USING MULTIVARIATE STATISTICS

SECOND EDITION

Barbara G. Tabachnick
Linda S. Fidell

California State University, Northridge

HarperCollins*Publishers*

Contents

Chapter 9 Multivariate Analysis of Variance and Covariance 371

Preface

This second edition differs from the first in several respects. The first edition illustrated multivariate analyses through BMDP and SPSS. As a result of grumbling from our East Coast colleagues (and the thought of more potential sales), we added SAS to this edition. We also added SYSTAT to illustrate the analyses through another flexible and powerful package currently available for microcomputers. We ran the examples on the most current versions of each of the packages so the output and options shown are those available as of spring 1988.

There are two new chapters in this new edition, on multiway frequency analysis and profile analysis of repeated measures. Both of these techniques are growing in popularity and use. Their inclusion should give you even greater flexibility in your analysis of data. The chapter on screening data prior to analysis (Chapter 4) has been completely rewritten and now includes comprehensive computer output along with flow diagrams for screening both grouped and ungrouped data.

Our students continue to be our best teachers of what needs to be said about these statistical techniques. After teaching several more semesters with the first edition as text, we learned where changes needed to be made and what changes were likely to work better. As a result, the narrative has been completely rewritten, first sentence to last. Our goal, as with the first edition, is clarity of presentation. We try to tell you what your options are when you do an analysis, how to set up the analysis, and how to interpret and write up your results. We do not explain in gruesome detail the mathematical origins of the procedure, or, for the most part, differences of opinion among those who are developing the analyses. Where such differences of opinion exist, we have chosen the alternative that we think most likely to lead you to the safest and best analysis of your data. The battle over alternatives is being fought out in the journals, not here.

Because of this orientation, some have accused this book of being a cookbook,

to which we plead both guilty and not guilty. "Guilty" because we do not focus on the mathematical underpinning of the procedures; "not guilty" because the fourth section of Chapters 5 through 12 has a small example worked out by hand that ties the calculations into those for univariate statistics, where possible. For those who want to understand how the analysis develops, the relevant material is in those sections plus the appendix on matrix algebra. Just as the chapters can be read in any order, or excluded, so also the material in the fourth sections and appendix can be used or omitted, at the interest and discretion of the reader.

Most people do not have to understand the intricacies of the internal-combustion engine to drive to work, and we believe that most people do not have to understand the subtleties of matrix algebra and calculus to analyze their data. In both cases it is nicer and safer if you do thoroughly understand the material, but in both cases there are also signals that something might be amiss that, if attended, help prevent disaster. Look to your dashboard, and Chapter 4 as well as Section 3 of Chapters 5 through 12, respectively, for those signals.

For those who would like to play with the setups and data for the large examples in their own computing facilities, we have developed a floppy diskette, available from the authors upon request (with a small donation for the floppy and mailing). Contact us—we'd like to hear from you anyway.

Many people have contributed to the production of the second edition: buyers of the first edition who created the demand, the staff at Harper & Row—especially Ellen MacElree—who were patient and helpful, Barbara Nicholson and Kenya Moore who typed the first edition onto floppies so that we could edit it, and the following reviewers who provided very helpful reviews of both the second edition and the new chapters: Dale Berger, Claremont Graduate School; Robert Gardiner, University of Western Ontario; Lawrence Gordon, University of Vermont; Mark Leary, Wake Forest University; Henry Morlock, SUNY-Plattsburgh; Thomas Nygren, Ohio State University; and Paul Spector, University of South Florida. We especially want to thank Curt Dommeyer, William Vincent, and Jim Dole, who provided a colleague's-eye view of the first edition, and Betty Rose and David Barber, who gave us a very helpful student's-eye view.

And very most especially we want to thank our husbands (and, in one case, children) who still haven't divorced us!

Barbara G. Tabachnick
Linda S. Fidell

Chapter 5

Multiple Regression

5.1 GENERAL PURPOSE AND DESCRIPTION

Regression analyses are a set of statistical techniques that allow one to assess the relationship between one DV and several IVs. For example, can reading ability in primary grades (the DV) be predicted from several IVs such as gender and preschool measures of perceptual and motor development? The terms regression and correlation are used more or less interchangeably to label these procedures although the term regression is often used when the intent of the analysis is prediction, and the term correlation is used when the intent is simply to measure the degree of association between the DV and the IVs.

Regression techniques can be applied to a data set in which the IVs are correlated with one another and with the DV to varying degrees. One can, for instance, assess the relationship between IVs such as education, income, and socioeconomic status and a DV such as occupational prestige. Because regression techniques can be used when the IVs are correlated, they are helpful both in experimental research (when, for instance, correlation among IVs is created by unequal numbers of cases in cells) and in observational or survey research where nature has "manipulated" correlated variables. The flexibility of regression techniques is, then, of special importance to the researcher who is interested in real-world or very complicated problems that cannot be meaningfully reduced to orthogonal designs in a laboratory setting.

Multiple regression is an extension of bivariate regression (see Chapter 3) in which several IVs instead of just one are combined to predict a value on a DV for each subject. The result of regression is an equation that represents the best prediction of a DV from several continuous or dichotomous IVs. The regression equation takes the following form:

$$Y' = A + B_1X_1 + B_2X_2 + \cdots + B_kX_k$$

where Y' is the predicted value on the DV, A is the Y intercept (the value of Y when all the X values are zero), the X's represent the various IVs (of which there are k), and the Bs are the coefficients assigned to each of the IVs during regression. Although the intercept and the coefficients are the same for a whole sample, a different Y' value is predicted for each subject as a result of inserting the subject's own X values into the equation.

The goal of regression is to arrive at the set of B values, called regression coefficients, for the IVs that bring the Y values predicted from the equation as close as possible to the Y values obtained by measurement. The regression coefficients that are computed accomplish two intuitively appealing and highly desirable goals: they minimize (the sum of the squared) deviations between predicted and obtained Y values and they optimize the correlation between the predicted and obtained Y values for the data set. In fact, one of the important statistics derived from a regression analysis is the multiple correlation coefficient, the Pearson product moment correlation coefficient between the obtained and predicted Y values (see Section 5.4.1).

Regression techniques consist of standard multiple regression, hierarchical regression, and statistical (stepwise) or setwise regression. Differences between these techniques involve the way variables enter the equation. What happens to variance shared by variables and who determines the order in which variables enter the equation?

5.2 KINDS OF RESEARCH QUESTIONS

The primary goal of regression analysis is usually to investigate the relationship between a DV and several IVs. As a preliminary step, one determines how strong the relationship is between DV and IVs; then, with some ambiguity, one assesses the importance of each of the IVs to the relationship.

A more complicated goal might be to investigate the relationship between a DV and some IVs with the effect of other IVs statistically eliminated. Researchers often use regression to perform what is essentially a covariates analysis in which they ask if some critical variable (or variables) adds anything to a prediction equation for a DV after other IVs—the covariates—have already entered the equation. For example, does gender add to prediction of mathematical performance after extent and type of mathematical training have been accounted for?

Another strategy is to compare the ability of several competing sets of IVs to predict a DV. Can use of Valium be predicted better by a set of health variables or by a set of attitudinal variables?

All too often, regression is used to find the best prediction equation for some phenomenon regardless of the meaning of the equation, a goal met by statistical (stepwise) regression. In statistical regression, of which there are several varieties, statistical criteria alone, computed from a single sample, determine which IVs enter the equation and the order in which they enter.

Regression analyses can be used with either continuous or dichotomous IVs. But a variable that is initially discrete can also be used if it is converted into a set

of dichotomous variables by dummy variable coding. For example, consider an initially discrete religious affiliation variable on which 1 stands for Protestant, 2 for Catholic, 3 for Jewish, and 4 for none or other. The variable may be converted into a set of three new IVs (Protestant vs. non-Protestant, Catholic vs. non-Catholic, Jewish vs. non-Jewish), one IV for each degree of freedom. When the new set is entered as a group, the variance due to the original discrete variable is analyzed, and, in addition, one can examine effects of the newly created dichotomous components. Dummy variable coding is covered in glorious detail by Cohen and Cohen (1975, pp. 173–188).

ANOVA (Chapter 3) is a special case of regression where main effects and interactions have been dummy variable coded. ANOVA problems can be handled through multiple regression, but multiple regression problems often cannot readily be converted into ANOVA because of correlations among IVs and the presence of continuous IVs. If analyzed through ANOVA, continuous IVs have to be rendered discrete (e.g., high, medium, and low), a process that often results in loss of information. In regression, the full range of continuous IVs is maintained.

As a statistical tool, regression is very helpful in answering a number of practical questions, as discussed in Sections 5.2.1 through 5.2.8.

5.2.1 Degree of Relationship

How good is the regression equation? Does the regression equation really provide better-than-chance prediction? Is the multiple correlation really any different from zero when allowances for naturally occurring fluctuations in such correlations are made? For example, can one reliably predict reading ability given knowledge of perceptual development, motor development, and gender? The statistical procedures described in Section 5.6.2.1 allow you to determine if your multiple correlation is reliably different from zero.

5.2.2 Importance of IVs

If the multiple correlation is different from zero, you may want to ask which of the IVs is important in the equation and which of the IVs can be deleted. For example, is knowledge of motor development helpful in predicting reading ability, or can we do just as well with knowledge of only gender and perceptual development? The methods in Section 5.6.1 help you to evaluate the relative importance of various IVs to a regression solution.

5.2.3 Adding IVs

Suppose that you have just computed a regression equation and you want to know whether or not you can improve your prediction of the DV by addition of one or more IVs to the equation. For example, is prediction of a child's reading ability enhanced by adding a variable reflecting parental interest in reading to the three IVs already included in the equation? A test for improvement of the multiple correlation after

addition of one new variable is given in Section 5.6.1.2 and for improvement after addition of several new variables in Section 5.6.2.3.

5.2.4 Changing IVs

Although the regression equation is a linear equation (that is, it does not contain squared values, cubed values, cross products of variables, and the like), the researcher may include nonlinear relationships in the analysis by redefining IVs. Curvilinear relationships can be made available for analysis by squaring or raising to higher powers original IVs. Interaction can be made available for analysis by creating a new IV that is a cross product of two or more original IVs.

For an example of a curvilinear relationship, suppose a child's reading ability increases with increasing parental interest up to a point, and then levels off. Greater parental interest does not result in greater reading ability. If the square of parental interest were added as an IV, better prediction of a child's reading ability could be achieved.

Inspection of a scatterplot between predicted and obtained Y values (known as residuals analysis—see Section 5.3.2.4) may reveal that the relationship between the DV and the IVs is not exclusively linear but has more complicated components such as curvilinearity. To improve prediction, one may have to include new IVs as described above.

There is danger, however, in too liberal use of powers or cross products of IVs; the sample data may be overfit to the extent that results no longer generalize to a population. Procedures for using regression for nonlinear curve fitting are discussed in Cohen and Cohen (1975) and McNeil, Kelly, and McNeil (1975).

5.2.5 Contingencies among IVs

You may be interested in the way that one IV behaves in the context of one, or a set, of the other IVs. Hierarchical regression can be used to hold the effects of several IVs statistically "constant" while examining the relationship between an especially interesting IV and the DV. For example, after adjustment for differences in perceptual and motor development, does gender predict reading ability? This procedure is described in Section 5.5.2.

5.2.6 Comparing Sets of IVs

Is prediction of a DV from one set of IVs better than prediction from another set of IVs? For example, is prediction of reading ability based on perceptual and motor development and gender as good as prediction from family income and parental educational attainments? Section 5.6.2.5 demonstrates a method for comparing the solutions given by two sets of predictors.

5.2.7 Predicting DV Scores for Members of a New Sample

One of the more important applications of regression involves predicting scores on a DV for subjects for whom only data on IVs are available. This application is fairly frequent in personnel selection for employment or graduate training and the like. Over a fairly long period, a researcher collects data on a DV, say, success in graduate school, and on several IVs, say, undergraduate GPA, GRE verbal scores, and GRE math scores. If the IVs are strongly related to the DV, then for a new sample of applicants to graduate school, regression coefficients are applied to IV scores to predict success in graduate school ahead of time. Admission to the graduate school may, in fact, be based on prediction of success from regression.

The generalizability of a regression solution to a new sample may also be checked within a single large sample by a procedure called cross validation. A regression equation is developed from a portion of a sample and then applied to the other portion of the sample. If the solution generalizes, the regression equation predicts DV scores better than chance for the new cases, as well.

5.2.8 Causal Modeling

Path analysis, or causal analysis, is a special application of regression analysis in which questions about the minimum number of relationships (causal influences) and their directions are asked by observing the sizes of regression coefficients with and without certain variables entered into equations. For example, one can investigate the direct and indirect influence of intelligence on reading ability in the context of perceptual and motor development and gender. We have not included a description of path analysis in this book but refer the interested reader to Asher (1976) or Heise (1975).

An increasingly popular set of new techniques, called structural analyses (Bentler, 1980), involves path analysis of a set of IVs and DVs as latent variables (factors) all considered simultaneously. Analysis is facilitated by computer programs such as LISREL (Joreskog and Sorbom, 1978) and EQS (Bentler, 1985). Although the method is not included in this book, you should be aware that many researchers are finding this combination of canonical (Chapter 6) and factor analysis (Chapter 12) useful in attempting causal inference from nonexperimental or quasi-experimental research.

5.3 LIMITATIONS TO REGRESSION ANALYSES

5.3.1 Theoretical Issues

Regression analyses reveal relationships among variables but do not imply that the relationships are causal. Demonstration of causality is a logical and experimental, rather than statistical, problem. An apparently strong relationship between variables could stem from many sources, including the influence of other, currently unmeasured

variables. One can make an airtight case for causal relationship among variables only by showing that manipulation of some of them is followed inexorably by change in others when all other variables are controlled.

Another problem for logic rather than statistics is inclusion of variables. Which IV should be used and how it is to be measured? Which IVs should be examined and how are they to be measured? If one already has some IVs in an equation, which IVs should be added to the equation for the most improvement in prediction? The answers to these questions can be provided by theory, astute observation, or good hunches, but they will not be provided by statistics.

There are, however, some general considerations for choosing IVs. Regression will be best when each IV is strongly correlated with the DV but uncorrelated with other IVs. A general goal of regression, then, might be to select the fewest IVs necessary to provide good prediction of a DV where each IV predicts a substantial and independent segment of the variability in the DV.

There are other considerations to selection of variables. If the goal of research is manipulation of some DV (say body weight), it is strategic to use as IVs those variables that can be manipulated (e.g., caloric intake, physical activity) rather than those that cannot (e.g., genetic predisposition). Or, if one is interested in predicting a variable such as annoyance caused by noise for a neighborhood, it is strategic to use cheaply obtained sets of IVs (e.g., neighborhood characteristics published by the Census Bureau) rather than expensively obtained ones (e.g., attitudes from in-depth interviews) if both sets of variables predict equally well.

It should be clearly understood that a regression solution is extremely sensitive to the combination of variables that is included in it. Whether or not an IV appears particularly important in a solution depends on the nature of other IVs in the set. If the IV of interest is the only one that assesses some important facet of the DV, the IV will appear important; if the IV of interest is only one of several that assess the same important facet of the DV, it will appear less important. If some important facet of the DV is assessed by none of the IVs, the solution is weakened. An optimal set of IVs is the smallest, uncorrelated set that "covers the waterfront" with respect to the DV.

5.3.2 Practical Issues

In addition to theoretical considerations, use of multiple regression requires that several practical matters be attended, as described in Sections 5.3.2.1 through 5.3.2.4.

5.3.2.1 Ratio of Cases to IVs The cases-to-IVs ratio has to be substantial or the solution will be perfect—and meaningless. With more IVs than cases, one can find a regression solution that completely predicts the DV for each case, but only as an artifact of the cases-to-IV ratio. If either standard multiple or hierarchical regression is used, one would like to have 20 times more cases than IVs. That is, if you plan to include 5 IVs, it would be lovely to measure 100 cases. In fact, because of the width of the errors of estimating correlation with small samples, power may be unacceptably low no matter what the cases-to-IVs ratio if you have fewer than 100

cases. However, *a bare minimum requirement is to have at least 5 times more cases than IVs*—at least 25 cases if 5 IVs are used.

Be sure to verify that you have as many cases as you think you have. By default, regression programs delete cases for which there are missing values on any of the variables. Consult Chapter 4 if you have missing values and wish to rescore rather than delete them.

A higher cases-to-IV ratio is needed when the DV is skewed, effect size is anticipated to be small, or substantial measurement error is expected from unreliable variables. If the DV is not normally distributed and transformations are not undertaken, more cases are required. The size of anticipated effect is also relevant because more cases are needed to demonstrate a small effect than a large one. Finally, if substantial measurement error is expected from somewhat unreliable variables, more cases are needed.

It is also possible to have too many cases, however. As the number of cases becomes quite large, almost any multiple correlation will depart significantly from zero, even one that predicts negligible variance in the DV. For both statistical and practical reasons, then, one wants to measure the smallest number of cases that has a decent chance of revealing a significant relationship if, indeed, one is there.

If stepwise regression is to be used, more cases are needed. A cases-to-IV ratio of 40 to 1 is reasonable because stepwise regression often produces a solution that does not generalize beyond the sample unless the sample is large. An even larger sample is needed in stepwise regression if cross validation (deriving the solution with some of the cases and testing it on the remaining cases) is to be used to test the generalizability of the solution.

If you cannot measure as many as five cases for each IV, there are some strategies that may help. You can delete some IVs or create one (or more than one) IV that is a composite of several others. The new, composite IV is used in the analysis in place of the original IVs. Lastly, you can employ a setwise regression strategy, as available in BMDP9R and SAS RSQUARE and described in Section 5.5.3. In this frankly exploratory strategy, regressions are computed for all possible subsets of IVs and the researcher decides which regression is best.

5.3.2.2 Outliers Extreme cases have too much impact on the regression solution and should be deleted or rescored to reduce their influence. Consult Chapter 4 for a summary of procedures for detecting and dealing with univariate and multivariate outliers.

In regression, cases are evaluated for univariate extremeness with respect to the DV and each IV. Univariate outliers show up in initial screening runs (e.g., with BMDP2D or SPSS[x] FREQUENCIES) as cases far from the mean on either plots or z scores. Multivariate outliers are sought using either statistical methods such as Mahalanobis distance which looks at the combination of IVs (through BMDPAM, SPSS[x] REGRESSION, or SYSTAT MGLH as described in Chapter 4) or graphical methods that look at the combination of IVs in the context of the DV.

Graphical methods for identifying multivariate outliers use residuals to identify cases for which there is a poor fit between the obtained and predicted DV scores. A residuals plot (described in Section 5.3.2.4) is requested in an initial regression run;

extreme cases produce very large residuals which are some distance from the rest of the cases. That is, multivariate outliers produce points well outside the swarm of points for the remaining cases. Offending cases are then identified by requesting case by case output of the measures plotted.

Another graphical method based on residuals uses leverage on the X axis and residuals on the Y axis; the combination of leverage and residuals is called influence. *Residuals* identify outliers in the solution; *leverage* identifies outliers among the IVs; *influence* identifies cases that are too influential. As in routine residuals plots, outlying cases fall outside the swarm of points produced by the remainder of the cases.

Several varieties of residuals, leverage, and influence are available for plotting in the four statistical packages. For example, *residuals* are available in raw or standardized form—with or without the outlying case deleted. *Leverage* measures are Mahalanobis distance or variations of the diagonal elements of a "hat" matrix, called HATDIAG, RHAT, and h_i. The diagonal elements of the hat matrix are similar to Mahalanobis distance, but on a different scale so that significance tests based on a χ^2 distribution do not apply. Most *influence* measures are variations on Cook's distance, including modified Cook's distance, DFFITS, and DBETAS. Cook's distance is a measure of the change in regression coefficients produced by leaving out a case; cases with scores larger than 1.00 are suspected of being outliers. Tables 5.11 through 5.13 show which programs in the four computer packages produce these measures; consult the program manuals for a more comprehensive list.

5.3.2.3 Multicollinearity and Singularity Calculation of regression coefficients requires inversion of the matrix of correlations among the IVs (Equation 5.6), an inversion that is impossible if IVs are singular and unstable if they are multicollinear, as discussed in Chapter 4. Singularity and multicollinearity can be identified through perfect or very high squared multiple correlations (SMC) among IVs, where each IV in turn serves as DV while the others are IVs, or very low tolerances (1 − SMC). In regression, these conditions are also signaled by a very large (relative to the scale of the variable) standard error for a regression coefficient.

Most multiple regression programs have default values for tolerance that protect the user against inclusion of multicollinear variables. If the default values for the programs are in place, variables that are very highly correlated with variables already in the equation are not entered. This makes sense both statistically and logically because the variables threaten the analysis due to instability of regression coefficients and also are not needed because of their correlations with other variables.

However, you may want to make your own choice about which variable is deleted on logical rather than statistical grounds by considering such issues as the reliability of the variables. You may want to delete the least reliable variable rather than the variable identified by the program with very low tolerance. With a less reliable IV deleted from the set of IVs, the tolerance for the IV in question may be sufficient for entry.

If multicollinearity is detected but you want to maintain your set of IVs anyway, ridge regression might be considered. Ridge regression is a controversial procedure that attempts to stabilize estimates of regression coefficients by inflating the variance that is analyzed. For a more thorough description of ridge regression, see Chapter

7 of Dillon and Goldstein (1984). Although originally greeted with enthusiasm (cf. Price, 1977), serious questions about the procedure have been raised by Rozeboom (1979) and others. If, after consulting this literature, you still want to employ ridge regression, it is available through BMDP2R (see Appendix C–7 of the BMDP manual, Dixon, 1985).

5.3.2.4 Normality, Linearity, Homoscedasticity, and Independence of Residuals

Examination of residuals scatterplots provides a test of assumptions of normality, linearity, and homoscedasticity between predicted DV scores and errors of prediction. Assumptions of analysis are that the residuals (differences between obtained and predicted DV scores) are normally distributed about the predicted DV scores, that residuals have a straight line relationship with predicted DV scores, and that the variance of the residuals about predicted DV scores is the same for all predicted scores. When these assumptions are met, the residuals will appear as plot (a) in Figure 5.1.

Residuals scatterplots may be examined in lieu of initial screening runs or after them. If residuals scatterplots are examined in lieu of initial screening, and the assumptions of analysis are deemed met, further screening of variables is unnecessary. That is, if the residuals show normality, linearity, and homoscedasticity, if no outliers are in evidence, if the number of cases is sufficient, and if there is no evidence of multicollinearity or singularity, then regression requires only one run. (Parenthetically, we might note that we have never, in many years of multivariate analyses with many data sets, found this to be the case.) If, on the other hand, the residuals scatterplot from an initial run looks yucky, further screening via the procedures in Chapter 4 is warranted.

Residuals scatterplots are provided by all the statistical programs discussed in this chapter. All provide a scatterplot in which one axis is predicted scores and the other axis is errors of prediction. Which axis is which, however, and whether or not the predicted scores and residuals are standardized differ from program to program.

SPSS and BMDP provide the plots directly in their regression programs. In SPSS, both predicted scores and errors of prediction are standardized; in BMDP they are not. For SYSTAT and SAS, you have to save the predicted values and residuals, and then plot them using SYSTAT GRAPH or SAS PLOT, in either standardized or unstandardized form. In any event, it is the overall shape of the scatterplot that is of interest. If all assumptions are met, the residuals will be nearly rectangularly distributed with a concentration of scores along the center. As mentioned above, Figure 5.1(a) illustrates a distribution in which all assumptions are met.

The assumption of normality is that errors of prediction are normally distributed around each and every predicted DV score. The residuals scatterplot should reveal a pileup of residuals in the center of the plot at each value of predicted score, and a normal distribution of residuals trailing off symmetrically from the center. Figure 5.1(b) illustrates a failure of normality, with a skewed distribution of residuals.

A second test for the normality of the distribution of residuals is available through SPSS and BMDP programs for regression. The test is a normal probability plot of residuals in which their expected normal values are plotted against their actual normal values. Expected normal values are estimates of the z score a score should

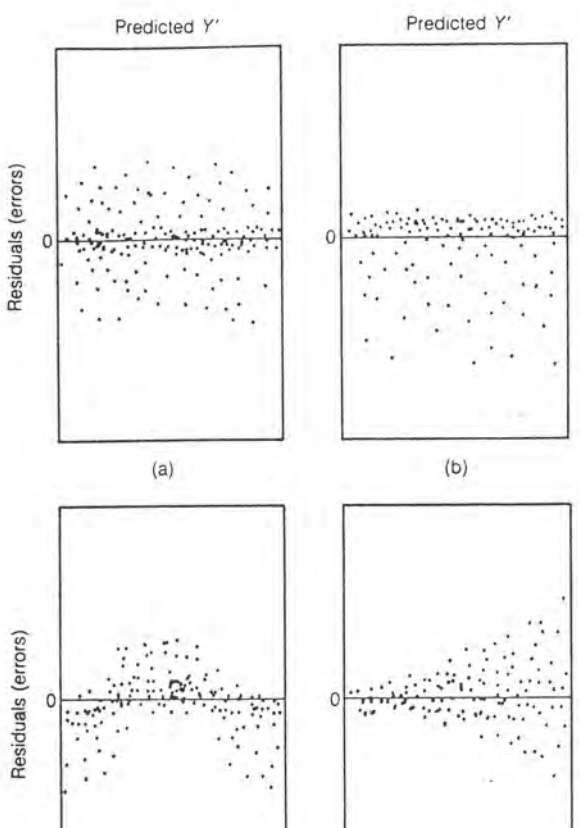


Figure 5.1 Plots of predicted values of the DV (Y') against residuals, showing (a) assumptions met, (b) failure of normality, (c) nonlinearity, and (d) heteroscedasticity.

have, given its rank in the original distribution if the original distribution is normal. If the expected normal values of residuals correspond to actual normal values (i.e., if the distribution of residuals is normal), the points will fall along a straight line running from the bottom left to the upper right corners of the graph. Distributions that are not normal will deviate from the straight line by curving above or below it in specific ways, depending on how the residuals are skewed. Bock (1975, pp. 156–160) illustrates the effects of several deviations from normality in the normal probability plot of residuals, as well as other details for the interested reader. The "fix" for the failure of normality of residuals is transformation of variables, after viewing distributions of individual variables.

Linearity of relationship between predicted DV scores and errors of prediction

is also assumed. If nonlinearity is present, the overall shape of the scatterplot will be curved instead of rectangular, as seen in Figure 5.1(c). In this illustration, errors of prediction are generally in a negative direction for low and high predicted scores and in a positive direction for medium predicted scores. Typically, nonlinearity of residuals can be made linear by transforming IVs (or the DV) so that there is a linear relationship between each IV and the DV. If, however, there is a curvilinear relationship between an IV and the DV, it may be necessary to include the square of the IV in the set of IVs.

Failure of linearity of residuals in regression does not invalidate an analysis so much as weaken it. A curvilinear relationship between the DV and an IV is a perfectly good relationship that is not completely captured by a linear correlation coefficient. The power of the analysis is reduced to the extent that the analysis does not have available the full extent of the relationships among the IVs and the DV.

The assumption of homoscedasticity is the assumption that the standard deviations of errors of prediction are approximately equal for all predicted DV scores. Heteroscedasticity also does not invalidate the analysis so much as weaken it. Homoscedasticity means that the band enclosing the residuals is approximately equal in width at all values of the predicted DV. Typical heteroscedasticity is a case in which the band becomes wider at larger predicted values, as illustrated in Figure 5.1(d). In this illustration, the errors of prediction increase as the size of the prediction increases. Heteroscedasticity may occur when some of the variables are skewed and others are not. Transformation of the variables may reduce or eliminate heteroscedasticity.

Another assumption of regression is that errors of prediction are independent of one another. In some instances, this assumption is violated as a function of something associated with the order of cases. For example, subjects who are interviewed early in a survey might be expected to exhibit more variability of response because of interviewer inexperience with the questionnaire.

Nonindependence of errors associated with order of cases can be checked in SPSS^x REGRESSION, BMDP9R, or SYSTAT MGLH by entering cases in order and requesting a plot of residuals against sequence of cases. The Durbin-Watson statistic is a measure of autocorrelation of errors over the sequence of cases, and, if significant, indicates nonindependence of errors. Details on the use of this statistic and a test for its significance are given by Wesolowsky (1976). If nonindependence is found, consult Dillon and Goldstein (1984) for the options available to you.

5.4 FUNDAMENTAL EQUATIONS FOR MULTIPLE REGRESSION

A data set appropriate for multiple regression consists of a sample of research units (e.g., graduate students) for whom scores are available on a number of IVs and on one DV. A small sample of hypothetical data with three IVs and one DV is illustrated in Table 5.1.

Table 5.1 contains six students, with scores on three IVs: a measure of professional motivation (MOTIV), a composite rating of qualifications for admissions

TABLE 5.1 SMALL SAMPLE OF HYPOTHETICAL DATA FOR ILLUSTRATION OF MULTIPLE CORRELATION

	IVs			DV
Case No.	MOTIV (X_1)	QUAL (X_2)	GRADE (X_3)	COMPR (Y)
1	14	19	19	18
2	11	11	8	9
3	8	10	14	8
4	13	5	10	8
5	10	9	8	5
6	10	7	9	12
Mean	11.00	10.17	11.33	10.00
Standard deviation	2.191	4.834	4.367	4.517

to graduate training (QUAL), and a composite rating of performance in graduate courses (GRADE). The DV is a rating of performance on graduate comprehensive exams (COMPR). We ask how well we can predict COMPR from scores on MOTIV, QUAL, and GRADE.

In actually addressing that question, a sample of six cases is highly inadequate, but the sample is adequate to illustrate calculation of multiple correlation and to demonstrate some analyses by canned computer programs. The reader is encouraged to work problems involving these data by hand as well as by available computer programs. Setup and selected output for this example through SPSS^x REGRESSION, BMDP1R, SAS REG, and SYSTAT MGLH appear in Section 5.4.3.

A variety of ways is available to develop the "basic" equation for multiple correlation.

5.4.1 General Linear Equation

One way of developing multiple correlation is to obtain the prediction equation for Y' in order to compare the predicted value of the DV with obtained Y.

$$Y' = A + B_1X_1 + B_2X_2 + \cdots + B_kX_k \tag{5.1}$$

where Y' is the predicted value of Y, A is the value of Y' when all X's are zero, B_1 to B_k represent regression coefficients, and X_1 to X_k represent the IVs.

The best-fitting regression coefficients produce a prediction equation for which squared differences between Y and Y' are at a minimum. Because squared errors of prediction—$(Y - Y')^2$—are minimized, this solution is called a least-squares solution.

In the sample problem, $k = 3$. That is, there are three IVs available to predict the DV, COMPR.

$$(\text{COMPR})' = A + B_M(\text{MOTIV}) + B_Q(\text{QUAL}) + B_G(\text{GRADE})$$

To predict a student's COMPR score, the available IV scores (MOTIV, QUAL, and GRADE) are multipled by their respective regression coefficients. The coefficient-by-score products are summed and added to the intercept, or baseline, value (A).

Differences among the observed values of the DV (Y), the mean of Y ($\bar{Y}$), and the predicted values of Y (Y') are summed and squared, yielding estimates of variation attributable to different sources. Total sum of squares for Y is partitioned into a sum of squares due to regression and a sum of squares left over or residual.

$$SS_y = SS_{reg} + SS_{res} \tag{5.2}$$

Total sum of squares of Y

$$SS_y = \Sigma (Y - \bar{Y})^2$$

is, as usual, the sum of squared differences between each individual's observed Y score and the mean of Y over all N cases. The sum of squares for regression

$$SS_{reg} = \Sigma (Y' - \bar{Y})^2$$

is the portion of the variation in Y that is explained by use of the IVs as predictors. That is, it is the sum of squared differences between predicted Y' and the mean of Y because the mean of Y is the best prediction for the value of Y in the absence of any useful IVs. Sum of squares residual

$$SS_{res} = \Sigma (Y - Y')^2$$

is the sum of squared differences between observed Y and the predicted scores, Y', and represents errors in prediction.

The squared multiple correlation is

$$R^2 = \frac{SS_{reg}}{SS_y} \tag{5.3}$$

The squared multiple correlation, R^2, is the proportion of sum of squares for regression in the total sum of squares for Y.

The squared multiple correlation is, then, the proportion of variation in the DV that is predictable from the best linear combination of the IVs. The multiple correlation itself is the correlation between the obtained and predicted Y values; that is, $R = r_{yy'}$.

Total sum of squares (SS_y) is calculated directly from the observed values of the DV. For example, in the sample problem, where the mean on the comprehensive examination is 10,

$$\begin{aligned} SS_C &= (18 - 10)^2 + (9 - 10)^2 + (8 - 10)^2 \\ &\quad + (8 - 10)^2 + (5 - 10)^2 + (12 - 10)^2 \\ &= 102 \end{aligned}$$

To find the remaining sources of variation, it is necessary to solve the prediction equation (Equation 5.1) for Y', which means finding the best-fitting A and B_i. The

most direct method of deriving the equation involves thinking of multiple correlation in terms of individual correlations.

5.4.2 Matrix Equations

Another way of looking at R^2 is in terms of the correlations between each of the IVs and the DV. The squared multiple correlation is the sum across all IVs of the product of the correlation between the DV and IV and the (standardized) regression coefficient for the IV.

$$R^2 = \sum_{i=1}^{k} r_{yi}\,\beta_i \tag{5.4}$$

where each r_{yi} is the correlation between the DV and the ith IV, and β_i is the standardized regression coefficient, or beta weight. The standardized regression coefficient is the regression coefficient that would be applied to the standardized X_i value—the z score of the X_i value—to predict standardized Y'.

Because r_{yi} are calculated directly from the data, computation of R^2 involves finding the standardized regression coefficients (β_i) for the k IVs.

Derivation of the k equations in k unknowns is beyond the scope of this book. However, solution of these equations is easily illustrated using matrix algebra. For those who are not familiar with matrix algebra, the rudiments of it are available in Appendix A. Sections A.4 (matrix multiplication) and A.5 (matrix inversion or division) are the only portions of matrix algebra necessary to follow the next few steps.

In matrix form:

$$R^2 = \mathbf{R}_{yi}\mathbf{B}_i \tag{5.5}$$

where $\mathbf{R}_{yi}$ is the row matrix of correlations between the DV and the k IVs, and $\mathbf{B}_i$ is a column matrix of standardized regression coefficients for the same k IVs.

The standardized regression coefficients can be found by inverting the matrix of correlations among IVs and multiplying that inverse by the matrix of correlations between the DV and the IVs.

$$\mathbf{B}_i = \mathbf{R}_{ii}^{-1}\mathbf{R}_{iy} \tag{5.6}$$

$\mathbf{B}_i$ is the column matrix of standardized regression coefficients, $\mathbf{R}_{ii}^{-1}$ is the inverse of the matrix of correlations among the IVs, and $\mathbf{R}_{iy}$ is the column matrix of correlations between the DV and the IVs. Because multiplication by an inverse is the same as division, the column matrix of correlations between the IVs and the DV is divided by the correlation matrix of IVs.

These equations,[1] then, are used to calculate R^2 for the sample COMPR data from Table 5.1. All the required correlations are in Table 5.2.

[1] The equations can be solved in terms of $\boldsymbol{\Sigma}$ (variance-covariance) matrices or $\mathbf{S}$ (sum-of-squares and cross-product) matrices as well as correlation matrices. If $\boldsymbol{\Sigma}$ or $\mathbf{S}$ matrices are substituted for corresponding $\mathbf{R}$ matrices, the regression coefficients are nonstandardized coefficients, as in Equation 5.1.

TABLE 5.2 CORRELATIONS AMONG IVs AND THE DV FOR SAMPLE DATA IN TABLE 5.1

		$\mathbf{R}_{ii}$			$\mathbf{R}_{iy}$
		MOTIV	QUAL	GRADE	COMPR
	MOTIV	1.00000	.39658	.37631	.58613
	QUAL	.39658	1.00000	.78329	.73284
	GRADE	.37631	.78329	1.00000	.75043
$\mathbf{R}_{yi}$	COMPR	.58613	.73284	.75043	1.00000

TABLE 5.3 INVERSE OF MATRIX OF INTERCORRELATIONS AMONG IVs FOR SAMPLE DATA IN TABLE 5.1

	MOTIV	QUAL	GRADE
MOTIV	1.20255	−0.31684	−0.20435
QUAL	−0.31684	2.67113	−1.97305
GRADE	−0.20435	−1.97305	2.62238

Procedures for inverting a matrix are amply demonstrated elsewhere (e.g., Cooley and Lohnes, 1971; Harris, 1975) and are typically available in computer installations. Because the procedure is extremely tedious by hand, and becomes increasingly so as the matrix becomes larger, the inverted matrix for the sample data is presented without calculation in Table 5.3.

Using Equation 5.6, the $\mathbf{B}_i$ matrix is found by postmultiplying the $\mathbf{R}_{ii}^{-1}$ matrix by the $\mathbf{R}_{iy}$ matrix.

$$\mathbf{B}_i = \begin{bmatrix} 1.20255 & -0.31684 & -0.20435 \\ -0.31684 & 2.67113 & -1.97305 \\ -0.20435 & -1.97305 & 2.62238 \end{bmatrix} \begin{bmatrix} .58613 \\ .73284 \\ .75043 \end{bmatrix} = \begin{bmatrix} 0.31931 \\ 0.29117 \\ 0.40221 \end{bmatrix}$$

so that $\beta_M = 0.319$, $\beta_Q = 0.291$, and $\beta_G = 0.402$. Then, using Equation 5.5, we obtain

$$R^2 = [.58613 \quad .73284 \quad .75043] \begin{bmatrix} 0.31931 \\ 0.29117 \\ 0.40221 \end{bmatrix} = .70237$$

In this example, 70% of the variance in graduate comprehensive exam scores is predictable from knowledge of motivation, admission qualifications, and graduate course performance.

Once the standardized regression coefficients are available, they are used to write the equation for the predicted values of COMPR (Y'). If z scores are used throughout, the beta weights (β_i) are used to set up the prediction equation. The equation is similar to Equation 5.1 except that there is no A (intercept).

If the equation is needed in raw score form, the coefficients must first be transformed to unstandardized B_i coefficients.

$$B_i = \beta_i \left(\frac{S_y}{S_i}\right) \tag{5.7}$$

Unstandardized coefficients (B_i) are found by multiplying standardized coefficients (beta weights $-$ β_i) by the ratio of standard deviations of the DV and IV, where S_i is the standard deviation of the ith IV and S_y is the standard deviation of the DV.

and

$$A = \bar{Y} - \sum_{i=1}^{k} (B_i \bar{X}_i) \tag{5.8}$$

The intercept is the mean of the DV less the sum of the means of the IVs multiplied by their respective unstandardized coefficients.

For the sample problem of Table 5.1:

$$B_M = 0.319 \left(\frac{4.517}{2.191}\right) = 0.658$$

$$B_Q = 0.291 \left(\frac{4.517}{4.834}\right) = 0.272$$

$$B_G = 0.402 \left(\frac{4.517}{4.366}\right) = 0.416$$

$$A = 10 - [(0.658)(11.00) + (0.272)(10.17) + (0.416)(11.33)] = -4.72$$

The prediction equation for raw COMPR scores, once scores on MOTIV, QUAL, and GRADE are known, is

$$(\text{COMPR})' = -4.72 + 0.658\,(\text{MOTIV}) + 0.272\,(\text{QUAL}) + 0.416\,(\text{GRADE})$$

If a graduate student has ratings of 12, 14, and 15, respectively, on MOTIV, QUAL, and GRADE, the predicted rating on COMPR is

$$(\text{COMPR})' = -4.72 + 0.658(12) + 0.272(14) + 0.416(15) = 13.22$$

5.4.3 Computer Analyses of Small Sample Example

Tables 5.4 through 5.7 demonstrate setup and selected output for computer analyses of the data in Table 5.1, using default values. Table 5.4 illustrates SPSS^x REGRESSION. The simplest BMDP program, BMDP1R, is illustrated in Table 5.5. Tables 5.6 and 5.7 show runs through SAS REG and SYSTAT MGLH, respectively.

In SPSS^x REGRESSION all of the variables, IVs and DV, are listed in the

TABLE 5.4 SETUP AND SELECTED SPSS* REGRESSION OUTPUT FOR STANDARD MULTIPLE REGRESSION ON SAMPLE DATA IN TABLE 5.1

```
        TITLE          SMALL SAMPLE MULTIPLE REGRESSION
        FILE HANDLE    TAPE50
        DATA LIST      FILE=TAPE50    FREE
                       /SUBJNO,MOTIV,QUAL,GRADE,COMPR
        REGRESSION     VARS=MOTIV TO COMPR/
                       DEP=COMPR/ENTER/


        * * * *  M U L T I P L E   R E G R E S S I O N  * * * *

VARIABLE LIST NUMBER 1   LISTWISE DELETION OF MISSING DATA

EQUATION NUMBER 1    DEPENDENT VARIABLE..  COMPR

BEGINNING BLOCK NUMBER  1.  METHOD:  ENTER

VARIABLE(S) ENTERED ON STEP NUMBER  1..    GRADE
                                    2..    MOTIV
                                    3..    QUAL


MULTIPLE R           .83807       ANALYSIS OF VARIANCE
R SQUARE             .70235                    DF       SUM OF SQUARES      MEAN SQUARE
ADJUSTED R SQUARE    .25588       REGRESSION    3           71.64007          23.88002
STANDARD ERROR      3.89615       RESIDUAL      2           30.35993          15.17997

                                  F =      1.57313     SIGNIF F =  .4114


---------- VARIABLES IN THE EQUATION ----------

VARIABLE            B         S B        BETA          T     SIG T

GRADE            .41603      .64619      .40221      .644    .5857
MOTIV            .65827      .87213      .31931      .755    .5292
QUAL             .27205      .58911      .29116      .462    .6896
(CONSTANT)     -4.72180     9.06565                  -.521    .6544


FOR BLOCK NUMBER  1   ALL REQUESTED VARIABLES ENTERED.
```

VARS = sentence of the REGRESSION instruction. Then the DV is specified. ENTER is the instruction that specifies standard multiple regression. Because this is standard multiple regression, results are given for only one step. At the left are R, R^2, adjusted R^2 (see Section 5.6.3) and STANDARD ERROR, the standard error of the predicted score, Y'. To the right is the ANOVA table, showing details of the F test of the hypothesis that multiple regression is zero (see Section 5.6.2.1). Below are the regression coefficients and their significance tests, including B weights, the standard error of B (S B), β weights (BETA), t-tests for the coefficients, and their significance levels (SIG T). The term (CONSTANT) refers to the intercept (A).

BMDP1R is limited to standard multiple regression so no special instruction is necessary to specify that procedure. The REGRESS instruction identifies the DV and the IVs. BMDP1R provides the same information as SPSS* but in slightly different format and with some different labels. For example, β is referred to as STD. REG COEFF and B weights are in the column labeled COEFFICIENT. Other differences are seen by comparing values in Tables 5.4 and 5.5. An additional statistic is TOLERANCE (cf. Chapter 4), but adjusted R^2 is not reported. BMDP1R also provides

TABLE 5.5 SETUP AND SELECTED BMDP1R OUTPUT FOR STANDARD MULTIPLE REGRESSION ON SAMPLE DATA IN TABLE 5.1

```
/PROBLEM    TITLE IS 'SMALL SAMPLE MULTIPLE REGRESSION'.
/INPUT      VARIABLES=5.  FORMAT IS  FREE.
            FILE = 'TAPE50'.
/VARIABLE   NAMES ARE SUBJNO,MOTIV,QUAL,GRADE,COMPR.
/REGRESS    DEPENDENT=COMPR.
            INDEPENDENT=MOTIV TO GRADE.
/END
```

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM
1 SUBJNO	3.50000	1.87083	.53452	1.00000	6.00000
2 MOTIV	11.00000	2.19089	.19917	8.00000	14.00000
3 QUAL	10.16667	4.83391	.47547	5.00000	19.00000
4 GRADE	11.33333	4.36654	.38528	8.00000	19.00000
5 COMPR	10.00000	4.51664	.45166	5.00000	18.00000

MULTIPLE R .8381 STD. ERROR OF EST. 3.8961
MULTIPLE R-SQUARE .7024

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	P(TAIL)
REGRESSION	71.6401	3	23.8800	1.573	.4114
RESIDUAL	30.3599	2	15.1800		

VARIABLE		COEFFICIENT	STD. ERROR	STD. REG COEFF	T	P(2 TAIL)	TOLERANCE
INTERCEPT		-4.72180					
MOTIV	2	.65827	.87213	.319	.755	.5292	.83157
QUAL	3	.27205	.58911	.291	.462	.6896	.37437
GRADE	4	.41603	.64619	.402	.644	.5857	.38133

TABLE 5.6 SETUP AND SAS REG OUTPUT FOR STANDARD MULTIPLE REGRESSION ON SAMPLE DATA OF TABLE 5.1

```
DATA SSAMPLE;
INFILE TAPE50;
INPUT SUBJNO MOTIV QUAL GRADE COMPR;
PROC REG;
MODEL COMPR = MOTIV QUAL GRADE;
```

DEP VARIABLE: COMPR

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	3	71.64007	23.88002	1.573	0.4114
ERROR	2	30.35993	15.17997		
C TOTAL	5	102			

ROOT MSE	3.896148	R-SQUARE	0.7024
DEP MEAN	10	ADJ R-SQ	0.2559
C.V.	38.96148		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|
INTERCEP	1	-4.7218	9.065646	-0.521	0.6544
MOTIV	1	0.6582658	0.8721309	0.755	0.5292
QUAL	1	0.2720479	0.5891141	0.462	0.6896
GRADE	1	0.4160342	0.6461905	0.644	0.5857

univariate statistics for each of the variables (see Chapter 4 for definition of the coefficient of variation).

SAS REG is also a program designed specifically for standard multiple regression. The variables for the regression equation are specified in the MODEL

TABLE 5.7 SETUP AND SYSTAT MGLH OUTPUT FOR STANDARD MULTIPLE REGRESSION ON SAMPLE DATA IN TABLE 5.1

```
USE TAPE50
MODEL COMPR=CONSTANT+MOTIV+QUAL+GRADE
ESTIMATE

DEP VAR:  COMPR      N:   6   MULTIPLE R:  .838   SQUARED MULTIPLE R:  .702
ADJUSTED SQUARED MULTIPLE R:  .256   STANDARD ERROR OF ESTIMATE:        3.896

  VARIABLE    COEFFICIENT   STD ERROR    STD COEF  TOLERANCE     T    P(2 TAIL)

CONSTANT          -4.722       9.066       0.000  1.0000000   -0.521   0.654
   MOTIV           0.658       0.872       0.319   .8315651    0.755   0.529
    QUAL           0.272       0.589       0.291   .3743735    0.462   0.690
   GRADE           0.416       0.646       0.402   .3813334    0.644   0.586

                        ANALYSIS OF VARIANCE

  SOURCE   SUM-OF-SQUARES   DF  MEAN-SQUARE     F-RATIO      P

REGRESSION        71.640     3       23.880       1.573     (0.41)
  RESIDUAL        30.360     2       15.180
```

statement, with the DV on the left side of the equation and the IVs on the right. In the ANOVA table the sum of squares for regression is called MODEL and residual is called ERROR. Total SS and df are in the row labeled C TOTAL. Below the ANOVA is the standard error of the estimate, shown as the square root of the error term, MS_{res} (ROOT MSE). Also printed are the mean of the DV (DEP MEAN), R^2, adjusted R^2, and the coefficient of variation (C.V.—defined as 100 times the standard error of the estimate divided by the mean of the DV). The section labeled PARAMETER ESTIMATES includes the usual B coefficients in the PARAMETER ESTIMATE column, their standard errors, t-tests for those coefficients (T FOR HO: PARAMETER = 0) and significance levels (PROB > |T|). Standardized regression coefficients are not printed unless requested.

Setup for SYSTAT MGLH is similar to SAS REG, except that the IV side of the equation includes CONSTANT (specifying the intercept) and terms are connected by + signs. The listing of information about R is fully labeled, followed by information about the regression coefficients—B (COEFFICIENT), standard error of B (STD ERROR), β (STD COEF), TOLERANCE, t-test and significance level of t. The ANOVA table is printed last.

Additional features of these programs are discussed in Section 5.7.

5.5 MAJOR TYPES OF MULTIPLE REGRESSION

There are three major analytic strategies in multiple regression: standard multiple regression, hierarchical regression, and statistical (stepwise and setwise) regression. Differences among the strategies involve what happens to overlapping variability due to correlated IVs and who determines the order of entry of IVs into the equation.

Consider the Venn diagram in Figure 5.2(a) in which there are three IVs and one DV. IV_1 and IV_2 both correlate substantially with the DV and with each other. IV_3 correlates to a lesser extent with the DV and to a negligible extent with IV_2. R^2

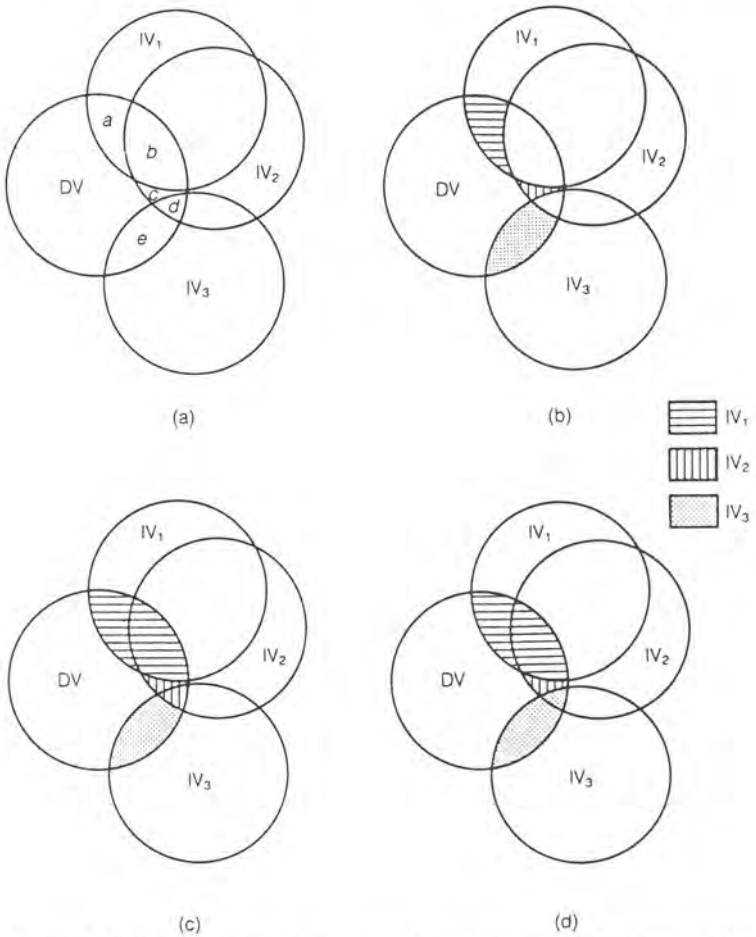


Figure 5.2 Venn diagrams illustrating (a) overlapping variance sections; and allocation of overlapping variance in (b) standard multiple regression, (c) hierarchical regression, and (d) stepwise regression.

for this situation is the area $a + b + c + d + e$. Area a comes unequivocally from IV_1; area c unequivocally from IV_2; area e from IV_3. However, there is ambiguity regarding areas b and d. Both areas could be predicted from either of two IVs; area b from either IV_1 or IV_2, area d from either IV_2 or IV_3. To which IV should the contested area be assigned? The interpretation of analysis can be drastically affected by choice of strategy because the apparent importance of the various IVs to the solution changes.

5.5.1 Standard Multiple Regression

The standard model is the one used in the solution for the small sample graduate student data in Table 5.1. In the standard, or simultaneous, model all IVs enter into

the regression equation at once; each one is assessed as if it had entered the regression after all other IVs had entered. Each IV is evaluated in terms of what it adds to prediction of the DV that is different from the predictability afforded by all the other IVs.

Consider Figure 5.2(b). The darkened areas of the figure indicate the variability accorded each of the IVs when the procedure of Section 5.6.1.1 is used. IV_1 "gets credit" for area *a*, IV_2 for area *c*, and IV_3 for area *e*. That is, each IV is assigned only the area it contributes uniquely. The overlapping areas, *b* and *d*, contribute to R^2, but are not assigned to any of the individual IVs.

In standard multiple regression, it is possible for a variable like IV_2 to appear unimportant in the solution when it actually is highly correlated with the DV. If the area of that correlation is whittled away by other IVs, the unique contribution of the IV is often very small despite a substantial correlation with the DV. For this reason, both the full correlation and the unique contribution of the IV need to be considered in interpretation.

Standard multiple regression is handled in the SPSS^x package by the REGRESSION program (NEW REGRESSION in the update of the original SPSS program), as are all other types of multiple regression. A selected part of output is given in Table 5.4 for the sample problem of Table 5.1. (Full interpretation of program output is available in substantive examples presented later in this chapter.)

Within the BMD series, standard multiple regression is routinely handled by BMDP1R. A selected sample of output from that program is given in Table 5.5. Standard multiple regression is also available through the more elaborate stepwise or setwise programs—BMDP2R or BMDP9R. SAS REG and SYSTAT MGLH are used for standard analyses, as illustrated in Tables 5.6 and 5.7.

5.5.2 Hierarchical Multiple Regression

In hierarchical regression, IVs enter the equation in an order specified by the researcher. Each IV is assessed in terms of what it adds to the equation at its own point of entry. Consider the example in Figure 5.2(c). Assume that the researcher assigns IV_1 first entry, IV_2 second entry, and IV_3 third entry. In assessing importance of variables by the procedure of Section 5.6.1.2, IV_1 "gets credit" for areas *a* and *b*, IV_2 for areas *c* and *d*, and IV_3 for area *e*. Each IV is assigned the variability, unique and overlapping, left to it at its own point of entry. Notice that the apparent importance of IV_2 would increase dramatically if it were assigned first entry and, therefore, "got credit" for *b*, *c*, and *d*.

The researcher normally assigns order of entry of variables according to logical or theoretical considerations. For example, IVs that are presumed (or manipulated) to be causally prior are given higher priority of entry. For instance, height might be considered prior to amount of training in assessing success as a basketball player and accorded a higher priority of entry. Variables with greater theoretical importance could also be given early entry.

Or the opposite tack could be taken. The research could enter manipulated or other variables of major importance on later steps, with "nuisance" variables given higher priority for entry. The lesser, or nuisance, set is entered first; then the major set is evaluated for what it adds to the prediction over and above the lesser set. For

example, we might want to see how well we can predict reading speed (the DV) from intensity and length of a speed reading course (the major IVs) while holding constant initial differences in reading speed (the nuisance IV). This is the basic analysis of covariance problem in regression format.

IVs can be entered one at a time or in blocks. The analysis proceeds in steps, with information about variables both in and out of the equation given in computer output at each step. Finally, after all variables are entered, summary statistics are provided along with the information available at the last step.

In the SPSS^x package, hierarchical regression is performed by the REGRESSION program (NEW REGRESSION in earlier SPSS versions). In Table 5.8, selected output is shown for the sample problem of Table 5.1, with higher priority given to admission qualifications and course performance, and lower priority given to motivation.

In the BMDP package, hierarchical regression is run with BMDP2R. Output from BMDP2R is illustrated in the next section. Neither SAS or SYSTAT provide programs for developing hierarchical models in a single run. IVs can be forced into the regression, but their order cannot be specified. However, each step in the hierarchy can be evaluated by a separate run using either SAS REG or SYSTAT MGLH.

5.5.3 Statistical (Stepwise) and Setwise Regression

Statistical regression (sometimes generically called stepwise regression) is a rather controversial procedure, in which order of entry of variables is based solely on statistical criteria. The meaning or interpretation of the variables is not relevant. Decisions about which variables are included and which omitted from the equation are based solely on statistics computed from the particular sample drawn; minor differences in these statistics can have profound effect on the apparent importance of an IV.

Consider the example in Figure 5.2(d). IV_1 and IV_2 both correlate substantially with the DV; IV_3 correlates less strongly. The choice between IV_1 and IV_2 for first entry is based on which of the two IVs has the higher full correlation with the DV, even if the higher correlation shows up in the second or third decimal place. Let's say that IV_1 has the higher correlation with the DV and enters first. It "gets credit" for areas a and b. At the second step, IV_2 and IV_3 are compared, where IV_2 has available to add to prediction areas c and d, and IV_3 has areas e and d. At this point IV_3 contributes more strongly to R^2 and enters the equation. IV_2 is now assessed for whether or not its remaining area, c, contributes significantly to R^2. If it does, IV_2 enters the equation; otherwise it does not despite the fact that it is almost as highly correlated with the DV as the variable that entered first. For this reason, interpretation of a statistical regression equation is hazardous unless the researcher takes special care to remember the message of the initial DV–IV correlations.

There are actually three versions of statistical regression, forward selection, backward deletion, and stepwise regression. In forward selection, the equation starts out empty and IVs are added one at a time provided they meet the statistical criteria for entry. Once in the equation, an IV stays in. In backward deletion, the equation starts out with all IVs entered and they are deleted one at a time if they do not contribute significantly to regression. Stepwise regression is a compromise between

TABLE 5.8 SETUP AND SELECTED SPSS* REGRESSION OUTPUT FOR HIERARCHICAL MULTIPLE REGRESSION ON SAMPLE DATA IN TABLE 5.1

```
TITLE          SMALL SAMPLE HIERARCHICAL MULTIPLE REGRESSION
FILE HANDLE    TAPE50
DATA LIST      FILE=TAPE50    FREE
               /SUBJNO MOTIV QUAL GRADE COMPR
REGRESSION     VARIABLES=MOTIV TO COMPR/
               STATISTICS=DEFAULTS,HISTORY/
               DEPENDENT=COMPR/
               ENTER QUAL GRADE/ENTER MOTIV/


                    * * * *  M U L T I P L E   R E G R E S S I O N  * * * *

VARIABLE LIST NUMBER 1   LISTWISE DELETION OF MISSING DATA

EQUATION NUMBER 1    DEPENDENT VARIABLE..  COMPR

BEGINNING BLOCK NUMBER  1.  METHOD:  ENTER    QUAL    GRADE

VARIABLE(S) ENTERED ON STEP NUMBER  1..    GRADE
                                    2..    QUAL


MULTIPLE R              .78586          ANALYSIS OF VARIANCE
R SQUARE                .61757                      DF       SUM OF SQUARES      MEAN SQUARE
ADJUSTED R SQUARE       .36262          REGRESSION   2            62.99218          31.49609
STANDARD ERROR         3.60591          RESIDUAL     3            39.00782          13.00261

                                        F =       2.42229     SIGNIF F =  .2365


---------------- VARIABLES IN THE EQUATION ----------------  ------------ VARIABLES NOT IN THE EQUATION ----------

VARIABLE     B        SE B       BETA       T       SIG T     VARIABLE     BETA      IN       PARTIAL      MIN      TOLER      T       SIG
T

GRADE      .47216      .59408     .45647      .795     .4848   MOTIV       .31931     .47085     .37437     .755
.5292
QUAL       .35065      .53664     .37529      .653     .5601
(CONSTANT) 1.08387    4.44060                 .244     .8229


FOR BLOCK NUMBER  1   ALL REQUESTED VARIABLES ENTERED.
```

TABLE 5.8 (Continued)

```
                    * * * *  M U L T I P L E   R E G R E S S I O N  * * * *

EQUATION NUMBER 1    DEPENDENT VARIABLE..  COMPR

BEGINNING BLOCK NUMBER  2.  METHOD:  ENTER      MOTIV

VARIABLE(S) ENTERED ON STEP NUMBER  3..   MOTIV

MULTIPLE R            .83807        ANALYSIS OF VARIANCE
R SQUARE              .70235                    DF       SUM OF SQUARES      MEAN SQUARE
ADJUSTED R SQUARE     .25588        REGRESSION   3            71.64007          23.88002
STANDARD ERROR       3.89615        RESIDUAL     2            30.35993          15.17997

                                    F =      1.57313      SIGNIF F =  .4114

---------------------- VARIABLES IN THE EQUATION -----------------------

VARIABLE            B         SE B        BETA          T    SIG T

GRADE             .41603      .66619      .40221      .644    .5857
QUAL              .27205      .58911      .29116      .462    .6896
MOTIV             .65827      .87213      .31931      .755    .5292
(CONSTANT)      -4.72180     9.06565                  -.521    .6544

FOR BLOCK NUMBER  2   ALL REQUESTED VARIABLES ENTERED.

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

                                        SUMMARY TABLE
                                        -------------

STEP    MULTR      RSQ    ADJRSQ    F(EQU)   SIGF    RSQCH      FCH  SIGCH          VARIABLE  BETAIN  CORREL  LABEL
   1                                                                          IN:  GRADE    .7504   .7504
   2    .7859     .6176    .3626     2.422    .236    .6176    2.422   .236  IN:  QUAL     .3753   .7328
   3    .8381     .7024    .2559     1.573    .411    .0848     .570   .529  IN:  MOTIV    .3193   .5861
```

the two other procedures in which the equation starts out empty and IVs are added one at a time if they meet statistical criteria, but they may also be deleted at any step where they no longer contribute significantly to regression. Of the three procedures, stepwise regression is considered the surest path to the best prediction equation.

Statistical regression is typically used to develop a subset of IVs that is useful in predicting the DV, and to eliminate those IVs that do not provide additional prediction to the IVs already in the equation. For this reason, statistical regression may have some utility if the only aim of the researcher is a prediction equation. Even so, the sample from which the equation is drawn should be large and representative because of the following considerations.

The procedure's controversy lies primarily in capitalization on chance and overfitting of data. Decisions about which variables are included and which omitted from the equation are dependent on potentially minor differences in statistics computed from a single sample, where some variability in the statistics from sample to sample is expected. In this way, the technique capitalizes on chance differences within a single sample. For similar reasons, the equation derived from a sample is too close to the sample and may not generalize well to the population. In this way, statistical regression may overfit the data.

For statistical regression, cross validation with a second sample is highly recommended. At the very least, separate analyses of two halves of an available sample should be conducted, with conclusions limited to results that are consistent for both analyses. (This caution applies to setwise regression as well.)

Further, a statistical analysis may not lead to the optimum solution in terms of R^2. Several IVs considered together may increase R^2, while any one of them considered alone does not. In simple statistical regression, none of the IVs enters.

By specifying that IVs enter in blocks, one can set up combinations of hierarchical and statistical regression. A block of high-priority IVs is set up to compete among themselves stepwise for order of entry; then a second block of IVs compete among themselves for order of entry. The regression is hierarchical over blocks, but statistical within blocks.

An alternative to statistical regression in choosing an optimal subset of variables is setwise regression. In setwise regression, separate regressions are computed for all IVs singly, all possible pairs of IVs, all possible trios of IVs and so forth until the best subset of IVs is identified according to some criterion (such as maximum R^2) from among all possible subsets. One computer program in the BMD series, BMDP9R, is designed to do this, as is one program in the SAS system, RSQUARE. This alternative enables the researcher to find IVs that significantly improve R^2 only when considered as a group.

The SPSS^x REGRESSION program handles stepwise regression in a manner similar to that of hierarchical regression. Both SPSS^x REGRESSION and BMDP2R provide a variety of statistical criteria and stepping options for statistical regression. For the data in Table 5.1, output from a BMDP2R run with default values is shown in Table 5.9.

SAS STEPWISE is the procedure provided by the SAS system for statistical regression. SYSTAT MGLH allows statistical regression as well, but minimal output is provided. For full information at each step, it is necessary to do separate runs.

TABLE 5.9 SETUP AND SELECTED BMDP2R OUTPUT FOR STEPWISE REGRESSION ON SAMPLE DATA IN TABLE 5.1

```
          /PROBLEM      TITLE IS 'SMALL SAMPLE STEPWISE REGRESSION'.
          /INPUT        VARIABLES ARE 5.  FORMAT IS FREE.
                        FILE IS 'TAPE50'.
          /VARIABLE     NAMES ARE SUBJNO,MOTIV,QUAL,GRADE,COMPR.
                        USE = MOTIV TO COMPR.
          /REGRESS      DEPENDENT=COMPR.
                        INDEPENDENT=MOTIV TO GRADE.
          /END


STEP NO.    0
----------

STD. ERROR OF EST.      4.5166

ANALYSIS OF VARIANCE
                  SUM OF SQUARES      DF     MEAN SQUARE
        RESIDUAL    102.00000          5     20.40000

                    VARIABLES IN EQUATION FOR COMPR

                              STD. ERROR   STD REG
    VARIABLE     COEFFICIENT  OF COEFF     COEFF     TOLERANCE
(Y-INTERCEPT        10.00000 )

                    VARIABLES NOT IN EQUATION

    F                          PARTIAL                    F
  TO REMOVE LEVEL.  VARIABLE    CORR.    TOLERANCE   TO ENTER LEVEL
                  .
                  . MOTIV      2 0.58613   1.00000      2.09      1


STEP NO.    1
----------
VARIABLE ENTERED    4 GRADE

MULTIPLE R              0.7504
MULTIPLE R-SQUARE       0.5631
ADJUSTED R-SQUARE       0.4539

STD. ERROR OF EST.      3.3376

ANALYSIS OF VARIANCE
                  SUM OF SQUARES      DF     MEAN SQUARE      F RATIO
        REGRESSION  57.440580          1     57.44056           5.16
        RESIDUAL    44.559420          4     11.13986

                                                          VARIABLES NOT IN EQUATION
```

STEP NO. 1

VARIABLE ENTERED 4 GRADE

MULTIPLE R 0.7504
MULTIPLE R-SQUARE 0.5631
ADJUSTED R-SQUARE 0.4539

STD. ERROR OF EST. 3.3376

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	57.440560	1	57.44056	5.16
RESIDUAL	44.559420	4	11.13986	

VARIABLES IN EQUATION FOR COMPR

VARIABLE	COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT	1.20280)					
GRADE 4	0.77622	0.3418	0.750	1.00000	5.16	1

***** F LEVELS(4.000, 3.900) OR TOLERANCE INSUFFICIENT FOR FURTHER STEPPING

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
MOTIV	2	0.49600	0.85839	0.98	1

STEPWISE REGRESSION COEFFICIENTS

VARIABLES STEP	0 Y-INTCPT	2 MOTIV	3 QUAL	4 GRADE
0	10.0000*	1.2083	0.6847	0.7762
1	1.2028*	0.7295	0.3507	0.7762*

NOTE - 1) REGRESSION COEFFICIENTS FOR VARIABLES IN THE EQUATION ARE INDICATED BY AN ASTERISK
2) THE REMAINING COEFFICIENTS ARE THOSE WHICH WOULD BE OBTAINED IF THAT VARIABLE WERE TO ENTER IN THE NEXT STEP

SUMMARY TABLE

STEP NO.	VARIABLE ENTERED	VARIABLE REMOVED	MULTIPLE R	RSQ	CHANGE IN RSQ	F TO ENTER	F TO REMOVE	NO.OF VAR. INCLUDED
1	4 GRADE		0.7504	0.5631	0.5631	5.16		1

5.5.4 Choosing among Regression Strategies

To simply assess relationships among variables, and answer the basic question of multiple correlation, the method of choice is standard multiple regression. As a matter of fact, unless there is good reason to use some other technique, standard multiple regression is recommended. Reasons for using hierarchical regression are theoretical or for testing hypotheses.

Hierarchical regression allows the researcher to control the advancement of the regression process. Importance of IVs in the prediction equation is determined by the researcher according to logic or theory. Explicit hypotheses are tested about proportion of variance attributable to some IVs after variance due to IVs already in the equation is accounted for.

Although there are similarities in programs used and output produced for hierarchical and statistical regression, there are fundamental differences in the way that IVs enter the prediction equation and in the interpretations that can be made from the results. In hierarchical regression the researcher controls entry of variables, while in statistical and setwise regression statistics computed from sample data control order of entry. Statistical and setwise regression are, therefore, model-building rather than model-testing procedures. As exploratory techniques, they may be useful for such purposes as eliminating variables that are clearly superfluous in order to tighten up future research. However, clearly superfluous IVs will show up in any of the procedures.

When multicollinearity or singularity are present, setwise regression can be indispensable in identifying multicollinear variables, as indicated in Chapter 4.

For the example of Section 5.4, in which performance on graduate comprehensive exam (COMPR) is predicted from professional motivation (MOTIV), qualifications for graduate training (QUAL), and performance in graduate courses (GRADE), the differences among regression strategies might be phrased as follows. If standard multiple regression is used, two fundamental questions are asked: (1) What is the size of the overall relationship between COMPR and MOTIV, QUAL, and GRADE? and (2) How much of the relationship is contributed uniquely by each IV? If hierarchical regression is used, with QUAL and GRADE entered before MOTIV, the question is, Does MOTIV significantly add to prediction of COMPR after differences among students in QUAL and GRADE have been statistically eliminated? If statistical regression is used, one asks, What is the best linear combination of IVs to predict the DV in this sample? And use of setwise regression leads to the query, What is the size of R^2 from each one of the IVs, from all possible combinations of two IVs, and from all three IVs for this sample?

5.6 SOME IMPORTANT ISSUES

5.6.1 Importance of IVs

If the IVs are uncorrelated with each other, assessment of the contribution of each of them to multiple regression is straightforward. IVs with bigger correlations or

higher standardized regression coefficients are more important to the solution than those with lower (absolute) values. (Because unstandardized regression coefficients are in a metric that depends on the metric of the original variables, their sizes are harder to interpret.)

If the IVs are correlated with each other, assessment of the importance of each of them to regression is more ambiguous. When IVs are correlated with each other, correlations and regression coefficients are somewhat redundant and misleading. The correlation between an IV and the DV reflects variance shared with the DV, but some of that variance may be predictable from other IVs. Regression coefficients can be misleading when an IV has a large regression coefficient not because it directly predicts the DV, but because it predicts the DV well after another IV suppresses irrelevant variance (see Section 5.6.4).

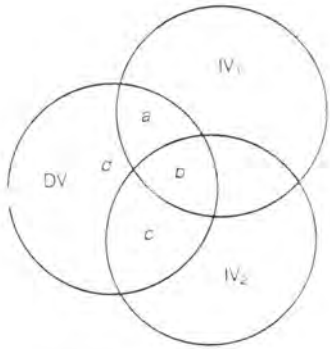
To get the most straightforward answer regarding the importance of an IV to regression, one needs to consider the type of regression it is, and both the full and unique relationship between the IV and the DV. This section reviews several of the issues to be considered when assessing the importance of an IV to standard multiple, hierarchical, or statistical regression. In all cases, one needs to compare the total relationship of the IV with the DV, the unique relationship of the IV with the DV, and the correlations of the IVs with each other in order to get a complete picture of the function of an IV in regression. The total relationship of the IV with the DV (correlation) and the correlations of the IVs with each other are given in the correlation matrix. The unique contribution of an IV to predicting a DV is generally assessed by either partial or semipartial correlation.

For standard multiple and hierarchical regression, the relationships between correlation, partial correlation, and semipartial correlation are given in Figure 5.3 for a simple case of one DV and two IVs. In the figure, (squared) correlation, partial correlation, and semipartial correlation coefficients are defined as areas created by overlapping circles. Area $a + b + c + d$ is the total area of the DV and reduces to a value of 1 in many equations. Area b is the segment of the variability of the DV that can be explained by either IV_1 or IV_2, and is the segment that creates the ambiguity.

In partial correlation, the contribution of the other IVs is taken out of both the IV and the DV. In semipartial correlation, the contribution of other IVs is taken out of only the IV. Thus (squared) semipartial correlation expresses the unique contribution of the IV to the total variance of the DV. Squared semipartial correlation (sr_i^2) is a very useful measure of importance of an IV. The interpretation of sr_i^2 differs, however, depending on the type of multiple regression employed.

5.6.1.1 Standard Multiple and Setwise Regression In standard multiple regression and setwise regression sr_i^2 for an IV is the amount by which R^2 is reduced if that IV is deleted from the regression equation. That is, sr_i^2 represents the unique contribution of the IV to R^2 in that set of IVs.

When the IVs are correlated, squared semipartial correlations do not necessarily sum to multiple R^2. The sum of sr_i^2 is usually smaller than R^2 (although under some rather extreme circumstances, the sum can be larger than R^2). When the sum is smaller, the difference between R^2 and the sum of sr_i^2 for all IVs represents shared



	Standard Multiple	Hierarchical
r^2	IV_1 $(a + b)/(a + b + c + d)$ IV_2 $(c + b)/(a + b + c + d)$	$(a + b)/(a + b + c + d)$ $(c + b)/(a + b + c + d)$
sr^2	IV_1 $a/(a + b + c + d)$ IV_2 $c/(a + b + c + d)$	$(a + b)/(a + b + c + d)$ $c/(a + b + c + d)$
pr^2	IV_1 $a/(a + d)$ IV_2 $c/(c + d)$	$(a + b)/(a + b + d)$ $c/(c + d)$

Figure 5.3 Areas representing squared correlation, squared semipartial correlation, and squared partial correlation in standard multiple and hierarchical regression (where IV_1 is given priority over IV_2).

variance, variance that is contributed to R^2 by two or more IVs. It is rather common to find substantial R^2 with sr_i^2 for all IVs quite small.

Table 5.10 summarizes procedures for finding sr_i^2 (and pr_i^2) for both standard multiple and hierarchical regression through SPSS^x, BMDP, SAS, and SYSTAT.

SPSS^x and SAS provide sr_i^2 as part of their output. If you use SPSS^x REGRESSION or SPSS NEW REGRESSION, sr_i is optionally available as PART COR by requesting STATISTICS = ZPP. SAS REG provides two semipartial correlations with different error terms. For standard multiple regression, the appropriate form is SCORR2, which uses the Type II error term.

For BMDP1R, BMDP9R, or SYSTAT MGLH, sr_i^2 is most easily calculated using Equation 5.9.

$$sr_i^2 = \frac{T_i^2}{\mathrm{df}_{\mathrm{res}}}(1 - R^2) \tag{5.9}$$

The squared semipartial correlation (sr_i^2) for the ith IV is calculated from the T_i ratio for that IV, the residual degrees of freedom ($\mathrm{df}_{\mathrm{res}}$), and R^2.

In BMDP1R, R^2 is labeled MULTIPLE R-SQUARE and $\mathrm{df}_{\mathrm{res}}$ is found in the analysis of variance table (cf. Table 5.5). With BMDP9R, information for calculating sr_i^2 is available for the best subset, with $\mathrm{df}_{\mathrm{res}}$ shown as DENOMINATOR DEGREES OF FREEDOM. R^2 is shown as SQUARED MULTIPLE CORRELATION. SYSTAT

TABLE 5.10 PROCEDURES FOR FINDING sr_i^2 AND pr_i^2 THROUGH SPSS*, SAS, BMDP, AND SYSTAT FOR STANDARD MULTIPLE AND HIERARCHICAL REGRESSION

	sr_i^2	pr_i^2
Standard Multiple Regression		
SPSS* REGRESSION	STATISTICS = ZPP PART COR	STATISTICS = ZPP PARTIAL
SAS REG	SCORR2	PCORR2
BMDP1R, 9R	Use Eq. 5.9	Use Eq. 5.10
SYSTAT MGLH	Use Eq. 5.9	Use Eq. 5.10
Hierarchical Regression		
SPSS* REGRESSION	RSQCH in SUMMARY TABLE	Use Eq. 5.10
BMDP2R	CHANGE IN RSQ	Use Eq. 5.10
SAS STEPWISE REG	PARTIAL R**2 SCORR1	PCORR1
SYSTAT MGLH	Get by subtraction of R^2 in sequential runs	Use Eq. 5.10

MGLH offers R^2 as SQUARED MULTIPLE R (cf. Table 5.7) and shows df_{res} in the analysis of variance table. (SAS RSQUARE offers neither significance tests for IVs nor residual degrees of freedom for any equation.)

From the output shown in Table 5.5 the squared semipartial correlation for GRADE is

$$sr_i^2 = \frac{(0.644)^2}{2}(1 - .7024) = .06$$

In this set of IVs, then, R^2 would be reduced by .06 if GRADE were deleted from the equation.

If partial correlations squared (pr_i^2) are desired, they are available (in unsquared form) in output in SPSS and SAS. The statistic PARTIAL is reported in SPSS* REGRESSION or SPSS NEW REGRESSION. SAS REG provides two types of partial correlation coefficients, differing by the error term chosen. As for sr_i^2, the Type II error term is appropriate for standard multiple regression, and pr_i^2 is requested as PCORR2.

For BMDP and SYSTAT, squared partial correlations are calculated from T_i and df_{res} as shown in Equation 5.10.

$$pr_i^2 = \frac{T_i^2}{T_i^2 + df_{res}} \tag{5.10}$$

In all standard multiple regression programs, T_i is the significance test for sr_i^2, B_i, and β_i, as discussed in Section 5.6.2.

5.6.1.2 Hierarchical or Statistical Regression In these two forms of regression, sr_i^2 is interpreted as the amount of variance added to R^2 by each IV at the point that it enters the equation. The research question is, How much does this IV add to multiple R^2 after IVs with higher priority have contributed their share to prediction of the DV? Thus, the apparent importance of an IV is very likely to depend on its point of entry into the equation.

The sr_i^2 do sum to R^2 in hierarchical and statistical regression. Consult Figure 5.2 if you want to review this point.

As reviewed in Table 5.10, SPSS, BMDP, and SAS programs provide squared semipartial correlations as part of routine output for hierarchical and statistical regression. For SPSS, sr_i^2 is RSQCH for each IV in the SUMMARY TABLE (see Table 5.8). For BMDP2R, sr_i^2 is CHANGE IN RSQ for each IV (see Table 5.9). SAS STEPWISE provides sr_i^2 as PARTIAL R**2. If you use SAS REG for hierarchical regression by ordering IVs in successive MODEL statements, you can find the appropriate type of sr_i^2 by requesting SCORR1.

For SYSTAT MGLH, R^2 is provided for each sequential run. You can calculate sr_i^2 by subtraction between subsequent runs.

If desired, partial correlations can be found from T_i, as in Equation 5.10. These partial correlations are the same as for standard multiple regression. SAS REG offers PCORR1 as the form of partial correlation appropriate for hierarchical multiple regression.

5.6.2 Statistical Inference

This section covers significance tests for multiple regression and for regression coefficients for individual IVs. A test, F_{inc}, is also described for evaluating the statistical significance of adding two or more IVs to a prediction equation in hierarchical or statistical analysis. Calculations of confidence limits for unstandardized regression coefficients and procedures for comparing the predictive capacity of two different sets of IVs conclude the section.

It should be noted that when the researcher is using either statistical or setwise regression as an exploratory tool, inferential procedures of any kind may be inappropriate. Inferential procedures require that the researcher have a hypothesis to test. When statistical or setwise regression is used to snoop data, there may be no hypothesis, even though the statistics themselves are available.

5.6.2.1 Test for Multiple *R* The overall inferential test in multiple regression is whether the sample of scores is drawn from a population in which multiple R is zero. This is equivalent to the null hypothesis that all correlations between DV and IVs and all regression coefficients are zero. With large N, the test of this hypothesis becomes trivial because it is almost certain to be rejected. Testing multiple R is more interesting with fewer cases.

For standard multiple and hierarchical regression, the test of this hypothesis is presented in all computer outputs as analysis of variance. For hierarchical regression (and for standard multiple regression performed through stepwise programs), the analysis of variance table at the last step gives the relevant information. The F ratio

for mean square regression over mean square residual tests the significance of multiple R. Mean square regression is the sum of squares for regression in Equation 5.2 divided by k degrees of freedom; mean square residual is the sum of squares for residual in the same equation divided by $(N - k - 1)$ degrees of freedom.

If you insist on inference in statistical or setwise regression, adjustments are necessary because all potential IVs do not enter the equation and sample R^2 is not distributed as F. Therefore the analysis of variance table at the last step (or for the "best" equation) is misleading; the reported F is biased so that the F ratio actually reflects a Type I error rate in excess of alpha.

Wilkinson and Dallal (1981) have developed tables for critical R^2 when forward selection procedures are used for statistical addition of variables and the selection stops when the F-to-enter for the next variable falls below some preset value. Appendix C contains Table C.5 that shows how large multiple R^2 must be to be statistically significant at .05 or .01 levels, given N, k, and F, where N is sample size, k is the number of potential IVs, and F is the minimum F-to-enter that is specified. F-to-enter values that can be chosen are 2, 3, or 4.

For example, for a statistical regression in which there are 100 subjects, 20 potential IVs, and an F-to-enter value of 3 chosen for the solution, a multiple R^2 of approximately .19 is required to be considered significantly different from zero at $\alpha = .05$ (and approximately .26 at $\alpha = .01$). Wilkinson and Dallal report that linear interpolation on N and k works well. However, they caution against extensive extrapolation for values of N and k beyond those given in the table. In hierarchical regression this table is also used to find critical R^2 values if a post hoc decision is made to terminate regression as soon as R^2 reaches statistical significance, if the appropriate F-to-enter is substituted for R^2 probability value as the stopping rule.

Wilkinson and Dallal recommend forward selection procedures in favor of other selection methods (e.g., stepwise selection, subset regression). They argue that, in practice, results using different procedures are not likely to be substantially different. Further, forward selection is computationally simple and efficient, and allows straightforward specification of stopping rules. If you are able to specify in advance the number of variables you wish to select, an alternative set of tables is provided by Wilkinson (1979) to evaluate significance of multiple R^2 with forward selection.

After looking at the data, you may wish to test the significance of some subsets of IVs in predicting the DV where a subset may even consist of a single IV. If several a posteriori tests like this are desired, Type I errors become increasingly likely. Larzelere and Mulaik (1977) recommend the following conservative F test to keep Type I error rate below α for all combinations of IVs:

$$F = \frac{R_s^2/k}{(1 - R_s^2)/(N - k - 1)} \tag{5.11}$$

where R_s^2 is the squared multiple (or bivariate) correlation to be tested for significance, and k is the total number of IVs. Obtained F is compared with tabled F, with k and $(N - k - 1)$ degrees of freedom. That is, the critical value of F for each subset is the same as the critical value for the overall multiple R.

In the sample problem of Section 5.4, the bivariate correlation between MOTIV and COMPR is tested a posteriori as follows:

$$F = \frac{.58613^2/3}{(1 - .58613^2)/2} = 0.349, \qquad \text{with df} = 3, 2$$

5.6.2.2 Test of Regression Components In standard multiple regression, for each IV the same significance test evaluates B_i, β_i, sr_i and pr_i. The test is straightforward, and results are given in computer output. In SPSS NEW REGRESSION or SPSS^x, F_i tests the unique contribution of each IV and appears in the output section labeled VARIABLES IN THE EQUATION (see Table 5.4). Degrees of freedom are 1 and df_{res}, which appears in the accompanying analysis of variance table. In BMDP1R, SYSTAT MGLH, and SAS REG, T_i values are given for each IV, tested with df_{res} from the analysis of variance table (see Tables 5.5 to 5.7). If BMDP2R is used for standard multiple regression, F_i values are available at the last step of the analysis as F-TO-REMOVE; they are interpreted as for SPSS.

It is important in interpretation to recall the limitations of these significance tests. The significance tests are sensitive only to the unique variance an IV adds to R^2. A very important IV that shares variance with another IV in the analysis may be nonsignificant although the two IVs in combination are responsible in large part for the size of R^2. An IV that is highly correlated with the DV but has a nonsignificant regression coefficient may have suffered just such a fate. For this reason, it is important to report and interpret r_{iy} in addition to sr_i^2 and F_i for each IV, as shown in Table 5.18.

For setwise regression through BMDP9R the individual IVs are best interpreted as if for standard multiple regression where each variable enters last. The T_i values given for each variable in the "best" subset are interpreted the same as those for BMDP1R noted earlier.

For statistical and hierarchical regression, assessment of contribution of variables is more complex, and appropriate significance tests may not appear in the computer output. First, there is inherent ambiguity in the testing of each variable. In statistical and hierarchical regression, tests of sr_i^2 and pr_i^2 are not the same as tests of the regression coefficients (B_i and β_i). Regression coefficients are independent of order of entry of the IVs, whereas sr_i^2 and pr_i^2 depend directly on order of entry. Because sr_i^2 reflects "importance" as typically of interest in hierarchical or statistical regression, tests based on sr_i^2 are discussed here.[2]

SPSS, BMDP, and SAS provide significance tests for sr_i^2 in summary tables. For SPSS, the test is FCH—F ratio for change in R^2—that is accompanied by a significance value, SIGCH. In BMDP2R, the F ratio for sr_i^2 is reported in the summary table as F TO ENTER. No significance value is given, but you can evaluate this F with numerator df = 1 and denominator df_{res} for the last step in the equation. For SAS, the relevant values are F in the summary table and, for significance, PROB>F.

[2] For combined standard-hierarchical regression, it might be desirable to use the "standard" method for all IVs simply to maintain consistency. If so, be sure to report that the F test is for regression coefficients.

If you use SYSTAT MGLH, you need to calculate F for your sr_i^2 as found by subtraction (cf. Section 5.6.1.2) using the following equation:

$$F_i = \frac{sr_i^2}{(1 - R^2)/\mathrm{df}_{\mathrm{res}}} \tag{5.12}$$

The F_i for each IV is based on sr_i^2 (the squared semipartial correlation), multiple R^2 at the final step; and residual degrees of freedom from the analysis of variance table for the final step.

5.6.2.3 Test of Added Subset of IVs For hierarchical and statistical regression, one can test whether a block of two or more variables significantly increases R^2 above the R^2 predicted by a set of variables already in the equation.

$$F_{\mathrm{inc}} = \frac{(R_{wi}^2 - R_{wo}^2)/M}{(1 - R_{wi}^2)/\mathrm{df}_{\mathrm{res}}} \tag{5.13}$$

where F_{inc} is the incremental F ratio; R_{wi}^2 is the multiple R^2 achieved with the new block of IVs in the equation; R_{wo}^2 is the multiple R^2 without the new block of IVs in the equation; M is the number of IVs in the block; and $\mathrm{df}_{\mathrm{res}} = (N - k - 1)$ is residual degrees of freedom in the final analysis of variance table. Both R_{wi}^2 and R_{wo}^2 are found in the SUMMARY TABLE of any computer output. If you are using SPSS^x REGRESSION or SPSS NEW REGRESSION, F_{inc} is available through the TEST command.

The null hypothesis of no increase in R^2 is tested as F with M and $\mathrm{df}_{\mathrm{res}}$ degrees of freedom. If the null hypothesis is rejected, the new block of IVs does significantly increase the explained variance.

Although this is a poor example because only one variable is in the new block, we can use the hierarchical example in Table 5.8 to test whether MOTIV adds significantly to the variance contributed by the first two variables to enter the equation, QUAL and GRADE.

$$F_{\mathrm{inc}} = \frac{(.70235 - .61757)/1}{(1 - .70235)/2} = 0.570 \qquad \text{with df} = 1, 2$$

(Because only one variable was entered, this test is the same as the test provided for MOTIV in Table 5.8 under the heading VARIABLES NOT IN THE EQUATION. When the T value of .755 is squared, it is the same as F_{inc} above. Thus, F_{inc} can be used when there is only one variable in the block, but the information is already provided in the output.)

5.6.2.4 Confidence Limits around *B* To estimate population values, confidence limits for unstandardized regression coefficients (B_i) are calculated. Standard errors of unstandardized regression coefficients, unstandardized regression coefficients, and

the critical value of t for the desired confidence level (based on $N - 2$ degrees of freedom, where N is the sample size) are used in Equation 5.14.

$$CL_{B_i} = B_i \pm SE_{B_i}(t_{\alpha/2}) \tag{5.14}$$

The $(1 - \alpha)$ confidence limits for the unstandardized regression coefficient for the ith IV (CL_{B_i}) is the regression coefficient (B_i) plus or minus the standard error of the regression coefficient (SE_{B_i}) times the critical value of t, with $(N - 2)$ degrees of freedom at the desired level of α.

If 95% confidence limits are desired, they are given in SPSS[x] REGRESSION output as 95 CONFDNCE INTRVL B under the segment of output titled VARIABLES IN THE EQUATION. With other output or when 99% confidence limits are desired, Equation 5.14 is used. Unstandardized regression coefficients and the standard errors of those coefficients appear in the sections labeled VARIABLES IN THE EQUATION. Standard errors are called STD ERROR OF COEFF in BMDP2R and are called STD ERROR in SYSTAT MGLH, SAS, BMDP1R, and BMDP9R.

For the example in Table 5.4, the 95% confidence limits for GRADE, with df = 4, are

$$CL_{B_G} = 0.416 \pm 0.646(2.78) = 0.416 \pm 1.796 = -1.380 \leftrightarrow 2.212$$

If the confidence interval contains zero, one retains the null hypothesis that the population regression coefficient is zero.

5.6.2.5 Comparing Two Sets of Predictors It is sometimes of interest to know whether one set of IVs predicts a DV better than another set of IVs. For example, can ratings of current belly dancing ability be better predicted by personality tests or by past dance and musical training?

The procedure for finding out is fairly convoluted, but if you have a large sample and you have access to BMDP or are willing to enter a pair of predicted scores for each subject in your sample, a test for the significance of the difference between two "correlated correlations" (both correlations are based on the same sample and share a variable) is available (Steiger, 1980). (If sample size is small, nonindependence among predicted scores for cases can result in serious violation of the assumptions underlying the test.)

As suggested in Section 5.4.1, a multiple correlation can be thought of as a simple correlation between obtained DVs and predicted DVs; that is, $R = r_{yy'}$. If there are two sets of predictors, $Y_{a'}$ and $Y_{b'}$ (where, for example $Y_{a'}$ is prediction from personality scores and $Y_{b'}$ is prediction from past training), a comparison of their relative effectiveness in predicting Y is made by testing for the significance of the difference between $r_{yy'_a}$ and $r_{yy'_b}$. For simplicity, let's call these r_{ya} and r_{yb}.

The trick is that to test the difference we need to know the correlation between the predicted scores from set A (personality) and those from set B (training), that is, $r_{ya'yb'}$ or, simplified, r_{ab}. This is where the BMDP file manipulation procedures or hand entering become necessary.

The z test for the difference between r_{ya} and r_{yb} is

$$\bar{Z}^* = (z_{ya} - z_{yb}) \sqrt{\frac{N - 3}{2 - 2\bar{s}_{ya,yb}}} \tag{5.15}$$

where N is, as usual, the sample size,

$$z_{ya} = \tfrac{1}{2} \ln \left(\frac{1 + r_{ya}}{1 - r_{ya}} \right) \quad \text{and} \quad z_{yb} = \tfrac{1}{2} \ln \left(\frac{1 + r_{yb}}{1 - r_{yb}} \right)$$

and

$$\bar{s}_{ya,yb} = \frac{(r_{ab})(1 - \bar{r}^2 - \bar{r}^2) - \tfrac{1}{2}(\bar{r}^2)(1 - \bar{r}^2 - \bar{r}^2 - r_{ab}^2)}{(1 - \bar{r}^2)^2}$$

where $\bar{r} = \tfrac{1}{2}(r_{ya} + r_{yb})$.

So, for the example, if the correlation between currently measured ability and ability as predicted from personality scores is .40 ($R_a = r_{ya} = .40$), the correlation between currently measured ability and ability as predicted from past training is .50 ($R_b = r_{yb} = .50$), and the correlation between ability as predicted from personality and ability as predicted from training is .10 ($r_{ab} = .10$), and $N = 103$,

$$\bar{r} = \tfrac{1}{2}(.40 + .50) = .45$$

$$\bar{s}_{ya,yb} = \frac{(.10)(1 - .45^2 - .45^2) - 1/2(.45)^2(1 - .45^2 - .45^2 - .10^2)}{(1 - .45^2)^2}$$

$$= .0004226$$

$$z_{ya} = \tfrac{1}{2} \ln \left(\frac{1 + .40}{1 - .40} \right) = .42365$$

$$z_{yb} = \tfrac{1}{2} \ln \left(\frac{1 + .50}{1 - .50} \right) = .54931$$

and, finally,

$$\bar{z}^* = (.42365 - .54931) \sqrt{\frac{103 - 3}{2 - .000845}} = -0.88874$$

Because $\bar{z}^*$ is within the critical values of ± 1.96 for a two-tailed test, there is no statistically significant difference between multiple R when predicting Y from Y'_a or

Y'_b. That is, there is no statistically significant difference in predicting current belly dancing ability from past training vs. personality tests.

Steiger (1980) presents additional significance tests for situations where both the DV and the IVs are different, but from the same sample, and for comparing the difference between any two correlations within a correlation matrix.

5.6.3 Adjustment of R^2

Just as simple r_{xy} from a sample is expected to fluctuate around the value of the correlation in the population, sample R is expected to fluctuate around the population value. But multiple R never takes on negative value, so all chance fluctuations are in the positive direction and add to the magnitude of R. As in any sampling distribution, the magnitude of chance fluctuations increases as the sample size decreases. Therefore, R tends to be overestimated, and the smaller the sample the greater the overestimation. For this reason, in estimating the population value of R, adjustment is made for expected inflation in sample R.

All the programs discussed in Section 5.7 routinely provide adjusted R^2. Wherry (1931) provides a simple equation for this adjustment, which he calls $\bar{R}^2$:

$$\bar{R}^2 = 1 - (1 - R^2)\left(\frac{N - 1}{N - k - 1}\right) \tag{5.16}$$

where N is the sample size, k is the number of IVs, and R^2 is the squared multiple correlation. For the same sample problem,

$$\bar{R}^2 = 1 - (1 - .70235)(5/2) = .25588$$

as printed out for SPSS^x, Table 5.4.

For statistical regression, Cohen and Cohen (1975) recommend k based on the number of IVs considered for inclusion, rather than on the number of IVs selected by the program. They also suggest the convention of reporting $\bar{R}^2 = 0$ when the value spuriously becomes negative.

When the number of subjects is 60 or fewer and there are numerous IVs (say, more than 20), Equation 5.16 may provide inadequate adjustment for R^2. The adjusted value may be off by as much as .10 (Cattin, 1980). In these situations of small N and numerous IVs, Equation 5.17 (Browne, 1975) provides further adjustment.

$$R_s^2 = \frac{(N - k - 3)\,\bar{R}^4 + \bar{R}^2}{(N - 2k - 2)\,\bar{R}^2 + k} \tag{5.17}$$

The adjusted R^2 for small samples is a function of the number of cases, N, the number of IVs, k, and the $\bar{R}^2$ value as found from Equation 5.16.

When N is less than 50, Cattin (1980) provides an equation that produces even less bias but requires far more computation.

5.6.4 Suppressor Variables

Occasionally you may find an IV that is useful in predicting the DV and in increasing the multiple R^2 by virtue of its correlations with other IVs. This IV is called a suppressor variable because it "suppresses" variance that is irrelevant to prediction of the DV. In a full discussion of suppressor variables, Cohen and Cohen (1975) describe and provide examples of several varieties of suppression.

In psychology, for instance, one might administer as IVs two paper-and-pencil tests, a test of ability to list dance patterns and a test of test-taking ability. By itself the first test poorly predicts the DV (say, belly dancing ability). The second IV is a suppressor variable because it removes variance due to ability in taking tests, and thus enhances prediction of the DV by the first IV.

In output, the presence of a suppressor variable is identified by the pattern of regression coefficients and correlations of each IV with the DV. Compare the simple correlation between each IV and the DV in the correlation matrix available from output with the standardized regression coefficient (beta weight) for the IV. If the beta weight is significantly different from zero, either one of the following two conditions signals the presence of a suppressor variable: (1) the absolute value of the simple correlation between IV and DV is substantially smaller than the beta weight for the IV, or (2) the simple correlation and beta weight have opposite signs.

If you find that a suppressor variable is present, you need to search for it among IVs for which regression coefficients and correlations are congruent. You do this by systematically leaving each congruent IV out of the equation and examining the changes in regression coefficients for the IV(s) that had discrepancy between regression coefficient and correlation in the original equation.

Once suppressor variables are identified, they are properly interpreted as variables that enhance importance of other IVs by virtue of suppression of irrelevant variance in other IVs or in the DV.

5.7 COMPARISON OF PROGRAMS

The popularity of multiple regression is reflected in the abundance of applicable programs. BMDP and SAS each have three separate programs for standard, statistical, and setwise regression. SPSS^x has a single, highly flexible program for multiple regression. In SYSTAT, multiple regression is handled through a multivariate general linear model program.

Direct comparisons of multiple and statistical/hierarchical programs are summarized in Tables 5.11 and 5.12, respectively. Table 5.13 compares the two programs designed for setwise (subsets) regression. Some of these features are elaborated in Sections 5.7.1 through 5.7.4.

5.7.1 SPSS Package

The original SPSS program for regression was SPSS REGRESSION. Beginning with Release 9.0, the more elaborate NEW REGRESSION program became available (Hull

TABLE 5.11 COMPARISON OF PROGRAMS FOR STANDARD MULTIPLE REGRESSION

Feature	SPSS* REGRESSION[a]	BMDP1R	SYSTAT MGLH	SAS REG
Input				
Correlation matrix input	Yes	Yes	Yes	Yes
Covariance matrix input	Yes	Yes	Yes	Yes
Missing data options	Yes	Yes	No	No
Regression through the origin	Yes	No[b]	Yes	Yes
Tolerance option	Yes	Yes	No	Yes
Test of equality of subsamples	No	Yes	No	No
Post hoc hypotheses[c]	Yes	No	Yes	Yes
Optional error terms	No	No	Yes	No
Weighted least squares	No	Yes	Yes	Yes
Regression output				
Expanded variable labels	Yes	No	No	No
Analysis of variance for regression	Yes	Yes	Yes	Yes
Multiple R	Yes	Yes	Yes	No
R^2	Yes	Yes	Yes	Yes
Standard error for R	STANDARD ERROR	STD. ERROR OF ESTIMATE	STANDARD ERROR OF ESTIMATE	ROOT MSE
Adjusted R^2	Yes	Yes	Yes	Yes
Coefficient of variation	No	No	No	Yes
Correlation matrix	Yes	Yes	No[d]	Yes
Significance levels of correlation matrix	Yes	No	No	Yes
Sum-of-squares and cross-products (SSCP) matrix	Yes	No	No[d]	Yes
Inverse of SSCP matrix	No	No	No	Yes
Covariance matrix	Yes	Yes	No[d]	No
Means and standard deviations	Yes	Yes	No[e]	Yes
Matrix of correlation coefficients if some not computed	Yes	No[b]	N.A.	N.A.
N for each correlation coefficient	Yes	No[b]	N.A.	N.A.

Minimum and maximum of variables	No	Yes	No[e]	Yes
Sums for each variable	No	No	No	Yes
Coefficient of variation for each variable	No	Yes	No	No
Unstandarized regression coefficients	B	COEFFICIENT	COEFFICIENT	PARAMETER ESTIMATE
Standard error of regression coefficient	SE B	STD. ERROR	STD. ERROR	STANDARD ERROR
F or *t* test of regression coefficient	T(F optional)	T	T	T
Intercept (constant)	Yes	Yes	Yes	Yes
Standardized regression coefficient	BETA	STD. REG. COEFF.	STD. COEFF.	STANDARDIZED ESTIMATE
Approx. standard error of β	Yes	No	No	No
Partial correlation	Yes	No	Yes	Yes[f]
Semipartial correlation	PART COR	No	No	Yes[f]
Tolerance	Yes	Yes	Yes	Yes
Sums of squares for *B*	No	No	No	Yes[g]
Variance inflation factors	No	No	No	Yes
Variance-covariance matrix for unstandardized *B* coefficients	Yes	No	No	Yes
Correlation matrix for unstandardized *B* coefficients	No	Yes	Yes	Yes
95% confidence interval for *B*	Yes	No	No	No
Sweep matrix	Yes	No	No	Yes
Condition number bounds	Yes	No	No	No
Eigenvalues	No	No	Yes	Yes
Condition indices	No	No	Yes	Yes
Variance proportions	No	No	Yes	Yes
Hypothesis matrices	No	No	Yes	Yes
Residuals				
Predicted scores, residuals and standardized residuals	Yes	Yes	Yes[h]	Yes
95% confidence interval for predicted value	No	No	No	Yes[g]

TABLE 5.11 (Continued)

Feature	SPSS[x] REGRESSION[a]	BMDP1R	SYSTAT MGLH	SAS REG
Plot of standardized residuals against predicted scores	Yes	Yes	No[i]	No[j]
Normal plot of residuals	Yes	Yes	No[i]	No[j]
Durbin-Watson statistic	Yes	Yes	Yes	Yes
Standard error of predicted values	Yes	No	Yes[h]	Yes
Mahalanobis and/or Cook's distance	Yes	No[k]	Yes[h]	Yes
Influence diagnostics	No	No	No	Yes
Heteroscedasticity test	No	No	No	Yes
Histograms	Yes	No[l]	No[i]	No
Casewise plots	Yes	No	No	Yes
Partial plots	Yes[m]	Yes	No	Yes
Other plots available	Yes	Yes	No[i]	No[j]
Leverage	Yes	No	Yes[h]	No
Summary statistics for residuals	Yes	No	No	Yes
Save predicted values/residuals	Yes	Yes	Yes	Yes

Note: BMDP2R (cf. Table 5.12) and SAS GLM (cf. Table 9.11) can also be used for standard multiple regression.

[a] These features are available in SPSS NEW REGRESSION except as noted.

[b] Available through BMDPAM or BMDP8D.

[c] Does *not* use Larzelere and Mulaik (1977) correction.

[d] Available through SYSTAT CORR.

[e] Available through SYSTAT STATS.

[f] Use Type II for standard MR. Values in output are *squared* correlations.

[g] Optional types.

[h] Saved into a file.

[i] Available through SYSTAT GRAPH.

[j] Available through SAS PLOT.

[k] Available through BMDPAM, BMDP4M, or BMDP2R.

[l] Available through BMDP5D.

[m] Not available in SPSS NEW REGRESSION.

TABLE 5.12 COMPARISON OF PROGRAMS FOR STATISTICAL AND/OR HIERARCHICAL REGRESSION

Feature	SAS STEPWISE	SPSS* REGRESSION*	BMDP2R	SYSTAT MGLH
Input				
Correlation matrix input	No	Yes	Yes	Yes
Covariance matrix input	No	Yes	Yes	Yes
Missing data options	No	Yes	No[b]	No
Specify stepping algorithm	Yes	Yes	Yes	No
Specify tolerance	No	Yes	Yes	No
Specify F to enter	No	Yes	Yes	No
Specify F to remove	No	Yes	Yes	No
Specify probability of F to enter and/or remove	Yes	Yes	Yes	Yes
Specify maximum number of steps	No	Yes	Yes	No
Regression through the origin	No	Yes	Yes	Yes
Force variables into equation	INCLUDE	ENTER	FORCE	FORCE
Weighted least squares	No[c]	No	Yes	Yes
Specify order of entry (hierarchy)	No	Yes	Yes	No
IV sets for entry in single step	No	Yes	Yes	No
Regression output				
Expanded variable labels	No	Yes	No	No
Analysis of variance for regression, each step	Yes	Yes	Yes	No[d]
Multiple R, each step	No	Yes	Yes	No[d]
R^2, each step	R SQUARE	R SQUARE	MULTIPLE R-SQUARE	No[d]
Standard error for R, each step	No[c]	STANDARD ERROR	STD. ERROR OF EST.	No[d]
Adjusted R^2, each step	No	Yes	Yes	No[d]
Mallow's C_p, each step	Yes	No	No	No
Sum-of-squares and cross-products matrix	No[c]	Yes	No	No[c]
Correlation matrix	No[c]	Yes	Yes	No[c]

TABLE 5.12 (Continued)

Feature	SAS STEPWISE	SPSS[a] REGRESSION[a]	BMDP2R	SYSTAT MGLH
Correlation significance levels	No[c]	Yes	No	No
Covariance matrix	No	Yes	Yes	No
Correlation matrix of regression coefficients	No[c]	No	Yes	Yes
Covariance matrix of regression coefficients	No[c]	Yes	No	No
Means and standard deviations	No[c]	Yes	Yes	No
Matrix of correlation coefficients if some not computed	N.A.	Yes	No[b]	N.A.
N for each correlation coefficient	N.A.	Yes	No[b]	N.A.
Coefficients of variation	No	No	Yes	Yes
Minimums and maximums	No[c]	No	Yes	No
Smallest and largest *z*	No	No	Yes	No
Skewness and kurtosis	No	No	Yes	No
95% confidence interval for *B*	No	Yes	No	No
Sweep matrix	No[c]	Yes	No	No
Condition number bounds	Yes	Yes	No	No
Eigenvalues	No	No	No	Yes
Condition indices	No	No	No	Yes
Variance proportions	No	No	No	Yes
Variables in equation (each step)				
Unstandardized regression coefficients	B VALUE	B	COEFFICIENT	No[d]
Standard error of regression coefficient	STD ERROR B	SE B	STD. ERROR OF COEFF	No[d]
95% confidence interval for *B*	No	Yes	No	No
Standard regression coefficient	No[c]	BETA	STD REG COEFF	No[d]

F (or *T*) to remove	F	T (F optional)	F TO REMOVE	No
Intercept	INTERCEPT	(CONSTANT)	(Y-INTERCEPT)	No[d]
Variables not in equation (each step)				
Standardized regression coefficient for entering	No	BETA IN	No	No
Partial correlation coefficient for entering	No	PARTIAL	PARTIAL CORR.	No
Squared partial correlation for entering	MODEL R**2	No	No	No
Tolerance	Yes	Yes	Yes	No[d]
F (or *T*) to enter	F	T	F TO ENTER	No
Table of stepwise regression coefficients, *F* to enter and remove, and partial correlations	No	No	Yes	No[d]
Summary table				
Multiple *R*	No	MULTR	Yes	Yes
R^2	MODEL R**2	RSQ	MULTIPLE RSQ	RSQUARE
Change in R^2 (squared semipartial correlation)	PARTIAL R**2	RSQCH	CHANGE IN RSQ	No
Adjusted R^2	No	ADJRSQ	No	No
F to enter equation	No	F(EQU)	F TO ENTER	No
F to remove from equation	F	FCH	F TO REMOVE	No
Significance of *T* to enter or remove	Yes	Yes	No	No
Standardized regression coefficient	No	BETA IN	No	No
Mallow's C_p	Yes	No	No	No
Residuals				
Predicted scores, residuals, and standardized residuals	No[c]	Yes	Yes	Yes[f]
Plot of standardized residuals against predicted scores	No	Yes	Yes	No[g]
Normal plot of residuals	No	Yes	Yes	No[g]

TABLE 5.12 (Continued)

Feature	SAS STEPWISE	SPSS[a] REGRESSION[a]	BMDP2R	SYSTAT MGLH
Durbin-Watson statistic	No[c]	Yes	No	Yes
Standardized error of predicted values	No[c]	Yes	No	Yes[f]
Mahalanobis and/or Cook's distance	No[c]	Yes	Yes	Yes[f]
Summary statistics for residuals	No[c]	Yes	No	No
Leverage	No	Yes	Yes	Yes
Casewise plots	No[c]	Yes	Yes	No
Partial plots	No[c]	Yes	Yes	No
Save predicted values/residuals	No[c]	Yes	Yes	Yes
Misc. additional diagnostics	No	Yes	Yes	No
Other plots available	No	Yes	Yes	No[g]

[a] These features are available in SPSS NEW REGRESSION except as noted.

[b] Available through BMDPAM or BMDP8D.

[c] Available through SAS REG.

[d] Available by running separate standard multiple regressions for each step.

[e] Available through SYSTAT CORR.

[f] Saved to file.

[g] Available through SYSTAT GRAPH.

TABLE 5.13 COMPARISON OF PROGRAMS FOR SETWISE REGRESSION

Feature	BMDP9R	SAS RSQUARE
Input		
Correlation matrix input	Yes	Yes
Covariance matrix input	Yes	Yes
SSCP matrix input	No	Yes
Specify tolerance	Yes	No
Specify model criterion	Yes	No
Regression through the origin	Yes	Yes
Include variables in all models	No	Yes
Weighted least squares	Yes	Yes
Specify penalty for additional variables	Yes	No
Specify number of subsets	Yes	Yes
Specify minimum subset size	No	Yes
Specify maximum subset size	Yes	Yes
Regression output		
Means and standard deviations	Yes	Yes
Largest and smallest values	Yes	No
Skewness and kurtosis	Yes	No
Correlation matrix	Yes	Yes
Covariance matrix	Yes	No
Correlation of regression coefficients	Yes	No
Covariance of regression coefficients	Yes	No
Shaded, sorted correlation matrix	Yes	No
For each subset		
Variables in subset	Yes	Yes
R^2	Yes	Yes
Adjusted R^2	Yes	Yes
Model criteria	Yes[a]	Yes
Regression coefficients (B)	COEFFICIENT	PARAMETER ESTIMATES
T statistics for each variable	Yes	No
Residual mean square	Yes[b]	Yes[c]
Error sum of squares	No	Yes
For best subset (additional to above)		
Standard errors of B	Yes	No
Standardized coefficients (β)	Yes	No
Tolerances	Yes	No
Contributions to R^2 (sr^2)	Yes	No
Multiple R	Yes	No
F statistics with df and significance	Yes	No
Residuals		
Histograms of standardized residuals	Yes	No
Plots of residuals against predicted scores	Yes	No
Normal plot of residuals	Yes	No
Plots of variables against residuals	Yes	No

[a] Chosen criterion only.

[b] For "best" subset only.

[c] Several varieties.

and Nie, 1981). With few modifications, the latter program appears as REGRESSION in the SPSS^x series (SPSS, Inc., 1986). The distinctive feature of these programs is flexibility. The major improvement in SPSS^x is that the manual now approaches comprehensibility for this program. SPSS^x REGRESSION is the program summarized in Tables 5.11 and 5.12. With noted exceptions, the summary applies to SPSS NEW REGRESSION as well.

The SPSS programs offer four options for treatment of missing data (described in the manuals). Provision for expanded variable labels enhances recall of output that has not been looked at in a while. Data can be input raw or as correlation matrices. And as in all SPSS procedures, analysis can be limited to subsets of cases.

A special option is available so that correlation matrices are printed only when one or more of the correlations cannot be calculated. Also convenient is the optional printing of semipartial correlations for standard multiple regression and 95% confidence intervals for regression coefficients.

The statistical procedure offers forward, backward, and stepwise selection of variables. User-modifiable statistical criteria for statistical selection are (1) the maximum number of steps for model building, (2) the minimum F ratio for adding or maximum F for deleting a variable in the equation, (3) the minimum probability of F for adding or maximum probability of F for removing a variable in the equation, and (4) tolerance. Tolerance is the proportion of variance of a potential IV that is not explained by IVs already in the equation. The smaller the tolerance value, the less the restriction placed on entering variables. One avoids multicollinearity by maintaining reasonable tolerance levels for entry, and thereby disallowing entry of variables that add virtually nothing to predictability.

A series of ENTER subcommands can be used for hierarchical regression. Each ENTER subcommand is evaluated in order; the IV or IVs listed after each ENTER subcommand are evaluated in that order. Within a single subcommand, variables are entered in order of decreasing tolerance. If there is more than one IV in the subcommand, they are treated as a block in the summary table where changes in the equation are evaluated.

Extensive analysis of residuals is available. For example, a table of predicted scores and residuals can be requested and accompanied by a plot of standardized residuals against standardized predicted values of the DV (z scores of the Y' values). Plots of standardized residuals against sequenced cases are also available. For a sequenced file, one can request a Durbin-Watson statistic, which is used for a test of autocorrelation between adjacent cases. In addition, you can request Mahalanobis distance for the ten most extreme cases as a convenient way of evaluating outliers. This is the only program within the SPSS packages that offers Mahalanobis distance. Partial residual plots (partialing out all but one of the IVs) are available in SPSS^x REGRESSION but not in SPSS NEW REGRESSION.

The flexibility of input for the SPSS^x REGRESSION program does not carry over into output. The only difference between standard and statistical or hierarchical regression is in the printing of a progression of steps. Otherwise the statistics and parameter estimates are identical. These values, however, have different meanings, depending on the type of analysis. For example, you can request semipartial cor-

relations (called PART COR) through the ZPP statistics. But these apply only to standard multiple regression. As pointed out in Section 5.6.1, semipartial correlations for statistical or hierarchical analysis appear in the summary table (obtained through the HISTORY statistic) as RSQ CHANGE. It is best not to request a summary table for standard multiple regression because it contains no useful information that does not appear elsewhere.

5.7.2 BMD Series

The BMDP package (Dixon, 1985) offers separate programs for the various types of regression. All the programs do standard multiple regression, but the simplest applicable program is BMDP1R. For statistical and hierarchical regression, the program is BMDP2R. BMDP9R is available as a setwise program. All these programs exclude all cases with missing data, unless the values are estimated by, perhaps, some of the procedures available through BMDPAM or BMDP8D.

For standard multiple regression, BMDP1R offers analysis of "case combinations" or subsamples. Reduction of residuals due to grouping serves as a test of the equality of regression among the groups. Covariance and correlation matrices are available as output or can be used as input. In addition to tables of predicted values and residuals, the BMDP1R program allows a number of plots of residuals. Although not available in BMDP1R, Mahalanobis distance as well as other measures of extreme cases can be found using any one of several other BMDP programs, including BMDP2R.

The statistical/hierarchical regression program in the BMDP series, BMDP2R, combines several features of other statistical programs and adds some that are unavailable elsewhere. The user can specify the stepping algorithm (how the program adds or deletes variables in the equation as it progresses) in addition to specifying statistical criteria, as described in the BMDP manual (Dixon, 1985). Specifiable statistical criteria are tolerance, F to enter, F to remove, and maximum number of steps, as for SPSS. At the end of the output for all steps, a table of stepwise regression coefficients is available with the full regression equation for each step. The intercept (A) is given, along with unstandardized regression coefficients (B_i) for each of the variables in the equation at the next step. For each regression coefficient, F to enter and remove, and partial correlations are given.

In BMDP2R, the same input and output matrices are available as in BMDP1R. Residuals tables and plots, however, are far more extensive than BMDP1R. There are 21 diagnostic variables (including Mahalanobis distance) that can be printed or plotted, and 17 of these can be saved. For each plot, you can specify through options the number of extreme cases for which values are printed.

BMDP9R is a setwise program that searches among all possible subsets of variables for those subsets that are best. "Best" is defined by the user from among criteria such as maximum R^2. The researcher can specify the number of best subsets to be identified. Because of extensive output about the contribution of IVs to various subsets, BMDP9R is especially useful in identifying a combination of variables causing outlying cases to be unusual, as illustrated in Section 4.2.1.4. The extent

of residuals analysis in BMDP9R is somewhere between that for BMDP1R and BMDP2R. A more extensive list of residuals and other diagnostics (including Mahalanobis distance) can be saved to a file than is printed through BMDP9R.

Cross validation is simplified in BMDP2R and BMDP9R by a case-weighting procedure, where cases to be omitted from computation of the regression equation are given a weight of zero.

BMDP offers a number of additional programs for regression analysis. These cover procedures such as nonlinear regression, analysis of variance and covariance, periodic regression, and asymptotic regression.

5.7.3 SAS System

Like BMDP, SAS has separate programs for different types of regression analysis (SAS Institute, Inc., 1985). Standard multiple regression is handled through REG, statistical (though not hierarchical) through STEPWISE, and setwise through RSQUARE. In addition, GLM can be used for regression analysis; it is more flexible and powerful than any of the other SAS programs, but also more difficult to use.

SAS REG handles correlation or covariance matrix input but has no options for dealing with missing data. A case is deleted if it contains any missing values.

In SAS REG, two types of partial and semipartial correlations are available. But beware; the value reported as SEMI-PARTIAL CORR is actually squared semipartial correlation (sr^2). The sr_i^2 and pr_i^2 appropriate for standard multiple regression are those that use TYPE II (partial) sums of squares (which can also be printed). TYPE I sums of squares are used for evaluating hierarchical models where order of entry of IVs follows order in the MODEL statement. Since STEPWISE does not handle hierarchical regression, SAS REG is used for hierarchical analysis within SAS.

Tables of residuals and other diagnostics (including Cook's value but not Mahalanobis distance) are extensive but are saved to file rather than printed. After being saved to file, the diagnostics and residuals can be plotted through SAS PLOT. The only plots printed within REG are partial residual plots where all IVs except one are partialed out.

SAS STEPWISE is a limited program for statistical multiple regression. The program does not accept matrix input, nor can order of entry of IVs be specified. The usual forward, backward, and stepwise criteria for selection are available, in addition to two enhancements on forward selection. The only statistical criteria available are the probability of F to enter and F to remove an IV from the equation. Summary table information is adequate but misleading; squared semipartial correlations are labeled PARTIAL R**2.

Few of the optional features of SAS REG are available in STEPWISE. For example, descriptive statistics and matrices cannot be printed out. One useful feature is Mallow's C_p which, as described in the manual, can be used as a criterion for selecting a model produced by statistical regression. No diagnostics or residuals are available.

SAS RSQUARE does setwise regression but differs in several ways from BMDP9R. Instead of specifying a criterion for selecting subsets, RSQUARE evaluates

all subsets, and then, for each subset, prints out values for the various optimizing criteria. No "best" subset is identified. You make the choice of the best subset by comparing values on your chosen criterion. You can reduce output by specifying the number of subsets to be evaluated as well as minimum and maximum subset sizes.

In SAS RSQUARE, descriptive statistics and matrix output are limited. The only information given about each IV in a subset is the regression coefficient (B_i). As a result, the program is not useful for identifying the combination of IVs on which outlying cases differ. No residuals/diagnostics analysis is available.

5.7.4 SYSTAT System

In SYSTAT (Wilkinson, 1986), multiple regression is done through MGLH, a multivariate general linear program that handles ANOVA, MANOVA, canonical correlation, discriminant function analysis, and several types of regression. While SYSTAT MGLH has a variety of features for standard multiple regression, handling of statistical regression is limited, and there is no direct way to specify hierarchical or setwise regression.

Matrix input is accepted in SYSTAT MGLH, but processing of output of matrices or descriptive statistics requires the use of other programs in the SYSTAT package. The program is especially convenient for testing post hoc hypotheses, allowing you to specify your own error terms should you be unhappy with the ones normally provided. Residuals and other diagnostics are handled by saving values to a file, which can then be printed out through SYSTAT DATA or plotted through SYSTAT GRAPH. In this way, you can take a look at Cook's value, but not Mahalanobis distance, for each case. Or, you can find summary values for residuals and diagnostics through SYSTAT STATS, the program for descriptive statistics.

For statistical regression, SYSTAT MGLH provides a summary table and instructions to use the chosen predictors in a model to estimate coefficients. If you follow the instructions, you have a standard multiple regression at the last step. If you want to see results at any other step, you request a standard multiple regression for that step. The only stepping criterion available is the one that other programs label STEPWISE.

Although you can force inclusion of some IVs in the equation, you cannot specify their order of entry. To do hierarchical analysis, then, you must run a separate standard multiple regression for each step in the hierarchy and then compute summary statistics (cf. Sections 5.6.1 and 5.6.2).

5.8 COMPLETE EXAMPLES OF REGRESSION ANALYSIS

To illustrate applications of regression analysis, variables are chosen from among those measured in the research described in Appendix B, Section B.1. Two analyses are reported here, both with number of visits to health professionals (TIMEDRS) as the DV and both using the SPSS^x REGRESSION program.

The first example is a standard multiple regression between the DV and num-

ber of physical health symptoms (PHYHEAL), number of mental health symptoms (MENHEAL), and stress from acute life changes (STRESS). From this analysis, one can assess the degree of relationship between the DV and IVs, the proportion of variance in the DV predicted by regression, and the relative importance of the various IVs to the solution.

The second example demonstrates hierarchical regression with the same DV and IVs. The first step of the analysis is entry of PHYHEAL to determine how much variance in number of visits to health professionals can be accounted for by differences in physical health. The second step is entry of STRESS to determine if there is a significant increase in R^2 when differences in stress are added to the equation. The final step is entry of MENHEAL to determine if differences in mental health are related to number of visits to health professionals after differences in physical health and stress are statistically accounted for.

5.8.1 Evaluation of Assumptions

Because both analyses use the same variables, this evaluation is appropriate for both.

5.8.1.1 Ratio of Cases to IVs With 465 respondents and 3 IVs, the cases-to-IV ratio is 155:1, well above the minimum requirements for regression. There are no missing data.

5.8.1.2 Normality, Linearity, Homoscedasticity, and Independence of Residuals In these examples, we choose to examine the residuals to see if screening is necessary, rather than screening prior to examination of residuals. The initial run through SPSS^x REGRESSION uses untransformed variables in a standard multiple regression to produce the scatterplot of residuals against predicted DV scores that appears in Figure 5.4.

Notice the execrable overall shape of the scatterplot that violates all the assumptions regarding distribution of residuals. Comparison of Figure 5.4 with Figure 5.1(a) (in Section 5.3.2.4) suggests further analysis of the distributions of the variables. (It was noticed in passing, although we tried not to look, that R^2 for this analysis was significant, but only .22).

SPSS^x FREQUENCIES is used to examine the distributions of the variables, as seen in Table 5.14. All the variables have significant positive skewness (see Chapter 4), which explains, at least in part, the problems in the residuals scatterplot. Logarithmic and square root transformations are applied as appropriate, and the transformed distributions checked once again for skewness. Thus TIMEDRS and PHYHEAL, with logarithmic transformations, become LTIMEDRS and LPHYHEAL, and STRESS, with square root transformation, becomes SSTRESS. In the case of MENHEAL, application of the milder square root transformation makes the variable significantly negatively skewed, so no transformation is undertaken.

Table 5.15 shows output from FREQUENCIES for one of the transformed variables, LPHYHEAL, the worst prior to transformation. Transformations similarly reduce skewness in the other two transformed variables.

The residuals scatterplot from SPSS^x REGRESSION following regression with

```
TITLE          LARGE SAMPLE REGRESSION-OUTLIERS & RESIDUALS
FILE HANDLE    REGRESS
DATA LIST      FILE=REGRESS      FREE
               /SUBJNO,TIMEDRS,PHYHEAL,MENHEAL,STRESS
VAR LABELS     TIMEDRS 'VISITS TO HEALTH PROFESSIONALS'
REGRESSION     DESCRIPTIVES/
               VARS=TIMEDRS TO STRESS/
               DEP=TIMEDRS/ENTER/
               RESIDUALS=OUTLIERS(MAHAL)/
               SCATTERPLOT (*RESID,*PRED)/
```

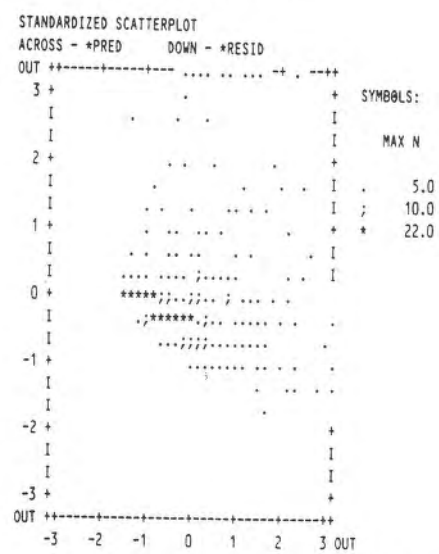


Figure 5.4 SPSS* REGRESSION setup and residuals scatterplot for original variables.

the transformed variables appears as Figure 5.5. Notice that, although the scatterplot is still not perfectly rectangular, its shape is considerably improved over that in Figure 5.4.

5.8.1.3 Outliers Multivariate outliers are sought using the transformed IVs as part of an SPSS* REGRESSION run in which the Mahalanobis distance of each case to the centroid of all cases is computed. The ten cases with the largest distance are printed (see Table 5.16). Mahalanobis distance is distributed as a chi square (χ^2) variable, with degrees of freedom equal to the number of IVs. To determine which cases are multivariate outliers, one looks up critical χ^2 at the desired alpha level (Table C.4). In this case, critical χ^2 at $\alpha = .001$ for 3 df is 16.266. Any case with a value larger than 16.266 in the *MAHAL column of the output is a multivariate outlier among the IVs. None of the cases has a value in excess of 16.266. (If outliers are found, the procedures detailed in Chapter 4 are followed to reduce their influence.)

5.8.1.4 Multicollinearity and Singularity The REGRESSION run of Table 5.17 resolves doubts about possible multicollinearity and singularity among the transformed

TABLE 5.14 SETUP AND OUTPUT FOR EXAMINING DISTRIBUTIONS OF VARIABLES THROUGH SPSS' FREQUENCIES

```
TITLE          LARGE SAMPLE REGRESSION-FREQUENCIES
FILE HANDLE    REGRESS
DATA LIST      FILE=REGRESS      FREE
                 /SUBJNO,TIMEDRS,PHYHEAL,MENHEAL,STRESS
FREQUENCIES    VARIABLES=TIMEDRS TO STRESS/
               HISTOGRAM/
               FORMAT=NOTABLE/
               STATISTICS=ALL/


TIMEDRS   VISITS TO HEALTH PROFESSIONALS

     COUNT   MIDPOINT     ONE SYMBOL EQUALS APPROXIMATELY  4.00 OCCURRENCES

        86          0  **********************
       187          4  ************************************************
        74          8  *******************
        44         12  ***********
        27         16  *******
        13         20  ***
         7         24  **
         6         28  **
         3         32  *
         4         36  *
         2         40  *
         1         44
         1         48
         2         52  *
         3         56  *
         3         60  *
         0         64
         0         68
         0         72
         1         76
         1         80
                   I....+....I....+....I....+....I....+....I....+....I....+....I
                   0         40         80        120        160        200
                             HISTOGRAM FREQUENCY

MEAN          7.901      STD ERR        .508      MEDIAN        4.000
MODE          2.000      STD DEV      10.948      VARIANCE    119.870
KURTOSIS     13.101      S E KURT      1.996      SKEWNESS      3.248
S E SKEW       .113      RANGE        81.000      MINIMUM           0
MAXIMUM      81.000      SUM        3674.000

VALID CASES    465      MISSING CASES     0


PHYHEAL

     COUNT     VALUE     ONE SYMBOL EQUALS APPROXIMATELY  2.00 OCCURRENCES

        59      2.00  *****************************
        93      3.00  ***********************************************
        78      4.00  ***************************************
        69      5.00  **********************************
        64      6.00  ********************************
        33      7.00  ****************
        29      8.00  **************
        18      9.00  *********
         8     10.00  ****
         6     11.00  ***
         4     12.00  **
         2     13.00  *
         1     14.00  *
         1     15.00  *
                   I.........I.........I.........I.........I.........I.........I
                   0        20        40        60        80        100
                             HISTOGRAM FREQUENCY

MEAN          4.972      STD ERR        .111      MEDIAN        5.000
MODE          3.000      STD DEV      2.388       VARIANCE      5.704
KURTOSIS      1.124      S E KURT      1.996      SKEWNESS      1.031
S E SKEW       .113      RANGE        13.000      MINIMUM       2.000
MAXIMUM      15.000      SUM        2312.000

VALID CASES    465      MISSING CASES     0
```

TABLE 5.14 (Continued)

```
MENHEAL

    COUNT       VALUE     ONE SYMBOL EQUALS APPROXIMATELY  1.00 OCCURRENCE

       27 ✓        0   ****************************
       41        1.00  ******************************************
       35        2.00  ************************************
       37        3.00  **************************************
       46        4.00  ***************************************************
       45        5.00  **************************************************
       48        6.00  ****************************************************
       33        7.00  **********************************
       31        8.00  ********************************
       20        9.00  ********************
       24       10.00  *************************
       15       11.00  ***************
       23       12.00  ************************
       13       13.00  *************
        6       14.00  ******
        9       15.00  *********
        7       16.00  *******
        2       17.00  **
        3       18.00  ***
                    I.........I.........I.........I.........I.........I
                    0         10        20        30        40        50
                              HISTOGRAM FREQUENCY

MEAN          6.123      STD ERR        .194      MEDIAN        6.000
MODE          6.000      STD DEV       4.194      VARIANCE     17.586
KURTOSIS      -.292      S E KURT      1.996      SKEWNESS       .602
S E SKEW      .113       RANGE        18.000      MINIMUM           0
MAXIMUM      18.000      SUM        2847.000

VALID CASES    465       MISSING CASES     0

STRESS

    COUNT   MIDPOINT     ONE SYMBOL EQUALS APPROXIMATELY  1.50 OCCURRENCES

       36         20   *************************
       63         64   ******************************************
       58        108   ***************************************
       65        152   ********************************************
       53        196   ************************************
       48        240   ********************************
       41        284   ****************************
       38        328   *************************
       21        372   **************
       16        416   ***********
        6        460   ****
        5        504   ***
        8        548   *****
        3        592   **
        1        636   *
        0        680
        1        724   *
        0        768
        0        812
        0        856
        1        900   *
                    I....+....I....+....I....+....I....+....I....+....I
                    0         15        30        45        60        75
                              HISTOGRAM FREQUENCY

MEAN        204.217      STD ERR       6.297      MEDIAN      178.000
MODE              0      STD DEV     135.793      VARIANCE  18439.662
KURTOSIS      1.801      S E KURT      1.996      SKEWNESS      1.043
S E SKEW       .113      RANGE       920.000      MINIMUM           0
MAXIMUM     920.000      SUM       94961.000

VALID CASES    465       MISSING CASES     0
```

TABLE 5.15 SETUP AND OUTPUT FOR EXAMINING DISTRIBUTION OF TRANSFORMED VARIABLE THROUGH SPSS* FREQUENCIES

```
TITLE          LARGE SAMPLE REGRESSION-FREQUENCIES
FILE HANDLE    REGRESS
DATA LIST      FILE=REGRESS      FREE
                 /SUBJNO,TIMEDRS,PHYHEAL,MENHEAL,STRESS
VAR LABELS     TIMEDRS 'VISITS TO HEALTH PROFESSIONALS'
COMPUTE        LTIMEDRS = LG10(TIMEDRS + 1)
COMPUTE        LPHYHEAL = LG10(PHYHEAL)
COMPUTE        SSTRESS = SQRT(STRESS)
FREQUENCIES    VARIABLES=LTIMEDRS,LPHYHEAL,SSTRESS/
               HISTOGRAM/
               FORMAT=NOTABLE/
               STATISTICS=ALL/
```

```
LTIMEDRS

     COUNT   MIDPOINT    ONE SYMBOL EQUALS APPROXIMATELY  1.50 OCCURRENCES

        42         0  ****************************
         0         .1
         0         .2
        44         .3  *****************************
         0         .4
        67         .5  *********************************************
        52         .6  ***********************************
        41         .7  ***************************
        53         .8  ************************************
        25         .9  ****************
        34        1.0  ***********************
        33        1.1  **********************
        24        1.2  ****************
        16        1.3  ***********
        11        1.4  *******
         6        1.5  ****
         6        1.6  ****
         3        1.7  **
         6        1.8  ****
         2        1.9  *
         0        2.0
                  I....+....I....+....I....+....I....+....I....+....I....+....I
                  0         15        30        45        60        75
                            HISTOGRAM FREQUENCY

MEAN          .741     STD ERR       .019     MEDIAN        .699
MODE          .477     STD DEV       .415     VARIANCE      .172
KURTOSIS     -.177     S E KURT     1.996     SKEWNESS      .228
S E SKEW      .113     RANGE        1.914     MINIMUM          0
MAXIMUM      1.914     SUM        344.698
```

IVs. All variables enter the equation without violating the default value for tolerance (cf. Chapter 4). Further, the highest correlation, between MENHEAL and LPHYHEAL, is .511. (If multicollinearity is indicated, SMCs between each IV serving as DV with the other two as IVs are found through REGRESSION or any of the other appropriate programs and redundant IVs dealt with as discussed in Chapter 4.)

5.8.2 Standard Multiple Regression

SPSS* REGRESSION is used to compute a standard multiple regression between LTIMEDRS (the transformed DV), and MENHEAL, LPHYHEAL, and SSTRESS (the IVs).

The output of this REGRESSION run appears as Table 5.17. Included are descriptive statistics, the values of R, R^2, and adjusted R^2, and a summary of the analysis of variance for regression. The significance level for R is found in the source table with $F(3, 461) = 92.90$, $p < .001$. In the portion labeled VARIABLES IN

Figure 5.5 Residuals scatterplot following regression with transformed variables. Output from SPSS[x] REGRESSION. See Table 5.16 for setup.

THE EQUATION are printed unstandardized and standardized regression coefficients with their significance levels and 95% confidence intervals, and three correlations: total (CORREL), semipartial (PART COR), and partial.

The significance levels for the regression coefficients (see Table 5.17, VARIABLES IN THE EQUATION) are evaluated against 1 and 461 df. Only two of the IVs, SSTRESS and LPHYHEAL, contribute significantly to regression, with F values

TABLE 5.16 SETUP AND OUTPUT FROM SPSS[x] REGRESSION SHOWING MULTIVARIATE OUTLIERS

```
TITLE          LARGE SAMPLE STANDARD MULTIPLE REGRESSION
FILE HANDLE    REGRESS
DATA LIST      FILE=REGRESS      FREE
                 /SUBJNO,TIMEDRS,PHYHEAL,MENHEAL,STRESS
VAR LABELS     TIMEDRS 'VISITS TO HEALTH PROFESSIONALS'
COMPUTE        LTIMEDRS = LG10(TIMEDRS + 1)
COMPUTE        LPHYHEAL = LG10(PHYHEAL)
COMPUTE        SSTRESS = SQRT(STRESS)
REGRESSION     DESCRIPTIVES/
               VARS=MENHEAL,LTIMEDRS TO SSTRESS/
               STATISTICS=DEFAULTS,ZPP,CI,F/
               DEP=LTIMEDRS/ENTER/
               RESIDUALS=OUTLIERS(MAHAL)/
               SCATTERPLOT (*RESID,*PRED)/
```

```
OUTLIERS - MAHALANOBIS' DISTANCE

  CASE #        *MAHAL

    403       14.13530
    125       11.64940
    198       10.56870
     52       10.54759
    446       10.22483
    159        9.35066
     33        8.62835
    280        8.58686
    405        8.43122
    113        8.35261
```

TABLE 5.17 STANDARD MULTIPLE REGRESSION ANALYSIS OF LTIMEDRS (THE DV) WITH MENHEAL, SSTRESS, AND LPHYHEAL (THE IVs). SELECTED OUTPUT. SETUP SHOWN IN TABLE 5.16

```
              MEAN   STD DEV  LABEL

MENHEAL      6.123     4.194
LTIMEDRS      .741      .415
LPHYHEAL      .648      .206
SSTRESS     13.400     4.972

N OF CASES =    465

CORRELATION:

            MENHEAL   LTIMEDRS   LPHYHEAL    SSTRESS

MENHEAL       1.000       .355       .511       .383
LTIMEDRS       .355      1.000       .586       .359
LPHYHEAL       .511       .586      1.000       .317
SSTRESS        .383       .359       .317      1.000


          * * * *  M U L T I P L E   R E G R E S S I O N  * * * *

EQUATION NUMBER 1    DEPENDENT VARIABLE..  LTIMEDRS

BEGINNING BLOCK NUMBER  1.  METHOD:  ENTER

VARIABLE(S) ENTERED ON STEP NUMBER  1..    SSTRESS
                                    2..    LPHYHEAL
                                    3..    MENHEAL


MULTIPLE R            .61382        ANALYSIS OF VARIANCE
R SQUARE              .37678                    DF      SUM OF SQUARES      MEAN SQUARE
ADJUSTED R SQUARE     .37272    REGRESSION       3          30.14607          10.04869
STANDARD ERROR        .32888    RESIDUAL       461          49.86410           .10817

                                F =      92.90143      SIGNIF F = .0000

---------------------------------- VARIABLES IN THE EQUATION ----------------------------------

                                    95 CONFDNCE                          PART
VARIABLE          B        SE B      INTRVL BB      BETA     CORREL       COR    PARTIAL         F    SIG F

SSTRESS       .01571      .00336    .00910    .02232    .18811    .35905    .17173    .21256    21.815    .0000
LPHYHEAL     1.03997      .08718    .86864   1.21130    .51641    .58575    .43859    .48565   142.289    .0000
MENHEAL      .00188       .00440   -.00677    .01053    .01902    .35513    .01573    .01993     .183    .6689
(CONSTANT)  -.15504       .05826   -.26952   -.04056                                            7.083    .0081

FOR BLOCK NUMBER  1   ALL REQUESTED VARIABLES ENTERED.
```

of 21.82 and 142.29, respectively. The 95% confidence intervals for these IVs also do not contain zero as similar indications of their significance.

Semipartial correlations (cf. Equation 5.9) are labeled PART COR in the VARIABLES IN THE EQUATION section. These values, when squared, indicate the amount by which R^2 would be reduced if an IV were omitted from the equation. The sum for the two significant IVs ($.17173^2 + .43859^2 = .2219$) is the amount of R^2 attributable to unique sources. The difference between R^2 and unique variance ($.37678 - .2219 = .1548$) represents variance that SSTRESS, LPHYHEAL, and MENHEAL jointly contribute to R^2.

Information from this analysis is summarized in Table 5.18, in a form that might be appropriate for publication in a professional journal.

It is noted from the correlation matrix in Table 5.17 that MENHEAL correlates with LTIMDRS ($r = .355$) but does not contribute significantly to regression. If

TABLE 5.18 STANDARD MULTIPLE REGRESSION OF HEALTH AND STRESS VARIABLES ON NUMBER OF VISITS TO HEALTH PROFESSIONALS

Variables	LTIMEDRS (DV)	LPHYHEAL	SSTRESS	MENHEAL	*B*	β	sr^2 (unique)
LPHYHEAL	.59				1.040**	0.52	.19
SSTRESS	.36	.32			0.016**	0.19	.03
MENHEAL	.36	.51	.38		0.002	0.02	
				Intercept = − 0.155			
Means	0.74	0.65	13.40	6.12			
Standard deviations	0.42	0.21	4.97	4.19			
						R^2 = .38[a]	
						Adjusted R^2 = .37	
						R = .61**	

** $p < .01$.

[a] Unique variability = .22; shared variability = .16.

TABLE 5.19 CHECKLIST FOR STANDARD MULTIPLE REGRESSION

1. Issues
   a. Ratio of cases to IVs
   b. Normality, linearity, and homoscedasticity of residuals
   c. Outliers
   d. Multicollinearity and singularity
2. Major analyses
   a. Multiple R^2, F ratio
   b. Adjusted multiple R^2, overall proportion of variance accounted for
   c. Significance of regression coefficients
   d. Squared semipartial correlations
3. Additional analyses
   a. Post hoc significance of correlations
   b. Unstandardized (B) weights, confidence limits
   c. Standardized (β) weights
   d. Unique versus shared variability
   e. Suppressor variables
   f. Prediction equation

Equation 5.11 is used a posteriori to evaluate the significance of the correlation coefficient,

$$F = \frac{(.355)^2/3}{(1 - .355^2)/(465 - 3 - 1)} = 22.16$$

the correlation between MENHEAL and LTIMDRS differs reliably from zero; $F(3, 461) = 22.16$, $p < .01$.

Thus, although the bivariate correlation between MENHEAL and LTIMEDRS is reliably different from zero, the relationship seems to be mediated by, or redundant to, the relationship between LTIMDRS and other IVs in the set. Had the researcher measured only MENHEAL and LTIMEDRS, however, the significant correlation might have led to stronger conclusions than are warranted about the relationship between mental health and number of visits to health professionals.

Table 5.19 contains a checklist of analyses performed in standard multiple regression. An example of a Results section in journal format appears next.

Results

A standard multiple regression was performed between number of visits to health professionals as the dependent variable and physical health, mental health, and stress as independent variables. Analysis was performed using SPSS* REGRESSION with an

assist from SPSS* FREQUENCIES for evaluation of assumptions.

Results of evaluation of assumptions led to transformation of the variables to reduce skewness in their distributions, reduce the number of outliers, and improve the normality, linearity, and homoscedasticity of residuals. A square root transformation was used on the measure of stress. Logarithmic transformations were used on number of visits to health professionals and on physical health. One IV, mental health, was positively skewed without transformation and negatively skewed with it; it was not transformed. With the use of a $p < .001$ criterion for Mahalanobis distance no outliers among the cases were found. No cases had missing data and no suppressor variables were found, N = 465.

Table 5.18 displays the correlations between the variables, the unstandardized regression coefficients (B) and intercept, the standardized regression coefficients (β), the semipartial correlations (sr^2) and R, R^2, and adjusted R^2. R for regression was significantly different from zero, $F(3, 461) = 92.90$, $p < .001$. For the two regression coefficients that differed significantly from zero, 95% confidence limits were calculated. The confidence limits for square root of stress were 0.0091 to 0.0223, and those for log of physical health were 0.8686 to 1.2113.

Insert Table 5.18 about here

Only two of the IVs contributed significantly to prediction of number of visits to health professionals as logarithmically transformed, log of physical health scores ($sr^2 = .19$) and square root of acute stress scores ($sr^2 = .03$). The three IVs in combination

contributed another .15 in shared variability. Altogether, 38% (37% adjusted) of the variability in visits to health professionals was predicted by knowing scores on these three IVs.

Although the correlation between log of visits to health professionals and mental health was .36, mental health did not contribute significantly to regression. Post hoc evaluation of the correlation revealed that it was significantly different from zero, $F(3, 461) = 22.16$, $p < .01$. Apparently the relationship between the number of visits to health professionals and mental health is an indirect result of the relationships between physical health, stress, and visits to health professionals.

5.8.3 Hierarchical Regression

The second example involves the same three IVs entered one at a time in an order determined by the researcher. LPHYHEAL is the first IV to enter, followed by SSTRESS and then MENHEAL. The main research question is whether or not information regarding differences in mental health can be used to predict visits to health professionals after differences in physical health and in acute stress are statistically eliminated. In other words, do people go to health professionals for more numerous mental health symptoms if they have physical health and stress similar to other people?

Table 5.20 shows setup and selected portions of the output for hierarchical analysis using the SPSS^x REGRESSION program. Notice that a complete regression solution is provided at the end of each of steps 1 through 3. The significance of the bivariate relationship between LTIMEDRS and LPHYHEAL is assessed at the end of step 1, $F(1, 463) = 241.83$, $p < .001$. The bivariate correlation is .59, accounting for 34% of the variance. After step 2, with both LPHYHEAL and SSTRESS in the equation, $F(2, 462) = 139.51$, $p < .01$, $R = .61$, and $R^2 = .38$. With the addition of MENHEAL, $F(3, 461) = 92.90$, $R = .61$, and $R^2 = .38$. Increments in R^2 at each step are read directly from the RSQCH column of the SUMMARY TABLE. Thus, $sr^2_{\text{LPHYHEAL}} = .34$ $sr^2_{\text{SRAHE}} = .03$, and $sr^2_{\text{MENHEAL}} = .00$

Significance of the addition of SSTRESS to the equation is indicated in the output at STEP NUMBER 2 (the step where SSTRESS entered) as F for SSTRESS in the segment labeled VARIABLES IN THE EQUATION. Because the F value of 24.772 exceeds critical F with 1 and 461 df, SSTRESS is making a significant contribution to the equation at this step.

Similarly, significance of the addition of MENHEAL to the equation is indicated

at STEP NUMBER 3, where the F for MENHEAL is .183. Because this F value does not exceed critical F with 1 and 461 df, MENHEAL is not significantly improving R^2 at its point of entry.

The significance levels of the squared semipartial correlations are also available in the summary table as FCH, with probability value SIGCH for evaluating the significance of the added IV.

Thus there is no reliable increase in prediction of LTIMEDRS by addition of MENHEAL to the equation if differences in LPHYHEAL and SSTRESS are already accounted for. Apparently the answer is no to the question: Do people go to health professionals for more numerous mental health symptoms if they have physical health and stress similar to others? A summary of information from this output appears in Table 5.21.

Table 5.22 is a checklist of items to consider in hierarchical regression. An example of a Results section in journal format appears next.

TABLE 5.20 SETUP AND SELECTED OUTPUT FOR SPSS* HIERARCHICAL REGRESSION

```
            TITLE          LARGE SAMPLE HIERARCHICAL MULTIPLE REGRESSION
            FILE HANDLE    REGRESS
            DATA LIST      FILE=REGRESS      FREE
                             /SUBJNO,TIMEDRS,PHYHEAL,MENHEAL,STRESS
            VAR LABELS     TIMEDRS 'VISITS TO HEALTH PROFESSIONALS'
            COMPUTE        LTIMEDRS = LG10(TIMEDRS + 1)
            COMPUTE        LPHYHEAL = LG10(PHYHEAL)
            COMPUTE        SSTRESS = SQRT(STRESS)
            REGRESSION     VARS=MENHEAL,LTIMEDRS TO SSTRESS/
                           STATISTICS=DEFAULTS,HISTORY,CI,F/
                           DEP=LTIMEDRS/ENTER LPHYHEAL/
                                        ENTER SSTRESS/
                                        ENTER MENHEAL/


VARIABLE LIST NUMBER 1   LISTWISE DELETION OF MISSING DATA

EQUATION NUMBER 1    DEPENDENT VARIABLE..  LTIMEDRS

BEGINNING BLOCK NUMBER  1.  METHOD:  ENTER       LPHYHEAL


VARIABLE(S) ENTERED ON STEP NUMBER  1..     LPHYHEAL

MULTIPLE R            .58575          ANALYSIS OF VARIANCE
R SQUARE              .34310                          DF      SUM OF SQUARES      MEAN SQUARE
ADJUSTED R SQUARE     .34168          REGRESSION       1          27.45151          27.45151
STANDARD ERROR        .33692          RESIDUAL       463          52.55866           .11352

                                      F =     241.82595      SIGNIF F =  .0000


------------------------------ VARIABLES IN THE EQUATION ------------------------------

VARIABLE              B         SE B     95 CONFDNCE INTRVL B       BETA          F  SIG F

LPHYHEAL         1.17961      .07586      1.03055     1.32868      .58575   241.826  .0000
(CONSTANT)       -.02354      .05160      -.12485      .07786                 .208  .6484

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE     BETA IN  PARTIAL  MIN TOLER         F  SIG F

MENHEAL      .07528   .07982     .73850     2.963  .0859
SSTRESS      .19278   .22559     .89956    24.772  .0000

FOR BLOCK NUMBER  1   ALL REQUESTED VARIABLES ENTERED.
```

TABLE 5.20 (Continued)

```
EQUATION NUMBER 1    DEPENDENT VARIABLE..  LTIMEDRS

BEGINNING BLOCK NUMBER  2.  METHOD:  ENTER       SSTRESS


VARIABLE(S) ENTERED ON STEP NUMBER  2..    SSTRESS

MULTIPLE R            .61362       ANALYSIS OF VARIANCE
R SQUARE              .37653                    DF       SUM OF SQUARES      MEAN SQUARE
ADJUSTED R SQUARE     .37383       REGRESSION    2           30.12626          15.06313
STANDARD ERROR        .32859       RESIDUAL    462           49.88391           .10797

                                   F =     139.50724      SIGNIF F = .0000


---------- VARIABLES IN THE EQUATION ----------

VARIABLE          B        SE B     95 CONFDNCE INTRVL B      BETA          F    SIG F

LPHYHEAL       1.05658     .07800      .90330     1.20986     .52465    183.486   .0000
SSTRESS         .01610     .00323      .00974      .02246     .19278     24.772   .0000
(CONSTANT)     -.15950     .05726     -.27203     -.04697                 7.758   .0056


---------- VARIABLES NOT IN THE EQUATION ----------

VARIABLE     BETA IN  PARTIAL  MIN TOLER        F   SIG F

MENHEAL       .01902   .01993     .68426      .183   .6689


FOR BLOCK NUMBER  2   ALL REQUESTED VARIABLES ENTERED.

* * * *  M U L T I P L E   R E G R E S S I O N  * * * *

EQUATION NUMBER 1    DEPENDENT VARIABLE..  LTIMEDRS

BEGINNING BLOCK NUMBER  3.  METHOD:  ENTER       MENHEAL


VARIABLE(S) ENTERED ON STEP NUMBER  3..    MENHEAL

MULTIPLE R            .61382       ANALYSIS OF VARIANCE
R SQUARE              .37678                    DF       SUM OF SQUARES      MEAN SQUARE
ADJUSTED R SQUARE     .37272       REGRESSION    3           30.14607          10.04869
STANDARD ERROR        .32888       RESIDUAL    461           49.86410           .10817

                                   F =      92.90143      SIGNIF F = .0000


---------- VARIABLES IN THE EQUATION ----------

VARIABLE          B        SE B     95 CONFDNCE INTRVL B      BETA          F    SIG F

LPHYHEAL       1.03997     .08718      .86864     1.21130     .51641    142.289   .0000
SSTRESS         .01571     .00336      .00910      .02232     .18811     21.815   .0000
MENHEAL         .00188     .00440     -.00677      .01053     .01902      .183    .6689
(CONSTANT)     -.15504     .05826     -.26952     -.04056                 7.083   .0081


FOR BLOCK NUMBER  3   ALL REQUESTED VARIABLES ENTERED.

* * * * * * * * * * * * * * * * * * * * * * * * * * * * *

                                   SUMMARY TABLE
                                   -------------

STEP   MULTR    RSQ    ADJRSQ   F(EQU)   SIGF   RSQCH     FCH  SIGCH          VARIABLE  BETAIN  CORREL
   1   .5857   .3431   .3417   241.826   .000   .3431  241.826   .000  IN:  LPHYHEAL    .5857   .5857
   2   .6136   .3765   .3738   139.507   -.000  .0334   24.772   .000  IN:  SSTRESS     .1928   .3590
   3   .6138   .3768   .3727    92.901   -.000  .0002     .183   .669  IN:  MENHEAL     .0190   .3551
```

TABLE 5.21 HIERARCHICAL REGRESSION OF HEALTH AND STRESS VARIABLES ON NUMBER OF VISITS TO HEALTH PROFESSIONALS

Variables	LTIMEDRS (DV)	LPHYHEAL	SSTRESS	MENHEAL	*B*	β	sr^2 (incremental)
LPHYHEAL	.59				1.040	0.52	.34**
SSTRESS	.36	.32			0.016	0.19	.03**
MENHEAL	.36	.51	.38		0.002	0.02	.00
				Intercept = −0.155			
Means	0.74	0.65	13.40	6.12			
Standard deviation	0.42	0.21	4.97	4.19			
						R^2 = .38	
						Adjusted R^2 = .37	
						R = .61**	

** $p < .01$.

TABLE 5.22 CHECKLIST OF HIERARCHICAL REGRESSION ANALYSIS

1. Issues
   a. Ratio of cases to IVs
   b. Normality, linearity, and homoscedasticity of residuals
   c. Outliers
   d. Multicollinearity and singularity
2. Major analyses
   a. Multiple R^2, F ratio
   b. Adjusted R^2, proportion of variance accounted for
   c. Squared semipartial correlations
   d. Significance of regression coefficients
   e. Incremental F
3. Additional analyses
   a. Unstandardized (B) weights, confidence limits
   b. Standardized (β) weights
   c. Prediction equation
   d. Post hoc significance of correlations
   e. Suppressor variables

Results

Hierarchical regression was employed to determine if addition of information regarding stress and then mental health symptoms improved prediction of visits to health professionals beyond that afforded by differences in physical health. Analysis was performed using SPSS* REGRESSION with an assist from SPSS* FREQUENCIES in evaluation of assumptions.

Results of evaluation of assumptions led to transformation of the variables to reduce skewness in their distributions, reduce the number of outliers, and improve the normality, linearity, and homoscedasticity of residuals. A square root transformation was used on the measure of stress. Logarithmic transformations were used on the number of visits to health professionals and physical health. One IV, mental health, was positively skewed without transformation and negatively skewed with it; it was not transformed. With

the use of a $p < .001$ criterion for Mahalanobis distance no outliers among the cases were identified. No cases had missing data and no suppressor variables were found, $N = 465$.

Table 5.21 displays the correlations between the variables, the unstandardized regression coefficients (B) and intercept, the standardized regression coefficients (β), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2 after entry of all three IVs. R was significantly different from zero at the end of each step. After step 3, with all IVs in the equation, $R = .61$, $F(3, 461) = 92.90$, $p < .01$.

Insert Table 5.21 about here

After step 1, with log of physical health in the equation, $R^2 = .34$, $F_{inc}(1, 461) = 241.83$, $p < .001$. After step 2, with square root of stress added to prediction of log of visits to health professionals by log of physical health, $R^2 = .38$, $F_{inc}(1, 461) = 24.77$, $p < .01$. Addition of square root of stress to the equation results in a significant increment in R^2. After step 3, with mental health added to prediction of visits by log of physical health and square root of stress, $R^2 = .38$ (adjusted $R^2 = .37$), $F_{inc}(1, 461) = 0.18$. Addition of mental health to the equation did not reliably improve R^2.

5.9 SOME EXAMPLES FROM THE LITERATURE

The following examples have been included to give you a taste for the ways in which regression can be, and has been, used by researchers.

Multiple regression is the statistical tool used in judgment analysis, a procedure developed to identify important variables underlying policy decisions. In a demonstration involving diagnosis of learning disabilities, Beatty (1977) gave diagnosticians (judges) test profiles for a sample of children. Judges ranked the children in order

of severity of learning disability, and these rankings served as the DV.[3] Beatty used as IVs the age of the child and the set of five test scores on which judges had based their rankings.

Judges were analyzed individually, and then in combination, so that judges using similar policies could be grouped. For each group of judges with a distinct policy, it was possible to evaluate how successfully the rankings could be predicted and the relative importance of IVs in influencing the diagnostic rankings. In Beatty's demonstration, the three scores on the Wechsler Intelligence Scale for Children (Verbal, Performance, and Full Scale) contributed the most to multiple correlation for all judges. Some judges seemed to have more predictable policies than others; for one judge, 96% of the variance in rankings was explained by knowledge of the IVs.

In evaluating a negative income tax project in New Jersey, Nicholson and Wright (1977) argue for inclusion of "participant's understanding of the treatment" as one of the IVs in policy experimentation. At each stage of their analysis, separate multiple regression equations were set up for each of six DVs (all related to a family's earnings and hours worked). (These six DVs could have been combined into a single canonical correlation evaluated against the IVs; see Chapter 6.) Separate equations were also run on the basis of ethnicity. IVs for all equations included (1) pretreatment value of the DV; (2) variables related to family age, education, and size; (3) variables related to site and whether or not families received welfare; (4) a dummy variable indicating whether a family was in the treatment or the control group; and (5) variables related to specification of which experimental plan a family was in, with eight combinations of guaranteed income and tax levels.

As the main focus of the evaluation, the regression coefficient for the difference between experimental and control groups was evaluated for statistical significance in each of the regression equations. Then Nicholson and Wright reran the regression equations with 11 binary variables related to participant's knowledge about the tax plan. Finally, a third set of equations was run that included terms related to interactions between knowledge levels and estimated effects of the guarantee and tax levels. It was found that the multiple correlation increased significantly with the addition of variables representing participant knowledge, and increased again when interaction variables were included.

In general, estimates of disincentive effects for families with high understanding were larger than for those who lacked understanding of the system. That is, among families who understood the system, those in the experimental group worked fewer hours and earned less than those in the control group. The difference between experimental and control was less, and in some cases in the opposite direction, for families who showed little understanding of the policy.

S. Fidell (1978) applied stepwise regression procedures to predict annoyance among groups of people. Respondents were interviewed at 24 sites across the United States, stratified by noise exposure and population density. The proportion of respondents at each site who had been highly annoyed by neighborhood noise exposure was predicted from both situational (demographic and physical) variables and atti-

[3] Opinions differ regarding application of regression to rank order data. However, since rank order data produce rectangular distributions with neither skewness nor outliers, the application may be considered justified.

tudinal variables. Since the prevalence of shared attitudes in the community was of greater interest than the intensity of individual beliefs, the unit of analysis was the site rather than the individual respondent. Primary interest in prediction also dictated stepwise entry of variables, many of which were available from census data. Excellent prediction was available from several combinations of variables, with one equation of 3 IVs accounting for 88% of the variance in community annoyance at exposure to noise.

Species density as a function of environmental measures in the United States was studied by Schall and Pianka (1978) in statistical regression analysis. Separate analyses were run with birds and lizards as DVs. The 11 IVs were the same for both analyses, consisting of such environmental variables as highest and lowest elevation, sunfall, average annual precipitation and several temperature measures. Units of measurement (cases) were 895 quadrants, each 1° of latitude by 1° of longitude. Environmental measures were found to be much better predictors of lizard density (multiple $R^2 = .82$) than of bird density (multiple $R^2 = .34$). There is also suggestion that different environmental factors predict density for lizards than for birds. Sunfall, entering the regression first for lizards, by itself accounted for .67 of the variance in density. For birds, sunfall appeared to be acted on by a suppressor variable. Simple correlation with bird density was virtually zero, but sunfall entered fifth in the regression and added 9% of predictable variance. Although no cross validation was reported, discrepancies in the two analyses are large enough to be considered important. (Hierarchical regression would have produced results that could be interpreted with more confidence.)

Maki, Hoffman, and Berk (1978) used standard multiple regression to study the effectiveness of a water conservation campaign. Two parallel analyses were done on sales and production of water (DVs) over 126 months, with months acting as the unit of study. IVs were moratorium (whether or not it and/or conservation measures were in effect), rainfall, season, population, and dollars spent on public relations (as a measure of intensity of campaign).

Results of the two analyses were similar, with both showing significant effects of the moratorium-conservation efforts. Money spent on the campaign was not statistically significant in either campaign. Rainfall was a significant predictor of production, but not of sales. Population and season predicted both sales and production.

Chapter 9

Multivariate Analysis of Variance and Covariance

9.1 GENERAL PURPOSE AND DESCRIPTION

Multivariate analysis of variance is a generalization of analysis of variance to a situation in which there are several DVs. For example, suppose a researcher is interested in the effect of different types of treatment on several types of anxiety: 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.

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 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 sex 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.

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 relaxation training or waiting-list control, but only on test anxiety. 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 conditions, it may reveal differences not shown in separate ANOVAs. Such a situation is shown in Figure 9.1 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 sometimes be more powerful than separate ANOVAs.

But there are no free lunches in statistics, just as there are none in life. 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

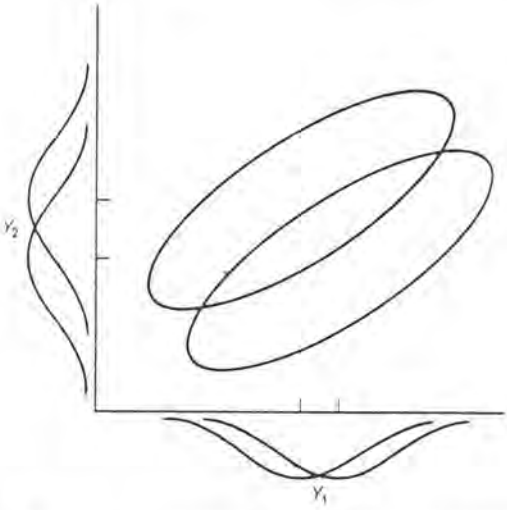


Figure 9.1 Advantage of MANOVA, which combines DVs, over ANOVA. Each axis represents a DV; frequency distributions projected to axes show considerable overlap, while ellipses, showing DVs in combination, do not.

than ANOVA. Thus, our recommendation is to avoid MANOVA except when there is compelling need to measure several DVs.[1]

Multivariate analysis of covariance (MANCOVA) is the multivariate extension of ANCOVA (Chapter 8). MANCOVA asks if there are statistically reliable 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 in the composite score differs in the three treatment groups, after adjusting for preexisting differences in the three types of anxiety.

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 8 for a discussion of the logical difficulties of using covariates this way.)

ANCOVA is used after MANOVA (or MANCOVA) in 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 strength of association, and 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. Typically, MANOVA is 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 function analysis (Chapter 11) 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 function analysis asks if there is some combination of variables that reliably separates groups. But it should be noted that there is no mathematical distinction between MANOVA and discriminant function analysis. At a practical level, computer programs for discriminant analysis are more informative but are also, for

[1] In several years of working with students, we have found that, once the possibility of measuring several DVs is raised, they each and every one become necessary. That is, we have been royally unsuccessful at talking our students out of MANOVA, despite its disadvantage.

the most part, limited to one-way designs. Therefore, analysis of one-way MANOVA is deferred to Chapter 11 and the present chapter covers factorial MANOVA and MANCOVA.

9.2 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 9.2.1 through 9.2.8.

9.2.1 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 9.4.1 and 9.4.3 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 females only 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 sex 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 sex of subject, the two tests independent of each other.

9.2.2 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 9.4.1 and 9.4.3. In the example, the test of interaction asks, Is the pattern of response to the three types of treatment the same for men as it is for women? If the interaction

is significant, it indicates that one type of treatment "works best" for women while another type "works best" 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.

9.2.3 Importance of DVs

If there are significant differences for one or more 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 9.1, 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 9.5.2.

9.2.4 Adjusted Marginal and Cell Means

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 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 is based on both adjusted and sample means, as illustrated in Section 9.7.

9.2.5 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 question for treatment. Similarly, if the interaction of sex 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 9.5.3, 3.2.6, 8.5.2.3, and 10.5.1.

9.2.6 Strength of Association

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? If, for example, there is a reliable main effect of treatment, you can compute the proportion of variance in the composite DV score that is associated with differences in treatment, as described in Section 9.4.1. Procedures are also available for finding the strength of association between the IV and an individually significant DV, as demonstrated in Section 9.7.

9.2.7 Effects of Covariates

When covariates are used, the researcher normally wants to assess their utility. Do the covariates provide reliable 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 9.7.3.1.

9.2.8 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 four times (instead of measuring three different kinds of anxiety once), before, immediately after, 3 months 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 four testing occasions as four DVs.

As discussed in Sections 8.5.2.1 and 3.2.3, repeated measures ANOVA has the often-violated assumption of homogeneity of covariance.[2] 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), or profile analysis (Chapter 10).

9.3 LIMITATIONS TO MULTIVARIATE ANALYSIS OF VARIANCE AND COVARIANCE

9.3.1 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

[2] An additional assumption of repeated-measures ANOVA is homogeneity of variance over within-subjects IVs, tested in SPSS MANOVA as the FMAX test of "homogeneity of variance over measures." If the test fails (i.e., is significant), MANOVA is preferred over repeated-measures ANOVA.

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 instituted. 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; care must be taken in interpreting causality from a significant IV-DV relationship.

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 stepdown F is the method chosen to assess the importance of DVs (see Section 9.5.2.2). 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 hierarchical multiple regression the significance of an IV is likely to depend on its position in the hierarchy.

When MANCOVA is used, the same limitations apply as in ANCOVA. Consult Sections 8.3.1, 8.5.3, and 8.5.4 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.

9.3.2 Practical Issues

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

9.3.2.1 Unequal Sample Sizes and Missing Data Problems associated with unequal cell sizes are discussed in Section 9.5.4.2. Problems caused by incomplete data (and solutions to them) are discussed in Chapters 4 and 8 (particularly Section 8.3.2.1). 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 a lot of cells. There are two

reasons for the requirement. First, 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 F's (and a very unhappy researcher).

The second reason for having more cases than DVs in every cell is associated with the assumption of homogeneity of variance-covariance matrices (see Section 9.3.2.4). 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 as an artifact of a cases-to-DVs ratio that is too low.

9.3.2.2 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 the violations are not due to outliers.

Mardia (1971) shows that MANOVA is also robust to modest violation of normality if the violation is created by skewness rather than by outliers. *A sample size that produces 20 degrees of freedom*[3] *for error in the univariate case should ensure robustness of the test, as long as sample sizes are equal and two-tailed tests are used.* Even with unequal n and only a few DVs, a sample size of about 20 in the smallest cell should ensure robustness. With small, unequal samples, normality is assessed by reliance on judgment. Are the individual DVs expected to be fairly normally distributed in the population? If not, would some transformation be expected to produce normality (cf. Chapter 4)?

9.3.2.3 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 separately for each cell of the design and transform or eliminate them.* Report the transformation or deletion of outlying cases. Screening runs for within-cell univariate and multivariate outliers are shown in Sections 8.7.1.4 and 9.7.1.4.

9.3.2.4 Homogeneity of Variance-Covariance Matrices The multivariate generalization of homogeneity of variance for individual DVs is homogeneity of variance-

[3] See Chapter 3 for a review of calculation of degrees of freedom for univariate F

covariance matrices.[4] The assumption is that variance-covariance matrices within each cell of the design are sampled from the same population variance-covariance matrix and can reasonably be pooled to create a single estimate of error.[5] 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 in T^2 (Hakstian, Roed, and 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 α 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 α 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 may be retained with confidence but indications of mean differences are suspect. *Use Pillai's criterion instead of Wilks' Lambda (see Section 9.5.1) to evaluate multivariate significance* (Olson, 1979). Or equalize sample sizes by random deletion of cases, if power can be maintained at reasonable levels.

9.3.2.5 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. *With a small number of DVs and covariates, examine all within-cell scatterplots between pairs of DVs, pairs of covariates, and pairs of DV-covariate combinations* through SPSS SCATTERGRAM, SPSS[x] PLOT, or BMDP6D.

With a large number of DVs and/or covariates, however, a complete check is unwieldly; spot check the bivariate relationships for which nonlinearity is likely. Or, use the procedure for checking linearity with large numbers of variables as available through BMDP1V and described in Section 8.3.2.6. One DV (preferably the one with the lowest entry priority) is designated DEPENDENT, and the remainder are designated INDEPENDENT (covariates) in an ANOVA run and residuals plots are examined as described in Section 5.3.2.4. 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.

[4] In MANOVA, homogeneity of variance for each of the DVs is also assumed. See Section 9.3.2.5 for discussion and recommendations.

[5] Don't confuse this assumption with the assumption of homogeneity of covariance that is relevant to repeated measures ANOVA or MANOVA, as discussed in Section 8.5.2.1.

9.3.2.6 Homogeneity of Regression In stepdown analysis (Section 9.5.2.2) and in MANCOVA (Section 9.4.3) it is assumed that the relationship between covariates and DVs in one group is the same as the relationship in other groups and that using the average regression to adjust for covariates in all groups is reasonable.

1. 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 8.3.2.7 and 8.5.5 for alternatives to MANCOVA where heterogeneity of regression is found.

In stepdown analysis, 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.

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 9.7.1.6.

9.3.2.7 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 8.3.2.8.

In stepdown analysis where all but the lowest-priority DV act as covariates in assessing other DVs, unreliability of any of the DVs (say $r_{xx} < .8$) 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 9.5.2) and report known or suspected unreliability of covariates and high-priority DVs in your Results section.

9.3.2.8 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. But if there is some compelling theoretical reason to retain all DVs, a principal components analysis (cf. Chapter 12) is done on the pooled within-cell correlation matrix, and component scores are entered as an alternative set of DVs.

BMDP4V and BMDP7M 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). The redundant DV is identified by using the within-cell correlation matrix (from MANOVA) as input to the REGRESSION program and then performing several regressions, with each DV in turn serving as

DV with all other DVs serving as IVs in analysis. If R^2 approaches .99 for some DV, that DV is redundant. Alternatively, the within-cell correlation matrix is input to BMDP4M or to the SPSS^X FACTOR program where SMCs (called EST COMMUNALITY in SPSS) are reported.

9.4 FUNDAMENTAL EQUATIONS FOR MULTIVARIATE ANALYSIS OF VARIANCE AND COVARIANCE

9.4.1 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 9.1. 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 9.4.3.

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 the two subtests differ as a function of degree of disability?

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, absence of the effect, at least among the no treatment group, would lead us to question the adequacy of classification.

The sample size of three children per cell is probably 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. Setup and selected output for this example appear in Section 9.4.2 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

TABLE 9.1 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)

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. Determinants[6] 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 8).

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

$$\sum_i \sum_j (Y_{ij} - \text{GM})^2 = n \sum_j (\bar{Y}_j - \text{GM})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_j)^2 \qquad (9.1)$$

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

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

For factorial designs, or those 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).

$$n_{km} \sum_k \sum_m (\text{DT}_{km} - \text{GM})^2 = n_k \sum_k (\text{D}_k - \text{GM})^2 + n_m \sum_m (T_m - \text{GM})^2 + \left[n_{km} \sum_k \sum_m (\text{DT}_{km} - \text{GM})^2 - n_k \sum_k (\text{D}_k - \text{GM})^2 - n_m \sum_m (T_m - \text{GM})^2 \right] \qquad (9.2)$$

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

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

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 $$SS_{tot} = SS_D + SS_T + SS_{DT}$$

The full partition for this factorial between-subjects design is

$$\sum_i \sum_k \sum_m (Y_{ikm} - GM)^2 = n_k \sum_k (D_k - GM)^2 + n_m \sum_m (T_m - GM)^2 + \left[n_{km} \sum_k \sum_m (DT_{km} - GM)^2 - n_k \sum_k (D_k - GM)^2 - n_m \sum_m (T_m - GM)^2 \right] + \sum_i \sum_k \sum_m (Y_{ikm} - DT_{km})^2 \quad (9.3)$$

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 9.1, column matrices of Y scores for the three children in the first cell of the design (mild disability with treatment) are

$$\mathbf{Y}_{111} = \begin{bmatrix} 115 \\ 108 \end{bmatrix} \quad \begin{bmatrix} 98 \\ 105 \end{bmatrix} \quad \begin{bmatrix} 107 \\ 98 \end{bmatrix}$$

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.

$$\mathbf{D}_1 = \begin{bmatrix} 95.83 \\ 96.50 \end{bmatrix} \qquad \mathbf{D}_2 = \begin{bmatrix} 88.83 \\ 88.00 \end{bmatrix} \qquad \mathbf{D}_3 = \begin{bmatrix} 82.67 \\ 77.17 \end{bmatrix}$$

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

$$\mathbf{T}_1 = \begin{bmatrix} 99.89 \\ 96.33 \end{bmatrix} \qquad \mathbf{T}_2 = \begin{bmatrix} 78.33 \\ 78.11 \end{bmatrix}$$

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 (Y_{ikm}). Thus for the first child in the example

$$(\mathbf{Y}_{111} - \mathbf{GM}) = \begin{bmatrix} 115 \\ 108 \end{bmatrix} - \begin{bmatrix} 89.11 \\ 87.22 \end{bmatrix} = \begin{bmatrix} 25.89 \\ 20.78 \end{bmatrix}$$

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. 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 9.3:

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

$$+ n_m \sum_m (\mathbf{T}_m - \mathbf{GM})(\mathbf{T}_m - \mathbf{GM})'$$

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

$$- n_i \sum_i (\mathbf{D}_i - \mathbf{GM})(\mathbf{D}_i - \mathbf{GM})'$$

$$\left. - n_m \sum_m (\mathbf{T}_m - \mathbf{GM})(\mathbf{T}_m - \mathbf{GM})' \right]$$

$$+ \sum_i \sum_k \sum_m (\mathbf{Y}_{ikm} - \mathbf{DT}_{km})(\mathbf{Y}_{ikm} - \mathbf{DT}_{km})'$$

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 ($S_{S(DT)}$).

For the example in Table 9.1, the four resulting cross-products matrices[7] are

$$S_D = \begin{bmatrix} 520.78 & 761.72 \\ 761.72 & 1126.78 \end{bmatrix} \quad S_T = \begin{bmatrix} 2090.89 & 1767.56 \\ 1767.56 & 1494.22 \end{bmatrix}$$

$$S_{DT} = \begin{bmatrix} 2.11 & 5.28 \\ 5.28 & 52.78 \end{bmatrix} \quad SS_{S(DT)} = \begin{bmatrix} 544.00 & 31.00 \\ 31.00 & 539.33 \end{bmatrix}$$

Notice that all these matrices are symmetric, with the elements top left to bottom right representing sums of squares (that, when divided by degrees of freedom, produce variances), and with the other elements representing sums of cross products (that, when divided by degrees of freedom, produce covariances). That is, 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 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 Wilks' Lambda criterion is used (see Section 9.5.1 for alternative criteria). These ratios follow the general form

$$\Lambda = \frac{|S_{error}|}{|S_{effect} + S_{error}|} \tag{9.4}$$

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

$$S_{DT} + S_{S(DT)} = \begin{bmatrix} 2.11 & 5.28 \\ 5.28 & 52.78 \end{bmatrix} + \begin{bmatrix} 544.00 & 31.00 \\ 31.00 & 539.33 \end{bmatrix}$$

$$= \begin{bmatrix} 546.11 & 36.28 \\ 36.28 & 592.11 \end{bmatrix}$$

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

TABLE 9.2 MULTIVARIATE ANALYSIS OF VARIANCE OF WRAT-R AND WRAT-A SCORES

Source of variance	Wilks' Lambda	Hypoth. df	Error df	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	.27170

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

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

$$|\mathbf{S}_{S(DT)}| = 292436.34$$
$$|\mathbf{S}_{D} + \mathbf{S}_{S(DT)}| = 1145629.58$$
$$|\mathbf{S}_{T} + \mathbf{S}_{S(DT)}| = 2123390.88$$
$$|\mathbf{S}_{DT} + \mathbf{S}_{S(DT)}| = 322042.38$$

At this point a source table, similar to the source table for ANOVA, is useful, as presented in Table 9.2. 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 9.4. For example, for the interaction between disability and treatment, Wilks' Lambda is

$$\Lambda = \frac{|\mathbf{S}_{S(DT)}|}{|\mathbf{S}_{DT} + \mathbf{S}_{S(DT)}|} = \frac{292436.34}{322042.38} = .908068$$

Tables for evaluating Wilks' Lambda directly are not common, though they do appear in Harris (1975). However, an approximation to F has been derived that closely fits Lambda. The last three columns of Table 9.2, then, represent the approximate F values and their associated degrees of freedom.

The following procedure for calculating approximate F is based on Wilks' Lambda and the various degrees of freedom associated with it.

$$\text{Approximate } F(\text{df}_1, \text{df}_2) = \left(\frac{1 - y}{y}\right)\left(\frac{\text{df}_2}{\text{df}_1}\right)$$

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

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

Λ is defined in Equation 9.4, and s is[8]

$$s = \sqrt{\frac{p^2(df_{effect})^2 - 4}{p^2 + (df_{effect})^2 - 5}} \tag{9.5}$$

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, $df_{error} = dt(n - 1)$

Thus

$$s = \sqrt{\frac{(2)^2(2)^2 - 4}{(2)^2 + (2)^2 - 5}} = 2$$

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

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

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

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

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

[8] When $df_{effect} = 1$, $S = 1$. When $p = 1$, we have univariate ANOVA.

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 9.2, significance is indicated at the highest level of alpha reached, following standard practice.

A measure of strength of association is readily available from Wilks' Lambda.[9] For MANOVA:

$$\eta^2 = 1 - \Lambda \tag{9.6}$$

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 9.4, 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 9.6. 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.[10] The square root of η^2 ($\eta = .93$) is a form of correlation between WRAT scores and assignment to treatment.

9.4.2 Computer Analyses of Small Sample Example

Tables 9.3 through 9.6 show setup and selected minimal output for SPSS[x] MANOVA, BMDP4V, SYSTAT MGLH, and SAS GLM, respectively.

In SPSS[x] MANOVA (Table 9.3) simple MANOVA source tables, resembling those of ANOVA, are printed out when PRINT = SIGNIF(BRIEF) is requested. After interpretive material including the design matrix is printed, the TESTS OF

[9] An alternative measure of strength of association 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 6, and the relation between canonical correlation and MANOVA is discussed briefly in Chapter 13.

[10] 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 a measure of the relative importance of an effect.

TABLE 9.3 MANOVA ON SMALL SAMPLE EXAMPLE THROUGH SPSS* MANOVA (SETUP AND OUTPUT)

```
          TITLE          SMALL SAMPLE MANOVA
          FILE HANDLE    TAPE81
          DATA LIST      FILE=TAPE81      FREE
                         /SUBJNO,TREATMNT,DISABILITY,WRATR,WRATA,IQ,CELLNO
          VALUE LABELS   TREATMNT 1 'TREATMENT' 2 'CONTROL'/
                         DISABILITY 1 'MILD' 2 'MODERATE' 3 'SEVERE'
          MANOVA         WRATR, WRATA BY TREATMNT(1,2), DISABLTY(1,3)/
                         PRINT=SIGNIF(BRIEF)/
                         NOPRINT=PARAMETERS(ESTIM)/
                         DESIGN/

* * * * * * * * * * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * * * * * * * * * *
        18 CASES ACCEPTED.
        0 CASES REJECTED BECAUSE OF OUT-OF-RANGE FACTOR VALUES.
        0 CASES REJECTED BECAUSE OF MISSING DATA.
        6 NON-EMPTY CELLS.
        1 DESIGN WILL BE PROCESSED.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
CORRESPONDENCE BETWEEN EFFECTS AND COLUMNS OF BETWEEN-SUBJECTS DESIGN 1

 STARTING  ENDING
  COLUMN    COLUMN   EFFECT NAME

     1         1     CONSTANT
     2         2     TREATMNT
     3         4     DISABLTY
     5         6     TREATMNT BY DISABLTY

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * * * * * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * * * * * * * * * *

TESTS OF SIGNIFICANCE FOR WITHIN CELLS     USING SEQUENTIAL SUMS OF SQUARES   62

SOURCE OF VARIATION              WILKS LAMBDA       MULT. F   HYPOTH. DF   ERROR DF   SIG. OF F

CONSTANT                              .00204    2687.77918         2.00      11.00           0
TREATMNT                              .13772      34.43570         2.00      11.00        .000
DISABLTY                              .25526       5.38602         4.00      22.00        .004
TREATMNT BY DISABLTY                  .90807        .27170         4.00      22.00        .893
```

TABLE 9.4 MANOVA ON SMALL SAMPLE EXAMPLE THROUGH BMDP4V (SETUP AND SELECTED OUTPUT)

```
            PROBLEM      TITLE IS 'SMALL SAMPLE MANOVA'.
            /INPUT       VARIABLES ARE 7.  FORMAT IS FREE.   FILE='TAPE8'.
            /VARIABLE    NAMES ARE SUBJNO,TREATMNT,DISABLTY,WRATR,WRATA,IQ,GROUPNO.
                         USE = TREATMNT,DISABLTY,WRATR,WRATA.
            /BETWEEN     FACTORS ARE TREATMNT, DISABLTY.
                         CODES(1) ARE 1,2.
                         NAMES(1) ARE TREATMENT, CONTROL.
                         CODES(2) ARE 1 TO 3.
                         NAMES(2) ARE MILD, MODERATE, SEVERE.
            /WEIGHTS     BETWEEN ARE EQUAL.
            /END
            ANALYSIS PROC IS FACTORIAL./
            /END

--- ANALYSIS SUMMARY ---


THE FOLLOWING EFFECTS ARE COMPONENTS OF THE SPECIFIED
LINEAR MODEL FOR THE BETWEEN DESIGN.  ESTIMATES AND TESTS
OF HYPOTHESES FOR THESE EFFECTS CONCERN PARAMETERS OF THAT MODEL.

         OVALL: GRAND MEAN
         T: TREATMNT
         D: DISABLTY
         TD

*****************************************************************

*****************************************************************

EFFECT   VARIATE       STATISTIC              F          DF          P
-----------------------------------------------------------------------
 OVALL: GRAND MEAN
        -ALL----
                TSQ=       5864.25          2687.78     2,   11  0.0000
        WRATR
                SS=     142934.222222
                MS=     142934.222222       3152.96     1,   12  0.0000
        WRATA
                SS=     136938.888889
                MS=     136938.888889       3046.85     1,   12  0.0000
 T: TREATMNT
        -ALL----
                TSQ=       75.1324            34.44     2,   11  0.0000
        WRATR
                SS=        2090.888889
                MS=        2090.888889        46.12     1,   12  0.0000
        WRATA
                SS=        1494.222222
                MS=        1494.222222        33.25     1,   12  0.0001
 D: DISABLTY
        -ALL----
                LRATIO=  0.255263             5.39     4,  22.00 0.0035
                TRACE=     2.89503
                TZSQ=      34.7404
                 P APPROXIMATION FOR ABOVE STATISTIC UNAVAILABLE - DF IS ZERO
                MXROOT=  0.742748
                 P APPROXIMATION FOR ABOVE STATISTIC UNAVAILABLE
        WRATR
                SS=         520.777778
                MS=         260.388889         5.79     2,   12  0.0178
        WRATA
                SS=        1126.777778
                MS=         563.388889        12.54     2,   12  0.0012

 TD
        -ALL----
                LRATIO=  0.908068             0.27     4,  22.00 0.8930
                TRACE=   0.100954
                TZSQ=      1.21144
                 P APPROXIMATION FOR ABOVE STATISTIC UNAVAILABLE - DF IS ZERO
                MXROOT=  0.892854E-01
                 P APPROXIMATION FOR ABOVE STATISTIC UNAVAILABLE
        WRATR
                SS=           2.111111
                MS=           1.055556         0.02     2,   12  0.9770
        WRATA
                SS=          52.777778
                MS=          26.388889         0.59     2,   12  0.5711
```

TABLE 9.4 (Continued)

```
ERROR
        WRATR
                SS=          504.00000000
                MS=           45.33333333
        WRATA
                SS=          539.33333333
                MS=           44.94444444
```

TABLE 9.5 MANOVA ON SMALL SAMPLE EXAMPLE THROUGH SYSTAT MGLH (SETUP AND SELECTED OUTPUT)

```
          USE TAPE8:
          CATEGORY TREATMNT=2,DISABLTY=3
          MODEL WRATR,WRATA=CONSTANT+TREATMNT+DISABLTY+TREATMNT*DISABLTY
          ESTIMATE
          HYPOTHESIS EFFECT=TREATMNT*DISABLTY
          TEST
          HYPOTHESIS EFFECT=DISABLTY
          TEST
          HYPOTHESIS EFFECT=TREATMNT
          TEST

NUMBER OF CASES PROCESSED:    18

TEST FOR EFFECT CALLED:
               TREATMNT
             BY
               DISABLTY

UNIVARIATE F TESTS

    VARIABLE        SS        DF        MS              F              P

   WRATR             2.111     2         1.056          0.023          0.977
      ERROR        544.000    12        45.333
   WRATA             52.778    2        26.389          0.587          0.571
      ERROR        539.333    12        44.944

MULTIVARIATE TEST STATISTICS

          WILKS' LAMBDA =       0.908
            F-STATISTIC =       0.272     DF =   4,  22      PROB =       0.893

           PILLAI TRACE =       0.092
            F-STATISTIC =       0.290     DF =   4,  24      PROB =       0.882

 HOTELLING-LAWLEY TRACE =       0.101
            F-STATISTIC =       0.252     DF =   4,  20      PROB =       0.905

                THETA =   0.089  S =  2, M = -.5, N =  4.5 PROB =       1.000


TEST FOR EFFECT CALLED:
               DISABLTY

UNIVARIATE F TESTS

    VARIABLE        SS        DF        MS              F              P

   WRATR           520.778     2       260.389          5.744          0.018
      ERROR        544.000    12        45.333
   WRATA          1126.778     2       563.389         12.535          0.001
      ERROR        539.333    12        44.944

MULTIVARIATE TEST STATISTICS

          WILKS' LAMBDA =       0.255
            F-STATISTIC =       5.386     DF =   4,  22      PROB =       0.004

           PILLAI TRACE =       0.750
            F-STATISTIC =       3.604     DF =   4,  24      PROB =       0.019

 HOTELLING-LAWLEY TRACE =       2.895
            F-STATISTIC =       7.238     DF =   4,  20      PROB =       0.001

                THETA =   0.743  S =  2, M = -.5, N =  4.5 PROB =       0.002
```

TABLE 9.5 (Continued)

```
TEST FOR EFFECT CALLED:
                    TREATMNT

UNIVARIATE F TESTS

    VARIABLE         SS          DF        MS              F              P

    WRATR         2090.889        1     2090.889        46.123          0.0001
        ERROR      544.000       12       45.333
    WRATA         1494.222        1     1494.222        33.246          0.0001
        ERROR      539.333       12       44.944

MULTIVARIATE TEST STATISTICS

            WILKS  LAMBDA =       0.138
              F-STATISTIC =      34.436      DF =   2,  11       PROB =       0.0001

             PILLAI TRACE =       0.862
              F-STATISTIC =      34.436      DF =   2,  11       PROB =       0.0001

  HOTELLING-LAWLEY TRACE =        6.261
              F-STATISTIC =      34.436      DF =   2,  11       PROB =       0.0001
```

TABLE 9.6 MANOVA ON SMALL SAMPLE EXAMPLE THROUGH SAS GLM (SETUP AND OUTPUT)

```
DATA SSAMPLE;
INFILE TAPE81;
INPUT SUBJNO TREATMNT DISABLTY WRATR WRATA IQ CELLNO;
PROC GLM;
  CLASS TREATMNT DISABLTY;
  MODEL WRATR WRATA = TREATMNT DISABLTY TREATMNT*DISABLTY/NOUNI;
  MANOVA H=_ALL_/SHORT;
```

```
                                         SAS

                          GENERAL LINEAR MODELS PROCEDURE

                              CLASS LEVEL INFORMATION

                              CLASS      LEVELS    VALUES

                              TREATMNT      2      1 2

                              DISABLTY      3      1 2 3

                      NUMBER OF OBSERVATIONS IN DATA SET = 18

                 CHARACTERISTIC ROOTS AND VECTORS OF: E INVERSE * H, WHERE
      H = TYPE III SS&CP MATRIX FOR: TREATMNT      E = ERROR SS&CP MATRIX

          CHARACTERISTIC   PERCENT       CHARACTERISTIC VECTOR  V'EV=1
               ROOT
                                              WRATR            WRATA

            6.26103637     100.00       0.03206535       0.02680051
            0.00000000       0.00      -0.02856728       0.03375300

 MANOVA TEST CRITERIA AND EXACT F STATISTICS FOR THE HYPOTHESIS OF NO OVERALL TREATMNT EFFECT

                              S=1     M=0     N=5

 STATISTIC                      VALUE             F            NUM DF        DEN DF       PR > F

 WILKS' LAMBDA               0.13772140        34.436             2            11        0.0001
 PILLAI'S TRACE              0.86227860        34.436             2            11        0.0001
 HOTELLING-LAWLEY TRACE      6.26103637        34.436             2            11        0.0001
 ROY'S GREATEST ROOT         6.26103637        34.436             2            11        0.0001
```

TABLE 9.6 (Continued)

CHARACTERISTIC ROOTS AND VECTORS OF: E INVERSE * H,
WHERE H = TYPE III SS&CP MATRIX FOR: DISABLTY E = ERROR SS&CP 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

S=2 M=-0.5 N=5

STATISTIC	VALUE	F	NUM DF	DEN DF	PR > F
WILKS' LAMBDA	0.2552626	5.386	4	22	0.0035
PILLAI'S TRACE	0.7504811	3.604	4	24	0.0195
HOTELLING-LAWLEY TRACE	2.895034	7.238	4	20	0.0009
ROY'S GREATEST ROOT	2.887241	17.323	2	12	0.0003

NOTE: F STATISTIC FOR ROY'S GREATEST ROOT IS AN UPPER BOUND
F STATISTIC FOR WILKS' LAMBDA IS EXACT

CHARACTERISTIC ROOTS AND VECTORS OF: E INVERSE * H, WHERE
H= TYPE III SS&CP MATRIX FOR: TREATMNT*DISABLTY E = ERROR SS&CP 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	-0.00434581

MANOVA TEST CRITERIA AND F APPROXIMATIONS FOR THE HYPOTHESIS OF NO OVERALL TREATMNT*DISABLTY EFFECT

S=2 M=-0.5 N=5

STATISTIC	VALUE	F	NUM DF	DEN DF	PR > F
WILKS' LAMBDA	0.9080679	0.272	4	22	0.8930
PILLAI'S TRACE	0.09219163	0.290	4	24	0.8816
HOTELLING-LAWLEY TRACE	0.1009535	0.252	4	20	0.9048
ROY'S GREATEST ROOT	0.09803883	0.588	2	12	0.5706

NOTE: F STATISTIC FOR ROY'S GREATEST ROOT IS AN UPPER BOUND
F STATISTIC FOR WILKS' LAMBDA IS EXACT

SIGNIFICANCE FOR WITHIN CELLS USING SEQUENTIAL SUMS OF SQUARES source table is shown.[11] WITHIN CELLS refers to the pooled within-cell error SSCP matrix (Section 9.4.1), the error term chosen by default for MANOVA. For the example, the two-way MANOVA source table consists of CONSTANT (test of the grand mean) followed by the two main effects and the interaction. For each source, you are given Wilks' Lambda, multivariate F with numerator and denominator degrees of freedom (HYPOTH. DF and ERROR DF, respectively), and the probability level achieved for the significance test.

BMDP4V (Table 9.4) requires a USE= instruction if there are more variables in your data set than are used in the MANOVA. After the IVs are listed in the BETWEEN instruction (and in the WITHIN instruction, if applicable) the remaining variables serve as DVs. There is a great deal of output showing the program's interpretation of the data and the analysis, and the cell and marginal means, most of which has been deleted from Table 9.4. Shown is a detailed source table starting

[11] Note that later versions of SPSS' MANOVA use UNIQUE rather than SEQUENTIAL SUMS OF SQUARES as default to correct for unequal-n, cf. Section 8.5.2.2.

with the tests for the grand mean. Results for each effect are reported in a common format: the multivariate test for -ALL---- (the DVs combined) is followed by univariate tests for each DV. For the example, the multivariate test of T: TREATMNT is a form of Hotelling's T^2 (because TREATMNT has only two levels), for which you are given the value of TSQ, the associated F ratio (34.44), numerator and denominator degrees of freedom (2, 11), and the significance level attained (rounded to four digits). For each DV, BMDP4V prints out sum of squares and mean square (SS and MS) for the effect, the univariate F ratio, numerator and denominator df, and significance level.

For an effect with more than one df (e.g., D: DISABLTY and TD, the interaction), several multivariate tests of significance are provided: Wilks' Lambda (LRATIO), Hotelling-Lawley TRACE, a form of Hotelling's T^2 labeled TZSQ, and Roy's gcr (MXROOT), described in Section 9.5.1. Because of the small n in this example, significance values for TZSQ and MXROOT are not printed out. Note that, unlike other programs reviewed here, the same F ratio is estimated for Wilks' Lambda and Hotelling-Lawley TRACE.

The last source of variance in the source table is ERROR. This section contains error SS and MS for each DV, in turn. These mean square values form the denominators in the univariate F tests above.

In SYSTAT MGLH (Table 9.5) the CATEGORY instruction defines the IVs. An equation is formed in the MODEL statement with DVs on the left of the equation and effects on the right. The program then requires a separate HYPOTHESIS instruction for each main effect and interaction to be tested.

Separate source tables are provided for each effect in the output. Each source table begins with univariate F tests for each of the DVs (giving the usual SS, DF, MS, F, and P level) followed by four fully labeled multivariate tests. The last, THETA, is Roy's gcr, with its three df parameters.

In SAS GLM (Table 9.6) IVs are defined in a CLASS instruction while the MODEL instruction defines the DVs and the effects to be considered. The NOUNI instruction suppresses printing of descriptive statistics and univariate F tests. MANOVA H = _ALL_ 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 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 (as discussed in Chapters 6, 12, and 13), and the three df parameters (Section 9.4.1). Each source table shows results of four multivariate tests, fully labeled.

9.4.3 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, preexperimental IQ scores (listed in parentheses in Table 9.1) 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 $\mathbf{GM}$—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

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

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 9.4.1. 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 $\mathbf{S}_T^{(XY)}$).

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

$$\mathbf{S}^* = \mathbf{S}^{(Y)} - \mathbf{S}^{(YX)}(\mathbf{S}^{(X)})^{-1}\mathbf{S}^{(XY)} \tag{9.7}$$

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 S* matrices. Ratios of determinants are formed to test hypotheses about main effects and interactions by using Wilks' Lambda criterion (Equation 9.4). 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 9.7.

One new item in this source table that is not in the MANOVA table of Section 9.4.1 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 for error is used for each covariate so that df_2 and s of Equation 9.5 are modified. For MANCOVA, then,

$$s = \sqrt{\frac{(p+q)^2(df_{effect})^2 - 4}{(p+q)^2 + (df_{effect})^2 - 5}} \tag{9.8}$$

TABLE 9.7 MULTIVARIATE ANALYSIS OF COVARIANCE OF WRAT-R AND WRAT-A SCORES

Source of variance	Wilks' Lambda	Hypoth. df	Error df	Multivariate *F*
Covariate	.58485	2.00	10.00	3.54913
Treatment	.10180	2.00	10.00	44.11554**
Disability	.25399	4.00	20.00	4.92112*
Treatment by disability	.93896	4.00	20.00	0.15997

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

where q is the number of covariates and all other terms are defined as in Equation 9.5. And

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

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 strength of association as shown in Equation 9.6.

9.5 SOME IMPORTANT ISSUES

9.5.1 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 (1 df), 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 ($df > 1$), the F values are often different but all three statistics are either 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 11) 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, 1975). Most researchers, however, use one of the pooled statistics to test the effect (Olson, 1976).

Wilks' Lambda, defined in Equation 9.4 and Section 9.4.1, 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. Equations describing these criteria are available in the SPSS Update 7–9 manual (Hull and Nie, 1981) and the SPSS MANOVA manual (Cohen and Burns, 1977).

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 9.3.2.4), 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 9.6).

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, with only tentative interpretation.

9.5.2 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 for 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.

9.5.2.1 Univariate *F* If pooled within-group correlations among DVs are zero, 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 strength of association. However, because of inflated Type I error rate due to multiple testing, more stringent α levels are required.

Because there are multiple ANOVA's, 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.

$$\alpha = 1 - (1 - \alpha_1)(1 - \alpha_2) \cdots (1 - \alpha_p) \tag{9.9}$$

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 pth, 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 α level according to Equation 9.9 is .039, acceptably below .05 overall. Or, if alpha is set at .02 for 2 DVs, and at .001 for the other 2 DVs, overall alpha is .042, also below .05.

When DVs are correlated, there are 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.

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, SAS GLM, and SYSTAT MGLH; the pooled within-group variance-covariance matrix is available through BMDP7M, and entries can be converted to correlations as described in Chapter 1.

In the example of Tables 9.1 and 9.2, there is a significant multivariate effect of treatment (and of disability, although, as previously noted, the disability effect 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 9.8 and 9.9, respectively. The pooled within-group correlation between WRAT-R and WRAT-A is .057 with 12 df. Because the DVs are uncorrelated, univariate F with adjustment of α for multiple tests is appropriate. There are two DVs, so each is set at α at .025.[12] 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.

9.5.2.2 Stepdown Analysis[13] The problem of correlated univariate F tests with correlated DVs is resolved by stepdown analysis (Bock, 1966; Bock and Haggard, 1968). Stepdown analysis of DVs is analogous to testing the importance of IVs in multiple regression by hierarchical analysis. Priorities are assigned to DVs according to theoretical or practical considerations.[14] The highest-priority DV is tested in univariate ANOVA, with appropriate adjustment of alpha. The rest of the DVs are tested

[12] 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.

[13] 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.

[14] 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.

TABLE 9.8 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 9.9 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	

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 9.5.2.1.

For the example, let's 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 9.8. Because the main effect of disability is not interesting and the interaction is not statistically significant in MANOVA (Table 9.2), 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 9.10.[15] 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 WRAT-R.

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 the effects of scholastic variables. The attitudinal variables serve as

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

TABLE 9.10 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	

DVs in MANCOVA while the scholastic variables serve as covariates. This type of analysis is demonstrated in the hierarchical discriminant function analysis of Chapter 11.

9.5.2.3 Choosing among Strategies for Assessing DVs You may find the procedures of Sections 11.6.3 and 11.6.4 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 10.5.1 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 very little correlation among the DVs, univariate *F* with adjustment for Type I error is acceptable. When DVs are correlated, 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 *F*'s 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 from univariate analysis, 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 9.7.

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 *F*'s but significance levels are not given.

An alternative to attempting interpretation of either univariate or stepdown F is interpretation of discriminant functions, as discussed in Section 11.6.3. This process is facilitated when SPSS[x] MANOVA, SAS GLM, or SYSTAT MGLH is used because information about the discriminant functions is provided as a routine part of the output.

9.5.3 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 comparisions 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.[16] Review Sections 3.2.6, 8.5.2.3, and 10.5.1 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, the degrees of freedom are those associated with the interaction.

9.5.4 Design Complexity

In between-subjects designs with 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 9.5.2 and 9.5.3.

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

9.5.4.1 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

[16] Comparisons and trend analysis can also be performed on a composite DV as illustrated in Tabachnick and Fidell (1983) Section 8.5.2 but the results for a composite DV are less interpretable.

3 and 8 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 produce a within-subjects IV with as many levels as occasions (Chapter 3). Alternatively, occasions can be treated as separate DVs—one DV per occasion (Section 9.2.8). In this latter view, if there is more than one DV measured on each occasion, the design is said to be doubly multivariate. (There is no distinction between the two views when there are only two levels of the within-subjects IV.) Section 10.5.3 has an example of 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. Generalizations from the example are reasonably straightforward.

9.5.4.2 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 not independent. There are a number of ways to adjust for overlap in sums of squares (cf. Woodward and Overall, 1975), as discussed in some detail in Section 8.5.2.2, particularly Table 8.11. Both the problem and the solutions generalize to MANOVA.

All the MANOVA programs described in Section 9.6 adjust for unequal n. SPSS MANOVA offers both Method 1 adjustment (METHOD = SSTYPE(UNIQUE)) and Method 3 adjustment (METHOD = SSTYPE(SEQUENTIAL)). Method 3 adjustment with survey data through SPSS^x MANOVA is shown in Section 9.7.2. BMDP4V allows very flexible specification of method through choice of cell-weighting procedure. In SAS GLM Method 1 (called TYPE III or TYPE IV) is the default among four options available. For SYSTAT MGLH, default is also Method 1; alternative methods require specification of custom sums of squares for error terms.

9.6 COMPARISON OF PROGRAMS

SPSS^x, BMDP, SAS, and SYSTAT all have highly flexible and full-featured MANOVA programs, as seen in Table 9.11. One-way between-subjects MANOVA is also available through discriminant function programs, as discussed in Chapter 11.

9.6.1 SPSS Package

Within the original SPSS series of programs (Nie et al., 1975), no program was specifically designed for MANOVA (although one-way between-subjects MANOVA could be analyzed through DISCRIMINANT). The new MANOVA, however, is a comprehensive, flexible program that gives SPSS and SPSS^x extended capability for both MANOVA and discriminant function analyses, as seen in Table 9.11.

The program offers several methods of adjustment for unequal n and is the only program that performs stepdown analysis as an option (Section 9.5.2.2). Should Pillai's criterion be desired as the multivariate significance test (Section 9.5.1), it

and several other alternatives are available through SPSS MANOVA. The program is also reasonably simple to use for comparisons on both between- and within-subjects IVs (Section 10.5.1).

Except for outliers,[17] tests of assumptions are readily available in MANOVA. Multicollinearity and homogeneity of variance-covariance matrices are readily tested through within-cell correlations and homogeneity of dispersion matrices, respectively.

For between-subjects designs, Bartlett's test of sphericity 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 9.5.2.1). In repeated measures designs, the sphericity test evaluates the homogeneity of covariance 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 (Section 8.5.5). Although this test of homogeneity of covariance is provided, adjustment for violation of the assumption is available only in the CDC CYBER version of the program. The FMAX test for homogeneity of variance among DVs also facilitates the choice between repeated-measures univariate ANOVA and MANOVA.

For designs with covariates or when stepdown F is used, homogeneity of regression is readily tested through control language illustrated in Table 9.14. If the assumption is violated, the manuals describe procedures for ANCOVA with separate regression estimates. Adjusted cell and marginal means can be obtained, although in different ways. The PMEANS procedure provides adjusted cell means. For adjusted marginal means, however, the CONSPLUS procedure is used, illustrated in Sections 9.7.2 and 9.7.3.

Finally, a principal components analysis can be performed on the DVs, 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.

9.6.2 BMD Series

The MANOVA program in the BMDP series is P4V. The program is also known as URWAS (University of Rochester Weighted Analysis of Variance System). It is unique among the programs reviewed here in that it uses a cell weighting system to adjust for unequal n and to specify hypotheses to be tested. Cell weighting can be used for situations where population sizes are unequal even though sample sizes are equal.

For repeated measures designs, both the Huynh-Feldt and the Greenhouse-Geisser-Imhof corrections for degrees of freedom are available to adjust for violation of the assumption of homogeneity of covariance.

Statistics appropriate for doubly multivariate analysis (in which levels of a

[17] The CYBER implentation of MANOVA provides graphic screening for multivariate outliers, but no formal statistical evaluation of them. Further, outliers are not assessed with respect to their own group.

TABLE 9.11 COMPARISON OF PROGRAMS FOR MULTIVARIATE ANALYSIS OF VARIANCE AND COVARIANCE

Feature	BMDP4V	SPSS MANOVA	SPSS[x] MANOVA	SYSTAT MGLH	SAS GLM
Input					
Post hoc comparisons	No	No	No	No	Yes
Variety of strategies for unequal n	Yes	Yes	Yes	Yes	Yes
Alternative languages or procedures to specify design	Yes	Yes	Yes	Yes	Yes
Specific comparisons	Yes	Yes	Yes	Yes	Yes
Tests of simple effects (complete)	Yes	No[a]	Yes	No	No
Output					
Standard source table	Yes	PRINT = SIGNIF(BRIEF)		No	No
F matrix of cell comparisons	No	No	No	No	No
Design matrix	Yes	Yes	Yes	Yes[b]	Yes[c]
Cell covariance matrices	No	Yes	Yes	No	No
Cell covariance matrix determinants	No	Yes	Yes	No	No
Cell correlation matrices	No	Yes	Yes	No	No
Cell SSCP matrices	No	Yes	Yes	No	No
Cell SSCP determinants	No	Yes	Yes	No	No
Marginal standard deviation for factorial design	Yes	No	No	No	No
Marginal means for factorial design	Yes	Yes	Yes	No	No
Weighted marginal means	Yes	Yes	Yes	No	No
Cell means	Yes	Yes	Yes	No	Yes
Cell standard deviations	Yes	Yes	Yes	No	No
Cell and marginal variance	Yes	No	No	No	No
Confidence interval around cell means	No	Yes	Yes	No	No
Cell and marginal minimum and maximum values	Yes	No	No	No	No
Adjusted cell means	No	PMEANS	PMEANS	No	LS MEANS
Adjusted marginal means	No	Yes[d]	Yes[d]	No	LS MEANS
Wilks' Lambda (U statistic) with approx. F statistic	LRATIO	Yes	Yes	Yes	LRATIO
Test for multivariate outliers	No[e]	No[a]	No	Yes	No
Canonical (discriminant function) statistics[f]	Eigenvalues and eigenvectors	Yes	Yes	Yes	Yes

Univariate F tests	Yes	Yes	Yes	Yes	Yes
Averaged univariate F tests	No	Yes	Yes	No	No
Stepdown F tests (DVs)	No[g]	Yes	Yes	No[g]	No[g]
Criteria other than Wilks'	Yes	Yes	Yes	Yes	Yes
Greenhouse-Geisser-Imhof and Huynh-Feldt statistic for heterogeneity of covariance	Yes	No[a]	No	No	Yes
Bartlett test of sphericity for homogeneity of covariance	No	Yes	Yes	No	No
Tests for univariate homogeneity of variance	No	Yes	Yes	No	No
Test for homogeneity of covariance matrices	No	Box's M	Box's M	No	No
Principal components analysis of residuals	No	Yes	Yes	Yes	No
Hypothesis SSCP matrices	Yes	No	No	Yes	Yes
Inverse of hypothesis SSCP matrices	No	No	No	No	Yes
Pooled with-cell error SSCP matrix	Yes	Yes	Yes	Yes	Yes
Pooled within-cell covariance matrix	No	Yes	Yes	Yes	Yes
Pooled within-cell correlation matrix	No	Yes	Yes	Yes	Yes
Determinants of pooled within-cell correlation matrix	No	Yes	Yes	No	No
Homogeneity of regression	No	Yes	Yes	Yes	No
ANCOVA with separate regression estimates	No	Yes	Yes	No	No
Tukey's test for nonadditivity	No	Yes	Yes	No	No
Effect sizes	No	No[a]	No	No	No
Observed power values	No	No[a]	No	No	No
Tolerance	No	No	No	No	Yes
R^2 for model	No	No	No	No	Yes
R^2 for each DV	No	No	No	Yes	No
Coefficient of variation	No	No	No	No	Yes
Normalized plots for each DV and covariate	No	Yes	Yes	No[h]	No
Predicted values and residuals for each case	No	Yes	Yes	Yes	Yes
Confidence limits for predicted values	No	No	No	No	Yes
Residuals plots	No	Yes	Yes	No[h]	No[i]

[a] Available in CDC CYBER SPSS-6000 Version 9.0.

[b] Available by saving MODEL.

[c] Prints transformations matrices that define contrasts.

[d] Available through CONSPLUS procedure; see Section 9.7.

[e] Available through BMDP4M or BMDP7M.

[f] Discussed more fully in Chapter 11.

[g] Can be done through successive ANCOVAs as subproblems.

[h] Available through GRAPH program.

[i] Available through PLOT procedure.

within-subjects IV are treated as separate DVs) are automatically provided when multiple DVs are specified (cf. Chapter 10). Weighted marginal means are given that are useful if weights other than sample sizes are used (see Section 8.5.2.2). Adjusted (for covariates) cell and marginal means are planned but have not been implemented at the time of this writing. Three multivariate test criteria, as well as univariate F tests, are automatically provided. Specific comparisons include a simple effects procedure as described in Section 10.5.1.

For a factorial design in which tests of all main effects and interactions are desired, a simple statement so indicates. For complex and/or nonorthogonal designs, Model Description Language is available, in which the user specifies the design matrix of contrast coefficients and/or other design goodies. Finally, a structural-equation procedure allows the user to write out an equation that specifies the design and desired tests.

Neither stepdown analysis nor tests for homogeneity of regression are available in BMDP4V, although tests for MANOVA stepdown analysis can be done through BMDP1V, as described in Chapter 8. No direct evidence is available to evaluate multicollinearity, but the program does deal with the problem. If a sum-of-squares and products matrix is multicollinear, it is conditioned and a pseudoinverse calculated. And if there is any problem with the accuracy of a statistic, warning messages regarding precision are printed out.

9.6.3 SYSTAT System

In SYSTAT, the MGLH program serves for all varieties of linear modeling—multiple regression, ANOVA and ANCOVA, MANOVA and MANCOVA, and discriminant function analysis. Among the SYSTAT programs, MGLH is by far the most flexible and complex.

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. Several criteria are provided for tests of multivariate hypotheses, along with a great deal of flexibility in specifying these hypotheses. Mahalanobis distance is provided for evaluation of multivariate outliers, but the values are the square root of those typically given. To apply a χ^2 criterion to evaluate significance of outliers, you must first square the distance value produced by SYSTAT.

For MANCOVA or stepdown F, tests are provided for homogeneity of regression. However, there is no provision in MGLH or elsewhere in the package for adjusted cell or marginal means. Other univariate statistics are not provided in the program, but they can be obtained through the STATS module.

For repeated-measures designs, there is no test for homogeneity of covariance. Instead, the manual (Wilkinson, 1986) recommends the doubly multivariate approach (cf. Chapter 10) that does not require satisfaction of this assumption.

Like SPSS^x 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.

9.6.4 SAS System

MANOVA in SAS is done through the PROC GLM. This general linear model program, like that of SYSTAT MGLH, has great flexibility in testing models and specific comparisons. Four types of adjustment for unequal-n are available, called TYPE I through TYPE IV estimable functions (cf. 8.5.2.2); this program is considered by some to have provided the archetypes of the choices available for unequal-n adjustment. Cell means, adjusted cell means, and marginal means are printed out with the LSMEANS instruction. SAS has no test for multivariate outliers, however.

SAS GLM provides Greenhouse-Geisser and Huynh-Feldt adjustments to degrees of freedom and corresponding significance values for violation of homogeneity of covariance in repeated measures designs. 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 8.5.1).

Of the programs reviewed, GLM has the most extensive selection of options for testing contrasts. In addition to a CONTRAST procedure for specifying comparisons, numerous post hoc procedures are available.

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.

9.7 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?

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.

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).

Omnibus MANOVA (Section 9.7.2) asks whether these DVs are associated with

the two IVs (femininity and masculinity) or their interaction. Then 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 9.7.3), 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.

9.7.1 Evaluation of Assumptions

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

9.7.1.1 Unequal Sample Sizes and Missing Data SPSS^x 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 9.12 for output on the CONTROL variable). The run reveals the presence of a case for which the CONTROL score is missing. No datum is missing on any

TABLE 9.12 SETUP AND SELECTED SPSS^x FREQUENCIES OUTPUT FOR MANOVA VARIABLES SPLIT BY GROUPS

```
            FILE HANDLE    MANOVA
            DATA LIST      FILE=MANOVA
                           /CASENO,ANORM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
                            SEL2,INTEXT,NEUROTIC (A4,6F3.0,F8.5,F4.1,F3.0)
            VALUE LABELS   ANORM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS'
                           FEM,MASC 1 'LOW' 2 'HIGH'
            MISSING VALUES CONTROL(0)
            SORT CASES BY ANORM
            SPLIT FILE BY ANORM
            FREQUENCIES    VARIABLES=ESTEEM TO NEUROTIC/
                           FORMAT=NOTABLE/
                           HISTOGRAM=NORMAL/
                           STATISTICS=ALL/


CONTROL

     COUNT       VALUE     ONE SYMBOL EQUALS APPROXIMATELY  1.20 OCCURRENCES

        26        5.00   ***************I*****
        54        6.00   ***********************************I********
        51        7.00   *********************************************
        18        8.00   ***************
        20        9.00   *********I*******
         3       10.00   *I*
                      I.........I.........I.........I.........I.........I
                      0         12        24        36        48        60
                                HISTOGRAM FREQUENCY

 ANORM:   2   FEMNINE
CONTROL


 MEAN            6.773       STD ERR          .097       MEDIAN          7.000
 MODE            6.000       STD DEV         1.266       VARIANCE        1.603
 KURTOSIS        -.381       S E KURT        1.989       SKEWNESS         .541
 S E SKEW         .185       RANGE           5.000       MINIMUM         5.000
 MAXIMUM        10.000       SUM          1165.000

 VALID CASES      172       MISSING CASES      1
```

of the 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 hierarchical 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).

9.7.1.2 Multivariate Normality The sample size of 368 includes over 35 cases for each cell of the 2 × 2 between-subjects design, far more than the 20 df for error suggested to assure multivariate normality of the sampling distribution of means, even with unequal sample sizes. Further, the distributions for the full run (of which CONTROL in Table 9.12 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.

9.7.1.3 Linearity The full output for the run of Table 9.12 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^x PLOT would have been used with the SORT and SPLIT FILE setup in Table 9.12.

9.7.1.4 Outliers BMDPAM is used to check for univariate and multivariate outliers (cf. Table 8.16) within each of the four groups. Using a cutoff criterion of $p < .001$, no outliers are found.

9.7.1.5 Homogeneity of Variance-Covariance Matrices As a preliminary check for robustness, sample variances (in the full run of Table 9.12) for each DV are compared across the four groups. For no DV does the ratio of largest to smallest variance approach 20:1. 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 sample 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 9.15) produces $F(63, 63020) = 1.07$, $p > .05$, confirming homogeneity of variance-covariance matrices.

9.7.1.6 Homogeneity of Regression Because 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 9.13 shows the SPSS MANOVA deck setup 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 step.

TABLE 9.13 TEST FOR HOMOGENEITY OF REGRESSION FOR LARGE SAMPLE MANOVA STEPDOWN ANALYSIS (INPUT AND SELECTED OUTPUT FOR LAST TWO TESTS FROM SPSS' MANOVA)

```
TITLE          HOMOGENEITY OF REGRESSION FOR LARGE SAMPLE MANOVA
FILE HANDLE    MANOVA
DATA LIST      FILE=MANOVA
               /CASENO,ANDRM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
                SEL2,INTEXT,NEUROTIC (A8,6F3.0,F8.5,F8.1,F3.0)
VALUE LABELS   ANDRM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS'/
               FEM,MASC 1 'LOW' 2 'HIGH'
MISSING VALUES  CONTROL(0)
MANOVA  ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM(1,2) MASC(1,2)/
        PRINT=SIGNIF(BRIEF)/ NOPRINT=PARAMETERS(ESTIM)/
     ANALYSIS=ATTROLE/
     DESIGN=ESTEEM,FEM,MASC,FEM BY MASC, ESTEEM BY FEM = ESTEEM
          BY MASC + ESTEEM BY FEM BY MASC/
     ANALYSIS=NEUROTIC/
     DESIGN=ESTEEM,ATTROLE,FEM,MASC,FEM BY MASC,CONTIN(ESTEEM,ATTROLE)
          BY FEM + CONTIN(ESTEEM,ATTROLE) BY MASC + CONTIN
          (ESTEEM,ATTROLE) BY FEM BY MASC/
     ANALYSIS=INTEXT/
     DESIGN=ESTEEM,ATTROLE,NEUROTIC,FEM,MASC,FEM BY MASC,CONTIN(ESTEEM,
          ATTROLE,NEUROTIC) BY FEM + CONTIN(ESTEEM,ATTROLE,NEUROTIC)
          BY MASC + CONTIN(ESTEEM,ATTROLE,NEUROTIC) BY FEM BY MASC/
     ANALYSIS=CONTROL/
     DESIGN=ESTEEM,ATTROLE,NEUROTIC,INTEXT,FEM,MASC,FEM BY MASC,
          CONTIN(ESTEEM,ATTROLE,NEUROTIC,INTEXT) BY FEM +
          CONTIN(ESTEEM,ATTROLE,NEUROTIC,INTEXT) BY MASC +
          CONTIN(ESTEEM,ATTROLE,NEUROTIC,INTEXT) BY FEM BY MASC/
     ANALYSIS=SEL2/
     DESIGN=ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,FEM,MASC,FEM BY MASC,
          CONTIN(ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL) BY FEM +
          CONTIN(ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL) BY MASC +
          CONTIN(ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL) BY FEM BY MASC/
```

TESTS OF SIGNIFICANCE FOR CONTROL USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	442.61116	388	1.27187		
CONSTANT	16632.27174	1	16632.27174	13077.00892	0
ESTEEM	85.72041	1	85.72041	67.39708	0
ATTROLE	4.22285	1	4.22285	3.32018	.069
NEUROTIC	46.07474	1	46.07474	36.22595	0
INTEXT	.85481	1	.85481	.67209	.413
FEM	.05797	1	.05797	.04558	.831
MASC	.00005	1	.00005	.00004	.995
FEM BY MASC	.41040	1	.41040	.32267	.570
CONTIN(ESTEEM ATTROLE NEUROTIC INTEXT) B Y FEM + CONTIN(ESTEEM ATTROLE NEUROTIC I NTEXT) BY MASC + CONTIN(ESTEEM ATTROLE NEUROTIC INTEXT) BY FEM BY MASC	16.77587	12	1.39799	1.09932	.358

TESTS OF SIGNIFICANCE FOR SEL2 USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	220340.09519	344	640.52353		
CONSTANT	617835.10457	1	617835.10457	964.57831	0
ESTEEM	938.10858	1	938.10858	1.46450	.227
ATTROLE	17.20327	1	17.20327	.02686	.870
NEUROTIC	.12133	1	.12133	.00019	.989
INTEXT	755.15901	1	355.15901	1.17887	.278
CONTROL	1407.82269	1	1407.82269	2.19792	.139
FEM	20.02234	1	20.02234	.03126	.860
MASC	384.09808	1	384.09808	.59966	.439
FEM BY MASC	297.09000	1	297.09000	.46382	.496
CONTIN(ESTEEM ATTROLE NEUROTIC INTEXT CO NTROL) BY FEM + CONTIN(ESTEEM ATTROLE NE UROTIC INTEXT CONTROL) BY MASC + CONTIN(ESTEEM ATTROLE NEUROTIC INTEXT CONTROL) BY FEM BY MASC	16017.21610	15	934.48108	1.45893	.118

Table 9.13 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 SOURCE OF VARIATION, so that for SEL2 the F value for homogeneity of regression is $F(15, 344) = 1.45893$, $p > .01$. (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. Deck setups for all tests are shown in Table 9.14. 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 are also shown in Table 9.14. 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.

9.7.1.7 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.

9.7.1.8 Multicollinearity and Singularity The determinant of the pooled within-cells correlation matrix is found (through SPSS^x MANOVA setup in Table 9.15) to be 0.81. This is sufficiently different from zero that neither multicollinearity nor singularity is judged to be a problem.

9.7.2 Multivariate Analysis of Variance

Deck setup and partial output of omnibus MANOVA produced by SPSS MANOVA appear in Table 9.15. The order of IVs listed in the MANOVA statement together with METHOD = SSTYPE(SEQUENTIAL) sets up the priority for testing FEM before MASC in this unequal-n design.[18] 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.[19] Both main effects are highly significant, but there is no statistically reliable interaction. If desired, strength of association for the composite DV for each main effect is found using Equation 9.6.

[18] Check carefully with your computer center to see what the default is for adjustment for unequal n in MANOVA; in some versions the sequential method (Method 3) is default. On most versions, particularly more recent ones, Method 1 is default. We recommend that you take positive control of unequal n adjustment by using appropriate control language rather than relying on a changing default method.

[19] 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 9.14 TESTS OF HOMOGENEITY OF REGRESSION FOR LARGE SAMPLE MANCOVA AND STEPDOWN ANALYSIS (SETUP AND PARTIAL OUTPUT FOR OVERALL TESTS AND LAST STEPDOWN TEST FROM SPSS' MANOVA)

```
TITLE        HOMOGENEITY OF REGRESSION FOR LARGE SAMPLE MANCOVA
FILE HANDLE  MANOVA
DATA LIST    FILE=MANOVA
             /CASENO,ANORM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
              SEL2,INTEXT,NEUROTIC (A4,6F3.0,F8.5,F4.1,F3.0)
VALUE LABELS ANORM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS'/
             FEM,MASC 1 'LOW' 2 'HIGH'
MISSING VALUES  CONTROL(0)
MANOVA   ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM(1,2) MASC(1,2)/
         PRINT=SIGNIF(BRIEF)/
         NOPRINT=PARAMETERS(ESTIM)/
       ANALYSIS=ESTEEM,INTEXT,NEUROTIC/
       DESIGN=CONTROL,ATTROLE,SEL2,FEM,MASC,FEM BY MASC,
              CONTIN(CONTROL,ATTROLE,SEL2) BY FEM +
              CONTIN(CONTROL,ATTROLE,SEL2) BY MASC +
              CONTIN(CONTROL,ATTROLE,SEL2) BY FEM BY MASC/
       ANALYSIS=ESTEEM/
       DESIGN=CONTROL,ATTROLE,SEL2,FEM,MASC,FEM BY MASC,
              CONTIN(CONTROL,ATTROLE,SEL2) BY FEM +
              CONTIN(CONTROL,ATTROLE,SEL2) BY MASC +
              CONTIN(CONTROL,ATTROLE,SEL2) BY FEM BY MASC/
       ANALYSIS=INTEXT/
       DESIGN=CONTROL,ATTROLE,SEL2,ESTEEM,FEM,MASC,FEM BY MASC,
              CONTIN(CONTROL,ATTROLE,SEL2,ESTEEM) BY FEM +
              CONTIN(CONTROL,ATTROLE,SEL2,ESTEEM) BY MASC +
              CONTIN(CONTROL,ATTROLE,SEL2,ESTEEN,INTEXT) BY FEM BY MASC/
       ANALYSIS=NEUROTIC/
       DESIGN=CONTROL,ATTROLE,SEL2,ESTEEM,INTEXT,FEM,MASC,FEM BY MASC,
              CONTIN(CONTROL,ATTROLE,SEL2,ESTEEM,INTEXT) BY FEM +
              CONTIN(CONTROL,ATTROLE,SEL2,ESTEEM,INTEXT) BY MASC +
              CONTIN(CONTROL,ATTROLE,SEL2,ESTEEM,INTEXT) BY FEM BY MASC/
```

TESTS OF SIGNIFICANCE FOR WITHIN+RESIDUAL USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	WILKS LAMBDA	MULT. F	HYPOTH. DF	ERROR DF	SIG. OF F
CONSTANT	.02360	4826.62570	3.00	350.00	0
CONTROL	.74523	39.88542	3.00	350.00	0
ATTROLE	.93181	8.53755	3.00	350.00	.000
SEL2	.99517	.56566	3.00	350.00	.638
FEM	.95317	5.73190	3.00	350.00	.001
MASC	.85120	20.39038	3.00	350.00	0
FEM BY MASC	.99734	.31119	3.00	350.00	.817
CONTIN(CONTROL ATTROLE SEL2) BY FEM + CO NTIN(CONTROL ATTROLE SEL2) BY MASC + CO NTIN(CONTROL ATTROLE SEL2) BY FEM BY MASC	.93292	.91150	27.00	1022.82	.586

TESTS OF SIGNIFICANCE FOR NEUROTIC USING SEQUENTIAL SUMS OF SQUARES

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIG. OF F
WITHIN+RESIDUAL	6662.67324	344	19.36824		
CONSTANT	27930.53261	1	27930.53261	1442.07931	0
CONTROL	1487.85072	1	1487.85072	76.81911	0
ATTROLE	50.58323	1	50.58323	2.61166	.107
SEL2	1.67428	1	1.67428	.08644	.769
ESTEEM	535.03736	1	535.03736	27.62447	0
INTEXT	29.26997	1	29.26997	1.51124	.220
FEM	4.32678	1	4.32678	.22340	.637
MASC	.92184	1	.92184	.04760	.827
FEM BY MASC	4.94321	1	4.94321	.25522	.614
CONTIN(CONTROL ATTROLE SEL2 ESTEEM INTEX T) BY FEM + CONTIN(CONTROL ATTROLE SEL2 ESTEEM INTEXT) BY MASC + CONTIN(CONTROL ATTROLE SEL2 ESTEEM INTEXT) BY FEM BY MASC	420.18577	15	28.01245	1.44631	.124

TABLE 9.15 MULTIVARIATE ANALYSIS OF VARIANCE OF ESTEEM, CONTROL, ATTROLE, SELF, INTEXT, AND NEUROTIC, AS A FUNCTION OF (TOP TO BOTTOM) FEMINITY BY MASCULINITY INTERACTION, MASCULINITY, AND FEMININITY (SETUP AND SELECTED OUTPUT FROM SPSS' MANOVA)

```
TITLE           LARGE SAMPLE MANOVA
FILE HANDLE     MANOVA
DATA LIST       FILE=MANOVA
                /CASENO,ANORM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
                 SELF,INTEXT,NEUROTIC (A4,6F3.0,F8.5,F8.1,F3.0)
VALUE LABELS    ANORM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS'/
                FEM,MASC 1 'LOW' 2 'HIGH'
MISSING VALUES    CONTROL(0)
MANOVA    ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SELF BY FEM(1,2) MASC(1,2)/
          NOPRINT=PARAMETERS(ESTIM)/
          PRINT=SIGNIF(STEPDOWN), ERROR(COR),
                HOMOGENEITY(BARTLETT,COCHRAN,BOXM)/
          METHOD=SSTYPE(SEQUENTIAL)/
          DESIGN/
```

```
EFFECT .. FEM BY MASC

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 1, M = 2 , N = 178 1/2)

TEST NAME          VALUE     APPROX. F     HYPOTH. DF     ERROR DF     SIG. OF F

PILLAIS            .00815      .49230          6.00       359.00          .814
HOTELLINGS         .00823      .49230          6.00       359.00          .814
WILKS              .99194      .49230          6.00       359.00          .814
ROYS               .00816

EFFECT .. MASC

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 1, M = 2 , N = 178 1/2)

TEST NAME          VALUE     APPROX. F     HYPOTH. DF     ERROR DF     SIG. OF F

PILLAIS            .24363     19.27301          6.00       359.00             0
HOTELLINGS         .32211     19.27301          6.00       359.00             0
WILKS              .75637     19.27301          6.00       359.00             0
ROYS               .24363

EFFECT .. FEM

MULTIVARIATE TESTS OF SIGNIFICANCE (S = 1, M = 2 , N = 178 1/2)

TEST NAME          VALUE     APPROX. 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
```

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 *F*'s, and stepdown *F*'s 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 9.16. (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 *F*'s is required to correctly interpret the pattern of stepdown *F*'s. 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 9.17 for the three effects in turn: FEM by MASC,

TABLE 9.16 POOLED WITHIN-CELL CORRELATIONS AMONG SIX DVs (SELECTED OUTPUT FROM SPSS' MANOVA—SEE TABLE 9.15 FOR SETUP)

WITHIN CELLS CORRELATIONS WITH STD. DEVS. ON DIAGONAL

	ESTEEM	ATTROLE	NEUROTIC	INTEXT	CONTROL	SEL2
ESTEEM	3.53338					
ATTROLE	.14528	5.22718				
NEUROTIC	.35772	.05070	4.96517			
INTEXT	-.16444	.01111	-.00923	3.58729		
CONTROL	.34789	-.03051	.38737	-.08315	1.26680	
SEL2	-.03510	.01635	-.01501	.05526	-.08417	25.50087

TABLE 9.17 UNIVARIATE ANAYSES 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 9.15 FOR SETUP)

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	14115.12120	36.79594	38.77781	.94888	.331
NEUROTIC	.20238	8973.67662	.20238	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	236708.96634	353.58143	650.29936	.54372	.461

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	0
ATTRO E	1426.75675	14115.12120	1426.75675	38.77781	36.79313	0
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	236708.96634	1105.38196	650.29936	1.69980	.193

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	14115.12120	610.88860	38.77781	15.75356	.000
NEUROTIC	44.05442	8973.67662	44.05442	24.65296	1.78688	.182
INTEXT	87.75996	4684.17900	87.75996	12.86862	6.81958	.009
CONTROL	2.83106	584.14258	2.83106	1.60479	1.76414	.185
SEL2	9.00691	236708.96634	9.00691	650.29936	.01385	.906

MASC, and FEM. F values are substantial for all DVs but SEL2 for MASC and for ESTEEM, ATTROLE, and INTEXT for FEM.

Finally, 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. 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 9.5.2.2), 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 9.18.

TABLE 9.18 STEPDOWN ANALYSES OF SIX ORDERED DVs FOR (TOP TO BOTTOM) FEM BY MASC INTERACTION, MASCULINITY, AND FEMININITY (SELECTED OUTPUT FROM SPSS' MANOVA—SEE TABLE 9.15 FOR SETUP)

EFFECT .. FEM BY MASC (CONT.)

ROY-BARGMAN STEPDOWN F - TESTS

VARIABLE	HYPOTH. MS	ERROR MS	STEP-DOWN F	HYPOTH. DF	ERROR DF	SIG. OF F
ESTEEM	17.48685	12.48474	1.40068	1	364	.237
ATTROLE	26.85653	38.06383	.85302	1	363	.420
NEUROTIC	2.69735	21.61699	.12478	1	362	.724
INTEXT	.26110	12.57182	.02079	1	361	.885
CONTROL	.01040	1.28441	.31955	1	360	.572
SEL2	297.09000	652.80588	.45510	1	359	.500

EFFECT .. MASC (CONT.)

ROY-BARGMAN STEPDOWN F - TESTS

VARIABLE	HYPOTH. MS	ERROR MS	STEP-DOWN F	HYPOTH. DF	ERROR DF	SIG. OF F
ESTEEM	979.60086	12.48474	78.46383	1	364	0
ATTROLE	726.51682	38.06383	19.13935	1	363	.000
NEUROTIC	4.14529	21.61699	.19176	1	362	.662
INTEXT	139.88354	12.57182	11.13471	1	361	.001
CONTROL	.00082	1.28441	.00064	1	360	.980
SEL2	406.59619	652.80588	.62284	1	359	.431

EFFECT .. FEM (CONT.)

ROY-BARGMAN STEPDOWN F - TESTS

VARIABLE	HYPOTH. MS	ERROR MS	STEP-DOWN 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	2.21545	21.61699	.10247	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

For purposes of journal reporting, critical information from Tables 9.17 and 9.18 is consolidated into a single table with both univariate and stepdown analyses, as shown in Table 9.19. The α level established for each DV is reported along with the significance levels for stepdown *F*. 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 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.)

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 9.20 contains deck setup and control language for these marginal means as produced through SPSS* MANOVA. In the table, level of effect is identified under PARAMETER and mean

TABLE 9.19 TESTS OF FEMININITY, MASCULINITY, AND THEIR INTERACTION

IV	DV	Univariate *F*	df	Stepdown *F*	df	α
Femininity	ESTEEM	8.13*	1/364	8.13**	1/364	.01
	ATTROLE	15.75*	1/364	19.15**	1/363	.01
	NEUROTIC	1.79	1/364	0.10	1/362	.01
	INTEXT	6.82*	1/364	3.82	1/361	.01
	CONTROL	1.76	1/364	0.05	1/360	.01
	SEL2	0.01	1/364	0.02	1/359	.001
Masculinity	ESTEEM	78.46*	1/364	78.46**	1/364	.01
	ATTROLE	36.79*	1/364	19.14**	1/363	.01
	NEUROTIC	7.28*	1/364	0.19	1/362	.01
	INTEXT	25.44*	1/364	11.13**	1/361	.01
	CONTROL	7.39*	1/364	0.00	1/360	.01
	SEL2	1.70	1/364	0.62	1/359	.001
Femininity by masculinity interaction	ESTEEM	1.40	1/364	1.40	1/364	.01
	ATTROLE	0.95	1/364	0.65	1/363	.01
	NEUROTIC	0.01	1/364	0.12	1/362	.01
	INTEXT	0.00	1/364	0.02	1/361	.01
	CONTROL	0.56	1/364	0.32	1/360	.01
	SEL2	0.54	1/364	0.46	1/359	.001

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

TABLE 9.20 ADJUSTED MARGINAL MEANS FOR ESTEEM: ATTROLE WITH ESTEEM AS A COVARIATE; AND INTEXT WITH ESTEEM, ATTROLE, AND NEUROTIC AS COVARIATES (SETUP AND SELECTED OUTPUT FROM SPSS' MANOVA)

```
TITLE          ADJUSTED MEANS FOR LARGE SAMPLE MANOVA
FILE HANDLE    MANOVA
DATA LIST      FILE=MANOVA
               /CASENO,ANDRM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
                SEL2,INTEXT,NEUROTIC (A4,6F3.0,F8.5,F4.1,F3.0)
VALUE LABELS   ANDRM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS'/
               FEM,MASC 1 'LOW' 2 'HIGH'
MISSING VALUES  CONTROL(0)
MANOVA    ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM(1,2) 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

CONSPLUS FEM                                CONSPLUS MASC

 PARAMETER          COEFF.                   PARAMETER          COEFF.

         1     16.5700934579                         2     17.1508559923
         2     15.4137931034                         3     13.7138143624

ESTIMATES FOR ATTROLE ADJUSTED FOR 1 COVARIATE

CONSPLUS FEM                                CONSPLUS MASC

 PARAMETER          COEFF.                   PARAMETER          COEFF.

         1     32.7152449841                         2     35.3913780733
         2     35.8485384127                         3     32.1158759823
```

TABLE 9.20 (Continued)

```
ESTIMATES FOR INTEXT ADJUSTED FOR 3 COVARIATES

CONSPLUS MASC

 PARAMETER          COEFF.

       2      11.0021422402
       3      12.4757067316

Note: COEFF. = adjusted marginal means; first parameter=low, second parameter=high
```

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 9.21 where means for NEUROTIC and CONTROL are found for the main effect of MASC adjusted for FEM.

Strength of association between the main effect and each DV with which it has a significant stepdown relationship is evaluated as η^2 (Equations 3.25, 3.26, 8.7, 8.8, or 8.9). The information you need for calculation of η^2 is available in SPSS MANOVA stepdown tables (see Table 9.18) but not in a convenient form; mean squares are given in the tables but you need sums of squares for calculation of η^2. Sums of squares are found by multiplying a mean square by its associated degrees of freedom. For this example, appropriate information is summarized in Table 9.22. Once sums of squares are computed, total sum of squares for a DV is calculated by adding the sums of squares for all effects and error.

TABLE 9.21 UNADJUSTED MARGINAL MEANS FOR NEUROTIC AND CONTROL (SETUP AND SELECTED OUTPUT FROM SPSS® MANOVA)

```
TITLE           'UNADJUSTED MEANS FOR LARGE SAMPLE MANOVA'
FILE HANDLE     MANOVA
DATA LIST       FILE=MANOVA
                /CASENO,ANORM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
                 SEL2,INTEXT,NEUROTIC (A6,6F3.0,F8.5,F4.1,F3.0)
VALUE LABELS    ANORM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS'/
                FEM,MASC 1 'LOW' 2 'HIGH'
MISSING VALUES     CONTROL(0)
MANOVA     ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM(1,2) MASC(1,2)/
           PRINT=PARAMETERS (ESTIM)/
           ANALYSIS=NEUROTIC/DESIGN=FEM, CONSPLUS MASC/
           ANALYSIS=CONTROL/ DESIGN=FEM, CONSPLUS MASC/


ESTIMATES FOR NEUROTIC

CONSPLUS MASC

 PARAMETER          COEFF.

       2       9.3709383047
       3       7.8861041130


ESTIMATES FOR CONTROL

CONSPLUS MASC

 PARAMETER          COEFF.

       2       6.8916331030
       3       6.5125816027

Note: COEFF. = marginal means; first parameter = low, second parameter = high
```

TABLE 9.22 SUMMARY OF ADJUSTED SUMS OF SQUARES AND η^2 FOR EFFECTS OF FEMININITY, MASCULINITY, AND THEIR INTERACTION ON ESTEEM; ATTROLE WITH ESTEEM AS A COVARIATE; AND INTEXT WITH ESTEEM, ATTROLE, AND NEUROTIC AS COVARIATES

Source of variance	ESTEEM		ATTROLE		INTEXT	
	SS′	η^2	SS′	η^2	SS′	η^2
F	101.47	.02	728.77	.05	47.99	—
M	979.60	.17	728.52	.05	139.98	.03
F by *M*	17.49	—	24.86	—	0.26	—
Error	4544.45		13817.17		4538.43	
Total	5643.01		15299.32		4726.67	

Note: See Table 9.20 for MS values. SS = (MS)(DF).

For example, ATTROLE is significantly related to FEM (F). For ATTROLE, the adjusted sum of squares is the mean square for hypothesis (HYPOTH. MS) times the degrees of freedom (HYPOTH. DF) from the FEM section of Table 9.18.

$$SS'_F = (728.76735)(1) = 728.77$$

Adjusted sums of squares for the remaining effects and error are necessary for computation of sum of squares total; these values are also available, but scattered around through Table 9.18. The adjusted sum of squares for MASC (middle part of Table 9.18) is

$$SS'_M = (728.51682)(1) = 728.52$$

and for FEM BY MASC (top part of Table 9.18) is

$$SS'_{F \text{ by } M} = (24.85653)(1) = 24.86$$

Error mean square and degrees of freedom are duplicated in all three segments of Table 9.18. Adjusted sum of squares for error, then, is

$$SS'_{error} = (38.06383)(363) = 13817.17$$

Finally, η^2 for ATTROLE is found by computing

$$\eta^2 = \frac{SS'_F}{SS'_{total}} = \frac{SS'_F}{SS'_F + SS'_M + SS'_{F \text{ by } M} + SS'_{error}}$$

$$= \frac{728.77}{728.77 + 728.52 + 24.86 + 13817.17} = \frac{728.77}{15299.32} = .05$$

A checklist for MANOVA appears in Table 9.23. An example of a Results section, in journal format, follows for the study just described.

TABLE 9.23 CHECKLIST FOR MULTIVARIATE ANALYSIS OF VARIANCE

1. Issues
   a. Unequal sample sizes and mising 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
   a. Multivariate strength of association
   b. Importance of DVs
      (1) Within-cell correlations, stepdown F, univariate F
      (2) Strength of association for significant F
      (3) Means or adjusted marginal and/or cell means for significant F
3. 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 $\alpha = .001$. 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 moderate association between masculinity scores (low vs. high) and the combined DVs, $\eta^2 = .24$. The association was less substantial between femininity and the DVs, $\eta^2 = .08$. [*F and lambda are from Table 9.15; η^2 is calculated according to Equation 9.6.*]

To investigate the impact of each main effect on the individual DVs, a 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 9.19. An experimentwise error rate of 5 percent was achieved by the apportionment of alpha as shown in the last column of Table 9.19 for each of the DVs.

Insert Table 9.19 about here

A unique contribution to predicting differences between those low and high on femininity was made by self-esteem, stepdown $F(1, 364) = 8.13$, $p < .01$, $\eta^2 = .02$. Self-esteem was scored inversely, so women with higher femininity scores showed greater self-esteem (mean self-esteem = 15.41) than those with lower femininity (mean self-esteem = 16.57). After the pattern of differences measured by self-esteem was entered, a difference was also found on attitude toward role of women, stepdown $F(1, 363) = 19.15$, $p < .01$, $\eta^2 = .05$. Women with higher femininity scores had more conservative attitudes toward women's

role (adjusted mean attitude = 35.85) than those lower in femininity (adjusted mean attitude = 32.72). Although a univariate comparison revealed that those higher in femininity also were more extraverted, 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-extravert—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 $F(1, 364) = 78.46$, $p < .01$, $\eta^2 = 17$. Women scoring high in masculinity had higher self-esteem (mean self-esteem = 13.71) than those scoring low (mean self-esteem = 17.16). With differences due to self-esteem already entered, ATTROLE made a unique contribution, stepdown $F(1, 363) = 19.14$, $p < .01$, $\eta^2 = .05$. Women scoring lower in masculinity had more conservative attitudes toward the proper role of women (adjusted mean attitude = 35.39) than those scoring higher (adjusted mean attitude = 32.12). Introversion-extraversion, adjusted by self-esteem, attitudes toward women's role, and neuroticism also made a unique contribution to the composite DV, stepdown $F(1, 361) = 11.13$, $p < .01$, $\eta^2 = .03$. Women with higher masculinity were more extraverted (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 9.20, η^2 values are from Table 9.22. Means adjusted for main effects but not other DVs for univariate interpretation are in Table 9.21.*]

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 9.16.

Insert Table 9.16 about here

9.7.3 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.

Deck setup and partial output of omnibus MANCOVA as produced by SPSS* MANOVA appear in Table 9.24. As in MANOVA, Method 3 adjustment for unequal n is used with MASC adjusted for FEM. And, as in MANOVA, both main effects are highly significant but there is no interaction.

9.7.3.1 Assessing Covariates Under EFFECT..WITHIN CELLS 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). 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 setup of Table 9.24 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 9.25.

TABLE 9.24 MULTIVARIATE ANALYSIS OF COVARIANCE OF ESTEEM, INTEXT, AND NEUROTIC AS A FUNCTION OF (TOP TO BOTTOM) FEM BY MASC INTERACTION, MASCULINITY, FEMININITY, COVARIATES ARE ATTROLE, CONTROL, AND SEL2 (SETUP AND SELECTED OUTPUT FROM SPSS' MANOVA)

```
TITLE          'LARGE SAMPLE MANCOVA'
FILE HANDLE    MANOVA
DATA LIST      FILE='MANOVA'
               /CASENO,ANDRM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
                SEL2,INTEXT,NEUROTIC (A4,6F3.0,F8.5,F4.1,F3.0)
VALUE LABELS   ANDRM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS' /
               FEM,MASC 1 'LOW' 2 'HIGH'
MISSING VALUES CONTROL(0)
MANOVA         ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM(1,2) MASC(1,2)/
               ANALYSIS=ESTEEM,INTEXT,NEUROTIC WITH CONTROL,ATTROLE,SEL2/
               NOPRINT=PARAMETERS(ESTIM)/
               PRINT=SIGNIF(STEPDOWN), ERROR(COR),
                     HOMOGENEITY(BARTLETT,COCHRAN,BOXM)/
               METHOD=SSTYPE(SEQUENTIAL)/
               DESIGN/
```

EFFECT .. WITHIN CELLS 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	10.00377	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				

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				

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	.03255	4.66837	3.00	359.00	.003
HOTELLINGS	.03901	4.66837	3.00	359.00	.003
WILKS	.96745	4.66837	3.00	359.00	.003
ROYS	.03755				

At the top of Table 9.25 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 9.7.3.2). At the bottom of Table 9.25 under REGRESSION ANALYSIS FOR WITHIN CELLS 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

TABLE 9.25 UNIVARIATE, STEPDOWN, AND MULTIPLE REGRESSION ANALYSES FOR THREE DVs WITH THREE COVARIATES (SELECTED OUTPUT FROM SPSS* MANOVA—SEE TABLE 9.24 FOR SETUP)

```
UNIVARIATE F-TESTS WITH (3,361) D. F.

VARIABLE    SQ. MUL. R    MUL. R  ADJ. R-SQ.  HYPOTH. MS   ERROR MS          F  SIG. OF F

ESTEEM         .16502     .38134     .13832   220.28068   10.75791   20.47616          0
INTENT         .00932     .09655     .00109    14.55535   12.85461    1.13231       .336
NEUROTIC       .15425     .39274     .14722   461.38686   21.02359   21.94615          0


EFFECT .. WITHIN CELLS     REGRESSION (CONT.)

ROY-BARGMAN STEPDOWN F - TESTS

VARIABLE      HYPOTH. MS     ERROR MS   STEP-DOWN F   HYPOTH. DF   ERROR DF   SIG. OF F

ESTEEM        220.28068     10.75791      20.47616            3        361           0
INTENT          6.35936     12.60679        .50444            3        360        .679
NEUROTIC      239.64208     18.72842      12.16164            3        359        .000


REGRESSION ANALYSIS FOR WITHIN CELLS     ERROR TERM

DEPENDENT VARIABLE .. ESTEEM

COVARIATE          B          BETA     STD. ERR.    T-VALUE   SIG. OF T  LOWER 95 CL  UPPER 95 CL

CONTROL    .9817325741   .3518752398     .13625    7.20542           0       .71078      1.24967
ATTROLE    .0886059312   .1561581235     .02762    3.20773        .001       .03428       .14293
SEL2      -.0011127953  -.0080312308     .00677    -.16446        .869      -.01442       .01219

DEPENDENT VARIABLE .. INTENT

COVARIATE          B          BETA     STD. ERR.    T-VALUE   SIG. OF T  LOWER 95 CL  UPPER 95 CL

CONTROL   -.2232208679  -.0798274507     .14894   -1.49877        .135      -.51611       .06967
ATTROLE    .0045581881   .0079125744     .00319     .15096        .880      -.05482       .06394
SEL2       .0068216006   .0484927597     .00740     .92231        .357      -.00772       .02137

DEPENDENT VARIABLE .. NEUROTIC

COVARIATE          B          BETA     STD. ERR.    T-VALUE   SIG. OF T  LOWER 95 CL  UPPER 95 CL

CONTROL   1.5312818992   .3906873702     .19047    8.03955           0      1.15671      1.90585
ATTROLE    .0497076488   .0623418410     .03861    1.28727        .199      -.02623       .12565
SEL2       .0032824106   .0168383493     .00946     .34703        .729      -.01532       .02188
```

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.

9.7.3.2 Assessing DVs Procedures for evaluating DVs, now adjusted for covariates, follow those specified in Section 9.7.2 for MANOVA. Correlations among all DVs, among covariates, and between DVs and covariates are informative so all the correlations in Table 9.16 are still relevant.[20]

Univariate *F*'s are now adjusted for covariates. The univariate ANCOVAs produced by the SPSS* MANOVA run specified in Table 9.24 are shown in Table 9.26. Although significance levels are misleading, there are substantial *F* values for ESTEEM and INTEXT for MASC (adjusted for FEM) and for FEM.

[20] 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 9.26 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 9.24 FOR SETUP)

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	.02590	4640.51295	.02590	12.85461	.00202	.964
NEUROTIC	1.52636	7589.51604	1.52636	21.02359	.07260	.788

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	0
INTEXT	264.44545	4640.51295	264.44545	12.85461	20.57204	.000
NEUROTIC	35.82929	7589.51604	35.82929	21.02359	1.70424	.193

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	74.94312	4640.51295	74.94312	12.85461	5.83006	.016
NEUROTIC	26.81431	7589.51604	26.81431	21.02359	1.27544	.259

TABLE 9.27 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 9.24 FOR SETUP)

EFFECT .. FEM BY MASC (CONT.)

ROY-BARGMAN STEPDOWN F - TESTS

VARIABLE	HYPOTH. MS	ERROR MS	STEP-DOWN 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	4.98321	19.72942	.25255	1	359	.617

EFFECT .. MASC (CONT.)

ROY-BARGMAN STEPDOWN F - TESTS

VARIABLE	HYPOTH. MS	ERROR MS	STEP-DOWN F	HYPOTH. DF	ERROR DF	SIG. OF F
ESTEEM	533.61774	10.75791	49.60237	1	361	0
INTEXT	137.74436	12.60679	10.92621	1	360	.001
NEUROTIC	1.07421	19.72942	.05445	1	359	.816

EFFECT .. FEM (CONT.)

ROY-BARGMAN STEPDOWN F - TESTS

VARIABLE	HYPOTH. MS	ERROR MS	STEP-DOWN F	HYPOTH. DF	ERROR DF	SIG. OF F
ESTEEM	107.44454	10.75791	9.98749	1	361	.00
INTEXT	47.36159	12.60679	3.75683	1	360	.05
NEUROTIC	4.23502	19.72942	.21466	1	359	.64

TABLE 9.28 TESTS OF COVARIATES, FEMININITY, MASCULINITY, AND INTERACTION

Effect	DV	Univariate *F*	df	Stepdown *F*	df	α
Covariates	ESTEEM	20.48*	3/361	20.48**	3/361	.02
	INTEXT	1.13	3/361	0.50	3/360	.02
	NEUROTIC	21.95*	3/361	12.16**	3/359	.01
Femininity	ESTEEM	9.99*	1/361	9.99**	1/361	.02
	INTEXT	5.83*	1/361	3.76	1/360	.02
	NEUROTIC	1.28	1/361	0.21	1/359	.01
Masculinity	ESTEEM	49.60*	1/361	49.60**	1/361	.02
	INTEXT	20.57*	1/361	10.93**	1/360	.02
	NEUROTIC	1.70	1/361	0.05	1/359	.01
Femininity by masculinity interaction	ESTEEM	0.67	1/361	0.67	1/361	.02
	INTEXT	0.00	1/361	0.03	1/360	.02
	NEUROTIC	0.07	1/361	0.25	1/359	.01

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

TABLE 9.29 ADJUSTED MARGINAL MEANS FOR ESTEEM ADJUSTED FOR THREE COVARIATES, AND INTEXT ADJUSTED FOR ESTEEM PLUS THREE COVARIATES (SETUP AND SELECTED OUTPUT FROM SPSS* MANOVA)

```
TITLE          ADJUSTED MEANS FOR LARGE SAMPLE MANOVA
FILE HANDLE    MANOVA
DATA LIST      FILE=MANOVA
               /CASENO,ANDRM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
                SEL2,INTEXT,NEUROTIC (A6,6F3.0,F8.5,F4.1,F3.0)
VALUE LABELS   ANDRM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS'/
               FEM,MASC 1 'LOW' 2 'HIGH'
MISSING VALUES  CONTROL(0)
MANOVA    ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SEL2 BY FEM(1,2) 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

 CONSPLUS FEM                                  CONSPLUS MASC

  PARAMETER          COEFF.                     PARAMETER          COEFF.

          1     16.6136356957                           2     16.9194927002
          2     15.3959424542                           3     14.2190000415


ESTIMATES FOR INTEXT ADJUSTED FOR 4 COVARIATES

 CONSPLUS MASC

  PARAMETER          COEFF.

          2     11.0067954086
          3     12.4700349085
```

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

TABLE 9.30 MARGINAL MEANS FOR INTEXT ADJUSTED FOR THREE COVARIATES ONLY (SETUP AND SELECTED OUTPUT FROM SPSS' MANOVA)

```
TITLE          'ADJUSTED MEANS FOR LARGE SAMPLE MANOVA'
FILE HANDLE    MANOVA
DATA LIST      FILE='MANOVA'
               /CASEND,ANORM,FEM,MASC,ESTEEM,CONTROL,ATTROLE,
                SELF,INTEXT,NEUROTIC (A4,6F3.0,F8.5,F4.1,F3.0)
VALUE LABELS   ANORM 1 'UNDIFF' 2 'FEMNINE' 3 'MASCLNE' 4 'ANDROGNS' /
               FEM,MASC 1 'LOW' 2 'HIGH'
MISSING VALUES CASEND TO ATTROLE(0)
MANOVA    ESTEEM,ATTROLE,NEUROTIC,INTEXT,CONTROL,SELF BY FEM(1,2) MASC(1,2)/
          PRINT='PARAMETERS (ESTIM)'/
       ANALYSIS='INTEXT WITH CONTROL,ATTROLE,SELF,ESTEEM'/
       DESIGN='CONSPLUS FEM'/

ESTIMATES FOR INTEXT ADJUSTED FOR 4 COVARIATES

CONSPLUS FEM

 PARAMETER          COEFF.

         1      11.08284034548
         2      11.91221113039
```

Note: COEFF. = adjusted marginal means; first parameter = low; second parameter = high.

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

TABLE 9.31 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
   a. Multivariate strength of association
   b. Importance of DVs
      (1) Within-cell correlations, stepdown F, univariate F
      (2) Strength of association for significant F
      (3) Adjusted marginal and/or cell means for significant F
3. Additional analyses
   a. Assessment of covariates
   b. Interpretation of IV-covariates interaction (if homogeneity of regression violated for stepdown analysis)
   c. Post hoc comparisons

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 9.27. 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 9.26 and 9.27, as well as some information from Table 9.25, appears in Table 9.28, along with apportionment of the .05 α error to the various tests.

For the DVs associated with significant main effects, interpretation requires associated marginal means. Table 9.29 contains setup 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. Setup and marginal means for the main effect of FEM on INTEXT (univariate but not stepdown effect) appear in Table 9.30. A checklist for MANCOVA appears in Table 9.31. An example of a Results section, as might be appropriate for journal presentation, follows.

Results

A 2 × 2 between-subjects multivariate analysis of covariance was performed on three dependent variables associated with personality of respondents: self-esteem, introvert-extravert, 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).

SPSS MANOVA was used for the analyses with the hierarchical 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.

With the use of Wilks' criterion, the combined DVs were significantly related to the combined covariates, approximate $\underline{F}(9, 873) = 10.86$, $\underline{p} < .01$, to femininity, $\underline{F}(3, 359) = 4.67$, $\underline{p} < .01$, and to masculinity, $\underline{F}(3, 359) = 20.59$, $\underline{p} < .001$ but not to the interaction, $\underline{F}(3, 359) = 0.31$, $\underline{p} > .05$. There was a moderate association between DVs and covariates, with $\eta^2 = .23$. A somewhat smaller association was found between combined DVs and the main effect of masculinity, $\eta^2 = .15$, and the association between the main effect of femininity and the combined DVs was smaller yet, $\eta^2 = .04$. [*F and lambda are from Table 9.24; η^2 is calculated from Equation 9.6.*]

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 β value of .35 for locus of control was significantly different from zero, $\underline{t}(361) = 7.21$, $\underline{p} < .001$, as was the β value of .16 for attitudes toward women's role, $\underline{t}(361) = 3.21$, $\underline{p} < .01$. None of the covariates provided adjustment to the introversion-extraversion scale. For neuroticism, only locus of control reached statistical significance, with $\beta = .39$, $\underline{t}(361) = 8.04$, $\underline{p} < .001$. For none of the DVs did socioeconomic status provide significant adjustment.

Effects of masculinity and femininity on the DVs after adjustment for covariates were investigated in univariate and 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 9.28. 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.

Insert Table 9.28 about here

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 $F(1, 361) = 9.99$, $p < .01$, $\eta^2 = .02$. With self-esteem scored inversely, women with higher femininity scores showed greater self-esteem after adjustment for covariates (adjusted mean self-esteem = 15.40) than those scoring lower on femininity (adjusted mean self-esteem = 16.61). Univariate analysis revealed that a reliable 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 9.29 and 9.30; η^2 is calculated as in Section 9.7.2.*]

Lower- versus higher-masculinity women differed in self-esteem, the highest-priority DV, after adjustment for covariates, stepdown $F(1, 361) = 49.60$, $p < .01$, $\eta^2 = .10$. Greater self-esteem was found among higher-masculinity women (adjusted mean = 14.22) than among lower-masculinity

women (adjusted mean = 16.92). The measure of introversion and extraversion, adjusted for covariates and self-esteem, was also related to differences in masculinity, stepdown $F(1, 360) = 10.93$, $p < .01$, $\eta^2 = .03$. Women scoring higher on the masculinity scale were more extraverted (adjusted mean extraversion 12.47) than those showing lower masculinity (adjusted mean extraversion = 11.01).

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 9.16.

Insert Table 9.16 about here

9.8 SOME EXAMPLES FROM THE LITERATURE

9.8.1 Examples of MANOVA

Wade and Baker (1977) surveyed clinical psychologists on their reasons for decisions to use tests. Psychologists were divided into low and high test users, as the two levels of the IV. Six "reasons," each rated on a 5-point scale of importance, served as the DVs. Separate MANOVAs were performed on users of objective and projective tests. After finding significant multivariate effects—that is, that high and low test users rated importance of reasons differently—for both objective and projective users, the researchers performed univariate ANOVAs on each of the reasons.

They found that for objective test users, the distinguishing reasons were reliability and validity, and agency requirements, for which high users gave higher ratings of importance than did low users. For the projective test users, high users rated as more important than low users graduate training experience, previous experience with tests, and reliability and validity. A MANOVA testing reasons for use as a function of eight major therapeutic orientations was not statistically significant.

In a study of romantic attraction, Giarrusso (1977) investigated the effects of physical attractiveness (low vs. high) and similarity of attitudes (similar vs. different) of a target date in a 2 × 2 between-groups MANOVA. Male students rated the target date on degree of liking, desire to date, comparability with previous dates, goodness of personality, and whether the target date liked them. These five ratings served as DVs. A significant multivariate effect of physical attractiveness was found, but not of similarity of attitudes or of the interaction between attractiveness and attitudes. After an a priori ordering of DVs, it was found that only the highest-priority variable, degree of liking, was significantly affected by physical attractiveness. That is, the more attractive the target date, the more she was liked. The remaining variables showed no statistically significant differences once the effect on degree of liking was partialed out.

Junior high school students were evaluated in terms of their behavioral changes after being exposed in class for one week to materials on "responsibility," in a study by Singh, Greer, and Hammond (1977). IVs were (1) whether or not materials were presented; (2) grade level: 7, 8, and 9; and (3) ability levels, three levels based on the School and College Ability Tests. The three DVs consisted of an attitude test designed to accompany the "responsibility" materials, an experimenter-designed essay, and objective tests, which tapped comprehension of various aspects of responsibility. With the use of a 2 × 3 × 3 between-groups multivariate analysis of variance on gain scores, only the main effect of program was found to be statistically significant. Bonferroni-type confidence intervals constructed for the three DVs, used instead of stepdown analysis, indicated that the attitude test alone accounted for the significant multivariate effect, but the authors felt that the magnitude of change was insufficient to support the program as implemented in the study.

Concerned that policy and decision makers seldom use evaluation information, Brown, Braskamp, and Newman (1978) investigated effects on readers' reactions to use of jargon and to objective vs. subjective statements in evaluation reports. After reading one of four statements (with presence vs. absence of jargon, and subjective vs. objective statements factorially combined), readers judged the reports on two dependent measures, difficulty and technicality, and judged the writer of the evaluation on nine measures (including such variables as thoroughness and believability). Separate MANOVAs were performed on two types of DVs—judgment of report and judgment of writer. Significant multivariate effects were interpreted in terms of cell means for DVs with significant univariate F's, rather than with the combination of univariate and stepdown analysis recommended here.

The only significant multivariate effect was that of jargon on ratings of the report. Both report DVs, technicality and difficulty, produced significant univariate F's, indicating that reports high in jargon were considered to be more technical and more difficult. Ratings of the report writer were not affected by jargon, objectivity, or the interaction. Further, a univariate ANOVA revealed that extent of agreement with the recommendations of the evaluation report was independent of use of jargon, objectivity, or the interaction.

Jakubczak (1977) was interested in age differences in regulation of caloric intake of rats. For the first experiment, he used as IVs three age levels, three levels of dilution of food with cellulose, two levels of previous food deprivation experience,

and six daysets, a within-subjects factor. DVs were caloric intake and body weight. Significant multivariate main effects were found on all factors, in addition to a number of two- and three-way interactions. Univariate ANOVAs, rather than stepdown and univariate analyses, were then used for interpreting effects on individual DVs. An age-related decrement in caloric intake and body weight was found in response to dilution with cellulose. Previous deprivation experience did not affect this relationship.

In a second experiment, two age levels, two levels of dilution by water, and nine daysets were used as factors. As an additional IV, rats were or were not given quinine in their diets. Older rats behaved like younger rats in response to dilution of the diet by water, maintaining caloric intake and weight. With addition of quinine, however, older rats decreased caloric intake and body weight to a greater degree than younger rats.

For both experiments, MANCOVAs run with initial levels of caloric intake and body weight as covariates revealed no difference in decisions with respect to within-subjects effects, as could be expected (cf. Section 8.5.2.1). Jakubczak made no mention of changes in between-subjects effects with the use of covariates.

Willis and Wortman (1976) studied public acceptance of randomized control-group designs in medical experimentation. Factors were sex of subject, three levels of science emphasis in describing the medical experiment, and three levels of treatment scarcity situations. For the last IV, descriptions of the medical experiment explicitly mentioned nonscarcity (i.e., the proposed treatment was available to all), mentioned scarcity (the treatment was expensive and scarce), or failed to mention anything about scarcity. Dependent measures were nine opinion questions, answered on bipolar scales. MANOVA revealed significant multivariate effects for two of the three IVs, scarcity of treatment and sex of subject. None of the interactions was statistically significant. DVs were analyzed in terms of cell means for scarcity and sex associated with those DVs with significant univariate F's. For post hoc comparisons, among the three scarcity levels, multivariate F's were appropriately adjusted for inflated α error. It was concluded that the most favorable reactions occurred in situations in which scarcity was not mentioned at all, and that women tended to be less accepting than men of placebos and withholding of treatment.

9.8.2 Examples of MANCOVA

Cornbleth (1977) studied the effect of a protected hospital ward on geriatric patients. IVs were ward assignment (protected or unprotected) and identification of the patient as a wanderer or nonwanderer. These two IVs were factorially combined, forming a 2×2 between-subjects design. Separate analyses were run for two periods in time (rather than including time period as a within-subjects factor). DVs were two physical measures, five cognitive measures, and six psychosocial measures. Separate MANOVAs were done for each of these three categories of functioning. In additional analyses, preassessment measures on the appropriate DVs were used as covariates. Therefore, there were three MANOVAs and three MANCOVAs. Contribution of the DVs to multivariate effects were evaluated through examination of standardized discriminant function coefficients (cf. Chapter 11). In addition, means adjusted for covariates were reported for all DVs.

On physical variables, a significant multivariate interaction was found (marginal at time 1, $p < .05$ at time 2). The protected ward produced greater range of motion for wanderers, while the regular nonprotected ward facilitated performance for nonwanderers. In addition, independent of ward assignment, wanderers showed a lower level of psychosocial functioning than nonwanderers at both time points.

Modification of perceived locus of control in high school students was studied by Bradley and Gaa (1977). Subjects were blocked on sex and on two levels of prior achievement, and then randomly assigned to one of three treatment groups, goal-setting, conference (serving as a placebo), and control. Five variables measuring aspects of perceived locus of control were used as DVs. The covariate was a teacher-developed achievement pretest. In the report of results, no mention was made of main effects of the blocking IVs or any interactions between them and the manipulated IV.

Three separate analyses were reported, reflecting specific comparisons among groups rather than a single omnibus analysis of the three groups. These three comparisons were not mutually orthogonal, and they required more degrees of freedom than the omnibus test. However, no mention was made of adjusting (by Bonferroni procedures or others) for inflated α produced by multiple testing. Further, although it was reported that three MANCOVAs were performed, no multivariate statistics were presented. Instead, univariate F's for each DV were reported for each of the three analyses. It was concluded that goal-setting treatment was effective in promoting more internal orientation among students, but only on measures related to locus of control in an academic situation.

Appendix A

A Skimpy Introduction to Matrix Algebra

The purpose of this appendix is to provide readers with sufficient background to follow, and duplicate as desired, calculations illustrated in the fourth sections of Chapters 5 through 12. The purpose is not to provide a thorough review of matrix algebra or even to facilitate an in-depth understanding of it. The reader who is interested in more than calculational rules has several excellent discussions available, particularly those in Tatsuoka (1971) and Rummel (1970).

Most of the algebraic mnaipulations with which the reader is familiar—addition, subtraction, multiplication, and division—have counterparts in matrix algebra. In fact, the algebra that most of us learned is a special case of matrix algebra involving only a single number, a scalar, instead of an ordered array of numbers, a matrix. Some generalizations from scalar algebra to matrix algebra seem "natural" (i.e., matrix addition and subtraction) while others (multiplication and division) are convoluted. Nonetheless, matrix algebra provides an extremely powerful and compact method for manipulating sets of numbers to arrive at desirable statistical products.

The matrix calculations illustrated here are calculations performed on square matrices. Square matrices have the same number of rows as columns. Sums-of-squares and cross-products matrices, variance-covariance matrices, and correlation matrices are all square. In addition, these three very commonly encountered matrices are symmetrical, having the same value in row 1, column 2, as in column 1, row 2, and so forth. Symmetrical matrices are mirror images of themselves about the main diagonal (the diagonal going from top left to bottom right in the matrix).

There is a more complete matrix algebra that includes nonsquare matrices as well. However, once one proceeds from the data matrix, which has as many rows as research units (subjects) and as many columns as variables, to the sum-of-squares and cross-products matrix, as illustrated in Section 1.5, most calculations illustrated in this book involve square, symmetrical matrices. A further restriction on this appendix is to limit the discussion to only those manipulations used in the fourth sections of Chapters 5 through 12. For purposes of numerical illustration, two very simple matrices, square, but not symmetrical (to eliminate any uncertainty regarding which elements are involved in calculations), will be defined as follows:

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} 3 & 2 & 4 \\ 7 & 5 & 0 \\ 1 & 0 & 8 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} r & s & t \\ u & v & w \\ x & y & z \end{bmatrix} = \begin{bmatrix} 6 & 1 & 0 \\ 2 & 8 & 7 \\ 3 & 4 & 5 \end{bmatrix}$$

A.1 ADDITION OR SUBTRACTION OF A CONSTANT TO A MATRIX

If one has a matrix, A, and wants to add or subtract a constant, k, to the elements of the matrix, one simply adds (or subtracts) the constant to every element in the matrix.

$$\mathbf{A} + k = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} + k$$

$$= \begin{bmatrix} a + k & b + k & c + k \\ d + k & e + k & f + k \\ g + k & h + k & i + k \end{bmatrix} \tag{A.1}$$

If $k = -3$, then

$$\mathbf{A} + k = \begin{bmatrix} 0 & -1 & 1 \\ 4 & 2 & -3 \\ -2 & -3 & 5 \end{bmatrix}$$

A.2 MULTIPLICATION OR DIVISION OF A MATRIX BY A CONSTANT

Multiplication or division of a matrix by a constant is a straightforward process.

$$k\mathbf{A} = k \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$$k\mathbf{A} = \begin{bmatrix} ka & kb & kc \\ kd & ke & kf \\ kg & kh & ki \end{bmatrix} \tag{A.2}$$

and

$$\frac{1}{k}\mathbf{A} = \begin{bmatrix} \frac{a}{k} & \frac{b}{k} & \frac{c}{k} \\ \frac{d}{k} & \frac{e}{k} & \frac{f}{k} \\ \frac{g}{k} & \frac{h}{k} & \frac{i}{k} \end{bmatrix} \tag{A.3}$$

Numerically, if $k = 2$, then

$$k\mathbf{A} = \begin{bmatrix} 6 & 4 & 8 \\ 14 & 10 & 0 \\ 2 & 0 & 16 \end{bmatrix}$$

A.3 ADDITION AND SUBTRACTION OF TWO MATRICES

These procedures are straightforward, as well as useful. If matrices **A** and **B** are as defined at the beginning of this appendix, one simply performs the addition or subtraction of corresponding elements.

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} + \begin{bmatrix} r & s & t \\ u & v & w \\ x & y & z \end{bmatrix}$$

$$= \begin{bmatrix} a + r & b + s & c + t \\ d + u & e + v & f + w \\ g + x & h + y & i + z \end{bmatrix} \tag{A.4}$$

and

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} a - r & b - s & c - t \\ d - u & e - v & f - w \\ g - x & h - y & i - z \end{bmatrix} \tag{A.5}$$

For the numerical example:

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 3 & 2 & 4 \\ 7 & 5 & 0 \\ 1 & 0 & 8 \end{bmatrix} + \begin{bmatrix} 6 & 1 & 0 \\ 2 & 8 & 7 \\ 3 & 4 & 5 \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 9 & 3 & 4 \\ 9 & 13 & 7 \\ 4 & 4 & 13 \end{bmatrix}$$

Calculation of a difference between two matrices is required when, for instance, one desires a residuals matrix, the matrix obtained by subtracting a reproduced matrix from an obtained matrix (as in factor analysis, Chapter 12). Or, if the matrix that is subtracted happens to consist of columns with appropriate means of variables inserted in every slot, then the difference between it and a matrix of raw scores produces a deviation matrix.

A.4 MULTIPLICATION, TRANSPOSES, AND SQUARE ROOTS OF MATRICES

Matrix multiplication is both unreasonably complicated and undeniably useful. Note that the ijth element of the resulting matrix is a function of row i of the first matrix and column j of the second.

$$\mathbf{AB} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} r & s & t \\ u & v & w \\ x & y & z \end{bmatrix}$$

$$= \begin{bmatrix} ar + bu + cx & as + bv + cy & at + bw + cz \\ dr + eu + fx & ds + ev + fy & dt + ew + fz \\ gr + hu + ix & gs + hv + iy & gt + hw + iz \end{bmatrix} \tag{A.6}$$

Numerically,

$$\mathbf{AB} = \begin{bmatrix} 3 & 2 & 4 \\ 7 & 5 & 0 \\ 1 & 0 & 8 \end{bmatrix} \begin{bmatrix} 6 & 1 & 0 \\ 2 & 8 & 7 \\ 3 & 4 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \cdot 6 + 2 \cdot 2 + 4 \cdot 3 & 3 \cdot 1 + 2 \cdot 8 + 4 \cdot 4 & 3 \cdot 0 + 2 \cdot 7 + 4 \cdot 5 \\ 7 \cdot 6 + 5 \cdot 2 + 0 \cdot 3 & 7 \cdot 1 + 5 \cdot 8 + 0 \cdot 4 & 7 \cdot 0 + 5 \cdot 7 + 0 \cdot 5 \\ 1 \cdot 6 + 0 \cdot 2 + 8 \cdot 3 & 1 \cdot 1 + 0 \cdot 8 + 8 \cdot 4 & 1 \cdot 0 + 0 \cdot 7 + 8 \cdot 5 \end{bmatrix}$$

$$= \begin{bmatrix} 34 & 35 & 34 \\ 52 & 47 & 35 \\ 30 & 33 & 40 \end{bmatrix}$$

Regrettably, $\mathbf{AB} \neq \mathbf{BA}$ in matrix algebra. Thus

$$\mathbf{BA} = \begin{bmatrix} 6 & 1 & 0 \\ 2 & 8 & 7 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 3 & 2 & 4 \\ 7 & 5 & 0 \\ 1 & 0 & 8 \end{bmatrix}$$

$$\mathbf{BA} = \begin{bmatrix} 25 & 17 & 24 \\ 69 & 44 & 64 \\ 42 & 26 & 52 \end{bmatrix}$$

If another concept of matrix algebra is introduced, some useful statistical properties of matrix algebra can be shown. The transpose of a matrix is indicated by a prime (′) and stands for a rearrangement of the elements of the matrix such that the first row becomes the first column, the second row the second column, and so forth. Thus

$$\mathbf{A}' = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 7 & 1 \\ 2 & 5 & 0 \\ 4 & 0 & 8 \end{bmatrix} \tag{A.7}$$

When transposition is used in conjunction with multiplication, then some advantages of matrix multiplication become clear, namely,

$$\mathbf{AA}' = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$$

$$= \begin{bmatrix} a^2 + b^2 + c^2 & ad + be + cf & ag + bh + ci \\ ad + be + cf & d^2 + e^2 + f^2 & dg + eh + fi \\ ag + bh + ci & dg + eh + fi & g^2 + h^2 + i^2 \end{bmatrix} \tag{A.8}$$

The elements in the main diagonal are the sums of squares while those off the diagonal are cross products.

Had **A** been multiplied by itself, rather than by a transpose of itself, a different result would have been achieved.

$$\mathbf{AA} = \begin{bmatrix} a^2 + bd + cg & ab + be + ch & ac + bf + ci \\ da + ed + fg & db + e^2 + fh & dc + ef + fi \\ ga + hd + ig & gb + he + ih & gc + hf + i^2 \end{bmatrix}$$

If $\mathbf{AA} = \mathbf{C}$, then $\mathbf{C}^{1/2} = \mathbf{A}$. That is, there is a parallel in matrix algebra to squaring and taking the square root of a scalar, but it is a complicated business because of the complexity of matrix multiplication. If, however, one has a matrix **C** from which a square root is desired (as in canonical correlation, Chapter 6), one searches for a matrix, **A**, which, when multiplied by itself, produces **C**. If, for example,

$$\mathbf{C} = \begin{bmatrix} 27 & 16 & 44 \\ 56 & 39 & 28 \\ 11 & 2 & 68 \end{bmatrix}$$

then

$$\mathbf{C}^{1/2} = \begin{bmatrix} 3 & 2 & 4 \\ 7 & 5 & 0 \\ 1 & 0 & 8 \end{bmatrix}$$

A.5 MATRIX "DIVISION" (INVERSES AND DETERMINANTS)

If you liked matrix multiplication, you'll love matrix inversion. Logically, the process is analogous to performing division for single numbers by finding the reciprocal of the number and multiplying by the reciprocal: if $a^{-1} = 1/a$, then $(a)(a^{-1}) = a/a = 1$. That is, the reciprocal of a scalar is a number that, when multiplied by the number itself, equals 1. Both the concepts and the notation are similar in matrix algebra, but they are complicated by the fact that a matrix is an array of numbers.

To determine if the reciprocal of a matrix has been found, one needs the matrix equivalent of the 1 as employed in the preceding paragraph. The identity matrix, **I**, a matrix with 1's in the main diagonal and zeros elsewhere, is such a matrix. Thus

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.9}$$

Matrix division, then, becomes a process of finding $\mathbf{A}^{-1}$ such that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \tag{A.10}$$

One way of finding $\mathbf{A}^{-1}$ requires a two-stage process, the first of which consists of finding the determinant of **A**, noted $|\mathbf{A}|$. The determinant of a matrix is sometimes said to represent the generalized variance of the matrix, as most readily seen in a 2 × 2 matrix. Thus we define a new matrix as follows:

$$\mathbf{D} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

where

$$|\mathbf{D}| = ad - bc \tag{A.11}$$

If **D** is a variance-covariance matrix where a and d are variances while b and c are covariances, then $ad - bc$ represents variance minus covariance. It is this property of determinants that makes them useful for hypothesis testing (see, for example, Chapter 9, Section 9.4, where Wilks' Lambda is used in MANOVA).

Calculation of determinants becomes rapidly more complicated as the matrix gets larger. For example, in our 3 by 3 matrix,

$$|\mathbf{A}| = a(ei - fh) + b(fg - di) + c(dh - eg) \tag{A.12}$$

Should the determinant of **A** equal zero, then the matrix cannot be inverted because the next operation in inversion would involve division by zero. Multicollinear or singular matrices (those with variables that are linear combinations of one another, as discussed in Chapter 4) have zero determinants that prohibit inversion.

A full inversion of **A** is

$$\begin{aligned} \mathbf{A}^{-1} &= \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} \\ &= \frac{1}{|\mathbf{A}|} \begin{bmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{bmatrix} \end{aligned} \tag{A.13}$$

Please recall that because **A** is not a variance-covariance matrix, a negative determinant is possible, even somewhat likely. Thus, in the numerical example,

$$\begin{aligned} |\mathbf{A}| &= 3(5 \cdot 8 - 0 \cdot 0) + 2(0 \cdot 1 - 7 \cdot 8) + 4(7 \cdot 0 - 5 \cdot 1) \\ &= -12 \end{aligned}$$

and

$$\begin{aligned} \mathbf{A}^{-1} &= \frac{1}{-12} \begin{bmatrix} 5 \cdot 8 - 0 \cdot 0 & 4 \cdot 0 - 2 \cdot 8 & 2 \cdot 0 - 4 \cdot 5 \\ 0 \cdot 1 - 7 \cdot 8 & 3 \cdot 8 - 4 \cdot 1 & 4 \cdot 7 - 3 \cdot 0 \\ 7 \cdot 0 - 5 \cdot 1 & 2 \cdot 1 - 3 \cdot 0 & 3 \cdot 5 - 2 \cdot 7 \end{bmatrix} \\ &= \begin{bmatrix} {}^{40}/_{-12} & {}^{-16}/_{-12} & {}^{-20}/_{-12} \\ {}^{-56}/_{-12} & {}^{20}/_{-12} & {}^{28}/_{-12} \\ {}^{-5}/_{-12} & {}^{2}/_{-12} & {}^{1}/_{-12} \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} -3.33 & 1.33 & 1.67 \\ 4.67 & -1.67 & -2.33 \\ .42 & -.17 & -.08 \end{bmatrix}$$

Confirm that, within rounding error, Equation A.10 is true. Once the inverse of **A** is found, "division" by it is accomplished whenever required by using the inverse and performing matrix multiplication.

A.6 EIGENVALUES AND EIGENVECTORS: PROCEDURES FOR CONSOLIDATING VARIANCE FROM A MATRIX

We promised you a demonstration of computation of eigenvalues and eigenvectors for a matrix, so here it is. Like a Chinese dinner, however, you may well find that this discussion satisfies your appetite for only a couple of hours. During that time, round up Tatsuoka (1971), get the cat off your favorite chair, and prepare for an intelligible, if somewhat lengthy, description of the same subject.

Most of the multivariate procedures rely on eigenvalues and their corresponding eigenvectors (also called characteristic roots and vectors) in one way or another because they consolidate the variance in a matrix (the eigenvalue) while providing the linear combination of variables (the eigenvector) to do it. The coefficients applied to variables to form linear combinations of variables in all the multivariate procedures are rescaled elements from eigenvectors. The variance that the solution "accounts for" is associated with the eigenvalue, and is sometimes called so directly.

Calculation of eigenvalues and eigenvectors is best left up to a computer with any realistically sized matrix. For illustrative purposes, a 2 × 2 matrix will be used here. The logic of the process is also somewhat difficult, involving several of the more abstract notions and relations in matrix algebra, including the equivalence between matrices, systems of linear equations with several unknowns, and roots of polynomial equations.

Solutions of an eigenproblem involves solution of the following equation:

$$(\mathbf{D} - \lambda \mathbf{I})V = 0 \tag{A.14}$$

where λ is the eigenvalue and V the eigenvector to be sought. Expanded, this equation becomes

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

or

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

or, by applying Equation A.5,

$$\begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 \tag{A.15}$$

If one considers the matrix **D**, whose eigenvalues are sought, a variance-covariance matrix, one can see that a solution is desired to "capture" the variance in **D** while rescaling the elements in **D** by v_1 and v_2 to do so.

It is obvious from Equation A.15 that a solution is always available when v_1 and v_2 are zero. A nontrivial solution may also be available when the determinant of the leftmost matrix in Equation A.15 is zero.[1] That is, if (following Equation A.11)

$$(a-\lambda)(d-\lambda) - bc = 0 \tag{A.16}$$

then there may exist values of λ and values of v_1 and v_2 that satisfy the equation and are not zero. However, expansion of Equation A.16 gives a polynomial equation, in λ, of degree 2:

$$\lambda^2 - (a+d)\lambda + ad - bc = 0 \tag{A.17}$$

Solving for the eigenvalues, λ, requires solving for the roots of this polynomial. If the matrix has certain properties (see footnote 1), there will be as many positive roots to the equation as there are rows (or columns) in the matrix.

If Equation A.17 is rewritten as $x\lambda^2 + y\lambda + z = 0$, the roots may be found by applying the following equation:

$$\lambda = \frac{-y \pm \sqrt{y^2 - 4xz}}{2x} \tag{A.18}$$

For a numerical example, consider the following matrix.

$$\mathbf{D} = \begin{bmatrix} 5 & 1 \\ 4 & 2 \end{bmatrix}$$

Applying Equation A.17, we obtain

$$\lambda^2 - (5+2)\lambda + 5 \cdot 2 - 1 \cdot 4 = 0$$

or

$$\lambda^2 - 7\lambda + 6 = 0$$

The roots to this polynomial may be found by Equation A.18 as follows:

$$\lambda_1 = \frac{-(-7) + \sqrt{(-7)^2 - 4 \cdot 1 \cdot 6}}{2 \cdot 1}$$

$$= 6$$

[1] Re. d Tatsuoka (1971); a matrix is said to be positive definite when all $\lambda_i > 0$, positive semidefinite when all $\lambda_i \geq 0$, and ill-conditioned when some $\lambda_i < 0$.

and

$$\lambda_2 = \frac{-(-7) - \sqrt{(-7)^2 - 4 \cdot 1 \cdot 6}}{2 \cdot 1}$$

$$= 1$$

(The roots could also be found by factoring to get $[\lambda - 6][\lambda - 1]$.)

Once the roots are found, they may be used in Equation A.15 to find v_1 and v_2, the eigenvector. There will be one set of eigenvectors for the first root and a second set for the second root. Both solutions require solving sets of two simultaneous equations in two unknowns, to wit, for the first root, 6, and applying Equation A.15.

$$\begin{bmatrix} 5-6 & 1 \\ 4 & 2-6 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

or

$$\begin{bmatrix} -1 & 1 \\ 4 & -4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

so that

$$-1v_1 + 1v_2 = 0$$

and

$$4v_1 - 4v_2 = 0$$

When $v_1 = 1$ and $v_2 = 1$, a solution is found.

For the second root, 1, the equations become

$$\begin{bmatrix} 5-1 & 1 \\ 4 & 2-1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

or

$$\begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

so that

$$4v_1 + 1v_2 = 0$$

and

$$4v_1 + 1v_2 = 0$$

When $v_1 = -1$ and $v_2 = 4$, a solution is found. Thus the first eigenvalue is 6, with [1, 1] as a corresponding eigenvector, while the second eigenvector is 1, with [−1, 4] as a corresponding eigenvector.

Because the matrix was 2×2, the polynomial for eigenvalues was quadratic and there were two equations in two unknowns to solve for eigenvectors. Imagine the joys of a matrix 15×15, a polynomial with terms to the 15th power for the first half of the solution and 15 equations in 15 unknowns for the second half. A little more appreciation for the computer center, please, next time you visit!