambda	df_1	df_2	Multivariate F
Treatment	.13772	2.00	11.00	34.43570**
Disability	.25526	4.00	22.00	5.38602*
Treatment by disability	.90807	4.00	22.00	0.27170

*$p < .01$.
**$p < .001$.

The following procedure for calculating approximate F (Rao, 1952) is based on Wilks' lambda and the various degrees of freedom associated with it.

$$\text{Approximate } F(df_1, df_2) = \left(\frac{1-y}{y}\right)\left(\frac{df_2}{df_1}\right) \tag{7.5}$$

where df_1 and df_2 are defined below as the degrees of freedom for testing the F ratio, and y is

$$y = \Lambda^{1/s} \tag{7.6}$$

Λ is defined in Equation 7.4, and s is[7]

$$s = \min(p, df_{effect}) \tag{7.7}$$

where p is the number of DVs, and df_{effect} is the degrees of freedom for the effect being tested. And

$$df_1 = p(df_{effect})$$

and

$$df_2 = s\left[(df_{error}) - \frac{p - df_{effect} + 1}{2}\right] - \left[\frac{p(df_{effect}) - 2}{2}\right]$$

where df_{error} is the degrees of freedom associated with the error term.

For the test of interaction in the sample problem, we have

$p = 2$ the number of DVs

$df_{effect} = 2$ the number of treatment levels minus 1 times the number of disability levels minus 1 or $(t - 1)(d - 1)$

$df_{error} = 12$ the number of treatment levels times the number of disability levels times the quantity $n - 1$ (where n is the number of scores per cell for each DV)—that is, $df_{error} = dt(n - 1)$

[7]When $p = 1$, we have univariate ANOVA.

Thus

$$s = \min(p, df_{effect}) = 2$$

$$y = .908068^{1/2} = .952926$$

$$df_1 = 2(2) = 4$$

$$df_2 = 2\left[12 - \frac{2-2+1}{2}\right] - \left[\frac{2(2)-2}{2}\right] = 22$$

$$\text{Approximate } F(4, 22) = \left(\frac{.047074}{.952926}\right)\left(\frac{22}{4}\right) = 0.2717$$

This approximate F value is tested for significance by using the usual tables of F at selected α. In this example, the interaction between disability and treatment is not statistically significant with 4 and 22 df, because the observed value of 0.2717 does not exceed the critical value of 2.82 at $\alpha = .05$.

Following the same procedures, the effect of treatment is statistically significant, with the observed value of 34.44 exceeding the critical value of 3.98 with 2 and 11 df, $\alpha = .05$. The effect of degree of disability is also statistically significant, with the observed value of 5.39 exceeding the critical value of 2.82 with 4 and 22 df, $\alpha = .05$. (As noted previously, this main effect is not of research interest, but does serve to validate the classification procedure.) In Table 7.2, significance is indicated at the highest level of α reached, following standard practice.

A measure of effect size is readily available from Wilks' lambda.[8] For MANOVA:

$$\eta^2 = 1 - \Lambda \tag{7.8}$$

This equation represents the variance accounted for by the best linear combination of DVs as explained below.

In a one-way analysis, according to Equation 7.4, Wilks' lambda is the ratio of (the determinant of) the error matrix and (the determinant of) the total sum-of-squares and cross-products matrix. The determinant of the error matrix—Λ—is the variance not accounted for by the combined DVs so $1 - \Lambda$ is the variance that is accounted for.

Thus, for each statistically significant effect, the proportion of variance accounted for is easily calculated using Equation 7.8. For example, the main effect of treatment:

$$\eta_T^2 = 1 - \Lambda_T = 1 - .137721 = .862279$$

In the example, 86% of the variance in the best linear combination of WRAT-R and WRAT-A scores is accounted for by assignment to levels of treatment. The square root of η^2 ($\eta = .93$) is a form of correlation between WRAT scores and assignment to treatment.

However, unlike η^2 in the analogous ANOVA design, the sum of η^2 for all effects in MANOVA may be greater than 1.0 because DVs are recombined for each effect. This lessens the appeal of an interpretation in terms of proportion of variance accounted for, although the size of η^2 is still a measure of the relative importance of an effect.

[8]An alternative measure of effect size is canonical correlation, printed out by some computer programs. Canonical correlation is the correlation between the optimal linear combination of IV levels and the optimal linear combination of DVs where optimal is chosen to maximize the correlation between combined IVs and DVs. Canonical correlation as a general procedure is discussed in Chapter 12, and the relation between canonical correlation and MANOVA is discussed briefly in Chapter 17.

Another difficulty in using this form of η^2 is that effects tend to be much larger in the multivariate than in the univariate case. Therefore, a recommended alternative, when $s > 1$ is

$$\text{partial } \eta^2 = 1 - \Lambda^{1/s} \tag{7.9}$$

Estimated effect size is reduced to .63 with the use of partial η^2 for the current data, a more reasonable assessment. Confidence limits around effect sizes are in Section 7.6.

7.4.2 Computer Analyses of Small-Sample Example

Tables 7.3 through 7.5 show syntax and selected minimal output for SPSS MANOVA, SPSS GLM, and SAS GLM, respectively.

In SPSS MANOVA (Table 7.3) simple MANOVA source tables, resembling those of ANOVA, are printed out when PRINT=SIGNIF(BRIEF) is requested. After interpretive material is printed (not shown), the source table is shown, labeled `Tests using UNIQUE sums of squares and WITHIN+RESIDUAL`. `WITHIN+RESIDUAL` refers to the pooled within-cell error SSCP matrix (Section 7.4.1) plus any effects not tested, the error term chosen by default for MANOVA.

For the example, the two-way MANOVA source table consists of the two main effects and the interaction. For each source, you are given Wilks' lambda, `Approximate` (multivariate) `F` with numerator and denominator degrees of freedom (`Hyp. DF` and `Error DF`, respectively), and the probability level achieved for the significance test.

Syntax for SPSS GLM is similar to that of MANOVA, except that levels of IVs are not shown in parentheses. METHOD, INTERCEPT, and CRITERIA instructions are produced by the menu system by default.

Output consists of a source table that includes four tests of the multivariate effects, Pillai's, Wilks', Hotelling's, and Roy's (see Section 7.5.2 for a discussion of these tests). All are identical when there are only two levels of a between-subjects IV. The results of Wilks' Lambda test match those of SPSS MANOVA in Table 7.3. This is followed by univariate tests on each of the DVs, in the

TABLE 7.3 MANOVA on Small-Sample Example through SPSS MANOVA (Syntax and Output)

```
MANOVA
 WRATR WRATA BY TREATMNT(1,2) DISABLTY(1,3)
 /PRINT=SIGNIF(BRIEF)
 /DESIGN = TREATMNT DISABLTY TREATMNT*DISABLTY.

* * * * * * A n a l y s i s   o f   V a r i a n c e-design 1 * * * * * *

Multivariate Tests of Significance
Tests using UNIQUE sums of squares and WITHIN+RESIDUAL error term
Source of Variation     Wilks    Approx F    Hyp. DF    Error DF    Sig of F

TREATMNT                .138      34.436       2.00      11.000        .000
DISABLTY                .255       5.386       4.00      22.000        .004
TREATMNT * DISABLTY     .908        .272       4.00      22.000        .893
```

TABLE 7.4 MANOVA on Small-Sample Example through SPSS GLM (Syntax and Selected Output)

```
GLM
 wratr wrata BY treatmnt disablty
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /CRITERIA = ALPHA(.05)
 /DESIGN = treatmnt disablty treatmnt*disablty.
```

General Linear Model

Between-Subjects Factors

		Value Label	N
Treatment type	1.00	Treatment	9
	2.00	Control	9
Degree of disability	1.00	Mild	6
	2.00	Moderate	6
	3.00	Severe	6

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.998	2687.779[a]	2.000	11.000	.000
	Wilks' Lambda	.002	2687.779[a]	2.000	11.000	.000
	Hotelling's Trace	488.687	2687.779[a]	2.000	11.000	.000
	Roy's Largest Root	488.687	2687.779[a]	2.000	11.000	.000
Treatmnt	Pillai's Trace	.862	34.436[a]	2.000	11.000	.000
	Wilks' Lambda	.138	34.436[a]	2.000	11.000	.000
	Hotelling's Trace	6.261	34.436[a]	2.000	11.000	.000
	Roy's Largest Root	6.261	34.436[a]	2.000	11.000	.000
Disablty	Pillai's Trace	.750	3.604	4.000	24.000	.019
	Wilks' Lambda	.255	5.386[a]	4.000	22.000	.004
	Hotelling's Trace	2.895	7.238	4.000	20.000	.001
	Roy's Largest Root	2.887	17.323[b]	2.000	12.000	.000
Treatmnt * disablty	Pillai's Trace	.092	.290	4.000	24.000	.882
	Wilks' Lambda	.908	.272[a]	4.000	22.000	.893
	Hotelling's Trace	.101	.252	4.000	20.000	.905
	Roy's Largest Root	.098	.588[b]	2.000	12.000	.571

[a]Exact statistic
[b]The statistic is an upper bound on F that yields a lower bound on the significance level.
[c]Design: Intercept+Treatmnt+Disablty+Treatmnt * Disablty

TABLE 7.4 Continued

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	WRAT - Reading	2613.778[a]	5	522.756	11.531	.000
	WRAT - Arithmetic	2673.778[b]	5	534.756	11.898	.000
Intercept	WRAT - Reading	142934.222	1	142934.222	3152.961	.000
	WRAT - Arithmetic	136938.889	1	136938.999	3046.848	.000
Treatmnt	WRAT - Reading	2090.889	1	2090.889	46.123	.000
	WRAT - Arithmetic	1494.222	1	1494.222	33.246	.000
Disablty	WRAT - Reading	520.778	2	260.389	5.744	.018
	WRAT - Arithmetic	1126.778	2	563.389	12.535	.001
Treatmnt * Disablty	WRAT - Reading	2.111	2	1.056	.023	.977
	WRAT - Arithmetic	52.778	2	26.389	.587	.571
Error	WRAT - Reading	544.000	12	45.333		
	WRAT - Arithmetic	539.333	12	44.944		
Total	WRAT - Reading	146092.000	18			
	WRAT - Arithmetic	140152.000	18			
Corrected Total	WRAT - Reading	3157.778	17			
	WRAT - Arithmetic	3213.111	17			

[a]R Squared = .828 (Adjusted R Squared = .756)
[b]R Squared = .832 (Adjusted R Squared = .762)

table labeled Tests of Between-Subjects Effects. The format of the table follows that of univariate ANOVA (see Table 6.5). Note that interpretation of MANOVA through univariate ANOVAs is *not* recommended (cf. Section 7.5.3.1).

In SAS GLM (Table 7.5) IVs are defined in a `class` instruction and the `model` instruction defines the DVs and the effects to be considered. The `nouni` instruction suppresses printing of descriptive statistics and univariate F tests. The `manova h = _all_` instruction requests tests of all main effects and interactions listed in the `model` instruction, and `short` condenses the printout.

The output begins with some interpretative information (not shown), followed by separate sections for `TREATMNT`, `DISABLTY`, and `TREATMNT*DISABLTY`. Each source table is preceded by information about characteristic roots and vectors of the error SSCP matrix (not shown—these are discussed in Chapters 9, 12, and 13), and the three df parameters (Section 7.4.1). Each source table shows results of four multivariate tests, fully labeled (cf. Section 7.5.2).

7.4.3 Multivariate Analysis of Covariance

In MANCOVA, the linear combination of DVs is adjusted for differences in the covariates. The adjusted linear combination of DVs is the combination that would be obtained if all participants had the same scores on the covariates. For this example, pre-experimental IQ scores (listed in parentheses in Table 7.1) are used as covariates.

In MANCOVA the basic partition of variance is the same as in MANOVA. However, all the matrices—$\mathbf{Y}_{ikm}$, $\mathbf{D}_k$, $\mathbf{T}_m$, $\mathbf{DT}_{km}$, and **GM**—have three entries in our example; the first entry is the covariate (IQ score) and the second two entries are the two DV scores (WRAT-R and WRAT-A). For example, for the first child with mild disability and treatment, the column matrix of covariate and DV scores is

$$\mathbf{Y}_{111} = \begin{bmatrix} 110 \\ 115 \\ 108 \end{bmatrix} \begin{matrix} \text{(IQ)} \\ \text{(WRAT-R)} \\ \text{(WRAT-A)} \end{matrix}$$

As in MANOVA, difference matrices are found by subtraction, and then the squares and cross-products matrices are found by multiplying each difference matrix by its transpose to form the **S** matrices.

At this point another departure from MANOVA occurs. The **S** matrices are partitioned into sections corresponding to the covariates, the DVs, and the cross-products of covariates and DVs. For the example, the cross-products matrix for the main effect of treatment is

$$\mathbf{S}_T = \left[\begin{array}{c|cc} [2.00] & [64.67 & 54.67] \\ \hline \begin{bmatrix} 64.67 \\ 54.67 \end{bmatrix} & \begin{bmatrix} 2090.89 & 1767.56 \\ 1767.56 & 1494.22 \end{bmatrix} \end{array}\right]$$

The lower right-hand partition is the $\mathbf{S}_T$ matrix for the DVs (or $\mathbf{S}_T^{(Y)}$) and is the same as the $\mathbf{S}_T$ matrix developed in Section 7.4.1. The upper left matrix is the sum of squares for the covariate (or $\mathbf{S}_T^{(X)}$). (With additional covariates, this segment becomes a full sum-of-squares and cross-products matrix.) Finally, the two off-diagonal segments contain cross-products of covariates and DVs (or $\mathbf{S}_T^{(XY)}$).

TABLE 7.5 MANOVA on Small-Sample Example through SAS GLM (Syntax and Selected Output)

```
proc glm data=SASUSER.SS_MANOV;
   class TREATMNT DISABLTY;
   model WRATR WRATA=TREATMNT DISABLTY TREATMNT*DISABLTY / nouni;
   manova h=_all_ / short;
run;
```

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of NO Overall TREATMNT Effect
H = Type III SSCP Matrix for TREATMNT
E = Error SSCP Matrix

S=1 M=0 N=4.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' lambda	0.13772139	34.44	2	11	<.0001
Pillai's Trace	0.86227861	34.44	2	11	<.0001
Hotelling-Lawley Trace	6.26103637	34.44	2	11	<.0001
Roy's Greatest Root	6.26103637	34.44	2	11	<.0001

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for DISABLTY
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 WRATR	WRATA
2.88724085	99.73	0.02260839	0.03531017
0.00779322	0.27	-0.03651215	0.02476743

MANOVA Test Criteria and F Approximations for the Hypothesis of NO Overall DISABLTY Effect
H = Type III SSCP Matrix for DISABLTY
E = Error SSCP Matrix

S=2 M=-0.5 N=4.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' lambda	0.25526256	5.39	4	22	0.0035
Pillai's Trace	0.75048108	3.60	4	24	0.0195
Hotelling-Lawley Trace	2.89503407	7.79	4	12.235	0.0023
Roy's Greatest Root	2.88724085	17.32	2	12	0.0003

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' lambda is exact.

(continued)

TABLE 7.5 Continued

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for TREATMNT*DISABLTY
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 WRATR	WRATA
0.09803883	97.11	0.00187535	0.04291087
0.00291470	2.89	0.04290407	-0.00434641

MANOVA Test Criteria and F Approximations for
the Hypothesis of NO Overall TREATMNT*DISABLTY Effect
H = Type III SSCP Matrix for TREATMNT*DISABLTY
E = Error SSCP Matrix

S=2 M=-0.5 N=4.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.90806786	0.27	4	22	0.8930
Pillai's Trace	0.09219163	0.29	4	24	0.8816
Hotelling-Lawley Trace	0.10095353	0.27	4	12.235	0.8908
Roy's Greatest Root	0.09803883	0.59	2	12	0.5706

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

Adjusted or $\mathbf{S}^*$ matrices are formed from these segments. The $\mathbf{S}^*$ matrix is the sums-of-squares and the cross-products of DVs adjusted for effects of covariates. Each sum of squares and each cross-product is adjusted by a value that reflects variance due to differences in the covariate.

In matrix terms, the adjustment is

$$\mathbf{S}^* = \mathbf{S}^{(Y)} - \mathbf{S}^{(YX)}(\mathbf{S}^{(X)})^{-1}\mathbf{S}^{(XY)} \tag{7.10}$$

The adjusted cross-products matrix $\mathbf{S}^*$ is found by subtracting from the unadjusted cross-products matrix of DVs ($\mathbf{S}^{(Y)}$) a product based on the cross-products matrix for covariate(s) ($\mathbf{S}^{(X)}$) and cross-products matrices for the relation between the covariates and the DVs ($\mathbf{S}^{(YX)}$ and $\mathbf{S}^{(XY)}$).

The adjustment is made for the regression of the DVs (Y) on the covariates (X). Because $\mathbf{S}^{(XY)}$ is the transpose of $\mathbf{S}^{(YX)}$, their multiplication is analogous to a squaring operation. Multiplying by the inverse of $\mathbf{S}^{(X)}$ is analogous to division. As shown in Chapter 3 for simple scalar numbers, the regression coefficient is the sum of cross-products between X and Y, divided by the sum of squares for X.

An adjustment is made to each **S** matrix to produce $\mathbf{S}^*$ matrices. The $\mathbf{S}^*$ matrices are 2×2 matrices, but their entries are usually smaller than those in the original MANOVA **S** matrices. For the example, the reduced $\mathbf{S}^*$ matrices are

$$\mathbf{S}_D^* = \begin{bmatrix} 388.18 & 500.49 \\ 500.49 & 654.57 \end{bmatrix} \qquad \mathbf{S}_T^* = \begin{bmatrix} 2059.50 & 1708.24 \\ 1708.24 & 1416.88 \end{bmatrix}$$

$$\mathbf{S}_{DT}^* = \begin{bmatrix} 2.06 & 0.87 \\ 0.87 & 19.61 \end{bmatrix} \qquad \mathbf{S}_{S(DT)}^* = \begin{bmatrix} 528.41 & -26.62 \\ -26.62 & 324.95 \end{bmatrix}$$

Note that, as in the lower right-hand partition, cross-products matrices may have negative values for entries other than the major diagonal which contains sums of squares.

Tests appropriate for MANOVA are applied to the adjusted $\mathbf{S}^*$ matrices. Ratios of determinants are formed to test hypotheses about main effects and interactions by using Wilks' lambda criterion (Equation 7.4). For the example, the determinants of the four matrices needed to test the three hypotheses (two main effects and the interaction) are

$$|\mathbf{S}_{S(DT)}^*| = 171032.69$$

$$|\mathbf{S}_D^* + \mathbf{S}_{S(DT)}^*| = 673383.31$$

$$|\mathbf{S}_T^* + \mathbf{S}_{S(DT)}^*| = 1680076.69$$

$$|\mathbf{S}_{DT}^* + \mathbf{S}_{S(DT)}^*| = 182152.59$$

The source table for MANCOVA, analogous to that produced for MANOVA, for the sample data is in Table 7.6.

One new item in this source table that is not in the MANOVA table of Section 7.4.1 is the variance in the DVs due to the covariate. (With more than one covariate, there is a line for combined covariates and a line for each of the individual covariates.) As in ANCOVA, one degree of freedom

TABLE 7.6 Multivariate Analysis of Covariance of WRAT-R and WRAT-A Scores

Source of Variance	Wilks' Lambda	df_1	df_2	Multivariate *F*
Covariate	.58485	2.00	10.00	3.54913
Treatment	.13772	2.00	10.00	44.11554**
Disability	.25526	4.00	22.00	4.92112*
Treatment by Disability	.90807	4.00	22.00	0.15997

*$p < .01$.

**$p < .001$.

for error is used for each covariate so that df_2 and s of Equation 7.5 are modified. For MANCOVA, then,

$$s = \min(p + q, df_{effect}) \tag{7.11}$$

where q is the number of covariates and all other terms are defined as in Equation 7.7.

$$df_2 = s\left[(df_{error}) - \frac{(p+q) - df_{effect} + 1}{2}\right] - \left[\frac{(p+q)(df_{effect}) - 2}{2}\right]$$

Approximate F is used to test the significance of the covariate–DV relationship as well as main effects and interactions. If a significant relationship is found, Wilks' lambda is used to find the effect size as shown in Equations 7.6 or 7.9.

7.5 Some Important Issues

7.5.1 MANOVA vs. ANOVAs

MANOVA works best with highly negatively correlated DVs and acceptably well with moderately correlated DVs in either direction (about $|.6|$). For example, two DVs, such as time to complete a task and number of errors, might be expected to have a moderate negative correlation and are best analyzed through MANOVA. MANOVA is less attractive if correlations among DVs are very highly positive or near zero (Woodward et al., 1990).

Using very highly positively correlated DVs in MANOVA is wasteful. For example, the effects of the Head Start program might be tested in a MANOVA with the WISC and Stanford-Binet as DVs. The overall multivariate test works acceptably well, but after the highest priority DV is entered in stepdown analysis, tests of remaining DVs are ambiguous. Once that DV becomes a covariate, there is no variance remaining in the lower priority DVs to be related to IV main effects or interactions. Univariate tests also are highly misleading, because they suggest effects on different behaviors when actually there is one behavior being measured repeatedly. Better strategies are to pick a single DV (preferably the most reliable) or to create a composite score (an average if the DVs are commensurate or a principal component score if they are not) for use in ANOVA.

MANOVA also is wasteful if DVs are uncorrelated—naturally, or if they are factor or component scores. The multivariate test has lower power than the univariate and there is little difference between univariate and stepdown results. The only advantage to MANOVA over separate ANOVAs on each DV is control of familywise Type I error. However, this error rate can be controlled by applying a Bonferroni correction (cf. Equation 7.12) to each test in a set of separate ANOVAs on each DV, although that could potentially result in a more conservative analysis than MANOVA.

Sometimes there is a mix of correlated and uncorrelated DVs. For example, there may be a set of moderately correlated DVs related to performance on a task and another set of moderately correlated DVs related to attitudes. Separate MANOVAs on each of the two sets of moderately correlated DVs are likely to produce the most interesting interpretations as long as appropriate adjustments are made for familywise error rate for the multiple MANOVAs. Or one set might serve as covariates in a single MANCOVA.

7.5.2 Criteria for Statistical Inference

Several multivariate statistics are available in MANOVA programs to test significance of main effects and interactions: Wilks' lambda, Hotelling's trace criterion, Pillai's criterion, as well as Roy's gcr criterion. When an effect has only two levels ($s = 1$, 1 df in the univariate sense), the F tests for Wilks' lambda, Hotelling's trace, and Pillai's criterion are identical. And usually when an effect has more than two levels ($s > 1$ and df > 1 in the univariate sense), the F values are slightly different, but either all three statistics are significant or all are nonsignificant. Occasionally, however, some of the statistics are significant while others are not, and the researcher is left wondering which result to believe.

When there is only one degree of freedom for effect, there is only one way to combine the DVs to separate the two groups from each other. However, when there is more than one degree of freedom for effect, there is more than one way to combine DVs to separate groups. For example, with three groups, one way of combining DVs may separate the first group from the other two while the second way of combining DVs separates the second group from the third. Each way of combining DVs is a dimension along which groups differ (as described in gory detail in Chapter 9) and each generates a statistic.

When there is more than one degree of freedom for effect, Wilks' lambda, Hotelling's trace criterion, and Pillai's criterion pool the statistics from each dimension to test the effect; Roy's gcr criterion uses only the first dimension (in our example, the way of combining DVs that separates the first group from the other two) and is the preferred test statistic for a few researchers (Harris, 2001). Most researchers, however, use one of the pooled statistics to test the effect (Olson, 1976).

Wilks' lambda, defined in Equation 7.4 and Section 7.4.1, is a likelihood ratio statistic that tests the likelihood of the data under the assumption of equal population mean vectors for all groups against the likelihood under the assumption that population mean vectors are identical to those of the sample mean vectors for the different groups. Wilks' lambda is the pooled ratio of error variance to effect variance plus error variance. Hotelling's trace is the pooled ratio of effect variance to error variance. Pillai's criterion is simply the pooled effect variances.

Wilks' lambda, Hotelling's trace, and Roy's gcr criterion are often more powerful than Pillai's criterion when there is more than one dimension but the first dimension provides most of the separation of groups; they are less powerful when separation of groups is distributed over dimensions. But Pillai's criterion is said to be more robust than the other three (Olson, 1979). As sample size decreases, unequal n's appear, and the assumption of homogeneity of variance-covariance matrices is violated (Section 7.3.2.2), the advantage of Pillai's criterion in terms of robustness is more important. When the research design is less than ideal, then Pillai's criterion is the criterion of choice.

In terms of availability, all the MANOVA programs reviewed here provide Wilks' lambda, as do most research reports, so that Wilks' lambda is the criterion of choice unless there is reason to use Pillai's criterion. Programs differ in the other statistics provided (see Section 7.7).

In addition to potentially conflicting significance tests for multivariate F is the irritation of a nonsignificant multivariate F but a significant univariate F for one of the DVs. If the researcher measures only one DV—the right one—the effect is significant, but because more DVs are measured, it is not. Why doesn't MANOVA combine DVs with a weight of 1 for the significant DV and a weight of zero for the rest? In fact, MANOVA comes close to doing just that, but multivariate F is often not as powerful as univariate or stepdown F and significance can be lost. If this happens, about the best one can do is report the nonsignificant multivariate F and offer the univariate and/or stepdown result as a guide to future research.

7.5.3 Assessing DVs

When a main effect or interaction is significant in MANOVA, the researcher has usually planned to pursue the finding to discover which DVs are affected. But the problems of assessing DVs in significant multivariate effects are similar to the problems of assigning importance to IVs in multiple regression (Chapter 5). First, there are multiple significance tests so some adjustment is necessary for inflated Type I error. Second, if DVs are uncorrelated, there is no ambiguity in assignment of variance to them, but if DVs are correlated, assignment of overlapping variance to DVs is problematical.

7.5.3.1 Univariate F

If pooled within-group correlations among DVs are zero (and they never are unless they are formed by principal components analysis), univariate ANOVAs, one per DV, give the relevant information about their importance. Using ANOVA for uncorrelated DVs is analogous to assessing importance of IVs in multiple regression by the magnitude of their individual correlations with the DV. The DVs that have significant univariate *F*s are the important ones, and they can be ranked in importance by effect size. However, because of inflated Type I error rate due to multiple testing, more stringent alpha levels are required.

Because there are multiple ANOVAs, a Bonferroni type adjustment is made for inflated Type I error. The researcher assigns alpha for each DV so that alpha for the set of DVs does not exceed some critical value.

$$\alpha = 1 - (1 - \alpha_1)(1 - \alpha_2)\ldots(1 - \alpha_p) \tag{7.12}$$

The Type I error rate (α) is based on the error rate for testing the first DV (α_1), the second DV (α_2), and all other DVs to the p^{th}, or last, DV (α_p).

All the alphas can be set at the same level, or more important DVs can be given more liberal alphas. For example, if there are four DVs and α for each DV is set at .01, the overall alpha level according to Equation 7.12 is .039, acceptably below .05 overall. Or if α is set at .02 for 2 DVs, and at .001 for the other 2 DVs, overall α is .042, also below .05. A close approximation if all α_i are to be the same is:

$$\alpha_i = \alpha_{fw}/p$$

where α_{fw} is the family-wise error rate (e.g., .05) and p is the number of tests

Correlated DVs pose two problems with univariate *F*s. First, correlated DVs measure overlapping aspects of the same behavior. To say that two of them are both "significant" mistakenly suggests that the IV affects two different behaviors. For example, if the two DVs are Stanford-Binet IQ and WISC IQ, they are so highly correlated that an IV that affects one surely affects the other. The second problem with reporting univariate *F*s for correlated DVs is inflation of Type I error rate; with correlated DVs, the univariate *F*s are not independent and no straightforward adjustment of the error rate is possible. In this situation, reporting univariate ANOVAs violates the spirit of MANOVA. However, this is still the most common method of interpreting the results of a MANOVA.

Although reporting univariate *F* for each DV is a simple tactic, the report should also contain the pooled within-group correlations among DVs so the reader can make necessary interpretive

adjustments. The pooled within-group correlation matrix is provided by SPSS MANOVA and SAS GLM.

In the example of Table 7.2, there is a significant multivariate effect of treatment (and of disability, although, as previously noted, it is not interesting in this example). It is appropriate to ask which of the two DVs is affected by treatment. Univariate ANOVAs for WRAT-R and WRAT-A are in Tables 7.7 and 7.8, respectively. The pooled within-group correlation between WRAT-R and WRAT-A is .057 with 12 df. Because the DVs are relatively uncorrelated, univariate F with adjustment of α for multiple tests might be considered appropriate (but note the stepdown results in the following section). There are two DVs, so each is set at alpha .025.[9] With 2 and 12 df, critical F is 5.10; with 1 and 12 df, critical F is 6.55. There is a main effect of treatment (and disability) for both WRAT-R and WRAT-A.

7.5.3.2 *Roy-Bargmann Stepdown Analysis*[10]

The problem of correlated univariate F tests with correlated DVs is resolved by stepdown analysis (Bock, 1966; Bock & Haggard, 1968). Stepdown analysis of DVs is analogous to testing the importance of IVs in multiple regression by sequential analysis. Priorities are assigned to DVs according to theoretical or practical considerations.[11] The highest-priority DV is tested in univariate ANOVA,

TABLE 7.7 Univariate Analysis of Variance of WRAT-R Scores

Source	SS	df	MS	*F*
D	520.7778	2	260.3889	5.7439
T	2090.8889	1	2090.8889	46.1225
DT	2.1111	2	1.0556	0.0233
S(*DT*)	544.0000	12	45.3333	

TABLE 7.8 Univariate Analysis of Variance of WRAT-A Scores

Source	SS	df	MS	*F*
D	1126.7778	2	563.3889	12.5352
T	1494.2222	1	1494.2222	33.2460
DT	52.7778	2	26.3889	0.5871
S(*DT*)	539.5668	12	44.9444	

[9]When the design is very complicated and generates many main effects and interactions, further adjustment of α is necessary in order to keep overall α under .15 or so, across the ANOVAs for the DVs.

[10]Stepdown analysis can be run in lieu of MANOVA where a significant stepdown F is interpreted as a significant multivariate effect for the main effect or interaction.

[11]It is also possible to assign priority on the basis of statistical criteria such as univariate F, but the analysis suffers all the problems inherent in stepwise regression, discussed in Chapter 5.

with appropriate adjustment of alpha. The rest of the DVs are tested in a series of ANCOVAs; each successive DV is tested with higher-priority DVs as covariates to see what, if anything, it adds to the combination of DVs already tested. Because successive ANCOVAs are independent, adjustment for inflated Type I error due to multiple testing is the same as in Section 7.5.3.1.

For the example, we assign WRAT-R scores higher priority since reading problems represent the most common presenting symptoms for learning disabled children. To keep overall alpha below .05, individual alpha levels are set at .025 for each of the two DVs. WRAT-R scores are analyzed through univariate ANOVA, as displayed in Table 7.7. Because the main effect of disability is not interesting and the interaction is not statistically significant in MANOVA (Table 7.2), the only effect of interest is treatment. The critical value for testing the treatment effect (6.55 with 1 and 12 df at $\alpha = .025$) is clearly exceeded by the obtained F of 46.1225.

WRAT-A scores are analyzed in ANCOVA with WRAT-R scores as covariate. The results of this analysis appear in Table 7.9.[12] For the treatment effect, critical F with 1 and 11 df at $\alpha = .025$ is 6.72. This exceeds the obtained F of 5.49. Thus, according to stepdown analysis, the significant effect of treatment is represented in WRAT-R scores, with nothing added by WRAT-A scores.

Note that WRAT-A scores show significant univariate but not stepdown F. Because WRAT-A scores are not significant in stepdown analysis does not mean they are unaffected by treatment but rather that no unique variability is shared with treatment after adjustment for differences in WRAT-R. This result occurs despite the relatively low correlation between the DVs.

This procedure can be extended to sets of DVs through MANCOVA. If the DVs fall into categories, such as scholastic variables and attitudinal variables, one can ask whether there is any change in attitudinal variables as a result of an IV, after adjustment for differences in scholastic variables. The attitudinal variables serve as DVs in MANCOVA while the scholastic variables serve as covariates.

7.5.3.3 *Using Discriminant Analysis*

Discriminant analysis, as discussed more fully in Chapter 9, provides information useful in assessing DVs (DVs are predictors in the context of discriminant analysis). A structure (loading) matrix is produced which contains correlations between the linear combination of DVs that maximizes treatment differences and the DVs themselves. DVs that correlate highly with the combination are more important to discrimination among groups.

TABLE 7.9 Analysis of Covariance of WRAT-A Scores, with WRAT-R Scores as the Covariate

Source	SS	df	MS	F
Covariate	1.7665	1	1.7665	0.0361
D	538.3662	2	269.1831	5.5082
T	268.3081	1	268.3081	5.4903
DT	52.1344	2	26.0672	0.5334
$S(DT)$	537.5668	11	48.8679	

[12]A full stepdown analysis is produced as an option through SPSS MANOVA. For illustration, however, it is helpful to show how the analysis develops.

Discriminant analysis also can be used to test each of the DVs in the standard multiple regression sense; the effect on each DV is assessed after adjustment for all other DVs. That is, each DV is assessed as if it were the last one to enter an equation. This is demonstrated in Section 9.6.4.

7.5.3.4 Choosing among Strategies for Assessing DVs

You may find the procedures of Sections 9.6.3 and 9.6.4 more useful than univariate or stepdown F for assessing DVs when you have a significant multivariate main effect with more than two levels. Similarly, you may find the procedures described in Section 8.5.2 helpful for assessment of DVs if you have a significant multivariate interaction.

The choice between univariate and stepdown F is not always easy, and often you want to use both. When there is no correlation among the DVs, univariate F with adjustment for Type I error is acceptable. When DVs are correlated, as they almost always are, stepdown F is preferable on grounds of statistical purity, but you have to prioritize the DVs and the results can be difficult to interpret.

If DVs are correlated and there is some compelling priority ordering of them, stepdown analysis is clearly called for, with univariate Fs and pooled within-cell correlations reported simply as supplemental information. For significant lower-priority DVs, marginal and/or cell means adjusted for higher-priority DVs are reported and interpreted.

If the DVs are correlated but the ordering is somewhat arbitrary, an initial decision in favor of stepdown analysis is made. If the pattern of results from stepdown analysis makes sense in the light of the pattern of univariate results, interpretation takes both patterns into account with emphasis on DVs that are significant in stepdown analysis. If, for example, a DV has a significant univariate F but a nonsignificant stepdown F, interpretation is straightforward: The variance the DV shares with the IV is already accounted for through overlapping variance with one or more higher-priority DVs. This is the interpretation of WRAT-A in the preceding section and the strategy followed in Section 7.6.

But if a DV has a nonsignificant univariate F and a significant stepdown F, interpretation is much more difficult. In the presence of higher-order DVs as covariates, the DV suddenly takes on "importance." In this case, interpretation is tied to the context in which the DVs entered the stepdown analysis. It may be worthwhile at this point, especially if there is only a weak basis for ordering DVs, to forgo evaluation of statistical significance of DVs and resort to simple description. After finding a significant multivariate effect, unadjusted marginal and/or cell means are reported for DVs with high univariate Fs but significance levels are not given.

An alternative to attempting interpretation of either univariate or stepdown F is interpretation of loading matrices in discriminant analysis, as discussed in Section 9.6.3.2. This process is facilitated when SPSS MANOVA or SAS GLM is used because information about the discriminant functions is provided as a routine part of the output. Alternatively, a discriminant analysis may be run on the data.

Another perspective is whether DVs differ significantly in the effects of IVs on them. For example: Does treatment affect reading significantly more than it affects arithmetic? Tests for contrasts among DVs have been developed in the context of meta-analysis with its emphasis on comparing effect sizes. Rosenthal (2001) demonstrates these techniques.

7.5.4 Specific Comparisons and Trend Analysis

When there are more than two levels in a significant multivariate main effect and when a DV is important to the main effect, the researcher often wants to perform specific comparisons or trend

analysis of the DV to pinpoint the source of the significant difference. Similarly, when there is a significant multivariate interaction and a DV is important to the interaction, the researcher follows up the finding with comparisons on the DV. Specific comparisons may also be done on multivariate effects. These are often less interpretable than comparisons on individual DVs, unless DVs are all scaled in the same direction, or are based on factor or principal component scores. Review Sections 3.2.6, 6.5.4.3, and 8.5.2 for examples and discussions of comparisons. The issues and procedures are the same for individual DVs in MANOVA as in ANOVA.

Comparisons are either planned (performed in lieu of omnibus F) or post hoc (performed after omnibus F to snoop the data). When comparisons are post hoc, an extension of the Scheffé procedure is used to protect against inflated Type I error due to multiple tests. The procedure is very conservative but allows for an unlimited number of comparisons. Following Scheffé for ANOVA (see Section 3.2.6), the tabled critical value of F is multiplied by the degrees of freedom for the effect being tested to produce an adjusted, and much more stringent, F. If marginal means for a main effect are being contrasted, the degrees of freedom are those associated with the main effect. If cell means are being contrasted, our recommendation is to use the degrees of freedom associated with the interaction.

Various types of contrasts on individual DVs are demonstrated in Sections 8.5.2.1 and 8.5.2.3. The difference between setting up contrasts on individual DVs and setting up contrasts on the combination is that all DVs are included in the syntax. Table 7.10 shows syntax for trend analysis and user-specified orthogonal contrasts on the main effect of DISABLTY for the small-sample example. The coefficients illustrated for the orthogonal contrasts actually are the trend coefficients. Note that SPSS GLM requires fractions in part of the LMATRIX command to produce the right answers.

Use of this syntax also provides univariate tests of contrasts for each DV. None of these contrasts are adjusted for post hoc analysis. The usual corrections are to be applied to minimize inflated Type I error rate unless comparisons are planned (cf. Sections 3.2.6.5, 6.5.4.3, and 8.5.2).

7.5.5 Design Complexity

When between-subjects designs have more than two IVs, extension of MANOVA is straightforward as long as sample sizes are equal within each cell of the design. The partition of variance continues to follow ANOVA, with a variance component computed for each main effect and interaction. The pooled variance-covariance matrix due to differences among subjects within cells serves as the single error term. Assessment of DVs and comparisons proceed as described in Sections 7.5.3 and 7.5.4.

Two major design complexities that arise, however, are inclusion of within-subjects IVs and unequal sample sizes in cells.

7.5.5.1 Within-Subjects and Between-Within Designs

The simplest design with repeated measures is a one-way within-subjects design where the same subjects are measured on a single DV on several different occasions. The design can be complicated by addition of between-subjects IVs or more within-subjects IVs. Consult Chapters 3 and 6 for discussion of some of the problems that arise in ANOVA with repeated measures.

Repeated measures is extended to MANOVA when the researcher measures several DVs on several different occasions. The occasions can be viewed in two ways. In the traditional sense, occasions is a within-subjects IV with as many levels as occasions (Chapter 3). Alternatively, each occasion can

TABLE 7.10 Syntax for Orthogonal Comparisons and Trend Analysis

Type of Comparison	Program	Syntax	Section of Output	Name of Effect
Orthogonal	SPSS GLM	GLM WRATR WRATA BY TREATMNT DISABLTY /METHOD = SSTYPE(3) /INTERCEPT = INCLUDE /CRITERIA = ALPHA(.05) /LMATRIX "LINEAR" DISABLTY 1 0 -1 TREATMNT*DISABLTY 1/2 0 -1/2 1/2 0 -1/2 /LMATRIX "QUADRATIC" DISABLTY 1 -2 1 TREATMNT*DISABLTY 1/2 -2/2 1/2 1/2 -2/2 1/2 /DESIGN = TREATMNT DISABLTY TREATMNT*DISABLTY.	**Custom Hypothesis Tests:** **Multivariate Test Results**	Wilks' Lambda
	SPSS MANOVA	MANOVA WRATR WRATA BY TREATMNT (1, 2) DISABLTY (1, 3) /METHOD = UNIQUE /PARTITION (DISABLTY) /CONTRAST(DISABLTY)=SPECIAL(1 1 1, 1 0-1, 1-2 1) /DESIGN = TREATMNT DISABLTY(1) DISABLTY(2) TREATMNT BY DISABLTY.	EFFECT... DISABLTY(2) EFFECT... DISABLTY(1)	Wilks' Lambda
	SAS GLM	PROC GLM DATA=SASUSER.SS_MANOV; CLASS TREATMNT DISABLTY; MODEL WRATR WRATA = TREATMNT DISABLTY TREATMNT*DISABLTY; CONTRAST 'LINEAR' DISABLTY 1 0 -1; CONTRAST 'QUADRATIC' DISABLTY 1 -2 1; manova h=_all_/short; run;	MANOVA Test Criteria.... No Overall linear (quadratic) Effect	Wilks' Lambda

(continued)

TABLE 7.10 Continued

Type of Comparison	Program	Syntax	Section of Output	Name of Effect
Trend Analysis	SPSS GLM	No special syntax; done as any other user-specified contrasts.	EFFECT... DISABLTY(2)	Wilks' Lambda
	SPSS MANOVA	MANOVA WRATR WRATA BY TREATMNT(1,2) DISABLTY(1,3) /METHOD = UNIQUE /PARTITION (DISABLTY) /CONTRAST(DISABLTY)= POLYNOMIAL (1,2,3) /DESIGN = TREATMNT DISABLTY(1) DISABLTY(2) TREATMNT BY DISABLTY.	EFFECT... DISABLTY(1)	
	SAS GLM	No special syntax for between-subjects IVs.		

be treated as a separate DV—one DV per occasion (Section 7.2.8). In this latter view, if there is more than one DV measured on each occasion, the design is said to be doubly multivariate—multiple DVs are measured on multiple occasions. (There is no distinction between the two views when there are only two levels of the within-subjects IV.)

Section 8.5.3 discusses a doubly-multivariate analysis of a small data set with a between-subjects IV (PROGRAM), a within-subjects IV (MONTH), and two DVs (WTLOSS and ESTEEM), both measured three times. A complete example of a doubly-multivariate design is in Section 8.6.2.

It also is possible to have multiple DVs, but treat the within-subjects IV univariately. This is useful when (1) there are only two levels of the within-subjects IV, (2) there is no concern with violation of sphericity (Sections 3.2.3 and 8.5.1), or (3) a trend analysis is planned to replace the omnibus tests of the within-subjects IV and any interactions with the within-subjects IV. All programs that do doubly-multivariate analysis also show univariate results, therefore the syntax is the same as that used in Section 8.5.3.

7.5.5.2 Unequal Sample Sizes

When cells in a factorial ANOVA have an unequal number of scores, the sum of squares for effect plus error no longer equals the total sum of squares, and tests of main effects and interactions are correlated. There are a number of ways to adjust for overlap in sums of squares (cf. Woodward & Overall, 1975), as discussed in some detail in Section 6.5.4.2, particularly Table 6.10. Both the problem and the solutions generalize to MANOVA.

All the MANOVA programs described in Section 7.7 adjust for unequal *n*. SPSS MANOVA offers both Method 1 adjustment (METHOD = UNIQUE), which is default, and Method 3 adjustment (METHOD = SEQUENTIAL). Method 3 adjustment with survey data through SPSS MANOVA is shown in Section 7.6.2. Method 1—called SSTYPE(3)—is the default among four

options in SPSS GLM. In SAS GLM, Method 1 (called TYPE III or TYPE IV) also is the default among four options available.

7.6 Complete Examples of Multivariate Analysis of Variance and Covariance

In the research described in Appendix B, Section B.1, there is interest in whether the means of several of the variables differ as a function of sex role identification. Are there differences in self-esteem, introversion-extraversion, neuroticism, and so on associated with a woman's masculinity and femininity? Files are MANOVA.*.

Sex role identification is defined by the masculinity and femininity scales of the Bem Sex Role Inventory (Bem, 1974). Each scale is divided at its median to produce two levels of masculinity (high and low), two levels of femininity (high and low), and four groups: Undifferentiated (low femininity, low masculinity), Feminine (high femininity, low masculinity), Masculine (low femininity, high masculinity), and Androgynous (high femininity, high masculinity). The design produces a main effect of masculinity, a main effect of femininity, and a masculinity–femininity interaction.[13]

DVs for this analysis are self-esteem (ESTEEM), internal versus external locus of control (CONTROL), attitudes toward women's role (ATTROLE), socioeconomic level (SEL2), introversion-extraversion (INTEXT), and neuroticism (NEUROTIC). Scales are coded so that higher scores generally represent the more "negative" trait: low self-esteem, greater neuroticism, etc.

Omnibus MANOVA (Section 7.6.2) asks whether these DVs are associated with the two IVs (femininity and masculinity) or their interaction. The Roy-Bargmann stepdown analysis, in conjunction with the univariate F values, allows us to examine the pattern of relationships between DVs and each IV.

In a second example (Section 7.6.3), MANCOVA is performed with SEL2, CONTROL, and ATTROLE used as covariates and ESTEEM, INTEXT, and NEUROTIC used as DVs. The research question is whether the three personality DVs vary as a function of sex role identification (the two IVs and their interaction) after adjusting for differences in socioeconomic status, attitudes toward women's role, and beliefs regarding locus of control of reinforcements.

7.6.1 Evaluation of Assumptions

Before proceeding with MANOVA and MANCOVA, we must assess the variables with respect to practical limitations of the techniques.

7.6.1.1 Unequal Sample Sizes and Missing Data

SPSS FREQUENCIES is run with SORT and SPLIT FILE to divide cases into the four groups. Data and distributions for each DV within each group are inspected for missing values, shape, and variance (see Table 7.11 for output on the CONTROL variable for the Feminine group). The run reveals the presence of a case for which the CONTROL score is missing. No datum is missing on any of the

[13]Some would argue with the wisdom of considering masculinity and femininity separate IVs, and of performing a median split on them to create groups. This example is used for didactic purposes.

TABLE 7.11 Syntax and Selected SPSS FREQUENCIES Output for MANOVA Variables Split by Group

```
MISSING VALUES CONTROL (0)
SORT CASES BY ANDRM.
SPLIT FILE
 SEPARATE BY ANDRM.
FREQUENCIES
 VARIABLES=ESTEEM CONTROL ATTROLE SEL2 INTEXT NEUROTIC /FORMAT NOTABLE
 /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN SKEWNESS SESKEW KURTOSIS
 SEKURT
 HISTOGRAM NORMAL
 /ORDER=ANALYSIS.
```

Frequencies Groups-4 = Feminine

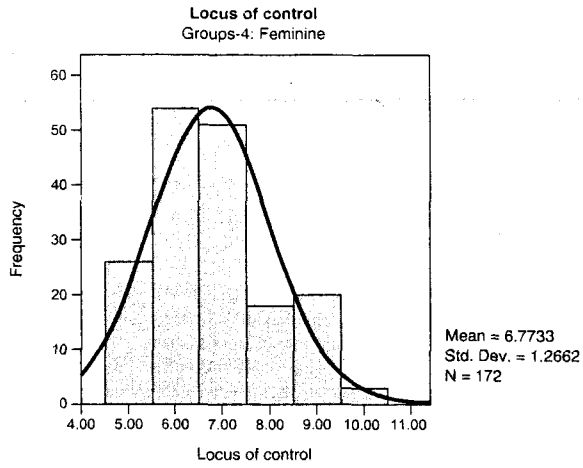
Statistics[a]

		Self esteem	Locus of control	Attitude toward role of women	Socio-economic level	Introversion-extroversion	Neuroticism
N	Valid	173	172	173	173	173	173
	Missing	0	1	0	0	0	0
Mean		16.4913	6.7733	37.0520	40.402643	11.3266	8.9653
Std. Deviation		3.48688	1.26620	6.28145	24.659579	3.66219	5.10688
Variance		12.158	1.603	39.457	608.095	13.412	26.080
Skewness		.471	.541	.076	−.235	−.327	.238
Std. Error of Skewness		.185	.185	.185	.185	.185	.185
Kurtosis		.651	−.381	−.204	−1.284	−.335	−.689
Std. Error of Kurtosis		.367	.368	.367	.367	.367	.367
Minimum		9.00	5.00	22.00	.00000	2.00	.00
Maximum		28.00	10.00	55.00	81.00000	20.00	23.00

[a]Groups-4 = Feminine

TABLE 7.11 Continued

Histogram



other DVs for the 369 women who were administered the Bem Sex Role Inventory. Deletion of the case with the missing value, then, reduces the available sample size to 368.

Sample sizes are quite different in the four groups: There are 71 Undifferentiated, 172 Feminine, 36 Masculine, and 89 Androgynous women in the sample. Because it is assumed that these differences in sample size reflect real processes in the population, the sequential approach to adjustment for unequal n is used with FEM (femininity) given priority over MASC (masculinity), and FEM by MASC (interaction between femininity and masculinity).

7.6.1.2 Multivariate Normality

The sample size of 368 includes over 35 cases for each cell of the 2 × 2 between-subjects design, more than the 20 df for error suggested to assure multivariate normality of the sampling distribution of means, even with unequal sample sizes; there are far more cases than DVs in the smallest cell. Further, the distributions for the full run (of which CONTROL in Table 7.11 is a part) produce no cause for alarm. Skewness is not extreme and, when present, is roughly the same for the DVs.

Two-tailed tests are automatically performed by the computer programs used. That is, the F test looks for differences between means in either direction.

7.6.1.3 Linearity

The full output for the run of Table 7.11 reveals no cause for worry about linearity. All DVs in each group have reasonably balanced distributions so there is no need to examine scatterplots for each pair of DVs within each group. Had scatterplots been necessary, SPSS PLOT would have been used with the SORT and SPLIT FILE syntax in Table 7.11.

7.6.1.4 Outliers

No univariate outliers were found using a criterion $z = |3.3|$ ($\alpha = .001$) with the minimum and maximum values in the full output of Table 7.11. SPSS REGRESSION is used with the split file in

place to check for multivariate outliers within each of the four groups (Table 7.12). The RESIDUALS= OUTLIERS(MAHAL) instruction produces the 10 most outlying cases for each of the groups. With six variables and a criterion $\alpha = .001$, critical $\chi^2 = 22.458$; no multivariate outliers are found.

7.6.1.5 *Homogeneity of Variance-Covariance Matrices*

As a preliminary check for robustness, sample variances (in the full run of Table 7.11) for each DV are compared across the four groups. For no DV does the ratio of largest to smallest variance approach 10:1. As a matter of fact, the largest ratio is about 1.5:1 for the Undifferentiated versus Androgynous groups on CONTROL.

Sample sizes are widely discrepant, with a ratio of almost 5:1 for the Feminine to Masculine groups. However, with very small differences in variance and two-tailed tests, the discrepancy in sam-

TABLE 7.12 Mahalanobis Distance Values for Assessing Multivariate Outliers (Syntax and Selected Portion of Output from SPSS REGRESSION)

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT CASENO
 /METHOD=ENTER ESTEEM CONTROL ATTROLE SEL2 INTEXT NEUROTIC
 /RESIDUALS=OUTLIERS(MAHAL).
```

Regression

Groups-4 = Undifferentiated

Outlier Statistics[a,b]

		Case Number	Statistic
Mahal. Distance	1	32	14.975
	2	71	14.229
	3	64	11.777
	4	5	11.577
	5	41	11.371
	6	37	10.042
	7	55	9.378
	8	3	9.352
	9	1	9.318
	10	25	8.704

[a]Dependent Variable: CASENO
[b]Groups-4 = Undifferentiated

Groups-4 = Masculine

Outlier Statistics[a,b]

		Case Number	Statistic
Mahal. Distance	1	277	14.294
	2	276	11.773
	3	249	11.609
	4	267	10.993
	5	251	9.175
	6	253	8.276
	7	246	7.984
	8	271	7.917
	9	278	7.406
	10	273	7.101

[a]Dependent Variable: CASENO
[b]Groups-4 = Masculine

TABLE 7.12 Continued

Groups-4 = Feminine

Outlier Statistics[a,b]

		Case Number	Statistic
Mahal. Distance	1	209	19.348
	2	208	15.888
	3	116	14.813
	4	92	14.633
	5	167	14.314
	6	233	13.464
	7	150	13.165
	8	79	12.794
	9	138	12.181
	10	179	11.916

[a]Dependent Variable: CASENO
[b]Groups-4 = Feminine

Groups-4 = Androgynous

Outlier Statistics[a,b]

		Case Number	Statistic
Mahal. Distance	1	326	19.622
	2	301	15.804
	3	315	15.025
	4	312	10.498
	5	288	10.381
	6	347	10.101
	7	285	9.792
	8	338	9.659
	9	318	9.203
	10	302	8.912

[a]Dependent Variable: CASENO
[b]Groups-4 = Androgynous

ple sizes does not invalidate use of MANOVA. The very sensitive Box's M test for homogeneity of dispersion matrices (performed through SPSS MANOVA as part of the major analysis in Table 7.15) produces $F(63, 63020) = 1.07$, $p > .05$, supporting the conclusion of homogeneity of variance-covariance matrices.

7.6.1.6 Homogeneity of Regression

Because Roy-Bargmann stepdown analysis is planned to assess the importance of DVs after MANOVA, a test of homogeneity of regression is necessary for each step of the stepdown analysis. Table 7.13 shows the SPSS MANOVA syntax for tests of homogeneity of regression where each DV, in turn, serves as DV on one step and then becomes a covariate on the next and all remaining steps (the split file instruction first is turned off).

Table 7.13 also contains output for the last two steps where CONTROL serves as DV with ESTEEM, ATTROLE, NEUROTIC, and INTEXT as covariates, and then SEL2 is the DV with ESTEEM, ATTROLE, NEUROTIC, INTEXT, and CONTROL as covariates. At each step, the relevant effect is the one appearing last in the column labeled `Source of Variation`, so that for SEL2 the F value for homogeneity of regression is $F(15, 344) = 1.46$, $p > .01$. (The more stringent cutoff is used here because robustness is expected.) Homogeneity of regression is established for all steps.

For MANCOVA, an overall test of homogeneity of regression is required, in addition to stepdown tests. Syntax for all tests is shown in Table 7.14. The ANALYSIS sentence with three DVs specifies the overall test, while the ANALYSIS sentences with one DV each are for stepdown analysis. Output for the overall test and the last stepdown test is also shown in Table 7.14. Multivariate output is printed for the overall test because there are three DVs; univariate results are given for the stepdown tests. All runs show sufficient homogeneity of regression for this analysis.

TABLE 7.13 Test for Homogeneity of Regression for MANOVA Stepdown Analysis (Syntax and Selected Output for Last Two Tests from SPSS MANOVA)

```
SPLIT FILE
 OFF.
MANOVA ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM, MASC(1,2)
    /PRINT=SIGNIF(BRIEF)
    /ANALYSIS=ATTROLE
    /DESIGN=ESTEEM,FEM,MASC,FEM BY MASC, ESTEEM BY FEM BY MASC
    /ANALYSIS=NEUROTIC
    /DESIGN=ATTROLE,ESTEEM,FEM,MASC,FEM BY MASC, POOL(ATTROLE,ESTEEM)
        BY FEM + POOL(ATTROLE,ESTEEM) BY MASC + POOL(ATTROLE,
        ESTEEM)BY FEM BY MASC/
    /ANALYSIS=INTEXT
    /DESIGN=NEUROTIC,ATTROLE,ESTEEM,FEM,MASC,FEM BY MASC, POOL(NEUROTIC,
        ATTROLE,ESTEEM) BY FEM + POOL(NEUROTIC,ATTROLE,ESTEEM)
        BY MASC + POOL(NEUROTIC,ATTROLE,ESTEEM) BY FEM BY MASC
    /ANALYSIS=CONTROL
    /DESIGN=INTEXT,NEUROTIC,ATTROLE,ESTEEM FEM,MASC FEM BY MASC,
        POOL(INTEXT,NEUROTIC,ATTROLE,ESTEEM) BY FEM +
        POOL(INTEXT,NEUROTIC,ATTROLE,ESTEEM) BY MASC +
        POOL(INTEXT,NEUROTIC,ATTROLE,ESTEEM) BY FEM BY MASC
    /ANALYSIS=SEL2
    /DESIGN=CONTROL,INTEXT,NEUROTIC,ATTROLE,ESTEEM,FEM,MASC,FEM BY MASC,
        POOL(CONTROL,INTEXT,NEUROTIC,ATTROLE,ESTEEM) BY FEM +
        POOL(CONTROL,INTEXT,NEUROTIC,ATTROLE,ESTEEM) BY MASC +
        POOL(CONTROL,INTEXT,NEUROTIC,ATTROLE,ESTEEM) BY FEM BY MASC.
```

Tests of Significance for CONTROL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	442.61	348	1.27		
INTEXT	2.19	1	2.19	1.72	.190
NEUROTIC	42.16	1	42.16	33.15	.000
ATTROLE	.67	1	.67	.52	.470
ESTEEM	14.52	1	14.52	11.42	.001
FEM	2.80	1	2.80	2.20	.139
MASC	3.02	1	3.02	2.38	.124
FEM BY MASC	.00	1	.00	.00	.995
POOL(INTEXT NEUROTIC ATTROLE ESTEEM) BY FEM + POOL(INTEXT NEUROTIC ATTROLE ESTEEM) BY MASC + POOL(INTEXT NEUROTIC ATTROLE ESTEEM) BY FEM BY MASC	19.78	12	1.65	1.30	.219

TABLE 7.13 Continued

Tests of Significance of SEL2 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	220340.10	344	640.52		
CONTROL	1525.23	1	1525.23	2.38	.124
INTEXT	.99	1	.99	.00	.969
NEUR	262.94	1	262.94	.41	.522
ATT	182.98	1	182.98	.29	.593
EST	157.77	1	157.77	.25	.620
FEM	1069.23	1	1069.23	1.67	.197
MASC	37.34	1	37.34	.06	.809
FEM BY MASC	1530.73	1	1530.73	2.39	.123
POOL (CONTROL INTEXT NEUROTIC ATTROLE ESTEEM) BY FEM + POOL(CONTROL INTEXT NEUROTIC ATTROLE ESTEEM) BY MASC + POOL(CONTROL INTEXT NEUROTIC ATTROLE ESTEEM) BY FEM BY MASC)	14017.22	15	934.48	1.46	.118

TABLE 7.14 Tests of Homogeneity of Regression for MANCOVA and Stepdown Analysis (Syntax and Partial Output for Overall Tests and Last Stepdown Test from SPSS MANOVA)

```
MANOVA   ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM MASC(1,2)
      /PRINT=SIGNIF(BRIEF)
   /ANALYSIS=ESTEEM,INTEXT,NEUROTIC
   /DESIGN=CONTROL,ATTROLE,SEL2,FEM,MASC,FEM BY MASC,
      POOL(CONTROL,ATTROLE,SEL2) BY FEM +
      POOL(CONTROL,ATTROLE,SEL2) BY MASC +
      POOL(CONTROL,ATTROLE,SEL2) BY FEM BY MASC
   /ANALYSIS=ESTEEM
   /DESIGN=CONTROL,ATTROLE,SEL2,FEM,MASC,FEM BY MASC,
      POOL(CONTROL,ATTROLE,SEL2) BY FEM +
      POOL(CONTROL,ATTROLE,SEL2) BY MASC +
      POOL(CONTROL,ATTROLE,SEL2) BY FEM BY MASC
   /ANALYSIS=INTEXT
   /DESIGN=ESTEEM,CONTROL,ATTROLE,SEL2,FEM,MASC,FEM BY MASC,
      POOL(ESTEEM,CONTROL,ATTROLE,SEL2) BY FEM +
      POOL(ESTEEM,CONTROL,ATTROLE,SEL2) BY MASC +
      POOL(ESTEEM,CONTROL,ATTROLE,SEL2) BY FEM BY MASC
   /ANALYSIS=NEUROTIC
   /DESIGN=INTEXT,ESTEEM,CONTROL,ATTROLE,SEL2,FEM,MASC,FEM BY MASC.
      POOL(INTEXT,ESTEEM,CONTROL,ATTROLE,SEL2) BY FEM+
      POOL(INTEXT,ESTEEM,CONTROL,ATTROLE,SEL2) BY MASC +
      POOL(INTEXT,ESTEEM,CONTROL,ATTROLE,SEL2) BY FEM BY MASC.
```

(continued)

TABLE 7.14 Continued

* * * * * * Analysis of Variance — design 1 * * * * * *

Multivariate Tests of Significance
Tests using UNIQUE sums of squares and WITHIN+RESIDUAL error term

Source of Variation	Wilks	Approx F	Hyp. DF	Error DF	Sig of F
CONTROL	.814	26.656	3.00	350.000	.000
ATTROLE	.973	3.221	3.00	350.000	.023
SEL2	.999	.105	3.00	350.000	.957
FEM	.992	.949	3.00	350.000	.417
MASC	.993	.824	3.00	350.000	.481
FEM BY MASC	.988	1.414	3.00	350.000	.238
POOL(CONTROL ATTROLE SEL2) BY FEM + POOL(CONTROL ATTROLE SEL2) BY MASC + POOL(CONTROL ATTROLE SEL2) BY FEM BY MASC	.933	.911	27.00	1022.823	.596

* * * * * * Analysis of Variance — design 1 * * * * * *

Tests of Significance for NEUROTIC using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	6662.67	344	19.37		
INTEXT	82.19	1	82.19	4.24	.040
ESTEEM	308.32	1	308.32	15.92	.000
CONTROL	699.04	1	699.04	36.09	.000
ATTROLE	1.18	1	1.18	.06	.805
SEL2	2.24	1	2.24	.12	.734
FEM	.07	1	.07	.00	.952
MASC	74.66	1	74.66	3.85	.050
FEM BY MASC	1.65	1	1.65	.09	.770
POOL(INTEXT ESTEEM CONTROL ATTROLE SEL2) BY FEM + POOL(INTEXT ESTEEM CONTROL ATTROLE SEL2) BY MASC + POOL(INTEXT ESTEEM CONTROL ATTROLE SEL2) BY FEM BY MASC	420.19	15	28.01	1.45	.124

7.6.1.7 *Reliability of Covariates*

For the stepdown analysis in MANOVA, all DVs except ESTEEM must be reliable because all act as covariates. Based on the nature of scale development and data collection procedures, there is no reason to expect unreliability of a magnitude harmful to covariance analysis for ATTROLE, NEUROTIC, INTEXT, CONTROL, and SEL2. These same variables act as true or stepdown covariates in the MANCOVA analysis.

7.6.1.8 *Multicollinearity and Singularity*

The log-determinant of the pooled within-cells correlation matrix is found (through SPSS MANOVA syntax in Table 7.15) to be −.4336, yielding a determinant of 2.71. This is sufficiently different from zero that multicollinearity is not judged to be a problem.

7.6.2 Multivariate Analysis of Variance

Syntax and partial output of omnibus MANOVA produced by SPSS MANOVA appear in Table 7.15. The order of IVs listed in the MANOVA statement together with METHOD=SEQUENTIAL sets up

TABLE 7.15 Multivariate Analysis of Variance of Composite of DVs (ESTEEM, CONTROL, ATTROLE, SEL2, INTEXT, and NEUROTIC), as a Function of (Top to Bottom) FEMININITY by MASCULINITY Interaction, MASCULINITY, and FEMININITY (Syntax and Selected Output from SPSS MANOVA)

```
MANOVA      ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM,MASC(1,2)
        /PRINT=SIGNIF(STEPDOWN), ERROR(COR),
          HOMOGENEITY(BARTLETT,COCHRAN,BOXM)
        /METHOD=SEQUENTIAL
        /DESIGN FEM MASC FEM BY MASC.

EFFECT.. FEM BY MASC
Multivariate Tests of Significance (S = 1, M = 2 , N = 178 1/2)

Test Name          Value    Exact F Hypoth. DF   Error DF  Sig. of F

Pillais           .00816     .49230       6.00     359.00       .814
Hotellings        .00823     .49230       6.00     359.00       .814
Wilks             .99184     .49230       6.00     359.00       .814
Roys              .00816
Note.. F statistics are exact.
EFFECT.. MASC
Multivariate Tests of Significance (S = 1, M = 2 , N = 178 1/2)

Test Name          Value    Exact F Hypoth. DF   Error DF  Sig. of F

Pillais           .24363   19.27301       6.00     359.00       .000
Hotellings        .32211   19.27301       6.00     359.00       .000
Wilks             .75637   19.27301       6.00     359.00       .000
Roys              .24363
Note.. F statistics are exact.

EFFECT.. FEM
Multivariate Tests of Significance (S = 1, M = 2 , N = 178 1/2)

Test Name          Value    Exact F Hypoth. DF   Error DF  Sig. of F

Pillais           .08101    5.27423       6.00     359.00       .000
Hotellings        .08815    5.27423       6.00     359.00       .000
Wilks             .91899    5.27423       6.00     359.00       .000
Roys              .08101
Note.. F statistics are exact.
```

the priority for testing FEM before MASC in this unequal-n design. Results are reported for FEM by MASC, MASC, and FEM, in turn. Tests are reported out in order of adjustment where FEM by MASC is adjusted for both MASC and FEM, and MASC is adjusted for FEM.

Four multivariate statistics are reported for each effect. Because there is only one degree of freedom for each effect, three of the tests—Pillai's, Hotelling's, and Wilks'—produce the same F.[14] Both main effects are highly significant, but there is no statistically significant interaction. If desired, effect size for the composite DV for each main effect is found using Equation 7.8 (shown in SPSS MANOVA as Pillai's value) or 7.9. In this case, full and partial η^2 are the same for each of the three effects because $s = 1$ for all of them. Confidence limits for effect sizes are found by entering values from Table 7.15 (`Exact F, Hypoth. DF, Error DF`, and the percentage for the desired confidence interval) into Smithson's (2003) NoncF.sav and running it through NoncF3.sps. Results are added to NoncF.sav, as seen in Table 7.16. (Note that partial η^2 also is reported as `r2`.) Thus, for the main effect of FEM, partial $\eta^2 = .08$ with 95% confidence limits from .02 to .13. For the main effect of MASC, partial $\eta^2 = .24$ with 95% confidence limits from .16 to .30. For the interaction, partial $\eta^2 = .01$ with 95% confidence limits from .00 to .02.

Because omnibus MANOVA shows significant main effects, it is appropriate to investigate further the nature of the relationships among the IVs and DVs. Correlations, univariate Fs, and stepdown Fs help clarify the relationships.

The degree to which DVs are correlated provides information as to the independence of behaviors. Pooled within-cell correlations, adjusted for IVs, as produced by SPSS MANOVA through PRINT = ERROR(COR), appear in Table 7.17. (Diagonal elements are pooled standard deviations.) Correlations among ESTEEM, NEUROTIC, and CONTROL are in excess of .30 so stepdown analysis is appropriate.

Even if stepdown analysis is the primary procedure, knowledge of univariate Fs is required to correctly interpret the pattern of stepdown Fs. And, although the statistical significance of these F values is misleading, investigators frequently are interested in the ANOVA that would have been produced if each DV had been investigated in isolation. These univariate analyses are produced automatically by SPSS MANOVA and shown in Table 7.18 for the three effects in turn: FEM by MASC, MASC, and FEM. F values are substantial for all DVs except SEL2 for MASC and ESTEEM, ATTROLE, and INTEXT for FEM.

Finally, Roy-Bargmann stepdown analysis, produced by PRINT=SIGNIF(STEPDOWN), allows a statistically pure look at the significance of DVs, in context, with Type I error rate controlled.

TABLE 7.16 Data Set Output from NoncF3.sps for Effect Size (r2) with 95% Confidence Limits (lr2 and ur2) for Interaction, MASC, and FEM, Respectively

	fval	df1	df2	conf	lc2	ucdf	uc2	lcdf	power	r2	lr2	ur2
1	.4923	6	359	.950	.000	.186	5.875	.02	.1306	.01	.00	.02
2	19.2730	6	359	.950	70.0	.975	160.0	.02	1.000	.24	.16	.30
3	5.2742	6	359	.950	9.08	.975	52.34	.02	.9910	.08	.02	.13

[14]For more complex designs, a single source table containing all effects can be obtained through PRINT=SIGNIF(BRIEF) but the table displays only Wilks' lambda.

TABLE 7.17 Pooled Within-Cell Correlations among Six DVs (Selected Output from SPSS MANOVA—See Table 7.15 for Syntax)

WITHIN+RESIDUAL Correlations with Std. Devs. on Diagonal

	ESTEEM	ATTROLE	NEUROTIC	INTEXT	CONTROL	SEL2
ESTEEM	3.533					
ATTROLE	.145	6.227				
NEUROTIC	.358	.051	4.965			
INTEXT	-.164	.011	-.009	3.587		
CONTROL	.348	-.031	.387	-.083	1.267	
SEL2	-.035	.016	-.015	.055	-.084	25.501

TABLE 7.18 Univariate Analyses of Variance of Six DVs for Effects of (Top to Bottom) FEM by MASC Interaction, Masculinity, and Femininity (Selected Output from SPSS MANOVA—See Table 7.15 for Syntax)

EFFECT.. FEM BY MASC (Cont.)
Univariate F-tests with (1,364) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
ESTEEM	17.48685	4544.44694	17.48685	12.48474	1.40066	.237
ATTROLE	36.79594	14115.1212	36.79594	38.77781	.94889	.331
NEUROTIC	.20239	8973.67662	.20239	24.65296	.00821	.928
INTEXT	.02264	4684.17900	.02264	12.86862	.00176	.967
CONTROL	.89539	584.14258	.89539	1.60479	.55795	.456
SEL2	353.58143	236708.966	353.58143	650.29936	.54372	.461

EFFECT.. MASC (Cont.)
Univariate F-tests with (1,364) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
ESTEEM	979.60086	4544.44694	979.60086	12.48474	78.46383	.000
ATTROLE	1426.75675	14115.1212	1426.75675	38.77781	36.79313	.000
NEUROTIC	179.53396	8973.67662	179.53396	24.65296	7.28245	.007
INTEXT	327.40797	4684.17900	327.40797	12.86862	25.44235	.000
CONTROL	11.85923	584.14258	11.85923	1.60479	7.38991	.007
SEL2	1105.38196	236708.966	1105.38196	650.29936	1.69980	.193

EFFECT.. FEM (Cont.)
Univariate F-tests with (1,364) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
ESTEEM	101.46536	4544.44694	101.46536	12.48474	8.12715	.005
ATTROLE	610.88860	14115.1212	610.88860	38.77781	15.75356	.000
NEUROTIC	44.05442	8973.67662	44.05442	24.65296	1.78698	.182
INTEXT	87.75996	4684.17900	87.75996	12.86862	6.81968	.009
CONTROL	2.83106	584.14258	2.83106	1.60479	1.76414	.185
SEL2	9.00691	236708.966	9.00691	650.29936	.01385	.906

For this study, the following priority order of DVs is developed, from most to least important: ESTEEM, ATTROLE, NEUROTIC, INTEXT, CONTROL, SEL2. Following the procedures for stepdown analysis (Section 7.5.3.2), the highest-priority DV, ESTEEM, is tested in univariate ANOVA. The second-priority DV, ATTROLE, is assessed in ANCOVA with ESTEEM as the covariate. The third-priority DV, NEUROTIC, is tested with ESTEEM and ATTROLE as covariates, and so on, until all DVs are analyzed. Stepdown analyses for the interaction and both main effects are in Table 7.19.

For purposes of journal reporting, critical information from Tables 7.18 and 7.19 is consolidated into a single table with both univariate and stepdown analyses, as shown in Table 7.20. The alpha level established for each DV is reported along with the significance levels for stepdown *F*. The final three columns show partial η^2 with 95% confidence limits for all stepdown effects, described later.

For the main effect of FEM, ESTEEM and ATTROLE are significant. (INTEXT would be significant in ANOVA but its variance is already accounted for through overlap with ESTEEM, as noted

TABLE 7.19 Stepdown Analyses of Six Ordered DVs for (Top to Bottom) FEM by MASC Interaction, Masculinity, and Femininity (Selected Output from SPSS MANOVA—see Table 7.16 for Syntax)

Roy-Bargman Stepdown F-tests

Variable	Hypoth. MS	Error MS	Stepdown F	Hypoth. DF	Error DF	Sig. of F
ESTEEM	17.48685	12.48474	1.40066	1	364	.237
ATTROLE	24.85653	38.06383	.65302	1	363	.420
NEUROTIC	2.69735	21.61699	.12478	1	362	.724
INTEXT	.26110	12.57182	.02077	1	361	.885
CONTROL	.41040	1.28441	.31952	1	360	.572
SEL2	297.09000	652.80588	.45510	1	359	.500

Roy-Bargman Stepdown F-tests

Variable	Hypoth. MS	Error MS	Stepdown F	Hypoth. DF	Error DF	Sig. of F
ESTEEM	979.60086	12.48474	78.46383	1	364	.000
ATTROLE	728.51682	38.06383	19.13935	1	363	.000
NEUROTIC	4.14529	21.61699	.19176	1	362	.662
INTEXT	139.98354	12.57182	11.13471	1	361	.001
CONTROL	.00082	1.28441	.00064	1	360	.980
SEL2	406.59619	652.80588	.62284	1	359	.431

Roy-Bargman Stepdown F-tests

Variable	Hypoth. MS	Error MS	Stepdown F	Hypoth. DF	Error DF	Sig. of F
ESTEEM	101.46536	12.48474	8.12715	1	364	.005
ATTROLE	728.76735	38.06383	19.14593	1	363	.000
NEUROTIC	2.21946	21.61699	.10267	1	362	.749
INTEXT	47.98941	12.57182	3.81722	1	361	.052
CONTROL	.05836	1.28441	.04543	1	360	.831
SEL2	15.94930	652.80588	.02443	1	359	.876

TABLE 7.20 Tests of Femininity, Masculinity, and Their Interaction

IV	DV	Univariate *F*	df	Stepdown *F*	df	*a*	Partial η^2	CL around Partial η^2 per *a* Lower	Upper
Femininity	ESTEEM	8.13[a]	1/364	8.13**	1/364	.01	.02	.00	.07
	ATTROLE	15.75[a]	1/364	19.15**	1/363	.01	.05	.01	.12
	NEUROTIC	1.79	1/364	0.10	1/362	.01	.00	.00	.01
	INTEXT	6.82[a]	1/364	3.82	1/361	.01	.01	.00	.05
	CONTROL	1.76	1/364	0.05	1/360	.01	.00	.00	.01
	SEL2	0.01	1/364	0.02	1/359	.001	.00	.00	.00
Masculinity	ESTEEM	78.46[a]	1/364	78.46**	1/364	.01	.18	.09	.27
	ATTROLE	36.79[a]	1/364	19.14**	1/363	.01	.05	.01	.12
	NEUROTIC	7.28[a]	1/364	0.19	1/362	.01	.00	.00	.02
	INTEXT	25.44[a]	1/364	11.13**	1/361	.01	.03	.00	.09
	CONTROL	7.39[a]	1/364	0.00	1/360	.01	.00	.00	.00
	SEL2	1.70	1/364	0.62	1/359	.001	.00	.00	.04
Femininity by masculinity interaction	ESTEEM	1.40	1/364	1.40	1/364	.01	.00	.00	.04
	ATTROLE	0.95	1/364	0.65	1/363	.01	.00	.00	.03
	NEUROTIC	0.01	1/364	0.12	1/362	.01	.00	.00	.01
	INTEXT	0.00	1/364	0.02	1/361	.01	.00	.00	.00
	CONTROL	0.56	1/364	0.32	1/360	.01	.00	.00	.03
	SEL2	0.54	1/364	0.46	1/359	.001	.00	.00	.04

[a]Significance level cannot be evaluated but would reach $p < .01$ in univariate context.

**$p < .01$.

in the pooled within-cell correlation matrix.) For the main effect of MASC, ESTEEM, ATTROLE, and INTEXT are significant. (NEUROTIC and CONTROL would be significant in ANOVA, but their variance is also already accounted for through overlap with ESTEEM, ATTROLE, and, in the case of CONTROL, NEUROTIC and INTEXT.)

For the DVs significant in stepdown analysis, the relevant adjusted marginal means are needed for interpretation. Marginal means are needed for ESTEEM for FEM and for MASC adjusted for FEM. Also needed are marginal means for ATTROLE with ESTEEM as a covariate for both FEM, and MASC adjusted for FEM; lastly, marginal means are needed for INTEXT with ESTEEM, ATTROLE, and NEUROTIC as covariates for MASC adjusted for FEM. Table 7.21 contains syntax and selected output for these marginal means as produced through SPSS MANOVA. In the table, level of effect is identified under `PARAMETER` and mean is under `Coeff`. Thus, the mean for ESTEEM at level 1 of FEM is 16.57. Marginal means for effects with univariate, but not stepdown, differences are shown in Table 7.22 where means for NEUROTIC and CONTROL are found for the main effect of MASC adjusted for FEM.

Effect size for each DV is evaluated as partial η^2 (Equation 3.25, 3.26, 6.7, 6.8, or 6.9). The information you need for calculation of η^2 is available in SPSS MANOVA stepdown tables (see Table 7.19) but not in a convenient form; mean squares are given in the tables but you need sums of squares for calculation of η^2. Smithson's (2003) program (NoncF3.sps) calculates confidence limits for effect sizes

TABLE 7.21 Adjusted Marginal Means for ESTEEM; ATTROLE with ESTEEM as a Covariate; and INTEXT with ESTEEM, ATTROLE, and NEUROTIC as Covariates (Syntax and Selected Output from SPSS MANOVA)

```
MANOVA      ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM,MASC(1,2)
      /PRINT=PARAMETERS(ESTIM)
    /ANALYSIS=ESTEEM /DESIGN=CONSPLUS FEM
            /DESIGN=FEM,CONSPLUS MASC
    /ANALYSIS=ATTROLE WITH ESTEEM /DESIGN=CONSPLUS FEM
                    /DESIGN=FEM, CONSPLUS MASC
    /ANALYSIS=INTEXT WITH ESTEEM,ATTROLE,NEUROTIC
        /DESIGN=FEM, CONSPLUS MASC.


Estimates for ESTEEM
--- Individual univariate .9500 confidence intervals

CONSPLUS FEM

Parameter      Coeff.  Std. Err.    t-Value   Sig. t Lower -95%  CL- Upper

      1   16.5700935     .37617   44.04964   .00000   15.83037   17.30982
      2   15.4137931     .24085   63.99636   .00000   14.94016   15.88742

Estimates for ESTEEM
--- Individual univariate .9500 confidence intervals

CONSPLUS MASC

Parameter      Coeff.  Std. Err.    t-Value   Sig. t Lower -95%  CL- Upper

      2   17.1588560     .24196   70.91464   .00000   16.68304   17.63468
      3   13.7138144     .32770   41.84820   .00000   13.06939   14.35824

Estimates for ATTROLE adjusted for 1 covariate
--- Individual univariate .9500 confidence intervals

CONSPLUS FEM

Parameter      Coeff.  Std. Err.    t-Value   Sig. t Lower -95%  CL- Upper

      1   32.5743167     .61462   52.99941   .00000   31.36568   33.78295
      2   35.9063146     .39204   91.58908   .00000   35.13538   36.67725

Estimates for ATTROLE adjusted for 1 covariate
--- Individual univariate .9500 confidence intervals

CONSPLUS MASC

Parameter      Coeff.  Std. Err.    t-Value   Sig. t Lower -95%  CL- Upper

      2   35.3849271     .44123   80.19697   .00000   34.51726   36.25260
      3   32.1251995     .60108   53.44537   .00000   30.94316   33.30723
```

TABLE 7.21 Continued

```
Estimates for INTEXT adjusted for 3 covariates
- Individual univariate .9500 confidence intervals

CONSPLUS MASC

Parameter      Coeff.   Std. Err.     t-Value    Sig. t Lower -95%  CL- Upper

      2   11.0013930      .25372    43.36122    .00000    10.50245   11.50033
      3   12.4772029      .35546    35.10172    .00000    11.77818   13.17623
```

Note: Coeff. = adjusted marginal mean; first parameter = low, second parameter = high.

TABLE 7.22 Unadjusted Marginal Means for Neurotic and Control (Syntax and Selected Output from SPSS MANOVA)

```
MANOVA      ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM,MASC(1,2)
        /PRINT=PARAMETERS(ESTIM)
        /ANALYSIS=NEUROTIC /DESIGN=FEM, CONSPLUS MASC
        /ANALYSIS=CONTROL /DESIGN=FEM, CONSPLUS MASC.

Estimates for NEUROTIC
--- Individual univariate .9500 confidence intervals

CONSPLUS MASC

Parameter      Coeff.   Std. Err.     t-Value    Sig. t Lower -95%  CL- Upper

      2    9.37093830      .33937    27.61309    .00000     8.70358   10.03830
      3    7.89610411      .45962    17.17971    .00000     6.99227    8.79994

Estimates for CONTROL
--- Individual univariate .9500 confidence intervals

CONSPLUS MASC

Parameter      Coeff.   Std. Err.     t-Value    Sig. t Lower -95%  CL- Upper

      2    6.89163310      .08665    79.53388    .00000     6.72124    7.06203
      3    6.51258160      .11735    55.49504    .00000     6.28181    6.74336
```

Note: Coeff. = unadjusted marginal mean; first parameter = low, second parameter = high.

and also calculates the effect size itself from F (stepdown or otherwise), df for effect (df1) and error (df2), and the percentage associated with the desired confidence limits. These four values are entered into the data sheet (NoncF.sav). The remaining columns of NoncF.sav are filled in when NoncF3.sps is run. The relevant output columns are r2 (equivalent to partial η^2 of Equation 6.9), lr2 and ur2, the lower and upper confidence limits, respectively, for the effect size. Table 7.23 shows the input/output data set for all of the stepdown effects following the order in Table 7.20, e.g., 1 = ESTEEM for FEM, 2 = ATTROLE for FEM, 3 = NEUROTIC for FEM and so on. Values filled into the first three columns

are from Table 7.20. The value of .99 or .999 filled in for the confidence limits reflects the chosen α level for each effect.

A checklist for MANOVA appears in Table 7.24. An example of a Results section, in journal format, follows for the study just described.

TABLE 7.23 Data Set Output for Stepdown Effects from NoncF3.sps for Effect Size (r2) with 95% Confidence Limits (lr2 and ur2)

	fval	df1	df2	conf	lc2	ucdf	uc2	lcdf	power	r2	lr2	ur2
1	8.1300	1	364	.990	.0208	.9950	29.5824	.0050	.5146	.02	.00	.07
2	19.1500	1	363	.990	3.1051	.9950	48.7540	.0050	.9403	.05	.01	.12
3	.0100	1	362	.990	.0000	.0796	.4300	.0642	.0052	.00	.00	.00
4	3.8200	1	361	.990	.0000	.9486	20.5773	.0050	.1955	.01	.00	.05
5	.0500	1	360	.990	.0000	.1768	2.1500	.0614	.0061	.00	.00	.01
6	.0200	1	359	.999	.0000	.1124	.8600	.0733	.0006	.00	.00	.00
7	78.4600	1	364	.990	37.6976	.9950	133.7230	.0050	1.0000	.18	.09	.27
8	19.1400	1	363	.990	3.1034	.9950	48.7472	.0050	.9402	.05	.01	.12
9	.1900	1	362	.990	.0000	.3368	8.1700	.0072	.0094	.00	.00	.02
10	11.1300	1	361	.990	.5444	.9950	35.1617	.0050	.6985	.03	.00	.09
11	.0000	1	360	.990	.0000	.1587	.0000	.1587	.0015	.00	.00	.00
12	.6200	1	359	.999	.0000	.5684	16.6625	.0005	.0035	.00	.00	.04
13	1.4000	1	364	.990	.0000	.7625	14.1422	.0050	.0519	.00	.00	.04
14	.6500	1	363	.990	.0000	.5794	11.4359	.0050	.0227	.00	.00	.03
15	.1200	1	362	.990	.0000	.2708	5.1600	.0227	.0077	.00	.00	.01
16	.0200	1	361	.990	.0000	.1124	.8600	.0733	.0054	.00	.00	.00
17	.3200	1	360	.990	.0000	.4280	9.8200	.0050	.0128	.00	.00	.03
18	.4600	1	359	.999	.0000	.5019	15.7550	.0005	.0025	.00	.00	.04

TABLE 7.24 Checklist for Multivariate Analysis of Variance

1. Issues
   a. Unequal sample sizes and missing data
   b. Normality of sampling distributions
   c. Outliers
   d. Homogeneity of variance-covariance matrices
   e. Linearity
   f. In stepdown, when DVs act as covariates
      (1) Homogeneity of regression
      (2) Reliability of DVs
   g. Multicollinearity and singularity
2. Major analyses: Planned comparisons or omnibus *F*, when significant. Importance of DVs
   a. Within-cell correlations, stepdown *F*, univariate *F*
   b. Effect sizes with confidence interval for significant stepdown *F*
   c. Means or adjusted marginal and/or cell means for significant *F*, with standard deviations, standard errors, or confidence intervals
3. Multivariate effect size(s) with confidence interval(s) for planned comparisons or omnibus *F*
4. Additional analyses
   a. Post hoc comparisons
   b. Interpretation of IV–covariates interaction (if homogeneity of regression violated)

Results

A 2 × 2 between-subjects multivariate analysis of variance was performed on six dependent variables: Self-esteem, attitude toward the role of women, neuroticism, introversion-extraversion, locus of control, and socioeconomic level. Independent variables were masculinity (low and high) and femininity (low and high).

SPSS MANOVA was used for the analyses with the sequential adjustment for nonorthogonality. Order of entry of IVs was femininity, then masculinity. Total N of 369 was reduced to 368 with the deletion of a case missing a score on locus of control. There were no univariate or multivariate within-cell outliers at $p < .001$. Results of evaluation of assumptions of normality, homogeneity of variance-covariance matrices, linearity, and multicollinearity were satisfactory.

With the use of Wilks' criterion, the combined DVs were significantly affected by both masculinity, $F(6, 359) = 19.27$, $p < .001$, and femininity, $F(6, 359) = 5.27$, $p < .001$, but not by their interaction, $F(6, 359) = 0.49$, $p > .05$. The results reflected a modest association between masculinity scores (low vs. high) and the combined DVs, partial $\eta^2 = .24$ with 95% confidence limits from .16 to .30. The association was even less substantial between femininity and the DVs, partial $\eta^2 = .08$ with 95% confidence limits from .02 to .13. For the nonsignificant interaction, $\eta^2 = .01$ with 95% confidence limits from .00 to .02. [*F and Pillai's value (partial* η^2*) are from Table 7.15; confidence limits for* η^2 *are found through NoncF3.sps.*]

To investigate the impact of each main effect on the individual DVs, a Roy-Bargmann stepdown analysis was performed on the prioritized DVs. All DVs were judged to be sufficiently reliable to warrant stepdown analysis. In stepdown analysis each DV was analyzed, in turn, with higher-priority DVs treated as covariates and with the highest-priority DV tested in a univariate ANOVA. Homogeneity of regression was achieved for all components of the stepdown analysis.

Results of this analysis are summarized in Table 7.20. An experimentwise error rate of 5% was achieved by the apportionment of alpha as shown in the last column of Table 7.20 for each of the DVs.

A unique contribution to predicting differences between those low and high on femininity was made by self-esteem, stepdown $F(1, 364) = 8.13$, $p < .01$, $\eta^2 = .02$ with 99% confidence limits from .00 to .07. Self-esteem was scored inversely, so women with higher femininity scores showed greater self-esteem (mean self-esteem = 15.41, SE = 0.24) than those with lower femininity (mean self-esteem = 16.57, SE = 0.38). After the pattern of differences measured by self-esteem was entered, a difference was also found on attitude toward the role of women, stepdown $F(1, 363) = 19.15$, $p < .01$, $\eta^2 = .05$ with confidence limits from .01 to .12. Women with higher femininity scores had more conservative attitudes toward women's role (adjusted mean attitude = 35.90, SE = 0.35) than those lower in femininity (adjusted mean attitude = 32.57, SE = 0.61). Although a univariate comparison revealed that those higher in femininity also were more extroverted, univariate $F(1, 364) = 6.82$, this difference was already represented in the stepdown analysis by higher-priority DVs.

Three DVs—self-esteem, attitude toward role of women, and introvert-extrovert—made unique contributions to the composite DV that best distinguished between those high and low in masculinity. The greatest contribution was made by self-esteem, the highest-priority DV, stepdown $F(1, 364) = 78.46$, $p < .01$, $\eta^2 = 18$ with confidence limits from .09 to .27. Women scoring high in masculinity had higher self-esteem (mean self-esteem = 13.71, SE = 0.33) than those scoring low (mean self-esteem = 17.16, SE = 0.24). With differences due to self-esteem already entered, attitudes toward the role of women made a unique contribution, stepdown $F(1, 363) = 19.14$, $p < .01$, $\eta^2 = .05$ with confidence limits from .01 to .12. Women scoring lower in masculinity had more conservative attitudes toward the

proper role of women (adjusted mean attitude = 35.39, SE = 0.44) than those scoring higher (adjusted mean attitude = 32.13, SE = 0.60). Introversion-extraversion, adjusted by self-esteem, attitudes toward women's role, and neuroticism also made a unique contribution to the composite DV, stepdown $F(1, 361) = 11.13$, $p < .01$, $\eta^2 = .03$ with confidence limits from .00 to .09. Women with higher masculinity were more extroverted (mean adjusted introversion-extraversion score = 12.48) than lower masculinity women (mean adjusted introversion-extraversion score = 11.00). Univariate analyses revealed that women with higher masculinity scores were also less neurotic, univariate $F(1, 364) = 7.28$, and had a more internal locus of control, univariate $F(1, 364)$ 7.39, differences that were already accounted for in the composite DV by higher-priority DVs. [*Means adjusted for main effects and for other DVs for stepdown interpretation are from Table 7.21,* partial η^2 *values and confidence limits are from Table 7.23. Means adjusted for main effects but not other DVs for univariate interpretation are in Table 7.22.*]

High-masculinity women, then, have greater self-esteem, less conservative attitudes toward the role of women, and more extraversion than women scoring low on masculinity. High femininity is associated with greater self-esteem and more conservative attitudes toward women's role than low femininity. Of the five effects, however, only the association between masculinity and self-esteem shows even a moderate proportion of shared variance.

Pooled within-cell correlations among DVs are shown in Table 7.17.

The only relationships accounting for more than 10% of variance are between self-esteem and neuroticism ($r = .36$), locus of control and self-esteem ($r = .35$), and between neuroticism and locus of control ($r = .39$). Women who are high in neuroticism tend to have lower self-esteem and more external locus of control.

7.6.3 Multivariate Analysis of Covariance

For MANCOVA the same six variables are used as for MANOVA but ESTEEM, INTEXT, and NEUROTIC are used as DVs and CONTROL, ATTROLE, and SEL2 are used as covariates. The research question is whether there are personality differences associated with femininity, masculinity, and their interaction after adjustment for differences in attitudes and socioeconomic status.

Syntax and partial output of omnibus MANCOVA as produced by SPSS MANOVA appear in Table 7.25. As in MANOVA, Method 3 adjustment for unequal n is used with MASC adjusted for FEM and the interaction is adjusted for FEM and MASC. And, as in MANOVA, both main effects are highly significant but there is no interaction. Effect sizes for the three effects are Pillai's values. Entering `Approx. F` and appropriate df and percentage values into the NoncF.sav program and running NoncF3.sps, 95% confidence limits for these effect sizes are .00 to 08 for FEM, .08 to .21 for MASC, and .00 to .01 for the interaction.

7.6.3.1 Assessing Covariates

Under `EFFECT..WITHIN+RESIDUAL Regression` is the multivariate significance test for the relationship between the set of DVs (ESTEEM, INTEXT, and NEUROTIC) and the set of covariates (CONTROL, ATTROLE, and SEL2) after adjustment for IVs. Partial η^2 is calculated through the NoncF3.sps algorithm (Pillai's criterion is inappropriate unless $s = 1$) using `Approx. F` and appropriate df and is found to be .10 with 95% confidence limits from .06 to .13.

Because there is multivariate significance, it is useful to look at the three multiple regression analyses of each DV in turn, with covariates acting as IVs (see Chapter 5). The syntax of Table 7.25 automatically produces these regressions. They are done on the pooled within-cell correlation matrix, so that effects of the IVs are eliminated.

The results of the DV-covariate multiple regressions are shown in Table 7.26. At the top of Table 7.26 are the results of the univariate and stepdown analysis, summarizing the results of multiple regressions for the three DVs independently and then in priority order (see Section 7.6.3.2). At the bottom of Table 7.26 under `Regression analysis for WITHIN+RESIDUAL error term` are the separate regressions for each DV with covariates as IVs. For ESTEEM, two covariates, CONTROL and ATTROLE, are significantly related but SEL2 is not. None of the three covariates is related to INTEXT. Finally, for NEUROTIC, only CONTROL is significantly related. Because SEL2 provides no adjustment to any of the DVs, it could be omitted from future analyses.

7.6.3.2 Assessing DVs

Procedures for evaluating DVs, now adjusted for covariates, follow those specified in Section 7.6.2 for MANOVA. Correlations among all DVs, among covariates, and between DVs and covariates are informative so all the correlations in Table 7.17 are still relevant.[15]

Univariate Fs are now adjusted for covariates. The univariate ANCOVAs produced by the SPSS MANOVA run specified in Table 7.25 are shown in Table 7.27. Although significance levels are misleading, there are substantial F values for ESTEEM and INTEXT for MASC (adjusted for FEM) and for FEM.

For interpretation of effects of IVs on DVs adjusted for covariates, comparison of stepdown Fs with univariate Fs again provides the best information. The priority order of DVs for this analysis is

[15]For MANCOVA, SPSS MANOVA prints pooled within-cell correlations among DVs (called criteria) adjusted for covariates. To get a pooled within-cell correlation matrix for covariates as well as DVs, you need a run in which covariates are included in the set of DVs.

TABLE 7.25 Multivariate Analysis of Covariance of Composite of DVs (ESTEEM, INTEXT, and NEUROTIC) as a Function of (Top to Bottom) FEM by MASC Interaction, Masculinity, and Femininity; Covariates are ATTROLE, CONTROL, and SEL2 (Syntax and Selected Output from SPSS MANOVA)

```
MANOVA      ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM,MASC(1,2)
        /ANALYSIS=ESTEEM,INTEXT,NEUROTIC WITH CONTROL,ATTROLE,SEL2
        /PRINT=SIGNIF(STEPDOWN), ERROR(COR),
           HOMOGENEITY(BARTLETT,COCHRAN,BOXM)
        /METHOD=SEQUENTIAL
        /DESIGN FEM MASC FEM BY MASC.

EFFECT .. WITHIN+RESIDUAL Regression
Multivariate Tests of Significance (S = 3, M = -1/2, N = 178 1/2)

Test Name          Value  Approx. F Hypoth. DF   Error DF  Sig. of F

Pillais          .23026   10.00372        9.00    1083.00        .000
Hotellings       .29094   11.56236        9.00    1073.00        .000
Wilks            .77250   10.86414        9.00     873.86        .000
Roys             .21770

EFFECT.. FEM BY MASC
Multivariate Tests of Significance (S = 1, M = 1/2, N = 178 1/2)

Test Name          Value  Approx. F Hypoth. DF   Error DF  Sig. of F

Pillais          .00263     .31551        3.00     359.00        .814
Hotellings       .00264     .31551        3.00     359.00        .814
Wilks            .99737     .31551        3.00     359.00        .814
Roys             .00263
Note.. F statistics are exact.

EFFECT.. MASC
Multivariate Tests of Significance (S = 1, M = 1/2, N = 178 1/2)

Test Name          Value  Approx. F Hypoth. DF   Error DF  Sig. of F

Pillais          .14683   20.59478        3.00     359.00        .000
Hotellings       .17210   20.59478        3.00     359.00        .000
Wilks            .85317   20.59478        3.00     359.00        .000
Roys             .14683
Note.. F statistics are exact.

EFFECT.. FEM
Multivariate Tests of Significance (S = 1, M = 1/2, N = 178 1/2)

Test Name          Value  Approx. F Hypoth. DF   Error DF  Sig. of F

Pillais          .03755    4.66837        3.00     359.00        .003
Hotellings       .03901    4.66837        3.00     359.00        .003
Wilks            .96245    4.66837        3.00     359.00        .003
Roys             .03755
Note.. F statistics are exact.
```

TABLE 7.26 Assessment of Covariates: Univariate, Stepdown, and Multiple Regression Analyses for Three DVs with Three Covariates (Selected Output from SPSS MANOVA—see Table 7.25 for Syntax)

EFFECT.. WITHIN+RESIDUAL Regression (Cont.)
Univariate F-tests with (3,361) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F
ESTEEM	660.84204	3883.60490	220.28068	10.75791	20.47616
INTEXT	43.66605	4640.51295	14.55535	12.85461	1.13231
NEUROTIC	1384.16059	7589.51604	461.38686	21.02359	21.94615

Variable	Sig. of F
ESTEEM	.000
INTEXT	.336
NEUROTIC	.000

Roy-Bargman Stepdown F-tests

Variable	Hypoth. MS	Error MS	Stepdown F	Hypoth. DF	Error DF	Sig. of F
ESTEEM	220.28068	10.75791	20.47616	3	361	.000
INTEXT	6.35936	12.60679	.50444	3	360	.679
NEUROTIC	239.94209	19.72942	12.16164	3	359	.000

Regression analysis for WITHIN+RESIDUAL error term
--- Individual Univariate .9500 confidence intervals
Dependent variable .. ESTEEM Self-esteem

COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t
CONTROL	.98173	.32005	.136	7.205	.000
ATTROLE	.08861	.15008	.028	3.208	.001
SEL2	-.00111	-.00723	.007	-.164	.869

Regression analysis for WITHIN+RESIDUAL error term
Dependent variable .. ESTEEM Self-esteem

COVARIATE	Lower -95% CL-	Upper
CONTROL	.714	1.250
ATTROLE	.034	.143
SEL2	-.014	.012

Dependent variable .. INTEXT Introversion-extroversion

COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t
CONTROL	-.22322	-.07655	.149	-1.499	.135
ATTROLE	.00456	.00812	.030	.151	.880
SEL2	.00682	.04662	.007	.922	.357

TABLE 7.26 Continued

COVARIATE	Lower -95% CL-	Upper
CONTROL	-.516	.070
ATTROLE	-.055	.064
SEL2	-.008	.021

Dependent variable .. NEUROTIC Neuroticism

COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t
CONTROL	1.53128	.39102	.190	8.040	.000
ATTROLE	.04971	.06595	.039	1.287	.199
SEL2	.00328	.01670	.009	.347	.729

COVARIATE	Lower -95% CL-	Upper
CONTROL	1.157	1.906
ATTROLE	-.026	.126
SEL2	-.015	.022

TABLE 7.27 Univariate Analyses of Covariance of Three DVs Adjusted for Three Covariates for (Top to Bottom) FEM by MASC Interaction, Masculinity, and Femininity (Selected Output from SPSS MANOVA—see Table 7.25 for Syntax)

EFFECT.. FEM BY MASC (Cont.)
Univariate F-tests with (1,361) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
ESTEEM	7.21931	3883.60490	7.21931	10.75791	.67107	.413
INTEXT	2.59032	464051.295	2.59032	1285.46065	.00202	.964
NEUROTIC	1.52636	7589.51604	1.52636	21.02359	.07260	.788

EFFECT.. MASC (Cont.)
Univariate F-tests with (1,361) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
ESTEEM	533.61774	3883.60490	533.61774	10.75791	49.60237	.000
INTEXT	26444.5451	464051.295	26444.5451	1285.46065	20.57204	.000
NEUROTIC	35.82929	7589.51604	35.82929	21.02359	1.70424	.193

EFFECT.. FEM (Cont.)
Univariate F-tests with (1,361) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
ESTEEM	107.44454	3883.60490	107.44454	10.75791	9.98749	.002
INTEXT	7494.31182	464051.295	7494.31182	1285.46065	5.83006	.016
NEUROTIC	26.81431	7589.51604	26.81431	21.02359	1.27544	.259

ESTEEM, INTEXT, and NEUROTIC. ESTEEM is evaluated after adjustment only for the three covariates. INTEXT is adjusted for effects of ESTEEM and the three covariates; NEUROTIC is adjusted for ESTEEM and INTEXT and the three covariates. In effect, then, INTEXT is adjusted for four covariates and NEUROTIC is adjusted for five.

Stepdown analysis for the interaction and two main effects is in Table 7.28. The results are the same as those in MANOVA except that there is no longer a main effect of FEM on INTEXT after adjustment for four covariates. The relationship between FEM and INTEXT is already represented by the relationship between FEM and ESTEEM. Consolidation of information from Tables 7.27 and 7.28, as well as some information from Table 7.26, appears in Table 7.29, along with apportionment of the .05 alpha error to the various tests and effect sizes with their confidence limits based on the α error chosen.

For the DVs associated with significant main effects, interpretation requires associated marginal means. Table 7.30 contains syntax and adjusted marginal means for ESTEEM and for INTEXT (which is adjusted for ESTEEM as well as covariates) for FEM and for MASC adjusted for FEM. Syntax and marginal means for the main effect of FEM on INTEXT (univariate but not stepdown effect) appear in Table 7.31.

Effect sizes and their confidence limits for stepdown effects are found through Smithson's (2003) program as for MANOVA. Table 7.32 shows the input/output for that analysis using values from Table 7.28. Values chosen for confidence limits reflect apportionment of α. A checklist for MANCOVA appears in Table 7.33. An example of a Results section, as might be appropriate for journal presentation, follows.

TABLE 7.28 Stepdown Analyses of Three Ordered DVs Adjusted for Three Covariates for (Top to Bottom) FEM by MASC Interaction, Masculinity, and Femininity (Selected Output from SPSS MANOVA—see Table 7.25 for Syntax)

Roy-Bargman Stepdown F-tests

Variable	Hypoth. MS	Error MS	StepDown F	Hypoth. DF	Error DF	Sig. of F
ESTEEM	7.21931	10.75791	.67107	1	361	.413
INTEXT	.35520	12.60679	.02817	1	360	.867
NEUROTIC	4.94321	19.72942	.25055	1	359	.617

Roy-Bargman Stepdown F-tests

Variable	Hypoth. MS	Error MS	StepDown F	Hypoth. DF	Error DF	Sig. of F
ESTEEM	533.61774	10.75791	49.60237	1	361	.000
INTEXT	137.74436	12.60679	10.92621	1	360	.001
NEUROTIC	1.07421	19.72942	.05445	1	359	.816

Roy-Bargman Stepdown F-tests

Variable	Hypoth. MS	Error MS	StepDown F	Hypoth. DF	Error DF	Sig. of F
ESTEEM	107.44454	10.75791	9.98749	1	361	.002
INTEXT	47.36159	12.60679	3.75683	1	360	.053
NEUROTIC	4.23502	19.72942	.21466	1	359	.643

TABLE 7.29 Tests of Covariates, Femininity, Masculinity (Adjusted for Femininity), and Interaction

IV	DV	Univariate *F*	df	Stepdown *F*	df	*a*	Partial η^2	CL around Partial η^2 per *a* Lower	Upper
Covariates	ESTEEM	20.48[a]	3/361	20.48**	3/361	.02	.15	.07	.22
	INTEXT	1.13	3/361	0.50	3/360	.02	.00	.00	.02
	NEUROTIC	21.95[a]	3/361	12.16**	3/359	.01	.09	.03	.17
Femininity	ESTEEM	9.99[a]	1/361	9.99**	1/361	.02	.03	.00	.08
	INTEXT	5.83[a]	1/361	3.76	1/360	.02	.01	.00	.05
	NEUROTIC	1.28	1/361	0.21	1/359	.01	.00	.00	.02
Masculinity	ESTEEM	49.60[a]	1/361	49.60**	1/361	.02	.12	.06	.20
	INTEXT	20.57[a]	1/361	10.93**	1/360	.02	.03	.00	.08
	NEUROTIC	1.70	1/361	0.05	1/359	.01	.00	.00	.01
Femininity by masculinity interaction	ESTEEM	0.67	1/361	0.67	1/361	.02	.00	.00	.03
	INTEXT	0.00	1/361	0.03	1/360	.01	.00	.00	.00
	NEUROTIC	0.07	1/361	0.25	1/359	.01	.00	.00	.03

[a]Significance level cannot be evaluated but would reach $p < .02$ in univariate context.

**$p < .01$.

TABLE 7.30 Adjusted Marginal Means for Esteem Adjusted for Three Covariates and INTEXT Adjusted for ESTEEM Plus Three Covariates (Syntax and Selected Output from SPSS MANOVA)

```
MANOVA      ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM,MASC(1,2)
    /PRINT=PARAMETERS(ESTIM)
    /ANALYSIS=ESTEEM WITH CONTROL,ATTROLE,SEL2
    /DESIGN=CONSPLUS FEM /DESIGN=FEM,CONSPLUS MASC
    /ANALYSIS=INTEXT WITH CONTROL,ATTROLE,SEL2,ESTEEM
    /DESIGN=FEM, CONSPLUS,MASC.

Estimates for ESTEEM adjusted for 3 covariates
--- Individual univariate .9500 confidence intervals

CONSPLUS FEM

 Parameter      Coeff.   Std. Err.     t-Value    Sig. t Lower -95%  CL- Upper

       1   16.7151721      .34268    48.77769    .00000    16.04128   17.38906
       2   15.3543164      .21743    70.61815    .00000    14.92674   15.78189

Estimates for ESTEEM adjusted for 3 covariates
--- Individual univariate .9500 confidence intervals

CONSPLUS MASC

 Parameter      Coeff.   Std. Err.     t-Value    Sig. t Lower -95%  CL- Upper

       2   16.9175545      .22684    74.57875    .00000    16.47146   17.36365
       3   14.2243700      .31940    44.53415    .00000    13.59625   14.85249
```

(continued)

TABLE 7.30 Continued

```
Estimates for INTEXT adjusted for 4 covariates
--- Individual univariate .9500 confidence intervals

CONSPLUS MASC

 Parameter      Coeff.  Std. Err.     t-Value    Sig. t Lower -95%  CL- Upper

       2   11.0058841      .25416    43.30276    .00000    10.50606   11.50571
       3   12.4718467      .35617    35.01620    .00000    11.77141   13.17228
```

Note: Coeff. = adjusted marginal mean; first parameter = low, second parameter = high.

TABLE 7.31 Marginal Means for INTEXT Adjusted for Three Covariates Only (Syntax and Selected Output from SPSS MANOVA)

```
MANOVA ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM,MASC(1,2)
        /PRINT=PARAMETERS(ESTIM)
        /ANALYSIS=INTEXT WITH CONTROL,ATTROLE,SEL2, ESTEEM
        /DESIGN=CONSPLUS FEM.

Estimates for INTEXT adjusted for 4 covariates
--- Individual univariate .9500 confidence intervals

CONSPLUS FEM

 Parameter      Coeff.  Std. Err.     t-Value    Sig. t Lower -95%  CL- Upper

       1   11.1014711      .35676    31.11753    .00000    10.39989   11.80305
       2   11.9085923      .22495    52.93984    .00000    11.46623   12.35096
```

Note: Coeff. = adjusted marginal mean; first parameter = low, second parameter = high.

TABLE 7.32 Data Set Output from NoncF3.sps for Effect Size (r2) with 95% Confidence Limits.

	fval	df1	df2	conf	lc2	ucdf	uc2	lcdf	power	r2	lr2	ur2
1	20.4800	3	361	.980	27.6000	.9900	102.9400	.0100	1.0000	.15	.07	.22
2	.5000	3	360	.980	.0000	.3175	9.1211	.0100	.0485	.00	.00	.02
3	12.1600	3	359	.990	10.0225	.9950	72.9125	.0050	.9960	.09	.03	.17
4	9.9900	1	361	.980	.6643	.9900	30.2627	.0100	.7187	.03	.00	.08
5	3.7600	1	360	.980	.0000	.9467	18.2345	.0100	.2610	.01	.00	.05
6	.2100	1	359	.990	.0000	.3530	9.0300	.0052	.0099	.00	.00	.02
7	49.6000	1	361	.980	21.4579	.9900	89.1734	.0100	1.0000	.12	.06	.20
8	10.9300	1	360	.980	.9188	.9900	31.8934	.0100	.7655	.03	.00	.08
9	.0500	1	359	.999	.0000	.1768	2.1500	.0614	.0007	.00	.00	.01
10	.6700	1	361	.980	.0000	.5864	9.8825	.0100	.0397	.00	.00	.03
11	.0300	1	360	.980	.0000	.1374	1.2900	.0726	.0111	.00	.00	.00
12	.2500	1	359	.999	.0000	.3826	10.7500	.0027	.0015	.00	.00	.03

TABLE 7.33 Checklist for Multivariate Analysis of Covariance

1. Issues
   a. Unequal sample sizes and missing data
   b. Normality of sampling distributions
   c. Outliers
   d. Homogeneity of variance-covariance matrices
   e. Linearity
   f. Homogeneity of regression
      (1) Covariates
      (2) DVs for stepdown analysis
   g. Reliability of covariates (and DVs for stepdown)
   h. Multicollinearity and singularity
2. Major analyses: Planned comparisons or omnibus *F;* when significant: Importance of DVs
   a. Within-cell correlations, stepdown *F,* univariate *F*
   b. Effect size with its confidence interval for significant stepdown *F*
   c. Adjusted marginal and/or cell means for significant *F,* and standard deviations or standard errors or confidence intervals
3. Multivariate effect size(s) with confidence interval(s) for planned comparisons or omnibus *F.*
4. Additional analyses
   a. Assessment of covariates
   b. Interpretation of IV-covariates interaction (if homogeneity of regression violated for stepdown analysis)
   c. Post hoc comparisons

Results

SPSS MANOVA was used for the analyses with the sequential adjustment for nonorthogonality. Order of entry of IVs was femininity, then masculinity. Total $N = 369$ was reduced to 368 with the deletion of a case missing a score on locus of control. There were no univariate or multivariate within-cell outliers at $\alpha = .001$. Results of evaluation of assumptions of normality, homogeneity of variance-covariance matrices, linearity, and multicollinearity were satisfactory. Covariates were judged to be adequately reliable for covariance analysis.

A 2 × 2 between-subjects multivariate analysis of covariance was performed on three dependent variables associated with personality of respondents: self-esteem, introversion-extraversion, and neuroticism. Adjustment was made for three covariates: attitude toward role of women, locus of control, and socioeconomic status. Independent variables were masculinity (high and low) and femininity (high and low).

With the use of Wilks' criterion, the combined DVs were significantly related to the combined covariates, approximate $F(9, 873) = 10.86$, $p < .01$, to femininity, $F(3, 359) = 4.67$, $p < .01$, and to masculinity, $F(3, 359) = 20.59$, $p < .001$ but not to the interaction, $F(3, 359) = 0.31$, $p > .05$. There was a modest association between DVs and covariates, partial $\eta^2 = .10$ with confidence limits from .06 to .29. A somewhat larger association was found between combined DVs and the main effect of masculinity, $\eta^2 = .15$ with confidence limits from .08 to .21, but the association between the main effect of femininity and the combined DVs was smaller, $\eta^2 = .04$ with confidence limits from .00 to .08. Effect size for the nonsignificant interaction was .00 with confidence limits from .00 to .01. [*F is from Table 7.25; partial* η^2 *and their confidence limits are found through Smithson's NoncF3.sps for main effects, interaction, and covariates.*]

To investigate more specifically the power of the covariates to adjust dependent variables, multiple regressions were run for each DV in turn, with covariates acting as multiple predictors. Two of the three covariates, locus of control and attitudes toward women's role, provided significant adjustment to self-esteem. The *B* value of .98 (confidence interval from .71 to 1.25) for locus of control was significantly different from zero, $t(361) = 7.21$, $p < .001$, as was the *B* value of .09 (confidence interval from .03 to .14) for attitudes

toward women's role, $t(361) = 3.21$, $p < .01$. None of the covariates provided adjustment to the introversion-extraversion scale. For neuroticism, only locus of control reached statistical significance, with $B = 1.53$ (confidence interval from 1.16 to 1.91), $t(361) = 8.04$, $p < .001$. For none of the DVs did socioeconomic status provide significant adjustment. [*Information about relationships for individual DVs and CVs is from Table 7.26.*]

Effects of masculinity and femininity on the DVs after adjustment for covariates were investigated in univariate and Roy-Bargmann stepdown analysis, in which self-esteem was given the highest priority, introversion-extraversion second priority (so that adjustment was made for self-esteem as well as for the three covariates), and neuroticism third priority (so that adjustment was made for self-esteem and introversion-extraversion as well as for the three covariates). Homogeneity of regression was satisfactory for this analysis, and DVs were judged to be sufficiently reliable to act as covariates. Results of this analysis are summarized in Table 7.29. An experimentwise error rate of 5% for each effect was achieved by apportioning alpha according to the values shown in the last column of the table.

After adjusting for differences on the covariates, self-esteem made a significant contribution to the composite of the DVs that best distinguishes between women who were high or low in femininity, stepdown $F(1, 361) = 9.99$, $p < .01$, $\eta^2 = .03$ with confidence limits from .00 to .08. With self-esteem scored inversely, women with higher femininity scores showed greater self-esteem after adjustment for covariates (adjusted mean self-esteem = 15.35, SE = 0.22) than those scoring lower on femininity (adjusted mean self-esteem = 16.72, SE = 0.34). Univariate analysis revealed that a statistically significant difference was also present on the introversion-extraversion measure,

with higher-femininity women more extraverted, univariate $F(1, 361) = 5.83$, a difference already accounted for by covariates and the higher-priority DV. [*Adjusted means are from Tables 7.30 and 7.31;* partial η^2 *and confidence limits are from Table 7.32.*]

Lower- versus higher-masculinity women differed in self-esteem, the highest-priority DV, after adjustment for covariates, stepdown $F(1, 361) = 49.60$, $p < .01$, $\eta^2 = .12$ with confidence limits from .06 to .20. Greater self-esteem was found among higher-masculinity women (aPjusted mean = 14.22, SE = 0.32) than among lower-masculinity women (adjusted mean = 16.92, SE = 0.23). The measure of introversion and extraversion, adjusted for covariates and self-esteem, was also related to differences in masculinity, stepdown $F(1, 360) = 10.93$, $p < .01$, $\eta^2 = .03$ with confidence limits from .00 to .08. Women scoring higher on the masculinity scale were more extraverted (adjusted mean extraversion 12.47, SE = 0.36) than those showing lower masculinity (adjusted mean extraversion = 11.01, SE = 0.25).

High-masculinity women, then, are characterized by greater self-esteem and extraversion than low-masculinity women when adjustments are made for differences in socioeconomic status, attitudes toward women's role, and locus of control. High-femininity women show greater self-esteem than low-femininity women with adjustment for those covariates.

Pooled within-cell correlations among dependent variables and covariates are shown in Table 7.19. The only relationships accounting for more than 10% of variance are between self-esteem and neuroticism ($r = .36$), locus of control and self-esteem ($r = .35$), and between neuroticism and locus of control ($r = .39$). Women who are high in neuroticism tend to have lower self-esteem and are more likely to attribute reinforcements to external sources.

7.7 Comparison of Programs

SPSS, SAS, and SYSTAT all have highly flexible and full-featured MANOVA programs, as seen in Table 7.34. One-way between-subjects MANOVA is also available through discriminant function programs, as discussed in Chapter 9.

7.7.1 SPSS Package

SPSS has two programs, MANOVA (available only through syntax) and GLM. Features of the two programs are quite different, so that you may want to use both programs for an analysis.

Both programs offer several methods of adjustment for unequal n and several statistical criteria for multivariate effects. In repeated-measures designs, the sphericity test offered by both programs evaluates the sphericity assumption; if the assumption is rejected (that is, if the test is significant), one of the alternatives to repeated-measures ANOVA—MANOVA, for instance—is appropriate. There are also the Greenhouse-Geisser, Huynh-Feldt and lower-bound epsilons for adjustment of df for sphericity. SPSS MANOVA and GLM do the adjustment and provide significance levels for the effects with adjusted df.

SPSS MANOVA has several features that make it superior to any of the other programs reviewed here. It is the only program that performs Roy-Bargmann stepdown analysis as an option (Section 7.5.3.2). Use of other programs requires a separate ANCOVA run for each DV after the one of highest priority. SPSS MANOVA also is the only program that has special syntax for pooling covariates to test homogeneity of regression for MANCOVA and stepdown analysis (Section 7.6.1.6). If the assumption is violated, the manuals describe procedures for ANCOVA with separate regression estimates, if that is your choice. Full simple effects analyses are easily specified using the MWITHIN instruction (Section 8.5.2). SPSS MANOVA also is easier to use for user-specified comparisons. Bivariate collinearity and homogeneity of variance-covariance matrices are readily tested in SPSS MANOVA through within-cell correlations and homogeneity of dispersion matrices, respectively. Multicollinearity is assessed through the determinant of the within-cells correlation matrix (cf. Section 4.1.7).

Both programs provide complete descriptive statistics for unadjusted means and standard deviations, however, adjusted means for marginal and cell effects are more easily specified in SPSS GLM through the EMMEANS instruction. SPSS MANOVA provides adjusted cell means easily, but marginal means require rather convoluted CONSPLUS instructions, as seen in Section 7.6. SPSS GLM provides leverage values (that are easily converted to Mahalanobis distance) to assess multivariate outliers.

For between-subjects designs, both programs offer Bartlett's test of sphericity, which tests the null hypothesis that correlations among DVs are zero; if they are, univariate F (with Bonferroni adjustment) is used instead of stepdown F to test the importance of DVs (Section 7.5.3.1).

A principal components analysis can be performed on the DVs through SPSS MANOVA, as described in the manuals. In the case of multicollinearity or singularity among DVs (see Chapter 4), principal components analysis can be used to produce composite variables that are orthogonal to one another. However, the program still performs MANOVA on the raw DV scores, not the component scores. If MANOVA for component scores is desired, use the results of PCA and the COMPUTE facility to generate component scores for use as DVs.

TABLE 7.34 Comparison of Programs for Multivariate Analysis of Variance and Covariance[a]

Feature	SPSS GLM	SPSS MANOVA	SAS GLM	SYSTAT ANOVA, GLM, and MANOVA[g]
Input				
Variety of strategies for unequal *n*	Yes	Yes	Yes	Yes
Specify tolerance	EPS	No	SINGULAR	Yes
Specify exact tests for multivariate effects	No	No	MSTAT= EXACT	No
Output				
Standard source table for Wilks' lambda	No	PRINT = SIGNIF (BRIEF)	No	No
Cell covariance matrices	No	Yes	No	No
Cell covariance matrix determinants	No	Yes	No	No
Cell correlation matrices	No	Yes	No	No
Cell SSCP matrices	No	Yes	No	No
Cell SSCP determinants	No	Yes	No	No
Unadjusted marginal means for factorial design	Yes	Yes	Yes	No
Unadjusted cell means	Yes	Yes	Yes	No
Unadjusted cell standard deviations	Yes	Yes	Yes	No
Confidence interval around unadjusted cell means	No	Yes	No	No
Adjusted cell means	EMMEANS	PMEANS	LSMEANS	PRINT MEDIUM
Standard errors for adjusted cell means	EMMEANS	No	LSMEANS	PRINT MEDIUM
Adjusted marginal means	EMMEANS	Yes[b]	LSMEANS	No
Standard errors for adjusted marginal means	EMMEANS	No	LSMEANS	No
Wilks' lambda with approximate *F* statistic	Yes	Yes	Yes	Yes
Criteria other than Wilks'	Yes	Yes	Yes	Yes
Multivariate influence/leverage statistics by cell	No	No	No	Data file
Canonical (discriminant function) statistics[c]	No	Yes	Yes	Yes
Univariate *F* tests	Yes	Yes	Yes	Yes
Averaged univariate *F* tests	No	Yes	No	No
Stepdown *F* tests (DVs)	No	Yes	No	No
Sphericity test	Yes	Yes	No	Yes[h]
Adjustment for failure of sphericity	Yes	Yes	Yes	No
Tests for univariate homogeneity of variance	Yes	Yes	Yes[c]	No
Test for homogeneity of covariance matrices	Box's *M*	Box's *M*	No	No

TABLE 7.34 Continued

Feature	SPSS GLM	SPSS MANOVA	SAS GLM	SYSTAT ANOVA, GLM, and MANOVA[g]
Output *(continued)*				
Principal component analysis of residuals	No	Yes	No	Yes
Hypothesis SSCP matrices	Yes	No	Yes	Yes
Hypothesis covariance matrices	No	Yes	No	No
Inverse of hypothesis SSCP matrices	No	No	Yes	Yes
Pooled within-cell error SSCP matrix	Yes	Yes	Yes	Yes
Pooled within-cell covariance matrix	No	Yes	No	Yes
Pooled within-cell correlation matrix	No	Yes	Yes	Yes
Total SSCP matrix	No	No	No	Yes
Determinants of pooled within-cell correlation matrix	No	Yes	No	No
Covariance matrix for adjusted cell means	No	No	Data file	No
Effect size for univariate tests	ETASQ	POWER	No	No
Power analysis	OPOWER	POWER	No[f]	No
SMCs with effects for each DV	No	No	No	Yes
Confidence intervals for multivariate tests	No	Yes	No	No
Post hoc tests with adjustment	POSTHOC	Yes[d]	Yes	Yes
Specific comparisons	Yes	Yes	Yes	Yes
Tests of simple effects (complete)	No	Yes	No	No
Homogeneity of regression	No	Yes	No	Yes
ANCOVA with separate regression estimates	No	Yes	No	No
Regression coefficient for each covariate	PARAMETER	Yes	No	PRINT LONG
Regression coefficient for each cell	Yes	No	Yes	No
R^2 for model	Yes	No	Yes	No
Coefficient of variation	No	No	Yes	No
Normalized plots for each DV and covariate	No	Yes	No	No
Predicted values and residuals for each case	Data file	Yes	Yes	Data file
Confidence limits for predicted values	No	No	Yes	No
Residuals plots	No	Yes	No	No

[a]Additional features are discussed in Chapter 6 (ANCOVA).
[b]Available through CONSPLUS procedure, see Section 7.6.
[c]Discussed more fully in Chapter 9.
[d]Bonferroni and Scheffé confidence intervals.
[e]One-way design only.
[f]Available in a separate program: GLMPOWER.
[g]MANOVA, added to SYSTAT in Version 11, differs from GLM only in its menu access.
[h]Not available in "long" output.

7.7.2 SAS System

MANOVA in SAS is done through the PROC GLM. This general linear model program, like SPSS GLM, has great flexibility in testing models and specific comparisons. Four types of adjustments for unequal-*n* are available, called `TYPE I` through `TYPE IV` estimable functions (cf. 6.5.4.2); this program is considered by some to have provided the archetypes of the choices available for unequal-*n* adjustment. Adjusted cell and marginal means are printed out with the `LSMEANS` instruction. SAS tests multivariate outliers by adding leverage values (which may be converted to Mahalanobis distance) to the data set (cf. Section 6.6.1.4). Exact tests of multivariate effects may be requested in place of the usual *F* approximation.

SAS GLM provides Greenhouse-Geisser and Huynh-Feldt adjustments to degrees of freedom and significance tests for effects using adjusted df. There is no explicit test for homogeneity of regression, but because this program can be used for any form of multiple regression, the assumption can be tested as a regression problem where the interaction between the covariate(s) and IV(s) is an explicit term in the regression equation (Section 6.5.3).

There is abundant information about residuals, as expected from a program that can be used for multiple regression. Should you want to plot residuals, however, a run through the PLOT procedure is required. As with most SAS programs, the output requires a fair amount of effort to decode until you become accustomed to the style.

7.7.3 SYSTAT System

In SYSTAT, the GLM, ANOVA, and MANOVA programs may be used for simple, fully factorial MANOVA, however GLM and MANOVA are recommended for more complex designs for their numerous features and flexibility, and because they are not much more difficult to set up.

Model 1 adjustment for unequal *n* is provided by default, along with a strong argument as to its benefits. Other options are available, however, by specification of error terms or a series of sequential regression analyses. Several criteria are provided for tests of multivariate hypotheses, along with a great deal of flexibility in specifying these hypotheses. Leverage values are saved in a data set by request, and may be converted to Mahalanobis distance as per Equation 4.3 to assess multivariate outliers.

The program provides cell least squares means and their standard errors, adjusted for covariates, if any. Other univariate statistics are not provided in the program, but they can be obtained through the STATS module.

Like SPSS MANOVA, principal components analysis can be done on the pooled within-cell correlation matrix. But also like the SPSS program, the MANOVA is performed on the original scores.