Statistica Sinica **7**(1997), 647-667

ROBUST ANALYSIS OF ONE-WAY REPEATED MEASURES DESIGNS WITH MULTIPLE REPLICATIONS PER CELL

M. Mushfiqur Rashid and Ansuman Bagchi

Food and Drug Administration and Worcester Polytechnic Institute

Abstract: We develop a robust inference for one-way repeated measures designs with multiple replications per cell assuming exchangeability of the errors within each subject. R-estimators of the treatment effects are obtained by minimizing a dispersion function. We develop asymptotically equivalent test procedures based on drop in dispersion, and two quadratics which depend on R-estimates and the gradient vector. Multiple comparison procedures are developed based on the R-estimators. Test results based on a baseball data set concerning three different base running methods are presented and compared with normal theory and Friedman rank sum techniques. Asymptotic relative efficiencies of the rank tests, with respect to the normal-theory counterpart, are discussed. Comparisons with alternative robust tests are also discussed. A small scale simulation study is conducted to investigate the small sample behavior of the rank-based tests.

Key words and phrases: Asymptotic linearity, dispersion function, multiple comparisons, R-estimates.

1. Introduction

In this paper we propose a rank based inference for one-way repeated measures (RM) designs with multiple replications (MR) per cell. We assume that the components of the error vector for each subject are exchangeable continuous random variables and the error vectors corresponding to different subjects are independently identically distributed. We use a sum of Jaeckel (1972) type dispersion functions based on intra-subject ranks of residuals to obtain R-estimates and rank tests.

Some distribution-free test procedures for one-way RM designs with MR per cell are available based on intra-subject rankings; however, these procedures possess certain limitations. One can use Benard and Van Elteren's (1953) Friedman-type rank test for testing the equality of the treatment effects when the number of subjects is large. One can also use Mack and Skillings's (1980) Friedman-type rank test when the number of observations per cell is large. Brunner and Dette (1992) developed procedures for the two factor mixed model with unequal cell frequencies using intra-block ranks. Their statistic for testing the equality of the treatment effects, when there are no interactions involved in the model and the cell frequencies are equal, reduces to the Benard-Van Elteren statistic; however, none of the aforementioned procedures produce R-estimators analogous to Rashid, Aubuchon and Bagchi (1993) and Rashid (1995). As a result, multiple comparisons cannot be carried out based on sensible estimators.

Hochberg and Tamhane (1987), p.213 discussed parametric analysis for one-way RM designs with MR per cell. See also Hocking (1973) and Scheffe (1956) for further discussions. However, the parametric analysis assumes the multi-normality of each error vector with equicorrelated covariance matrix. In many situations the multi-normality assumption of the error vectors may not be valid.

The purpose of this paper is to develop robust inference for one-way RM designs with MR per cell. We will define a rank based dispersion function for each subject using Jaeckel's (1972) rank based dispersion function. For each subject, the dispersion function will be based on the residuals and the intra-subject ranks of the residuals. Then we will define an overall dispersion function for the proposed design. Analogous to the sum of squared errors (SSE) of the linear model theory, this dispersion function will play a key role for rank based inference for one way RM designs with MR per cell. In Section 2, we discuss the model with assumptions. In Section 3, we will obtain rank estimators of the treatment effects by minimizing the dispersion function. The asymptotic distributions of the rank estimators are developed under the assumption that n, the number of subjects, is large. In Section 4, three different test statistics using drop in dispersion, gradient of the dispersion function and R-estimators will be developed for testing $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$, where p is the number of treatments (fixed) and α_i is the effect of the ith treatment. In Section 5, we develop multiple comparison procedures using the R-estimators. In Section 6, a consistent estimator of τ, a scale parameter, is obtained. An illustration of the aforementioned methodology is presented in Section 7. Section 8 contains the ARE of the proposed rank tests with respect to the normal theory counterpart. Section 9 compares the rank tests with other nonparametric competitors. A report of a small scale simulation study is given in Section 10. Concluding remarks are given in Section 11.

2. The Model

Consider a one-way RM designs in which there are n subjects and p levels of a repeated measures factor. Further, each level of the repeated measures factor is applied m times to each subject. We assume that there is a washout period between administering any two treatments, i.e., there are are no carry-over effects. Let Y_{ijk} be the response of the jth subject corresponding to the ith treatment on the kth occasion. A model for this kind of experiment can be written in the form

$$Y_{ijk} = \mu + \alpha_i + \epsilon_{ijk}, \quad i = 1, \ldots, p; \; j = 1, \ldots, n; \quad k = 1, \ldots, m, \tag{2.1}$$

where μ is the overall mean, α_i is the effect of the ith level of the repeated measure factor such that $\sum_{i=1}^{p}\alpha_i = 0$, and ϵ_{ijk} is the random error term. Let

$$\underline{Y}_j = (Y_{1j1}, \ldots, Y_{1jm}, Y_{2j1}, \ldots, Y_{2jm}, \ldots, Y_{pj1}, \ldots, Y_{pjm})'$$

be the observation vector corresponding to jth subject and

$$\underline{\epsilon}_j = (\epsilon_{1j1}, \ldots, \epsilon_{1jm}, \epsilon_{2j1}, \ldots, \epsilon_{2jm}, \ldots, \epsilon_{pj1}, \ldots, \epsilon_{pjm})' \tag{2.2}$$

be the corresponding error vector.

In parametric inference, one assumes that $\underline{\epsilon}_j$'s are independently multinormally distributed with mean $\underline{0}$ vector and covariance matrix

$$\sigma^2[(1-\rho)I_{pm\times pm} + \rho\underline{1}_{pm\times 1}\underline{1}'_{pm\times 1}], \tag{2.3}$$

where σ^2 is the variance of each component and ρ $(-\frac{1}{mp-1} < \rho < 1)$ is the equicorrelation coefficient between any two components within the jth subject, $I_{p\times p}$ is an identity matrix of order p and $\underline{1}_p$ is a $p \times 1$ vector of unity. Model (2.1) with assumption (2.3) will be called a *parametric repeated measures model* in this paper.

Model (2.1) under the assumption (2.3) is considered by Hochberg and Tamhane (1987). For $m \geq 2$, one uses the F-statistic

$$\{n\sum_{i=1}^{p}(\bar{Y}_{i..}-\bar{Y}_{...})^2/(p-1)\}/\Big[\{\sum_{j=1}^{n}\sum_{i=1}^{p}(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})^2\}/\{(n-1)(p-1)\}\Big] \tag{2.4}$$

which follows an F distribution with $p-1$ and $(p-1)(n-1)$ degrees of freedom under the equality of the treatment effects. However, as mentioned earlier in many instances the multinormality assumption is unreasonable, and in such cases one would prefer to use a distribution-free procedure or an asymptotically distribution-free procedure. In this paper, we assume that the following assumptions hold for model (2.1):

(B.1) The elements of $\underline{\epsilon}_j$ $(j = 1, \ldots, n)$ are exchangeable random variables with c.d.f. $F(\epsilon_{1j1}, \ldots, \epsilon_{1jm}, \ldots, \epsilon_{pj1}, \ldots, \epsilon_{pjm})$, and $\underline{\epsilon}_j$'s are i.i.d. and continuous random vectors;

(B.2) the bivariate density $f(.,.)$ of any two components of $\underline{\epsilon}_j$'s corresponding to the joint c.d.f. $F(., \ldots, .)$ in (B.1) is continuous in R^2;

(B.3) $\int_{-\infty}^{\infty} f(\epsilon, \epsilon)\, d\epsilon < \infty$.

Model (2.1) with assumptions (B.1) - (B.3) will be called a *non-parametric repeated measures model* throughout the paper. Note that $\int_{-\infty}^{\infty} f(\epsilon, \epsilon)\, d\epsilon$ is the probability density function of $\epsilon_{ijk} - \epsilon_{i'jl}$ $(i \neq i'$ or $k \neq l)$ at the median (0). Further, note that if the components of the $\underline{\epsilon}_j$ are independent, the assumption (B.3) implies that the marginal distribution of the ϵ_{ijk} has the finite Fisher information (see Hettmansperger (1984)).

It is well known that the normal theory repeated measures model can be analyzed by a mixed model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk},\ \beta_j \sim NI[0, \rho\sigma^2],\ e_{ijk} \sim NI[0, (1-\rho)\sigma^2],\ \rho > 0, \quad (2.5)$$

where NI stands for normally and independently distributed, β_j (jth subject effect, $j = 1, \ldots, n$) and e_{ijk} ($i = 1, \ldots, p;\ j = 1, \ldots, n;\ k = 1, \ldots, m$) are independently distributed. However, our non-parametric repeated measures model cannot be written in mixed model form since the sum of β_j and e_{ijk} may not have the same distribution as Y_{ijk}. For example, the student's t variable does not have the reproductive property. As a result we will not be able to use a model of the form (2.5) and define a dispersion function (Jaeckel (1972)) of the design based on joint ranks of the entire residual vector $(Y_{ijk} - \mu - \alpha_i - \beta_j, i = 1, \ldots, p; j = 1, \ldots, n; k = 1, \ldots, m)$. So, even though in parametric inference a repeated measures model is analyzed by a mixed model (in fact indirectly analyzed by a two factor fixed effects model with i.i.d. errors), the two factor fixed effects model (with i.i.d. errors) approach is not applicable to the non-parametric repeated measures model. Therefore, the results for the linear model with i.i.d. errors based on the joint rankings (Hettmansperger (1984), Chapter 5) are not applicable to the non-parametric repeated measures model. Since, in the non-parametric repeated measures model, the elements of each error vector $\underline{\epsilon}_j$ are exchangeable random variables, we will rank the residuals within each subject separately. We will show that our results have higher efficiency than the parametric counterpart for heavy-tailed error distributions.

3. R-Estimators and Their Asymptotic Distributions

In this section we develop rank estimators using a rank (intra-subject) based dispersion function proposed by Jaeckel (1972). The dispersion function for the jth subject using Wilcoxon scores is

$$D_j(\underline{\alpha}) = \{\sqrt{12}/(mp+1)\} \sum_{i=1}^{p} \sum_{k=1}^{m} [R(Y_{ijk} - \alpha_i) - (mp+1)/2]\,[Y_{ijk} - \alpha_i], \quad (3.1)$$

where $R(Y_{ijk} - \alpha_i)$ stands for the intra-subject rank of the residual, $Y_{ijk} - \alpha_i$, corresponding to the kth replication of the ith treatment on the jth subject. By Theorem 1 of Jaeckel (1972), $D_j(\underline{\alpha})$ is a non-negative, continuous, location free and convex function of $\underline{\alpha}$. The combined dispersion function is given by

$$D(\underline{\alpha}) = \{\sqrt{12}/(mp+1)\} \sum_{j=1}^{n} \sum_{i=1}^{p} \sum_{k=1}^{m} [R(Y_{ijk} - \alpha_i) - (mp+1)/2]\,[Y_{ijk} - \alpha_i]. \quad (3.2)$$

One can use the Nelder and Mead algorithm (see Olsson (1974)) to minimize $D(\underline{\alpha})$.

Combined dispersion functions (a sum of Jaeckel-type dispersion functions) similar to (3.2) have been used in the literature to make inferences for repeated measures designs with single or no observations per cell. See, for example, Rashid, Aubuchon and Bagchi (1993), Rashid (1995), and Rashid and Nandram (1995).

Note that $D(\underline{\alpha})$ is also non-negative, continuous, location free and convex in $\underline{\alpha}$. Hence a rank estimate $\hat{\underline{\alpha}}$ of $\underline{\alpha}$ can be obtained by minimizing the dispersion function $D(\underline{\alpha})$. Let $\underline{\alpha}^0$ be the true value of $\underline{\alpha}$. In the following we develop the asymptotic distribution of $\sqrt{n}(\hat{\underline{\alpha}} - \underline{\alpha}^0)$. We need the asymptotic distribution $\sqrt{n}(\hat{\underline{\alpha}} - \underline{\alpha}^0)$ in order to develop large sample rank tests. In order to develop the asymptotic distribution theory we need a linear approximation to the gradient of $D(\underline{\alpha})$ and a quadratic approximation to $D(\underline{\alpha})$.

First we consider the linear approximation to the gradient vector. It is easy to show that negative of the vector of partial derivatives of $D(\underline{\alpha})$ at $\underline{\alpha}$ (or a negative sub-gradient $-\delta_{\underline{\alpha}} D(\underline{\alpha})$, in case $D(\underline{\alpha})$ is not differentiable at $\underline{\alpha}$) is given by the vector $S(\underline{\alpha})$ with the ith element

$$s_i(\underline{\alpha}) = \{\sqrt{12}/(mp+1)\} \sum_{j=1}^{n} \sum_{k=1}^{m} \left[R(Y_{ijk} - \alpha_i) - (mp+1)/2\right],\ i = 1, \ldots, p. \quad (3.3)$$

We set $\sum = [pI_{p\times p} - \underline{1}_p\, \underline{1}_p']$. We assume that the true value of $\underline{\alpha}$ is $\underline{\alpha}^0 = \underline{0}$. Under assumptions (B.1) - (B.3), as n becomes large, using Rashid and Bagchi (1993) it can be shown that

$$n^{-1/2} \left\{ S\left(n^{-1/2}\underline{\alpha}\right) \; - \; S\left(\underline{0}\right) \right\} \xrightarrow{\mathcal{P}} [-m^2\sqrt{n}/\{\tau(mp+1)\}] \textstyle\sum \underline{\alpha}. \quad (3.4)$$

The parameter $\tau = 1/[\sqrt{12}\int_{-\infty}^{\infty} f(\epsilon,\epsilon)d\epsilon]$ in the above expression plays a role similar to $\sigma^2(1-\rho)$ in parametric repeated measures analysis. In fact, the parameter ρ measures subject effect and if $\rho = 0$, there is none.

Using (3.4) we construct a linear approximation to the gradient of the dispersion function as follows:

$$n^{-1/2} S\left(\underline{\alpha}\right) \; = \; n^{-1/2} S(\underline{0}) - [m^2\sqrt{n}/\{\tau(mp+1)\}] \textstyle\sum \underline{\alpha} + o_p(1), \quad (3.5)$$

where $o_p(1)$ tends to zero in probability uniformly for all vectors $\underline{\alpha}$ such that $\sqrt{n}\|\underline{\alpha}\| \le c$, for any $c > 0$. The expression (3.5) can be proved assuming (3.4) and from the fact that $S(\underline{\alpha})$ has monotone components in $\underline{\alpha}$ since $D(\underline{\alpha})$ is a convex function of $\underline{\alpha}$. (See Rashid and Bagchi (1993) for details. See also a related result in Rashid, Aubuchon and Bagchi (1993) for single group repeated measures balanced incomplete block designs.) Jureckova (1969) proved a similar result based on the joint rankings for the linear model with i.i.d. errors.

Next, we consider a quadratic approximation to $D(\underline{\alpha})$. Let

$$D(\underline{\alpha}) = \{\sqrt{12}/(mp+1)\} \sum_{i=1}^{p} \sum_{j=1}^{n} \sum_{k=1}^{m} \left[R(Y_{ijk} - \alpha_i) - (mp+1)/2\right] \left[Y_{ijk} - \alpha_i\right]$$

and

$$Q(\underline{\alpha}) = D(\underline{0}) + [m^2 n/\{2\tau(mp+1)\}]\underline{\alpha}'\sum\underline{\alpha} - \underline{\alpha}'S(\underline{0}). \tag{3.6}$$

It is clear that $Q(\underline{\alpha})$ is a quadratic in $\underline{\alpha}$ and has the property that $Q(\underline{0}) = D(\underline{0})$. Also the gradient of $Q(\underline{\alpha})$ is an approximation to that of $D(\underline{\alpha})$. Further, under assumptions (B.1) - (B.3), for every $c > 0$ and every $\epsilon > 0$,

$$\lim_{n\to\infty} P_0\Big\{ \sup_{\|\sqrt{n}\underline{\alpha}\|\le c} |D(\underline{\alpha}) - Q(\underline{\alpha})| \ge \epsilon \Big\} = 0. \tag{3.7}$$

The expression in (3.7) can be proved assuming (3.5) and using arguments similar to those given in Jaeckel (1972). Then the function

$$Q(\underline{\alpha}) = D(\underline{0}) + [nm^2/\{2\tau(mp+1)\}]\underline{\alpha}'\sum\underline{\alpha} - \underline{\alpha}'S(\underline{0})$$

is a quadratic approximation to $D(\underline{\alpha})$, satisfying

$$D(\underline{\alpha}) \;=\; Q(\underline{\alpha}) + o_p(1) \tag{3.8}$$

as implied by (3.7). Hence, minima of $D(\underline{\alpha})$ and $Q(\underline{\alpha})$ coincide asymptotically.

It is shown in Rashid and Bagchi (1993) that the statements

$$\text{(3.4), (3.5) and (3.7) are equivalent.} \tag{3.9}$$

(See Heiler and Willers (1988), and Rashid, Aubuchon and Bagchi (1993) for similar conclusions, respectively, for the linear model with i.i.d. errors and repeated measures balanced incomplete block designs.)

Finally, suppose $\underline{\widetilde{\alpha}}$ minimizes $Q(\underline{\alpha})$. Then

$$\underline{\widetilde{\alpha}} = [\tau(mp+1)/(nm^2)]\ \sum\nolimits^- S(\underline{0}), \tag{3.10}$$

where $\sum^-$ is a generalized inverse of $\sum$ satisfying $\sum^-\sum\sum^- = \sum^-$. But, by a multivariate Central Limit Theorem (Mardia, Kent and Bibby (1979), p. 51)

$$n^{-1/2}[S(\underline{0})] \xrightarrow{\mathcal{D}} MVN\left(\underline{0}, \{m^2/(mp+1)\}\sum\right) \text{ as } n\to\infty. \tag{3.11}$$

Similar to Jaeckel (1972), it follows from (3.7) that

$$\sqrt{n}(\underline{\hat{\alpha}} - \underline{\tilde{\alpha}}) \xrightarrow{\mathcal{P}} \underline{0} \text{ as } n\to\infty. \tag{3.12}$$

Therefore, under assumptions (B.1) - (B.3) and the true value $\underline{\alpha}^0$,

$$\sqrt{n}(\underline{\hat{\alpha}} - \underline{\alpha}^0) \xrightarrow{\mathcal{D}} MVN\left(\underline{0},\ \{\tau^2(mp+1)/m^2\}\ \sum\nolimits^-\right) \tag{3.13}$$

as $n \to \infty$.

4. Rank Tests

In this section we develop three different statistics based on rank estimates of $\underline{\alpha}$ for testing hypothesis of the form $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$.

Assuming (3.7), (3.11), (3.12) and (3.13), and using arguments similar to those given in Rashid and Bagchi (1993) it can be shown that under $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$,

$$\begin{aligned} D^*(\tau) &= 2[D(\underline{0}) - D(\hat{\underline{\alpha}})]/\tau = 2[Q(\underline{0}) - Q(\tilde{\underline{\alpha}})]/\tau + o_p(1) \\ &= [(mp+1)/(nm^2)][S(\underline{0})]'\textstyle\sum^-[S(\underline{0})] + o_p(1). \end{aligned}$$

Let $\hat{\tau}$ be a consistent estimator of τ. Hence, by (3.11)

$$D^*(\hat{\tau}) = 2[D(\underline{0}) - D(\hat{\underline{\alpha}})]/\hat{\tau} \tag{4.1}$$

has a chi-square distribution with $p-1$ degrees of freedom as $n \to \infty$.

McKean and Hettmansperger (1976) proved (4.1) for the linear model with i.i.d. errors. Rashid, Aubuchon and Bagchi (1993) proved (4.1) for the balanced incomplete repeated measures designs with exchangeable errors within each subject. The D^* test is analogous to likelihood ratio tests.

Now, we develop a test statistic based on the gradient vector assuming that the null hypothesis $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ is true. Since by (3.11), $n^{-1/2}S(\underline{0})$ has an asymptotic multi-normal distribution, we can consider a quadratic form involving it and obtain the following statistic

$$S^* = \{(mp+1)/(nm^2)\}[S(\underline{0})]'\textstyle\sum^-[S(\underline{0})] = [12/\{npm^2(mp+1)\}]\displaystyle\sum_{i=1}^{p} R_{i..}^2 - 3n(mp+1).$$

Under assumptions (B.1) - (B.3) and H_0,

$$S^* \xrightarrow{\mathcal{D}} \chi^2_{p-1} \tag{4.2}$$

as $n \to \infty$, since the rank of $\sum$ is $p-1$. Here $R_{i..}$ is the sum of the ranks of the observations corresponding to the ith treatment. The S^* statistic can be called the generalized Friedman's statistic or Benard-Van Elteren (1953) statistic. It is also a scores test.

A third approach to testing $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ is based directly on the full model R-estimate $\hat{\underline{\alpha}}$ determined by minimizing $D(\underline{\alpha})$. Assuming (3.13) it can be shown that under assumptions (B.1) - (B.3) and $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$,

$$W^*(\hat{\tau}) = [(nm^2p)/\{\hat{\tau}^2(mp+1)\}](\hat{\underline{\alpha}})'\textstyle\sum(\hat{\underline{\alpha}}) = (nm^2p)/\{\hat{\tau}^2(mp+1)\}\sum_{i=1}^{p}(\hat{\alpha}_i - \alpha^*)^2 \tag{4.3}$$

converges to a chi-square distribution with $p-1$ degrees of freedom, where $\alpha^* = (1/p)\sum_{i=1}^{p}\hat{\alpha}_i$. Note that $W^*(\hat{\tau})$ is a Wald-type test. To make the D^* and W^*

tests operational, we need a consistent estimate of τ. We develop a consistent estimate of τ in Section 6.

5. Multiple Comparisons

In this section we develop multiple comparison procedures based on the R-estimators. First we consider a Tukey-type procedure. From (3.13), for large n, we have

$$\widehat{\mathrm{Var}}_0[\sqrt{n}(\hat{\alpha}_i - \hat{\alpha}_{i'})] = 2\hat{\tau}^2(mp+1)/(pm^2) \ \ (i \neq i'). \tag{5.1}$$

Therefore, using (5.1) and following Hochberg and Tamhane (1987), p. 83, $100(1-\gamma)\%$ confidence intervals for all contrasts $\underline{b}'\underline{\alpha}$ are given by

$$\underline{b}'\hat{\underline{\alpha}} \ \pm \ q_{\gamma,p,\infty} \ \hat{\tau} \ [(mp+1)/(npm^2)]^{1/2} \ \sum_{i=1}^{p} |b_i|/2, \tag{5.2}$$

where $q_{\gamma,p,\infty}$ is the upper (100γ)th percentile for the range of p independent $N(0,1)$ random variables, and b_i is the ith element of $\underline{b}$. Similarly, we can produce least significant difference (LSD) and Bonferroni's procedures.

Finally, we develop multiple comparisons of test treatments with a control treatment. Without any loss of generality, let us assume that treatment 1 is the control treatment. Then the R-estimates of the $p-1$ contrasts $\alpha_2 - \alpha_1, \alpha_3 - \alpha_1, \ldots, \alpha_p - \alpha_1$ will be $\hat{\alpha}_2 - \hat{\alpha}_1, \hat{\alpha}_3 - \hat{\alpha}_1, \ldots, \hat{\alpha}_p - \hat{\alpha}_1$ where $\hat{\alpha}_i$'s are obtained by minimizing $D(\underline{\alpha})$. Therefore, from (3.13), $\rho_{(\hat{\alpha}_i - \hat{\alpha}_1, \hat{\alpha}_{i'} - \hat{\alpha}_1)} = 1/2$. The estimates $\hat{\alpha}_i - \hat{\alpha}_1$ $(i = 2, 3, \ldots, p)$ along with the estimate $\hat{\tau}$ for τ can be used in Dunnett's (1955) procedure to obtain the following $100(1-\gamma)\%$ simultaneous two-sided intervals:

$$\alpha_i - \alpha_1 \ \in \ \left[\hat{\alpha}_i - \hat{\alpha}_1 \pm |M|^{(\gamma)}_{p-1,\rho} \ \{2\hat{\tau}^2(mp+1)/(npm^2)\}^{1/2}\right], \tag{5.3}$$

where $|M|^{(\gamma)}_{p-1,\rho}$ is the upper γ point of the maximum absolute values of the components of $(p-1)$ $N(0,1)$ random variables with common correlation ρ. Similarly one can produce one-sided confidence intervals.

6. Estimation of τ

In this section we consider a method of estimating τ. It has been mentioned earlier that the parameter τ plays a role similar to $\sigma^3(1–\rho)$ in parametric repeated measures. Therefore, it is important to find a consistent estimator $\hat{\tau}$ of τ.

For the one-way repeated measures with multiple replications per cell $1/[\sqrt{12}\tau]$ is the density of $d_j^{(i,i')} = Y_{ijk} - Y_{i'j\ell}$ $(i \neq i' = 1, \ldots, p;\ \ k, \ell = 1, \ldots, m)$ at $\alpha_i - \alpha_{i'}$ for each $j = 1, \ldots, n$. We can consider $Y_{i1k} - Y_{i'1\ell}, Y_{i2k} - Y_{i'2\ell}, \ldots, Y_{ink} -$

$Y_{i'n\ell}$ as a random sample from an absolutely continuous symmetric distribution with median at $\alpha_i - \alpha_{i'}$. Let $d_{(j)}^{(i,i')}$ be the ordered $d_j^{(i,i')}$ $(j = 1, \ldots n)$.

Bloch and Gastwirth (1968) developed a consistent estimate of the reciprocal of the density function at the median using the ordered observations. Using Bloch and Gastwirth (1968), we construct the following consistent estimate of τ :

$$[n/\{2\sqrt{12}q\}]\left[d^{(i,i')}_{([\frac{n}{2}]+q)} - d^{(i,i')}_{([\frac{n}{2}]-q+1)}\right] \quad (i \neq i' = 1, \ldots, p). \tag{6.1}$$

Note that the estimate does not depend on the α_i's. We can produce $m^2p(p-1)$ consistent estimates of τ using (6.1).

Furthermore, we can consider $d_1^{(k,\ell)} = Y_{i1k} - Y_{i1\ell}, d_2^{(k,\ell)} = Y_{i2k} - Y_{i2\ell}, \ldots,$ $d_n^{(k,\ell)} = Y_{ink} - Y_{in\ell},\ k \neq \ell$, as a random sample from an absolutely continuous symmetric distribution with median at 0. Therefore,

$$[n/\{2\sqrt{12}q\}]\left[d^{(k,\ell)}_{([\frac{n}{2}]+q)} - d^{(k,\ell)}_{([\frac{n}{2}]-q+1)}\right] \quad (k \neq \ell) \tag{6.2}$$

is another set of $mp(m-1)$ consistent estimators of τ.

We average all the $mp(mp-1)$ consistent estimators in (6.1) and (6.2) to get an overall consistent estimator of τ. This overall estimate of τ is a weighted estimator, and we expect that this estimate will perform better than an unweighted estimator (based on only one sample).

Following Bloch and Gastwirth (1968), one may set $q = cn^{4/5}$. In our small sample study (see Section 10) $c = .5$ was chosen to estimate τ. The $\hat{\tau}$ has performed well in standardizing the D^* and W^* tests in the sense of maintaining a selected significance level. For all practical purposes, it is recommended that c be chosen to be .5.

7. An Illustration

For illustrative purposes, we consider an example using the Woodward (1970) data presented in Table 1. Woodward (1970), a shortstop of the 1970 Cincinnati Reds NL baseball team compared three methods of base running around the first base, entitled "round out" (treatment 1), "narrow angle" (treatment 2) and "wide angle" (treatment 3), in order to determine which one requires the least amount of time to reach the second base. For a detailed description of these methods, see Hollander and Wolfe (1973), Figure 1, page 142.

Woodward considered 22 baseball players for the experiment. Each player made two runs corresponding to each method and took rest between the runs. Each entry in Table 1 represents the time taken to run from a point on the first base line 35 ft from home plate to a point 15 ft short of second base.

Table 1 . Subjects' times for two trials of each round

Subjects	Round out		Narrow angle		Wide angle	
	1	2	1	2	1	2
1	5.4	5.4	5.5	5.5	5.5	5.6
2	5.9	5.8	5.7	5.7	5.7	5.8
3	5.3	5.1	5.6	5.6	5.5	5.5
4	5.5	5.6	5.5	5.5	5.3	5.5
5	5.9	5.9	5.9	5.8	5.7	5.7
6	5.5	5.4	5.5	5.6	5.7	5.5
7	5.4	5.4	5.4	5.4	5.3	5.4
8	5.4	5.5	5.5	5.5	5.3	5.4
9	5.2	5.3	5.0	5.1	4.9	5.1
10	5.8	5.9	5.9	5.7	5.7	5.7
11	5.2	5.3	5.3	5.1	5.1	5.1
12	5.7	5.6	5.5	5.6	5.4	5.5
13	5.6	5.6	5.4	5.3	5.5	5.4
14	5.1	5.0	4.9	5.1	4.9	5.0
15	5.5	5.5	5.5	5.5	5.4	5.4
16	5.5	5.4	5.6	5.5	5.5	5.5
17	5.5	5.6	5.6	5.5	5.4	5.3
18	5.5	5.4	5.5	5.5	5.5	5.6
19	5.5	5.5	5.5	5.4	5.3	5.2
20	5.7	5.6	5.5	5.7	5.4	5.4
21	5.7	5.7	5.7	5.6	5.6	5.5
22	6.2	6.4	6.3	6.3	6.2	6.3

Following the notation of our paper, $n = 22$, $p = 3$ and $m = 2$.

The least-squares estimates and the R-estimates of the treatment contrasts $\alpha_1 - \alpha_2$, $\alpha_2 - \alpha_3$, and $\alpha_1 - \alpha_3$ are presented in Table 2.

Table 2. Estimates of contrasts

Method	$\alpha_1 - \alpha_2$	$\alpha_2 - \alpha_3$	$\alpha_1 - \alpha_3$
Rank	0.0000	0.0999	0.1000
Least-squares	0.0136	0.0705	0.0841

Based on the discussions in Section 6, a consistent estimate of τ was found to be 0.1101. Consequently, the test statistics for various rank tests are calculated:

$$D^* = 20.67 \text{ (P-value } 0+), \qquad S^* = 18.27 \text{ (P-value .0001) and}$$
$$W^* = 20.79 \text{ (P-value } 0+)$$

all of which are highly significant. The normal theory F (defined in (2.4)) statistic (using SAS) was found to be 5.87 (P-value .0056), which is also significant.

In Table 3, we present confidence intervals for the difference of treatment effects, using results from Section 5 and results from normal theory, respectively.

Table 3. 95% Joint confidence intervals for pairwise comparisons based on R-estimators and least-squares estimators

Procedure Used	Method	Comparisons		
		$\alpha_1 - \alpha_2$	$\alpha_2 - \alpha_3$	$\alpha_1 - \alpha_3$
LSD	Rank	[-0.0497, 0.0497]	[0.0503, 0.1497]	[0.0503, 0.1497]
	Least	[-0.0395, 0.0667]	[0.0174, 0.1236]	[0.0310, 0.1372]
Tukey-type	Rank	[-0.0573, 0.0573]	[0.0427, 0.1573]	[0.0427, 0.1573]
	Least	[-0.0504, 0.0776]	[0.0065, 0.1345]	[0.0201, 0.1481]

The Tukey-type and LSD procedures based on the R-estimates imply that both the round out and narrow angle methods of base running differ from the wide angle at $\gamma = 0.05$; however, the round out method and the narrow angle method do not differ from each other. We come to the same conclusions from corresponding normal theory intervals. It appears the wide angle method takes the least time.

Next, we present the pairwise comparisons results based on Friedman rank-sum. One concludes in favor of the alternative hypothesis $\alpha_i \neq \alpha_{i'}$ if

$$|R_{i\cdot\cdot} - R_{i'\cdot\cdot}| > q_{\gamma,p,\infty} \left[nm^2p(mp+1)/12\right]^{1/2}, \tag{7.1}$$

where $q_{\gamma,p,\infty}$ is the upper $(1-\alpha)$th percentile for the range of p independent $N(0,1)$ random variables. We obtain

$$|R_{1\cdot\cdot} - R_{2\cdot\cdot}| = 36.0, \quad |R_{2\cdot\cdot} - R_{3\cdot\cdot}| = 39.0, \quad |R_{1\cdot\cdot} - R_{3\cdot\cdot}| = 75.0 \tag{7.2}$$

which were compared to the upper critical value (right hand side of equation 7.1) of 41.125 for $\gamma = 0.05$. It follows, that only the round out and the wide angle methods significantly differ at a 5% level. Note that we arrived at different conclusions from that of normal theory and R-estimator intervals.

8. Asymptotic Relative Efficiency

Under assumptions (B.1) - (B.3) and for large n, the test-statistics D^* and W^* can be written as $S^*(\underline{0}) + o_p(1)$ under the null hypothesis of the equality of the treatment effects. Also under assumptions (B.1) - (B.3) and the sequence of translation alternatives $H_n : y_{ijk} = \mu + \alpha_i/\sqrt{n} + \epsilon_{ijk}$, $(i = 1,\ldots,p,\ j = 1,\ldots,n$, and $k = 1,\ldots,m, \sum_{i=1}^{p} \alpha_i = 0)$ it can be shown that

$$E[n^{-1/2}S(\underline{0})] = -[m^2/\{\tau(mp+1)\}][pI_{p\times p} - \underline{1}_p\ \underline{1}_p']\underline{\alpha}. \tag{8.1}$$

Therefore, using (8.1) and (4.2), the non-centrality parameter of all three rank tests under the translation alternatives is

$$\{m^2p/(mp+1)\}\left[\int_{-\infty}^{\infty} f(\epsilon,\epsilon)\ d\epsilon\right]^2 \sum_{i=1}^{p} \alpha_i^2. \tag{8.2}$$

Under the translation alternatives, it can be shown that the F statistic converges to a chi-square distribution with $p-1$ degrees of freedom and non-centrality parameter

$$m\sum_{i=1}^{p} \alpha_i^2/\{\sigma^2(1-\rho)\}. \tag{8.3}$$

Therefore the ARE of all the rank tests with respect to the normal theory counterpart (F) under the translation alternatives is

$$[12\sigma^2(1-\rho)mp/(mp+1)]\left[\int_{-\infty}^{\infty} f(\epsilon,\epsilon)\ d\epsilon\right]^2, \tag{8.4}$$

where ρ is the equi-correlation coefficient and σ^2 is the variance of the error term. As a result, the ARE is improved compared to the one-way repeated measures models. (i.e., when m=1).

The ARE's corresponding to exchangeable multivariate normal (N), multivariate t (with k d.f.), multivariate logistic (standard form) and multivariate exponential distributions, respectively, are given by

$$3mp/\{\pi(mp+1)\},\ \ [6mp/\{\pi(mp+1)(k-2)\}]\left\{\Gamma\left(\frac{k+1}{2}\right)/\Gamma\left(\frac{k}{2}\right)\right\}^2,$$
$$\pi^2 mp/\{8(mp+1)\},\ \text{and}\ mp(1-\rho)(3+4\rho)^2/\{3(mp+1)\}. \tag{8.5}$$

From (8.5) we see that for the multivariate normal and t distributions the AREs are independent of ρ. However, for the bivariate exponential distribution (Johnson and Kotz (1972), p. 263)

$$f(u,v) = e^{-(u+v)}[1+\theta(2e^{-u}-1)(2e^{-v}-1)],\ \ |\theta| \le 1, 0 < u, v < \infty$$

the ARE depends on ρ, (increasing function of ρ when $0 \le \rho \le \frac{1}{4}$). In Table 4, we present ARE values for some specific choices of the parameters m, p and k.

Table 4. Asymptotic relative efficiencies ($m = 2$)

$p =$	2	3	4	5	6	7	8	9	10
Mult. normal	0.76	0.82	0.85	0.87	0.88	0.89	0.90	0.9047	0.9095
Mult. t (3 d.f.)	1.95	2.08	2.16	2.21	2.24	2.27	2.29	2.30	2.32
Mult. logis.($\rho = .5$)	0.99	1.06	1.10	1.12	1.14	1.15	1.16	1.17	1.18
Mult. expon.($\rho = .2$)	3.08	3.30	3.43	3.5	3.55	3.59	3.62	3.65	3.68

Note. Multi: Multivariate; t: student's t; logis: logistic; expon: exponential.

It is seen that the rank tests developed in this paper are relatively more efficient than their normal theory counterparts for heavy-tailed symmetric distributions (t, and logistic) and skewed distributions (exponential).

9. Comparisons with Alternative Robust Tests

In this section we compare the performances of the S^*, W^*, and D^* statistics with other nonparametric tests such as Mack and Skillings (1980) test. We also discuss Jaeckel's (1972) fixed effects model approach with reference to the model considered in this article.

9.1. Mack and Skillings (M-S) test

Mack and Skillings (1980) developed distribution free tests for main effects in a two-factor ANOVA with multiple replications per cell for the model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}, i = 1, \ldots, p, j = 1, \ldots, n, k = 1, \ldots, m, \tag{9.1}$$

where $\alpha_i\,(\sum_{i=1}^{t} \alpha_i = 0)$ is the main effect of the ith level of the first factor (A), $\beta_j\,(\sum_{j=1}^{n} \beta_j = 0)$ is the main effect of the jth level of the second factor (B), m is the replications per cell and e_{ijk} are i.i.d. random errors.

To obtain a distribution free test for $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$, M-S ranked the observations from smallest to largest within each level of the factor B. Let $R_{ij.}$ denote the sum of the ranks in the ith level of factor A and jth level of factor B. It turns out that the S^* test in this article is identical to both the M-S and Benard-Van Elteren tests (See Mack and Skillings (1980), p. 949.) Both the S^* and Benard-Van Elteren tests follow a chi-square distribution as $n \to \infty$ when m and p are fixed. However, the M-S test-statistic follows a chi-square distribution with $p-1$ degrees of freedom as $m \to \infty$ when p and n are fixed. As a result S^* also follows a chi-square distribution with $p-1$ degrees of freedom when $m \to \infty$. The M-S test (under i.i.d. errors) achieves Wilcoxon efficiency when compared with its normal theory counterpart. If we assume $m \to \infty$ rather than $n \to \infty$, then under the sequence translation alternatives $H_m : \alpha_i/[m]^{\frac{1}{2}}$ the S^* test for model (2.1) under assumptions (B.1) - (B.3) achieves the following efficiency when compared with its normal theory counterpart:

$$12\sigma^2(1-\rho)\left[\int_{-\infty}^{\infty} f(\epsilon,\epsilon)\; d\epsilon\right]^2 \tag{9.2}$$

which follows from the ARE expression in (8.4). The ARE's corresponding to exchangeable multivariate normal, multivariate t (with k d.f.), multivariate logistic (standard form) and multivariate exponential distributions, respectively, are given by

$$3/\pi,\ \ [6/\{\pi(k-2)\}]\left\{\Gamma\left(k/2+.5\right)/\Gamma\left(k/2\right)\right\}^2,\ \ \pi^2/8,\ \text{and}\ (1-\rho)(3+4\rho)^2/3. \tag{9.3}$$

Note that for most distributions (multivariate logistic, normal and t) the above ARE's are independent of ρ and the number of treatments. Thus, as $m \to \infty$, the S^* test of this article has efficiency equal to that of the M-S test, and therefore it achieves Wilcoxon efficiency (when the error distributions are multivariate normal, t and logistic) when compared with the normal theory F test.

Note that under H_0 both $D^*(\tau)$ and $W^*(\tau)$ also follow a chi-square distribution with $p-1$ degrees of freedom when n is fixed and $m \to \infty$. Thus $D^*(\tau)$ and $W^*(\tau)$ achieve the efficiency in (9.3) with respect to their parametric counterparts. We are not able to estimate τ by keeping n fixed and letting $m \to \infty$ because we need a random sample to estimate τ. In a repeated measures setting, the assumption $m \to \infty$ might not be feasible. Note that $m \to \infty$ is useful if we are interested in testing equality of the levels of the second factor B (subject effect in our case) for the i.i.d. error models. For this case the M-S procedure requires rankings to be done separately within the levels of the first factor (A). In repeated measures designs, it is possible that we are not interested in subject effects.

9.2. Jaeckel's (1972) univariate approach

In this subsection we consider Jaeckel's (1972) univariate (fixed effects) approach. One might want to analyze a replicated repeated measures model by the cell means as done in parametric inference. It is well known that the parametric repeated measures model based on the cell means can be analyzed by a mixed model:

$$\bar{Y}_{ij.} = \mu+\alpha_i+\beta_j+\bar{e}_{ij.},\ \beta_j \sim NI[0,\rho\sigma^2],\ \bar{e}_{ij.} \sim NI[0,(1-\rho)\sigma^2/m],\ \rho > 0, \quad (9.4)$$

where β_j $(j = 1,\ldots,n)$ and $\bar{e}_{ij.}$ $(i = 1,\ldots,p;\ j = 1,\ldots,n)$ are independently distributed. Note that (9.4) requires $\rho > 0$, whereas for an exchangeable model $\rho \in (-1/(p-1), 1)$, provided the $\underset{\sim}{\epsilon}_j$ has a finite covariance matrix. In the absence of normality of the $\bar{Y}_{ij.}$, (9.4) also requires that the sum of β_j and $\bar{e}_{ij.}$ has the same distributional form as the $\bar{Y}_{ij.}$, which is not valid for all continuous distributions. Thus the practice of writing a repeated measures model as a mixed model is not valid under nonparametric inference. Therefore, we cannot use Jaeckel's (1972) fixed effects model approach in analyzing our model (2.1) or model (9.4) (without normality) using the average of the replicates. Further, the use of the sample average of the replicates is justified in normal theory inference because the sample average is a sufficient statistic. Note that the sample average is not a sufficient statistic for all continuous distributions.

If one applies Jaeckel's (1972) method for the fixed effects model to estimate the random effects (e.g. β_j in models 9.4 and 2.5) in mixed models then one would estimate the random variable β_j by assuming that it is fixed. Thus, it is not clear how to use the results of Hettmansperger (1984), Chapter 5 in the case of repeated measures and mixed models. Analogous to normal theory inference, one might want to apply Jaeckel's (1972) results to the fixed effects version of model (2.5) (i.e. assuming the β_j is fixed and $\mathrm{Var}(e_{ijk}) = \mathrm{Var}(Y_{ijk}) = \sigma^2$). However, in that case the dependency of the data within each subject will be ignored. For example, the denominator of the drop in dispersion test will contain a scale parameter $\tau_1 = 1/[\sqrt{12}\int_{-\infty}^{\infty} g^2(x)\,dx]$ where $g(\cdot)$ is the p.d.f. of e_{ijk}, the random error term in the fixed effects version of model (2.5). It is worth noting that the power of the tests for repeated measures models (either normal theory or nonparametric inference) must be an increasing function of ρ, which is evidenced from our simulation study. See also Jensen (1982) for discussions concerning normal theory repeated measures models. However, the power function of the tests based on Jaeckel's (1972) dispersion function for the fixed effects version of model (2.5) will not contain a parameter that will show the dependency of the data within each subject.

Martin (1988), p.265 noted, "Although Draper (1988), p.243 mentions that the rank based methods using Jaeckels's (1972) dispersion function for fixed effects models may be extended to mixed models, I wonder really how natural the methods are even for estimating random effects, let alone variance components?" Concerning parametric inference about mixed models, Hocking (1985) noted that one should formulate a statistical model directly in terms of the covariance structure of observations. Samuels, Casella and McCabe (1991) noted that many people would find this difficult, because for them a linear representation like models (2.5) and (9.4) is easier to formulate than a set of assumptions about a covariance matrix. Note that in general Jaeckel's (1972) fixed effects model approach requires $N(= pnm) \to \infty$ as $m \to \infty$ (when Huber's condition is satisfied, see Heiler and Willers (1988) for a detailed discussion) and assumes mixed models as fixed effects models.

10. Small Sample Study

A small scale simulation study (with $n = 15$ and 20) was conducted to investigate the small sample behavior of the rank-based procedures developed in this article. The distributions considered were multivariate normal and multivariate t (with 3 and 8 degrees of freedom) with scale parameter $\sigma^2 = 1$ and $\rho =$ .2, .5 and .8. The data sets were generated for $p = 2, 3, 4,$ and 5, $m = 2$ and $n = 15$ and 20. We have run 1000 simulations on each distribution. In the following we describe the results of the simulation study.

First, we see how well the rank tests achieve the nominal level .05. The empirical levels are presented in Table 5.

Table 5. The empirical levels of the rank tests at the nominal .05 level when σ=1.0, $m = 2$

$(n \quad p)$	$F(.,\ldots,.)$	Test($\rho = .2$)			Test($\rho = .5$)			Test($\rho = .8$)		
		D^*	S^*	W^*	D^*	S^*	W^*	D^*	S^*	W^*
15 2	N	.056	.058	.063	.049	.060	.059	.058	$.071^+$	$.071^+$
	$t8$	.053	.063	.059	.053	.056	.056	.041	.048	.045
	$t3$	.051	.060	.051	.047	.053	.046	.054	.059	.050
15 3	N	.054	.050	.058	.058	.055	.060	.052	.054	.056
	$t8$	.050	.048	.048	.050	.050	.059	.047	.039	.051
	$t3$	.056	$.064^+$	.050	.045	.053	.045	.046	.049	.043
15 4	N	.060	.054	$.070^+$	.056	.040	$.066^+$	.048	.043	.059
	$t8$	.048	.050	.052	.042	.041	.049	.049	.046	.052
	$t3$	.049	.053	.041	.056	.057	.057	.045	.051	.052
15 5	N	.059	.050	$.065^+$	.057	.051	$.071^+$	.051	.050	.061
	$t8$	.051	.047	.057	.049	.048	.046	.049	.045	.056
	$t3$	.050	.057	.038	.039	.042	.035	.042	.042	.028
20 2	N	.051	.039	.058	.057	.043	.063	.048	.039	.050
	$t8$	.058	.050	.055	.050	.040	.057	.051	.043	.059
	$t3$	.053	.056	.054	.051	.044	.054	.045	.043	.050
20 3	N	.056	.048	$.067^+$	.037	.040	.051	.057	.057	$.067^+$
	$t8$	.052	.062	.052	.037	.052	$.030^-$	.053	.062	.048
	$t3$	.057	.058	.056	.045	.058	.041	.041	.038	.052
20 4	N	.060	.061	.063	.057	.051	$.071^+$	.051	.050	.061
	$t8$	.057	.056	.053	.050	.048	.055	.045	.046	.052
	$t3$	.042	.049	.041	.046	.042	.044	.051	.049	.050
20 5	N	.055	.039	.061	.044	.036	$.066^+$	.062	.048	.067
	$t8$	.038	.036	.052	.055	.046	.059	.062	.056	.063
	$t3$	.047	.051	.042	.053	.063	.051	.047	.058	.047

Note. N: Multivariate normal; $t3$: Multivariate t with 3 degrees of freedom; $t8$: Multivariate t with 8 degrees of freedom; $+$: Two standard deviations above the true level; $-$: Two standard deviations below the true level.

The pluses (minuses) in Table 5 indicate empirical levels above (below) two standard deviations at the proportion 0.05. We see that only the empirical levels of the D^* test are within the two standard deviations of the nominal level 0.05. For example, when $p = 2$, $n = 15$ and $\rho = .8$, the empirical levels of S^* and W^* are outside the two standard deviations interval (.0362, .0638) with the center at the proportion 0.05. Therefore, we recommend the D^* test when $n \geq 15$, and $p \geq 2$.

Next, we investigate the empirical powers of the S^*, W^*, D^* and F statistics for testing $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ versus $H_1 : \alpha_i \neq \alpha_{i'} \quad (i \neq i')$. Under the local alternatives, the rank tests in this article have a non-central chi-square distribution with $p - 1$ degrees of freedom and non-centrality parameter

$m^2 p \sum_{j=1}^{p} \sum_{i=1}^{p} \alpha_i^2/[(mp+1)\tau^2]$. The parametric F for testing the hypothesis of no treatment effects has a non-central F- distribution with non-centrality parameter $m \sum_{i=1}^{p} \alpha_i^2/[\sigma^2(1-\rho)]$. We draw the empirical power curves as a function of distance $(\sum_{i=1}^{p} \alpha_i^2 = d)$ when the nominal level is .05 and the distribution is multivariate t with 3 ($t3$) degrees of freedom. The power curves (Figure 1) are drawn for $n = 15$, $p = 3$ and $\rho = .5$ for S^*, W^*, D^* and F when the observations are generated from a multivariate $t3$. The alternatives $(\alpha_1, \alpha_2, \alpha_3)$ are chosen to be $(-.5, 0, .5)'$, $(-.7, 0, .7)'$, $(-.65, 0, .65)'$, $(-.4, 0, .4)$, $(-.2, -.1, .3)'$, $(-.6, 0, .6)'$ and $(-.3, 0, .3)'$.

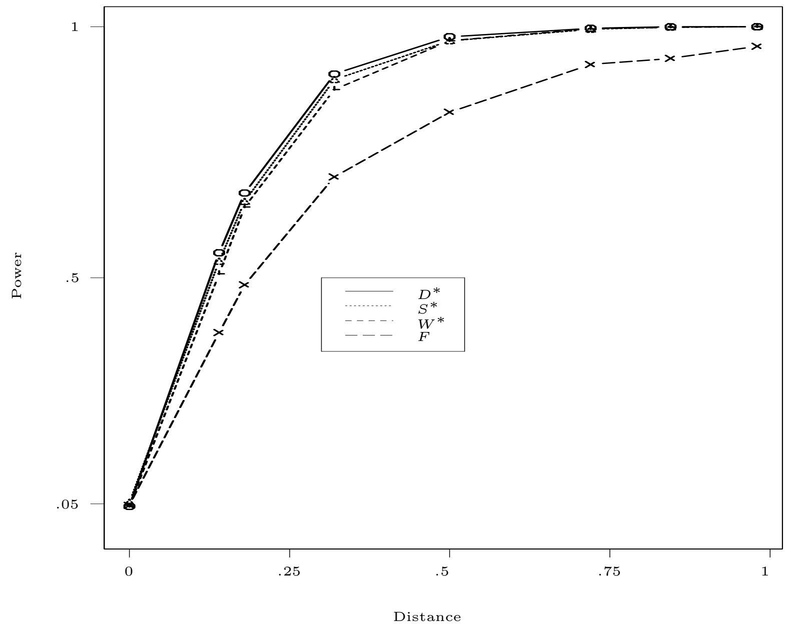

Figure 1. Empirical power curves of D^*, S^*, W^* and F tests when level = .05, $\rho = .5$, $\sigma = 1$, $n = 15$, $p = 3$, $m = 2$, distance $= \sum_{i=1}^{p}(\alpha_i - \bar{\alpha})^2$, and $\underset{\sim}{\epsilon}_j$ follows a multivariate t with 3 d.f.

We observe that the rank tests perform better than the F test. However, among the rank tests the D^* test has the highest empirical power. We have also investigated the empirical powers of the rank tests and the F test for multivariate $t8$ and multivariate normal distributions. When the data came from the $t8$ distribution, the rank tests again perform better than the F test. The D^* test has the highest empirical power among the rank tests. However, the F test is competitive with the rank tests. For the normal distribution the F test performed the best which is expected. The D^* test was the second best. Note that in general the empirical powers of all the tests are increasing functions of ρ. For illustrative purposes, we have plotted the empirical powers of the D^* test in Figure 2.

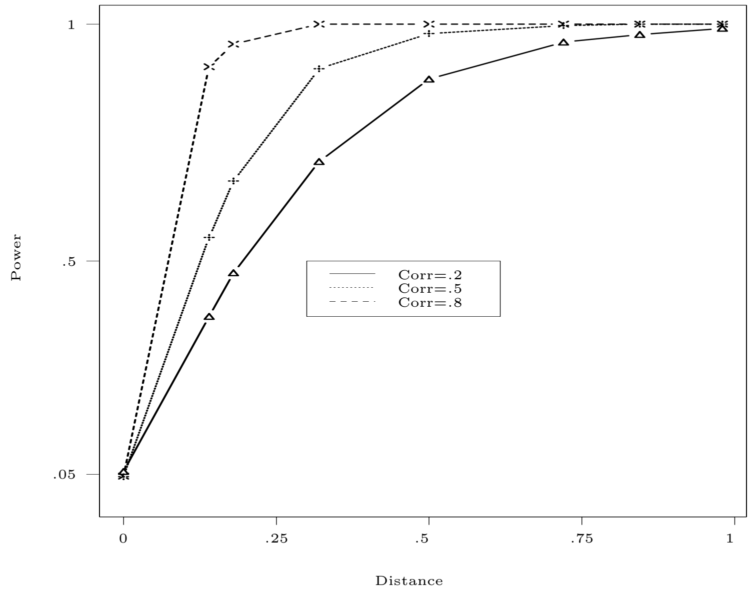

Figure 2. Empirical power curves of D^* test as a function of ρ (corr.) when level=.05, $n=15$, $p=3$, $\sigma=1$, $m=2$, distance $=\sum_{i=1}^{p}(\alpha_i-\bar{\alpha})^2$, and $\underline{\epsilon}_j$ follows a multivariate t with 3 d.f.

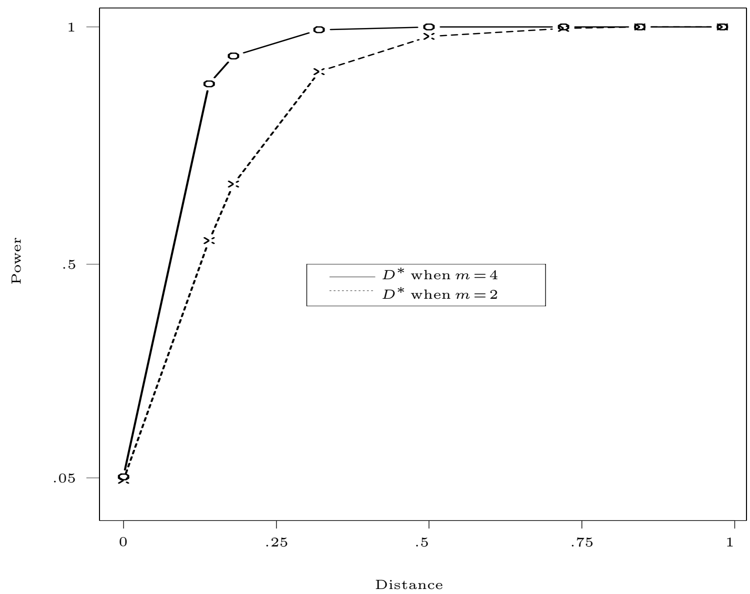

Figure 3. Empirical power curves of the D^* test when level = .05, ρ = .5, σ = 1, n = 15, p = 3, m = (2, 4), distance $= \sum_{i=1}^{p}(\alpha_i - \bar{\alpha})^2$, and $\underline{\epsilon}_j$ follows a multivariate t with 3 d.f.

On the basis of the study, we see that the D^* test is the most powerful test among the rank tests.

Consider now the effect of m on the empirical powers of the rank tests and the F test. For $m = 2$ and $m = 4$, we have plotted empirical power curves (Figure 3) of the D^* test when the distribution is multivariate $t3$, $n = 15$, $p = 3$, $\rho = .5$ with the same alternatives used in previous curves.

It is evident that the empirical powers increase with m.

Next, we investigate whether it is better to analyze the replicates rather than the cell means. We have plotted the empirical power curves of the D^* test based on the replicates and the cell means when the data were generated from the multivariate $t3$. We see from Figure 4 that the empirical powers based on the replicates are higher than those based on the cell means.

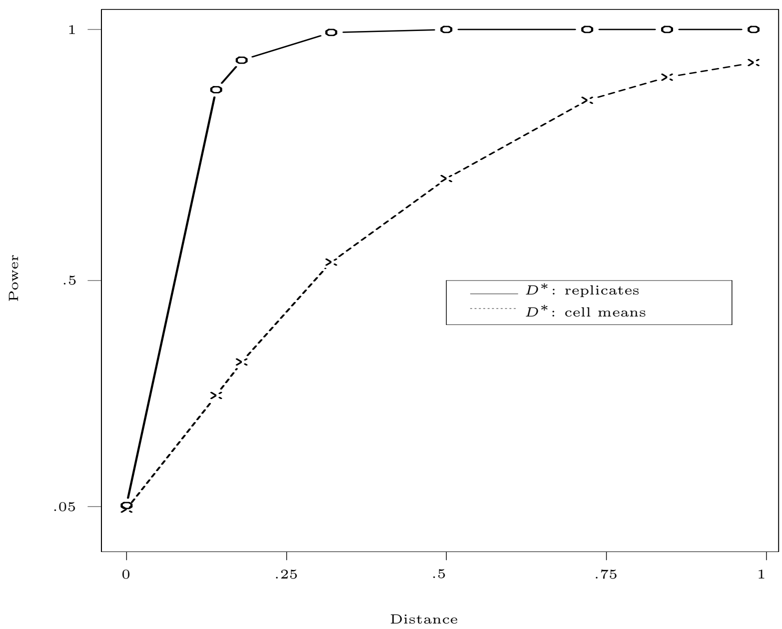

Figure 4. Empirical power curves of the D^* test based on the replicates ($m = 4$) and average of the replicates when Level = .05, $\rho = .5$, $\sigma = 1$, $n = 15$, $p = 3$, distance = $\sum_{i=1}^{p}(\alpha_i - \bar{\alpha})^2$, and $\underset{\sim}{\epsilon}_j$ follows a multivariate t with 3 d.f.

We have also examined the estimated standard error of the estimator of $\alpha_1 - \alpha_2$ and found that for multivariate normal distributions the estimated standard error of the least squares estimate of $\alpha_1 - \alpha_2$ is much less than that of the R-estimate of $\alpha_1 - \alpha_2$. However, for the multivariate t distributions the estimated standard error of the least squares estimate of $\alpha_1 - \alpha_2$ is higher than that of the R-estimators.

To summarize, the D^* test had the best performance among all rank tests. Specifically, for $n \geq 15$ and $p \geq 2$, it achieved the prescribed level of significance well and was the most powerful test. Our simulation study also showed that the empirical powers of the rank tests are better when the analysis is based on replicates (rather than cell means) and increase with cell size.

11. Concluding Remarks

In this article we have developed three rank tests analogous to the non-parametric linear model with i.i.d. errors. We have given a justification of the Benard-Van Elteren statistic from the non-parametric repeated measures model point of view. We have shown that our R-estimators are more robust than those of the normal theory estimators. Our results will have potential applications in biological, behavioral, medical and pharmaceutical research where replications are made for statistical analysis and accuracy.

The results of this article can be extended to unbalanced (different block sizes) and proper (same block size) designs with unequal or no replications per cell. These designs are under investigation and will appear in a future work.

Acknowledgements

The authors would like to thank Professor Myles Hollander of Florida State University for supplying the original data of Woodward (1970). The authors are thankful to Professor C. F. J. Wu (Chair Editor), an associate editor and two referees for useful comments and suggestions that have improved the quality of the paper. The first author's work was carried out while he was a faculty member at Worcester Polytechnic Institute.

References

Benard, A. and Van Elteren, P. (1953). A generalization of the method of m rankings. *Indag. Math.* **15**, 358-369.

Bloch, D. A. and Gastwirth, J. S. (1968). On a simple estimate of the reciprocal of the density function. *Ann. Math. Statist.* **39**, 1083-1085.

Brunner, E. and Dette, H. (1992). Rank procedures for the two-factor mixed model. *J. Amer. Statist. Assoc.* **87**, 884-888.

Draper, D. (1988). Rank-based robust analysis of linear models. 1. Exposition and Review. *Statist. Sci.* **3**, 239-271.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50**, 1096-1121.

Heiler, S. and Willers, R. (1988). Asymptotic normality of R-estimates in the linear model. *Statistics* **19**, 173-184.

Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks.* John Wiley, New York.

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures.* John Wiley, New York.

Hocking, R. R. (1973). A discussion of the two-way mixed model. *American Statistician* **27**, 148-152.

Hocking, R. R. (1985). The analysis of linear models. Montery, California: Brooks/Cole.

Hollander, M. and Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. John Wiley, New York.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43**, 1449-1458.

Jensen, D. R. (1982). Efficiency and robustness in the use of repeated measurements. *Biometrics* **38**, 813-825.

Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley, New York.

Jureckova, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.* **40**, 1889-1900.

Mack, G. A. and Skillings, J. H. (1980). A Friedman-Type rank test for main effects in a two-factor ANOVA. *J. Amer. Statist. Assoc.* **75**, 947-951.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, London, U. K.

Martin, R. D. (1988). Comment on Draper (1987). *Statist. Sci.* **3**, 264-266.

McKean, J. W. and Hettmansperger, T. P. (1976). Tests of hypotheses based on ranks in the general linear model. *Comm. Statist. Theory Methods* **5**, 693-709.

Olsson, D. M. (1974). A sequential simplex program for solving minimization problems. *J. Quality Technology* **6**, 53-57.

Rashid, M. M., Aubuchon, J. C. and Bagchi, A. (1993). Rank based analysis of repeated measures block designs. *Comm. Statist. Theory Methods* **22**, 2783-2811.

Rashid, M. M. and Bagchi, A. (1993). Robust inference for generalized one-way repeated measures. Technical Report, Mathematical Sciences Department, Worcester Polytechnic Institute, Worcester, MA 01609.

Rashid, M. M. (1995). Analysis of repeated measures incomplete block designs using *R*-estimators. *Statist. Probab. Lett.* **22**, 213-221.

Rashid, M. M. and Nandram, B. (1995). A rank-based analysis of one-way repeated measures designs with a changing covariate. *J. Statist. Plann. Inference* **46**, 249-263.

Samuels, M. L., Casella, G. and McCabe, G. P. (1991). Interpreting blocks and random factors. *J. Amer. Statist. Assoc.* **86**, 798-821.

Scheffe, H. (1956). A mixed model for the analysis of variance. *Ann. Math. Statist.* **27**, 23-36.

Woodward, W. F. (1970). A comparison of base running methods in baseball. *M. Sc. thesis, Florida State University.*

Division of Biometrics III, HFD 720, Center for Drug Evaluation and Research, Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, U.S.A.

E-mail: rashidm@cder.fda.gov

Mathematical Sciences Department, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, U.S.A.

E-mail: bagchi@wpi.edu

(Received December 1993; accepted June 1996)