Psychological Bulletin
1974, Vol. 81, No. 6, 335–343

SOME PROBLEMS IN THE NONORTHOGONAL ANALYSIS OF VARIANCE [1]

MARK I. APPELBAUM [2] AND ELLIOT M. CRAMER

University of North Carolina at Chapel Hill

Nonorthogonal analysis of variance has been much misunderstood by psychologists, and as a result there has been considerable controversy as to the appropriate methods of analysis. These problems traditionally associated with the nonorthogonal multifactor analysis of variance are rather easily resolved by viewing the analysis of variance (either orthogonal or nonorthogonal) as a series of model comparisons. From this point of view, the analysis of highly confounded designs is seen to yield results that correspond to those that a purely logical analysis would suggest. A logical flow of comparisons and decisions is developed for both the two- and three-factor designs that, although more complicated than procedures previously proposed, seems necessary for drawing proper inferences. It is further shown that there is no logical difference between orthogonal and nonorthogonal analysis of variance.

The nonorthogonal multifactor analysis of variance is perhaps the most misunderstood analytic technique available to the behavioral scientist, save factor analysis. Standard textbooks all but ignore it or bury it in such confused mathematics as to make it barely understandable. Recent articles (e.g., Joe, 1971; Overall & Spiegel, 1969; Rawlings, 1972; Werts & Linn, 1971; Williams, 1972) have attempted to clarify the situation and set guidelines for the analysis of nonorthogonal multifactor experiments; however, these articles have confused the issue with unnecessary proofs, antiquated "approximate" methods, and the implication that nonorthogonal designs are special cases to be avoided at all costs. So strong is the belief that there is something inherently difficult or strange about the nonorthogonal case that experimenters will on occasion go to unusual lengths, such as randomly discarding data from selected cells, in order to achieve an orthogonal design.

This article shows that there is no conceptual difference between orthogonal and nonorthogonal analysis of variance and that, indeed, the orthogonal design is a special case of the more general nonorthogonal design. By approaching the entire issue of the analysis of variance as one of model comparisons, the special problems encountered in the nonorthogonal case are rather easily understood. The closely related problem of deletion of variables in multiple regression analysis has been discussed by Cramer (1972). By treating the problem as one of comparisons of linear models, he has resolved the issue in an obvious manner. A similar approach with nonorthogonal designs leads to the same resolution.

The easy access to computer programs that perform the analysis of variance by a general linear model approach[3] makes possible the computations for this method of dealing with nonorthogonal multifactor designs and eliminates the need for approximate solutions.

A *nonorthogonal design* refers to any experimental design in which the numbers of observations are not equal in each and every cell. This definition even encompasses designs traditionally classified as proportional and in-

[1] This work was supported in part by Grant NE-G-00-3-0111 from the National Institute of Education, U.S. Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred. Support was also received from Grant MH-10007 from the National Institute of Mental Health, Public Health Service.

[2] Requests for reprints should be sent to either author, University of North Carolina, L. L. Thurstone Psychometric Laboratory, Chapel Hill, North Carolina 27514.

[3] For example, E. M. Cramer. *Revised MANOVA Program*. Chapel Hill: University of North Carolina Psychometric Laboratory, 1967.

cludes designs that are not complete factorial. Insofar as an experimental design may be considered a partially complete factorial design (e.g., randomized blocks, Latin squares, nested or hierarchical designs, etc.), the principles discussed in this article apply.

The term *method* is used to refer to the estimation procedure which is assumed to be the method of least squares. The concepts developed and discussed in this article apply only to least squares analyses and should not be applied to nonexact approaches, such as the unweighted-means analysis (Winer, 1971, 445–449), nor to cases that employ some other method of estimation of effects.

An *experimental design* refers to the plan of the experiment determined by the experimenter on the basis of his conception of some idealized state of nature. The minimum requirements for an experimental design are the specification of the experimental factors and the plan for random assignment of experimental units to treatments, including both the sampling plan and the determination of the number of units per treatment. It is the experimental design that implicitly specifies a set of possible models or idealized states whose appropriateness the researcher attempts to assess.

Hypotheses or *tests of hypotheses* are comparisons of various models. Within the analysis of variance, one attempts to assess the appropriateness of one model in comparison to another. To stress this point, significance tests are referred to as *model comparisons*. Unfortunately, the standard approaches to the analysis of variance overlook this consideration and have led to much unnecessary confusion.

Finally, one must consider those situations that may produce a nonorthogonal design. First there is the case where the design is intentionally planned and executed as nonorthogonal. Such designs may be preferred in cases where contrasts of particular cells are desired or where greater precision of estimation is required in some cells than in others. The second case occurs when a design (orthogonal or nonorthogonal) is not executed as planned. That is, once the random assignment of experimental units to treatments has been made, data are not obtained on some units. In this case, one of two different situations may have occurred and different approaches may be required. It may be that the cause of the loss of experimental units is a random phenomenon or one unrelated to the experimental treatments. Death of experimental animals, no-show of college subjects, and the like, often may be viewed as random phenomena and one is, in effect, left with a random sample of a random sample that is itself a random sample. In this case the methods of nonorthogonal analysis of variance to be discussed apply.

The situation that may cause difficulty is when the cause of loss of experimental units cannot be considered a random phenomenon. This loss may either be related to the experimental treatments, as when a combination of treatments cause the death of some subjects, or logically unrelated to the experimental treatments, as when one set of treatment combinations is run late in the afternoon and causes an increase in the no-show rate. In such a case, there would seem to be no remedy short of pretending that the missing observations are random. Perhaps the definitive statement was made by Cochran and Cox (1957) when they observed that the only complete solution to the problem of missing data is not to have any. The following method leads to correct analyses and interpretation of designs that are (*a*) planned as nonorthogonal or (*b*) become nonorthogonal due to the random processes of nature.

Models and the Method of Least Squares

Having decided to employ the method of least squares and having determined the design of the experiment on the basis of a belief as to the nature of the world (in some idealized sense), one is left only with the selection of possible models and model comparisons. The models selected are logically independent of the observed numbers of observations per cell. Although the analysis will be affected by the cell frequencies, the experimenter is free to choose the numbers of observations per cell. The models themselves are not expressed in terms of the number of units in any subpopulation and hence the cell frequencies can hardly matter in terms of the

correctness of the model. But then why all of the concern about nonorthogonal analysis? The answer is that the problem of nonorthogonal analysis really occurs at the level of model comparisons and proper interpretation of the results of such comparisons. The difficulty arises from the methods available to assess the "correctness" of the several models being compared.

In the one-way analysis of variance, it is clear that the least squares estimates of the population means are the sample means regardless of ns. If the total number of observations is held constant, it may easily be shown that the overall test of significance has maximum power when there are equal numbers of observations in cells. The only effect of varying the distribution of the n_i is to vary the power of the test of significance.

The situation is not nearly so simple when one moves to the case of factorial experiments. Consider a two-way analysis of variance (intentionally constructed to represent an extreme case) for which we have observed the cell means $\bar{x}_{ij}$ with cell frequencies n_{ij} as given in Table 1. Assume that the estimated within-cell standard deviation (error) is 15 in each cell (i.e., $MS_{\text{error}} = 225$).

What can one say, given the n_{ij} and $\bar{x}_{ij}$, about the presence of any main effects or interactions in this experiment? Given the answer to this question, further consider what one would suppose to be true of the populations? Relatively sophisticated users of analysis of variance often answer these questions in an inconsistent manner. Many base their answers to the first question on the means alone, judging that there is an A effect but no B effect or interaction. The obvious inequality in the numbers of observations per cell is troubling and the sophisticated respondent observes, in response to the second question, that the main diagonal means are much more stable than the off-diagonal means.

If one asks the further question, "What would and should an analysis of variance tell you about the true populations?", he is at the heart of the problem of nonorthogonal analysis of variance. Any analysis of variance must give information about the population means. Unequal numbers of observations in cells does not alter the character of this information, although it certainly does alter the precision of any statements.

Looking at the sample means of Table 1, it is apparent that if the population means are the same as these sample means (and this is the best guess), there is only an A effect present. These statements must, however, take into account the sampling variability of these sample means. Consider for a moment the 95% confidence intervals (Table 1) that may be generated about the four observed sample means. Because the samples themselves are independent random samples from four possibly different populations, the confidence intervals are in the same sense independent. From these confidence intervals, one may see it is reasonable that this sample could have come from a population with any of the following patterns of means (Table 2). These are but three among many possible sets, but notice that the first would be considered as one in which there was a main effect of A but no B or AB

TABLE 1

CELL MEANS, FREQUENCIES, AND 95% CONFIDENCE INTERVALS FOR TWO-WAY EXAMPLE WITH MARGINAL MEANS COLLAPSED OVER CLASSIFICATIONS

A \ B	1	2	
1	$\bar{X}_{11} = 10$ $N_{11} = 25$ $CI: 3.97 \leftrightarrow 16.03$	$\bar{X}_{12} = 10$ $N_{12} = 2$ $CI: -11.32 \leftrightarrow 31.32$	10
2	$\bar{X}_{21} = 20$ $N_{21} = 2$ $CI: -1.32 \leftrightarrow 41.32$	$\bar{X}_{22} = 20$ $N_{22} = 25$ $CI: 13.97 \leftrightarrow 26.03$	20
	10.7	19.3	

Note. The confidence intervals on sample means are based on the pooled MS error with 50 degrees of freedom.

TABLE 2

SOME "REASONABLE" POPULATION MEANS

A \ B	1	2
1	10	10
2	20	20

A \ B	1	2
1	10	20
2	10	20

A \ B	1	2
1	10	20
2	30	20

effect; the second would be considered as an example of a main effect for B but no A or AB effect; whereas the third would be indicative of a situation with interaction and main effects. The present authors thus believe that the conclusion one should logically draw from these sample values is that there are some effects, but that the data do not permit a definitive statement as to which. They further believe that a proper analysis of variance should lead one to draw such conclusions.

An Incorrect Approximate Analysis

Consider an erroneous analysis that many users may be inclined to perform. This is an analysis of each factor collapsing over the other. Although it does have some intuitive appeal and indeed may be useful in conjunction with other analyses, it will in general lead to incorrect conclusions about the population means when *used alone*. Suppose the design given in Table 1 is collapsed over the B classification leaving two levels of A with mean values of 10 and 20, as shown in the marginal values of Table 1. A one-way analysis would then lead to the conclusion that there is a significant main effect of A ($p = .017$). If the design is then collapsed over levels of A, one has the levels of B with means of 10.7 and 19.3, suggesting a B effect ($p = .042$) as well as an A effect. These analyses would be called "A ignoring B" and "B ignoring A." The use of the phrase A ignoring B is meant to indicate that in the two-way table the B classification is "ignored" and the design treated as if it were only a one-way classification with levels of A. Observations for a given level of A are considered replicates regardless of whether or not they correspond to the same level of B. One assumes no B or AB effects. When there is no B or AB effect, the observations at the several levels of the collapsed factor are replicates because the variability between levels of B is of the same order of magnitude as the variability within a level of B. If, however, in a nonorthogonal design there is a B or an AB effect, the estimated magnitude of the A effect (ignoring B) is in general affected by the number of observations in the cells and does not represent an unbiased estimate of any population value. Only when there are equal numbers of observations in the cells is the estimate of the magnitude of the A effect unaffected by the number of observations in the presence of a B or AB effect.

When one estimates the magnitude of effects, he may safely ignore other effects in the design only when they are null or when their estimates are independent of the effects in which he is interested. The condition of null effects depends only on the state of nature; the condition of independence of estimates depends on the actual design of the experiment. The conclusions drawn from this ignoring analysis of Table 1 will be incorrect under the (plausible) assumption that there is only one main effect in the population responsible for the results.

For the general nonorthogonal case, a different method is necessary in order to estimate treatment effects without bias and to provide unbiased tests of significance. These are tests of "A eliminating B" and "B eliminating A" with corresponding estimates of the effects. They take into account and eliminate the confounding effects of other factors when they are present. Thus a test of A eliminating B removes any confounding effects of factor B. If there is no B effect or if the design is orthogonal, there is no confounding due to B and nothing to eliminate; hence, the test will be identical to that of A ignoring B. The test of A eliminating B answers the question: Given the possibility of a B effect, is there evidence for an A effect *in addition to* any B effect that may be present? On the other hand, the test of A ignoring B answers the question: Is there any evidence for an A effect assuming there is no B effect or ignoring it if it is present? The estimate of the A effect corresponding to the test of A eliminating B is unbiased regardless of the existence of any B effect or of orthogonality in the design. It is always the "correct" estimate.

Model Comparisons and Tests of Effects

The eliminating method described previously involves fitting a model allowing for both A and B effects and then comparing the fit (i.e., the quality of the model) to that

of a model omitting one or more of the effects. For example, consider the following models that "predict" the response of a subject in the ij cell of a two-factor design:

1. $Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon.$
2. $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon.$
3. $Y_{ij} = \mu + \beta_j + \epsilon.$
4. $Y_{ij} = \mu + \alpha_i + \epsilon.$
5. $Y_{ij} = \mu + \epsilon.$

Model 1 is the most complete model for a two-factor design since it allows for an interactive effect (γ_{ij}). The other models are obtained from the first by dropping the interaction term and possibly one or both of the main effects. Those accustomed to only orthogonal analysis of variance will be inclined to regard Model 1 as capable of providing the parametric estimates needed for the other models, but this is not so in general. Each model represents a separate least squares estimation problem and may provide different estimates of the parameters involved. Only in the case of orthogonality will the estimated parameters for the different models be necessarily the same. Likewise, it is only for the orthogonal case that the estimated parameters within a model will be statistically independent of one another. This is the real meaning of orthogonality.

The analysis of a two-way factorial, either orthogonal or nonorthogonal, begins with the test of interaction. Parsimony dictates a preference for a main effects model if it is consistent with the data, hence one would compare a model allowing for main effects and interaction (Model 1) with one only allowing for main effects (Model 2). In a two-factor complete factorial experiment, this is the usual test of the two-way interaction. If one is able to reject the hypothesis of null interaction effects, the usual procedure would be to stop at this point with an interaction model. If, however, one is unable to reject this hypothesis (i.e., conclude that interaction effects are nonsignificant), one would proceed to tests of main effects.

When one allows the possibility of both an A and B effect in the population, one specifies a series of tests involving Model 2. Thus to test either effect, one must test it in that model, implying an alternative model in which it is absent. To test for an A effect, Model 2 and Model 3 are compared; to test for a B effect, Model 2 and Model 4 are compared. In each of these tests, one allows for the possible existence of the effect not being tested. In testing A, one is asking the question, "Given the possible existence of B in this model, is A needed?" This is the meaning of the term A eliminating B.

The judgment as to which model to accept is based on the relative magnitudes of the sum of squared errors produced by the competing models, and the F test gives a method for testing whether the models differ in this respect. This procedure is always correct in either the orthogonal or nonorthogonal case. In the orthogonal case, it will produce results identical to those produced by the ordinary computational methods.

Different tests of A and B effects may be obtained by beginning with different model assumptions. If it is assumed that there is no B effect, Model 4 is an appropriate model and it would be compared to Model 5 in order to test the existence of an A effect in Model 4. This test of A ignoring B is not a proper test, however, unless Model 4 is the correct model. Similarly, one may test B ignoring A by comparing Model 3 against Model 5, but here the test is proper only if Model 3 is appropriate. In the case of an orthogonal design, these tests will give us the same results as those tests involving Model 2, but although the results are computationally the same (due to independence of the estimates of the parameters involved), they are not logically the same in terms of comparing the same models.

An Example

Let us now apply this method to the data of Table 1 using the multivariate analysis of variance program. All the relevant statistical tests may be summarized in the following analysis of variance table (Table 3):

There is no evidence for an interaction; however, the small numbers of observations in two of the cells makes the power of this test low. Tests of A eliminating B and B eliminating A are clearly nonsignificant, whereas the tests of A ignoring B and B ignoring A, given previously, are both significant. All five of these statistical tests must

TABLE 3

ANALYSIS OF VARIANCE TABLES FOR COMPLETE ANALYSIS OF DATA IN TABLE 1

Source	*df*	*SS*	*MS*	*F*	*p*
AB	1	.00	.00	.00	1.000
A eliminating B	1	370.37	370.37	1.646	.205
B eliminating A	1	.00	.00	.00	1.000
A ignoring B	1	1,349.99	1,349.99	5.999	.017
B ignoring A	1	979.63	979.63	4.353	.042
Within cells	50	11,250.00	225.00		

be considered in order to draw proper conclusions about the population means. The tests of A eliminating B and B eliminating A do not provide any evidence regarding the existence of either A or B effects (although they clearly imply that both effects are not necessary jointly), whereas the tests of A ignoring B and B ignoring A separately provide evidence for either effect, depending on which test is considered. Keeping in mind the models that are compared, the eliminating tests indicate that there is no evidence for one effect in addition to the other. One must conclude then from this analysis that there must be some effect, either an A or B effect (but one cannot tell which), and there is clearly no evidence to suppose that both exist. Because of the substantially disproportionate numbers of observations in cells, the power of the eliminating tests is rather low and the effects are highly confounded. Indeed, this example approaches closely the completely confounded case in which all observations would be in the A_1B_1 and A_2B_2 cells. In the completely confounded case, the one degree of freedom between cells could be attributed to either an A effect or a B effect with no possibility of deciding between them.

The patterns of possible results from the analysis of a two-factor design with no interaction are given in Table 4. Pattern 1 indicates that A and B are both needed in the model because, given the presence of one, the other is still significant. Patterns 2 and 3 both illustrate cases for which a second main effect is not needed given the inclusion of the other, but the significant effect must be included (i.e., from Pattern 2 one would retain the A effect and from Pattern 3, the B effect). Pattern 4 is the case for which no main effects are included in the final model. These constitute the standard cases and are the only cases that may arise from an orthogonal design. The remaining patterns are unique to the nonorthogonal case. Pattern 5 is the seriously confounded situation in which only one effect need be included in the final model, but the choice of which effect to retain is indeterminate. Patterns 6 and 7 occur only in situations in which there is very serious confounding in the design. The significant main effect should be included in the final model. In these circumstances it is particularly important to ask why such a seriously confounded design was produced.

RECOMMENDED PROCEDURE FOR A TWO-WAY NONORTHOGONAL DESIGN

The present authors suggest the following procedure be employed in the analysis of a nonorthogonal two-factor design. This procedure is for the logical flow of decisions and conclusions that are made in such an analysis, but it does not dictate the actual order in which the computations need be performed.

1. Begin with the full model including main effects and interaction effects.
2. Test for a significant interaction; if this test is significant no tests of main effects are appropriate; however, one may wish to test certain contrasts in the cell means to aid in interpretation of the results. If the test is nonsignificant, eliminate the γ_{ij} terms from the model and proceed to Step 3 for tests of main effects.

TABLE 4

PATTERN OF RESULTS: TWO-WAY FACTORIAL WITHOUT INTERACTION

Test	Pattern						
	1	2	3	4	5	6	7
A eliminating B	s	s	ns	ns	ns	ns	ns
B eliminating A	s	ns	s	ns	ns	ns	ns
A ignoring B	x	x	x	ns	s	s	ns
B ignoring A	x	x	x	ns	s	ns	s

Note. s = significant; ns = nonsignificant; x = irrelevant.

3. Test A eliminating B and B eliminating A. If both tests are significant, adopt the model $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon$. If only one of the two tests is significant, adopt the model $Y_{ij} = \mu + \alpha_i + \epsilon$ (if A eliminating B is the significant test) or $Y_{ij} = \mu + \beta_j + \epsilon$ (if B eliminating A is the significant test). If neither is significant, proceed to Step 4.

4. Test A ignoring B and B ignoring A. If both are significant, retain either α_i or β_j but not both in the final model—the choice is indeterminate. In this case additional experimental evidence usually has to be obtained before much can be said about the meaning of the experiment. If only one of the two tests is significant, the significant effect should be retained, but the cautions referred to in the discussions of Patterns 6 and 7 should be adhered to diligently. If neither test is significant, no main effects should be included in the final model, this is, adopt the model $Y_{ij} = \mu + \epsilon$.

Extensions to Higher Order Designs

As a nonorthogonal design becomes more complex, the basic logical structure remains the same. In all cases, one is attempting to find the simplest model that adequately fits the data by comparing competing models. For the three-factor model, the process begins by tentatively adopting the full model,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk},$$

and then eliminating unnecessary terms. First one would test the third-order interaction, ABC, eliminating all second-order interactions and main effects. If the test of the ABC interaction is significant, one would accept the full model and proceed, if desired, to test specific contrasts in cell means to aid with interpretation. If, on the other hand, the third-order interaction is nonsignificant, one would drop the $(\alpha\beta\gamma)_{ijk}$ term from the model and proceed to investigate the second-order interaction terms in order to determine how many and which terms to include in the model.

It is convenient to now consider the procedure for main effects rather than second-order interactions. This is because some of the concepts carry over directly from the two-factor design and, given certain symmetries in the three-factor design, it is possible to then directly apply these concepts to tests of interaction terms. It should be emphasized that in the actual use of the process, tests of second-order interaction would *always* precede tests of main effects. The notation A|B,C is used to mean A eliminating B and C, whereas A|B implies A eliminating B but ignoring C. In testing for main effects, one is trying to determine how many effects must be included in the model and which ones they are. The only circumstance under which it would be necessary to include all of the main effects is one in which each main effect is significant, eliminating the other two (i.e., when the tests A|B,C; B|A,C; and C|A,B are all significant). If only two of the three tests are significant, the two significant effects would be retained and the third would be deleted from the model. Thus if all three of the tests or if two of the three tests are significant, the conclusions are quite direct—retain the significant effects.

If only one of these tests is significant, say A|B,C, the significant term should clearly be retained; however, it may be necessary to retain one of the other two effects. Having decided to keep the A effect, one must ask if either the B or the C effect is needed given the A effect, that is, to test B|A and C|A. If neither of these tests is significant, then clearly neither effect need be present given the A effect in the model. If one of the two is significant, say C|A, that term, C, should be included in the final model along with the A term. Should, however, both be significant, the situation is ambiguous. Previous tests have indicated that all three effects are not needed in the model and that the A effect must be in the model; therefore, the choice between B and C is indeterminate.

The potentially most complicated situation obtains when none of the three "doubly eliminating" tests are significant. It is still possible that one or two effects should be included in the model. In the two-factor design, it was reasoned that the significance of both

A|B and B|A indicated that both A and B should be included. In the three-factor design, there are three such pairs of tests involving A and B, A and C, and B and C (i.e., A|B and B|A; A|C and C|A; and B|C and C|B). The joint significance of any one of these pairs of tests indicates the need to include the relevant pairs of effects, but only two such effects may be included, the previous tests having excluded the possibility of all three effects being included in the model. If more than one pair of these tests shows significance, one is uncertain as to which pair of effects to include. Should no pair of effects be significant, one is left with the possibility of including only a single effect in the model. Thus, if any one effect were to appear significant (e.g., if the tests of A|B or A|C were significant), it would be included in the model. Should none of the single eliminating tests be significant, one would then examine the "doubly ignoring" tests, A, B, and C, as these may still indicate the necessity of including a single main effect. If none of these tests are significant, one would conclude that no main effects were necessary and would be left with the model $Y_{ijk} = \mu + \epsilon_{ijk}$. If but one of these tests is significant, that effect would be included in the final model. If two or more of the doubly ignoring tests are significant, one is again in an indeterminate situation and may arbitrarily choose one of the significant effects for the final model, but the choice is completely arbitrary.

The application of the main effect procedure to two-way interaction is straightforward, if one lets AB|AC,BC mean AB eliminating AC, BC, and all main effects. There is then the following symmetry in the three-factor case. Because there are three two-way interactions and three main effects in a three-factor model, the patterns of tests for main effects and for two-way interactions are exactly the same. Corresponding to tests of main effects A, B, and C, there are tests of interaction, AB, AC, and BC. For every main effect test, say A|B,C, there is a corresponding test, AB|AC,BC. Hence the preceding procedure is first applied to the three two-way interactions, eliminating all main effects and other two-way interactions (i.e., AB|AC,BC, AC|AB,BC, and BC|AB, AC), and would then be followed with parallel tests as needed. Should the conclusion be that there are no interactions, the procedure would then be applied to the main effects. If there are significant interactions, the factors involved should be also included as main effects. Should only one two-way interaction be included, the question of retaining the uninvolved main effect should be considered. To do this, the tests of that effect eliminating the other two main effects and the significant interaction should be performed. For example, if it were the BC interaction that were significant, one should perform the tests A|B, C,BC in order to determine if the A effect should be included in addition to the B, C, and BC effects.

Some Additional Comments

The methods discussed for both the two- and three-factor cases are based on the assumption that there is no a priori preference for explaining the data in terms of one factor above any others. Such a preference may exist in designs such as randomized blocks in which one would customarily not even consider the tests of treatments ignoring blocks; one assumes that there are block effects and is willing to consider the presence of treatment effects only if the test of treatment eliminating blocks is significant. Similar considerations may apply in a wide variety of cases and may simplify the process discussed here.

Another consideration is the number of tests involved in the complete procedure. Some of these tests will be highly correlated and some will be independent, depending on the degree and pattern of nonorthogonality. In the case of lack of knowledge of likely effects, one may perform preliminary combined tests (such as a test of pooled interaction) prior to doing individual tests. This would have to be moderated, however, by any knowledge that would a priori suggest the existence of specific effects.

REFERENCES

COCHRAN, W. G., & COX, G. M. *Experimental designs*. New York: Wiley, 1957.

CRAMER, E. M. Significance tests and tests of models

in multiple regression. *The American Statistician,* 1972, **26,** 26–30.

JOE, G. W. Comment on Overall and Spiegel's "Least squares analysis of experimental data." *Psychological Bulletin,* 1971, **75,** 364–366.

OVERALL, J. E. & SPIEGEL, D. K. Concerning least squares analysis of experimental data. *Psychological Bulletin,* 1969, **72,** 311–322.

RAWLINGS, R. R., JR. Note on nonorthogonal analysis of variance. *Psychological Bulletin,* 1972, **77,** 373–374.

WERTS, C. E. & LINN, R. L. Causal assumptions in various procedures for the least squares analysis of categorical data. *Psychological Bulletin,* 1971, **75,** 430–431.

WILLIAMS, J. D. Two way fixed effects analysis of variance with disproportionate cell frequencies. *Multivariate Behavioral Research,* 1972, **7,** 67–83.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

(Received July 3, 1973)