TIPLE LINEAR REGRESSION VIEWPOINTS
/OLUME 16, NUMBER 1, SPRING 1988

Multivariate Analysis Versus Multiple Univariate Analyses

Carl J. Huberty
University of Georgia

John D. Morris
Florida Atlantic University

Abstract

The argument for preceding multiple ANOVAs with a MANOVA to control for Type I error is challenged. Several situations are discussed in which multiple ANOVAs might be conducted. Three reasons for considering a multivariate analysis are discussed: to identify outcome variable system constructs, to select variable subsets, and to determine variable relative worth.

Paper presented at the annual meeting of the American Educational Research Association, Washington, April 1987.

Multivariate Analysis Versus
Multiple Univariate Analyses

The analyses discussed in this paper are those used in research situations where analysis of variance techniques are called for. These analyses are used to study the effects of "treatment" variables on outcome variables (in ex post facto well as experimental studies). With a single outcome variabl we speak of univariate analysis of variance (ANOVA); with multiple outcome variables it is multivariate analysis of variance (MANOVA).

With multiple outcome variables, the typical analysis approach used in the group-comparison context, at least in the behavioral sciences, is to either: (1) conduct multiple ANOVAs, or (2) conduct a MANOVA followed by multiple ANOVAs. The thesis of the current author is that the latter approach seldom appropriate, and the former approach is appropriate on in some special situations. The purpose of this paper is to provide a rationale for the stated thesis, and to present an argument for a truly multivariate analysis, when appropriate.

Type I Error Protection

An argument often given for conducting a MANOVA, as a preliminary to multiple ANOVAs, is to "control for Type I err probability" (see, e.g., Leary & Altmaier, 1980). The rationale typically given is that if the MANOVA yields significance, then one has a "license" to carry out the multiple ANOVAs, with the data interpretation being based on the results of the ANOVAs. It may be intuitively appealing to conclude that one would incorrectly reject a null ANOVA hypothesis less frequently if the null MANOVA hypothesis is initially rejected than if the latter were not rejected. This is the notion of a "protected (ANOVA) F test" (Bock, 1975, p. 422), an extension of Fisher's protected t test idea as applied to the study of contrasts in an ANOVA context.

If a researcher has a legitimate reason for testing univariate hypotheses, then he/she might consider either of two testing procedures. One is a simultaneous test procedure (STP) advocated by Bird and Hadzi-Pavlovic (1983) and programmed by O'Grady (1986). For the STP, as applied to the current MANOVA-ANOVAs context, the referent distribution for the ANOVA F values would be based on the MANOVA test statistic used. Bird and Hadzi-Pavlovic (1983, p. 168), however, point out that for the current context, the overall MANOVA test is not really a necessary prerequisite to simultaneous ANOVAs. Ryan (1980) makes the same point for the ANOVA-contrasts context. These two contexts may be combined to a MANOVA-ANOVAs-contrasts context in which it would be reasonable to go directly to the study of univariate group contrasts, if univariate hypotheses are the main concern (see next section.)

A second procedure for testing univariate hypotheses is to employ the usual univariate test statistics with a Bonferroni adjustment to the overall Type I error probability. How "overall" is defined is somewhat arbitrary. It could mean the probability of committing a Type I error across all tests conducted on the given data set. Or, it could mean the Type I error probability associated with an individual outcome variable when univariate questions are being studied. Whatever the choice (which can be a personal one, and one that is numerically nonconventional!), some error-splitting seems very reasonable. Assuming that Type I error probability for each in a set of m tests is constant, the alpha level for a given test may be determined by using either of two approaches. One approach is to use the additive Bonferroni inequality: for m tests, the alpha level for each test is given by the overall alpha level divided by m. A second approach is to use a multiplicative inequality: for m tests, the alpha level for each test is found by taking one minus the mth root of the complement of the overall alpha level. [See Games (1977).] The per-test alphas--constant across the m tests--found using the two approaches are, for most practical purposes, the same. Therefore, the simpler of the two approaches, namely the first one, is recommended when multiple tests are conducted.

In nearly all instances, outcome variables are interrelated. Thus, the ANOVA F tests are not independent; furthermore, contrast tests for individual outcome variables may not be independent. This lack of independence does not, however, present difficulties in determining the per-test alpha level to use. That this is the case may be seen by the following double inequality:

$$\text{overall alpha} < 1 - (1\text{-test alpha})^{m} < m \cdot \text{test alpha}.$$

It turns out that when conducting m tests, each at a constant alpha level, a considerably larger overall alpha level results. For example, 6 tests, each conducted using an alpha level of .05, yield an overall alpha level of .30 using the additive inequality, and about .26 using the multiplicative inequality (the middle of the double inequality above). The above double inequality ignores the extent of the outcome variable intercorrelations. If r is the constant correlation between all pairs of outcome variables, then the overall alpha level is approximately (Bird, 1975, p. 346)

$$1 - r^2(1 - \text{test alpha}) - (1 - r^2)\ (1 - \text{test alpha})^m.$$

Again, for 6 tests, each at an alpha level of .05, and a constant bi-variable correlation of .30, the overall alpha level is about .25.

While adjusting the individual test alphas in conducting multiple tests addresses the Type I error protection problem, a potential related problem emerges. For m tests and a test alpha equal to (1/m)th of the overall alpha, the statistical power of the multiple tests may be a concern if m is "large." One way of obtaining reasonable power values is to use an adequate sample size. Thus, in designing studies that incorporate multiple outcome variables, the sample size-to-variable ratio is an important consideration. The use of a liberal overall alpha is recommended; something like .20, or even higher in some situations. This whole issue becomes much more involved when group contrasts are studied for each outcome variable. Sound planning, good judgment, and reasonableness are clearly called for.

Merely conducting a MANOVA, obtaining significance at som level, and then conducting multiple ANOVAs, each at a conventional significance level, is hardly "controlling for Type I error probability." The notion that one completely controls for Type I error probability by first conducting an overall MANOVA or ANOVA is open to question (Bird & Hadzi-Pavlovic, 1983; Bray & Maxwell, 1982, p. 343; Ryan, 1980 since the alpha value for each follow-up test would be less than or equal to the alpha employed for the overall test only when the overall null hypothesis is true. (See, also, Wilkinson, 1975.) This notion does not have convincing empirical support in at least a MANOVA-ANOVAs context--the Hummel and Sligo (1971) and Hummel and Johnston (1986) studies notwithstanding.

When Multiple Univariate Analyses?

One situation in which multiple univariate analyses might be appropriate is as a means of screening outcome variables prior to a MANOVA. It behooves the researcher to screen out non-functional variables at the outset for various reasons; to enhance parsimony, to enhance estimated predictive accuracy, t abate collinearity, and so forth. Suppose a researcher has 15 sets of unimodally distributed outcome measures. A reasonable first analysis step would be to conduct 15 ANOVAs. A rule-of-thumb that seems appropriate is to delete any variable m further analysis if the associated ANOVA F-value is less n 1.00. In a two-factor design this rule would pertain to "all-effects" test--the test of the equality of all design l population means. A rationale for this rule is that such F-value implies that the variable is contributing nothing "noise" to the analysis. [An F-value of unity is ivalent to an eta-squared value of $df_h / (df_h + df_e)$.]

A second situation that would call for the use of multiple variate analyses is when the outcome variables are nceptually independent" (Biskin, 1980). [This is the ithesis of a situation involving a variable system, a notion cussed in the next section.] In such a situation one would interested in how a treatment variable affects each of the come variables. Here, there would be no interest in seeking linear combination of the outcome variables; an underlying nstruct" is of no concern. In particular, an underlying struct would perhaps be of little interest when each outcome iable is from a different domain. Dossey (1976), for mple, studied the effects of three treatment variables aching Strategy, Exemplification, Student Ability) on four come variables: Algebra Disjunctive Concept Attainment, metric Disjunctive Concept Attainment, Exclusive Disjunctive cept Attainment, and Inclusive Disjunctive Concept ainment. Considering these outcome variables as ceptually independent, four three-way ANOVAs were conducted.

The third situation in which multiple univariate analyses ht be appropriate is when the research being conducted is exploratory in nature. Such situations would exist when "new" treatment and outcome variables are being studied, and the effects of the former on the latter are being investigated so as to reach some tentative, nonconfirmatory conclusions. This approach might be of greater interest in status studies, as opposed to true experimental studies.

In the two latter situations it might be argued (via the "protected-test" argument) that the multiple tests on the individual outcome variables should be preceded by a MANOVA. As mentioned above, however, this is not necessary. If tests on individual outcome variables are the tests of basic interest, then going directly to the univariate analyses would seem reasonable. One can employ a simultaneous test procedure by referring to a MANOVA test statistic (with or without a Bonferroni adjustment), or multiple univariate analyses by referring to a univariate test statistic with a Bonferroni adjustment.

A fourth situation in which multiple univariate analyses may be appropriate is when some or all of the outcome variables under current study have been previously studied in univariate contexts. In this case separate univariate analysis results can be obtained for comparison purposes, in addition to a multivariate analysis if the latter is appropriate and desirable.

A fifth situation calling for multiple univariate analyses is where a researcher characteristic is considered. The researcher characteristic is a lack of understanding of, and/or appreciation for, multivariate methods. A lack of training and experience in multivariate methods may very well account for the lack of understanding/appreciation. Attempting to use analysis procedures with inadequate understanding is futile indeed. One possible solution to the lack-of-understanding problem (for non-dissertation research) is to contact a knowledgeable methodologist, stimulate his/her interest in the topic being investigated, offer him/her co-authorship, and complete the collaboration.

Finally, there is an evaluation design situation in which multiple univariate analyses might be conducted. This is when some evidence is needed to show that two or more groups of units are "equivalent" with respect to a number of descriptors. These analyses might be considered in an in situ design for the purpose of a comparative evaluation of a project. In this situation evidence of comparability may be obtained via multiple informal ("eye-ball") tests, or formal statistical tests.

Some six situations are presented that would seem appropriate for multiple univariate analyses. Multiple univariate analyses might be conducted: (1) to screen outcome variables prior to a multivariate analysis; (2) to study the effects of some treatment variable(s) on conceptually independent outcome variables; (3) to explore new treatment-outcome variable bivariate relationships; (4) to re-examine bivariate relationships within a multivariate context; (5) when a researcher is multivariately naive; and (6) to select a "comparison" group in designing a study.

Why a Multivariate Analysis?

Of course, the analysis strategy employed by a researcher is dependent, among other things, upon the questions he/she ha of the data on hand. And these questions are, or at least should be, derived from beliefs or theories of the researcher. With questions in mind, it is assumed that the researcher has judiciously chosen a collection of outcome variables that are relevant to his/her investigation. The interrelationship of these variables is an important consideration in deciding upon an analysis strategy. More specifically, does the collection of variables constitute, in some substantive sense, a system? Or, perhaps, are there subcollections that may constitute multiple systems?

A "system" of outcome variables may be loosely defined as a collection of interrelated variables that, at least potentially, determines one or more meaningful underlying variates or constructs. In a system one has several outcome variables which represent a small number of constructs--typically one or two. For example, Watterson et al. (1980) studied a system of five outcome measures on attitudes (based on interview and questionnaire data) that lead to two meaningful variates, political attitude and freedom of expression; Hackman and Taber (1979) studied a system of 21 outcome measures on student performance (based on interview data) that determined two meaningful variates, academic performance and personal growth.

A goal of a multivariate analysis is to identify and interpret the underlying construct(s). For such potential constructs to be meaningful, the judicious choice of outcome variables to study is necessary; the conceptual relationships among the variables must be considered in light of some overriding "theory." A multivariate analysis should enable the researcher to "get a handle" on some characteristics of his/her theory: What are the "emerging variables"?

These emerging variables are identified by considering some linear composites of the outcome variables, called canonical variates or <u>linear discriminant functions</u> (LDFs). Correlations--sometimes called structure correlations--between each outcome variable and each LDF are found. Just as in factor analysis, the absolute values of these correlations, or "loadings," are used in the identification process: those variables with high loadings are "tied together" to arrive at a label for each construct.[2] [See, however, Harris (1985, p. 319), for an opposing point of view regarding such a use of loadings.]

Sometimes a researcher is interested in studying multiple systems, or subsystems, of variables. Those subsystems may be studied for comparative purposos (see, e.g., Lunneborg & Lunneborg, 1977), or simply because different (conceptually independent?) constructs--based on different variable domains--are present (see, e.g., Elkins & Sultmann, 1981). In this case, a separate multivariate analysis for each subsystem would be conducted.

A primary reason, then, for conducting a multivariate analysis is to identify the variates or constructs that underlie the collection of outcome variables chosen for analysis. By doing so, one analyzes the collection as a *system,* taking into consideration the intercorrelations of the variables. This approach enables a researcher to seek answers to more general (more interesting?), complex questions; questions that reflect the real world of behavioral (or any other) science.[3] [See Dempster, (1971) for more on data structure.]

There are two *other* potential reasons for conducting a multivariate analysis. Either of these reasons is considered when the intercorrelations of the outcome variables are to be kept in mind. One potential reason is to determine if fewer variables than the total number initially chosen can adequately define a meaningful system. This is the so-called variable selection problem, and is discussed in some detail by Huberty (1986). This problem might be considered so as to seek a parsimonious interpretation of a system. It should be noted that this is not an imposed parsimony--as one might get with multiple univariate analyses--but a parsimony taking into consideration the intercorrelations of the outcome variables.

Another potential reason for conducting a multivariate analysis is to make an assessment of the relative contribution of the outcome variables to the resultant group differences, or to the resultant effects of the "treatment" variable(s). This is the so-called variable ordering problem. Although the assessment of variable importance is very difficult in all multi-variable analyses (including canonical correlation, factor analysis, cluster analysis), some reasonable indexes have been proposed for the MANOVA context (see Huberty, 1984). Of course, a meaningful ordering of variables that constitute a system can only be legitimately accomplished by taking the variable intercorrelations into consideration.

In a multiple-group situation, the study of system structure and of variable importance may lead to some interesting and informative conclusions. In the univariate case, group contrasts (pairwise or complex) are often of interest in addition to, or in lieu of, the omnibus inter-group comparison. Group contrasts may also be studied with multiple outcome variables--here we have multivariate group contrasts. The construct associated with one contrast may be characterized quite differently from that for another contrast. Also, the variable orderings for effects defined by two contrasts may be quite different. For a detailed discussion of this analysis strategy, see Huberty and Smith (1982).

None of the above three data analysis problems (system structure, variable selection, variable ordering) can be appropriately approached via multiple univariate analyses. As Gnanadesikan and Kettenring (1984, p. 323) put it, an objective of a multivariate analysis is to increase the "sensitivity of the analysis through the exploitation of the inter-correlations among the response variables so that indications that may not be noticeable in separate univariate analyses stand out more clearly in the multivariate analysis."

It should be pointed out that typically employed criteri for variate selection and variable ordering are sample- and system-specific. What is a good variable subset or a relatively good individual variable depends upon the collecti of the variables in the system being studied. How well the proposed selection and ordering criteria "hold up" over repeated sampling needs further empirical study. Of course, replication is highly desirable. The rank-order position of variable in a system of variables may change when new variabl are added to the system. Similarly for the composition of a good subset of variables. Hence, a conclusion regarding the goodness of a variable subset and/or the relative goodness of individual variables must be made with some caution (see Huberty, 1986, for elaboration).

Additional Comments

Some apparently "funny" results can occur when comparing multivariate analysis with multiple univariate analyses. Significant univariate results do not necessarily imply significant multivariate results (see, e.g., Cramer, 1975), a vice versa (see, e.g., Tatsuoka, 1971, pp. 13-24). Of course the meaning of "significant" in the two approaches may be different. Does rejecting a MANOVA null hypothesis lead to t same conclusion as rejecting one or more ANOVA null hypothese How does one compare a single P-value from MANOVA with the multiple P-values from the ANOVAs? Furthermore, how does one compare the power of a multivariate test with the power of a set of univariate tests? These types of comparisons are problematic, particularly because of "inconsistent" MANOVA - ANOVA results that may occur.

Ignoring the interrelatedness of a collection of outcome variables can lead to obtaining redundant information. For example, suppose Variable 1 yields univariate significance, and that Variable 2 is highly correlated with Variable 1. Significance yielded by Variable 2, then, would not be a new result. Van de Geer (1971, p. 271) points out that, "with separate analyses of variance for each variable, we never know how much the results are duplicating each other."

In summary, if a collection of outcome variables constitutes a potentially meaningfull _system,_ then a multivariate analysis called for. That is, a multivariate analysis should be conducted if interest is on potential underlying constructs. If not, then a multiple univariate analysis route would be taken (_without_ a preliminary multivariate analysis). If control over Type I error is of concern when conducting multiple univariate analyses, it is suggested that Bonferroni-adjusted probability values be considered.

References

Bird, K. D. (1975). Simultaneous contrast testing procedures for multivariate experiments. Multivariate Behavioral Research, 10, 343-351.

Bird, K. D., & Hadzi-Pavlovic, D. (1983). Simultaneous test procedures and the choice of a test statistic in MANOVA. Psychological Bulletin, 93, 167-178.

Biskin, B. H. (1980). Multivariate analysis in experimental counseling research. The Counseling Psychologist, 8, 69-72.

Bock, R. D. (1975). Multivariate statistical methods in behavioral research. New York: McGraw-Hill.

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

Cramer, E. M. (1975). The relation between Rao's paradox in discriminant analysis and regression analysis. Multivariate Behavioral Research, 10, 99-107.

Dempster, A. P. (1971). An overview of multivariate data analysis. Journal of Multivariate Analysis, 12, 316-346

Dossey, J. A. (1976). The relative effectiveness of four strategies for teaching algebric and geometric disjuncti concepts and for teaching inclusive and exclusive disjunctive concepts. Journal for Research in Mathemati Education, 7, 92-105.

Elkins, J., & Sultmann, W. F. (1981). ITPA and learning disability: A discriminant analysis. Journal of Learni Disabilities, 14, 88-92.

Gamos, P. A. (1977). An improved table for simultaneous control on g contrasts. Journal of the American Statistical Association, 72, 531-534.

Gnanadesikan, R., & Kettenring, J. R. (1984). A pragmatic review of multivariate methods in applications. In H. A. David & H. T. David (Eds.), Statistics: An appraisal (pp. 309-337). Ames, IA: Iowa State Universi Press.

Hackman, J. D., & Taber, T. D. (1979). Patterns of undergraduate performance related to success in college. American Educational Research Journal, 16, 117-138.

Harris, R. J. (1985). A primer of multivariate statistics. New York: Academic Press.

Huberty, C. J (1972). Regression analysis and 2-group discriminant analysis. Journal of Experimental Education, 41, 39-41.

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

Huberty, C. J (1986, April). Problems with stepwise methods--Better alternatives. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Huberty, C. J, & Smith, J. D. (1982). The study of effects in MANOVA. Multivariate Behavioral Research, 17, 417-432.

Hummel, T. J., & Johnston, C. B. (1986, April). An empirical comparison of size and power of seven methods for analyzing multivariate data in the two-sample case. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hummel, T. J., & Sligo, J. R. (1971). Empirical comparison of univariate and multivariate analysis of variance procedures. Psychological Bulletin, 76, 49-57.

Leary, M. R., & Altmaier, E. M. (1980). Type I error in counseling research: A plea for multivariate analyses. Journal of Counseling Psychology, 27, 611-615.

Lunneborg, C. E., & Lunneborg, P. W. (1977). Is there room for a third dimension in vocational interest differentiation? Journal of Vocational Behavior, 11, 120-127.

O'Grady, K. E. (1986). Simultaneous tests and confidence intervals. Behavior Research Methods, Instruments, & Computers, 18, 325-326.

Ryan, T. A. (1980). Comment on "Protecting the overall rate of Type I errors for pairwise comparisons with an omnibus test statistic." Psychological Bulletin, 88, 354-355.

Tatsuoka, M. M. (1971). Significance tests. Champaign, IL: Institute for Personality and Ability Testing.

Van de Geer, J. P. (1971). Introduction to Multivariate Analysis for the Social Sciences. San Francisco: Freeman.

Watterson, O. M., Joe, G. W., Cole, S. G., & Sells, S. B. (1980). Impression management and attitudes toward marihuana use. Multivariate Behavioral Research, 15, 139-156.

Wilkinson, L. (1975). Response variable hypotheses in the multivariate analysis of variance. Psychological Bulletin, 82, 408-412.

Footnotes

[1]In attempting to encourage graduate students to formally study multivariate methods, the current author has often been confronted with a response such as: "A researcher, keeping in mind some 'theory' underlying a research effort, poses his/her questions first, then seeks analyses to answer the questions. If multivariate analyses are imminent, then he/she can approach a 'statistician' for help." My argument, which only seldom is heeded, is that knowledge of multivariate techniques should enable the researcher to pose more interesting, relevant, and penetrating questions to begin with.

[2]It has been pointed out by Harris (1985, pp. 129, 257, 319) and proven by Huberty (1972) that in the two-group case, the squared LDF-variable correlations are proportional to the univariate F values. Thus, it might seem that if a system structure is to be identified via loadings, then multiple univariate analyses would suffice. In the multiple-group case where at least two LDFs result, however, the multiple constructs cannot be identified by multiple univariate analyses.

[3]The notion of a "construct" may be viewed as a varying one across different types of multivariate analyses. For the group-comparison or grouping-variable-effects situation on which we focus herein, the identified constructs are <u>extrinsic</u> to the set of outcome variables. That is, the optimization of the composites (i.e., LDFs) is based on something external to the outcome variables, namely, the maximization of effects.

Similarly for the optimization of composites (linear classification functions) in the context of predictive discriminant analysis (see Huberty, 1984) where classifica accuracy is maximized. On the other hand, in component analysis, for example, the identified constructs are <u>intri</u> to the set of outcome variables. That is, the optimizatio the composites (i.e., components) is based on something internal to the outcome variables, namely, the maximizatio accounted-for variance in the variable set. Furthermore, extrinsic-intrinsic, constructs-of-constructs situation co result when one conducts a MANOVA (or classification analy using component or factor scores as input.