Best, D. J.; Rayner, J. C. W.; Allingham, David "A statistical test for ranking data from partially balanced incomplete block designs" Journal of Sensory Studies Vol. 26, Issue 1, p. 81-84 (2011)

Available from: http://dx.doi.org/10.1111/j.1745-459X.2010.00324.x

Full title: A Statistical Test for Ranking Data from Partially-Balanced Incomplete Block Designs

Running title: Partially-Balanced Ranking Data

Authors: D.J. Best[1], J.C.W. Rayner[1], David Allingham[1,2]

Affiliations:
[1]School of Mathematical and Physical Sciences, The University of Newcastle, Callaghan, NSW, Australia, 2308
[2]CARMA, The University of Newcastle, Callaghan, NSW, Australia, 2308

Corresponding author:
D.J. Best
John.Best@newcastle.edu.au
Telephone +61-2-492-15747
Fax +61-2-492-16898

Abstract:
The analysis of partially-balanced incomplete block (PBIB) ranked data is discussed. Two examples are given to illustrate two alternative approaches. Analysis of PBIB ranking data is not covered in any of the standard sensory evaluation texts and this expository note is meant to help fill this gap. For some data sets the calculations for the first approach are simple enough to do by hand. The second approach that we consider assumes that computer software for general analysis of variance is available. Such analysis of variance software should cope with missing values via a regression method. A suggested multiple comparisons algorithm is also illustrated. R code is given to allow easy application of our first approach.

Practical applications:
Sensory fatigue can be a problem in some sensory evaluation trials. In the taste-test area this is sometimes called “palate paralysis”. To cope with this fatigue, balanced incomplete block designs can be employed. However, these restrict the sensory scientist to particular combinations of products, subjects and evaluations per subject. Sometimes such restrictions can be prohibitive and then partially-balanced designs which allow more freedom in the choice of these parameters can be used. Here we consider statistical analysis of ranking data from partially-balanced incomplete block designs.

Keywords:
multiple comparisons; palate paralysis; R code; sensory fatigue; tied ranks

1. Introduction

According to a number of authorities on the conduct of taste tests, it is not desirable to taste more than about four products at one tasting session. For example, Gacula et al. (2009, p. 169) say "When the panellists are judging several food items, taste fatigue occurs and may produce biased responses." Some authors refer to taste fatigue as palate paralysis. A commonly suggested solution is to use a balanced incomplete block (BIB) design. For example see Gacula et al. (2009).

For such a design, however, if one specifies the number of products to be tasted, say t, and the number to be evaluated at a tasting session, say k, then the number of tasters b, the number of repeat tastings of each product, say r, and the number of tasters who evaluate each pair of products, say λ, are all fixed by design, not the sensory scientist. To lessen this problem a solution is to use partially-balanced incomplete block designs. These are not usually discussed in sensory evaluation texts, perhaps because the computation involved in their analysis was once quite difficult. Modern computing has now made such computation routine.

The purpose of this expository note is to illustrate the Skillings and Mack approach (hereafter SM; Skillings and Mack, 1981, or Hollander and Wolfe, 1999) and analysis of variance (ANOVA) approaches to analysis of ranking data from partially-balanced incomplete block designs. We consider a simple, four-product data set and a more complicated six-product data set. Another important consideration is that partially-balanced incomplete designs reduce the number of blocks or judges needed for a reasonable design. Sections 2 and 3 illustrate the SM approach while Section 4 illustrates the ANOVA approach.

Before proceeding we emphasise that we are not generally advocating the use of partially-balanced designs in place of BIBs. The partially-balanced designs can be used if a BIB is not possible with the sensory scientist's available resources. However, if a BIB is compatible with the resources available, it should be used because (i) it is more statistically efficient and (ii) does not need a variance-covariance matrix to make pairwise comparisons.

We note that the two approaches used here can also be used for randomised complete block and BIB designs when there are missing values. Further, the two approaches can be used for any blocked unbalanced design.

2. A Simple Example

Suppose we have four products (t = 4) to be ranked two at a time (k = 2) by eight judges (b = 8), with the outcome shown in Table 1. Each product has been evaluated r = 4 times. Clearly this is an incomplete block design but it is not balanced as product 1 was not evaluated with product 3 and product 2 was not evaluated with product 4. This simple partially-balanced incomplete block design is also known as a cyclic design. If the design had been balanced an appropriate test is Durbin's well known rank test (see, for example, Gacula *et al*. 2009). Durbin's test statistic is calculated as a sum of squares but, because of the partial balance, the SM extension of Durbin's approach involves a quadratic form $a^T V^{-1} a$, say, where V is a variance-covariance matrix and a is a vector of adjusted sum of ranking totals for each product. The matrix $V = (v_{ij})$ is easily calculated. For products i and j ($i \neq j$), $-v_{ij}$ is the number of times products i and j are evaluated by the same judge. For each value of i the v_{ii} values are taken to be $-\sum_{i \neq j} v_{ij}$. Deleting the last row and column, for the Table 1 data we have

$$V = \begin{pmatrix} 4 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 4 \end{pmatrix}.$$

Using standard software, or by hand calculation, the inverse in this case is

$$V^{-1} = \begin{pmatrix} 0.375 & 0.25 & 0.125 \\ 0.25 & 0.5 & 0.25 \\ 0.125 & 0.25 & 0.375 \end{pmatrix}.$$

If R_i is the sum of the ranks of product i, then the adjusted sum, R_i^* say, and the ith element of a, is $R_i^* = \sqrt{12/(k+1)}\{R_i - (k+1)r/2\} = 2(R_i - 6)$ here. We find that $a^T = (4, -4, 4)$, where the fourth element has been dropped. Hence, $SM = a^T V^{-1} a = 8.0$ on $t - 1 = 3$ degrees of freedom. If the fourth row and column of V and the fourth element of a had not been dropped then a generalised inverse of V would have been needed. Finally, using the χ_3^2 approximation, a significant p-value of 0.046 is obtained.

Given that the p-value is less than 0.05, we might also be interested in multiple pairwise comparisons. The absolute values of the pairwise adjusted rank sum differences, that is $\left|R_i^* - R_j^*\right|$ for $i \neq j$, are given by the matrix

$$\begin{pmatrix} - & 8 & 0 & 8 \\ 8 & - & 8 & 0 \\ 0 & 8 & - & 8 \\ 8 & 0 & 8 & - \end{pmatrix}.$$

To get the corresponding least significant differences (LSDs), we need the standard errors of these differences. These can be found from the matrix V above. For example, $\text{var}(R_i^* - R_j^*) = v_{ii} + v_{jj} - 2v_{ij}$. Further, using a studentized rank statistic approach to pairwise comparisons, we also need the $(1 - \alpha)^{\text{th}}$ quantile for t products and infinite degrees of freedom. We call this quantile $q_{t,\infty}^{1-\alpha}$ and, for $\alpha = 0.10$ or 0.05, find it using, for example, Gacula *et al*. (2009, Table A6). Alternatively, using the statistical package R (R Development Core Team 2009), $q_{t,\infty}^{1-\alpha}$ may be found with the command "qtukey(1 – α, t, 10000)". For example, we find $q_{4,\infty}^{0.90} = 3.24$, and so the LSD for paired comparisons for the SM approach can be taken to be

$$\frac{1}{\sqrt{2}} q_{t,\infty}^{1-\alpha} \sqrt{\text{var}(R_i^* - R_j^*)},$$

for $i \neq j$. Observe that use of V implies pairwise comparisons are more difficult than with a BIB where only a single variance, not a matrix, is needed. This provides a further reason to use a BIB design if possible.

For the present data set and $\alpha = 0.05$ there are no significant differences. With $\alpha = 0.10$, the matrix of LSD values for pairs (i, j), $i \neq j$, to compare with $\left|R_i^* - R_j^*\right|$ is

$$\begin{pmatrix} - & 7.94 & 6.48 & 7.94 \\ 7.94 & - & 7.94 & 6.48 \\ 6.48 & 7.94 & - & 7.94 \\ 7.94 & 6.48 & 7.94 & - \end{pmatrix}.$$

Thus there are no significant pairwise comparisons at the 5% level. At the 10% level, $R_1^* - R_2^*$, $R_1^* - R_4^*$, $R_2^* - R_3^*$ and $R_3^* - R_4^*$ are significantly different: products 1 and 3 are not significantly different, and products 2 and 4 are not significantly different, and all other pairs of products are significantly different.

In our discussion above we have used a χ^2 approximation to find p-values. In our experience this is a common approach used by sensory scientists. However, we note that modern computing power allows more accurate p-values to be obtained and Bi (2009) discusses how to do this for BIB designs. It would be straightforward to adapt his approach for unbalanced designs.

3. A More Complex Example

Suppose we have t = 6 products, b = 6 judges, k = 4 products ranked (or scored and then ranked) by each judge and r = 4 ranks for each product. The data are shown in Table 2.

We find $SM = a^T V^{-1} a$ using a computer programme: an R package called `nppbib` (Allingham and Best 2010) performs this analysis (see the on-line supplement for details). Alternatively, an executable file to be run from the Microsoft Windows command line is available from the first author. We find a^T = (–7.75, –4.65, –2.32, 3.10, 4.65, 6.97), V as below and SM = 12.33 on 5 degrees of freedom with a p-value of 0.03, based on the χ_5^2 approximation. Note that for this example, and in Table 2, that the arithmetic mean of tied rankings is used.

$$V = \begin{pmatrix} 12 & -3 & -2 & -2 & -2 & -3 \\ -3 & 12 & -3 & -2 & -2 & -2 \\ -2 & -3 & 12 & -3 & -2 & -2 \\ -2 & -2 & -3 & 12 & -3 & -2 \\ -2 & -2 & -2 & -3 & 12 & -3 \\ -3 & -2 & -2 & -2 & -3 & 12 \end{pmatrix}$$

Observe that for a BIB design with t = 6 and k = 4, the sensory scientist would need 15 judges rather than six as in Table 2. Again, no significant pairwise differences occur for $\alpha = 0.05$. However, when using $\alpha = 0.10$ the first and sixth products are significantly different.

4. An Alternative Approach for Tied Data

In our more complex example of the previous section, a reasonable amount of the data were tied. In such cases we suggest not calculating SM, but rather

$$Q = b(k - 1)(t - 1)F/\{(bk - b - t + 1) + (t - 1)F\}$$

where F is the between products F-value from an ANOVA performed on the ranks. Computer software should again be used to perform the ANOVA, provided that it copes with missing values by using a regression approach. If there are no ties then $Q = SM$. We suggest the statistic Q is better adjusted to ties than SM is and plan to investigate this point more thoroughly as part of a future paper comparing <u>SM</u> and Q. The Q statistic is discussed in Desu and Raghavarao (2004, Chapter 5) and in Conover (1998, Chapter 5), where it can be derived from his equation (6) on p. 389, by noting that his T_1 is our Q and his T_2 is our F.

For the more complex data set, we find $Q = 13.96$ with a p-value of 0.02 using the χ^2_5 approximation. This compares with $SM = 12.33$ and p-value of 0.031 found previously.

For the simple data set, $Q = SM$ as there are no ties. Note that for the simple data set F is infinite, and so $Q = b(k - 1) = 8$ as before.

5. Conclusion

Two examples have illustrated how to calculate the SM and Q statistics for partially-balanced incomplete block designs. The data are ranks (or scores that are subsequently ranked) within blocks. Partial balance frees the sensory scientist from restrictions inherent in completely balanced designs. However, we advocate use of balanced designs when (i) resources allow their use and (ii) there are no missing values. Incomplete block designs are often needed in sensory evaluation applications because of sensory fatigue or palate paralysis.

Which of SM and Q is to be preferred? For smaller t and b and few tied ranks SM can be calculated by hand. For more tied data Q seems more likely to pick up significant differences but we plan to investigate this point in more detail in a future paper. Use the SM approach to give pairwise comparisons.

The discussion we have given for the partially-balanced incomplete block designs can easily be extended to any block design with ranked data or to balanced block designs with missing values.

We thank two referees and the editor for constructive suggestions.

References

ALLINGHAM, D and BEST, D.J. (2010). nppbib: Nonparametric Partially-Balanced Incomplete Block Design Analysis R package. Version 0.5-0. http://cran.r-project.org/.

BI, J. (2009) Computer-intensive methods for sensory data analysis, exemplified by Durbin's rank test. Food Quality and Preference. *20*, 195-202.

CONOVER, W.J. 1998. Practical Nonparametric Statistics. 3rd ed. Wiley, New York.

DESU, M.M. and RAGHAVARAO, D. 2004. Nonparametric Statistical Methods for Complete and Censored Data. Chapman and Hall/CRC, Boca Raton.

GACULA, M.C., SINGH, J., BI J. and ALTAN, S. 2009. Statistical Methods in Food and Consumer Research. Academic Press, New York.

HOLLANDER, M. and WOLFE, D.A. 1999. Nonparametric Statistical Methods. 2nd ed. Wiley, New York.

R DEVELOPMENT CORE TEAM. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

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

Table 1. Rankings from eight judges on four products

Judge	Products		Ranks	
1	1	2	2	1
2	2	3	1	2
3	3	4	2	1
4	1	4	2	1
5	1	2	2	1
6	2	3	1	2
7	3	4	2	1
8	1	4	2	1

Table 2. Rankings from 6 judges on 6 products

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