JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1974, 7, 639-641 NUMBER 4 (WINTER 1974)

"AN ANALYSIS-OF-VARIANCE MODEL FOR INTRASUBJECT REPLICATION DESIGN": SOME ADDITIONAL COMMENTS[1]

CARL E. THORESEN AND JANET D. ELASHOFF

STANFORD UNIVERSITY

The fixed effects ANOVA procedure utilized by Gentile, Roden, and Klein (1972) for single subjects is found inappropriate. Hartmann's proposal of a one-way fixed-effect ANOVA model is also considered. Time series analysis that takes serial correlation effects into account is recommended.

Gentile, Roden, and Klein (1972) identified an important problem in data analysis for the applied researcher. Often, the data from intensive studies of single subjects over time fails to provide clear-cut evidence of significant behavior change. Reliance on visual inspection as a basis for decision making is often invalid. White (1971), for example, demonstrated that individuals vary widely in their interpretation of data based on visual inspection—even to the point that some interpreted a trend as being accelerating and others judged the same trend to be decelerating. Huff (1954) showed how easily the eye could be misled by graphs and charts that distort the data. There is an obvious need for applied researchers to employ statistical techniques to make conclusions about what is happening to data within and between phases.

Gentile *et al.* (1972), acknowledging this problem, proposed a simple analysis-of-variance approach to study changes of the individual over time. There are several serious problems, however, in using a standard analysis of variance with such repeated measures data. As Hartmann (1974) points out, the basic assumptions of an analysis-of-variance model are typically violated when continuous data from the same subject are gathered over time. These assumptions include (1) a normal distribution of error components, (2) homogeneity of variance of error components, and (3) the independence of error components. Hartmann appropriately pointed out that the last assumption, that of dependence, is an assumption violated with fatal consequences. Serial correlation in the data tends to inflate the degrees of freedom involved and also lowers the variability within phases, thereby yielding a positively biased F ratio.

Hartmann also raised a crucial question about the marked limitations of relying on a mean value and deviations around a mean within a phase, rather than looking at the performance trend within a phase. Indeed, the major advantage of intensive designs is that they avoid the "static" reliance on mean performance and allow the investigator to examine change within a phase over time (Sidman, 1960; Thoresen, 1972). Applied researchers are well aware of the fact that two phases can have identical mean values, yet the slope or trend of the data in one phase can be sharply accelerating while that of a second phase can be dramatically decelerating. Hence, reliance on analytic models that consider only variability around a mean performance ignore what might be called the "dynamic" aspects of intensive designs. While Hartmann (1974) identified major problems with the Gentile *et al.* strategy, some additional observations are worth noting.

Gentile *et al.* err in assuming that the dependent variable, number of on-task behaviors, has a binomial distribution. First, it is most unlikely from the description of the experiment that two

[1]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.

successive observations of on- or off-task behavior are independent. In such an experiment, it would be preferable to use relative frequencies of on-task behavior by observation period or by task as the unit of analysis. The problem of non-independence from observation to observation or from treatment to treatment is not resolved by the combining of phases ($A_1 + A_2$, $B_1 + B_2$). Such a combination does *not* deal with the important problem of serial correlation effects *within* each phase. Any positive correlation of observations within a phase yields a positively biased F ratio. In addition, we can also expect the "true probability" of on-task behavior to change across time *during* a phase. Such a change violates the assumption necessary for the binomial, *i.e.,* each trial has the same probability of success. In fact, there is evidence presented by Gentile *et al.* that the "true probability" of success differs between phases for the same treatment. For example, the proportion of on-task behavior for James when compared for the first phase (A_1) and the fourth phase (A_2) yields a χ^2 value (4.39) significant at the 0.05 level. Hence, pooling the scores for A_1 and A_2 definitely violates the binomial assumption.

Interestingly, the analysis-of-variance model may not even be appropriate for the idealized coin-tossing experiment they describe. If the coin itself is not allowed to adapt to the surrounding temperature *before* beginning each phase, and if the warming-up or cooling-down phase is included within the data for a particular phase, the basic assumptions of the analysis-of-variance model are violated.

Other points merit comment. First, there is no logical basis for letting the number of observation periods vary so widely in each phase (see Table 1, p. 195). The range is approximately five-fold from 210 observations in one phase to almost 1000 observations in another phase. Second, conclusions about the effects of treatments for James and Lynn, the two subjects, hold only when these subjects are considered as a fixed effect. If these two were considered as a random sample of subjects, with generalizations to be made to a population of similar subjects, the F test for treatments would have been insignificant (3.48, d.f. = 2,2).

Finally, it should be noted that the proposed "t-test analysis", where only two treatments and one subject are involved, is identical to the analysis of variance.

HARTMANN'S REJOINDER

Hartmann offers an idealized model (Figure 1) for data involved in a reversal design. He appropriately points out that before using an ANOVA model one must first test for the assumption of independence, *i.e.,* serial correlation. In addition, there must also be a sufficient number of "stable" data points in each of the four treatment conditions. Some problems exist, however, with the Hartmann model. First, failure to find a significant serial correlation of Lag 1 (that is, is Observation No. 1 independent of Observation No. 2, No. 2 independent of No. 3?, and so on) does not guarantee independence. There may be a systematic bias within a phase represented by a Lag 5 relationship so that a teacher's behavior, for example, on Mondays and Fridays is highly correlated while Monday-Tuesday and Tuesday-Wednesday comparisons are not significantly correlated. In addition, tests of correlation coefficients are not very powerful unless sample sizes are large.

Hartmann's suggestion that the analysis incorporate only the "last *n* data points in each condition obtained during asymptotic responding", although a plausible suggestion, may present difficulties in many real situations. Typically, the data pattern within a phase is more likely to be accelerating, decelerating, or curvilinear. Thus, even if the regression of time on the dependent variable has a zero slope within a phase, it may not correspond to the last few data points within a phase. In practice, it is not easy to identify an interval when data are "stable".

The ANOVA model suggested by Shine and Bower (1971) also offers little solace to the applied researcher. These authors in effect propose a two-way fixed-effects analysis-of-variance model with one observation per cell. Its appropriateness is limited to a special case where responses are *in no way sequentially dependent within treatments,* although there may be restricted types of correlation patterns between treatments. Applied researchers seldom deal with behavior that is completely independent from observation to observation.

ALTERNATIVES TO ANOVA DESIGNS

A preferred strategy to ANOVA is based on various time-series analyses (*e.g.,* Gottman, McFall, and Barnett, 1969). The solution to "noisy" data about which the researcher wishes to make some inferences may be best found in analysis techniques that systematically take into account serial correlation effects. Glass, Willson, and Gottman (1973) offered an excellent methodological discussion of various intensive or time-series designs, especially concerning the problems of confounding factors with repeated measures. These authors, building on earlier efforts (*e.g.,* Box and Taio, 1965), offer what is called an "integrated moving average" method. This procedure allows the researcher to make probability statements about changes in level and slope between treatment phases. A recent example of this procedure is reported by Gottman and McFall (1972) in a study of self-monitoring effects in a high-school classroom. An alternative method has been suggested by White (1971), based on the use of median derived slopes to describe progress within and between phases. White utilized this method with a large number of classroom intervention studies to examine changes in level and slope between phases. The advantages of this median-based method over a standard regression analysis strategy are currently being examined. Some questions exist, for example, in whether the median procedure adequately deals with the effects of serial dependence.

A thorough discussion of these procedures and others is beyond this brief note. However, the applied researcher should know that some appropriate methods for analyzing intensive experiments are available in the literature. Hopefully, the next few years will see an expansion of efforts to develop appropriate statistical methodologies for intensive research designs.

REFERENCES

Box, G. E. T. and Tiao, G. C. A change in level of non-stationary time series. *Biometrika,* 1965, **52,** 181-192.

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. L., and Gottman, J. M. *Time series analysis in the behavioral sciences.* Boulder: Laboratory of Education Research, University of Colorado, 1973.

Gottman, J. M., McFall, R. M., and Barnett, J. T. Design and analyses of research using Time Series. *Psychological Bulletin,* 1969, **72,** 299-306.

Gottman, J. M. and McFall, R. M. Self-monitoring effects in a program for potential high school dropouts: a time series analysis. *Journal of Consulting and Clinical Psychology,* 1972, **39,** 273-281.

Hartmann, D. P. Forcing square pegs into round holes: some comments on "An analysis-of-variance model for the intrasubject replication design." *Journal of Applied Behavior Analysis,* 1974, **7,** 635-638.

Huff, D. *How to lie with statistics.* New York: Norton, 1954.

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

Sidman, M. *Tactics of scientific research.* New York: Free Press, 1960.

Thoresen, C. E. The intensive design: an intimate approach to research. *The Counseling Psychologist (in press).*

White, O. R. *Pragmatic approaches to the progress in the single case.* Unpublished doctoral dissertation, University of Oregon, 1971.

White, O. R. *A manual for the calculation and use of the median slope: a technique of progress estimation and prediction in the single case.* Working Paper No. 16. Regional Resource Center for Handicapped Children, University of Oregon, Eugene, Oregon, 1972.

Received 1 March 1973.
(Published without revision.)