Journal of Educational Statistics
Summer 1980, Volume 5, Number 2, pp. 169-176

ON "THE ANALYSIS OF RANKED DATA DERIVED FROM COMPLETELY RANDOMIZED FACTORIAL DESIGNS"

Larry E. Toothaker and Horng-shing Chang

University of Oklahoma

Key Words: Kruskal-Wallis, Factorial designs, Conservative, Inter-related, Monte-Carlo

ABSTRACT

Extensions of the Kruskal-Wallis procedure for a factorial design are reviewed and researched under various degrees and kinds of nonnullity. It was found that the distributions of these test statistics are a function of effects other than those being tested except under the completely null situation and their use is discouraged.

INTRODUCTION

The need to have satisfactory rank-sum methods for factorial designs has historically been one of the main detractions of the area of nonparametric statistics. Gaito (1959), Gardner (1975), and Siegel (1956) are illustrative of the comment from the applied researcher that suitable tests for main effects and interactions were needed to complete the package of those methods not making the normality assumption. However, few statistics textbooks offer the applied researcher rank-sum methods for two-factor designs with more than one observation per cell. If any tests are given for factorial designs (see Bradley,1968) they are less than suitable in that these methods will be low in efficiency, and no unified approach is given for all tests. Typically, the Kruskal-Wallis (or Friedman for matched data) tests are done on various sums or differences of the observations; that is, the researcher collapses the data over one main effect to test for the other main effect and then applies the Kruskal-Wallis (or Friedman) test to the resulting sum. Various differences, the number of which depends on the design, are taken to measure interaction, and the appropriate test is done on these differences. The obvious loss in power efficiency from not considering variability due to the other effects gives the major deficiency of these methods. The lack of a common procedure for each test in each design also detracts from the usefullness of these approaches.

Scheirer, Ray, and Hare (1976) also mention this problem, citing several sources which present solutions, including Mehra and Sen (1969) and Patel and Hoel (1973) who give tests on interaction effects from various models. The contribution by Mehra and Sen of a test for interaction is part of the area of "aligned ranks" which proceeded from a suggestion by Hodges and Lehmann (1962) and resulted in tests for main effects in an additive model (Mehra & Sarangi, 1967; Sen, 1968), a test for interaction in a completely randomized factorial design (McSweeney, 1967), and tests for interaction in mixed models or split-plot designs (Koch, 1969; Koch & Sen, 1968). Marascuilo and McSweeney (1977) and Puri and Sen (1971) are texts which cover aligned ranks methods. An additional rank procedure, which is a promising competitor to the aligned rank methods, is the rank transform procedure of using the usual parametric analysis of variance on the ranks (see Iman, 1974; Iman & Conover, 1976). Also, Scheirer et al.(1976) present another general procedure for factorial designs which is an extension of the Kruskal-Wallis (1952) procedure. The Scheirer et al. method allows for tests of main effects, interaction and linear contrasts.

The purpose of this paper is to review the Scheirer et al. method, and to show that the distributions of the test statistics for such main effects tests and interaction tests from a factorial design are dependent on the values of effects other than those being tested. Using Monte-Carlo methods, the tests due to Scheirer et al. are examined in the presence of nonnull main effects and interactions.

REVIEW

A two factor design will be used where there are J levels of one factor, K levels of the second factor, and I observations per cell, I > 1. Only the fixed model will be considered and is given as

$$X_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + e_{ijk}, \tag{2.1}$$

where $\Sigma\alpha_j = \Sigma\beta_k = \sum_j \alpha\beta_{jk} = \sum_k \alpha\beta_{jk} = 0$, and e_{ijk} is the random variable error component for each individual in each cell. Note the e_{ijk} are independently distributed with mean zero and variance σ_e^2 for each cell.

Scheirer et al. suggest using the usual ANOVA numerator sums of squares on the ranks (called RSS) of the raw observations, resulting in the H statistic for the A main effect of

$$H_A = RSS_A(12/N(N+1)) \tag{2.2}$$

for the B main effect of

$$H_B = RSS_B(12/N(N+1)), \tag{2.3}$$

and for the AB interaction

$$H_{AB} = RSS_{AB}(12/N(N+1)). \tag{2.4}$$

H_A, H_B, and H_{AB} are asymptotically distributed as chi-square with df of, respectively, (J-1), (K-1), and (J-1)(K-1), under the assumption that all A, B, and AB effects are zero. Scheirer et al. gave the results from a small Monte-Carlo study which showed adequacy of fit of the chi-square distribution for a main effects test and individual comparisons tests, even for small sample sizes. However, only the null case $X_{ijk}=\mu+e_{ijk}$ was examined. That is, the statistic was examined in the situation where all $\alpha_j = 0$, all $\beta_k = 0$, and all $\alpha\beta_{jk} = 0$. Work done by Reinach (1965) for a replication model where ranking is done for each replicate indicates that the test for any effect becomes more conservative when the magnitude of any other effect is increased. That is, the distributions of rank-sum tests of the type proposed by Scheirer et al. may depend on the values of effects other than those being tested and may be free from those effects only if all other effects are zero.

METHOD

A computer program was written to perform simulations of the Scheirer et al. statistics and the ANOVA, for comparison purposes, for a 3 x 2 factorial design with five observations per cell. Data was generated for a normal population using the method due to Box and Muller (1958), which transforms independent unit uniform psuedo-random numbers from a procedure due to Chen (1971). For the interaction effect, one null and two nonnull cases were examined, where the interaction effects in the nonnull cases were chosen so as to give power of approximately .60 and .90 for the ANOVA F-test when $\alpha = .05$. All main effects and interaction effects are given in Table I. Factorial combination of these cases gave 12 sets of means for which the empirical probability of rejecting H_o was recorded for one main effect test and the interaction test for 1,000 replications. The .10, .05, .025, and .01 levels of significance were used. Finally, all cases were examined for the same 1,000 replications so as to facilitate comparison of rejection rates for different combinations of cases.

TABLE I

Treatment and Interaction Effects for a 2X3 ANOVA

Case	α_j	β_k	$\alpha\beta_{jk}$
1	null (zero)	null	null
2	null	null	-.5887 0 .5887 .5887 0 -.5887
3	null	null	-.8327 0 .8327 .8327 0 -.8327
4	null	-.7539 .7539	null
5	null	-.7539 .7539	-.5887 0 .5887 .5887 0 -.5887
6	null	-.7539 .7539	-.8327 0 .8327 .8327 0 -.8327
7	-.8327 0 .8327	null	null
8	-.8327 0 .8327	null	-.5887 0 .5887 .5887 0 -.5887
9	-.8327 0 .8327	null	-.8327 0 .8327 .8327 0 -.8327
10	-.8327 0 .8327	-.7539 .7539	null
11	-.8327 0 .8327	-.7539 .7539	-.5887 0 .5887 .5887 0 -.5887
12	-.8327 0 .8327	-.7539 .7539	-.8327 0 .8327 .8327 0 -.8327

RESULTS

From the empirical probabilities reported in Table II, it is clear that only for the case where all effects are zero are the Scheirer et al. tests adequate in their fit to the independent chi-square distributions. For example, if α = .05 for

TABLE II

Empirical α and Power

Case	Test Statistic	Theoretical α .10	.05	.025	.01
1- 6	FM	.092	.037	.020	.011
1	HM	.090	.036*	.013*	.006
2	HM	.045*	.017*	.008*	.003*
3	HM	.023*	.006*	.003*	0 *
4	HM	.021*	.006*	.001*	0 *
5	HM	.013*	.004*	0 *	0 *
6	HM	.007*	0 *	0 *	0 *
7-12	FM	.949	.910	.823	.699
7	HM	.941	.881	.793	.640
8	HM	.892	.774	.658	.451
9	HM	.816	.661	.499	.290
10	HM	.805	.640	.469	.263
11	HM	.678	.502	.320	.175
12	HM	.566	.381	.222	.107
1,4,7,10	FI	.115	.053	.032	.012
1	HI	.106	.049	.024	.008
7	HI	.033*	.016*	.005*	0 *
4	HI	.024*	.011*	.002*	.001*
10	HI	.007*	.002*	0 *	0 *
2,5,8,11	FI	.727	.593	.472	.334
2	HI	.699	.560	.426	.290
8	HI	.455	.315	.194	.072
5	HI	.459	.287	.170	.078
11	HI	.245	.117	.045	.015
3,6,9,12	FI	.942	.894	.821	.690
3	HI	.935	.868	.782	.628
9	HI	.801	.658	.491	.313
6	HI	.787	.645	.493	.292
12	HI	.574	.399	.254	.108

Note: for Scheirer et al. tests, HM=Main Effect, HI=Interaction; for F-tests, FM and FI, respectively.
* More than $2\sigma_p$ from theoretical α ($\sigma_p = \sqrt{\alpha(1-\alpha)/1000}$)

the null case $X_{ijk} = \mu + e_{ijk}$, $\alpha_{HM} = .036$, and $\alpha_{HI} = .049$, but for nonnull effects for the other main effect, $\alpha_{HM} = .006$ and $\alpha_{HI} = .011$. For nonnull interaction effects only, $\alpha_{HM} = .017$ and .006 (low and high interaction power), while for nonnull effects for the other main effect plus nonnull interaction effects, $\alpha_{HM} = .004$ and zero (low and high interaction power). The effect on the power of the tests is equally devastating; for example, the power of HM is .381 when the power of the main effect F is .910 (high interaction power, nonnull case for the other main effect). Clearly, for the normal case, the Scheirer et al. tests become more conservative as a function of the magnitude of the other effects.

CONCLUSIONS

Although rank-sum procedures for factorial designs are attractive methodologies to include in a list of useful statistics, the tests due to Scheirer et al. neither control α at the set value nor are they powerful in the presence of effects other than those being tested. Hence, the researcher desiring rank-sum procedures for a factorial design would be wise to consider the better-known aligned rank methods, if the model for which the tests are proposed indeed fits the researchers data (e.g., additive model, model including replication effects etc.) or the rank transform. Alternatively, the researcher can rely upon the well-known robustness of the ANOVA to nonnormality and proceed with parametric analysis of the data. Under no circumstances could the tests due to Scheirer et al. be recommended for use.

REFERENCES

Box, G. E. P., & Muller, M. E. A note on the generation of random normal deviates. The Annals of Mathematical Statistics, 1958, 29, 610-611.

Bradley, J. V. Distribution-free statistical tests. Englewood Cliffs, N.J.; Prentice-Hall, 1968.

Chen, E. H. Random normal number generator for 32-bit-word computers. Journal of the American Statistical Association, 1971, 66, 400-403.

Gaito, J. Non-parametric methods in psychological research. Psychological Reports, 1959, 5, 115-125.

Gardner, P. L. Scales and statistics. Review of Educational Research, 1975, 45, 43-57.

Hodges, J. L., & Lehmann, E. L. Rank methods for combination of independent experiments in analysis of variance. The Annals of Mathematical Statistics, 1962, 33, 482-497.

Iman, R. L. A power study of a rank transform for the two-way classification model when interaction may be present. The Candadian Journal of Statistics, Section C: Applications, 1974, 2, 227-239.

Iman, R. L., & Conover, W. J. A comparison of several rank tests for the two-way layout, (SAND76-0631). Albuquerque, New Mexico; Sandia Laboratories, 1976.

Koch, G. G. Some aspects of the statistical analysis of "split-plot" experiments in completely randomized layouts. Journal of the American Statistical Association, 1969, 64, 485-505.

Koch, G. A., & Sen, P. K. Some aspects of the statistical analysis of the "mixed model." Biometrics, 1968, 24, 27-48.

Kruskal, W. H., & Wallis, W. A. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 1952, 47, 583-621.

Marascuilo, L. A., & McSweeney, M. Nonparametric and distribution-free methods for the social sciences, Monterey, Calif: Brooks-Cole, 1977.

McSweeney, M. An empirical study of two proposed nonparametric tests for main effects and interaction. Unpublished doctoral dissertation, University of California, Berkeley, 1967.

Mehra, K. L., & Sarangi, J. Asymptotic efficiency of certain rank tests for comparative experiments. The Annals of Mathematical Statistics, 1967, 38, 90-107.

Mehra, K. L., & Sen, P. K. On a class of conditionally distribution-free tests for interactions in factorial experiments. The Annals of Mathematical Statistics, 1969, 40, 658-664.

Patel, K. M., & Hoel, D. G. A non-parametric test for interaction in factorial experiments. Journal of the American Statistical Association, 1973, 68, 615-620.

Puri, M. L., & Sen, P. K. Nonparametric methods in multivariate analysis. New York: John Wiley & Sons, 1971.

Reinach, S. G. A nonparametric analysis for a multiway classification with one element per cell. South African Journal of Agricultural Science, 1965, 8, 941-960.

Scheirer, C. J., Ray, W. S., & Hare, N. The analysis of ranked data derived from completely randomized factorial designs. Biometrics, 1976, 32, 429-434.

Sen, P. K. On a class of aligned rank order tests in two way layouts. The Annals of Mathematical Statistics, 1968, 39, 1,115-1,124.

Siegel, S. Nonparametric statistics for the behavioural sciences. New York: McGraw-Hill, 1956.

AUTHORS

TOOTHAKER, LARRY E. Address: The University of Oklahoma, Department of Psychology, 455 West Lindsey, Room 705, Norman, OK 73019. Title: Professor of Psychology. Degrees: B.S. in Ed., University of Nebraska; M.S., Ph.D., University of Wisconsin, Madison. Specialization: Educational statistics, experimental design.

CHANG, HORNG-SHING. Address: The University of Oklahoma, Department of Psychology, 455 West Lindsey, Room 705, Norman, OK 73019. Title: Project Assistant. Degrees: B.S. in Psychology, National Taiwan University, M.A. in Psychology, University of South Carolina. Specialization: Applied Statistics. Theories of mental test scores.