Best, D. J. & Rayner, J. C. W. "An alternative to pages test permitting both tied and missing data" Published in the *Journal of Statistical Theory and Practice*, Vol. 9, Issue 3, p. 524-536, (2015).

Available from: http://dx.doi.org/10.1080/15598608.2014.940098

# An Alternative to Page's Test Permitting Both Tied and Missing Data

D.J. BEST[1] AND J.C.W. RAYNER[2,1]
[1] School of Mathematical and Physical Sciences, University of Newcastle, NSW 2308, Australia
[2] Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia

Data consisting of ranks within blocks are considered for randomized complete block layouts where treatment effects are expected to be ordered. Ranks with and without ties are considered as well as missing values. A small indicative test size study indicates both a new test and the Page test, modified by Thas et al. (2012) to easily permit ties, perform well. An advantage of the new test is that it can be easily applied to data with missing values. Three real examples are given. One of these illustrates a quadratic trend test. Comparisons between the new test and the Alvo and Cabilio (1995) extended Page test for missing values are given.

*AMS Subject Classification*: 62G10; 62G86.
*Keywords*: Analysis of variance; Mid-ranks; Ordered alternatives; Orthogonal trend analysis; Randomized blocks; Size and power study.

## 1. Introduction

Suppose that we have ranks within blocks, $Y_{ij}$, in a randomized block design. Also suppose that if the ranks within blocks are tied then they are given mid-ranks. With $t$ treatments we wish to test the null hypothesis of no treatment effects ($\tau_i$) against ordered alternatives $K_1$: $\tau_1 \leq \tau_2 \leq \ldots \leq \tau_t$ and $K_2$: $\tau_1 \geq \tau_2 \geq \ldots \geq \tau_t$, in each case with at least one inequality strict. The test of Page (1963) is a well-known rank test of the null hypothesis against either $K_1$ or $K_2$.

Thas et al. (2012) give a simple formula for the Page statistic, $L$ say, which is valid for tied and untied data. This is given by

$$L = \sqrt{c}\ \sum_{i=1}^{t} l_i \bar{Y}_i / d$$

in which $d^2 = \sum_{i=1}^{t} l_i^2$, $\bar{Y}_i$ is the mean of the ranks for treatment $i$, the $l_i$ are the usual linear trend coefficients (for completeness these are given in Appendix A) and $c = b(t-1)/(tV)$ in which $V = \sum_{i,j} Y_{ij}^2/(bt) - (k+1)^2/4$. For untied data $V = (t^2-1)/12$. It is known that $L$ has an asymptotic N(0, 1) distribution; see Appendix B. For data with ties we are not aware that use of N(0, 1) p-values in small samples have been investigated previously.

Here we wish to compare the test based on $L$ with that based on a new statistic, $T$ say, derived from the orthogonal trend analysis used in the ANOVA. Kuehl (2000, section 3.3), for example, gives details of orthogonal trend analysis. If $S^2$ is the error mean square from a randomized block ANOVA of the $Y_{ij}$ then an alternative test statistic to $L$ is

$$T = \sqrt{b}\ \sum_{i=1}^{t} \frac{l_i \bar{Y}_i}{dS}$$

in which, as usual for ANOVA, $T$ has an asymptotic $t_{df}$ distribution with $df = (b - 1)(t - 1)$. This alternative test statistic mimics the alternative Conover (1999, p.370) gives for the well-known Friedman test statistic for ranks within blocks. This alternative to the Friedman test statistic is also based on the standard ANOVA F test.

In section 2 we give examples where $L$ and $T$ are calculated. In section 3 we compare sizes for $L$ and $T$ using N(0, 1) and $t_{df}$ critical values respectively. It appears sizes for $L$ are close to nominal for all but very small values of $b$ and $t$. The sizes of $T$ are quite close to nominal for all $b$ and $t$ examined. Section 4 extends the discussion of the section 2 examples. In section 5 we look at an example where data are tied and there are missing values. The statistic $T$ is easily extended to permit the analysis of such data using standard statistical software packages. Section 6 briefly compares sizes and powers for an extension of $T$, denoted by $T^*$, with the Alvo and Cabilio (1995) test statistic which is not available in standard statistical software packages or described in standard texts.

## 2. Examples

(i) Toads *data (no ties)*

Cabilio and Peng (2008) quote data on heart pressure measurements of toads (toads are blocks in the section 1 description) during an induced dehydration period. It was expected that mean rankings would decrease over four time periods (these are treatments in our section 1 description). In our Table 1 we give ranks for the five toads with no missing values.

TABLE 1

*Ranks of heart pressure at four times*

| Toad | Time 1 | Time 2 | Time 3 | Time 4 |
|---|---|---|---|---|
| 1 | 1 | 3 | 4 | 2 |
| 2 | 1 | 2 | 3 | 4 |
| 6 | 1 | 2 | 3 | 4 |
| 7 | 2 | 3 | 1 | 4 |
| 8 | 1 | 2 | 4 | 3 |

For the ranks in Table 1 $L = 2.789$ with a one-tailed p-value of 0.003 using the N(0, 1) approximation and $T = 3.726$ with a one-tailed p-value of 0.001 using the $t_{12}$ distribution. Both $L$ and $T$ are highly significant.

(ii) *Lemonades data (with ties)*

Thas et al. (2012) consider sensory evaluation rankings for five tasters (blocks) and four lemonades (treatments) A, B, C and D where the lemonades were the same except their sugar content increased from A to B to C and then to D. It was expected that the rankings, given in Table 2, would increase as sugar increased. Tied rankings are allowed.

TABLE 2

*Taster rankings for four lemonades*

| Taster | A | B | C | D |
|---|---|---|---|---|
| 1 | 3 | 2 | 1 | 4 |
| 2 | 3 | 1.5 | 1.5 | 4 |
| 3 | 1 | 4 | 2 | 3 |
| 4 | 3 | 2 | 1 | 4 |
| 5 | 4 | 2 | 2 | 2 |

TABLE 3

*Estimated test sizes for tests based on L and T for nominal levels of 10%, 5% and 1% and no ties*

| | *L* | *T* | | *L* | *T* |
|---|---|---|---|---|---|
| $b = 5$ | | | $b = 10$ | | |
| $t = 3$ | 0.079 | 0.086 | $t = 3$ | 0.110 | 0.110 |
| | 0.040 | 0.055 | | 0.046 | 0.049 |
| | 0.007 | 0.010 | | 0.011 | 0.011 |
| $t = 5$ | 0.101 | 0.104 | $t = 5$ | 0.099 | 0.103 |
| | 0.051 | 0.053 | | 0.046 | 0.053 |
| | 0.007 | 0.012 | | 0.010 | 0.011 |
| $t = 7$ | 0.103 | 0.102 | $t = 7$ | 0.100 | 0.102 |
| | 0.048 | 0.052 | | 0.049 | 0.051 |
| | 0.009 | 0.011 | | 0.009 | 0.011 |
| | | | | | |
| $b = 20$ | | | $b = 30$ | | |
| $t = 3$ | 0.089 | 0.093 | $t = 3$ | 0.110 | 0.094 |
| | 0.048 | 0.050 | | 0.051 | 0.051 |
| | 0.010 | 0.010 | | 0.008 | 0.012 |
| $t = 5$ | 0.101 | 0.102 | $t = 5$ | 0.098 | 0.104 |
| | 0.052 | 0.053 | | 0.051 | 0.054 |
| | 0.009 | 0.011 | | 0.010 | 0.011 |
| $t = 7$ | 0.098 | 0.100 | $t = 7$ | 0.099 | 0.101 |
| | 0.040 | 0.051 | | 0.051 | 0.051 |
| | 0.010 | 0.011 | | 0.010 | 0.010 |

TABLE 4

*Estimated test sizes for tests based on L and T with ties allowed and for nominal levels of 10%, 5% and 1%*

| | *L* | *T* | | *L* | *T* |
|---|---|---|---|---|---|
| $b = 5$ | | | $b = 10$ | | |
| $t = 3$ | 0.098 | 0.099 | $t = 3$ | 0.099 | 0.103 |
| | 0.052 | 0.048 | | 0.050 | 0.051 |
| | 0.014 | 0.007 | | 0.011 | 0.009 |
| $t = 5$ | 0.099 | 0.102 | $t = 5$ | 0.100 | 0.104 |
| | 0.051 | 0.053 | | 0.049 | 0.053 |
| | 0.008 | 0.013 | | 0.010 | 0.011 |
| $t = 7$ | 0.103 | 0.103 | $t = 7$ | 0.103 | 0.103 |
| | 0.049 | 0.051 | | 0.049 | 0.051 |
| | 0.009 | 0.011 | | 0.010 | 0.011 |
| | | | | | |
| $b = 20$ | | | $b = 30$ | | |
| $t = 3$ | 0.099 | 0.101 | $t = 3$ | 0.099 | 0.100 |
| | 0.050 | 0.050 | | 0.048 | 0.051 |
| | 0.009 | 0.010 | | 0.009 | 0.010 |
| $t = 5$ | 0.103 | 0.103 | $t = 5$ | 0.097 | 0.103 |
| | 0.051 | 0.053 | | 0.052 | 0.053 |
| | 0.009 | 0.011 | | 0.009 | 0.010 |
| $t = 7$ | 0.101 | 0.102 | $t = 7$ | 0.099 | 0.101 |
| | 0.050 | 0.051 | | 0.051 | 0.052 |
| | 0.010 | 0.011 | | 0.009 | 0.010 |

For the Table 2 rankings $L = 0.408$ with a one-tailed p-value of 0.342 using the N(0, 1) approximation and $T = 0.484$ with a one-tailed p-value of 0.319 based on the $t_{12}$ distribution. Neither $L$ nor $T$ is significant at the usual significance levels.

## 3. Test Sizes and Powers

(i) *Test sizes for data without ties*

Simulations, using 100,000 samples were carried out to check the $L$ and $T$ critical values based on the N(0, 1) and $t_{df}$ distributions. Choices of $b$ and $t$ are shown in Table 3. The sizes were found using permutation tests and showed that for small $b$ and $t$ the test based on $T$ was better approximated by its asymptotic distribution than that based on $L$. For larger $b$ and $t$ both tests had sizes in good agreement with their nominal values. In Appendix B discussion is given supporting the use of $t_{df}$ as an approximate distribution of $T$.

(ii) *Test sizes for data with ties*

Sizes were calculated as in Brockhoff et al. (2004, section 4 and also the discussion in section 6). For each block and treatment one of the scores 1, 2, …, $t$ was randomly assigned with probability $1/t$. These values were then ranked by block with ties given mid-ranks. This was repeated 100,000 times for each $b$ and $t$ combination shown in Table 4. Test sizes were close to nominal for the tests based on both $L$ and $T$.

(iii) *Powers for data without ties*

We now describe a small indicative power study. As with the sizes, the powers were based on 100,000 simulations. Alternative (a) had cell probabilities (0.25, 0.25, 0.25, 0.25) for treatments 1 and 2 and (0.1, 0.2, 0.3, 0.4) for treatments 3 and 4. The alternative (b) cell probabilities were (0.1, 0.2, 0.3, 0.4) for treatment 1, (0.2, 0.2, 0.2, 0.4) for treatment 2 and (0.1, 0.1, 0.1, 0.7) for treatments 3 and 4. Alternative (c) cell probabilities were (0.25, 0.25, 0.25, 0.25) for treatments 1 and 2 and (0.1, 0.1, 0.3, 0.5) for treatments 3 and 4. With no ties the powers of the tests based on $T$ and $L$ were very similar.

TABLE 5

*Estimated powers for tests based on L and T for $\alpha = 10\%$, 5% and 1%, $b = 8$, $t = 4$ with no ties allowed and alternative cell probabilities (a), (b), (c)*

| Alternative | $\alpha\%$ | $T$ | $L$ |
|---|---|---|---|
| (a) | 10 | 0.39 | 0.33 |
| | 5 | 0.26 | 0.25 |
| | 1 | 0.10 | 0.09 |
| (b) | 10 | 0.41 | 0.41 |
| | 5 | 0.28 | 0.27 |
| | 1 | 0.11 | 0.10 |
| (c) | 10 | 0.55 | 0.55 |
| | 5 | 0.41 | 0.40 |
| | 1 | 0.18 | 0.17 |

(iv) *Powers for data with ties*

The set-up was as for the previous section but ties were allowed. The powers for the test based on $L$ were similar to the no ties case, but the powers of the test based on $T$ improved on the no ties case, and therefore here on the powers for the test based on $L$.

TABLE 6

*Estimated powers for tests based on L and T for α = 10%, 5% and 1%, b = 8, t = 4 with ties allowed and alternative cell probabilities (a), (b), (c)*

| Alternative | $\alpha$% | $T$ | $L$ |
|---|---|---|---|
| (a) | 10 | 0.42 | 0.39 |
| | 5 | 0.28 | 0.24 |
| | 1 | 0.11 | 0.06 |
| (b) | 10 | 0.49 | 0.40 |
| | 5 | 0.35 | 0.24 |
| | 1 | 0.14 | 0.05 |
| (c) | 10 | 0.61 | 0.57 |
| | 5 | 0.47 | 0.40 |
| | 1 | 0.23 | 0.14 |

## 4. Further Analysis of the Section 2 Examples

(i) *Toads data*

The ranks data given in Table 1 are for toads with no missing values. Table 7 shows extra data for three other toads.

TABLE 7

*Toads with missing ranked values*

| Toad | Time 1 | Time 2 | Time 3 | Time 4 |
|---|---|---|---|---|
| 3 | - | 1 | 2 | 3 |
| 4 | 1 | - | - | 2 |
| 5 | 1 | 2 | - | 3 |

We can find an extension of $T$, called $T^*$, for when there are missing values by using an ANOVA computer routine that allows for unbalanced data. For example, use the GLM command in MINITAB, the RGLM routine in IMSL, the 'Fit Model' platform in JMP or the 'general linear model' platform in SPSS. These give an error mean square, $S^2$, as before but now with $df = (n - 1) - (t - 1) - (b - 1)$ degrees of freedom where $n$, the total number of observations, is 28 for the extended toad data. Also we define $n_1 = 7$, $n_2 = 7$, $n_3 = 6$ and $n_4 = 8$ as the number of observations for each of the four times. Moreover we suggest not using $\bar{Y}_i$, the raw means for each time, but rather $\bar{Y}_i^*$, the adjusted means given by standard software such as that listed above. This gives our $T^*$ for missing value data as

$$T^* = \frac{\sum_{i=1}^{t} l_i \bar{Y}_i^*}{S\sqrt{\sum_{i=1}^{t} l_i^2 / n_i}}.$$

Notice that if there are no ties then $T = T^*$. Section 6 briefly examines this approximation. Here $T^*$ takes the value 4.900 with one-sided p-value 0.0001 based on the $t_{17}$ approximation. This is a smaller p-value than that obtained using only the data from complete blocks.

Another statistic that has been suggested for missing values data when an ordered alternative is thought important is the $L^*$ statistic of Alvo and Cabilio (1995). However critical values of $L^*$ exist only for certain values of t and for certain missing value layouts. When the appropriate normal approximation is used with this test statistic we denote the test and test statistic by $Z$. It is clearly more convenient for general use. For data with no missing values $Z = L$. Section 6 compares $Z$ and $T^*$. For the toad data including toads with missing values $Z$ takes the value 3.50 and with p-value 0.002.

(ii) *Lemonade data*

A more appropriate alternative for the lemonade data might be that as sugar content increases taster preference increases and then decreases as the drink goes from not sweet enough to too sweet. A quadratic contrast may then be appropriate and Thas et al. (2012) suggest the statistic

TABLE 8

*Rankings for ordered categories nasal discharge data*

| ID | Day 1 | Day 2 | Day 3 | Day 4 |
|---|---|---|---|---|
| 1 | 1.5 | 1.5 | 3.5 | 3.5 |
| 2 | 2.5 | 2.5 | 2.5 | 2.5 |
| 3 | 2.5 | 2.5 | 2.5 | 2.5 |
| 4 | 2.5 | 2.5 | 2.5 | 2.5 |
| 5 | 1.5 | 3.5 | 3.5 | 1.5 |
| 6 | 4 | 2 | 2 | 2 |
| 7 | 3 | 3 | 1 | 3 |
| 8 | 3 | 3 | 3 | 1 |
| 9 | 4 | 3 | 1.5 | 1.5 |
| 10 | 2 | 2 | 2 | 4 |
| 11 | 3 | 1 | 3 | 3 |
| 12 | 2 | 4 | 2 | 2 |
| 13 | 1.5 | 4 | 3 | 1.5 |
| 14 | 4 | 2 | 2 | 2 |
| 15 | - | 2.5 | 2.5 | 1 |
| 16 | 1 | 3 | 3 | 3 |
| 17 | 4 | 2 | 2 | 2 |
| 18 | 2.5 | 2.5 | 2.5 | 2.5 |
| 19 | 3 | - | 2 | 1 |
| 20 | 2.5 | 2.5 | 2.5 | 2.5 |
| 21 | 4 | 2 | 2 | 2 |
| 22 | 2 | 2 | 4 | 2 |
| 23 | 3.5 | 1.5 | 1.5 | 3.5 |
| 24 | 2.5 | 2.5 | 2.5 | 2.5 |
| 25 | 2.5 | 2.5 | 2.5 | 2.5 |
| 26 | 2.5 | 2.5 | 4 | 1 |
| 27 | 4 | 2 | 2 | 2 |
| 28 | 2.5 | 2.5 | 2.5 | 2.5 |
| 29 | 1.5 | 3.5 | 3.5 | 1.5 |
| 30 | 2.5 | 2.5 | 2.5 | 2.5 |
| 31 | 3 | 3 | 3 | 1 |
| 32 | 2.5 | 2.5 | 2.5 | 2.5 |

$$Q = \frac{\sqrt{c}}{d}\sum_{i=1}^{t} m_i \overline{Y}_i$$

where now $d = \sum_{i=1}^{t} m_i^2$. The $m_i$ values are given in Appendix A. We find $Q$ = 2.191 with an N(0, 1) one-tailed p-value of 0.014 and $T = \sum_{i=1}^{t} m_i \overline{Y}_i / (Sd/\sqrt{b})$ = 2.598 with a one-tailed p-value of 0.012 based on the $t_{12}$ approximation.

## 5. An Example with Both Ties and Missing Values

Davis (2002, Chapter 8) shows how to use Cochran-Mantel-Haenszel (CMH) statistics to analyze ordinal categorical data with missing values. For patients monitored on four consecutive days the ordered categories were no nasal discharge, mild nasal discharge, moderate nasal discharge, and severe symptoms. Davis assigns arbitrary scores 0, 1, 2 and 3 to the ordered categories and uses these to check for a linear trend over days. A sensible alternative is to use ranks. Table 8 shows the ranks for the Davis data. If only patients with complete data are used $T$ = – 1.641 with p-value 0.052 using the $t_{87}$ one-tailed approximation. If all the data, that is, patients with complete and incomplete data, are used, then $T^*$ = – 2.031 with p-value 0.023 based on the $t_{92}$ one-tailed approximation. Again, as for the toad data, it seems important to use all the data and not delete incomplete blocks. Section 6 indicates the $t_{df}$ approximation to the $T^*$ distribution is quite good, even for small samples. Table 11 in the following section indicates $Z$ is not well approximated by an N(0, 1) distribution when there are both ties and missing values. Hence we only quote $T^*$ probabilities here.

## 6. Sizes and Powers when there is Missing Data

(i) *Sizes for data without ties*

We will compare sizes of the tests based on $T^*$ and the $Z$ statistic of Alvo and Cabilio (1995) when $t_{df}$ and N(0, 1), respectively, critical values are used. Table 9 gives sizes for three missing values layouts which are either (i) the same layout as in the toads example given in Cabilio and Peng (2008) or in Alvo and Cabilio (1995), (ii) a layout with two extra missing values compared to (i) such that the first five blocks have no missing values, block six has ranks for times (treatments) 1 and 2 missing and blocks seven and eight have ranks for times 2 and 3 missing, (iii) a layout with one less missing value compared to (i) such that the first five blocks have no missing values, block six has the rank for time 1 missing, block seven has the rank for time 2 missing and block eight has the rank for time 3 missing.

The sizes in Table 9 are excellent for both $T^*$ and $Z$. Alvo and Cabilio (1995) also noted the sizes for $Z$ were good even, as here, in small samples. The sizes were found using permutation tests involving 100,000 simulations of samples of eight blocks and four treatments. Permutations, like the rankings, were done within blocks as required by the Page test.

(ii) *Powers for data without ties*

As the sizes were excellent for layouts (i), (ii) and (iii) we now again use the $t_{df}$ and N(0, 1) critical values. This implies the powers in Table 10 will not be biased by differences in sizes, which is a problem with some power studies.

TABLE 9

*Sizes when there are no ties, eight blocks, four treatments, $\alpha$ = 10%, 5% and 1% and missing patterns (i), (ii), (iii) described in the text*

| Missing value pattern | $\alpha$% | $T^*$ | $Z$ |
|---|---|---|---|
| (i) | 10 | 0.100 | 0.101 |
| | 5 | 0.047 | 0.048 |
| | 1 | 0.010 | 0.008 |
| (ii) | 10 | 0.101 | 0.104 |
| | 5 | 0.050 | 0.046 |
| | 1 | 0.011 | 0.008 |
| (iii) | 10 | 0.101 | 0.102 |
| | 5 | 0.052 | 0.052 |
| | 1 | 0.011 | 0.009 |

TABLE 10

*Powers when there are no ties, eight blocks, four treatments, $\alpha$ = 10%, 5% and 1%, alternatives (a), (b), (c) and missing values patterns (i), (ii), (iii) described in the text*

| Missing value pattern | alternative | $\alpha$% | $T^*$ | $Z$ |
|---|---|---|---|---|
| (i) | (a) | 10 | 0.36 | 0.35 |
| | | 5 | 0.22 | 0.22 |
| | | 1 | 0.08 | 0.06 |
| | (b) | 10 | 0.38 | 0.37 |
| | | 5 | 0.24 | 0.24 |
| | | 1 | 0.09 | 0.07 |
| | (c) | 10 | 0.50 | 0.50 |
| | | 5 | 0.35 | 0.35 |
| | | 1 | 0.15 | 0.13 |
| (ii) | (a) | 10 | 0.35 | 0.34 |
| | | 5 | 0.22 | 0.20 |
| | | 1 | 0.08 | 0.06 |
| | (b) | 10 | 0.38 | 0.35 |
| | | 5 | 0.24 | 0.21 |
| | | 1 | 0.09 | 0.06 |
| | (c) | 10 | 0.49 | 0.47 |
| | | 5 | 0.35 | 0.31 |
| | | 1 | 0.14 | 0.10 |
| (iii) | (a) | 10 | 0.38 | 0.36 |
| | | 5 | 0.24 | 0.23 |
| | | 1 | 0.09 | 0.07 |
| | (b) | 10 | 0.39 | 0.38 |
| | | 5 | 0.26 | 0.25 |
| | | 1 | 0.10 | 0.08 |
| | (c) | 10 | 0.52 | 0.51 |
| | | 5 | 0.38 | 0.37 |
| | | 1 | 0.16 | 0.14 |

TABLE 11

*Sizes when there are ties, eight blocks, four treatments, $\alpha$ = 10%, 5% and 1% and missing patterns (a), (b) , (c)*

| Missing value pattern | $\alpha$% | $T^*$ | $Z$ |
|---|---|---|---|
| (a) | 10 | 0.098 | 0.073 |
| | 5 | 0.048 | 0.031 |
| | 1 | 0.011 | 0.002 |
| (b) | 10 | 0.096 | 0.072 |
| | 5 | 0.050 | 0.029 |
| | 1 | 0.010 | 0.003 |
| (c) | 10 | 0.099 | 0.075 |
| | 5 | 0.051 | 0.033 |
| | 1 | 0.011 | 0.004 |

Although powers were comparable it should be noted that those for the test based on $T^*$ were never inferior to those for the test based on $Z$. The powers for layouts (i), (ii) and (iii) were similar and missing value data lessened the powers compared to those for the eight blocks with no missing values. See Table 5 and 10.

(iii) *Sizes for data with ties*

Table 9 gives sizes when there are no ties. Table 11 gives sizes when ties are allowed. The tied data was produced using the same technique as in section 3 (ii) above except that now not all blocks have $t$ values. When block $i$ has $n_i$ values then scores 1, 2, …, $n_i$ were produced with probabilities $1/n_i$. These scores were then ranked using mid-ranks for ties.

Alvo and Cabilio (1995) did not give an adjustment for ties and so in Table 11 we have just used the same $Z$ as for the no ties case except that the numerator now uses mid-ranks rather than ranks and the denominator or standard error of the numerator is unadjusted for ties. We did not expect Z to have good sizes, and indeed, this was the case. However, as for the no ties case, the sizes for $T^*$ are, excellent.

(iv) *Powers for data with ties*

For every alternative considered the powers for the test based on $T^*$ were superior to those for the test based on $Z$. See Table 12.

## 7. Conclusion

For almost all $t$ and $b$ studied the N(0, 1) approximation to $L$ and the $t_{df}$ approximation to $T$ were excellent. Only for untied data with $t = 3$ and $b = 5$ was the normal approximation slightly inferior to the $t_{df}$ approximation, and even then its use is not inappropriate.

An advantage of the $T$ statistics is that using existing software for unbalanced ANOVA gives $T^*$, an easy to calculate extension of $T$ for ranks data with missing values. The use of $T$ and $T^*$ was illustrated for the real toad and nasal discharge data sets and for artificial lemonade data based on actual data from the former CSIRO Food Research Laboratory at North Ryde. Use of a quadratic trend test was also illustrated. We note that $Z$ does not apply to data with both ties and missing values whereas $T^*$ does apply to such data. This is probably the main reason for $T^*$ having more power than $Z$ in Table 12 and for $T^*$ having better sizes than $Z$ in Table 11. However note that in Table 6 $T$ has slightly better power than $Z$ even though the Table 4 sizes were similar. Our results indicate that the tests based on $T$ and $T^*$ are superior alternatives to those based $L$ and $Z$.

We thank the referees for constructive comments that improved the paper.

TABLE 12

*Powers when there are ties, eight blocks, four treatments, $\alpha$ = 10%, 5% and 1%, alternatives (a), (b), (c) and missing values patterns (i), (ii), (iii) described in the text*

| Missing value pattern | alternative | $\alpha$% | $T^*$ | $Z$ |
|---|---|---|---|---|
| (i) | (a) | 10 | 0.39 | 0.34 |
| | | 5 | 0.26 | 0.20 |
| | | 1 | 0.09 | 0.04 |
| | (b) | 10 | 0.45 | 0.35 |
| | | 5 | 0.31 | 0.20 |
| | | 1 | 0.12 | 0.04 |
| | (c) | 10 | 0.57 | 0.50 |
| | | 5 | 0.42 | 0.33 |
| | | 1 | 0.19 | 0.10 |
| (ii) | (a) | 10 | 0.38 | 0.32 |
| | | 5 | 0.25 | 0.18 |
| | | 1 | 0.09 | 0.04 |
| | (b) | 10 | 0.44 | 0.33 |
| | | 5 | 0.31 | 0.17 |
| | | 1 | 0.12 | 0.03 |
| | (c) | 10 | 0.55 | 0.47 |
| | | 5 | 0.41 | 0.30 |
| | | 1 | 0.18 | 0.08 |
| (iii) | (a) | 10 | 0.40 | 0.34 |
| | | 5 | 0.27 | 0.21 |
| | | 1 | 0.10 | 0.05 |
| | (b) | 10 | 0.46 | 0.35 |
| | | 5 | 0.32 | 0.20 |
| | | 1 | 0.13 | 0.04 |
| | (c) | 10 | 0.58 | 0.50 |
| | | 5 | 0.44 | 0.34 |
| | | 1 | 0.21 | 0.10 |

## References

Alvo, M. and Cabilio, P. (1995). Testing ordered alternatives in the presence of incomplete data. *J. Amer. Statist. Assoc.*, **90**, 1015-1024.

Bickel, P.J. and Dobson, K.A. (1977). *Mathematical Statistics*. San Francisco: Holden-Day.

Brockhoff, P.B., Best, D.J. and Rayner, J.C.W. (2004). Partitioning Anderson's statistic for tied data. *J. Statist. Planning Infer.*, **121**, 93-111.

Cabilio, P. and Peng, J. (2008). Multiple rank based testing for ordered alternatives with incomplete data. *Statistics and Probability Letters*, **78**, 2609-2613.

Conover, W.J. (1999). *Practical Nonparametric Statistics* (3rd ed.). New York: Wiley.

Davis, C.S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. New York: Springer.
Kuehl, R.O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*. Pacific Grove, California: Duxbury Press.
Page, E. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *J. Amer. Statist. Assoc.*, **58**, 216-230.
Thas, O., Best, D.J. and Rayner, J.C.W. (2012). Using orthogonal trend contrasts for testing ranked data with ordered alternatives. *Statistica Neerlandica*, **66**(4), 452-471.

## Appendix A: Linear and Quadratic Coefficients

*Linear Coefficients*

| $t$ | $l_1, l_2, \ldots, l_t$ | $\sum_{i=1}^{t} l_i^2$ |
|---|---|---|
| 3 | – 1, 0, 1 | 2 |
| 4 | – 3, – 1, 1, 3 | 20 |
| 5 | – 2, – 1, 0, – 1, 2 | 10 |
| 6 | – 5, – 3, – 1, 1, 3, 5 | 70 |
| 7 | – 3, – 2, – 1, 0, 1, 2, 3 | 28 |

*Quadratic Coefficients*

| $t$ | $m_1, m_2, \ldots, m_t$ | $\sum_{i=1}^{t} m_i^2$ |
|---|---|---|
| 3 | 1, – 2, 1 | 6 |
| 4 | 1, –1, – 1, 1 | 4 |
| 5 | 2, –1, – 2, – 1, 2 | 14 |
| 6 | 5, – 1, – 4, – 4, – 1, 5 | 84 |
| 7 | 5, 0, – 3, – 4, – 3, 0, 5 | 84 |

## Appendix B: Asymptotic distributions of *L* and *T*

To reflect their dependence on the sample size various statistics such as $L$ will be denoted by $L_n$. All limits are as $n \to \infty$; they are either in probability or in law. Now

$$L_n = \frac{\sqrt{b(t-1)}}{d\sqrt{t}} \frac{\sum_{i=1}^{t} l_i \overline{Y}_i}{\sqrt{V_n}}$$

Using results found, for example, in Bickel and Dobson (1977), $V_n \xrightarrow{P} \sigma^2$, the variance of the ranks. It follows that $L_n \xrightarrow{L}$ constant $\sum_{i=1}^{t} l_i \overline{Y}_i$, this having the N(0, 1) distribution, as is well-known.

Similarly

$$T_n = \sqrt{b} \sum_{i=1}^{t} \frac{l_i \overline{Y}_i}{d S_n}.$$

Now $S_n^2$, the error mean square, $\xrightarrow{P}$ $\sigma^2$, so $T_n$ $\xrightarrow{L}$ constant $\sum_{i=1}^{t} l_i \overline{Y}_i$, this again having the standard normal distribution, as is well-known.

If the data are normal $T_n$ has the $t_{df}$ distribution, but that is not the case here, as the data are ranks. However the ANOVA is well-known to be robust to its assumptions. Moreover the $t_{df}$ distribution approaches the standard normal as the degrees of freedom increase. So it is reasonable to anticipate that $T_n$ approaches normality with the $t_{df}$ distribution an intermediate approximation. This is supported by the simulation reported in section 3.