DOCUMENT RESUME

ED 358 134 TM 019 917

AUTHOR	Thompson, Bruce
TITLE	The General Linear Model (As Opposed to the Classical Ordinary Sums of Squares) Approach to Analysis of Variance Should Be Taught in Introductory Statistical Methods Classes.
PUB DATE	3 Apr 93
NOTE	30p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Analysis of Variance; Correlation; Effect Size; Higher Education; Introductory Courses; *Least Squares Statistics; Mathematical Models; *Mathematics Instruction; *Methods Courses; *Research Methodology; *Statistics
IDENTIFIERS	*General Linear Model; Parametric Analysis; Statistical Package for the Social Sciences; Sum of Squares

ABSTRACT

It is argued that analysis of variance (ANOVA) and related methods should be taught using a general linear model (GLM) approach, rather than a classical ordinary sums of squares approach. The GLM approach emphasizes the linkages among conventional parametric methods, emphasizing that all classical parametric methods are least squares procedures that implicitly or explicitly use weights, focus on latent synthetic variables, and yield effect sizes analogous to "r" squared (are correlational). The case for teaching statistics using a GLM conceptual framework is based on the following four contentions: (1) a GLM instructional approach provides a unifying conceptual framework that better enables students to understand analytic methods; (2) a GLM approach provides a better match between the researcher's analytic model and the researcher's model of reality; (3) the GLM emphasis on planned contrasts helps students understand the critical role of reflective thought in good research; and (4) a GLM instructional approach helps students see that focusing on variance-accounted-for (or other) effect sizes, rather than statistical significance, is important in all analyses. Three tables present analysis examples. A 68-item list of references is included, and an appendix lists Statistical Package for the Social Sciences control cards. (Author/SLD)

aerapro1.wp1 4/3/93

THE GENERAL LINEAR MODEL (AS OPPOSED TO THE CLASSICAL ORDINARY SUMS OF SQUARES) APPROACH TO ANALYSIS OF VARIANCE SHOULD BE TAUGHT IN INTRODUCTORY STATISTICAL METHODS CLASSES

Bruce Thompson

Texas A&M University 77843-4225
and
Baylor College of Medicine

Paper presented at the annual meeting (session #37.04) of the American Educational Research Association, Atlanta, April 15, 1993.

ABSTRACT

The purpose of the present paper is to argue that ANOVA and related methods should be taught using a general linear model (GLM) approach, as against a classical ordinary sums of squares approach. The GLM approach (Cohen, 1968; Knapp, 1978) can be defined as one that emphasizes the linkages among conventional parametric methods (e.g., t-tests, ANOVA, ANCOVA, R). The GLM approach emphasizes that all classical parametric methods are least squares procedures that implicitly or explicitly (a) use weights, (b) focus on latent synthetic variables, and (c) yield effect sizes analogous to r^2, i.e., all classical *analytic* methods are *correlational* (Fan, 1992; Knapp, 1978; Thompson, 1988a).

The case for teaching statistics using a GLM conceptual framework is based on four contentions. First, a GLM instructional approach provides a unifying conceptual framework that better enables students to understand analytic methods in the context of how the methods are really alike and how they really differ. Second, the GLM approach to analysis tends to produce a better match between the researcher's analytic model and the researcher's model of reality, thus yielding more valid results, and so this approach should be emphasized in instruction. Third, the GLM emphasis on planned contrasts, as against omnibus tests followed by unplanned comparisons, helps students understand the critical role of reflective thought as the critical ingredient in good research. Fourth, a GLM instructional approach helps students see that focusing on variance-accounted-for (or other) effect sizes, as against statistical significance, is important in all analyses.

As defined by Gage (1963, p. 95), "Paradigms are models, patterns, or schemata. Paradigms are not the theories; they are rather ways of thinking or patterns for research." Tuthill and Ashton (1983, p. 7) note that

> A scientific paradigm can be thought of as a socially shared cognitive schema. Just as our cognitive schema provide us, as individuals, with a way of making sense of the world around us, a scientific paradigm provides a group of scientists with a way of collectively making sense of their scientific world.

But scientists usually do not consciously recognize the influence of their paradigms. As Lincoln and Guba (1985, pp. 19-20) note:

> If it is difficult for a fish to understand water because it has spent all its life in it, so it is difficult for scientists... to understand what their basic axioms or assumptions might be and what impact those axioms and assumptions have upon everyday thinking and lifestyle.

Even though researchers are usually unaware of paradigm influences, paradigms are nevertheless potent influences in that they tell us what we need to think about, and also the things <u>about which we need not think</u>. As Patton (1975, p. 9) suggests,

> Paradigms are normative, they tell the practitioner what to do without the necessity of long existential

> or epistemological consideration. But it is this aspect of a paradigm that constitutes both its strength and its weaknesses--its strength in that it makes action possible; its weakness in that the very reason for action is hidden in the unquestioned assumptions of the paradigm.

With respect to ANOVA, Kerlinger (1986, p. 203) has noted that, "The analysis of variance is not just a statistical method. It is an approach and a way of thinking." Viewed as a paradigm, CVA methods are also a way of not thinking.

The purpose of the present paper is to argue that OVA methods should be taught using a general linear model (GLM) approach, as against a classical ordinary sums of squares approach. The GLM approach (Cohen, 1968; Knapp, 1978) can be defined as one that emphasizes the linkages among conventional parametric methods (e.g., *t*-tests, ANOVA, ANCOVA, *R*). The GLM approach emphasizes that all classical parametric methods are least squares procedures that implicitly or explicitly (a) use weights, (b) focus on latent synthetic variables, and (c) yield effect sizes analogous to r^2, i.e., all classical *analytic* methods are *correlational* (Fan, 1992; Knapp 1978; Thompson, 1988a).

The case for teaching statistics using a GLM conceptual framework is based on four contentions:

1. A GLM instructional approach provides a unifying conceptual framework that better enables students to understand analytic methods in the context of how the methods are really alike and

how they really differ.

2. The GLM approach to analysis tends to produce a better match between the researcher's analytic model and the researcher's model of reality, thus yielding more valid results, and so this approach should be emphasized in instruction.
3. The GLM emphasis on planned contrasts, as against omnibus tests followed by unplanned comparisons, helps students understand the critical role of reflective thought as the critical ingredient in good research.
4. A GLM instructional approach helps students see that focusing on variance-accounted-for (or other) effect sizes, as against statistical significance, is important in all analyses.

Of course, in arguing against a classical sums of squares approach to instruction, it is important not to make an "is/ought" or "should/would" error. Arguing that something "ought" to be done in the future simply because something else that is incorrect "is" being done now, and perhaps otherwise "would" not be corrected, is logically inconsistent. As Strike (1979, p. 13) explains,

> To deduce a proposition with an "ought" in it from premises containing only "is" assertions is to get something in the conclusion not contained in the premises, something impossible in a valid deductive argument.

Hudson (1969) offers a book on the "is/ought" fallacy.

Contentions

Contention #1: <u>A GLM instructional approach provides a unifying</u>

conceptual framework that better enables students to understand analytic methods in the context of how the methods are really alike and how they really differ.

In a seminal article, Cohen (1968, p. 426) noted that ANOVA and ANCOVA are special cases of multiple regression analysis, and argued that in this realization "lie possibilities for more relevant and, therefore more powerful exploitation of research data." Since that time researchers have increasingly recognized that conventional multiple regression analysis of data as they were initially collected (no conversion of intervally scaled independent variables into dichotomies or trichotomies) does not discard information or distort reality, and that the general linear model

> ...can be used equally well in experimental or non-experimental research. It can handle continuous and categorical variables. It can handle two, three, four, or more independent variables... Finally, as we will abundantly show, multiple regression analysis can do anything the analysis of variance does--sums of squares, mean squares, *F* ratios--and more. (Kerlinger & Pedhazur, 1973, p. 3)

However, canonical correlation analysis, and not regression analysis, is the most general case of the classical parametric general linear model (Baggaley, 1981, p. 129; Fornell, 1978, p. 168). In an important article, Knapp (1978, p. 410) demonstrated this in some mathematical detail and concluded that "virtually all of the commonly encountered tests of significance can be treated as special cases of canonical correlation analysis." Thompson (1988a, 1991a) and Fan (1992) illustrate how canonical correlation analysis can be employed to implement all the parametric tests that canonical methods subsume as special cases.

The GLM approach inherently emphasizes on the truth that <u>all parametric analyses, including OVA analyses, are correlational</u>. Traditionally, researchers have frequently employed OVA methods when they had data from experimental designs. That was fine. However, many researchers then unconsciously associated the analytic method with the characteristics inuring from the design, and not from the analysis. It is experimental design that allows causal inferences, not the analytic method that is used with data from this design. This was not fine. Because the confusion was often unconscious, the illogic was even more powerful as an influence on analytic preferences.

Humphreys (1978, p. 873) notes that many researchers are prone to unconsciously and <u>erroneously</u> associate ANOVA with the power of experimental designs:

> The basic fact is that a measure of individual differences is not an independent variable, and it does not become one by categorizing the scores and treating the categories as if they defined a variable under experimental control in a factorially designed analysis of variance.

Similarly, Humphreys and Fleishman (1974, p. 468) note that categorizing variables in a nonexperimental design using an ANOVA analysis "not infrequently produces in both the investigator and his audience the illusion that he has experimental control over the independent variable. Nothing could be more wrong."

As Keppel and Zedeck (1989) argue,

> We maintain that whereas it is useful to speak in terms of experimental and correlational *designs*, it is unnecessary to maintain the distinction between experimental and correlational *statistics* (ANOVA and MRC, respectively), since the results are statistically identical. This latter point will be repeatedly stated and demonstrated throughout this book. (p. 4, emphasis in original)

Traditionally, confusion that OVA is an experimental analysis led to very frequent use of OVA methods within the social sciences (e.g., Edgington, 1974). Fortunately, the emergence of GLM instructional approaches has helped some researchers see that all analytic methods are correlational, and has led to less frequent use of OVA methods (Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Willson, 1982). Thus, the GLM approach provides a *unifying conceptual framework* within which students can compare and contrast analytic choices, and can better understand what they're really doing when they implement a given choice.

Contention #2: <u>The GLM approach to analysis tends to produce a better match between the researcher's analytic model and the researcher's model of reality, thus</u>

yielding more valid results, and so this approach should be emphasized in instruction.

As Thompson (1991b) noted,

> too few researchers recognize that in all analyses we inherently invoke both a presumptive model of reality and an analytic model. When the two don't match, the analysis doesn't help us understand the reality we believe exists. (p. 1072)

The OVA analytic model requires that all predictors be nominally scaled, even if they must be converted from original interval scale. Most researchers use balanced OVA designs so that their effects will be uncorrelated, and so that their analyses will be more robust to the violation of the homogeneity of variance assumption. Thus, the analytic model presumes that all predictors are nominally scaled, that all predictor main and interaction effects are uncorrelated, and that the distributions of scores on the nominal predictor variables are flat or rectangular.

Unfortunately, most researchers' models of reality do not match this analytic model very well. The most common nominally scaled predictor effects are experimental condition and gender; while gender is that very rare variable on which virtually everyone both knows their status and will actually even honestly report, gender may not be a particularly useful independent variable in most studies.

Even most experimental studies invoke intervally scaled "aptitude" variables (e.g., IQ scores in a study with academic achievement as a dependent variable), to conduct the aptitude-treatment interaction (ATI) analyses recommended so persuasively by Cronbach (1957, 1975) in his 1957 APA Presidential address. But as Cliff (1987, p. 130) notes, the practice of discarding variance on intervally scaled predictor variables to perform OVA analyses creates problems in almost all cases:

> Such divisions are not infallible; think of the persons near the borders. Some who should be highs are actually classified as lows, and vice versa. In addition, the "barely highs" are classified the same as the "very highs," even though they are different. Therefore, reducing a reliable variable to a dichotomy makes the variable more unreliable, not less.

Discarding variance is not generally good research practice (Thompson, 1988c, 1988d). As Kerlinger (1986, p. 558) explains,

> ...partitioning a continuous variable into a dichotomy or trichotomy throws information away... To reduce a set of values with a relatively wide range to a dichotomy is to reduce its variance and thus its possible correlation with other variables. A good rule of research data analysis, therefore, is: <u>Do not reduce continuous variables to partitioned variables</u> (dichotomies, trichotomies, etc.) unless compelled to do so by circumstances or the nature of the data (seriously skewed, bimodal,

etc.).

Kerlinger (1986, p. 558) notes that variance is the "stuff" on which all analysis is based. Discarding variance by categorizing intervally-scaled variables amounts to "squandering of information" (Cohen, 1968, p. 441). As Pedhazur (1982, pp. 452-453) notes,

> Categorization of attribute variables is all too frequently resorted to in the social sciences... It is possible that some of the conflicting evidence in the research literature of a given area may be attributed to the practice of categorization of continuous variables... Categorization leads to a loss of information, and consequently to a less sensitive analysis.

Contention #3: The GLM emphasis on planned contrasts, as against omnibus tests followed by unplanned comparisons, helps students understand the critical role of reflective thought as the critical ingredient in good research.

There are two reasons why researchers generally prefer the use of planned comparisons to the use of unplanned comparisons (cf. Benton, 1991; Tucker, 1991). First, as noted by numerous researchers, planned comparisons offer more power against making Type II errors:

> procedures recommended for *a priori* orthogonal comparisons are more powerful than procedures recommended for *a priori* nonorthogonal and *a*

posteriori comparisons. That is, the former procedures are more likely to detect real differences among means. (Kirk, 1968, p. 95)

The probability of test's detecting that... [the contrast's effect] is not zero [i.e., is statistically significant] is greater with a planned than with an unplanned comparison on the same sample means. Thus, for any particular comparison, the test is more powerful when planned than when post hoc. (Hays, 1981, p. 438)

Post hoc tests protect us from making too many Type I errors by requiring a bigger difference before declaring it to be significant than do planned comparisons. But this protection tends to be too conservative for planned comparisons, thereby lowering the power of the test. (Minium & Clarke, 1982, p. 322)

The tests of significance for a priori, or planned, comparisons are more powerful than those for post hoc comparisons. In other words, it is possible for a specific comparison to be not significant when tested by post hoc methods but significant when tested by a priori methods. (Pedhazur, 1982, pp.

304-305 (also Kerlinger & Pedhazur, 1973, p. 131))

> Post hoc comparisons must always follow the finding of a significant overall *F*-value... There are no limits to the number of combinations that can be tested post hoc, but none of these procedures has the power of planned comparison tests for detecting statistical significance. (Sowell & Casey, 1982, p. 119)

> The test of planned subhypotheses is more powerful than the test of post hoc subhypotheses. For this reason, we should make planned comparisons whenever possible in planning the design of research within the ANOVA context. (Glasnapp & Poggio, 1985, p. 474)

Second, and perhaps even more importantly, <u>planned comparisons tend to force the researcher to be more thoughtful in conducting research</u>, since the number of planned comparisons that can be tested is limited. As Snodgrass, Levy-Berger and Haydon (1985, p. 386) suggest, "The experimenter who carries out post hoc comparisons often has a rather diffuse hypothesis about what the effects of the manipulation should be." Keppel (1982, p. 165) notes that, "Planned comparisons are usually the motivating force behind an experiment. These comparisons are targeted from the start of the investigation and represent an interest in particular combinations of conditions--not in the overall experiment." In summary, as

Kerlinger (1986, p. 219) suggests, "While post hoc tests are important in actual research, especially for exploring one's data and for getting leads for future research, the method of planned comparisons is perhaps more important scientifically."

It is important to note that most researchers have fairly good notions of what their studies will show, at least when research is grounded in theoretical constructs or in previous empirical findings, so most researchers are able to suggest planned comparisons prior to data collection. Thus, Huberty and Morris (1988, p. 576) maintain that

> only very few research situations would preclude a researcher from specifying all contrasts of interest prior to an examination of the outcome measures and/or the outcome 'cell' means.

A Concrete Heuristic Example of Power

Just as some researchers benefit from seeing heuristic demonstrations that all parametric significance testing procedures are subsumed by and can be conducted with canonical correlation analysis (Thompson, 1988a, 1991a), it may be helpful to present a hypothetical analysis demonstrating that planned orthogonal comparisons have greater statistical power against Type II error than testing omnibus hypotheses and then exploring statistically significant effects with unplanned comparisons. The data presented in Table 1 can be utilized for this purpose. Table 2 presents a conventional one-way ANOVA keyout associated with the Table 1 data. Even if the researcher conducted unplanned post hoc tests in the absence of a statistically significant main effect, none of the unplanned tests would result in a statistically significant comparison for these data. However, as noted in Table 3, a statistically significant ($p < 0.01$) result is isolated for the planned contrast hypothesis that the mean attitude-toward-school score of the two school board members differs from the mean for the remaining 10 subjects.

INSERT TABLES 1 THROUGH 3 ABOUT HERE.

The Use of Planned Comparisons in Lieu of Omnibus Tests

Some researchers suggest that at least some unplanned comparisons can be made even if an omnibus effect is not statistically significant. For example, Spence, Cotton, Underwood and Duncan (1983, p. 215) suggest that,

> The Tukey *hsd* [honestly significant difference test] usually is performed only if the *F* obtained in the analysis of variance is significant, but it theoretically permissible to perform whatever the significance of *F*.

Similarly, Hays (1981, p. 434) notes:

> This statement is not to be interpreted to mean that post hoc comparisons are somehow illegal or immoral if the original *F* test is not significant at the required alpha level... What one cannot do is to attach an unequivocal probability statement to such post hoc comparisons, unless the conditions

underlying the method have been met.

However, the preponderant view regarding use of unplanned post hoc tests is expressed by Gravetter and Wallnau (1985, p. 423):

> These [a posteriori] tests attempt to control the overall alpha level by making the adjustments for the number of different samples (potential comparisons) in the experiment. To justify a posteriori tests, the F-ratio from the overall ANOVA must be significant.

On the other hand, with respect to the use of planned comparisons, "Most statisticians agree that planned t tests between means are appropriate, even when the overall F is insignificant" (Clayton, 1984, p. 193). Snodgrass, Levy-Berger and Haydon (1985, p. 386) concur:

> For planned comparisons, it is not necessary for the overall ANOVA to be significant in order to carry them out... Post hoc comparisons, on the other hand, may not be carried out unless the overall ANOVA is significant.

Gravetter and Wallnau (1985, p. 423) agree that, "Planned comparisons can be made even when the overall F-ratio is not significant."

In fact, "It is not necessary to perform an over-all test of significance prior to carrying out planned orthogonal t tests" (Kirk, 1968, p. 73, emphasis added). As Hays (1981, p. 426) suggests,

> The $\underline{F}$ test gives evidence to let us judge if all of a set of $\underline{J}$ - 1 such orthogonal comparisons are simultaneously zero in the populations. For this reason, if planned orthogonal comparisons are tested separately, the overall $\underline{F}$ test is not carried out, and vice versa.

Swaminathan (1989, p. 231, emphasis added) presents the same argument with respect to the MANOVA case:

> The often advocated procedure of following up the rejection of the null hypothesis with a more powerful multiple comparison procedure should be discouraged. First, the overall rejection of the null hypothesis does not guarantee any meaningful contrast among the means will be significant, as our example showed. Second..., significant contrasts may be found even when the null hypothesis would not have been rejected. Third, follow up multiple comparison procedures which are unrelated to the overall test result in an inflation of the experiment-wise error rate. If multiple comparisons are of primary interest, a suitable multiple comparison procedure can be used <u>without</u> first performing an overall test.

Given that planned tests have greater power against Type II error than either unplanned tests or omnibus tests, planned comparisons should be employed in most research studies using OVA methods. Planned tests should be employed in lieu of omnibus tests. Rosnow and Rosenthal (1989, p. 1281) quite rightly deplore the "overreliance on omnibus tests of diffuse hypotheses that although providing protection for some investigators from the dangers of 'data mining' with multiple tests performed as if each were the only one considered", because omnibus tests generally do not:

> tell us anything we really want to know. As Abelson (1962) pointed out long ago in the case of analysis of variance (ANOVA), the problem is that when the null hypothesis is accepted, it is frequently because of the insensitive omnibus character of the standard F-test as much as by reason of sizable error variance. All the while that a particular predicted pattern among the means is evident to the naked eye the standard F-test is often insufficiently illuminating to reject the null hypothesis that several means are statistically identical.

Planned contrasts (Rosnow & Rosenthal, 1989, p. 1281) encourage precision of thought and theory, and "usually result in increased power and greater clarity of substantive interpretation."

The problem with unplanned contrasts isn't that they make corrections for tests researchers look at and care about, <u>the problem is that these methods make corrections even for tests that researchers don't care about and which they refuse to consult or interpret</u>. As Thompson (1991b) notes, In our example 6 x 4 x 2 design, the Tukey test for the A-way main effect corrects for making all possible 15 ((6 x (6 - 1)) / 2) pairwise comparisons, even if we're only interested in four of them. Perhaps some people are willing to pay the price for looking in someone else's window, but very few of us want to go to jail for looking in a window we actually didn't look in, and in which we have absolutely no interest. We should be equally prudent as regards contrasts. (p. 507)

Contention #4: <u>A GLM instructional approach helps students see that focusing on variance-accounted-for (or other) effect sizes, as against statistical significance, is important in all analyses</u>.

The propensity to overinterpret significance tests continues, notwithstanding several decades of effort "to exorcise the null hypothesis" (Cronbach, 1975, p. 124). Thompson (1989a, p. 66) notes that

> few statistical procedures have caused more confusion within the research community than statistical significance testing... Because statistical significance is largely an artifact of sample size, significance decisions... must be interpreted in the context of sample size.

Rosnow and Rosenthal (1989, p. 1277) comment on contemporary overemphasis on significance tests:

> It may not be an exaggeration to say that for many PhD students, for whom the .05 alpha has acquired an almost ontological mystique, it can mean joy, a doctoral degree, and a tenure-track position at a major university if their dissertation *p* is less than .05.... [But] surely, God loves the .06 nearly as much as the .05 [level].

Thompson (1987b) explores the consequences of these problems. Even sophisticated authors of prominent textbooks are sometimes not quite sure what role significance tests should play in multivariate analysis (Thompson, 1987a, 1988f), though doctoral students may be disproportionately susceptible to excessive awe for significance tests (Eason & Daniel, 1989; Thompson, 1988b). Recent important treatments of these issues are also offered by Huberty (1987) and by Kupfersmid (1988).

Researchers who have had the fortunate experience of working with large samples (cf. Kaiser, 1976) soon realize that virtually all null hypotheses will be rejected, since "the null hypothesis of no difference is almost never exactly true in the population" (Thompson, 1987b, p. 14). As Meehl (1978, p. 822) notes, "As I believe is generally recognized by statisticians today and by thoughtful social scientists, the null hypothesis, taken literally, is always false." Thus Hays (1981, p. 293) argues that "virtually any study can be made to show significant results if one uses enough subjects." Thompson (1992b) summarizes the implication:

> Statistical significance testing can involve a

> tautological logic in which tired researchers, having collected data from hundreds of subjects, then conduct a statistical test to evaluate whether there were a lot of subjects, which the researchers already know, because they collected the data and [already] know they're tired. (p. 436)

These considerations suggest that researchers out to interpret results from their analyses by considering effect sizes as well as significance test results (Huberty, 1987), or by interpreting significance in the context of sample size (i.e., at what smaller sample size would this result have been no longer significant?--Thompson, 1989a), or by conducting analyses that investigate the replicability of results (Thompson, 1989b, in press). Replicability analyses include the cross-validation logics discussed by Thompson (1984, pp. 41-47, 1989b), or variants of bootstrap (Diaconis & Efron, 1983; Efron, 1979; Lunneborg, 1987, Thompson, 1988e, 1992a) or jackknife (e.g., Crask & Perreault, 1977; Daniel, 1989) methods.

The GLM perspective forces researchers to recognize that all conventional parametric analyses capitalize on sampling error, because all these methods are least squares methods that optimize a variance-accounted-for parameter estimate. The less enlightened researcher may express an ill-founded preference for OVA methods because of a misconception that analyses explicitly named after correlation coefficients capitalize on sampling error, while OVA and other analyses purportedly do not. Nothing could be more wrong. Evaluating the generalizability of analytic results is important whenever *any* parametric least-squares methods ($\underline{t}$-tests, OVA, etc.) are used, because "one tends to take advantage of chance [both sampling and measurement errors] in any situation where something is optimized from the data at hand" (Nunnally, 1978, p. 298).

It is ironic that researchers who are blinded by the paradigm influences which create an excessive reliance on significance tests are often hoisted on their own petards. The researcher desirous of statistically significant effects for substantive main and interaction effects will quite reasonably employ the largest sample possible so as to achieve the hoped-for results. Regrettably, large samples that tend to yield significance for substantive tests also tend to yield statistically significant results leading to rejection of assumption null hypotheses, as in the test of equality of dependent variable variances across groups required by the ANOVA homogeneity of variance assumption.

Few, if any, researchers would ever interpret a bivariate $\underline{r}$, a multiple $\underline{R}$, or a canonical correlation ($\underline{R}_C$) study without focusing attention on a variance-accounted-for statistic, such as $\underline{r}^2$, $\underline{R}^2$, or $\underline{R}_C^2$ adjusted for shrinkage (Thompson, 1990). A GLM perspective forces students to acknowledge that analogous statistics (e.g., eta^2, omega^2) are available for all analyses, including OVA analyses (Carter, 1979; Snyder & Lawson, in press).

It is inconsistent to insist that variance-accounted-for statistics must be interpreted for selected analytic methods, and to then ever decline to interpret variance-accounted-for (or some kind of effect size) estimates when using other analytic methods that are co-equal partners (Fan, 1992; Knapp, 1978) in the same family of GLM analyses. The GLM approach to instruction and to thinking can be a powerful paradigm to help researchers overcome the tendency of OVA paradigm to encourage researchers to not think.

References

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Benton, R.L. (1991). Statistical power considerations in ANOVA. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 119-132). Greenwich, CT: JAI Press.

Carter, D.S. (1979). Comparison of different shrinkage formulas in estimating population multiple correlation coefficients. Educational and Psychological Measurement, 39, 261-266.

Clayton, K. N. (1984). An introduction to statistics. Columbus, OH: Merrill.

Cliff, N. (1987). Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Crask, M.R., & Perreault, W.D., Jr. (1977). Validation of discriminant analysis in marketing research. Journal of Marketing Research, 14, 60-68.

Cronbach, L.J. (1957). The two disciplines of psychology. American Psychologist, 12, 671-684.

Cronbach, L.J. (1975). Beyond the two disciplines of psychology. American Psychologist, 30, 116-127.

Daniel, L.G. (1989, January). Use of the jackknife statistic to establish the external validity of discriminant analysis results. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 305 382)

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. Scientific American, 248(5), 116-130.

Eason, S.H., & Daniel, L.G. (1989, January). Trends and methodological practices in several cohorts of dissertations. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston. (ERIC Document Reproduction Service No. ED 306 299)

Edgington, E.S. (1974). A new tabulation of statistical procedures in APA journals. American Psychologist, 29, 25-26.

Elmore, P.B., & Woehlke, P.L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Fan, Xitao. (1992, April). Canonical correlation analysis as a general data-analytic model. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 348 383)

Fornell, C. (1978). Three approaches to canonical analysis. Journal of the Market Research Society, 20, 166-181.

Gage, N.L. (1963). Paradigms for research on teaching. In N.L. Gage (Ed.), Handbook of research on teaching (pp. 94-141). Chicago: Rand McNally.

Glasnapp, D. R., & Poggio, J. P. (1985). Essentials of statistical

analysis for the behavioral sciences. Columbus, OH: Merrill.
Goodwin, L.D., & Goodwin, W.L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.
Gravetter, F. J., & Wallnau, L. B. (1985). Statistics for the behavioral sciences. St. Paul, MN: West.
Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart and Winston.
Huberty, C.J. (1987). On statistical testing. Educational Researcher, 16(8), 4-9.
Huberty, C.J, & Morris, J.D. (1988). A single contrast test procedure. Educational and Psychological Measurement, 48, 567-578.
Hudson, W.D. (1969). The is/ought question. London: MacMillan.
Humphreys, L.G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. Journal of Educational Psychology, 70, 873-876.
Humphreys, L.G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-differences variables. Journal of Educational Psychology, 66, 464-472.
Kaiser, H.F. (1976). Review of Factor analysis as a statistical method. Educational and Psychological Measurement], 36, 586-589.
Keppel, G. (1982). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.
Keppel, G., & Zedeck, S. (1989). Data analysis for research designs. New York: W. H. Freeman.
Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.
Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.
Kirk, R. E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole. (pp. 69-98)
Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.
Kupfersmid, J. (1988). Improving what is published: A model in search of an editor. American Psychologist, 43, 635-642.
Lincoln, Y.S., & Guba, E.G. (1985). Naturalistic inquiry. Beverly Hills: SAGE.
Lunneborg, C.E. (1987). Bootstrap applications for the behavioral sciences. Seattle: University of Washington.
Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.
Minium, E. W., & Clarke, R. B. (1982). Elements of statistical reasoning. New York: John Wiley and Sons.
Nunnally, J.C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.
Patton, M.Q. (1975). Alternative evaluation research paradigm.

Grand Forks: University of North Dakota Press.
Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart and Winston.
Rosnow, R.L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284.
Snodgrass, J. G., Levy-Berger, G., & Haydon, M. (1985). Human experimental psychology. New York: Oxford University Press.
Snyder, P., & Lawson, S. (in press). Evaluating statistical significance using corrected and uncorrected magnitude of effect size estimates. In B. Thompson (Guest Editor) of the special issue on statistical significance testing, Journal of Experimental Education.
Sowell, E. J., & Casey, R. J. (1982). Research methods in education. Belmont, CA: Wadsworth.
Spence, J. T., Cotton, J. W., Underwood, B. J., & Duncan, C. P. (1983). Elementary statistics (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
Strike, K.A. (1979). An epistemology of practical research. Educational Researcher, 8(1), 10-16.
Swaminathan, H. (1989). Interpreting the results of multivariate analysis of variance. In B. Thompson (Ed.), Advances in social science methodology (pp. 205-232, Vol. 1). Greenwich, CT: JAI Press.
Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Beverly Hills: SAGE.
Thompson, B. (1987a). Review of Foundations of behavioral research (3rd ed.) by F. Kerlinger. Educational Research and Measurement, 47, 1175-1181.
Thompson, B. (1987b, April). The use (and misuse) of statistical significance testing: Some recommendations for improved editorial policy and practice. Paper presented at the annual meeting of the American Education Research Association, Washington, DC. (ERIC Document Reproduction Service No. ED 287 868)
Thompson, B. (1988a, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)
Thompson, B. (1988b, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 301 595)
Thompson, B. (1988c). Discarding variance: A cardinal sin in research. Measurement and Evaluation in Counseling and Development, 21, 3-4.
Thompson, B. (1988d). Misuse of chi-square contingency table test statistics. Educational and Psychological Research, 8, 39-49.
Thompson, B. (1988e). Program FACSTRAP: A program that computes

bootstrap estimates of factor structure. Educational and Psychological Measurement, 48, 681-686.

Thompson, B. (1988f). Review of Analyzing multivariate data by N. Cliff. Educational and Psychological Measurement, 48, 1129-1135.

Thompson, B.(1989a). Asking "what if" questions about significance tests. Measurement and Evaluation in Counseling and Development, 22, 66-68.

Thompson, B. (1989b). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. Measurement and Evaluation in Counseling and Development, 22, 2-6.

Thompson, B. (1990). Finding a correction for the sampling error in multivariate measures of relationship: A Monte Carlo study. Educational and Psychological Measurement, 50, 15-31.

Thompson, B. (1991a). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24(2), 80-95.

Thompson, B. (1991b). Review of Data analysis for research designs by G. Keppel & S. Zedeck. Educational and Psychological Measurement, 51, 500-510.

Thompson, B. (1992a). DISCSTRA: A computer program that computes bootstrap resampling estimates of descriptive discriminant analysis function and structure coefficients and group centroids. Educational and Psychological Measurement, 52, 905-911.

Thompson, B. (1992b). Two and one-half decades of leadership in measurement and evaluation. Journal of Counseling and Development, 70, 434-438.

Thompson, B. (in press). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality.

Tucker, M.L. (1991). A compendium of textbook views on planned versus post hoc tests. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 107-118). Greenwich, CT: JAI Press.

Tuthill, D., & Ashton, P. (1983). Improving educational research through the development of educational paradigms. Educational Researcher, 12(10), 6-14.

Willson, V. L. (1982, January). Misuses of regression approaches to ANOVA and ANCOVA. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 222 522)

Table 1
Hypothetical Data for Attitudes Toward School Study (n=12)

Group	LEVEL	ID	DV	Contrast C1	C2	C3	C4	C5
Students	1	1	10	-1	-1	-1	-1	-1
		2	20	-1	-1	-1	-1	-1
Teacher Aides	2	3	10	1	-1	-1	-1	-1
		4	20	1	-1	-1	-1	-1
Teachers	3	5	10	0	2	-1	-1	-1
		6	20	0	2	-1	-1	-1
Principals	4	7	10	0	0	3	-1	-1
		8	20	0	0	3	-1	-1
Superintendents	5	9	10	0	0	0	4	-1
		10	20	0	0	0	4	-1
Board Members	6	11	25	0	0	0	0	5
		12	35	0	0	0	0	5

Table 2
One-Way ANOVA Results

Source	SOS	df	Mean Square	F	p	eta Square
Between	375.0000	5	75.0000	1.5000	.3155	.55556
Error	300.0000	6	50.0000			
Total	675.0000	11				

Table 3
Planned Comparison Results

Contrast Source	SOS	df	Mean Square	F	p	eta Square
C1	.0000	1	.0000	0.0000		.00000
C2	.0000	1	.0000	0.0000		.00000
C3	.0000	1	.0000	0.0000		.00000
C4	.0000	1	.0000	0.0000		.00000
C5	375.0000	1	375.0000	12.5000	.0054	.55556
Error	300.0000	6	50.0000			
Total	675.0000	11				

APPENDIX A
Selected SPSS-X Control Cards

```
TITLE '*****OMNIBUS no  POSTHOC no  A PRIORI yes'
FILE HANDLE BT/NAME='APRIORI.DTA'
DATA LIST FILE=BT/LEV 1 DV 2-4
COMPUTE C1=0
COMPUTE C2=0
COMPUTE C3=0
COMPUTE C4=0
IF (LEV EQ 2)C1=1
IF (LEV EQ 1)C1=-1
IF (LEV EQ 3)C2=2
IF (LEV LT 3)C2=-1
IF (LEV EQ 4)C3=3
IF (LEV LT 4)C3=-1
IF (LEV EQ 5)C4=4
IF (LEV LT 5)C4=-1
IF (LEV EQ 6)C5=5
IF (LEV LT 6)C5=-1
REGRESSION VARIABLES=DV C1 TO C5/DESCRIPTIVES=ALL/
  CRITERIA=PIN(.95) POUT(.999) TOLERANCE(.00001)/DEPENDENT=DV/
  ENTER C5/ENTER C4/ENTER C3/ENTER C2/ENTER C1/
```