

# Nonparametric Tests for Two Factor Designs with an Application to Latin Squares

J.C.W. Rayner and D.J. Best

The University of Newcastle, Callaghan, NSW, 2308, AUSTRALIA  
*John.Rayner@newcastle.edu.au* and *John.Best@newcastle.edu.au*

## Abstract

We show how to construct nonparametric tests for two factor designs. These tests depend on whether or not the levels of the factors are ordered. Pearson's  $X^2$  statistic is decomposed into components of orders 1, 2, ... . These components may be further decomposed, the decomposition depending on the design. If neither factor is ordered, the components reflect linear, quadratic etc main and interaction effects. The approach is demonstrated with reference to the latin squares design.

*Keywords:* Randomised block design, Randomised complete block design, orthonormal polynomials, Pearson's  $X^2$

## 1. Introduction

The approach described here is based on components of the test statistic of the Pearson  $X^2$  test of independence. The first order component utilises ranks. Tests are available of higher order components, which can be thought of as being based on generalised ranks.

In a limited empirical assessment for the latin square design we find that our first order (rank-based) test consistently gives superior power to the parametric F test and our benchmark nonparametric test, the Conover rank transform test (see [3, p.419]).

The approach generalises readily to the development of multifactor nonparametric tests.

In section 2 we construct contingency tables and show how Pearson's  $X^2$  statistic  $X_p^2$  may be partitioned into components that reflect, for example, linear, quadratic and other effects. The components depend on how many factors have ordered levels. In section 3 we consider no factors ordered, and in section 4 at least one factor ordered. Section 5 gives a brief empirical assessment for the latin squares design.

## 2. Decomposition of the Pearson Statistic into Linear, Quadratic and Other Effects

We assume that we have observations  $x_{ij}$ ,  $i = 1, \dots, I$  and  $j = 1, \dots, J$ , in which  $i$  and  $j$  are the levels of

factors A and B respectively. All  $IJ = n$  observations are ranked and we count  $N_{rij}$ , the number of times rank  $r$  is assigned to the observation at level  $i$  of factor A and level  $j$  of factor B. For simplicity we assume throughout that there are no ties.

### 2.1 Singly Ordered Tables: Neither Factor Ordered

Initially it is assumed that only the ranks are ordered. With no ties  $\{N_{rij}\}$  defines a three-way singly ordered table of counts of zeros and ones. As in [2] and [4, section 10.2], Pearson's  $X^2$  statistic  $X_p^2$  may be partitioned into components  $Z_{uij}$  via

$$X_p^2 = \sum_{u=1}^{n-1} \sum_{i=1}^I \sum_{j=1}^J Z_{uij}^2$$

with  $Z_{uij} = \sum_{r=1}^n a_u(r) N_{rij} / \sqrt{np_{i..} p_{..j}}$ , in which  $\{a_u(r)\}$  is an orthonormal polynomial on  $\{p_{r..}\}$  with  $a_0(r) = 1$  for  $r = 1, \dots, n$ . Here the standard dot notation has been used, so that, for example,  $N_{...} = IJ = n$ , the number of times a rank has been assigned. Formally  $X_p^2$  also includes a term for Pearson's  $X^2$  for the unordered table formed by summing over  $r$ :  $\{N_{.ij}\}$ . However this table has every entry one, and  $X^2$  is zero. We also find that  $N_{.i.} = J$  and  $N_{.j.} = I$ . It follows that  $p_{i.} = 1/I$  and  $p_{.j} = 1/J$ , giving  $Z_{uij} = \sum_{r=1}^{IJ} a_u(r) N_{rij}$ .

For  $u = 1, 2, \dots, n - 1$  define

$$SS_u = \sum_{i=1}^I \sum_{j=1}^J Z_{uij}^2$$

so that  $X_p^2 = SS_1 + \dots + SS_{n-1}$ ; the  $SS_u$  give order  $u$  assessments of factor effects.

The  $\{Z_{uij}\}$  may be thought of as akin to Fourier coefficients: for each  $(i, j)$  pair  $Z_{uij}$  is the projection of  $x_{ij}$  into  $[n - 1]$  dimensional ‘order’ space, where the first dimension reflects, roughly, location, and the second reflects, roughly, dispersion, and so. Now  $Z_{1ij} = \sum_{r=1}^n (r - \mu) N_{rij} / \sigma$  in which  $\mu = (n + 1)/2$  and  $\sigma^2 = (n^2 - 1)/12$ . The linear or location statistic is  $SS_1 = \sum_{i,j} Z_{1ij}^2$ . As in [4, section 3.4] this is of the form of a Kruskal-Wallis test.

### 2.2 Doubly Ordered Tables: One Factor Ordered

Now assume that the first factor is ordered. To reflect this change write  $N_{rsj}$  for the number of times rank  $r$  is assigned to the factor combination  $(s, i)$ . As there are no ties  $\{N_{rsj}\}$  defines a three-way doubly ordered table of counts of zeros and ones. As in [2] and [4, section 10.2], Pearson’s  $X^2$  statistic  $X_p^2$  may be partitioned into components  $Z_{uvj}$  via

$$X_p^2 = \sum_{u=1}^{n-1} \sum_{v=1}^{I-1} \sum_{j=1}^J Z_{uvj}^2 + \sum_{v=1}^{I-1} \sum_{j=1}^J Z_{0vj}^2 + \sum_{u=1}^{n-1} \sum_{j=1}^J Z_{u0j}^2$$

with  $Z_{uvj} = \sum_{r=1}^n \sum_{s=1}^I a_u(r) b_v(s) N_{rsj} / \sqrt{np_{..j}}$ , in which  $\{a_u(r)\}$  is orthonormal on  $\{p_{r..}\}$  with  $a_0(r) = 1$  for  $r = 1, \dots, n$  and  $\{b_v(s)\}$  is orthonormal on  $\{p_{.s.}\}$  with  $b_0(s) = 1$  for  $s = 1, \dots, I$ . We find that  $N_{...} = n$ ,  $p_{r..} = 1/n$ ,  $p_{.s.} = 1/I$  and  $p_{.j} = 1/J$ , giving  $Z_{uvj} = \sum_{r=1}^n \sum_{s=1}^I a_u(r) b_v(s) N_{rsj} / \sqrt{I}$ . If for  $u = 0, 1, 2, \dots, n - 1$  and  $v = 0, 1, \dots, I - 1$ , but not  $(u, v) = (0, 0)$ ,  $SS_{uv} = \sum_{j=1}^J Z_{uvj}^2$ , we have  $X_p^2 = \sum_{u,v} SS_{uv}$ .

Analogous to [4, section 6.5] the  $Z_{1ij}$  are Page test statistics at each of the levels of factor B, and the  $Z_{uvj}$  are extensions of Page’s test statistic. Now  $SS_{uv} = \sum_j Z_{uvj}^2$  gives an aggregate assessment over the whole table of order  $(u, v)$  effects, generalised correlations in the sense of [5]. As above, the aggregation of all these order  $(u, v)$  effects is  $X_p^2$ .

### 2.3 Completely Ordered Tables: Both Factors Ordered

Finally assume that both factors are ordered. To reflect this change write  $N_{rst}$  for the number of times rank  $r$  is assigned to the factor combination  $(s, t)$ . With no ties  $\{N_{rst}\}$  defines a three-way completely

ordered table of counts of zeros and ones. As in [1] and [4, section 10.2], Pearson’s  $X^2$  statistic  $X_p^2$  may be partitioned into components  $Z_{uvw}$  via

$$X_p^2 = \sum_{u=1}^{n-1} \sum_{v=1}^{I-1} \sum_{w=1}^{J-1} Z_{uvw}^2 + \sum_{v=1}^{I-1} \sum_{w=1}^{J-1} Z_{0vw}^2 + \sum_{u=1}^{n-1} \sum_{w=1}^{J-1} Z_{u0w}^2 + \sum_{u=1}^{n-1} \sum_{v=1}^{I-1} Z_{uv0}^2$$

with  $Z_{uvw} \sqrt{np} = \sum_{r=1}^n \sum_{s=1}^I \sum_{t=1}^J a_u(r) b_v(s) c_w(t) N_{rst}$ , in which  $\{a_u(r)\}$  is orthonormal on  $\{p_{r..}\}$  with  $a_0(r) = 1$  for  $r = 1, \dots, n$ ,  $\{b_v(s)\}$  is orthonormal on  $\{p_{.s.}\}$  with  $b_0(s) = 1$  for  $s = 1, \dots, I$  and  $\{c_w(t)\}$  is orthonormal on  $\{p_{..t}\}$  with  $c_0(t) = 1$  for  $t = 1, \dots, J$ .

In our previous notation  $SS_{uvw} = Z_{uvw}^2$  for  $u = 0, 1, 2, \dots, n - 1$  and  $v = 0, 1, \dots, I - 1$ , and  $w = 0, 1, \dots, J - 1$ , but not  $(u, v, w) = (0, 0, 0)$ . Thus  $X_p^2 = \sum_{u,v,w} SS_{uvw}$ . The  $SS_{uvw}$  may be thought of as further extensions of the Page test statistic, this time to three dimensions. The  $SS_{uv0}$ ,  $SS_{u0w}$  and  $SS_{0vw}$  are the familiar two-dimensional generalised Page test statistics as, for example, in [4, section 6.5 and Chapter 8].

## 3. Factors Not Ordered

Recall now that in the two factor analysis of variance without replication with observations  $y_{ij}$ ,  $i = 1, \dots, I$  and  $j = 1, \dots, J$ , the total sum of squares  $SS_{\text{Total}} = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2$  may be arithmetically partitioned into sum of squares due to factor A, namely  $SS_A = J \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$ , due to factor B, namely  $SS_B = I \sum_j (y_{.j} - \bar{y}_{..})^2$ , and a residual or interaction sum of squares  $SS_{AB} = \sum_{i,j} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$ . Thus

$$SS_{\text{Total}} = SS_A + SS_B + SS_{AB}.$$

Here  $y_{i.} = \sum_j y_{ij}$  and  $\bar{y}_{i.} = y_{i.}/J$  etc as usual.

For each  $u = 1, 2, \dots, n - 1$  put  $y_{ij} = Z_{uij} = \sum_{r=1}^n a_u(r) N_{rij}$  in  $SS_{\text{Total}}$ . The order  $u$  factor A sum of squares is  $SS_{uA} = \sum_i Z_{ui.}^2 / J - Z_{u..}^2 / (IJ)$ . As in [4, section 3.4],  $SS_{1A}$  is the Kruskal-Wallis test statistic for factor A, and for general  $u$  the  $SS_{uA}$  are the component test statistics discussed there. Clearly the  $SS_{uB}$  are the parallel generalised Kruskal-Wallis test statistics for factor B, while the  $SS_{uAB}$  are nonparametric tests for generalised interaction effects. For example, for  $u = 2$ ,  $SS_{2AB}$  assesses whether or not the

quadratic (dispersion) factor A effects are the same at different levels of factor B.

*Examples.*

The completely randomised design can be accessed either by calculating the  $SS_{uA}$  directly or by partitioning the  $\{Z_{uij}\}$  as in the one factor ANOVA. However it is done, the usual Kruskal-Wallis test statistic and its extensions are obtained.

In the randomised block design factor A can be taken to be treatments and factor B blocks. Of course there is no interest in testing for a block effect or a treatment by block interaction effect. The treatment effect test is not the Friedman test, as observations are ranked overall, not merely on each block. From an overall ranking the ranks on each block may be derived, so there is more information assumed in this approach. This could result in more power when the test is applicable. In some situations only ranks within blocks are available.

**4. At least One Factor Ordered**

Suppose now that the first factor is ordered. The  $Z_{uvj} = \sum_{r=1}^n \sum_{s=1}^t a_u(r) b_v(s) N_{rsj} / \sqrt{I}$ , are generalised Page test statistics at each level of factor B. As in the Happiness example in [4, p. 147 and p. 188],  $X_p^2$  may be partitioned into meaningful components. An alternative is to sum over the levels of factor B and obtain  $Z_{uv}$ , generalised Page test statistics aggregating over factor B. This is appropriate when factor B is replicates, as in the completely randomised design, or blocks, as in the randomised block design

If both factors are ordered  $X_p^2$  is partitioned by the  $SS_{uvw}$  of section 2.3. These are new extensions of the Page test, this time to three dimensions.

**5. Latin Squares**

The parametric analysis of the  $t \times t$  latin square design partitions the total sum of squares into sum of squares of treatments, rows and columns and error. For the nonparametric analysis we assume that neither rows nor columns are ordered and investigate parallel partitions of the total sum of squares.

We count  $N_{rjk}$ , the number of times rank  $r$  is assigned to the treatment in row  $j$  and column  $k$ , with  $r = 1, \dots, t^2, j, k = 1, \dots, t$ . Note that treatment  $i, i = 1, \dots, t$ , occurs in cells  $(j, k)$  specified by the design. As long as we know any two of the treatment, row and column, we know the other. Hence a latin square

may be considered to be any of three two factor designs. This observation is utilised subsequently.

Throughout this section we assume that treatments are unordered, and that only the ranks are ordered. With no ties  $\{N_{rjk}\}$  defines a three way singly ordered table of counts of zeroes and ones.

As in section 2,  $X_p^2 = SS_1 + \dots + SS_{t^2-1}$  in which

$$SS_u = \sum_{j=1}^t \sum_{k=1}^t Z_{ujk}^2 \text{ for all } u$$

$$\text{with } Z_{ujk} = \sum_{r=1}^{t^2} a_u(r) N_{rjk}.$$

The factor A test statistic of order  $u = 1, \dots, t^2 - 1$ , can be denoted by  $SS_{uA}$ , a generalised Kruskal-Wallis test statistic. By letting the factors be in turn rows and columns, columns and treatments, and treatments and rows, we are able to show that

$$3 SS_u = 2 SS_{utreatments} + 2 SS_{urows} + 2 SS_{ucolumns} + SS_{utreatments \times rows} + SS_{utreatments \times columns} + SS_{urows \times columns}.$$

In most applications it is enough to know that  $SS_u = SS_{utreatments} + \text{residual}$ , but it is interesting to know that, parallel to the parametric partition, the residual could be used to assess rows and columns, and interactions between treatments, rows and columns. However, unlike the parametric case, this analysis applies for any order. We recognise that in most applications few users would be interested in treatment effects beyond orders two or three.

*Empirical Study*

We now briefly assess the power properties of some of the tests constructed. Treatments tests of orders one and two, with test statistics denoted by  $SS_{1T}$  and  $SS_{2T}$  respectively, are considered. We also consider tests formed from the table of counts  $\{N_{rsi}\}$  where the second category is treatments, assumed to be ordered. Then test statistics  $S_{uv}$  are constructed from  $\{N_{rs}\}$ , particularly the Page test based on  $S_{11}$  and the umbrella test based on  $S_{21}$ . These will be compared with the parametric F test (denoted by F) and the Conover rank transform test (denoted by CRT) that ranks the data and applies a parametric F test to the ranks.

Only the  $5 \times 5$  Latin square is considered, and rather than use asymptotic critical values 5% critical values are found using random permutations. The critical value for  $SS_{1T}$  is 8.9059 while that for the CRT test was 3.3642. Compare these with the asymptotic critical values of 9.4877 using the  $\chi_4^2$  distribution for the  $SS_{1T}$  test and 3.2592 using the  $F_{4,12}$  distribution for the CRT test. Not surprisingly these asymptotic critical values aren't practical for a

table of this size. However the critical value for the parametric F test is exact.

Table 5. Test sizes for competitor tests for various error distributions.

Error distn	CRT	SS <sub>1T</sub>	F	SS <sub>2T</sub>	S <sub>11</sub>	S <sub>21</sub>
Normal	0.050	0.049	0.050	0.050	0.051	0.049
Expon	0.050	0.050	0.040	0.049	0.050	0.051
U(0, 1)	0.052	0.052	0.055	0.049	0.052	0.051
Cauchy (t <sub>1</sub> )	0.049	0.049	0.017	0.050	0.051	0.051
t <sub>2</sub>	0.049	0.050	0.031	0.050	0.052	0.051
t <sub>3</sub>	0.050	0.050	0.041	0.050	0.051	0.051
Lognormal	0.049	0.049	0.032	0.049	0.052	0.050

Table 6. Powers for competitor tests for various error distributions with linear alternatives  $\alpha_i = (-1, -0.5, 0, 0.5, 1)$ .

Error distn	CRT	SS <sub>1T</sub>	F	SS <sub>2T</sub>	S <sub>11</sub>	S <sub>21</sub>
Normal	0.62	0.69	0.64	0.07	0.91	0.02
Expon	0.78	0.83	0.68	0.22	0.96	0.01
Cauchy (t <sub>1</sub> )	0.19	0.22	0.05	0.07	0.40	0.04
t <sub>2</sub>	0.32	0.37	0.20	0.07	0.62	0.03
t <sub>3</sub>	0.40	0.45	0.32	0.06	0.72	0.02
Lognormal	0.54	0.59	0.29	0.22	0.84	0.02

Table 7. Powers for competitor tests for various error distributions with quadratic alternatives  $\alpha_i = (1, 0, -2, 0, 1)$ .

Error distn	CRT	SS <sub>1T</sub>	F	SS <sub>2T</sub>	S <sub>11</sub>	S <sub>21</sub>
Normal	0.94	0.97	0.96	0.34	0.01	0.98
Expon	0.93	0.95	0.94	0.48	0.01	0.99
Cauchy (t <sub>1</sub> )	0.34	0.38	0.10	0.11	0.03	0.52
t <sub>2</sub>	0.59	0.66	0.44	0.16	0.03	0.77
t <sub>3</sub>	0.71	0.78	0.65	0.19	0.02	0.86
Lognormal	0.74	0.78	0.57	0.43	0.01	0.91

Table 8. Powers for competitor tests for various error distributions with complex alternatives  $\alpha_i = (0.5, -0.5, 0, 0.5, -0.5)$ .

Error distn	CRT	SS <sub>1T</sub>	F	SS <sub>2T</sub>	S <sub>11</sub>	S <sub>21</sub>
Normal	0.27	0.31	0.28	0.04	0.07	0.04
Expon	0.46	0.52	0.33	0.11	0.09	0.03
Cauchy (t <sub>1</sub> )	0.11	0.12	0.03	0.05	0.06	0.05
t <sub>2</sub>	0.16	0.17	0.09	0.05	0.07	0.05
t <sub>3</sub>	0.18	0.21	0.14	0.05	0.07	0.05
Lognormal	0.30	0.34	0.13	0.15	0.08	0.03

All simulations relate to 5% level tests with sample sizes of 25, and are based on 100,000 simulations. The error distributions are Normal, exponential, uniform (0, 1), Cauchy (t<sub>1</sub>), t<sub>2</sub>, t<sub>3</sub> and lognormal.

Using the simulated critical values we found the test sizes given in Table 5. They are remarkably close to the nominal significance level, as befits nonparametric tests. However the parametric F test fared less well, often having test size less than 5%. This will mean the corresponding powers will be less than if the nominal level was achieved. Nevertheless, this is how the test would be applied in practice.

The critical values used in Table 5 were also used to estimate powers in subsequent tables. These powers use the model  $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + E_{ijk}$  but with  $\beta_j = \gamma_k = 0$  for all  $j$  and  $k$  in this study. The uniform error distribution doesn't appear in Tables 6 to 8 as all powers are 1.00.

Even when normality holds, the test based on SS<sub>1T</sub> is, for the alternatives considered, slightly superior to the parametric F test, and clearly superior when normality doesn't hold. This linear effects test is also uniformly slightly superior to the Conover rank transform test. This is not due to a size difference as can be seen from Table 5. The Page and umbrella tests perform well when the alternative is constructed to reflect their designed strengths, but both are sometimes biased: their power is less than their test size. The performance of the test based on SS<sub>2T</sub> is disappointing, perhaps because powers have not been given for alternatives constructed to reflect their designed strengths.

References

[1] Beh, E. J. and Davy, P. J. (1998). Partitioning Pearson's chi-squared statistic for a completely ordered three-way contingency table. *Australian and NZ Journal of Statistics*, 40, 465-477.

[2] Beh, E. J. and Davy, P. J. (1999). Partitioning Pearson's chi-squared statistic for a partially ordered three-way contingency table. *Australian and NZ Journal of Statistics*, 41, 233-246.

[3] Conover, W.J. (1998). *Practical Nonparametric Statistics* (3<sup>rd</sup> ed.). New York: Wiley.

[4] Rayner, J.C.W. and Best, D.J. (2001). *A Contingency Table Approach to Nonparametric Testing*. Boca Raton: Chapman & Hall/CRC.

[5] Rayner, J.C.W. and Beh, Eric J. (2009). Towards a Better Understanding of Correlation. *Statistica Neerlandica*. 63(3), 324-333.