Australian & New Zealand Journal of Statistics

Aust. N. Z. J. Stat. **55**(3), 2013, 305–319 doi: 10.1111/anzs.12041

EXTENDED ANOVA AND RANK TRANSFORM PROCEDURES

J.C.W. Rayner[1,2,*] and D.J. Best[2]

University of Wollongong and University of Newcastle

Summary

The rank transform procedure is often used in the analysis of variance when observations are not consistent with normality. The data are ranked and the analysis of variance is applied to the ranked data. Often the rank residuals will be consistent with normality and a valid analysis results. Here we find that the rank transform procedure is equivalent to applying the intended analysis of variance to first order orthonormal polynomials on the rank proportions. Using higher order orthonormal polynomials extends the analysis to higher order effects, roughly detecting dispersion, skewness etc. differences between treatment ranks. Using orthonormal polynomials on the original observations yields the usual analysis of variance for the first order polynomial, and higher order extensions for subsequent polynomials. Again first order reflects location differences, while higher orders roughly detect dispersion, skewness etc. differences between the treatments.

Key words: complete randomised block designs; factorial designs; generalised ranks; orthonormal polynomials; ranks.

1. Introduction

Our focus is multifactor experimental designs including hierarchical designs and Latin square designs, typically modelled by a linear model that assumes normal errors and analysed by a standard analysis of variance (ANOVA). When normality is in doubt various remedies have been proposed in the literature. One such is the rank transform procedure whereby the ANOVA is applied to the ranks of the observations. See, for example, Conover & Imran (1981), Toothaker & Newman (1994), Marden & Muyot (1995) and Mansouri (1999).

The extended rank transform procedure for balanced designs developed here is based on constructing a table of counts for the mid-ranks of the data. From this contingency table a device of Beh & Davy (1998, 1999) is used to partition into components the test statistic X_P^2 of the Pearson test for independence. These components are shown to be equivalent to orthonormal polynomials. They may be analysed using the intended ANOVA. Smooth models are used to show that by applying the ANOVA to the orthonormal polynomials of different orders in turn, effects of order one, two, three, etc. may be assessed, giving a deeper scrutiny of the data than is possible by applying only the ANOVA to the observations. Order r reflects, roughly, rth moment differences between the treatments. The word 'roughly' here is

*Author to whom correspondence should be addressed.

[1]Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia.

e-mail: john.rayner@newcastle.edu.au

[2]School of Mathematical and Physical Sciences, University of Newcastle, NSW 2308, Australia.

important. It reflects the discussion late in section 2.4: an order r effect may reflect moment differences between the treatments of all orders *up to* order r.

For unbalanced designs the original motivation, that the ANOVA analyses of different orders are aspects of testing for independence of a particular contingency table, is lost. However, as for the balanced design analyses, smooth models may be used to show that applying the ANOVAs to the orthonormal polynomials of successive orders tests for important main and interaction effects of first, second, and higher orders.

An entirely parallel development is possible when, instead of using ranks, the original observations are analysed by the ANOVA. The first order orthonormal polynomial produces the usual ANOVA. The higher order orthonormal polynomials give extensions of order two, three, etc. These ANOVAs detect main and interaction effects reflecting differences in treatment dispersion, skewness etc. It is interesting to align conclusions of various orders about treatments with conclusions of various orders about ranks of treatments. See the examples in Sections 2.5 and 3.3.

We stress that although the construction of tables of counts for the mid-ranks of the data motivates the subsequent modelling and analysis, it is not necessary to actually construct such tables in order to apply the procedures. All that is required is to replace the observations (or their mid-ranks) by easily constructed orthonormal polynomials of the desired order and apply the intended ANOVA to these new observations.

The validity of the ANOVAs depends on the correctness of the assumptions. Nevertheless, in examples we have examined, the p-values based on the usual F tests are very close to those based on permutation tests, even when normality tests on the residuals have quite small p-values.

For the comfort of the reader most of the exposition here focuses on two-factor designs, but the ideas generalise readily to multifactor designs.

In Section 2 we focus on two-factor designs that are balanced in the sense of having an equal number of replicates in each cell. Section 3 permits unequal replicates in each cell. Each of these two sections concludes with an example. Section 4 outlines the procedure for multifactor designs. A brief concluding discussion follows.

2. The extended rank transform procedure for two-factor designs with equal replications

The rank transform procedure is mainly intended for when the normality assumption in an ANOVA is not valid. If the research interest is to assess whether or not treatments are similar, then the ANOVA only assesses equality of treatment means and the rank transform procedure assesses equality of treatment mean ranks. The approach here also assesses higher order effects, such as dispersion, skewness, etc.

It is assumed that transformations, possibly including alignment, of the original data have already been applied.

Replications need to be treated slightly differently from factors: the number of replicates depends on the number of observations at each combination of levels while the number of levels of each factor is, of course, fixed. However in this section we assume equal replicates; unequal replicates require a different treatment that will be deferred to the next section. Ties are handled using mid-ranks merely for the sake of being definite; other approaches could be used just as validly.

Independence is a natural nonparametric expression of lack of structure, such as linear, quadratic and other relationships between the variables. Indirectly it is this for which we seek to test. A convenient test for independence is the well-known Pearson test.

2.1. Decomposition of Pearson's X_{P}^2

Assume that we have n observations $x_{ijk}, i = 1, \ldots, I, j = 1, \ldots, J$ and $k = 1, \ldots, K$: there are K replicates at level i of factor A and level j of factor B. All observations are ranked and we count N_{rijk}, the number of times the rth of R mid-ranks is assigned to replicate k at level i of factor A and level j of factor B. Thus N_{rijk} is zero unless the rth mid-rank is assigned to this level combination, in which case it is one. Subsequently we give an arithmetic decomposition of the Pearson test statistic X_{P}^2 used to test for independence for the table $\{N_{rijk}\}$.

Standard dot notation has been used, so that, for example, $N_{\bullet\bullet\bullet\bullet} = IJK = n$, the number of times a rank has been assigned and the number of observations. For all r, i, j and k write $p_{rijk} = N_{rijk}/n$. We also note in passing that throughout the paper our methods can handle missing observations (usually a design issue) and $K = 1$ (often referred to as 'no replication').

It is assumed that none of the factors have ordered levels, or if they have then any ordering is ignored. As only the ranks are ordered, results for partially ordered tables in Beh & Davy (1999) and Rayner & Best (2001, section 10.2) may be applied. The contingency table $\{N_{rijk}\}$ defines a four-way singly ordered table for which Pearson's statistic X_{P}^2, calculated as $\sum_{\text{allcells}}(\text{observed} - \text{expected})^2/\text{expected}$, may be partitioned into components Z_{uijk} via

$$X_{\mathrm{P}}^2 = \sum_{u=0}^{R-1} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} Z_{uijk}^2.$$

Here $Z_{uijk} = \sum_{r=1}^{R} a_u(r) N_{rijk} (np_{\bullet i \bullet\bullet} p_{\bullet\bullet j \bullet} p_{\bullet\bullet\bullet k})^{-0.5}$, in which $\{a_u(r)\}$ is orthonormal on $\{p_{r\bullet\bullet\bullet}\}$ with $a_0(r) = 1$ for $r = 1, \ldots, R$. Note that $N_{\bullet i \bullet\bullet} = JK$, $N_{\bullet\bullet j \bullet} = IK$ and $N_{\bullet\bullet\bullet K} = IJ$. It follows that $p_{\bullet i \bullet\bullet} = 1/I$, $p_{\bullet\bullet j \bullet} = 1/J$ and $p_{\bullet\bullet\bullet k} = 1/K$, giving $Z_{uijk} = \sum_{r=1}^{IJk} a_u(r) N_{rijk}$. The Z_{uijk} may be thought of as akin to Fourier coefficients: for each (i, j, k) triple Z_{uijk} may be thought of as a projection of x_{ijk} into $(n - 1)$ dimensional 'order' space, where the first dimension reflects, with the caveat in section 2.4, location, and the second reflects, again with the same caveat, dispersion, and so on. We now show how to utilise these projections for testing purposes.

For $u = 0, 1, \ldots, R - 1$ we may define

$$SS_u = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} Z_{uijk}^2,$$

so that $X_{\mathrm{P}}^2 = SS_0 + SS_1 + \ldots + SS_{R-1}$. The term SS_0 is the Pearson statistic for the unordered three-way table formed by summing over r: $\{N_{\bullet rijk}\}$. For $u \geqslant 1$ the sum of squares SS_u gives an order u assessment of factor effects. Most interest will be on low order effects, specifically orders one, two and occasionally three.

For a given factor/replicate combination (i, j, k) only once is N_{rijk} non-zero and in this case it is one. This occurs for the unique value of r for which this combination takes its appropriate mid-rank. Thus for $u = 1, \ldots, R - 1$, $\{Z_{uijk} = \{a_u(r)\}$. In the next sections we argue that for each u the ANOVA should be applied to $\{Z_{uijk}\}$; justification for this is best seen in relation to the Zs but note that applying an ANOVA to $\{Z_{uijk}\}$ is equivalent to applying it to the order u orthonormal polynomial $\{a_u(r)\}$.

For this balanced design if there are no ties, then X_{P}^2 and the sums of squares SS_u are constant and thus these statistics cannot be used for inference (see Appendix A). If there are ties or unequal replicates, these simplifications do not hold.

2.2. An arithmetic partition

Recall that in the balanced two factor analysis of variance with equal replication and observations y_{ijk}, $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$, the total sum of squares $SS_{\mathrm{Total}} = \sum_{i,j,k}(y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2$ may be arithmetically partitioned into sum of squares due to factor A, namely $SS_{\mathrm{A}} = J\sum_i(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$, due to factor B, namely $SS_{\mathrm{B}} = I\sum_j(y_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$, an interaction sum of squares $SS_{\mathrm{AB}} = \sum_{i,j}(\bar{y}_{ij\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$, and an error sum of squares SS_{E}, usually obtained by difference using the identity

$$SS_{\mathrm{Total}} = SS_{\mathrm{A}} + SS_{\mathrm{B}} + SS_{\mathrm{AB}} + SS_{\mathrm{E}}.$$

For each $u = 1, 2, \ldots, n-1$ in turn put $y_{ijk} = Z_{uijk}$ in the ANOVA partition to produce order u sums of squares for factor A, factor B and factors A and B together.

As in section 2.1, N_{rijk} is an indicator function for the mid-rank of the combination (i, j, k). Thus when non-zero $Z_{1ijk} = a_1(r)N_{rijk}$ is the standardised rank of the level combination (i, j, k), and $Z_{1i\bullet\bullet}$ is the standardised rank sum for level i of factor A. As above, the SS_u partition the X_{P}^2 and can themselves be partitioned into SS_{uA}, SS_{uB} and SS_{uAB} with

$$SS_{u\mathrm{A}} = \sum_i Z_{ui\bullet\bullet}^2/(JK), SS_{u\mathrm{B}} = \sum_j Z_{u\bullet j\bullet}^2/(IK), \text{ and}$$
$$SS_{u\mathrm{AB}} = \sum_{i,j} Z_{uij\bullet}^2 - \sum_i Z_{ui\bullet\bullet}^2/(JK) - \sum_j Z_{u\bullet j\bullet}^2/(IK).$$

As in Rayner & Best (2001, section 3.4) SS_{1A} is identified as the Kruskal–Wallis test statistic for factor A, so that for general u the SS_{uA} may be considered to be extensions of the Kruskal–Wallis test statistic. Similarly the SS_{uB} and SS_{uAB} also may be considered to be extensions of the Kruskal–Wallis test statistic.

In as much as Z_{1ijk} is the standardised rank for the (i, j, k) combination and Z_{uijk} is the generalised order u rank for the (i, j, k) combination, the SS_{uA} involve generalised ranks. While SS_{1A} is a sum of squares due to the ranks of factor A, SS_{uA} is a sum of squares due to the generalised ranks of factor A. Similarly the SS_{uB} are sums of squares due to the generalised ranks of factor B, while the SS_{uAB} is a sum of squares due to generalised ranks of both factors.

To this point we have a purely arithmetic decomposition of X_{P}^2. While designs have been specified, models and hypotheses have not.

2.3. Modelling and analysis

Rayner & Beh (2009) give smooth models for the contingency tables such as $\{N_{rijk}\}$. In general the normality of the components for large sample sizes follows from the central limit theorem, but the smooth models are required to give distribution theory for certain sums of squares and to show that they are appropriate test statistics. For example, the degrees of freedom correspond to the number of parameters in the model.

Rayner & Beh (2009) discuss a product multinomial model where for each $r, r = 1, \ldots, R$, we have a multinomial with parameters 1 and $p_{rijk} = p_{r\bullet\bullet\bullet} \sum_{u=0}^{R-1} \theta_{uijk} a_u(r)$ in which the θ_{uijk} are real valued parameters with $\theta_{0ijk} = 1$. Here θ_{uijk} reflects the order u effect corresponding to replicate k of level i of factor A and level j of factor B and $\{a_u(r)\}$ is orthonormal on $\{p_{r\bullet\bullet\bullet}\}$ with $a_0(r) = 1$ for $r = 1, \ldots, R$. On taking expectation with respect to $\{p_{r\bullet\bullet\bullet}\}$ we find $E(Z_{uijk}) = \theta_{uijk}$.

It follows as in Rayner & Best (1986, Proof of Theorem 1) that the Z_{uijk} are all uncorrelated with unit variance under the null hypothesis, and, in general, asymptotically if we assume that all θ_{uijk} are $\mathrm{O}(n^{-0.5})$. For more details see Appendix B.

Reparametrise by putting, for $u = 1, \ldots, R-1$,

$$\theta_{uijk} = \mu_u + A_{ui} + B_{uj} + (AB)_{uij} \text{ in which } \sum\nolimits_{i=1}^{I} A_{ui} = \sum\nolimits_{j=1}^{J} B_{uj} = \sum\nolimits_{i=1}^{I} (AB)_{uij} (\text{for each } j) = \sum\nolimits_{j=1}^{J} (AB)_{uij} (\text{for each } i) = 0.$$

For each such u there are $IJ - 1$ independent parameters; the μ_u are identically zero and are only included to complete the analogy with traditional ANOVA models. It now follows that, by the usual development, asymptotically

$$\mathrm{E}(SS_{u\mathrm{A}}) = I - 1 + J \sum\nolimits_{i=1}^{I} A_{ui}^2,\ \mathrm{E}(SS_{u\mathrm{B}}) = J - 1 + I \sum\nolimits_{j=1}^{J} A_{uj}^2,$$
$$\text{and } \mathrm{E}(SS_{u\mathrm{AB}}) = (I-1)(J-1) + IJ \sum\nolimits_{i=1}^{I} (AB)_{uij}^2.$$

It follows that we can, for example, test H_{uA}: $(A_{ui}) = 0$ against K_{uA}: not $\mathrm{H}_{u\mathrm{A}}$ using $SS_{u\mathrm{a}}$ or F ratios in the ANOVA if the usual assumptions, such as the residuals being consistent with normality, are satisfied. As in Rayner & Best (2001, section 3) the $SS_{u\mathrm{A}}$, $SS_{u\mathrm{B}}$ and $SS_{u\mathrm{AB}}$ all asymptotically have χ^2 distributions. Being asymptotic, p-values using χ^2 distributions are less reliable than resampling p-values and p-values obtained from F ratios in a valid ANOVA.

2.4. Interpretation

It is helpful to note that, in general, the first four orthonormal polynomials of a random variable X are given by

$$a_0(x) = 1 \text{ for all } x, a_1(x) = (x-\mu)/\sqrt{\mu_2},$$
$$a_2(x) = \{(x-\mu)^2 - \mu_3(x-\mu)/\mu_2 - \mu_2\}/\sqrt{d} \text{ and}$$
$$a_3(x) = \{(x-\mu)^3 - a(x-\mu)^2 - b(x-\mu) - c\}/\sqrt{e},$$

in which, with $d = \mu_4 - \mu_3^2/\mu_2 - \mu_2^2$,

$$a = (\mu_5 - \mu_3\mu_4/\mu_2 - \mu_2\mu_3)/d, b = (\mu_4^2/\mu_2 - \mu_2\mu_4 - \mu_3\mu_5/\mu_2 + \mu_3^2)/d,$$
$$c = (2\mu_3\mu_4 - \mu_3^3/\mu_2 - \mu_2\mu_5)/d, \text{ and}$$
$$e = \mu_6 - 2a\mu_5 + (a^2 - 2b)\mu_4 + (2b - c)\mu_3 + (b^2 + 2ac)\mu_2 + c^2.$$

Here μ is the mean and μ_r, $r = 2, 3, \ldots$ are the central moments of X. If the explicit formulae are inconvenient, or further orthonormal functions are required, the recurrence formulae in Rayner, Thas & De Boeck (2008) may be used.

TABLE 1

Howell (2012) data.

Age	Counting	Rhyming	Adjective	Imagery	Intention
Old	9, 8, 6, 8, 10, 4, 6, 5, 7, 7	7, 9, 6, 6, 6, 11, 6, 3, 8, 7	11, 13, 8, 6, 14, 11, 13, 13, 10, 11	12, 11, 16, 11, 9, 23, 12, 10, 19, 11	10, 19, 14, 5, 10, 11, 14, 15, 11, 11
Young	8, 6, 4, 6, 7, 6, 5, 7, 9, 7	10, 7, 8, 10, 4, 7, 10, 6, 7, 7	14, 11, 18, 14, 13, 22, 17, 16, 12, 11	20, 16, 16, 15, 18, 16, 20, 22, 14, 19	21, 19, 17, 15, 22, 16, 22, 22, 18, 21

TABLE 2

p-values for the parametric analysis (ANOVA on the observations) and the rank transform procedure (ANOVA on the ranks) for the Howell data.

	p-values	
Source	Parametric analysis	Rank transform procedure
Age	0.000	0.000
Recall condition	0.000	0.000
Interaction	0.000	0.002
Shapiro–Wilk	0.027	0.273

Note that, since $a_1(X)$ is a location-scale transformation of X, and the ANOVA is location-scale invariant, the interpretations of the first order analyses applied to the data and to the ranks are identical to the parametric and rank transform interpretations.

Moreover, and most importantly, $\sum_j a_1(x_j)/n = (\bar{x} - \mu)/\sigma$ and first order effects reflect differences between the effect means and the estimate of the overall population mean. If a first order effect is not significant, then the effect and population means are consistent and $\sum_j a_2(x_j)/n$ is approximately proportional to $\sum_j (x_j - \mu)^2/n - \mu_2$. The second order effects then reflect differences between the effect variances and the population variance. If a first order effect is significant then the corresponding second order effects reflect differences in the effect and population moments *up to* the second order. Similar comments apply to higher order effects.

2.5. Howell example

The data in Table 1 are given by Howell (2012, p. 415), who only wanted to compare 'old' with 'young'. The response is the number, out of 27, of words recalled under different conditions. Subjects were classified as either old (aged between 55 and 65) or young (aged between 18 and 30) and 50 subjects from each age group were randomly assigned into five groups of 10 each. Each group recalled words under different conditions.

An initial investigation might first look at the usual parametric analysis; see the first column in Table 2. As the residuals are not consistent with normality at the 0.05 level according to the Shapiro–Wilk test, it would be reasonable to apply the rank transform procedure; see the second column in Table 2.

Both analyses have highly significant main effects and interactions significant at the 0.01 level.

TABLE 3

p-values for the extended ANOVA procedure (ANOVA on the orthonormal polynomials of the observations) and the extended rank transform procedure (ANOVA on the orthonormal polynomials of the ranks) for the Howell data. For each cell the first entry is from the ANOVA F test, the second from a permutation test.

Data as scores	*p*-values		
Source	First order	Second order	Third order
Age	0.000/0.000	0.073/0.075	0.155/0.145
Recall condition	0.000/0.000	0.003/0.003	0.772/0.775
Interaction	0.000/0.001	0.105/0.105	0.144/0.149
Shapiro–Wilk	0.027	0.000	0.001
Ranks as scores	*p*-values		
Source	First order	Second order	Third order
Age	0.000/0.000	0.008/0.008	0.729/0.731
Recall condition	0.000/0.000	0.065/0.067	0.084/0.084
Interaction	0.002/0.002	0.037/0.038	0.357/0.356
Shapiro–Wilk	0.273	0.000	0.046

Conclusions that apply to the ranks may or may not apply to the raw scores. Thus instead of analysing ranks when the residuals are not consistent with normality, the statistician may calculate permutation test p-values for the parametric analysis. For calculating permutation test p-values we use method 1 suggested in Manly (2007, p. 145). In all the examples we have examined, there has been a remarkable consistency between the ANOVA F test p-values and the permutation test p-values. This is true even when normality of the residuals has been seriously in doubt, the Shapiro–Wilk and Anderson–Darling tests producing very small p-values. This is a testament to the robustness of the ANOVA analyses employed. We have analysed highly categorised data; see, for example, Akritas, Arnold & Brunner (1997) where the data are counts that fall into just four classes. Again the F test and the permutation test p-values were in very close agreement.

If the brief from the experimenter was to assess if there were differences in the effects of all orders rather than just differences in location effects, then the extended analyses are relevant. Using first the data and second, the ranks, as scores, the p-values in Table 3 were obtained for analyses of first, second and third order.

As expected the first order results are identical to the initial parametric and rank-transform results. Although normality is dubious for the parametric analysis, the permutation test results suggest there is no reason to doubt those p-values.

There is remarkable consistency between the results for the data and for the ranks. At first order both *Age* and *Recall Condition* are significant at the 0.001 level and the interaction at the 0.01 level. At the third order no effects are significant at the 0.05 level.

There are, however, interesting second order effects. Using the data, *Recall Condition* is significant at the 0.01 level, *Age* is significant at the 0.1 but not the 0.05 level, and the interaction is not significant at the 0.1 level. Using ranks, *Age* and *Recall Condition* swap places: *Age* is now significant at the 0.01 level whereas *Recall Condition* is now significant at the 0.1 but not the 0.05 level. The interaction is now significant at the 0.05 level.

It is a question for the experimenter to determine if an inconsistency between results for the raw data and for the ranks is important.

As discussed in section 2.4, second order reflects a combination of location and dispersion effects: second order effects are not precisely variance effects. In the second order analysis of the data there is a weak *Age* effect. It seems the *Old* are less variable than the *Young* (standard deviations 4.007, 5.787). There is a strong *Recall Condition* effect. The second order ANOVA gives the greatest effect between *Counting* and *Adjective*, but the standard deviations have greatest difference between *Counting* (standard deviation 1.618) and *Intention* (standard deviation 4.902).

Again it is a matter for the experimenter to reflect on these matters. However the extended analyses alert both the statistician and the experimenter to effects that may not be immediately apparent.

3. The extended rank transform procedure for two factor designs with unequal replications

With unequal replicates the number of observations at each combination of levels depends on the particular combination. Independence, therefore, is not a good question: the table $\{N_{rijk}\}$ is clearly dependent. Thus, a different treatment to that of the previous section is required. We cannot consider testing for independence and constructing components. However, motivated by the Section 2 treatment, we can start with the components as defined in Section 2 and construct parallel and appropriate smooth models.

3.1. Components and partition

Assume that we have observations x_{ijk}, $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, n_{ij}$. Here n_{ij} is the number of replicates at level i of factor A and level j of factor B. All $n = \sum_i \sum_j n_{ij}$ observations are ranked and we count N_{rijk}, the number of times the rth of R mid-ranks is assigned to replicate k at level i of factor A and level j of factor B. Thus N_{rijk} is zero unless the rth mid-rank is assigned to this level combination, in which case it is one.

Motivated by the results of the previous section define the components $Z_{uijk} = \sum_{r=1}^{R} a_u(r) N_{rijk}$, in which $\{a_u(r)\}$ is orthonormal on $\{p_{r\bullet\bullet\bullet}\}$ with $a_0(r) = 1$ for $r = 1, \ldots, R$. For $u = 0, 1, \ldots, R-1$ we may define

$$SS_u = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n_{ij}} Z_{uijk}^2.$$

For $u \geq 1$ the sum of squares SS_u gives an order u assessment of the factor effects. As previously most interest will be on the effects of orders one, two and three.

In the unbalanced two factor analysis of variance with unequal replicates the observations are y_{ijk}, $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, n_{ij}$. There are several possible analyses from which to choose; all produce sums of squares due to factor A, factor B, interaction and error. For each $u = 1, 2, \ldots, R-1$ in turn put $y_{ijk} = Z_{uijk}$ in the chosen ANOVA to produce order u sums of squares for factor A, factor B and factors A and B together. As before, applying the ANOVA to the Z_{uijk} for a given u is equivalent to applying it the order u orthonormal polynomial.

3.2. Modelling and analysis

Rayner & Beh (2009) give smooth models for the contingency tables such as $\{N_{rijk}\}$ and in particular discuss a product multinomial model in which for each r, $r = 1, \ldots, R$, we have a multinomial with parameters 1 and $p_{rijk} = p_{r\bullet\bullet\bullet} \sum_{u=0}^{R-1} \theta_{uijk} a_u(r)$ in which the θ_{uijk} are real valued parameters with $\theta_{0ijk} = 1$. Here θ_{uijk} reflects the order u effect corresponding to replicate k of level i of factor A and level j of factor B and $\{a_u(r)\}$ is orthonormal on $p_{r\bullet\bullet\bullet}$ with $a_0(r) = 1$ for $r = 1, \ldots, R$. On taking expectation with respect to $p_{r\bullet\bullet\bullet}$ we find $\mathrm{E}(Z_{uijk}) = \theta_{uijk}$.

It follows as before that the Z_{uijk} are all uncorrelated with zero mean and unit variance under the null hypothesis, and in general asymptotically if we assume that all θ_{uijk} are $\mathrm{O}(n^{-0.5})$. The details are similar to the treatment in Appendix B.

Reparametrise by putting, for $u = 1, \ldots, R-1$,

$$\theta_{uijk} = \mu_u + A_{ui} + B_{uj} + (AB)_{uij} \text{ in which } \sum_{i=1}^{I} A_{ui} = \sum_{j=1}^{J} B_{uj} = \sum_{i=1}^{I} (AB)_{uij} \,(\text{for each } j) = \sum_{j=1}^{J} (AB)_{uij} \,(\text{for each } i) = 0.$$

For each such u there are $IJ - 1$ independent parameters as μ_u is identically zero. This model is hence identical to the usual ANOVA model with grand mean zero.

Subject to the validity of the assumptions it now follows that for the chosen ANOVA, the usual development for the F tests in the original parametric model is again appropriate for the reparametrised smooth model. It follows that we can test, for example, $\mathrm{H}_{u\mathrm{A}}$: $(A_{ui}) = 0$ against $\mathrm{K}_{u\mathrm{A}}$: not $\mathrm{H}_{u\mathrm{A}}$ using the appropriate F ratio from the ANOVA. If the usual diagnostics, such as an assessment of the normality of the ANOVA residuals, are satisfactory, the resulting p-values are valid; if not resampling p-values would be preferable. As previously, we do not recommend obtaining p-values by referring the sums of squares to their asymptotic χ^2 distributions.

3.3. Brunner and Puri example

The data in Table 4 are given by Brunner & Puri (2002, p. 353). Three drug regimens are applied in each of 2 years with varying numbers of replicates. Multiplicities of counts are indicated in parentheses.

Probably the most commonly applied ANOVA in this situation is due to Yates (1934). Kuehl (2000, section 6.9), for example, illustrates the Yates method. Appropriate sums of squares are the default in BMDP, are called 'type III' in SAS and 'adjusted' in MINITAB. An initial investigation might look at the usual parametric analysis and the rank transform procedure, with the results shown in Table 5.

The residuals from both analyses are consistent with normality. Both analyses agree that *Drugs* are significant at the 0.05 level and that there are no other significant effects at the usual levels of significance.

If the experimenter wishes to examine higher order effects extended analyses for both the original data and the ranks give the p-values in Table 6.

As expected, the first order ANOVA F test p-values are exactly as for the parametric and rank transform procedures. While the residuals for both first order analyses and third order extended rank transform procedures are consistent with normality, those for both

TABLE 4

Brunner & Puri (2002) data.

Drug	Year 1	Year 2
A	11(2), 12(1), 13(4), 14(2)	9(1), 11(1), 12(5), 13(2), 14(2), 15(1), 16(1)
B	11(2), 12(2), 13(1), 14(3), 15(1)	9(1), 10(1), 11(3), 12(1), 13(1), 15(1)
C	11(1), 12(1), 13(1), 14(2), 15(2), 17(1)	9(1), 12(1), 13(4), 14(3), 15(2), 17(1)

TABLE 5

p-values for the parametric analysis (ANOVA on the observations) and the rank transform procedure (ANOVA on the ranks) for the Brunner & Puri data.

	p-values	
Source	Parametric analysis	Rank transform procedure
Drug	0.033	0.031
Year	0.208	0.261
Interaction	0.441	0.416
Shapiro–Wilk	0.327	0.479

TABLE 6

p-values for the extended ANOVA procedure (ANOVA on the orthonormal polynomials of the observations) and the extended rank transform procedure (ANOVA on the orthonormal polynomials of the ranks) for the Brunner & Puri data. For each cell the first entry is from the ANOVA F test, the second from a permutation test.

Data as scores	*p*-values		
Source	First order	Second order	Third order
Drugs	0.033/0.032	0.374/0.379	0.931/0.934
Year	0.208/0.205	0.209/0.213	0.628/0.629
Interaction	0.441/0.444	0.528/0.535	0.783/0.781
Shapiro–Wilk	0.327	0.000	0.007

Ranks as scores	*p*-values		
Source	First order	Second order	Third order
Drugs	0.031/0.032	0.247/0.244	0.894/0.893
Year	0.261/0.254	0.268/0.271	0.873/0.874
Interaction	0.416/0.416	0.288/0.289	0.533/0.533
Shapiro–Wilk	0.479	0.004	0.262

second order analyses and the third order extended ANOVA procedure are not. Thus the validity of those analyses is in question. These conclusions are in agreement with unpresented Q–Q plots and application of the Anderson–Darling test. At this point the statistician could calculate permutation test *p*-values, and would find these are in close agreement with the F test *p*-values. It would seem this is again a reflection of the robustness of the ANOVA F tests and there is no reason to doubt any of the ANOVA F test *p*-values.

When using the extended rank transform procedure, results are very similar to those for the extended parametric analysis. If the analysis is exploratory one might simply conclude at this point that there are no statistically significant second and third order effects while at the first order only *Drugs* means are significantly different at the 0.05 level.

Brunner & Puri (2002) found p-values, using Wilcoxon scores and the modified Box approximation to better detect first order effects, to be 0.031, 0.263 and 0.403 for A, B and interaction respectively. Their p-values, using Mood scores and the modified Box approximation to better detect second order effects, are 0.244, 0.254 and 0.260. These are similar to ours. There are no statistically significant second or third order effects, although significant second order effects in particular would have been influential in conclusions drawn from the analysis.

4. Multi-factor designs

Suppose we have observations $x_{i_1\ldots i_m k}$, $i_1 = 1, \ldots, I_1, \ldots, i_m = 1, \ldots, I_m$ and $k = 1, \ldots, n_{i_1\ldots i_m}$ for an m-factor design with unequal replicates and no factors considered to be ordered. All $n = \sum_{i_1} \ldots \sum_{i_m} n_{i_1\ldots i_m}$ observations are ranked and $N_{ri_1\ldots i_m k}$ counts the number of times the rth of R mid-ranks is assigned to the kth replicate of the i_1th level of the first factor, $\ldots, i_m$th level of the mth factor. Thus $N_{ri_1\ldots i_m k}$ is zero unless the rth mid-rank is assigned to the level combination $(i_1, \ldots, i_m)$, in which case it is one. From Rayner & Beh (2009) the Pearson test statistic $X_P^2 = \sum_{\text{allcells}}(\text{observed } - \text{expected})^2/\text{expected}$ for testing independence in the table $\{N_{ri_1\ldots i_m k}\}$ is given by

$$X_P^2 = \sum_{u=0}^{R-1} \sum_{i_1=1}^{I_1} \cdots \sum_{i_m=1}^{I_m} \sum_{k=1}^{n_{i_1\ldots i_m}} Z_{ui_1\ldots i_m k}^2,$$

with $Z_{ui_1\ldots i_m k} = \sum_{r=1}^{R} a_u(r) N_{ri_1\ldots i_m k} (np_{\bullet i_1 \bullet \ldots \bullet} \cdots p_{\bullet\bullet\bullet \ldots i_m \bullet} p_{\bullet\bullet \ldots \bullet k})^{-0.5}$, in which $p_{ri_1\ldots i_m k} = N_{ri_1\ldots i_m k}/n$ with the marginal probabilities obtained by summing, and $\{a_u(r)\}$ is orthonormal on $\{p_{r\bullet\ldots\bullet}\}$. Thus X_P^2 is represented as the sum of the squares of components of all orders.

As noted above, for a given level combination $(i_1, \ldots, i_m)$, only once is $N_{ri_1\ldots i_m k}$ nonzero, and then it is one. This occurs for the unique value of r for which this level combination takes its appropriate mid-rank. Thus each $Z_{ui_1\ldots i_m k}$ takes precisely one of the values of $a_u(r)(np_{\bullet i_1 \bullet \ldots \bullet} p_{\bullet\bullet\bullet \ldots i_m \bullet} p_{\bullet\bullet \ldots \bullet k})^{-0.5}$ and $\{Z_{ui_1\ldots i_m k}\} = \{a_u(r)(np_{\bullet i_1 \bullet \ldots \bullet} p_{\bullet\bullet\bullet \ldots i_m \bullet} p_{\bullet\bullet \ldots \bullet k})^{-0.5}\}$.

For balanced tables in which $n_{i_1\ldots i_m} = K$ for all $i_1, \ldots, i_m$, $\{Z_{ui_1\ldots i_m k}\} = \{a_u(r)\}$. As for the tables considered previously, SS_0 is the Pearson statistic for the table $\{N_{\bullet i_1\ldots i_m k}\}$ while for $u = 1, \ldots, R-1$, SS_u can be defined as the sum of squares of the $Z_{ui_1\ldots i_m k}$, or equivalently, the orthonormal polynomial of order u. The fixed effects ANOVA can be used to partition SS_u into the sum of main effects, two and three factor interactions, and so on. Then a smooth model may be applied to derive test statistics of main effects and interactions of all orders.

For unbalanced tables we may apply the ANOVA to the values of a given set of orthonormal functions. A smooth model may be constructed to validate the procedure.

In multi-way tables there will be many possible assessments, and even if there are no effects in the model that generated the data, 0.05 of the test statistics can be expected to be significant at the 0.05 level. One strategy for dealing with this is to limit the number of assessments. This could mean only looking at orders 1, 2 and 3.

5. Conclusions

The extended rank transform procedure generates tests based on the scaled orthonormal polynomials. The order one procedure tests are the same as the well-known rank transform tests. The extension to tests of higher order is natural if dispersion, skewness and higher order moment effects are of interest.

If there are ties or the design is unbalanced, the motivation of constructing components from the Pearson test of independence is lost. Nevertheless, smooth models may be constructed that validate the rank transform procedure and its extensions to assessing higher order effects.

If the orthonormal polynomials are constructed on the original data instead of the ranks, parallel procedures result. That of first order is the original ANOVA, while those of higher orders are extensions that assess second, third, etc. orders, reflecting treatment differences in second, third, etc. order moments.

Appendix A: Two-factor balanced designs with no ties

In this appendix we show that under certain conditions the Pearson statistic and sums of squares SS_u may be constants that depend only on the design and not on the data. Naturally these statistics then cannot be used for inference.

It is assumed that there are two factors, the first with I levels and the second with J levels. At each combination of levels there are K replicates. Further suppose there are no ties and that $\{N_{rijk}\}$ defines a four-way singly ordered table of counts of zeros and ones with $n = \sum_{r,i,j,k} N_{rijk}$ observations. As in Beh & Davy (1999) and Rayner & Best (2001), Pearson's statistic X_{P}^2 may be partitioned into components Z_{uijk} via

$$X_{\mathrm{P}}^2 = \sum_{u=1}^{n-1} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} Z_{uijk}^2,$$

with $Z_{uijk} = \sum_{r=1}^{n} a_u(r) N_{rijk} (np_{\bullet i \bullet \bullet} p_{\bullet \bullet j \bullet} p_{\bullet \bullet \bullet k})^{-0.5}$, in which $\{a_u(r)\}$ is orthonormal on $\{p_{r\bullet\bullet\bullet}\}$ with $a_0(r) = 1$ for $r = 1, \ldots, n$. Formally, X_{P}^2 includes a term for Pearson's statistic for the unordered two-way table formed by summing over r: $\{N_{\bullet ijk}\}$. However this table has every entry equal to one, and Pearson's X^2 for it is zero. As in section 2.1 note that $N_{\bullet i \bullet \bullet} = JK$, $N_{\bullet \bullet j \bullet} = IK$ and $N_{\bullet \bullet \bullet k} = IJ$. It follows that $p_{\bullet i \bullet \bullet} = 1/I$, $p_{\bullet \bullet j \bullet} = 1/J$ and $p_{\bullet \bullet \bullet k} = 1/K$, giving $Z_{uijk} = \sum_{r=1}^{IJk} a_u(r) N_{rijk}$.

As there are no ties, first, since $p_{r\bullet\bullet\bullet} = 1/n$ (every rank is equally likely), $\{a_u(r)\}$ is orthonormal on $\{1/n\}$. Second, under the null hypothesis of no treatment effects $E(N_{rijk})$ is a constant independent of r, i, j and k. It follows that $\sum_{r,i,j,k} N_{rij} = n = \sum_{r,i,j,k} E(N_{rijk}) = E(N_{rijk}) \sum_{r,i,j,k} 1 = n^2 \times E(N_{rijk})$ (since $\sum_{r,i,j,k} 1 = nIJK = n^2$), so that $E(N_{rijk}) = 1/n$. Third, $N_{rijk} = 0$ or 1 for all r, i, j, k implies $N_{rijk}^2 = N_{rijk}$. Hence *no matter what the data*

$$X_P^2 = \sum_{r=1}^{n} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} N_{rijk}^2 / E(N_{rijk}) - n = n \sum_{r=1}^{n} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} N_{rijk} - n = n(n-1).$$

Since X_P^2 is not a random variable, it is not useful for inference.

We define

$$SS_u = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} Z_{uijk}^2,$$

for $u = 1, 2, \ldots, n-1$ so that $X_P^2 = SS_1 + \ldots + SS_{n-1}$.

We have from above that $Z_{uijk} = \sum_{r=1}^{n} a_u(r) N_{rijk}$. First, as $a_1(r) = (r-\mu)/\sigma$ in which $\mu = (n+1)/2$ and $\sigma^2 = (n^2-1)/12$, we have $Z_{1ijk} = \sum_{r=1}^{n}(r-\mu)N_{rijk}/\sigma$ and $SS_1 = \sum_{i,j,k} Z_{1ijk}^2$. For each (i,j,k) combination only once is N_{rijk} nonzero and it is one for this particular combination. So $Z_{1ijk} = (r-\mu)/\sigma$ for this combination and zero otherwise. Squaring and summing over all i, j and k we sum over every possible outcome, and hence all ranks. It follows that $SS_1 = \sum_{r=1}^{n}(r-\mu)^2/\sigma^2 = n$.

In general, using the orthonormality of $\{a_u(r)\}$,

$$SS_u = \sum\nolimits_{i,j,k} Z_{uijk}^2 = \sum_{r=1}^{n} a_u^2(r) = n\sum_{r=1}^{n} a_u^2(r) p_{r\bullet\bullet\bullet} = n.$$

This confirms that $X_P^2 = n(n-1)$.

Appendix B: Moments

Rayner & Beh (2009) discuss a product multinomial model in which for each r, $r = 1, \ldots, n$, we have a multinomial with parameters 1 and $p_{rijk} = p_{r\bullet\bullet\bullet}\sum_{u=0}^{n-1}\theta_{uijk}a_u(r)$ in which the θ_{uijk} are real valued parameters with $\theta_{0ijk} = 1$. Recall that $Z_{uijk} = \sum_{r=1}^{n} a_u(r)N_{rijk}$ and that $\{a_u(r)\}$ is orthonormal on $\{p_{r\bullet\bullet\bullet}\}$ with $a_0(r) = 1$ for $r = 1, \ldots, n$. Taking expectation with respect to $\{p_{r\bullet\bullet\bullet}\}$,

$$\begin{aligned}
\mathrm{E}(Z_{uijk}) &= \sum\nolimits_{r=1}^{n} a_u(r)E(N_{rijk}) = \sum\nolimits_{r=1}^{n} a_u(r)p_{rijk} = \sum\nolimits_{r=1}^{IJK} a_u(r)p_{r\bullet\bullet\bullet}\sum\nolimits_{v=0}^{n-1}\theta_{vijk}a_v(r)\\
&= \sum\nolimits_{v=0}^{n-1}\theta_{vijk}\sum\nolimits_{r=1}^{IJK} a_u(r)a_v(r)p_{r\bullet\bullet\bullet} = \theta_{rijk}.
\end{aligned}$$

Under the null hypothesis these are all zero; under the alternative all θ_{uijk} are $O(n^{-0.5})$ and hence $\mathrm{E}(Z_{uijk}))$ tends to zero. Next

$$\begin{aligned}
\mathrm{var}(Z_{uijk}) &= \sum\nolimits_r a_u^2(r)\mathrm{var}(N_{rijk}) + \sum\nolimits_{r\neq s} a_u(r)a_u(s)\mathrm{cov}(N_{rijk}, N_{sijk})\\
&= \sum\nolimits_r a_u^2(r)p_{rijk}(1-p_{rijk}) - \sum\nolimits_{r\neq s} a_u(r)a_u(s)p_{rijk}p_{sijk}\\
&= \sum\nolimits_r a_u^2(r)p_{rijk} - \sum\nolimits_{r,s} a_u(r)a_u(s)p_{rijk}p_{sijk}\\
&= \sum\nolimits_r a_u^2(r)p_{r\bullet\bullet\bullet}\sum\nolimits_{v=0}^{n-1}\theta_{vijk}a_v(r)\\
&\quad - \left\{\sum\nolimits_u a_u(r)p_{r\bullet\bullet\bullet}\sum\nolimits_v \theta_{vijk}a_v(r)\right\}\left\{\sum\nolimits_s a_u(s)p_{s\bullet\bullet\bullet}\sum\nolimits_v \theta_{vijk}a_v(s)\right\}.
\end{aligned}$$

Now $\sum_r a_u^2(r)p_{r\bullet\bullet\bullet}\sum_{v=0}^{n-1}\theta_{vijk}a_v(r) = \sum_v \theta_{vijk}\sum_u a_u(r)a_v(r)p_{r\bullet\bullet\bullet} = \sum_v \theta_{vijk}\delta_{uv} = \theta_{uijk}$. As above this is either zero or tends to zero.

Next, note that $\mathrm{E}\{a_u^2(u)\} = 1 < \infty$ so that $a_u^2(r) = \sum_{w=0}^{n-1} c_{uw}a_w(r)$ with $c_{u0} = 1$. Hence

$$\begin{aligned}\sum_r a_u^2(r)p_{r\bullet\bullet\bullet}\sum_{v=0}^{n-1}\theta_{vijk}a_v(r) &= \sum_{v=0}^{n-1}\theta_{vijk}\sum_u a_v(r)\Big\{\sum_u c_{uw}a_w(r)\Big\}p_{r\bullet\bullet\bullet}\\ &= \sum_{v=0}^{n-1}c_{uv}\theta_{vijk} = 1+\sum_{v=1}^{n-1}c_{uv}\theta_{vijk}.\end{aligned}$$

It follows that under the null hypothesis $\operatorname{var}(Z_{uijk})$ is one; under the alternative that all θ_{uijk} are $O(n^{-0.5})$, $\operatorname{var}(Z_{uijk})$ tends to one.

Similarly for $(i,j,k) \neq (i',j',k')\operatorname{cov}(Z_{uijk}, Z_{u'i'j'k'})$ is, apart from the expectations $E(Z_{uijk})$ and $E(Z_{u'i'j'k'})$ that are zero or tend to zero, $E(Z_{uijk}Z_{u'i'j'k'})$. Consider the case $u = u'$. We find

$$\begin{aligned}E(Z_{uijk}Z_{u'i'j'k'}) &= E\Big[\Big\{\sum_r a_u(r)N_{rij}\Big\}\Big\{\sum_s a_u(s)N_{si'j'k'}\Big\}\Big] = E\Big\{\sum_{r,s} a_u(r)a_u(s)N_{rijk}N_{si'j'k'}\Big\}\\ &= E\Big\{\sum_r a_u(r)a_u(r)N_{rijk}N_{ri'j'k'}\Big\} + E\Big\{\sum_{r\neq s} a_u(r)a_{u'}(s)N_{rijk}N_{si'j'k'}\Big\}.\end{aligned}$$

The first expression here involves a single multinomial, while the second involves two. As before, both are either zero or tend to zero. For example

$$\begin{aligned}E\Big\{\sum_{r\neq s} a_u(r)a_{u'}(s)N_{rijk}N_{si'j'k'}\Big\} &= \sum_{r\neq s} a_u(r)a_{u'}(s)E(N_{rijk}N_{si'j'k'})\\ &= -\sum_{r\neq s} a_u(r)a_{u'}(s)p_{rijk}p_{si'j'k'}\\ &= -\sum_r a_u(r)a_u(s)\Big\{p_{r\bullet\bullet\bullet}\sum_v \theta_{vijk}a_v(r)\Big\}\\ &\quad\times\Big\{p_{s\bullet\bullet\bullet}\sum_w \theta_{vi'j'k'}a_w(s)\Big\}\\ &= -\sum_{r,v,w,x,y}\theta_{vijk}\theta_{vi'j'k'}c_{uvx}c_{uwy}a_x(r)a_y(s)p_{r\bullet\bullet\bullet}p_{s\bullet\bullet\bullet} = 0,\end{aligned}$$

since, for example, $\sum_r a_u(r)p_{r\bullet\bullet\bullet} = 0$.

Similarly for $u \neq u'$, $\operatorname{cov}(Z_{uij}, Z_{u'i'j'})$ is zero.

References

AKRITAS, M.G., ARNOLD, S.F. & BRUNNER, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *J. Amer. Statist. Assoc.* **92**, 258–265.

BEH, E.J. & DAVY, P.J. (1998). Partitioning Pearson's chi-squared statistic for a completely ordered three-way contingency table. *Aust. N. Z. J. Stat.* **40**, 465–477.

BEH, E.J. & DAVY, P.J. (1999) Partitioning Pearson's chi-squared statistic for a partially ordered three-way contingency table. *Aust. N. Z. J. Stat.* **41**, 233–246.

BRUNNER, E. & PURI, M.L. (2002). A class of rank-score tests in factorial designs. *J. Statist. Plann. Inference* **103**, 331–360.

CONOVER, W.J. & IMRAN, R.L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *Amer. Stat.* **35**, 124–129.

HOWELL, D.C. (2012). *Statistical Methods for Psychology*, 8th edn. Belmont, California: Wadsworth Publishing.

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

MANLY, B.F.J. (2007). *Randomization, Bootstrap, and Monte Carlo Methods in Biology*, 3rd edn. London: Chapman & Hall.

MANSOURI, H. (1999) Multifactor analysis of variance based on the aligned rank transform technique. *Comput. Statist. Data Anal.* **29**, 177–189.

MARDEN, J.I. & MUYOT, M.E. (1995). Rank tests for main and interaction effects in the analysis of variance. *J. Amer. Statist. Assoc.* **90**, 1388–1398.

RAYNER, J.C.W. & BEH, E.J. (2009). Components of Pearson's statistic for at least partially ordered *m*-way contingency tables. *J. Appl. Math. Decis. Sci.* **2009**, Article ID 980706. doi:10.1155/2009/980706.

RAYNER, J.C.W. & BEST, D.J. (1986). Neyman-type smooth tests for location-scale families. *Biometrika* **73**, 437–446.

RAYNER, J.C.W. & BEST, D.J. (2001). *A Contingency Table Approach to Nonparametric Testing*. Boca Raton, Florida: Chapman & Hall/CRC.

RAYNER, J.C.W., THAS, O. & DE BOECK, B. (2008). A generalised Emerson recurrence relation. *Aust. N. Z. J. Stat.* **50**, 235–240.

TOOTHAKER, L.E. & NEWMAN, D. (1994). Nonparametric competitors to the two-way ANOVA. *J. Ed. Behav. Stat.* **19**, 237–273.

YATES, F. (1934). The analysis of multiple classifications with unequal numbers in different classes. *J. Amer. Statist. Assoc.* **29**, 51–66.