Biometrika (1991), 78, 3, *pp.* 697-701
Printed in Great Britain

A note on the rank transform for interactions

BY G. L. THOMPSON
Department of Statistical Science, Southern Methodist University, Dallas, Texas 75275, U.S.A.

SUMMARY

The asymptotic properties of the rank transform statistic for testing for interaction in a balanced two-way classification are studied. Necessary and sufficient conditions are obtained for the asymptotic distribution of this rank transform statistic to be chi-squared under the null hypothesis of no interaction. It is shown that the rank transform test statistic for interaction is asymptotically chi-squared, divided by its degrees of freedom, when there are exactly two levels of both main effects or when there is only one main effect. In the latter case, the test detects a nested effect instead of an interaction. In all other two-way layouts, there exist values for the main effects such that, under the null hypothesis of no interaction, the expected value of the rank transform test statistic goes to infinity as the sample size increases.

Some key words: Asymptotic distribution; Interaction; Rank transform.

1. INTRODUCTION

The rank transform procedure was proposed by Conover & Iman (1981) as a bridge between nonparametrics and classical analysis of variance. All the observations are ranked together without regard to row or column membership, and classical normal theory tests are applied to the ranks, instead of to the observations. This procedure has gained much popularity because it is very easy to implement and is initially very appealing.

Confusion about the performance of the rank transform statistic for interactions in a two-way layout stems from several seemingly contradictory simulation studies. For a limited case with small sample sizes and small main effects, the simulation studies of Iman (1974) and Conover & Iman (1976) show that this rank transform statistic performs well at detecting interactions. On the other hand, simulations by Blair, Sawilowsky & Higgins (1987) show that the type I error rates are unacceptably large if either the main effects or the sample size are large. This points to a critical need to study the asymptotic properties of the rank transform test for interaction. In contrast, the asymptotic properties of rank transform statistics column effects and row effects have been widely studied by Hora & Conover (1984), Iman, Hora & Conover (1984), Kepner & Robinson (1988), Hora & Iman (1988), Thompson & Ammann (1989, 1990) and Thompson (1991). The author does not know of theoretical studies of the small sample properties of the rank transform test.

This discussion is motivated by the fact that the critical points for the rank transform test are identical to the critical points for the normal theory test from the F-distribution. The asymptotic null distribution of the normal theory test for interaction is chi-squared, divided by its degrees of freedom. For the rank transform test to behave properly, it must have the same asymptotic null distribution as the normal theory test. Even more importantly, the test must be asymptotically distribution-free over a reasonably large class of distribution functions and the asymptotic null distribution must not depend on the values of the unknown main effects. It is shown that this is the case when either there is only one main effect, or when there are exactly two levels of both main effects. Otherwise, values of the main effects exist such that the expected value of the test statistic under the null hypothesis approaches infinity as the sample size increases. As a result, it becomes grossly liberal with large type I error rates even for large sample sizes.

These theoretical results proving that a commonly used rank transform statistic has unacceptable properties are important because despite the limited asymptotic results, the complete absence of theoretical small sample results, and contradictory simulation results, the rank transform procedure has become popular with social scientists, business professionals, and other researchers in both academic and industrial fields. Easily implemented tests that detect alternatives are attractive. Further contributing to the inappropriate use of the rank transform are two widely used manuals for statistical procedures that endorse the procedure without reservation. The 1985 release of SAS (SAS, 1985, p. 647) states:

> For example, a set of data may be passed through PROC RANK to obtain the ranks for a response variable that could then be fit to an analysis-of variance model using the ANOVA or GLM procedures.

The 1987 IMSL *User's Manual Stat/Library* also suggests applying analysis of variance tests to ranked data (IMSL, 1987). These endorsements are misleading. Because ranking is a nonlinear transformation, the rank transform does not always behave like its normal theory counterpart. Extreme care must be taken to assure that both tests at least have the same asymptotic null distribution. In § 2 the model and the rank transform statistic are defined. In § 3 the asymptotic properties are discussed.

2. Definitions and preliminary notation

Consider the model for a two-way layout with interaction:

$$X_{ijn} = \theta + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijn} \quad (i = 1, \ldots, I;\ j = 1, \ldots, J;\ n = 1, \ldots, N),$$

where ε_{ijn} are independent random variables with an absolutely continuous distribution function $F(x)$ such that $F(0) = \frac{1}{2}$. Because the object of this paper is to show that the rank transform has an undesirable property, we consider only the balanced case. Assume that $\alpha_. = 0$, $\beta_. = 0$, $(\alpha\beta)_{.j} = 0$ and $(\alpha\beta)_{i.} = 0$ where a dot in the subscript indicates summing over that index. The null hypothesis of no interaction effect is H_0: $(\alpha\beta)_{ij} = 0$ for all i and j; the alternative is H_a: $(\alpha\beta)_{ij} \neq 0$ for some i and j. Let $F_{ij}(x) = F(x - \theta - \alpha_i - \beta_j)$ be the distribution function of X_{ijn} under the null hypothesis, and let the average distribution function be $H(x) = (IJ)^{-1} \Sigma\Sigma\, F_{ij}(x)$. Let X_{ij} denote a random variable with distribution function $F_{ij}(x)$ and denote $H(X_{ij})$ by H_{ij}. Because $H(x)$ is bounded and increasing on the support of X_{ij}, it follows that $0 < \operatorname{var}(H_{ij}) < \infty$. To avoid trivial situations, assume that the supports of X_{ij} and X_{ab} intersect for all (i, j) and (a, b).

To define the rank transform statistic, let R_{ijn} be the rank of X_{ijn} among all of the IJN observations. We restrict our attention to Wilcoxon scores and let $a_{ijn} = R_{ijn}/(IJN + 1)$. Define

$$Q = \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{I} \left\{ a_{ij.} - \frac{1}{J} a_{i..} - \frac{1}{I} a_{.j.} + \frac{1}{IJ} a_{...} \right\}^2,$$

$$D = \frac{1}{IJN - IJ} \sum_{n=1}^{N} \sum_{j=1}^{J} \sum_{i=1}^{I} (a_{ijn} - N^{-1} a_{ij.})^2.$$

Then, the statistic

$$T = \frac{Q}{(IJ - I - J + 1)D}$$

is the classical normal theory test with the scored ranks, a_{ijn}, substituted in place of the observations.

3. Asymptotic properties of T under H_0

To determine when the asymptotic null distribution of T cannot be $\chi^2_{(IJ-I-J+1)}/(IJ - I - J + 1)$, first define $\mu_{ij} = NE(H_{ij})$. When applicable, we assume throughout that the second index runs faster than the first. Thus, define the vectors $a = (a_{11.}, \ldots, a_{IJ.})'$ and $\mu = (\mu_{11}, \ldots, \mu_{IJ})'$, and let

Γ be an $IJ \times IJ$ matrix whose rows and columns are indexed by the ordered pairs (i, j) and (r, s) $(i, r = 1, \ldots, I, j, s = 1, \ldots, J)$. The $(i, j), (r, s)$th element of Γ is

$$\operatorname{cov}\left\{H_{ij} - \frac{1}{IJ}\sum_{v=1}^{J}\sum_{u=1}^{I} F_{ij}(X_{uv}),\ H_{rs} - \frac{1}{IJ}\sum_{v=1}^{J}\sum_{u=1}^{I} F_{rs}(X_{uv})\right\}.$$

Let $\gamma^2_{(i,j)}$ be the (i, j)th diagonal element of Γ. Note that $0 < \operatorname{var}(H_{ij}) < \infty$ implies $0 < \gamma_{(i,j),(r,s)} < \infty$.

LEMMA 1. *Under the null hypothesis, $N^{-\frac{1}{2}}(a - \mu)$ converges in distribution to $N_{IJ}(0, \Gamma)$; in particular, $N^{-\frac{1}{2}}(a_{ij.} - \mu_{ij})/\gamma_{(i,j)}$ converges in distribution to $N(0, 1)$.*

Proof. The univariate result follows by applying Theorem 3.3 of Thompson & Ammann (1989) to the linear rank statistic $a_{ij.}$ with Wilcoxon scores and simplifying the expression for the variance. The proof of the multivariate result is very similar to the proof of Theorem 4.2 of Thompson & Ammann (1989). □

LEMMA 2. *Under the null hypothesis D converges in probability to the nonnegative, finite constant*

$$\sigma^2 = \frac{1}{3} - \frac{1}{IJ}\sum\sum\{E(H_{ij})\}^2 = \frac{1}{IJ}\sum\sum \operatorname{var}(H_{ij}). \quad (1)$$

Proof. This proof is almost identical to the proof of Theorem 5.3 of Thompson & Ammann (1989). Note that $D \geqslant 0$ for all N implies that $\frac{1}{3} \geqslant \sigma^2 \geqslant 0$. □

THEOREM 3. *Under the null hypothesis of no interaction as $N \to \infty$, $\lim E(T)$ is finite if and only if*

(i) *$E(H_{ij} - H_{aj})$ does not depend on j for all $1 \leqslant i, a \leqslant I$ and $1 \leqslant j \leqslant J$,*
(ii) *$E(H_{ij} - H_{ib})$ does not depend on i for all $1 \leqslant i \leqslant I$ and $1 \leqslant j, b \leqslant J$.*

Proof. It follows from Lemma 2 and Slutsky's theorem that $\lim E(T)$ is finite if and only if $\lim E(Q)$ is finite. Let $\delta(i, r) = 1$ or 0 indicating $i = r$ or $i \neq r$, and define an $IJ \times IJ$ matrix A with elements

$$\delta(i, r)\delta(j, s) - \frac{1}{I}\delta(j, s) - \frac{1}{J}\delta(i, r) + \frac{1}{IJ}$$

which are indexed exactly as the elements of Γ. Then, Q is the quadratic form $N^{-1}a'Aa$. Because A does not depend on N and because the elements of Γ converge to finite values, $\operatorname{tr}(A\Gamma)$ is finite, and

$$\lim_{N\to\infty} E(Q) = \operatorname{tr}(A\Gamma) + \lim_{N\to\infty} N^{-1}e'Ae, \quad (2)$$

where $e = (e_{11}, \ldots, e_{IJ})'$ is the vector with elements $e_{ij} = E(a_{ij.})$. Therefore, $\lim E(Q)$ is finite if and only if $e'Ae = O(N)$. Note that

$$e'Ae = \sum\sum\left(e_{ij} - \frac{1}{J}e_{i.} - \frac{1}{I}e_{.j} + \frac{1}{IJ}e_{..}\right)^2 = O(N)$$

is equivalent to

$$e_{ij} - \frac{1}{J}e_{i.} - \frac{1}{I}e_{.j} + \frac{1}{IJ}e_{..} = O(N^{\frac{1}{2}})$$

for all i and j. Theorem 3.3 of Thompson & Ammann (1989) and Lemma 1.5.5.A of Serfling (1980) imply that $\lim (e_{ij} - \mu_{ij})/\gamma_{(i,j)} = 0$. Because $0 < \gamma_{(i,j)} < \infty$, both e_{ij} and μ_{ij} converge to the same limit as N increases. Therefore, $e_{ij} - J^{-1}e_{i.} - I^{-1}e_{.j} + (IJ)^{-1}e_{..}$ is $O(N^{\frac{1}{2}})$ if and only if $u_{ij} - J^{-1}\mu_{i.} - I^{-1}\mu_{.j} + (IJ)^{-1}\mu_{..}$ is $O(N^{\frac{1}{2}})$, which is equivalent to $\nu_{ij} - J^{-1}\nu_{i.} - I^{-1}\nu_{.j} + (IJ)^{-1}\nu_{..} = 0$ for all i and j where $\nu_{ij} = E(H_{ij})$. To obtain (i), subtract $\nu_{aj} - J^{-1}\nu_{a.} - I^{-1}\nu_{.j} + (IJ)^{-1}\nu_{..} = 0$ from $\nu_{ij} - J^{-1}\nu_{i.} - I^{-1}\nu_{.j} + (IJ)^{-1}\nu_{..} = 0$. This gives $\nu_{ij} - \nu_{aj} = J^{-1}(\nu_{i.} - \nu_{a.})$ which does not depend on j. The result for (ii) is obtained similarly. □

When $E(T)$ converges to infinity, not only is T not asymptotically chi-squared, but it becomes very liberal for large samples. This is consistent with the simulation results of Blair et al. (1987) in which the nominal α-levels approach 1 as sample size increases. Thus, the rank transform should never be used to detect interactions if (i) and (ii) cannot be shown to hold. Note that Theorem 3 does not depend on the linearity of the model; it holds for any absolutely continuous distributions F_{ij} that satisfy (i) and (ii). It is of interest to determine when these two conditions hold.

Clearly, (i) and (ii) hold if $F_{ij}=F_i$ or $F_{ij}=F_j$, that is, if there is only one main effect. The condition $F_{ij}=F_i$ can be interpreted either as testing the null hypothesis of no nested effect, $H_{01}:(\alpha\beta)_{ij}=0$ for all i and j, in the model

$$X_{ijn}=\theta+\alpha_i+(\alpha\beta)_{ij}+\varepsilon_{ijn},$$

or as testing the null hypothesis of no column effect, $H_{02}:\beta_j+(\alpha\beta)_{ij}=0$ for all i and j, in the two-way layout

$$X_{ijn}=\theta+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\varepsilon_{ijn}.$$

With either interpretation, T converges in distribution to $\chi^2_{(IJ-I-J+1)}/(IJ-I-J+1)$ by Lemma 1 and Lemma 2. It is interesting to compare this test for nested effects with the test proposed by Akritas (1990). In Akritas's test, all the data are ranked together and then a function of the ranks, namely the ranks divided by an estimate of the standard deviation, is substituted into the classical F test for a nested effect. In the test T, all of the data are ranked together and the ranks are substituted into a different classical F test than for a nested effect. Hence, neither test is a rank transform test, but for different reasons. By contrast, Corollary 4 shows that T is a rank transform test when $I=J=2$.

COROLLARY 4. *When both main effects are present, conditions* (*i*) *and* (ii) *are satisfied for all values of* α_i *and* β_j *if and only if* $I=J=2$.

Proof. Assume that $I=J=2$. Conditions (i) and (ii) are equivalent to

$$E(H_{11}-H_{21})-E(H_{12}-H_{22})=0$$

which can be shown to be equivalent to

$$\int\{F(x+2\alpha+2\beta)+F(x-2\alpha-2\beta)-F(x+2\alpha-2\beta)-F(x-2\alpha+2\beta)\}f(x)\,dx=0 \qquad (3)$$

by expanding $H(x)$ as a sum, changing variables in the integrals, and cancelling terms. To show that (3) always holds, we show that $\int\{F(x+\delta)+F(x-\delta)\}f(x)\,dx$ is a constant function in δ by showing that its partial derivative with respect to δ is $\int\{f(x+\delta)-f(x-\delta)\}f(x)\,dx=0$. Hence, conditions (i) and (ii) hold. Conversely, if $J\geqslant 3$, a counterexample to the condition that $E(H_{1j}-H_{2j})$ does not depend on j is generated for symmetric distributions by letting $\alpha_1=-\alpha_2$, $\beta_1=-\beta_2$ and $\beta_j=0$ for $3\leqslant j\leqslant J$. Then

$$E(H_{11}-H_{21})=E(H_{12}-H_{22})\neq E(H_{13}-H_{23}).$$

Counterexamples for $I\geqslant 3$ and for nonsymmetric distributions are handled similarly. □

For $I=J=2$, computations show that $\operatorname{tr}(A\Gamma)=\sigma^2$ and it follows from Lemma 1, Lemma 2 and the proof of Theorem 3 that T is $\chi^2_{(IJ-I-J+1)}/(IJ-I-J+1)$. Power properties of this test are simulated by S. S. Sawilowsky in an unpublished report.

A fundamental difference between T and the classical F test is illustrated by the proof of Theorem 3. In the normal theory, the null hypothesis of no interaction is a system of $IJ-I-J+1$ linearly independent equations in terms of the IJ means $E(X_{ij})$. Because ranking is a nonlinear mapping, the same system of equations in terms of $E(a_{ij.})$ instead of $E(X_{ij})$ is no longer the null hypothesis of no interaction. Beyond the 2×2 case, the rank transform may be detecting the nonlinearity of the mapping from the data to the ranks. Similar points are discussed by Blair et al. (1987).

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation. The author appreciates the helpful comments and suggestions of the referees.

REFERENCES

AKRITAS, M. G. (1990). The rank transform method in some two-factor designs. *J. Am. Statist. Assoc.* **85**, 73-8.

BLAIR, R. C., SAWILOWSKY, S. S. & HIGGINS, J. J. (1987). Limitations of the rank transform statistic. *Comm. Statist.* B **16**, 1133-45.

CONOVER, W. J. & IMAN, R. L. (1976). On some alternative procedures using ranks for the analysis of experimental designs. *Comm. Statist. A* **5**, 1349-68.

CONOVER, W. J. & IMAN, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *Am. Statistician* **35**, 124-33.

HORA, S. C. & IMAN, R. L. (1988). Asymptotic relative efficiencies of the rank-transform procedure in randomized complete block designs. *J. Am. Statist. Assoc.* **83**, 462-70.

HORA, S. C. & CONOVER, W. J. (1984). The *F*-statistic in the two-way layout with rank-score transformed data. *J. Am. Statist. Assoc.* **79**, 668-73.

IMAN, R. L. (1974). A power study of a rank transform for the two-way classification model when interaction may be present. *Can. J. Statist.* **2**, 227-39.

IMAN, R. L., HORA, S. C. & CONOVER, W. J. (1984). Comparison of asymptotically distribution-free procedures for the analysis of complete blocks. *J. Am. Statist. Assoc.* **79**, 674-85.

IMSL (1987). *IMSL User's Manual Stat/Library.* Houston, Texas: IMSL.

KEPNER, J. J. & ROBINSON, D. L. (1988). Nonparametric methods for detecting treatment effects in repeated measures designs. *J. Am. Statist. Assoc.* **83**, 456-61.

SAS INSTITUTE, INC. (1985a). *SAS User's Guide: Statistics*, 5th ed. Cary, N.C.: SAS Institute.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

THOMPSON, G. L. (1991). A unified approach to rank tests for multivariate and repeated measures designs. *J. Am. Statist. Assoc.* **86**, 410-9.

THOMPSON, G. L. & AMMANN, L. P. (1989). Efficacies of rank transform statistics in two-way models with no interaction. *J. Am. Statist. Assoc.* **84**, 325-30.

THOMPSON, G. L. & AMMANN, L. P. (1990). Efficiencies of interblock rank statistics for repeated measures designs. *J. Am. Statist. Assoc.* **85**, 519-28.

[*Received June* 1990. *Revised January* 1991]