JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1974, 7, 643-645 NUMBER 4 (WINTER 1974)

CONCERNING THE STATISTICAL PROCEDURES ENUMERATED BY GENTILE et al.: ANOTHER PERSPECTIVE

H. J. KESELMAN[1] AND L. LEVENTHAL

UNIVERSITY OF MANITOBA

The thrust of this paper is to bring to the attention of operant researchers statistical procedures that are appropriate for correlated data. In addition to specifying these statistical procedures consideration is given to the question of using individual comparison statistics rather than omnibus F tests. Specifically, it is recommended that a more powerful test of the experimental hypotheses can be obtained by performing Bonferroni t statistics rather than analysis of variance F tests.

Gentile, Roden, and Klein (1972) and Shine and Bower (1971) maintained that operant behavioral data can be analyzed with between-subjects, (nonrepeated factors) analysis-of-variance (ANOVA) statistics, holding that successive measurements taken on the same experimental subject are not statistically related (correlated) but rather are as independent as are the outcomes of tossing a fair coin (Gentile *et al.*, 1972, p. 195). Although the present author does not find these rationales palatable, this paper is not a lengthy discourse against their viewpoints; it seeks rather to alert the operant researcher to statistical techniques that are appropriate for analyzing correlated data.

Gentile *et al.* (1972, p. 196) described an operant paradigm in which successive measurements are taken on two subjects for repeated administrations of a treatment variable. Schematically, the experimental design is given in Figure 1. The sources contributing to the magnitude of the dependent variable are given by (1)

$$Y_{ijk} = \mu + \eta_i + \alpha_j + \eta\alpha_{ij} + \beta_k + \eta\beta_{ik} + \alpha\beta_{jk} + \eta\alpha\beta_{ijk} + e_{ijk} \quad (1)$$

where Y_{ijk} is the ith score in the jk-th treatment cell,

μ is the mean of the treatment populations and is a constant for the Y_{ijk} observations,

η_i is a constant associated with observation i,

α_j is the effect of treatment j, which is a constant for all i within treatment j,

β_k is the effect of treatment k, which is a constant for all i within treatment k,

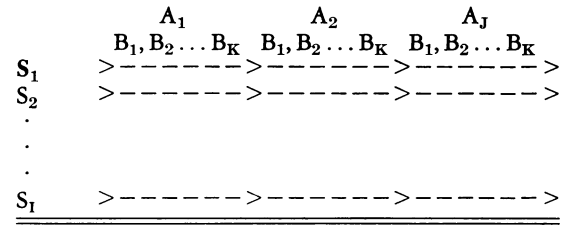
$\eta\alpha_{ij}$ is the nonadditive effect of the η_i and α_j treatments,

$\eta\beta_{ik}$ is the nonadditive effect of the η_i and β_k treatments,

$\alpha\beta_{jk}$ is the nonadditive effect of the α_j and β_k treatments,

$\eta\alpha\beta_{ijk}$ is the nonadditive effect of the η_i, α_j and β_k treatments, and

e_{ijk} is the error associated with the linear model, which is assumed to be *independently* normally distributed with mean equal to zero and variance equal to σ_e^2.



S = subject factor
A = repeated treatment factor
B = repeated trials factor

Fig. 1. Schematic representation of Klein's (1971) experimental layout.

[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.

The operant experimental hypothesis as expressed by Gentile *et al.* (1972, p. 196) is a specific *a priori* question, and therefore a planned linear comparison of the baseline and intervention conditions would provide a more powerful test than the omnibus F test on the repeated treatment factor. A simple t statistic as given in (2) would be appropriate.

$$t = \frac{\Psi}{\sqrt{MS_e[\Sigma c_{jk}^2/n_{jk}]}}, \qquad (2)$$

where Ψ, is a sample linear comparison of the repeated factor treatment means,

MS_e is the appropriate error term from the ANOVA summary table (*e.g.*, either MS_{IJ} or $MS_{e(pooled)}$)

c_{jk}'s are the weights associated with the linear comparison such that $\Sigma c_{jk} = 0.0$

n_{jk} specifies the numbers of observations for each sample mean involved in the linear comparison, and

t is a random variable distributed as student's t with ν_2 (error) degrees of freedom.

If the experimenter has formulated C *a priori* hypotheses, then each can be tested in the planned comparison sense and the probability of a Type I error can be controlled by adopting the Bonferroni or Dunn procedure (Miller, 1966, p. 67; Kirk, 1968, p. 79). The rationale of the Bonferroni procedure is to set a Type I probability risk for each linear comparison such that the sum of the risks will not exceed an overall Type I error probability stipulated *a priori* by the experimenter. Dunn (1961) tabled critical values when alpha is evenly divided among the C comparisons (α/C).

In addition to testing specific *a priori* comparisons, the experimenter also has the option of exploring additional global hypotheses with the omnibus ANOVA statistic. For model (1), if the experimenter can assume that the effects due to the three-way interaction are zero, then the three-way interaction can be used as an estimate of error variability (σ_e^2) to test the subjects, treatments $\times$ subjects, and trials $\times$ subjects effects. Given that the three-way interaction is assumed to be zero, then a pooled estimate of σ_e^2 can be derived (Kirk, 1968, pp. 214-217) to test the J, K, and J $\times$ K effects. Legitimate tests of the J, K, and J $\times$ K effects are also available without assuming the effect of the three-way interaction to be zero. (See Table 1).

The F-ratios in Table 1 are easily obtained by following the procedures for deriving expected mean squares E(MS) and forming F ratios enumerated in experimental design texts (Myers, 1972, p. 198; Winer, 1972, p. 371). The test statistic (2) and those in Table 1 are distributed as t and F variables under the assumptions that (1) the observations have been randomly sampled from jk normal populations with (2) homogeneous variances and covariances and (3) J and K represent fixed effects treatment factors.[2]

Any of the effects associated with the omnibus F tests can be investigated further with *post hoc* comparison procedures such as the Scheffé statistic (Scheffé, 1959).[3]

DISCUSSION

Analysis-of-variance and individual linear comparison procedures can be used by operant researchers to explore their data. The statistical tests presented here are appropriate when repeated measurements are taken on the experimental subjects. A crucial distinction between the statistical procedures enumerated in this article and those offered by Gentile *et al.* (1972) and by Shine and Bower (1971) are the error terms used in testing for the presence of treatment effects. Repeated measures statistics partition the error sources of variability, thereby generally permitting efficient and powerful tests. Therefore, disregarding the rationales offered by Gentile *et al.* and by Shine and Bower for the

[2]Procedures for testing the variance-covariance assumptions and consequent statistical adjustments can be found in Kirk, 1968, pp. 256-263; Winer, 1972, pp. 552-524.

[3]An excellent summary of multiple comparison procedures is given by Games (1971).

Table 1

Sources of variation, degrees of freedom, expected mean squares, and F-ratios for Model (1).

Source	*df*	*E(MS)*	*F*	*F'*
Between subjects	*I-1*			
Subject (I)	I-1	$\sigma_e^2 + JK\sigma_i^2$		MS_I/MS_e
Within subjects	*I(JK-1)*			
Treatments (J)	J-1	$\sigma_e^2 + IK\sigma_j^2 + K\sigma_{ij}^2$		MS_{IJ}/MS_e
J × I	(J-1) (I-1)	$\sigma_e^2 + K\sigma_{ij}^2$	MS_K/MS_{IK}	$MS_K/MS_{e(pooled)}$
Trials (K)	K-1	$\sigma_e^2 + IJ\sigma_k^2 + J\sigma_{ik}^2$		MS_{IK}/MS_e
K × I	(K-1) (I-1)	$\sigma_e^2 + j\sigma_{ik}^2$	MS_{JK}/MS_{IJK}	$MS_{JK}/MS_{e(pooled)}$
K × J	(K-1) (J-1)	$\sigma_e^2 + I\sigma_{jk}^2 + \sigma_{ijk}^2$		
K × J × I	(K-1) (J-1) (I-1)	$\sigma_e^2 + \sigma_{ijk}^2$		
Total	IJK-1			

F = omnibus test assuming $\sigma_{IJK}^2 \neq 0.0$
F' = omnibus tests assuming $\sigma_{IJK}^2 = 0.0$

statistics they enumerate, individual comparison and/or omnibus ANOVA F repeated measures tests are procedures that are not only *bona fide* as statistically appropriate, but, are most sensitive for locating treatment effects.

REFERENCES

Dunn, O. J. Multiple comparisons among means. *Journal of the American Statistical Association,* 1961, **56,** 52-64.

Games, P. A. Multiple comparisons of means. *American Educational Research Journal,* 1971, **8,** 531-565.

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.

Kirk, R. E. *Experimental design: procedures for the behavioral sciences.* California: Brooks/Cole Publishing Company, 1968.

Miller, R. G. Jr. *Simultaneous statistical inference.* New York: McGraw-Hill, 1966.

Myers, J. L. *Fundamentals of experimental design.* 2d ed.; Boston: Allyn Bacon, 1972.

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

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.

Winer, B. J. *Statistical principles in experimental design.* 2d ed.; Toronto: McGraw-Hill, 1971.

Received 29 May 1973.
(Published without revision.)