JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1974, 7, 635-638 NUMBER 4 (WINTER 1974)

FORCING SQUARE PEGS INTO ROUND HOLES: SOME COMMENTS ON "AN ANALYSIS-OF-VARIANCE MODEL FOR THE INTRASUBJECT REPLICATION DESIGN"

DONALD P. HARTMANN[1]

UNIVERSITY OF UTAH

This paper critically examines the application of fixed-effect one-way analysis-of-variance procedures to learning data from a single subject. Procedures more appropriate for data obtained from intrasubject replication designs are briefly described.

Gentile, Roden, and Klein (1972) described an analysis-of-variance (ANOVA) model for detecting treatment effects in "noisy" single-subject reversal designs.[2] Their efforts to integrate useful elements of general psychology with operant technology are to be applauded. Nonetheless, the statistical model they recommend, a fixed-effect one-way ANOVA, deserves comment.

Briefly, Gentile *et al.* recommend collapsing the data from the four conditions of a typical reversal design (A_1, B_1, A_2, and B_2) into two treatment levels, baseline and treatment. (Preferably each condition would contain an equal number of data points.) Trials within conditions are considered replications—analogous to subject replication in a simple ANOVA. Thus, the single subject is considered to function as a random response generator; *i.e.,* the data within the two treatment cells are assumed to be statistically independent and normally distributed about the treatment means. The analysis associated with this model is summarized in Table 1.

Before examining the applicability of this model to the $N = 1$ intrasubject replication design, consider a set of idealized data (presented in Figure 1) resulting from an application of such a design to a program designed to accelerate a selected target behavior. These fictitious data have the following characteristics: baseline 1 rates are low and stable; during treatment 1, the rate of responding is positively accelerated as the subject experiences the changed contingencies, and finally reaches a stable asymptote; baseline 2 is a mirror image of treatment 1; and finally, treatment 2 is essentially a duplication of the rate displayed during treatment 1.

Table 1

Summary table for a one-way fixed effects ANOVA applied to data from an $N = 1$ reversal design.

Source	*df*	*MS*	*F*
Treatment (X_A *vs.* X_B)	1	$MS_{Treatment}$	$MS_{Treatment}/MS_{Error}$
Error (within cells)	$4n-2$	MS_{Error}	

NOTE—Each of the four treatment conditions, A_1, B_1, A_2, and B_2 contains n observations.

Data such as these are seldom if ever seen, and certainly would not require the application of statistical procedures to detect treatment effects; only with the addition of background noise would such procedures be required.

Assuming that the form of the data shown in Figure 1 (or some simple transformation) is

[1]I wish to thank David Dodd, Donna Gelfand, and Tom Malloy for their critical reading of an earlier draft of this paper.

This is one in a series of articles available for $1.50 from the Business Manager, *Journal of Applied Behavior Analysis,* Department of Human Development, University of Kansas, Lawrence, Kansas 66045. Ask for Monograph #4.

[2]The critical comments made in this paper are also applicable to the application of two-way ANOVA procedures for $N > 1$, as discussed by Gentile *et al.* (1972).

prototypic of the data gathered by behavioral researchers, it can reasonably be asked how well the model suggested by Gentile *et al.* fits data of this general form. According to Hays (1963), the assumptions of the one-way fixed-effects ANOVA model include:

a. a normal distribution of error components;
b. homogeneity of variance of error components;
c. and independence of error components; *i.e.,* for any pair of observations i and j, the expected value of $r_{e_i e_j}$ must equal zero, whether e_i and e_j are selected from the same or different treatment conditions.

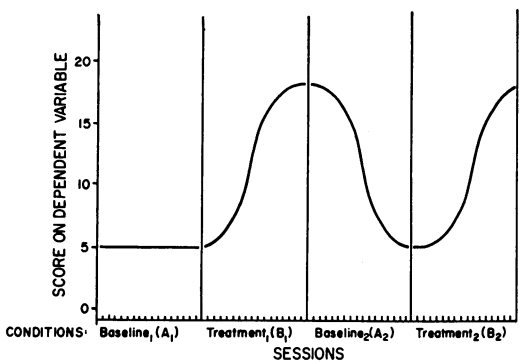

Fig. 1. Idealized data representing the application of reversal procedures to an acceleration problem.

All of these three assumptions appear to be violated with data resembling those presented in Figure 1. The distribution of error components in B_1, A_2, and B_2 would likely be more platykurtic than normal because of the large number of extreme scores in the flattened-out portions of the curves at the beginning and end of each condition except baseline 1. The error variances in A_1 and A_2 would be heterogeneous, as would be the error variances of the combined A conditions in comparison with the combined B conditions. In both cases, the heterogeneity would be due to the greater variability of scores in B_1, A_2, and B_2 in comparison with A_1.

Although these violations may be of relatively small concern (see, for example, Hays, 1963), it is violation of the third and final assumption that may provide a serious, and perhaps lethal threat to the use of the statistical model proposed by Gentile *et al.* According to Hays (1963), if the assumption of uncorrelated errors is not met (*i.e.,* if the data are sequentially dependent) "*very serious errors in inference can be made* (p. 379, author's italics)."

The within-cell dependencies that are of greatest concern are a result of the failure to consider the *systematic changes occurring across trials* within the B_1, A_2, and B_2 conditions. That is, the presence of a trial effect within conditions will result in positive serial correlations within conditions even though each trial score also has a random error component. The presence of this within-cell dependency poses a number of thorny problems for the use of the *F*-statistic. First, dependency obviously reduces the amount of independent information included in the data. This suggests that the number of degrees of freedom recommended by Gentile *et al.* is inflated and would result in a positively biased *F*-test. Second, nonindependence generally produces artificially lower variability, which also results in a positively biased *F*-test.

In addition to the problems caused by multiple violations of these assumptions, there is still another, and perhaps equally troublesome, problem with the ANOVA model proposed by Gentile *et al.* when applied to data resembling those shown in Figure 1. Again, the problem is a result of treating systematic changes within conditions as error. Thus, instead of calculating error variance in B_1, A_2, and B_2 about some linear, or, in the case of Figure 1, cubic trend, error variance in the proposed model is based on the deviation of the scores within conditions from the condition *mean, i.e.,* from the combined A and the combined B condition means respectively. This procedure greatly increases the magnitude of the MS_{Error} and consequently decreases the probability of detecting a true treatment effect. The problem is exacerbated by basing each condition mean on all data points within that condition, rather than just those obtained when asymptotic levels of performance have been reached.

SUMMARY AND RECOMMENDATIONS

In summary, the problems of increased error variance and reduced treatment variance, together with the various violations of assumptions, taken singly and in combination, suggest that the ANOVA model recommended by Gentile *et al.* should not be used with data resembling those shown in Figure 1. Its use should be restricted to data sets in which the aforementioned assumptions can be met, until such time as the nature and extent of the violations of the *F*-test are more fully examined.

If a researcher still insists on forcing square pegs into round holes by applying traditional ANOVA models to $N = 1$ data, then he should use either the relatively unexplored, but more sophisticated ANOVA model suggested by Shine and Bower (1971) and Shine (1973) or the variation of the one-way fixed-effect model described below.

This variation differs from the model suggested by Gentile *et al.* on two important points: its application is preceded by specific tests of the vexing independence assumption; and its use is restricted to data that are likely to meet both the independence assumption and the other previously noted ANOVA assumptions. The data requirements and statistical procedures for the proposed factorial model include:

1. Sufficient data so that an equal number of *stable* data points are available for each of the four treatment conditions. Thus, the analysis incorporates only the last n data points in each condition obtained during asymptotic responding; *i.e.*, when the regression of time on the dependent variable has zero slope. Because of the necessity to perform correlational analyses to test the independence assumptions (see below), at least 12 and preferably many more stable data points should be available for each condition.

2. Data that meet the independence assumptions. Two separate tests of independence are employed: first, on the serial correlations of at least lag one calculated on the n data points within each of the four conditions (see Holtzman, 1963); second, on the cross-serial correlations of at least lag zero calculated between trials across conditions, *i.e.*, $r_{A_1B_1}$, $r_{A_1A_2}$ . . . $r_{A_2B_2}$. If neither the serial correlations nor the cross-serial correlations are different from zero, it may be assumed that the data do not violate the independence assumption, and formal statistical analysis may be initiated.

3. Application of either an ANOVA set-up similar to the one summarized in Table 1 or a fixed-effects 2×2 completely randomized factorial model (summarized in Table 2) with n data points per condition. The data analysis procedures for both models are described in a straightforward manner in standard statistical texts. The design, as it is summarized in Table 2, allows testing of main effects due to treatments and to changes across time, and the interaction of treatment and time. For $N > 1$, a new fixed factor, subjects, would be added; the calculations and interpretations would follow those outlined for three factor factorial designs.

Table 2

Summary table for a fixed-effects 2×2 factorial ANOVA applied to data from a $N = 1$ reversal design.

Source	*df*	*MS*	*F*
Treatments (X_A *vs.* X_B)	1	MS_T	MS_T/MS_{Error}
Order (A_1+B_1 *vs.* A_2+B_2)	1	MS_O	MS_O/MS_{Error}
Treatment × Order	1	$MS_{T\times O}$	$MS_{T\times O}/MS_{Error}$
Error	$4(n-1)$	MS_{Error}	

NOTE—Each of the four treatment conditions, A_1, B_1, A_2, and B_2 contain n observations.

Researchers less tied to traditional ANOVA models who find it necessary to apply statistical procedures to reversal design data may want to consider a number of methods developed more specifically for time series. Foremost among these are the promising new generating-function procedures described in some detail by Gottman, McFall, and Barnett (1969). These procedures make "positive use of dependency observations (p. 302)"; they can be applied to one or more

subjects, and they have associated inferential statistics to provide evaluation of both between- and within-subject effects. Additional material on time series analysis can be found in Gottman (1973) and Glass, Willson, and Gottman (1973).

REFERENCES

Gentile, J. R., Roden, A. H., and Klein, R. D. An analysis-of-variance model for the intrasubject replication design. *Journal of Applied Behavior Analysis,* 1972, **5,** 193-198.

Glass, G. V., Willson, V. K., and Gottman, J. M. *The design and analysis of time series experiments.* Boulder, Colorado: Laboratory of Educational Research Press, 1973.

Gottman, J. M. *N*-of-1 and *N*-of-2 research in psychotherapy. *Psychological Bulletin,* 1973, **80,** 93-105.

Gottman, J. M., McFall, R. M., and Barnett, J. T. Design and analysis of research using time series. *Psychological Bulletin,* 1969, **72,** 299-306.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart, and Winston, 1963.

Holtzman, W. H. Statistical models for the study of change in the single case. In C. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1963. Pp. 199-211.

Shine, L. C. A multi-way analysis of variance for single-subject designs. *Educational and Psychological Measurement,* 1973, **33,** 633-636.

Shine, L. C. and Bower, S. M. A one-way analysis of variance for single-subject designs. *Educational and Psychological Measurement,* 1971, **31,** 105-113.

Received 28 September 1972.
(Published without revision.)