Psychological Bulletin
1975, Vol. 82, No. 2, 182–186

Equivalence of Orthogonal and Nonorthogonal Analysis of Variance

John E. Overall
University of Texas Medical Branch, Galveston

Douglas K. Spiegel
University of South Dakota

Jacob Cohen
New York University

Much of the controversy surrounding dummy variate multiple regression approaches to nonorthogonal analysis of variance would be cleared up if a criterion could be accepted for deciding what constitutes a proper generalization of the classical analysis of variance for orthogonal factorial designs. It is proposed that a general multiple regression solution be interpreted as testing analysis of variance effects only if it results in an estimation of the same parameters and tests of the same hypotheses that might otherwise be estimated and tested in an orthogonal design involving the same factors. A method which satisfies this criterion is identified, and a simple procedure for examining equivalence in orthogonal and nonorthogonal cases is suggested.

Although it is true that there appears to be considerable confusion among psychologists about the appropriate application of multiple regression and correlation methods for analysis of variance of data from nonorthogonal designs (Appelbaum & Cramer, 1974), it is the belief of the present authors that this confusion is more apparent than real. The appearance of confusion is lessened if one carefully discounts a few obviously erroneous and naive comments that have made their way into the literature and if one also recognizes the tendency of certain other writers to imply discrepancies where none exist. More convincingly, the feeling of doubt about this highly useful approach can be eliminated if a simple criterion can be presented whereby any interested reader can verify for himself what is and what is not an appropriate generalization of classical analysis of variance to nonorthogonal designs. Although the primary purpose of this article is to provide a definition and method of evaluating the appropriateness of analysis of variance interpretation for nonorthogonal designs, it is first necessary to clear the air by commenting very briefly on several recent articles that would otherwise cast a cloud of suspicion on the recommendations presented here.

Wolf and Cartwright (1974) claimed that a discrepancy between Cohen (1968) and Overall and Spiegel (1969) prompted them to discover a "correct" method for coding dummy variates and purportedly to demonstrate that the 1,0, − 1 effects coding advocated by Overall and Spiegel does not yield correct results in nonorthogonal analysis of variance. Not only is the perceived discrepancy between the previous authors a misunderstanding on the part of Wolf and Cartwright, but the results that they themselves presented clearly verify that the 1,0, − 1 effects coding advocated by Overall and Spiegel produces regression coefficients that are estimates of the usual analysis of variance parameters. Other computational and logical errors in their article can easily be recognized and have been cited by others (Bogartz, 1975). Thus, disregard of the erroneous conclusions by Wolf and Cartwright is appropriate and intentional. The reader can take some comfort in the fact that there is no substantive disagreement between Cohen (1968) or Cohen and Cohen (1975) and Overall and Spiegel (1969). Cohen has tended to emphasize the flexibility of multiple regression and correlation as a general data-analytic technique, while Overall and Spiegel have been more concerned with

This work was supported in part by Grant DHEW 5 RO1 MH14675-07.

Requests for reprints should be sent to John E. Overall, Psychometric Laboratory, University of Texas Medical Branch, Galveston, Texas 77550.

the delineation of a narrower set of methods that are more consistent with conventional analysis of variance interpretation. Yet, there is no substantive disagreement between these authors on coding or analysis as far as conventional analysis of variance effects are concerned.

Rock (1974) has added to the appearance of controversy by rediscovering and misinterpreting the fact noted by Overall and Spiegel (1969, p. 319) and by numerous other authors that the component sums of squares in nonorthogonal designs add to the total sum of squares only when Method 3 is used. He offers the curious conclusion that $SS_T(1 - R^2_{\alpha_i,\beta_j,\alpha\beta_{ij}})$ is identical for all methods, but that its complement, $SS_T(R^2_{\alpha_i,\beta_j,\alpha\beta_{ij}})$, is not the same for the different methods described by Overall and Spiegel. Evidently, from the remainder of his discussion, Rock means that the particular treatment effects are underestimated in Method 1 and 2 analyses because "the correlated components are dropped" (p. 1013). However, this is not actually the case either because the component sums of squares for Method 1 may add to either a larger or a smaller total than the component sums of squares for a Method 3 analysis of the same data. Thus, Rock's preference for Method 3 analysis should be discounted on technical grounds, even before one considers the criterion of equivalence of hypotheses that are tested in orthogonal and nonorthogonal designs.

Several writers have generated controversy by describing alternative methods which lead to similar results (Gocka, 1973; Overall & Spiegel, 1973a, 1973b; Rawlings, 1972). In a recent article, Appelbaum and Cramer (1974) included Overall and Spiegel (1969) among authors who 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" (p. 335). Whatever the faults of our previous work, any reader can verify that they do not include any one of those just mentioned. Such comment is all the more surprising because Appelbaum and Cramer then proceeded to outline a strategy for the comparison of models of varying complexity (pp. 340–341), which strategy is that described by Overall and Spiegel (1969) as Method 2 and which analysis is accomplished directly in a single pass by a computer program presented by Overall and Klett (1972). Rather than contributing to a consensus in the field, their article leaves the reader with further doubts that there is any agreement at all. Since we recognize that Appelbaum and Cramer have presented a lucid discussion of Method 2, much of the controversy is more apparent than real.

Authors who propose alternative computational methods while implicitly or explicitly disavowing the appropriateness of other equivalent solutions have done much to contribute to the apparent confusion. Bogartz (1975) questions the validity of concern over the "coding problem" while presenting only an alternative computational formula. He implies that the computational solution obviates entirely the problem of coding; yet the *X* matrix in his formula contains a 0–1 code vector for each individual, and by his own statement the transformation matrix *H* consists of the coefficients of linear contrasts among treatment means. That is exactly what dummy variate coding represents! [1] At most, he has merely presented an alternative computational method in which dummy variate coding is incorporated into transformation matrices that are applied to previously computed mean vectors. It would be interesting to see his solution for a complex nonorthogonal design involving 10 or more factors with almost invariably some empty cells. If successful, he will have coded (at least implicitly) a full rank design matrix, will have made decisions about contrasts which are tantamount to transformation to dummy variate coding, effects coding, contrast coding, or nonsense coding (Cohen & Cohen, 1975), and his results will be identical to those obtained alternatively from equivalent coding in a multiple regression solution.

Having hopefully dispelled some of the doubts about simple 1,0, – 1 effects coding

[1] Throughout this article *dummy variate* is employed in a generic sense to indicate artificial variates. Cohen prefers to distinguish between different types of artificial variate coding with dummy coding designated as one particular type.

for a multiple regression approach to nonorthogonal analysis of variance, let us now consider the problem of defining a solution for the nonorthogonal case which is consistent with the usual interpretation of analysis of variance effects in orthogonal designs. The original article and subsequent comments by Overall and Spiegel (1969, 1973a, 1973b) have had the aim of pointing out that several strategies which have been recommended by different authors for dealing with problems of nonorthogonality actually result in the testing of different hypotheses. The strategy that appeared most often to be recommended in applied statistics texts involves basically a "main effects" model with tests for interaction effects included as a safeguard against departures from additivity (Rao, 1965; Snedecor & Cochran, 1967; Winer, 1971). The analysis proposed by Appelbaum and Cramer (1974) follows this logic, and the analysis proposed by Rawlings (1972) yields identical results, with the possible exception of using a pooled error term (Overall & Spiegel, 1973a). Thus, on the basis of popular acceptance, the approach designated as Method 2 seemed to us the one most appropriate for analysis of experimental data involving disproportionate cell frequencies.

Rather than merely accepting popular usage as the basis for choosing the general linear regression solution which is most nearly equivalent to conventional analysis of variance for balanced experimental designs, the controversies that have followed have caused us to seek a more rational criterion for choice among alternative methods. The criterion we propose is that the method for the analysis of variance of data from nonorthogonal designs should estimate the same parameters and test the same hypotheses as can otherwise be estimated and tested in a balanced analysis of variance experimental design involving the same factors. We emphasize that it is the conventional analysis of variance model and effects that are under consideration; alternative models and functions of population parameters are considered by Cohen (1968) and Cohen and Cohen (1975).

The procedure which is proposed for verifying that a particular method of analysis satisfies the analysis of variance criterion is to generate data for orthogonal and nonorthogonal designs involving exactly the same α, β, and $\alpha\beta$ parameters and then to determine which method of analysis yields the same parameter estimates in the orthogonal and nonorthogonal cases. In Table 1 data are presented for a two-way analysis of variance design. In each cell there are three observations above the dotted line and either none or three below the dotted line. The three observations above the line in each cell can be considered to represent data for a balanced two-way design with three observations per cell. It will be noted that (not by coincidence) the three scores below the dotted line, in cells in which they are present, duplicate those above the line. This mechanism is employed to change the cell n without altering the mean. The reader will appreciate that duplication of certain scores does not invalidate the analysis of variance.

Any interested reader can verify that exactly the same parameter estimates result from an analysis of the orthogonal ($n = 3$) and nonorthogonal designs when one codes the research factors using 1,0, − 1 effects coding, represents the interaction vectors by multiplication, and uses Method 1 as described by Overall and Spiegel (1969). The same cell means can be calculated using the partial regression coefficients from either the orthogonal or nonorthogonal analysis as parameter estimates. The "adjusted means" for the A and B main effects in the nonorthogonal case are precisely the row and column means of the ($n = 3$) orthogonal design. The general regression constant in the nonorthogonal case is the unweighted mean of all the cell means and hence precisely the grand mean in the orthogonal case. Thus, the 1,0, − 1 effects coding in the nonorthogonal case provides estimates of the effects that one would expect for the same factors in a balanced experimental design. Timm and Carlson (in press) provide an analytic demonstration of this same conclusion for the more mathematically inclined reader.

Although the parameter estimates, and thus the effects being tested, are identical in the orthogonal and nonorthogonal cases, the nonorthogonal analysis involving larger total N yields larger F ratios than the orthogonal analysis for testing the same main effects and interactions present in Table 1. In a substantial series of previously unreported Monte Carlo studies, analyses of nonorthogonal designs have been compared with analyses of orthogonal designs obtained by randomly eliminating observations in certain cells of the nonorthogonal design to render all cell frequencies equal. The larger total N in the nonorthogonal analysis routinely provides greater power in testing the same hypotheses than can otherwise be tested by eliminating observations to artificially balance the design, as anticipated by Cohen (1969, chap. 8). On the other hand, for the same total N, equal cell frequencies provide greater power in tests of the same hypotheses than may otherwise be tested in a nonorthogonal design. For the same total N, the more disproportionate the cell frequencies, the lower the power in testing conventional analysis of variance effects. Nevertheless, the effects estimated and the hypotheses tested in the two cases are the same.

TABLE 1

Data for Orthogonal and Nonorthogonal Designs Generated for the Same Analysis of Variance Model

	B_1	B_2	B_3
A_1	3 6 9 - - - 3 6 9	6 4 7	9 12 14
A_2	5 6 8	6 5 7 - - - 6 5 7	3 5 8
A_3	10 12 9	11 13 15 - - - 11 13 15	12 8 10 - - - 12 8 10

Overall and Spiegel (1969, 1973a, 1973b) previously recommended Method 2 as being consistent with strategies advocated by the majority of authors of well-known statistics texts for the analysis of data from designs involving unequal and disproportionate cell frequencies. However, neither Method 2 nor 3 as described by Overall and Spiegel (1969) nor the frequency-weighted coding described by Gocka (1973) yield identical parameter estimates for the orthogonal and nonorthogonal cases represented in Table 1. Appelbaum and Cramer (1974) and Rawlings (1972) proposed strategies for the analysis of nonorthogonal designs that are similar to Method 2; thus, the proposals by these authors also result in analyses that fail to yield the same main effect parameters in orthogonal and nonorthogonal cases in which a true but nonsignificant interaction is present. On the basis of these types of comparisons of orthogonal and nonorthogonal cases, it is concluded that 1,0, − 1 effects coding of dummy variates and Method 1 analysis (which is fully simultaneous for all effects) is the appropriate nonorthogonal generalization of conventional analysis of variance for balanced factorial designs. The same type of coding followed by Method 2 analysis (which is simultaneous only for main effects) is appropriate for testing main effects, as is the equivalent Appelbaum and Cramer (1974) strategy, only if it can be assumed a priori that no true interactions exist. Method 3 analysis (which is hierarchical and uses an a priori ordering of all main effects and interactions), used with any method of coding, can never be expected to test the same hypotheses in a nonorthogonal design that can otherwise be tested in an orthogonal design involving the same factors. The criterion of equivalence in orthogonal and nonorthogonal designs that has been proposed here seems clearly to favor Method 1 with 1,0, − 1 effects coding of dummy variates as described by Overall and Spiegel (1969), and thus we find ourselves in agreement with Carlson and Timm (1974).

The intent of this discussion is to emphasize the equivalence of parameter estimates and tests of significance in orthogonal and nonorthogonal designs, which is not to be confused with the question of estimability of main effects in the presence of true interactions. With regard to what we intend by the estimation of main effect parameters that are equivalent for orthogonal and nonorthogonal designs, the following Monte Carlo demonstration can be convincing. Let one individual generate data according to the analysis of variance model and restrictions specified as follows:

$$X = \mu + \alpha + \beta + \alpha\beta + \epsilon ,$$

where ϵ is $N(0,\sigma_\epsilon^2)$ with restrictions $\Sigma\alpha = \Sigma\beta = \Sigma\alpha\beta = 0$. Give the data to someone else who does not know the parameter values used to generate the data. The second party can obtain unbiased estimates of the main effect and interaction parameters as long as the model and restrictions employed to analyze the data are the same as those used in generating them. This is true for either the orthogonal or nonorthogonal case. Thus, our position appears consistent with Scheffé (1959, p. 110), who stated that tests of main effects are valid regardless of the true values of the interactions, although the interpretation may be difficult if interactions are present. Again, it is not the point of this discussion to argue the question of estimability, but only to emphasize that the situation is directly equivalent in orthogonal and nonorthogonal cases.

The problem with which we have been concerned, in view of the growing use of general linear regression methods for the analysis of experimental data, is that different methods test hypotheses which are different from those tested in ordinary balanced experimental designs; yet all are described as analysis of variance and the effects that are tested are interpreted as main effects and "interactions" in the usual analysis of variance conception. We should not presume to deny the informed investigator the great flexibility of general linear regression methods for the analysis of experimental data using a variety of models (Cohen, 1968; Cohen & Cohen, 1975). He should, however, clearly understand the hypotheses that he is actually testing and be able to communicate their nature in a truly understandable manner to the general scientific audience to which his work is addressed. The terminology of conventional experimental design analysis of variance should not be used, without explicit qualification, to describe results from analyses that test hypotheses that are different from the main and interaction effects usually tested in balanced experimental designs.

REFERENCES

Appelbaum, M. I., & Cramer, E. M. Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin,* 1974, *81,* 335–343.

Bogartz, W. Coding dummy variables is a waste of time: Reply to Wolf and Cartwright, among others. *Psychological Bulletin,* 1975, *82,* 180.

Carlson, J. E., & Timm, N. H. Analysis of nonorthogonal fixed-effects designs. *Psychological Bulletin,* 1974, *81,* 563–570.

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

Cohen, J. *Statistical power analysis for the behavior sciences.* New York: Academic Press, 1969.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Gocka, E. F. Regression analysis of proportional cell data. *Psychological Bulletin,* 1973, *80,* 28–30.

Overall, J. E., & Klett, C. J. *Applied multivariate analysis.* New York: McGraw-Hill, 1972.

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

Overall, J. E., & Spiegel, D. K. Comment on Rawlings' nonorthogonal analysis of variance. *Psychological Bulletin,* 1973, *79,* 164–167. (a)

Overall, J. E., & Spiegel, D. K. Comment on "regression analysis of proportional cell data." *Psychological Bulletin,* 1973, *80,* 28–30. (b)

Rao, C. R. *Linear statistical inference and its applications.* New York: Wiley, 1965.

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

Rock, D. A. Appropriate method for least squares analysis of categorical data. *Psychological Bulletin,* 1974, *81,* 1012–1013.

Scheffé, H. *The analysis of variance.* New York: Wiley, 1959.

Snedecor, G. W., & Cochran, W. G. *Statistical methods.* Ames, Iowa: University of Iowa Press, 1967.

Timm, N. H., & Carlson, J. E. Analysis of variance through full rank models. *Multivariate Behavioral Research: Monograph,* in press.

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

Wolf, G., & Cartwright, B. Rules for coding dummy variables in multiple regression. *Psychological Bulletin,* 1974, *81,* 173–179.

(Received August 19, 1974)