AN ALTERNATIVE TO FRIEDMAN'S TEST PERMITTING BOTH TIED AND MISSING DATA

D.J. BEST[1] and J.C.W. RAYNER[2,1*]

University of Wollongong and University of Newcastle

Summary

Data consisting of ranks within blocks are considered for randomized block designs. Ties are permitted. Examples are given of Friedman's test and a test due to Conover, both appropriate when there are no missing values. Examples also given of the Skillings-Mack test and an extension of the Conover approach, both appropriate when there are missing values. An indicative empirical study suggests that compared with the Friedman and their Skillings-Mack competitors, the more convenient Conover test statistic and its extension are better approximated by their asymptotic distributions. Moreover the tests based on them are slightly more powerful than the Friedman and the Skillings-Mack competitors; there is a greater power advantage with tied data.

AMS Subject Classification: 62Gxx; 62Fxx

Keywords: Analysis of variance; randomized block designs; F tests; powers; ranks data; test sizes

* Author to whom correspondence should be addressed.
e-mail: John.Rayner@newcastle.edu.au
[1] School of Mathematical and Physical Sciences, University of Newcastle, NSW 2308, Australia
[2] National Institute for Applied Statistics Research Australia, University of Wollongong, NSW 2522, Australia

1. Introduction

Randomized block designs are commonly used in taste-test or sensory evaluation studies where blocks are judges or tasters and treatments are food or beverage products. Often judges are asked to rank the products for preference or to categorize the products on an ordinal five point scale. For example, *just about right* categories may be 'definitely not sweet enough', 'not quite sweet enough', 'just about right sweetness', 'a little too sweet' and 'definitely too sweet'. Similarly *likely to purchase* categories might be 'definitely would not purchase', 'unlikely to purchase', 'not sure if would purchase or not', 'likely to purchase' and 'definitely would purchase'. Taking ranks of this ordinal data is a sensible approach and use of Friedman's test for rankings data and ranks of ordinal data would commonly be used as part of a statistical analysis. Sometimes the ranks data contain lost, corrupted or otherwise missing values. This paper gives improved analysis for data with missing values. Sometimes due to possible sensory fatigue or palate paralysis a judge ranks or categorizes only a subset of the products and incomplete block designs are used. Such data can be regarded as randomized block data with missing values and again the improved analysis given here is appropriate.

Although we give sensory evaluation applications subsequently, we note that the analysis we give in this paper also applies in many other areas of study.

This introduction considers test statistics for ranked and possibly tied data from randomized blocks for the case of no missing values, while section 2 considers applications with missing values. Section 3 looks at sizes for two missing value statistics when asymptotic distributions are used to obtain p-values. Section 4 gives a small power study while section 5 briefly considers incomplete blocks.

We will illustrate the application of the Friedman test when there are tied data but no missing values using data from Sprent and Smeeton (2007, p. 223). This data set is for five tasters who rank four varieties of raspberries in order of preference; A is Malling Enterprise, B is Malling Jewel, C is Glen Clova and D is Norfolk Giant. Mid-ranks are given for ties. See Table 1.

We are interested in testing H_0: there are no treatment (variety) effects against H_A: at least two treatments differ. Table 2 counts how many times each ranking occurs in Table 1. As above, if we consider b blocks (tasters in Table 1) and t products then we can say the probability of each ranking is the count in Table 2 divided by bt.

In Table 2 r_s, $s = 1, \ldots, q$ denotes the sth ranking and c_s the associated count. The variance of a rank in Table 1 is

$$V = \left\{ \sum_{s=1}^{q} r_s^2 c_s / (bt) \right\} - (t+1)^2 / 4,$$

where q is the number of different values of s; $q = 5$ in Table 2. Notice that if there were no tied rankings then $c_s/(bt)$ is $1/t$ and $V = (t^2 - 1)/12$.

Friedman's well-known test statistic S is given by

$$S = \left\{ \frac{(t-1)}{t} \right\} \left\{ \frac{1}{bV} \right\} \left\{ \sum_{j=1}^{t} \left[R_j - \frac{b(t+1)}{2} \right]^2 \right\}$$

in which R_j is the sum of the ranks given to treatment (variety) j, $j = 1, \ldots, t$. For the raspberry data $S = 6.47$ and using the χ_3^2 approximation gives a p-value is 0.09. This agrees with Sprent and Smeeton (2007) who give an alternative formula for S.

Conover (1999, p. 370) suggests the test statistic

$$T_2 = \frac{(b-1)S}{b(t-1)-S}$$

in preference to S because its approximate distribution, $F_{(t-1),(b-1)(t-1)}$, is more accurate than the $\chi^2_{(t-1)}$ approximate distribution for S. The statistic T_2 is just the ANOVA F test statistic for treatments on the ranked data. For the Table 1 data $T_2 = 3.03$ and using the $F_{3,12}$ approximation results in a p-value of 0.07. This is a little smaller than the p-value based on the χ^2 approximation to S. Our experience is that T_2 generally gives a smaller p-value than S, particularly if there are tied ranks. In section 2 we will discuss generalizations of S and T_2 that cope with missing values.

2. Extension for Missing Values

The most common extension of S which copes with missing values is probably the Skillings-Mack statistic, T^*, defined in Skillings and Mack (1981). Also see Hollander et al. (2014, section 7.8). Within each block rank the observations from 1 to t_i, where t_i is the number of treatments in block i. If ties occur, mid-ranks are used.

Let r_{ij} be the rank for the jth treatment on the ith block, assuming the treatment is not missing. Otherwise put $r_{ij} = (t_i + 1)/2$. Compute the adjusted treatment sum for treatment j as

$$A_j^* = \sum_{i=1}^{b} \{(r_{ij} - \frac{t_i+1}{2})\sqrt{\frac{12}{t_i+1}}\}, j = 1, 2, \ldots, t-1.$$

To specify the variance-covariance matrix Σ* of the { A_j^* } first let m_{jk} be the number of blocks containing both treatment j and treatment k. Then the off-diagonal elements of Σ* are ($-m_{jk}$) and the diagonal element for row i is the negative of the sum of the off-diagonal elements for row j. The Skillings-Mack test statistic is

$$T^* = \mathrm{A}^{*\mathrm{T}}\, \Sigma^{-1} \mathrm{A}^*$$

in which A* = ($A_1^*, A_2^*, \ldots, A_{t-1}^*$)$^{\mathrm{T}}$. Observe that T^* has an approximate χ^2_{t-1} distribution.

To generalize T_2 to the missing values case compute an F value for an ANOVA using the adjusted sum of squares for treatments. Software for this sum of squares is commonly available. See, for example, the GLM command in MINITAB or the FIT MODEL command in JMP; These software packages also conveniently give multiple comparisons. We obtain the ANOVA shown in Table 3.

T_2 is just the F test using the adjusted treatments sum of squares and has an approximate $F_{(t-1),(b-1)(t-1)-m}$ distribution. For further discussion see, for example, Kuehl (2000, section 8.5).

Table 4 gives some data based on an example in Skillings and Mack (1981). See also Hollander et al. (2014, p. 346). For these data T^* = 15.49 with a χ^2_3 p-value of 0.0014.

To find T_2 suppose we use MINITAB with the data entered as in Table 5. Then use the command GLM C1 = C2 C3, to obtain the ANOVA table in Table 6. As is common in our experience, the p-value for T_2, 0.0000, is smaller than that for T^*.

Section 3 checks the F and χ^2 approximations to the null distributions of T_2 and T^* in the missing values case. The results in section 3 and other preliminary work suggest the F approximation to the null distribution of T_2 is good. Section 4 gives some limited but indicative powers.

3. Test sizes

Iman and Davenport (1980) have already looked at sizes for the no missing data case and so here we will only consider the missing data case; however we will refer to their findings. We consider (a) no data tied and (b) tied data occur.

(a) *No ties*

We use permutation test based on 100,000 samples with $t = 4$ and $b = 8$ and consider three different layouts of missing values. All have no missing values in the first five blocks. Layout (i) has block 6 with treatment 1 missing, block 7 with treatments 2 and 3 missing. Layout (ii) has two further missing values compared to (i): block six has ranks for treatments 1 and 2 missing and blocks seven and eight have ranks for treatments 2 and 3 missing. Layout (iii) has one less missing value compared to (i): the first five blocks have no missing values, block six has the rank for treatment 1 missing, block seven has the rank for treatment 2 missing and block eight has the rank for treatment 3 missing.

We also use these three layouts for case (b) sizes and for cases (a) and (b) for powers in section 4 following.

In Table 7 sizes for T^* are slightly lower than nominal and those for T_2 are slightly higher. This agrees with the no missing values results of Iman and Davenport (1980). However the sizes for T_2 are always closer to α, again mirroring the results of Iman and Davenport (1980) for the no missing values case.

(b) *Ties may occur*

Table 8 gives sizes where ties are allowed. When block i has t_i values then scores 1, 2, …, t_i were produced with probabilities $1/t_i$ for each treatment score. These scores were then ranked using mid-ranks for ties. This approach was given in Brockhoff et al. (2004, section 4). We conclude that for these missing values layouts, and when random allocation of ties is allowed, the sizes of T_2 are a considerable improvement on the T^* sizes: although the former are slightly large they are close to the nominal sizes, the latter are far too small.

4. Test powers

We use the same three layouts as for the sizes. Now we also consider three alternatives to the null. The Brockhoff et al. (2004) approach was again used but now treatments no longer had equal probability scores but rather treatment score probabilities as follows. As with the sizes, the powers were based on 100,000 simulations. Alternative (a) had treatment probabilities (0.25, 0.25, 0.25, 0.25) for treatments 1 and 2 and (0.1, 0.4, 0.2, 0.3) for treatments 3 and 4. The alternative (b) treatment 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) treatment 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. Cases (a) no ties and (b) ties allowed were again considered.

In Table 7 sizes for T_2 were greater than those for T^*. As Table 9 uses the same approximate distributions to get the critical values used in Table 7, we might expect the powers for the test based on T_2 to be greater than those based on T^*. This is, in fact, born out in Table 9. Recalling that in Table 7 the T_2 sizes were closer to the nominal than the T^* sizes there appears to be benefit in using T_2 rather than T^*.

The Table 10 powers favour T_2 over T^* even more than those in Table 9. This is hardly surprising given the substantially smaller sizes for T^* apparent in Table 8. Given the better sizes for T_2 in Table 8 and the substantially greater powers in Table 10 we suggest even more strongly that T_2 be used rather than T^* when there are missing values.

We have based our size and power comparisons on approximate chi-squared and F critical values rather than Monte Carlo critical values as the F approximation is quite good and because we feel many users will not get p-values via Monte Carlo.

5. Incomplete block designs

Other designs which are randomized block designs with missing values but have more structure are balanced incomplete block designs (BIBDs) and partially balanced incomplete block designs (PBIBDs). We have considered BIBDs in Best and Rayner (2014) and so will not consider these here. We will, however, give an example of ranked data from a PBIBD analysed using both Skillings-Mack and the extended Conover F statistics. See the data and design in Table 11. Note that (i) each product is ranked four times and (ii) the design is a cyclic design. As this is a cyclic design we could also have discussed a statistic of Alvo and Cabilio (1993) but we will not do so as it has a difficult approximate distribution and it does not deal with tied data. Another possibility is to use the classical Durbin test statistic to analyse PBIBD data such as that in Table 11. Again we do not investigate this approach here.

For these data the Skillings-Mack statistic is 12.33 with p-value 0.03 using the χ^2_5 approximation. For the extended Conover F statistic we find the ANOVA given in Table 12. Using the $F_{5,13}$ approximation gives a p-value of 0.001: somewhat smaller than the p-value for the Skillings-Mack test.

The powers in Table 10 support using the extended F test with $F_{5,13}$ approximation rather than the Skillings-Mack test with its χ^2_5 approximation. For this data set both approaches give p-values significant at the 5% level but this would not always be the case. At the 1% level the conclusions differ.

6. Conclusion

Conover (1999, p. 370) suggested replacing use of Friedman's rank test for ranks data in randomized blocks by an ANOVA test. We have considered an extension of this ANOVA test approach for testing ranks data in randomized blocks with missing values. Missing values might occur at random or as part of an incomplete block structured design. Based on our indicative results, the extension of the ANOVA test works well and we recommend its use rather than the Skillings-Mack test.

References

ALVO, M. & CABILIO, P. (1993). Rank correlations and the analysis of rank-based experimental designs. Chapter 8 in *Probability Models and Statistical Analyses for Ranking Data*. New York: Springer-Verlag.

BEST, D.J. & RAYNER, J.C.W. (2014). Conover's F test as an Alternative to Durbin's test. To appear in the *Journal of Modern Applied Statistical Methods*.

BROCKHOFF, P.B., BEST, D.J. & RAYNER, J.C.W. (2004). Partitioning Anderson's statistic for tied data. *J. Statist. Planning Infer*., 121, 93-111.

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

HOLLANDER, M., WOLFE, D.A. and CHICKEN, E. (2014). *Nonparametric Statistical Methods* (3rd edition). New York: Wiley.

IMAN R.L. & DAVENPORT J.M. (1980). Approximations to the critical region of the Friedman statistic. *Communications in Statistics - Theory and Methods*, 571-595.

KUEHL, R.O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*. Pacific Grove, CA: Duxbury Press.

SKILLINGS, J.H. & MACK, G.A. (1981). On the use of a Friedman-type statistic in balanced and unbalanced block designs. *Technometrics*, 23, 171-177.

SPRENT, P. & SMEETON, N. (2007). *Applied Nonparametric Statistical Methods*. (4th ed.) Boca Raton: Chapman and Hall/CRC.

TABLE 1

Preference ranks within tasters of four raspberry varieties A, B, C and D

Taster	A	B	C	D	Sum
1	3	2	1	4	10
2	3	1.5	1.5	4	10
3	1	4	2	3	10
4	3	2	1	4	10
5	4	2	2	2	10
Sum	14	11.5	7.5	17	50

TABLE 2

Counts of each ranking in Table 1

Rank (r_s)	1	1.5	2	3	4	Sum
Count (c_s)	3	2	6	4	5	20

TABLE 3

ANOVA for m missing values

Source	*df*	*SS*	*F*
Blocks (unadjusted)	$b - 1$	$\sum_{i=1}^{b} t_i(\bar{r}_{i\bullet} - \bar{r}_{\bullet\bullet})^2$	
Treatments (adjusted)	$t - 1$	*TSS* (from GLM or FIT MODEL)	$\dfrac{\{(b-1)(t-1)-m\}TSS}{(t-1)ESS}$
Error	$(b - 1)(t - 1) - m$	*ESS* (by difference)	
Total		$\sum_{i,j}(r_{ij} - \bar{r}_{\bullet\bullet})^2$	

TABLE 4

Preference ranks within judges for four food products A, B, C and D

Judge	A	B	C	D
1	1	3	2	4
2	1	3	2	4
3	2	1	3	4
4	1	2	3	4
5	1	4	2	3
6	2	3	1	-
7	-	2	1	-
8	1	3	2	4
9	1	-	2	3

TABLE 5

MINITAB data entry

ID	Rank (C1)	Product (C2)	Judge (C3)
1	1	1	1
2	1	1	2
3	2	1	3
⋮	⋮	⋮	⋮
30	3	4	5
31	4	4	8
32	3	4	9

TABLE 6

ANOVA for judges/products data

Source	*df*	*SS*	$F(T_2)$	P-value
Judges (unadjusted)	8	2.72		
products (adjusted)	3	23.64	14.52	0.0000
Error	20	10.86		
Total	31	37.22		

TABLE 7

Sizes for α = 10%, 5% and 1% when t = 4, b = 8, no tied data and missing data for layouts (i), (ii) and (iii) described in the text.

Layout	α%	T_2	T^*
(i)	10	0.105	0.092
	5	0.058	0.041
	1	0.012	0.004
(ii)	10	0.107	0.093
	5	0.059	0.039
	1	0.015	0.003
(iii)	10	0.104	0.090
	5	0.056	0.039
	1	0.012	0.004

TABLE 8

Sizes for α = 10%, 5% and 1% when t = 4, b = 8, possible tied data and missing data for layouts (i), (ii) and (iii) described in the text.

Layout	α%	T_2	T^*
(i)	10	0.106	0.042
	5	0.057	0.013
	1	0.013	4×10^{-4}
(ii)	10	0.112	0.040
	5	0.060	0.011
	1	0.015	3×10^{-4}
(iii)	10	0.106	0.043
	5	0.054	0.013
	1	0.012	0.001

TABLE 9

Powers for α = 10%, 5% and 1% when t = 4, b = 8, no tied data, missing data for layouts (i), (ii) and (iii) described in the text and alternative probabilities (a), (b) and (c) as described in the text.

Layout	Alternative	α%	T_2	T^*
(i)	(a)	10	0.198	0.177
		5	0.120	0.090
		1	0.037	0.015
	(b)	10	0.225	0.203
		5	0.142	0.111
		1	0.046	0.019
	(c)	10	0.292	0.270
		5	0.190	0.153
		1	0.069	0.033
(ii)	(a)	10	0.126	0.104
		5	0.071	0.046
		1	0.021	0.004
	(b)	10	0.213	0.182
		5	0.139	0.097
		1	0.049	0.013
	(c)	10	0.269	0.238
		5	0.179	0.132
		1	0.068	0.021
(iii)	(a)	10	0.123	0.109
		5	0.068	0.051
		1	0.020	0.006
	(b)	10	0.229	0.210
		5	0.142	0.115
		1	0.053	0.020
	(c)	10	0.297	0.276
		5	0.196	0.161
		1	0.081	0.034

TABLE 10

Powers for α = 10%, 5% and 1% when t = 4, b = 8, tied data, missing data for layouts (i), (ii) and (iii) described in the text and alternative probabilities (a), (b) and (c) as described in the text.

Layout	Alternative	α%	T_2	T^*
(i)	(a)	10	0.217	0.113
		5	0.133	0.047
		1	0.044	0.004
	(b)	10	0.274	0.101
		5	0.173	0.039
		1	0.061	0.003
	(c)	10	0.336	0.193
		5	0.229	0.092
		1	0.090	0.010
(ii)	(a)	10	0.127	0.051
		5	0.073	0.017
		1	0.021	0.000
	(b)	10	0.257	0.082
		5	0.165	0.029
		1	0.063	0.001
	(c)	10	0.312	0.164
		5	0.210	0.072
		1	0.082	0.006
(iii)	(a)	10	0.124	0.056
		5	0.069	0.020
		1	0.019	0.001
	(b)	10	0.273	0.114
		5	0.183	0.051
		1	0.011	0.005
	(c)	10	0.344	0.020
		5	0.237	0.020
		1	0.096	0.012

TABLE 11

Data from six judges on six food products ranked four at a time

Judge	Products				Ranks			
1	1	2	3	4	1	2	3.5	3.5
2	2	3	5	4	1	2	3.5	3.5
3	3	4	6	5	1	3	3	3
4	4	5	1	6	2	3	1	4
5	5	6	2	1	3.5	3.5	1	2
6	6	1	3	2	4	1	2	3

TABLE 12

ANOVA for six food products ranked by six judges four at a time

Source	df	SS	F
Judges (unadjusted)	5	0.00	-
Products (adjusted)	5	20.54	8.97
Error	13	5.96	-
Total	23	26.60	