2014

Extended analysis of at least partially ordered multi-factor ANOVA

J C W Rayner
University of Wollongong

D J. Best
University of Newcastle

O Thas
University of Wollongong

Extended analysis of at least partially ordered multi-factor ANOVA

Abstract

For multifactor experimental designs in which the levels of at least one of the factors are ordered we show how to construct components that provide a deep nonparametric scrutiny of the data. The components assess generalised correlations and the resulting tests include and extend the Page and umbrella tests. Application of the tests described is straightforward. Orthonormal polynomials on the ANOVA responses and the factors need to be constructed. Products of at least two of these orthonormal polynomials are then used as inputs into standard ANOVA routines. For example using the first order orthonormal polynomia l on factor A and the first/second order orthonormal polynomial on the ANOVA response will assess, if, for example, with increasing levels of factor A the response increases or increases and then decreases.

NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

06-14

EXTENDED ANALYSIS OF AL LEAST PARTIALLY ORDERED MULTI-FACTOR ANOVA

J.C.W. RAYNER, D.J. BEST AND O.THAS

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

EXTENDED ANALYSIS OF AT LEAST PARTIALLY ORDERED MULTI-FACTOR ANOVA

J.C.W. RAYNER[1,2]*, D.J. BEST[2] AND O. THAS [3,1]

Summary

For multifactor experimental designs in which the levels of at least one of the factors are ordered we show how to construct components that provide a deep nonparametric scrutiny of the data. The components assess generalised correlations and the resulting tests include and extend the Page and umbrella tests. Application of the tests described is straightforward. Orthonormal polynomials on the ANOVA responses and the factors need to be constructed. Products of at least two of these orthonormal polynomials are then used as inputs into standard ANOVA routines. For example using the first order orthonormal polynomial on factor A and the first/second order orthonormal polynomial on the ANOVA response will assess, if, for example, with increasing levels of factor A the response increases or increases and then decreases.

Key Words: Completely randomised designs; factorial designs; generalised correlations; orthonormal polynomials; Page test; umbrella test.

* Author to whom correspondence should be addressed.

e-mail: John.Rayner@newcastle.edu.au

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

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

[3] Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, B-9000 Gent, Belgium

1. Introduction

Our interest here is multifactor experimental designs, the simplest of which include the completely randomized, factorial and Latin square designs. Although nonparametric tests exist for these designs, most ignore any ordering of the levels of the factors. The only general nonparametric option is the rank transform procedure of Conover & Imran (1981). This paper addresses both of these issues.

Rayner & Best (2013) discuss designs in which the levels of all factors are not ordered, or the order is ignored. They generalise the rank transform. As there, here the approach is based on the construction of a contingency table from the data and the use a device of Beh & Davy (1998, 1999) to partition into components the Pearson X^2 statistic used to test for independence. A model is given that demonstrates that the components of the Pearson X^2 statistic are appropriate test statistics for nonparametric testing of relevant generalised correlations. The simplest of the tests that result from this approach are generalisations of familiar linear by linear (c.f. Page) and linear by quadratic (c.f. umbrella) tests. See, for example, Rayner & Best (2001, 2005).

Rayner & Best (2013) produce two sorts of tests. One extends the rank transform procedure that utilises ranks to assess differences between the treatment mean ranks of the. It uses what may be thought of as generalised ranks to construct tests that assess dispersion, skewness etc. differences between treatment ranks. The other class of tests is parallel to the first, and uses the data to produce tests that extend the usual ANOVA to assess equality of mean, dispersion, skewness etc. differences between treatments.

Our approach here is similar. Two sorts of tests are again available, using either the ranks or the data. We assess, in the sense of Rayner & Beh (2009a), generalised correlations between the factors that have ordered levels. It is the use of orthonormal polynomials that enables generalised correlations to be assessed. Many hypotheses may be tested by applying simple t and ANOVA F tests.

Since the hypotheses developed here are in terms of generalised correlations, it is appropriate to give a little background on these. For bivariate discrete random variables (X, Y) with $P(X = x_i, Y = y_j) = p_{ij}$ suppose that $\{a_u(X)\}$ are orthonormal polynomials on $\{p_{i\bullet}\}$ and $\{b_v(Y)\}$ are orthonormal polynomials on $\{p_{\bullet j}\}$. Then $\theta_{uv} = E[a_u(X)\, b_v(Y)]$ is the

(u, v)th bivariate *generalised correlation*. For trivariate random variables (X, Y, Z) with P($X = x_i$, $Y = y_j$, $Z = z_k$) = p_{ijk}, in addition to the first two sets of orthonormal polynomials, a third set of polynomials, $\{c_w(Z)\}$, orthonormal on $\{ p_{\bullet\bullet k} \}$ is required. Then θ_{uvw} = E[$a_u(X)$ $b_v(Y)$ $c_w(Z)$] is the (u, v, w)th trivariate generalised correlation. In the bivariate case most interest focuses on θ_{11}, which, when the original observations are used, is the Pearson correlation; when mid-ranks are used, it is the Spearman correlation. Also of interest is θ_{12}, that reflects the association between linear X and quadratic Y. In Rayner & Best (2001, section 8.1) this correlation reflects the effect of how, with increasing age, intelligence may increase and then decrease. In some scenarios it may be of interest to consider how, with increasing X, Y may vary as linear, quadratic or cubic functions. In the trivariate case θ_{11r} reflects how as both X and Y vary linearly Z behaves as an order r polynomial. Perhaps most interest in generalised correlations would be in assessing independence, where the non-zero generalised correlations identify how the independence model fails. However here we are most interested in generalisations of the Page and umbrella tests. Generalised correlations may similarly be defined for multivariate random variables (X_1, ..., X_k) for any k.

In the following we suggest it will be unusual to need orthonormal polynomials beyond the third. For the convenience of readers we record the initial orthonormal polynomials of a random variable X. Write μ for the mean of X and μ_r, r = 2, 3, … for the central moments of X. To avoid ambiguity set $a_0(x) = 1$ for all x. Then

$$a_1(x) = (x - \mu)/\mu_2,$$

$$a_2(x) = \{(x - \mu)^2 - \mu_3(x - \mu)/\mu_2 \ - \mu_2\}/\sqrt{d} \text{ in which } d = \ \mu_4 - \mu_3^2 / \mu_2 - \mu_2^2, \text{ and}$$

$$a_3(x) = \{(x - \mu)^3 - a(x - \mu)^2 - b(x - \mu) \ - c\}/\sqrt{e},$$

in which

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

$$e = \mu_6 - 2a\mu_5 + (a^2 - 2b)\mu_4 + 2(ab - c)\mu_3 + (b^2 + 2ac)\mu_2 + c^2.$$

The tests we propose using here are ANOVA F and t-tests, generally well-known for their robustness. However failure of the ANOVA assumptions may lead to invalid inference. See, for example, Keselman et al. (2002). In applying them we have checked their p-values with those obtained from permutation tests, and even when normality of the residuals is in serious doubt, for all the examples we have checked there was reasonable agreement. We suggest ANOVA F and t-test p-values can usually be used with confidence, especially if normality and variance homogeneity are confirmed. If not, there are remedies in the literature. Our preferred approach when a test gives borderline significance or the ANOVA assumptions are problematic, is to use permutation test p-values, specifically method 1 suggested in Manly (2007, p. 145).

The simplest designs we address are the completely randomised design with one factor and replicates, and the two-factor ANOVA with no replicates. If there are an equal number of replicates in each cell then up to a point to be clarified later, replicates can be treated as another unordered factor. With unequal replicates a slight modification is required. To demonstrate our approach we will consider the case of a balanced design with two factors and replication. We will work through the particular cases of the levels of no, one or two factors being ordered. Results for the completely randomised design can be inferred from those presented here.

In section 2 we consider equally replicated two-factor designs with one factor ordered. Section 3 considers equally replicated two-factor designs with both factors ordered. Section 4 addresses the modifications necessary when the designs are not equally replicated and multi-factor designs in general.

This section is concluded with an example featuring highly categorical data from Akritas et al. (1997), the theory for which is developed in section 2. The design here is an equally replicated two factor ANOVA with the levels of one factor ordered.

Drugs and Concentrations Example

Two drugs are administered in three concentrations. The outcomes are 0 (no changes), 1 (slight changes), 2 (distinct changes), 3 (severe changes). In each cell in Table 1 the entry after the '/' is the cell mean.

TABLES 1 & 2 NEAR HERE

The extended unordered analyses described in Rayner & Best (2013) yield the p-values in Table 2. The first set of p-values uses the ranks as scores and the first order analysis is the rank transform procedure. The second set uses the data as scores, with the first order analysis being the usual two-factor parametric analysis. The simplicity of the procedure lies in that the two-factor ANOVA is applied, instead of to the data (or their mid-ranks), to the first, second and third order sets of orthonormal polynomials. These give assessments of moments of orders one, two and three of the treatments. Essentially the data (or their mid-ranks) are being transformed using orthonormal functions. The resulting analyses are, in a sense to be clarified towards the end of section 2, uncorrelated.

Both sets of analyses are suspect, as the Shapiro-Wilk test finds the residuals are not consistent with normality at the 5% level. However, reflecting the robustness of ANOVA F tests, permutation tests give near identical p-values.

Both first order analyses find concentrations significant at the 0.1% level, drugs significant at the 5% level but not the 1% level, and the interaction not significant at all reasonable levels. There are no effects of order two or three that are significant at the 5% level.

TABLES 3 & 4 NEAR HERE

The extended ordered analyses developed in the body of this paper are given for these data in the Tables 3 and 4. We use the mid-ranks as response scores, 2, 5 and 10 as concentration scores and 1 and 2 as drug scores. In Tables 3 and 4 by 'order' we mean the order of the orthonormal function for that marginal. In Table 3, each cell contains a

sample generalised correlation for the entire table multiplied by $\sqrt{n}$. These are asymptotically standard normal, so those that are most extreme can be gleamed at a glance. The second entry in the table is the p-value from the one sample t-test for a mean of zero against two-sided alternatives; the third entry is the permutation test p-value using method 1 suggested in Manly (2007, p. 145). As was found for the unordered ANOVA in Rayner & Best (2013), there is reasonable agreement between the p-values produced by the two approaches. Good agreement with the permutation test p-values in Table 3 was found using the approximation given just before Table 9.

There is substantial generalised correlation of order (1, 1), suggesting that as concentration increases the outcome increases linearly. As the concentration increases from 2 to 5 and then to 10, the outcome means pass from 0.175 to 0.675 and then to 1.750. In addition the order (2, 2) generalised correlation is significantly different from zero. This is harder to interpret directly, but from a modelling perspective, the model to be developed in section 2 requires, as well as the table's marginal probabilities, the (1, 1) and (2, 2) generalised correlations.

Next, to assess whether or not the generalised correlations differ across drugs a one-way ANOVA with drugs as factor is applied to the sets of products of the orthonormal functions in turn. This tests whether the sample generalised correlations for each drug are consistent. Table 4 gives the p-values. It is certainly possible that a generalised correlation for the entire table that is consistent with zero differs across the levels of the unordered effect, but that is not the case here. None of the sample generalised correlations differ significantly from drug 1 to drug 2.

The treatment using the data as response scores is entirely parallel.

An important caveat on the conclusions here is that significant effects at a particular order (using orthonormal polynomials of order r say) may affect conclusions at higher orders. This effect is well explored in testing goodness of fit. See, for example, Rayner et al. (2009, section 5.3.3 and pp.196). In that context we argue that a significant component of order r may affect the significance or not of components up to order $2r$, but most attention should focus on components up to order r. The situation here requires both theoretical and empirical exploration that we defer to another time. We suggest that from a data analytic perspective, effects using orthonormal polynomials of order r be

interpreted as reflecting moment effects of that order, and effects jointly using orthonormal polynomials of order r and s be interpreted as reflecting generalised correlation effects of order (r, s).

2. Equally replicated two-way tables with one factor ordered

We now develop arithmetic decompositions of a Pearson statistic used to test independence in a contingency table constructed from the ANOVA data. These decompositions will subsequently be exploited to construct new nonparametric tests. We emphasise that the contingency table is not directly used in subsequent data analysis.

In some contingency tables certain statistics (the Pearson statistic and sums of squares SS_u, SS_{uv} and SS_{uvw} defined subsequently) are constant no matter what the data. Clearly these statistics cannot be used for inference. For similar results when no factors are ordered, see Rayner & Best (2013, Appendix 1).

First some notation is given that will be used both in this section and the next. Assume that we have n observations y_{ijk}, $i = 1, ..., I, j = 1, ..., J$ and $k = 1, \ldots, K$: there are K replicates at level i of factor A and level j of factor B. (By no replications we mean $K = 1$.) All observations are ranked and we count N_{rijk}, the number of times the rth of R distinct mid-ranks is assigned to replicate k at level i of factor A and level j of factor B. Thus $R = n$ if there are no ties. It follows that N_{rijk} is zero unless the rth mid-rank is assigned to this level/replicate 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 in the table $\{N_{rijk}\}$. This is a natural approach to take inasmuch as independence corresponds to a complete lack of structure, when all the generalised correlations for which we can test by decomposing the independence test statistic X_{P}^2 will be zero.

The case when the levels of all factors are unordered was considered in Rayner & Best (2013). When using mid-ranks in the decomposition of X_{P}^2 the first order component gives the rank transform procedure of Conover & Imran (1981). This assesses rank location effects. Higher order components give extensions of the rank transform

procedure and assess rank dispersion and rank skewness effects, and so on. However it is not necessary to use mid-ranks in the construction of $\{N_{rijk}\}$: any set of ordered scores may be used. If the original data are used, the first order component gives the original ANOVA. Use of the higher order components gives extensions of the ANOVA to assess dispersion, skewness effects and moment effects of higher order.

The same duality will be pursued here, where the levels of at least one of the factors are ordered.

Subsequently standard dot notation has been used, so that, for example, $N_{\bullet\bullet\bullet\bullet} = IJK = n$, which is both the number of times a rank has been assigned and the number of observations. For all r, i, j and k write $p_{rijk} = N_{rijk}/n$. 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 $p_{\bullet\bullet\bullet k} = 1/K$.

In this section it is assumed that the first factor is ordered and the second is not. The scenario when the levels of both factors are ordered is considered in section 3.

To reflect the fact that the levels of factor A are ordered we write N_{rsjk} for the number of times the rth of the R distinct ordered scores is assigned to the level/replicate combination (s, j, k); the subscript i is replaced by s. Then $\{N_{rsjk}\}$ defines a four-way doubly ordered table of counts of zeros and ones. As in Beh & Davy (1999) and Rayner & Best (2001, section 10.2), Pearson's independence test statistic X^2_{P} may be partitioned into components Z_{uvjk} via

$$X^2_{\mathrm{P}} = \sum_{u=1}^{R-1}\sum_{v=1}^{I-1}\sum_{j=1}^{J}\sum_{k=1}^{K} Z^2_{uvjk} + \sum_{v=1}^{I-1}\sum_{j=1}^{J}\sum_{k=1}^{K} Z^2_{0vjk} + \sum_{u=1}^{R-1}\sum_{j=1}^{J}\sum_{k=1}^{K} Z^2_{u0j}$$

with $Z_{uvjk} = \sum_{r=1}^{R}\sum_{s=1}^{I} a_u(r)b_v(s)N_{rsjk}/\sqrt{I}$, in which $\{a_u(r)\}$ is orthonormal on $\{p_{r\bullet\bullet\bullet}\}$ with $a_0(r) = 1$ for $r = 1, ..., R$, and $\{b_v(s)\}$ is orthonormal on $\{p_{\bullet s\bullet\bullet}\}$ with $b_0(s) = 1$ for $s = 1, ..., I$.

Perhaps the statistics of most interest are

$$SS_{uv} = \sum_{j=1}^{J}\sum_{k=1}^{K} Z_{uvjk}^2 ,$$

defined for $u = 0, 1, 2, ..., n - 1$ and $v = 0, 1, ..., I - 1$, but not $(u, v) = (0, 0)$. When $K = 1$ the Z_{uvjk} are proportional to the *generalised sample correlations* as in Rayner & Beh (2009a). Rayner & Best (2001, section 6.5) essentially define these Zs to be extensions of the Spearman test statistic. As there, $\sum_{u,v} Z_{uvjk}^2$ is Pearson's X_{P}^2 for the two-way tables corresponding to each level of factor B/replicate, and is an aggregation of the extended Spearman-type test statistics for the (j, k)th level/replicate combination. One interpretation of Z_{1vjk} is that for the (j, k)th level/replicate combination, as the treatments pass from 1 to I there is an effect of degree v. For example, if $v = 2$, then in passing from treatment 1 to treatment I there is either an increase then a decrease in the treatment effects, or a decrease and then an increase: an umbrella effect. From the above, $\sum_{u,v,j,k} Z_{uvjk}^2 = X_{\mathrm{P}}^2$; the Pearson test statistic is an aggregation of the extended Spearman-type test statistics for over all level/replicate combinations.

At this point, to motivate the subsequent treatment of multi-factor designs, we temporarily assume that instead of replicates we have a factor C with unordered levels. Later the model now developed is simplified to deal instead with replicates.

As in Rayner & Beh (2009b) we construct a smooth product multinomial model in which with the counts corresponding to the (j, k)th plane being multinomial with total count $n_{\bullet\bullet jk} = I$ and cell probabilities

$$p_{rsjk} = p_{r\bullet\bullet} p_{\bullet s\bullet} \sum_{u=0}^{R}\sum_{v=0}^{I} \theta_{uvjk} a_u(r) b_v(s)$$

for $r = 1, \ldots, R$, $s = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$ in which $\theta_{00jk} = \theta_{r0jk} = \theta_{0sjk} = 1$.

The generalised correlations θ_{uvjk} characterise the probability model $\{p_{rsjk}\}$ in the sense that knowledge of the marginal probabilities $p_{r\bullet\bullet}$ and $p_{\bullet s\bullet}$ and the θ_{uvjk} is

equivalent to knowledge of the p_{rsjk}. There are so many generalised correlations that it is not immediately clear which hypotheses concerning them are both practical and interesting. It would be remarkable if knowledge of just the $p_{r\bullet\bullet\bullet}$ and $p_{\bullet s\bullet\bullet}$ was sufficient for the four-way probabilities $\{p_{rsjk}\}$, with all the θ_{uvjk} consistent with zero. On the other hand, investigating all generalised correlations would seem counter-productive, especially as those of higher order are difficult to interpret usefully. As many users would have some intuition about the practical use of the low order bivariate generalised correlations, our approach is to focus on particular couples (u, v), expecting most users would be interested in u, v = 1 and 2, and occasionally 3.

Taking expectation with respect to $\{p_{rsjk}\}$ and exploiting the orthonormality,

$$\begin{aligned} \mathrm{E}[Z_{uvjk}] &= \sum_{r=1}^{R}\sum_{s=1}^{I} a_u(r)b_v(s)E[N_{rsjk}]/\sqrt{I} \;=\; n_{\bullet\bullet jk}\sum_{r=1}^{R}\sum_{s=1}^{I} a_u(r)b_v(s)p_{rsjk}/\sqrt{I} \\ &= I\sum_{r=1}^{R}\sum_{s=1}^{I} a_u(r)b_v(s)p_{r\bullet\bullet\bullet}p_{\bullet s\bullet\bullet}\sum_{r'}\sum_{s'}\theta_{r's'jk}a_{r'}(r)b_{s'}(s)/\sqrt{I} \\ &= \sqrt{I}\sum_{r'}\sum_{s'}\theta_{r's'jk}\delta_{ur'}\delta_{vs'} = \theta_{uvjk}\sqrt{I}. \end{aligned}$$

Thus $\mathrm{E}[Z_{uvjk}]$ is proportional to the generalised correlation of order (u, v) for the (j, k)th level combination.

The first tests of interest assess whether, for given (u, v), the $\{\theta_{uvjk}\}$ is consistent with zero: do the unordered factors have correlation structure of this order? The test may be based on the one sample t-test applied, for the given (u, v), to the $\{Z_{uvjk}\}$. The Zs are sums and, by the central limit theorem, are asymptotically normal. We seek to test if their mean is zero using a test statistic that is scale invariant.

Next we ask if the generalised correlations differ across levels of the unordered factors. For each (u, v) reparametrize using the ANOVA model for the two-factor ANOVA. Put $\theta_{uvjk}\sqrt{I} = \mu_{uv} + B_{uvi} + C_{uvj}$ in which $\sum_{i=1}^{I} B_{uvi} = \sum_{j=1}^{J} C_{uvj} = 0$. For each such ($u$, v) there are $I + J - 2$ independent parameters; the μ_{uv} are identically zero and are only included to complete the analogy with the two-factor ANOVA model. It now follows that by the usual development that we can test $\mathrm{H}_{uv\mathrm{B}}$: $(B_{uvj}) = 0$ against $\mathrm{K}_{uv\mathrm{B}}$: $(B_{uvj}) \neq 0$ and $\mathrm{H}_{uv\mathrm{C}}$: $(C_{uvj}) = 0$ against $\mathrm{K}_{uv\mathrm{C}}$: $(C_{uvj}) \neq 0$ using the ANOVA F ratios with data

$\{Z_{uvjk}\}$. If the usual assumptions, such as the residuals being consistent with normality, are satisfied, the appropriate F-test can be used. If they are not, then resampling p-values should be applied.

If instead of the factor C we had replicates then the model would be simpler: this 'factor' shouldn't be modelled. The model for the cell probabilities is then

$$p_{rsjk} = p_{r\bullet\bullet\bullet}p_{\bullet s\bullet\bullet}\sum_{u=0}^{R}\sum_{v=0}^{I}\theta_{uvj}a_u(r)b_v(s)$$

for the same values of r, s, j and k as previously. The development is as above with the generalised correlations of order (u, v) being tested for consistency with zero by one sample t-tests. To test if a given generalised correlation varies across levels of the unordered factor uses the completely randomized design. Put $\theta_{uvj}\sqrt{I} = \mu_{uv} + B_{uvi}$ in which $\sum_{i=1}^{I} B_{uvi} = 0$. For a given (u, v), testing H_{uvB}: $(B_{uvj}) = 0$ against K_{uvB}: $(B_{uvj}) \neq 0$, is equivalent to testing if $(\theta_{uvj}) = 0$ for $j = 1, \ldots, J$ against $(\theta_{uvj}) \neq 0$.

As in Rayner & Best (2013) it is helpful to note that for the three-factor model discussed in this section, for each (u, v), $\{Z_{uvjk}\} = \{a_u(r)\ b_v(s)\}$. To see this, recall that N_{rsjk} is the number of times the rth of R distinct ordered scores is assigned to the level/replicate combination (s, j, k), and hence is 0 or 1. The only time it is non-zero,

$$Z_{uvjk} = \sum_{r=1}^{R}\sum_{s=1}^{I} a_u(r)b_v(s)N_{rsjk} / \sqrt{I} = a_u(r)\ b_v(s)/\sqrt{I}.$$

This corresponds to the response y_{ijk} that for given j and k assigned the rth score overall to the sth level of factor A. The two-factor model with replicates is similar. Thus in both models it is sufficient to apply the appropriate ANOVA with data $\{a_u(r)\ b_v(s)/\sqrt{I}\}$. Since this ANOVA is location-scale invariant, it is sufficient, as in the drugs and concentrations example in section 1, to apply the ANOVA to $\{a_u(r)\ b_v(s)\}$, for the pairs (u, v) of interest.

That the Z_{uvjk} are uncorrelated follows as in Rayner & Best (2013, Appendix 2). We give no proof here. However this lack of correlation is another reason for using

orthonormal functions. It is also the genesis of the remark near Table 2, that the analyses of different orders are, in a sense, uncorrelated.

This analysis is implemented in the drugs and concentrations example at the conclusion of section 1.

3. Equally replicated two-way tables with both factors ordered

In this section it is assumed that both factors A and B are ordered. Write N_{rstk} for the number of times the rth of R distinct ordered scores is assigned to the level/replicate combination (s, t, k). Then $\{N_{rstk}\}$ defines a four-way triply ordered table of counts of zeros and ones. As in Beh & Davy (1998) and Rayner & Best (2001, section 10.2), Pearson's independence test statistic X^2_{P} may be partitioned into components Z_{uvwk} via

$$X^2_{\mathrm{P}} = \sum_{u=1}^{R-1}\sum_{v=1}^{I-1}\sum_{w=1}^{J-1}\sum_{k=1}^{K} Z^2_{uvwk} + \sum_{v=1}^{I-1}\sum_{w=1}^{J-1}\sum_{k=1}^{K} Z^2_{0vwk} + \sum_{u=1}^{R-1}\sum_{w=1}^{J-1}\sum_{k=1}^{K} Z^2_{u0wk} + \sum_{u=1}^{R-1}\sum_{v=1}^{I-1}\sum_{k=1}^{K} Z^2_{uv0k}$$

with $Z_{uvwk} = \sum_{r=1}^{R}\sum_{s=1}^{I}\sum_{t=1}^{J} a_u(r)b_v(s)c_w(t)N_{rstk}/\sqrt{n}$, in which $\{a_u(r)\}$ is orthonormal on $\{p_{r\bullet\bullet\bullet}\}$ with $a_0(r) = 1$ for $r = 1, ..., R$, $\{b_v(s)\}$ is orthonormal on $\{p_{\bullet s\bullet\bullet}\}$ with $b_0(s) = 1$ for $s = 1, ..., I$, and $\{c_w(t)\}$ is orthonormal on $\{p_{\bullet\bullet t\bullet}\}$ with $c_0(t) = 1$ for $t = 1, ..., J$. As before, the Z_{uvwk} are proportional to generalised sample correlations as in Rayner & Beh (2009a).

As with the previous notation, put $SS_{uvw} = \sum_{k=1}^{K} Z^2_{uvwk}$ for $u = 0, 1, 2, ..., R-1$, $v = 0, 1, ..., I-1$, and $w = 0, 1, ..., J-1$, but not $(u, v, w) = (0, 0, 0)$. Thus $X^2_P = \sum_{u,v,w} S_{uvw}$: the aggregation of all the order (u, v, w) effects SS_{uvw} is X^2_{P}.

From Rayner & Beh (2009b) a possible smooth model for $\{N_{rstk}\}$ is the multinomial with count total n and cell probabilities p_{uvwk} given by

$$p_{rstk} = p_{r\bullet\bullet\bullet}p_{\bullet s\bullet\bullet}p_{\bullet\bullet t\bullet}\sum_{u=0}^{R}\sum_{v=0}^{I}\sum_{w=0}^{J}\theta_{uvw}a_u(r)b_v(s)c_w(t),$$

in which $\theta_{000} = 1$ and $\theta_{u00} = \theta_{0v0} = \theta_{00w} = 0$ for all $r, s, t \geq 1$. The component Z_{uvwk} may be shown to satisfy $E[Z_{uvwk}] = \theta_{uvw} \sqrt{n}$.

For a given triple (u, v, w) we may test H_{uvw}: $\theta_{uvw} = 0$ against K_{uvw}: $\theta_{uvw} \neq 0$ to assess if each of the complex generalised correlations θ_{uvw} is consistent with zero. Parallel to our previous argument, for each (u, v, w), $\{Z_{uvwk}\} = \{a_u(r)\ b_v(s)\ c_w(t)\}$. For recall that N_{rstk} is an indicator variable that takes its only non-zero value, 1, when the response y_{stk} for a given k at the sth level of factor A and the tth level of factor B is assigned the rth score overall. Then $Z_{uvwk} = \sum_{r=1}^{n}\sum_{s=1}^{I}\sum_{t=1}^{J} a_u(r)b_v(s)c_w(t)N_{rstk} / \sqrt{n} = a_u(r)\ b_v(s)\ c_w(t)/\sqrt{n}$. There are K such values and we wish to test if their expected value, $E[Z_{uvwk}]$, and hence θ_{uvw}, is consistent with zero. An option consistent with previous practice here would be to use the one sample t-test.

Ants Example

The data in Table 5 come from Manly (2007, p.144) and relate to the number of ants consumed by two sizes of Eastern Horned Lizards over a four month period.

Month is significant using both the usual parametric and rank transform analyses, the first order effects in Table 6. The unordered extensions of Rayner & Best (2013) to orders two and three find no significant effects. All p-values in Table 6 are from ANOVA F tests, as permutation test p-values are very similar.

TABLES 5, 6 & 7 NEAR HERE

Using the data as scores we ask two questions. First we consider bivariate generalised correlations between the data and months by taking $w = 0$. Effectively ant size and replications are combined into replications. Second we ask if there are differences in these sample generalised correlations for large and small ants. This treats the levels of the factor size as unordered.

The first question asks if the generalised correlations θ_{uv0} consistent with being zero. To facilitate quick assessments the first entry in each cell in Table 7 gives the

sample generalised correlation multiplied by $\sqrt{n}$; these are asymptotically standard normal values. The second entry in each cell is the p-value from the one-sample t-test; the third is the corresponding permutation test p-value.

Both our first glance and closer scrutiny by the t-test confirm that the generalised sample correlations of order (1, 2) and (1, 3), and only these, are significantly different from zero at the 5% level. These suggest that both quadratic and cubic month effects are required to model the bivariate data. For example consider the following.

As θ_{12} and θ_{13} are the only non-zero θ in p_{rstk}, sum out both months and replicates. This results in a doubly ordered bivariate model for ants consumed and month:

$$p_{rs} = p_{r\bullet}p_{\bullet s}\{1 + \theta_{12}a_1(r)b_2(s) + \theta_{13}a_1(r)b_3(s)\}.$$

The condition probability function of $X|y$ is $p_{rs}/p_{\bullet s}$. It follows that

$$\mathrm{E}[a_1(X)|y] = \sum\nolimits_r a_1(r)p_{r\bullet}\{1 + \theta_{12}a_1(r)b_2(s) + \theta_{13}a_1(r)b_3(s)\}$$

$$= \sum\nolimits_x a_1(x)p_{r\bullet} + \sum\nolimits_r a_1(r)p_{r\bullet}\theta_{12}a_1(r)b_2(s) + \sum\nolimits_r a_1(x)p_{r\bullet}\theta_{13}a_1(x)b_3(s)\}$$

$$= \theta_{12}b_2(s) + \theta_{13}b_3(s),$$

using the orthonormality in the final step. Now $a_1(r) = (x - \mu)/\sigma$, so $x = \mu + \sigma a_1(r)$. Hence

$$\mathrm{E}[X|y] = \mu + \sigma\,\mathrm{E}[a_1(X)|y] = \mu + \sigma\{\theta_{12}b_2(s) + \theta_{13}b_3(s)\}.$$

For these data $\mu = 213.125$, $\sigma^2 = 97620.776$, $\theta_{12} = -0.453$ and $\theta_{13} = -0.602$. Plots of the data and the conditional means against months show the two-parameter model gives a reasonable fit. The conditional means for June to September are 155.825, 102.108, 607.058 and –12.492; to avoid the anomalous negative mean further θ_{rs} would need to be included in the model. However the above model is certainly indicative.

The second question seeks to compare θ_{uv1} with θ_{uv2} for each (u, v). This can be achieved using a one-way ANOVA with factor size or, equivalently a two-sample t-test of equality of the means for small and large ants, applied to the data $\{a_u(r)\ b_v(s)\}$. The p-values are in Table 8; the first entry in each cell is the p-value from the two-sample t-test for a mean of zero against two-sided alternatives; the second is the corresponding permutation test p-value.

It seems the only p-value that is not substantial is that of order (2, 3), and this suggests a weak effect at best. Thus it appears that the bivariate generalised correlations do not differ with size: large and small ants behave similarly with respect to their generalised correlations.

TABLES 8 & 9 NEAR HERE

We now calculate the trivariate generalised correlations θ_{uvw}, for convenience multiplied by $\sqrt{n}$, in which u refers to the data, v to months and w to size. Apart from θ_{000}, which is one by convention, only those correlations with at least two subscripts positive are defined. For $w = 0$, the generalised correlations multiplied by $\sqrt{n}$ are as in Table 7. For $w = 1$ see the first entries in each cell in Table 9. The correlations with $u = 0$ are all zero because, since $a_0(r) = 1$, they reflect the table $\{b_v(s)\ c_1(t)\}$ that is independent of the data.

The second entry is the ANOVA F test p-value; the third is the corresponding permutation test p-value. The agreement between these p-values is merely reasonable. We also tried referring $\hat{\theta}_{uvw}\sqrt{n-2}/\sqrt{1-\hat{\theta}^2_{uvw}}$ to the t_{n-2} distribution. This gave good agreement with the permutation test p-values in Tables 3, 7 and 9. However we could find no reasonable approximation for Tables 4 and 8.

None of these generalised correlations are significantly large, and so are not required to model the data. The only generalised correlations that are required are those of order (1, 2, 0) and (1, 3, 0) discussed previously.

The treatment using the ranks as scores is entirely parallel.

4. Summary: multi-factor possibly not equally replicated designs

Suppose we have an m-factor design with the first t factors having ordered levels while the remaining $m - t$ factors do not. Here $t = 0, 1, \ldots$ and $m = 1, 2, \ldots$. There are I_j levels of the jth factor, and $n_{s_1 \ldots s_t k_1 \ldots k_{m-t}}$ replicates of the $(s_1, \ldots, s_t, k_1, \ldots, k_{m-t})$th combination of levels. If $t = 0$ and all $n_{s_1 \ldots s_t k_1 \ldots k_{m-t}}$ are one (the no replication case) m should be at least 2. The appropriate ANOVA for the design may have dubious diagnostics, in which case a nonparametric test that makes weaker assumptions is sought. It may also be the case that the ANOVA gives only a location assessment of the model, and the scenario that generated the data motivates investigation of more comprehensive effects.

Suppose that $N_{r s_1 \ldots s_t k_1 \ldots k_{m-t} k}$ counts the number of times the rth of R distinct ordered scores is assigned to the $(s_1, \ldots, s_t, k_1, \ldots, k_{m-t}, k)$th level/replicate combination. The total number of observations is $n = N_{\bullet \ldots \bullet}$. To identify interesting and appropriate components, we could proceed by decomposing the Pearson test statistic in the equal replicates case, or by generalising results from the particular cases already discussed. Whatever the motivation, write $p_{r s_1 \ldots s_t k_1 \ldots k_{m-t}} = N_{r s_1 \ldots s_t k_1 \ldots k_{m-t}} / n$ and construct sets of orthonormal polynomials, $\{a_u(r)\}$ on $p_{r \bullet \ldots \bullet}$ and for $j = 1, \ldots, t$, $\{a_{v_w}(s_w)\}$ on $p_{\bullet \ldots \bullet s_j \bullet \ldots \bullet}$. The zeroth order polynomials are all identically one. For a given combination of the levels of the factors with ordered levels, namely $(v_1, \ldots, v_t)$, define

$$Z_{u v_1 \ldots v_t k_1 \ldots k_{m-t} k} = \sum_{r=1}^{R} \sum_{s_1=1}^{I_1} \ldots \sum_{s_t=1}^{I_t} a_u(r) a_{v_1}(s_1) \ldots a_{v_t}(s_t) N_{r s_1 \ldots s_t k_1 \ldots k_{m-t} k} \, .$$

We now give a smooth model for the table of counts $\{N_{r s_1 \ldots s_t k_1 \ldots k_{m-t} k}\}$. For a given level/replicate combination $(s_1, \ldots, s_t, k_1, \ldots, k_{m-t}, k)$ assume there is precisely one ordered score appropriate, such as the mid-rank or the datum. Hence for the $(v_1, \ldots, v_t)$th combination of the factors with ordered levels and the level/replicate combination $(k_1, \ldots, k_{m-t}, k)$ define a multinomial with parameters 1 and

$$p_{rs_1\ldots s_t k_1\ldots k_{m-t}} = p_{r\bullet\ldots\bullet}p_{\bullet s_1\bullet\ldots\bullet}\cdots p_{\bullet\ldots\bullet s_t\bullet\ldots\bullet}\sum_{u=1}^{R}\sum_{v_1=0}^{I_1}\cdots\sum_{v_t=0}^{I_t}\theta_{uv_1\ldots v_t k_1\ldots k_{m-t}k}a_u(r)a_{v_1}(s_1)\ldots a_{v_t}(s_t).$$

By the approach in Rayner & Best (2012, Appendix 2) E[$Z_{uv_1\ldots v_t k_1\ldots k_{m-t}k}$] is proportional to $\theta_{uv_1\ldots v_t k_1\ldots k_{m-t}}$, and as in Rayner & Beh (2009b) the *Z*s are efficient score statistics and appropriate test statistics for testing hypotheses about the $\theta_{uv_1\ldots v_t k_1\ldots k_{m-t}}$.

For all particular choices of (u, v_1, …, v_t), to test for all $\theta_{uv_1\ldots v_t k_1\ldots k_{m-t}} = 0$ (each against $\theta_{uv_1\ldots v_t k_1\ldots k_{m-t}} \neq 0$), reparameterize to the parameters of the class of ANOVAs appropriate for the design. For example, if the ANOVA was an m-way factorial model with replication, then use the ($m - t$)-way factorial model with main and interaction effects up to order $m - t$. Apply that ANOVA analysis to $\{a_u(r)a_{v_1}(s_1)\ldots a_{v_t}(s_t)\}$. This follows because the $N_{rs_1\ldots s_t k_1\ldots k_{m-t}k}$ are indicator functions, being either zero or one, when $Z_{uv_1\ldots v_t k_1\ldots k_{m-t}k}$ takes the value $a_u(r)a_{v_1}(s_1)\ldots a_{v_t}(s_t)$.

For some problems it may be helpful to ignore the ordering for some factors with ordered levels, to assess less complex generalised correlations.

In general we prefer to use permutation test p-values. However, if that is not possible then p-values based on ANOVA F tests are generally reasonable. As always conclusions are conditional on the assumptions made, and in the absence of exact p-values analysts will be aware some effects may be missed and spurious effects added. Moreover many tests are being made on the one set of data. Our view is that rather than correcting for this the analyst should consider the analysis to be a first pass at model building or preliminary data analysis, and cast recommendations in this light.

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 NZ 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 NZ J Stat.,* **41**, 233–246.

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

KESELMAN, H.J., CRIBBIE, R.A. & WILCOX, R.R. (2002). Pairwise multiple comparison tests when data are nonnormal. *Educational and Psychological Measurement,* **62**: 420.

MANLY, B.F.J. (2007). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Boca Raton: Chapman & Hall/CRC.

RAYNER, J.C.W. & BEH, E.J. (2009a). Towards a better understanding of correlation. *Stat Neerl*., **63**, 324–333.

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

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

RAYNER, J.C.W., BEST, D.J., D.J., BROCKHOFF, P.B. & RAYNER, G.D. (2005). *Nonparametrics For Sensory Science: A More Informative Approach*. Ames: Blackwell.

RAYNER, J.C.W. & BEST, D.J. (2013). Extended ANOVA and rank transform procedures. *Australian & NZ Journal of Statistics*, **55**, 305–319.

TABLE 1

Drugs and Concentrations data

Concentration	Drug 1	Drug 2
2	18 (0), 2 (1) / 0.1	16 (0), 3 (1), 2 / 0.25
5	12 (0), 6 (1), 2 (2) / 0.5	8 (0), 8 (1), 3 (2), 3 / 0.85
10	3 (0), 7 (1), 6 (2), 4 (3) / 1.55	1 (0), 5 (1), 8 (2), 6 (3) / 1.95

TABLE 2

Extended unordered analyses

	p-values with ranks as scores			p-values with data as scores		
Source	First Order	Second Order	Third Order	First Order	Second Order	Third Order
Drug	0.030	0.856	0.923	0.032	0.840	0.922
Concentration	0.000	0.123	0.979	0.000	0.091	0.717
Interaction	0.813	0.539	0.935	0.736	0.651	0.822
Shapiro-Wilk p-value	0.021	0.000	0.000	0.003	0.000	0.000

TABLE 3

Sample generalised correlations multiplied by √n, with one sample t-test and permutation test p-values

	Concentration order (v)	
Outcome order (u)	1	2
1	7.2902/ 0.000, 0.000	0.0503/ 0.957, 0.960
2	0.2921/ 0.785, 0.771	2.0493/ 0.027, 0.041
3	– 0.0263/ 0.982, 0.979	– 0.2095/ 0.803, 0.836

TABLE 4

Two sample t-test and permutation test p-values comparing generalised correlations across drugs

	Concentration (v)	
Outcome order (u)	1	2
1	0.701, 0.701	0.703, 0.703
2	0.339, 0.343	0.635, 0.636
3	0.921, 0.921	0.670, 0.673

TABLE 5

Ants data

Ants consumed	Lizard size	
Month	Small	Large
June	13, 242, 105	182, 21, 7
July	8, 59, 20	24, 312, 68
August	515, 488, 88	460, 1223, 990
September	18, 44, 21	140, 40, 27

TABLE 6

Extended unordered analyses

	p-values with ranks as scores			p-values with the data as scores		
Source	First Order	Second Order	Third Order	First Order	Second Order	Third Order
Size	0.233	0.955	0.768	0.051	0.618	0.784
Month	0.006	0.189	0.284	0.000	0.639	0.868
Interaction	0.379	0.328	0.686	0.062	0.279	0.793
Shapiro-Wilk p-value	0.982	0.657	0.920	0.086	0.482	0.144

TABLE 7

Extended order analysis of generalised correlations θ_{uv0}

	Month order (v)		
Ants consumed order (u)	1	2	3
1	0.7109/0.276, 0.510	– 2.2180/0.023, 0.016	– 2.9506/0.016, 0.001
2	0.2420/0.735, 0.820	0.7696/0.453, 0.468	1.1046/0.382, 0.292
3	– 0.7070/0.393, 0.512	0.5992/0.560, 0.588	0.3238/0.787, 0.764

TABLE 8

Two sample t-test and permutation test p-values comparing generalised correlations across ant size

$u\backslash v$	1	2	3
1	0.349, 0.401	0.173, 0.201	0.375, 0.423
2	0.963, 0.966	0.638, 0.651	0.094, 0.101
3	0.488, 0.528	0.822, 0.835	0.410, 0.447

TABLE 9

Genuine third order generalised correlations θ_{uv1} multiplied by $\sqrt{n}$, with approximate p-values obtained by the one sample t-test and permutation testing

$u\backslash v$	0	1	2	3
1	1.2237/0.229, 0.264	0.610/0.351, 0.571	– 1.2576/0.215, 0.251	– 1.0287/0.426, 0.251
2	0.5288/0.608, 0.620	0.0338/0.962, 0.976	– 0.4890/0.635, 0.646	– 2.0758/0.093, 0.038
3	0.3232/0.754, 0.772	0.5793/0.486, 0.590	– 0.2358/0.819, 0.833	1.0008/0.400, 0.348