Extensions to the Kruskal-Wallis Test and a Generalised Median Test with Extensions

J.C.W. RAYNER   john_rayner@uow.edu.au

Department of Applied Statistics, University of Wollongong, Northfields Avenue, NSW 2522, Australia

D.J. BEST   johnb@biomsyd.dfst.csiro.au

CSIRO Mathematical and Information Sciences, PO Box 52, North Ryde, NSW 2113, Australia

Abstract. The data for the tests considered here may be presented in two-way contingency tables with all marginal totals fixed. We show that Pearson's test statistic X_P^2 (P for Pearson) may be partitioned into useful and informative components. The first detects location differences between the treatments, and the subsequent components detect dispersion and higher order moment differences. For Kruskal-Wallis-type data when there are no ties, the location component is the Kruskal-Wallis test. The subsequent components are the extensions. Our approach enables us to generalise to when there are ties, and to when there is a fixed number of categories and a large number of observations. We also propose a generalisation of the well-known median test. In this situation the location-detecting first component of X_P^2 reduces to the usual median test statistic when there are only two categories. Subsequent components detect higher moment departures from the null hypothesis of equal treatment effects.

1.   Introduction

The idea of decomposing a test into orthogonal contrasts, as in the analysis of variance, has long been appreciated by statisticians as a way of making hypothesis tests more informative. In the authors' smooth goodness of fit work (see Rayner and Best, 1989), a similar approach is pursued. Omnibus test statistics are partitioned into smooth components. We define the components of a test statistic to be asymptotically pairwise independent, with each asymptotically having the chi-squared distribution, and such that their sum gives the original test statistic. The components provide powerful directional tests and permit a convenient and informative scrutiny of the data. This approach is applied to Spearman's test in Best and Rayner (1996) and Rayner and Best (1996a); Rayner and Best (1996b) gave an overview of this approach applied to several commonly used nonparametric tests, including the Friedman and Durbin tests.

Data for a generalisation of the median test that we subsequently propose, and for the Kruskal-Wallis test both with and without ties, may be presented in the form of two-way tables with fixed marginal totals. We derive the covariance matrix

of entries in such tables and then partition a multiple of X_P^2 into components that detect location and higher moment differences between rows.

For Kruskal-Wallis-type data when there are no ties, the location component is the Kruskal-Wallis test. Our approach enables us to generalise to when there are ties, and to when there is a fixed number of categories and a large number of observations. We also propose a generalisation of the well-known median test. The location detecting first component of X_P^2 reduces to the usual median test statistic when there are only two categories. Using more categories allows components other than this location component to be calculated. These additional components, that detect dispersion and higher moment effects, are not available when using the usual median test.

The structure of this paper is as follows. In the next section the model for two-way contingency tables with fixed marginal totals is given, and Pearson's X_P^2 is derived as a test statistic for the null hypothesis of like rows. In section three a multiple of X_P^2 is partitioned into components. The material in section 2 and 3 will be familiar to many readers, but is necessary background for the new work. In section four it is shown that when there are no ties the first component is the usual Kruskal-Wallis statistic. The non-location detecting components are our extensions. Section five generalises the treatment to when there are ties. Section six introduces a generalisation of the usual median X^2 test, which is thus identified as a location detecting test; the extensions permit dispersion and other effects to be detected.

2.   A Model and Pearson's X^2 Test

Suppose we have a two-way table of counts N_{ij}, with $i = 1, \ldots, r$ and $j = 1, \ldots, c$. The row and column totals, respectively $n_{i.}$, $i = 1, \ldots, r$ and $n_{.j}$, $j = 1, \ldots, c$ are known constants. Under the null hypothesis of simple random sampling, the likelihood was given by Roy and Mitra (1956) as

$$\left\{ \prod_{i=1}^{r} n_{i.} \right\} \left\{ \prod_{j=1}^{c} n_{.j} \right\} \Big/ \left\{ n_{..} \prod_{i=1}^{r} \prod_{j=1}^{c} n_{ij} \right\},$$

in which $n_{..} = \sum_i n_{i.} = \sum_j n_{.j}$ is the grand total of the observations. The models for tables with just one set of marginal totals fixed, or only the grand total fixed, are quite different from our model in which all row and column totals are fixed. See Lancaster (1969, chapter XI section 2, pp. 212-217). This likelihood can be expressed as a product of extended or multivariate hypergeometric probability functions:

$$\prod_{i=2}^{r} \left\{ \left[\prod_{j=1}^{c} {}^{n_{1j}+\ldots+n_{ij}} C_{n_{ij}} \right] \Big/ {}^{n_{1.}+\ldots+n_{i.}} C_{n_{i.}} \right\}.$$

To find moments of the N_{ij}, expectations may be taken with respect to the distribution of the second row conditional on knowledge of the column sums of the

first two rows, then conditional on the column sums of the first three rows, and so on. It suffices to know the moments of the extended hypergeometric distribution. Details are given in the Appendix. We find

$$E[N_{ij}] = n_{i.}n_{.j}/n_{..}, \quad i = 1, \ldots, r \quad \text{and} \quad j = 1, \ldots, c.$$

Write $\boldsymbol{N}_i = (N_{i1}, \ldots, N_{ic})^T$, $i = 1, \ldots, r$ and $\boldsymbol{N}^T = (\boldsymbol{N}_1^T, \ldots, \boldsymbol{N}_r^T)$, so that $\boldsymbol{N}$ is the vector of all the cell counts. The joint covariance matrix of $\boldsymbol{N}_i$ and $\boldsymbol{N}_j$ is, for $i \neq j$,

$$cov(\boldsymbol{N}_i, \boldsymbol{N}_j) = -\frac{n_{i.}n_{j.}}{n_{..}^2} \left\{ diag\left(\frac{n_{.r}n_{..}}{(n_{..}-1)}\right) - \left(\frac{n_{.r}n_{.s}}{n_{..}-1}\right) \right\}.$$

Write $f_j = n_{.j}/n_{..}, j = 1, \ldots, c$, and

$$\boldsymbol{R} = diag\left(\frac{n_{.r}n_{..}}{(n_{..}-1)}\right) - \left(\frac{n_{.r}n_{.s}}{n_{..}-1}\right).$$

The covariance matrix of $\boldsymbol{N}$ is $cov(\boldsymbol{N}) = \{diag(f_j) - (f_i f_j)\} \otimes \boldsymbol{R}$, where $\otimes$ is the direct or Kronecker product. See Lancaster (1969) for details about direct or Kronecker sums and products. Now define the standardised cell counts

$$Z_{ij} = (N_{ij} - E[N_{ij}])/\sqrt{E[N_{ij}]}, \quad i = 1, \ldots, r \quad \text{and} \quad j = 1, \ldots, c,$$

$$\boldsymbol{Z} = (Z_{11}, \ldots, Z_{1c}, \ldots, Z_{r1}, \ldots, Z_{rc})^T,$$

$\boldsymbol{I}_a$ the a by a identity matrix and $\boldsymbol{1}_a$ the a by 1 vector with every element 1. Then

$$cov(\boldsymbol{Z}) = \{\boldsymbol{I}_r - (\sqrt{[f_i f_j]})\} \otimes \boldsymbol{R}.$$

The matrix $\{\boldsymbol{I}_r - (\sqrt{[f_i f_j]})\}$ has $r - 1$ latent roots 1 and one latent root zero. The latent roots of $\boldsymbol{R}$ are difficult to find in general, but their asymptotic limits follow from Lancaster (1969, Chapter V.3). Lancaster showed that the quadratic form with vector the standardised cell counts and matrix essentially $\boldsymbol{R}$, is the familiar Pearson goodness of fit statistic, with asymptotic distribution χ_{c-1}^2. Hence the latent roots of $\boldsymbol{R}$ are asymptotically one $c - 1$ times and zero once. So under the null hypothesis of simple random sampling, $\boldsymbol{Z}$ has zero mean and covariance matrix $cov(\boldsymbol{Z})$, which asymptotically has $(r-1)(c-1)$ latent roots one, and the remaining $r + c - 1$ latent roots zero.

In the well known and often used 'classical' model, r and c are fixed and the total count $n_{..} \to \infty$. The test statistic X_P^2 is given by

$$X_P^2 = \sum_{i=1}^r \sum_{j=1}^c \left(N_{ij} - \frac{n_{i.}n_{.j}}{n_{..}} \right)^2 / \left(\frac{n_{i.}n_{.j}}{n_{..}} \right) = \boldsymbol{Z}^T \boldsymbol{Z}.$$

We now confirm that our model leads to this test statistic. Suppose $\boldsymbol{H}$ is orthogonal and diagonalises $cov(\boldsymbol{Z})$. Asymptotically we then have

$$\boldsymbol{H}^T \mathrm{cov}(\boldsymbol{Z})\boldsymbol{H} = \boldsymbol{I}_{(r-1)(c-1)} \oplus O_{(r+c-1)},$$

where $\oplus$ means direct or Kronecker sum. Define $\boldsymbol{Y} = \boldsymbol{H}^T\boldsymbol{Z}$. Now $\boldsymbol{Z}^T\boldsymbol{Z} = \boldsymbol{Y}^T\boldsymbol{Y}$, in which $\boldsymbol{Y}$, by the multivariate Central Limit Theorem, is asymptotically $N_{rc}(0, [\boldsymbol{I}_{(r-1)(c-1)} \oplus 0_{(r+c-1)}])$ under the null hypothesis of simple random sampling. It follows that under the null hypothesis, $X_P^2 = \boldsymbol{Z}^T\boldsymbol{Z} = \boldsymbol{Y}^T\boldsymbol{Y}$ asymptotically has the $\chi^2_{(r-1)(c-1)}$ distribution.

3. Partitioning Pearson's Statistic

We now show that X_P^2 may be partitioned into components, the sth of which detects sth moment departures from the null hypothesis of similarly distributed rows (treatments).

The elements Y_i of $\boldsymbol{Y}$ are such that $X_P^2 = \sum_{i=1}^{rc} Y_i^2$. There is some choice in defining the Y_i, as $\boldsymbol{H}$ is not yet fully specified. In doing so, our aim is to find Y_i that can be easily and usefully interpreted. To achieve one such partition, first suppose that $\{g_s(j)\}$ is the set of polynomials orthonormal on $\{n_{.j}/n_{..}\}$. See the Appendix for the definitions of the first two polynomials and the derivation of subsequent polynomials. This approach results, when there are no ties, in the first component being the Kruskal-Wallis test. Write g_s for the c by 1 vector with elements $g_s(j)$. Define $\boldsymbol{G}$ by

$$\boldsymbol{G} = [\boldsymbol{G} \quad \dots \quad \boldsymbol{G}_c]/\sqrt{c}$$

in which $\boldsymbol{G}_s$ is the rc by r matrix

$$\boldsymbol{G}_s = \begin{bmatrix} g_s & 0 & \dots & \\ 0 & g_s & & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_s \end{bmatrix}, \quad s = 1, \dots, c-1, \quad \text{and}$$

$$\boldsymbol{G}_c = \begin{bmatrix} 1_c & 0 & \dots & \\ 0 & 1_c & & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1_c \end{bmatrix}, \quad \text{is also } rc \text{ by } r.$$

Define $\boldsymbol{Y} = \sqrt{\frac{n-1}{n}}\,\boldsymbol{G}^T\boldsymbol{Z}$. The elements of $\boldsymbol{Y}$ may be considered in blocks of r, the sth block corresponding to the polynomial of order s. These blocks are asymptotically mutually independent. Write $\boldsymbol{Y}^T = (\boldsymbol{V}_1^T, \dots, \boldsymbol{V}_c^T)$, in which

$$\boldsymbol{V}_1 = (Y_1, \dots, Y_r)^T, \dots, \boldsymbol{V}_{c-1} = (Y_{(c-2)r+1}, \dots, Y_{(c-1)r})^T,$$

and $\boldsymbol{V}_c = 0$ (all the $\boldsymbol{V}_s$ are r by 1) so that

$$\left(\frac{n-1}{n}\right) X_P^2 = \left(\frac{n-1}{n}\right) \boldsymbol{Z}^T \boldsymbol{Z} = \boldsymbol{Y}^T \boldsymbol{Y} = \boldsymbol{V}_1^T \boldsymbol{V}_1 + \ldots + \boldsymbol{V}_{c-1}^T \boldsymbol{V}_{c-1}.$$

This partitions $\left(\frac{n-1}{n}\right) X_P^2$ into components $\boldsymbol{V}_s^T \boldsymbol{V}_s$, $s = 1, \ldots, c-1$. The $\boldsymbol{V}_s$ are asymptotically mutually independent and asymptotically $N_r(0, \boldsymbol{I}_{(r-1)} \oplus 0)$, so that the $\boldsymbol{V}_s^T \boldsymbol{V}_s$ are asymptotically mutually independent χ_{r-1}^2. Explicitly we have, for $s = 1, \ldots, c-1$,

$$\boldsymbol{V}_s = \frac{\sqrt{(n-1)}}{n} \boldsymbol{G}_s^T \boldsymbol{Z} = \frac{\sqrt{(n-1)}}{n} \left(\sum_{j=1}^c g_s(j) Z_{ij} \right).$$

Because $\boldsymbol{V}_s$ involves, through g_s, a polynomial of order s, the elements of $\boldsymbol{V}_s$ are polynomials of order s in the elements of $\boldsymbol{N}$. Under the null hypothesis $E[\boldsymbol{Z}] = 0$, but when this is not true $E[\boldsymbol{V}_s]$ involves moments up to order s of $\boldsymbol{Z}$. So for $s = 1, \ldots, r-1$, $\boldsymbol{V}_s^T \boldsymbol{V}_s$ detects sth moment departures from the null hypothesis of similarly distributed rows (treatments).

Instructors Example. See Conover (1980, p. 233). Three instructors assign grades in five categories according to the following table.

| | Grade | | | | | |
	A	B	C	D	E	Total
Instructor 1	4	14	17	6	2	43
Instructor 2	10	6	9	7	6	38
Instructor 3	6	7	8	6	1	28
Total	20	27	34	19	9	109

Conover (1980) found the Kruskal-Wallis statistic adjusted for ties to be 0.3209, which is to be compared with the χ_2^2 (5%) point of 5.991. We find the location detecting component $\boldsymbol{V}_1^T \boldsymbol{V}_1$ to have P-value 0.85, confirming, as Conover reported, that "none of the instructors can be said to grade higher or lower than the others on the basis of the evidence presented". However the dispersion detecting component $\boldsymbol{V}_2^T \boldsymbol{V}_2$ has P-value 0.01, indicating a significant variability difference. From the data it appears that the first instructor is less variable than the other two. In fact, $9.643 = (-2.113)^2 + (2.274)^2 + (-0.031)^2$, with the elements of $\boldsymbol{v}_2 = (-2.113, 2.274, -0.031)^T$ being values of approximately standard normally distributed contributions from instructors 1, 2 and 3 respectively. The first instructor is less variable than the third who is less variable than the second. This can be formalised by a LSD analysis. The residual $X_P^2 - \boldsymbol{V}_1^T \boldsymbol{V}_1 - \boldsymbol{V}_2^T \boldsymbol{V}_2$ has P-value 0.75, indicating no further effects in the data.

Partition of X_P^2 for instructor's data			
Statistics	Degrees of Freedom	Value	P-value
$\boldsymbol{V}_1^T \boldsymbol{V}_1$	2	0.324	0.85
$\boldsymbol{V}_2^T \boldsymbol{V}_2$	2	9.643	0.01
$X_P^2 - \boldsymbol{V}_1^T \boldsymbol{V}_1 - \boldsymbol{V}_2^T \boldsymbol{V}_2$	4	2.021	0.75
X_P^2	8	11.985	0.15

4.   The Kruskal-Wallis Test with No Ties

We now consider models that lead to the Kruskal-Wallis test when there are no ties. The latent roots of $cov(\boldsymbol{Z})$ will be found explicitly rather than asymptotically as in section 2. We show that X_P^2 is not an appropriate test statistic, but nevertheless, its components are. The first component is the Kruskal-Wallis test statistic, and the subsequent components provide informative extensions.

Suppose we have distinct observations x_{ij}, being the jth of n_i observations on the ith of t treatments. All $n = n_1 + ... + n_t$ observations are combined, ordered, ranked, and the sums R_i of the ranks obtained by the ith treatment calculated. The Kruskal-Wallis statistic is

$$H = \{12/[n(n+1)]\}\Sigma_i R_i^2/n_i - 3(n+1).$$

See for example, Conover (1980, section 5.2). The data may be presented as an t by n contingency table of counts $\{N_{ij}\}$, with $N_{ij} = 1$ if rank j is allotted to treatment i, and $N_{ij} = 0$ if rank j is allotted to some other treatment. The row and column totals are all fixed: the row totals are the treatment sample sizes, so that $n_{i.} = n_i$ for $i = 1, \ldots, t$, while the column totals are all one: $n_{.j} = 1$ for $i = 1, \ldots, n$. Such a table has $X_P^2 = (t-1)n$ no matter what the $\{N_{ij}\}$. Since X_P^2 is constant, it has P-value 1. Clearly X_P^2 is not a suitable test statistic.

The model of section 2 holds, except that now

$$\boldsymbol{R} = \frac{n}{n-1}\boldsymbol{I_n} - \frac{1}{n-1}\boldsymbol{1}_n\boldsymbol{1}_n^T.$$

This matrix has one latent root one and $n-1$ latent roots $n/(n-1)$. It follows that $cov(\boldsymbol{Z})$ has $(t-1)(n-1)$ latent roots $n/(n-1)$, and the remaining $t + n - 1$ latent roots zero. So if $\boldsymbol{H}$ is orthogonal and diagonalises $cov(\boldsymbol{Z})$ then

$$\boldsymbol{H}^T cov(\boldsymbol{Z})\boldsymbol{H} = \frac{n}{(n-1)}\left[\boldsymbol{I}_{(t-1)(n-1)} \oplus 0_{(t+n-1)}\right].$$

Define

$$\boldsymbol{Y} = \sqrt{\left(\frac{n-1}{n}\right)} \boldsymbol{H}^T \boldsymbol{Z}.$$

Then

$$\boldsymbol{Y}^T \boldsymbol{Y} = \left(\frac{n-1}{n}\right) \boldsymbol{Z}^T \boldsymbol{Z} = \left(\frac{n-1}{n}\right) X_P^2.$$

With r replaced by t and c replaced by n, this is the same as in section three.

As in section 2, we are interested in the distribution theory as $n \to \infty$. However there $\boldsymbol{Z}$ was an rc by 1 vector of fixed length; here $\boldsymbol{Z}$ is a tn by 1 vector. Fortunately, it is not the asymptotic distribution of $\boldsymbol{Z}$ that is required. First recall that X_P^2 has a fixed value, $(t-1)n$, for all tables, and so is not available as a test statistic. Second, as in section three, the multivariate Central Limit Theorem shows that each $\boldsymbol{V}_s$ is asymptotically $N_t(0, \boldsymbol{I}_{(t-1)} \oplus 0)$. Moreover consideration of all pairs $\boldsymbol{V}_s$, $\boldsymbol{V}_t$ shows that they are asymptotically jointly multivariate normal, and since their covariance matrix is zero, they are asymptotically pairwise independent. The $\boldsymbol{V}_s^T \boldsymbol{V}_s$ still partition $\left(\frac{n-1}{n}\right) X_P^2$. It is the pairwise independence and convenient χ_{t-1}^2 distribution of each $\boldsymbol{V}_s$ that makes data analysis so informative and convenient. What is lost by the unavailability of X_P^2, is demonstrated in the Employees Example below: there is no residual available to assess if there are higher moment differences between the treatments.

We now show that provided there are no ties, $\boldsymbol{V}_1^T \boldsymbol{V}_1$ is the Kruskal-Wallis statistic, so that the subsequent $\boldsymbol{V}_s^T \boldsymbol{V}_s$ provide extensions to the Kruskal-Wallis test. First note that the $\{g_s(j)\}$ is the set of polynomials orthonormal on the discrete uniform distribution, so that $g_1(j) = aj + b$, $j = 1, \ldots, n$, in which

$$a = \sqrt{12/(n^2 - 1)} \ \text{ and } \ b = -\sqrt{3(n+1)/(n-1)} = -\{(n+1)/2\}a.$$

The rank sum for treatment i, R_i, is $\displaystyle\sum_{j=1}^{n} jN_{ij}$, $i = 1, \ldots, t$. Now since $n_{.j} = 1$ for $j = 1, \ldots, n$,

$$\sum_j g_1(j)\sqrt{E[N_{ij}]} = \sqrt{n_i n} \sum_j g_1(j)g_0(j)(1/n) = 0$$

and

$$\begin{aligned}
\sum_j Z_{ij}g_1(j) &= \sqrt{[n/n_i]} \sum_j N_{ij}(aj + b) = \sqrt{[n/n_i]}\{aR_i + bn_i\} \\
&= a\sqrt{[n/n_i]}\left\{R_i - \frac{n+1}{2}n_i\right\}.
\end{aligned}$$

Now

$$\begin{aligned}
\boldsymbol{V}_1^T \boldsymbol{V}_1 &= Y_1^2 + \ldots + Y_t^2 = \frac{n-1}{n^2} \sum_{i=1}^t \left(\sum_{j=1}^n g_1(j) Z_{ij} \right)^2 \\
&= \frac{n-1}{n^2} a^2 n \sum_{i=1}^t \left\{ \frac{R_i}{\sqrt{n_i}} - \frac{n+1}{2} \sqrt{n_i} \right\}^2 = \frac{12}{n(n+1)} \sum_{i=1}^t \frac{R_i^2}{n_{i.}} - 3(n+1)
\end{aligned}$$

after some manipulation. This is the Kruskal-Wallis statistic, well known to be sensitive to location departures from the null hypothesis. Since $\boldsymbol{V}_s$ assesses sth moment departures between treatments, we have partitioned the statistic $(\frac{n-1}{n}) X_P^2$ into asymptotically pairwise independent components, $\boldsymbol{V}_s^T \boldsymbol{V}_s$, $s = 1, \ldots, n-1$, each with the χ_{t-1}^2 distribution, and such that the sth detects sth moment departures from the hypothesis of similarly distributed rows (treatments). Since the first of these is the Kruskal-Wallis statistic, the subsequent components provide extensions to the Kruskal-Wallis test.

Employees Example. Conover (1980, p. 238, exercise 2) gave an exercise in which 20 new employees are randomly assigned to four different job training programmes. At the end of their training the employees are ranked, with a low ranking reflecting a low job ability.

Programme	Ranks
1	2, 4, 6, 7, 10
2	1, 3, 8, 11, 12
3	5, 14, 16, 19, 20
4	9, 13, 15, 17, 18

The value of the Kruskal-Wallis statistic is 9.72, with χ_3^2 P-value 0.021, but Monte Carlo permutation test P-value 0.010. The latter is more likely to be accurate as the sample size is small. Further components are not significant. An LSD analysis can be used to show that programmes 1 and 2 and programmes 3 and 4 are equally effective, with 3 and 4 being superior.

5.   The Kruskal-Wallis Test with Ties

If there are ties, the data may be presented as an t by n^* contingency table of counts $\{N_{ij}\}$, with the row totals are fixed at the treatment sample sizes, so again $n_{i.} = n_i$, $i = 1, \ldots, t$, while the column totals are no longer all one. The covariance matrix of $\boldsymbol{Z}$ is

$$cov(\boldsymbol{Z}) = \{\boldsymbol{I_t} - (\sqrt{[f_i f_j]})\} \otimes \boldsymbol{R} \quad \text{and} \quad \boldsymbol{R} = diag\left(\frac{n_{.u} n_{..}}{(n_{..} - 1)} \right) - \left(\frac{n_{.u} n_{.v}}{n_{..} - 1} \right).$$

As in section 2, the latent roots of $\boldsymbol{R}$ are zero once and asymptotically one $n^* - 1$ times. It follows that $cov(\boldsymbol{Z})$ has $(t-1)(n^* - 1)$ latent roots asymptotically one, and the remaining $t + n^* - 1$ latent roots zero. With suitable modifications the partitioning of section three holds. For $s = 1, \ldots, n^* - 1$,

$$\boldsymbol{V}_s = \boldsymbol{G}_s^T \boldsymbol{Z} / \sqrt{n^*} = \sum_{j=1}^{n^*} g_s(j) Z_{ij} / \sqrt{n^*}.$$

Note that $\{g_s(j)\}$ is the set of polynomials orthonormal on $\{n_{\cdot j}/n_{\cdot\cdot}\}$, not on the discrete uniform as in the previous section when there were no ties. This is the partition derived in section 3 for X_P^2. So the first component of X_P^2 in the Instructors example is the Kruskal-Wallis statistic corrected for ties. The subsequent components are extensions to the Kruskal-Wallis test adjusted for ties. Note that for this example the model assumed in section 3, with fixed numbers of rows and columns, is more plausible than the model of this section, since $n^* = 5$ is hardly large.

6. Generalised Median Tests

Conover (1980, section 4.3) described the median test, in which random samples are taken from each of c populations. Each random sample is classified as above and below the grand median (the median of the combined random samples), forming an r by 2 contingency table with fixed marginal totals. The usual chi-squared test, based on X_P^2, is then applied to this contingency table.

If instead of the grand median, a 'grand quantile' is used, the resulting test is described as a quantile test: see Conover (1980, p. 174). These tests can be generalised by choosing c instead of two categories for the combined random samples, and so forming an r by c contingency table of counts N_{ij} of the number of observations for the ith sample in the jth category. This table has all row and column totals fixed and can be tested for row consistency using the results of the sections 2 and 3. The first three say, components of X_P^2 are of particular interest, indicating location, dispersion and skewness differences between treatments.

It is routine to show that the location component $\boldsymbol{V}_1^T \boldsymbol{V}_1$ of X_P^2 reduces to the median test statistic when observations are classified into just two categories. This is shown in the Appendix. The result identifies the median test as a location detecting test. To detect up to sth moment differences between the populations requires categorisation into $s + 1$ categories and the use of the $\boldsymbol{V}_2, \ldots, \boldsymbol{V}_s$ components. If there are as many categories as observations and each category has one observation, the test based on the location component is the Kruskal-Wallis test, which is known to be more powerful than the median test. Using more than two categories will result in less loss of information due to categorisation compared to the median test, and will permit assessment of higher moment differences between the treatments.

Corn Example. Conover (1980, p. 172) gave the example of four different methods of growing corn. He classified the data as greater than 89 and up to 88 and applied

the median test. In this form this does not conform to the fixed margins model. If the objective were to divide the data into groups of the lowest 18 and highest 16 observations, it would conform to the fixed margins model. We now classify the data into four approximately equal groups.

Using the median test, Conover reported a P-value "slightly less than 0.001": the method median yields are clearly different. We calculate $X_P^2 = 49.712$ on 9 degrees of freedom. In addition $V_1^T V_1 = 25.723$, $V_2^T V_2 = 19.972$ and $V_3^T V_3 = 2.574$, all on 3 degrees of freedom. The location and dispersion components and X_P^2 are all significant, with P-values all zero to three decimal places. The residual or skewness component has χ_3^2 P-value 0.45. The finer classification, compare to that employed by the median test, has uncovered a variability difference between the methods: methods 3 and 4 are significantly less variable than 1 and 2.

	First Quartile	Second Quartile	Third Quartile	Fourth Quartile	Total
Method 1	0	3	4	2	9
Method 2	1	6	3	0	10
Method 3	0	0	1	6	7
Method 4	8	0	0	0	8
Total	9	9	8	8	34

Appendix
The Orthogonal Polynomials
The first two polynomials, defined on $x_1, \ldots, x_c$ and orthonormal with regard to the weights $p_1, \ldots, p_c$, are $g_1(x_j)$ and $g_2(x_j)$, given explicitly by:

$$g_1(x_j) = (x_j - \mu)/\sqrt{\mu_2}$$

and

$$g_2(x_j) = a\{(x_j - \mu)^2 - \mu_3(x_j - \mu)/\mu_2 - \mu_2\}, \quad j = 1, \ldots, c,$$

in which

$$\mu = \sum_{j=1}^c x_j p_j, \quad \mu_r = \sum_{j=1}^c (x_j - \mu)^r p_j \quad \text{and} \quad a = \left(\mu_4 + \mu_3^2/\mu_2 - \mu_2^2\right)^{-0.5}.$$

The subsequent polynomials $g_3(x_j), \ldots, g_{c-1}(x_j)$ may be derived by using the useful recurrence relations in Emerson (1968). In the text we have taken, as in many applications, $x_j = j$, $j = 1, \ldots, c$.
Derivation of the Covariance Matrix of the Cell Counts
In section 2 the method used to find the moments of the N_{ij} is described. To find $E[N_{21}]$, we take $E[N_{21}|N_{1j} + N_{2j}, \quad j = 1, \ldots, c]$, then the conditional expectation of this expression with the sum of the first three columns being known, and so on. The successive expectations are

$$n_{2.}(N_{11} + N_{21})/(n_{1.} + n_{2.}),$$
$$\{n_{2.}/(n_{1.} + n_{2.})\}\{(n_{1.} + n_{2.})(N_{11} + N_{21} + N_{31})/(n_{1.} + n_{2.} + n_{3.})\},$$
$$\vdots$$
$$\{n_{2.}/(n_{1.} + n_{2.})\}\{(n_{1.} + n_{2.})/(n_{1.} + n_{2.} + n_{3.})\}\dots$$
$$\{(n_{1.} + \dots + n_{(c-1).})/(n_{1.} + \dots + n_{c.})\}\{n_{.1}\} = n_{2.}n_{.1}/n_{..}.$$

By symmetry $E[N_{ij}] = n_{i.}n_{.j}/n_{..}$, $i = 2, \dots, r$ and $j = 1, \dots, c$. By difference the expectations for the first row may be obtained, giving the familiar

$$E[N_{ij}] = n_{i.}n_{.j}/n_{..}, \quad i = 1, \dots, r \quad \text{and} \quad j = 1, \dots, c.$$

In the same way

$$E[N_{21}(N_{21} - 1)] = n_{2.}(n_{2.} - 1)n_{.1}(n_{.1} - 1)/\{n_{..}(n_{..} - 1)\},$$

from which we obtain $var(N_{21})$, and

$$var(N_{ij}) = n_{i.}\frac{n_{.j}}{n_{..}}\left(1 - \frac{n_{.j}}{n_{..}}\right)\left(\frac{n_{..} - n_{i.}}{n_{..} - 1}\right), \quad i = 1, \dots, r \quad \text{and} \quad j = 1, \dots, c.$$

Similarly

$$cov(N_{ij}, N_{ik}) = -n_{i.}\frac{n_{.j}}{n_{..}}\frac{n_{.k}}{n_{..}}\left(\frac{n_{..} - n_{i.}}{n_{..} - 1}\right), \quad i = 1, \dots, r \quad \text{and} \quad j \neq k = 1, \dots, c.$$

By symmetry

$$cov(N_{rj}, N_{sj}) = -n_{.j}\frac{n_{r.}}{n_{..}}\frac{n_{s.}}{n_{..}}\left(\frac{n_{..} - n_{.j}}{n_{..} - 1}\right), \quad i = 1, \dots, r \quad \text{and} \quad j \neq k = 1, \dots, c,$$

and by the expectation argument again

$$cov(N_{ir}, N_{js}) = \frac{n_{i.}n_{.r}}{n_{..}}\frac{n_{j.}n_{.s}}{n_{..}}\left(\frac{1}{n_{..} - 1}\right), \quad i \neq j = 1, \dots, r, \quad \text{and} \quad r \neq s = 1, \dots, c.$$

Write $\boldsymbol{N}_i = (N_{i1}, \dots, N_{ic})^T$, $i = 1, \dots, r$ and $\boldsymbol{N}^T = (\boldsymbol{N}_1^T, \dots, \boldsymbol{N}_r^T)$. The joint covariance matrix of $\boldsymbol{N}_i$ and $\boldsymbol{N}_j$ is, for $i \neq j$,

$$cov(\boldsymbol{N}_i, \boldsymbol{N}_j) = -\frac{n_{i.}}{n_{..}}\frac{n_{j.}}{n_{..}}\left\{diag\left(\frac{n_{.r}n_{..}}{(n_{..} - 1)}\right) - \left(\frac{n_{.r}n_{.s}}{n_{..} - 1}\right)\right\}.$$

Now since the $\{N_{ij}\}$ are such that the row and column totals are known constants, $cov(\boldsymbol{N}_i, \boldsymbol{N}_1 + \dots + \boldsymbol{N}_r) = 0$ for $i = 1, \dots, r$. So if we write $f_j = n_{.j}/n_{..}$, $j = 1, \dots, c$, and

$$\boldsymbol{R} = diag\left(\frac{n_{.r}n_{..}}{(n_{..}-1)}\right) - \left(\frac{n_{.r}n_{.s}}{n_{..}-1}\right),$$

then the covariance matrix of $\boldsymbol{N}_i$ is

$$cov(\boldsymbol{N_i}) = -\sum_{i \neq j} cov(\boldsymbol{N}_i, \boldsymbol{N}_j) = \sum_{i \neq j} f_i f_j \boldsymbol{R} = f_i(1 - f_i)\boldsymbol{R},$$

which agrees with direct calculation. So if $\otimes$ is the Kronecker product, the covariance matrix of $\boldsymbol{N}$ is

$$cov(\boldsymbol{N}) = \{diag(f_j) - (f_i f_j)\} \otimes R.$$

Recall that we have defined $Z_{ij} = (N_{ij} - E[N_{ij}])/\sqrt{E[N_{ij}]}$, $i = 1, \ldots, r$ and $j = 1, \ldots, c$, and $\boldsymbol{Z} = (Z_{11}, \ldots, Z_{1c}, \ldots, Z_{tl}, \ldots, Z_{tc})^T$. It follows that

$$cov(\boldsymbol{Z}) = \{\boldsymbol{I}_r - (\sqrt{[f_i f_j]})\} \otimes \boldsymbol{R}.$$

The location component $\boldsymbol{V}_1^T \boldsymbol{V}_1$ of X_P^2 reduces to the median test statistic

If there are b observations below a predetermined point in the combined sample, and a above it, then Conover (1980, p. 172) gave the X^2 Median test statistic as

$$T = \frac{n_{..}^2}{ab} \sum_{i=1}^{r} \left(N_{i1} - \frac{n_{i.}b}{n_{..}}\right)^2 /n_{i.}.$$

It is routine to show that $g_{1j} = (j - \mu)/\sigma$, $j = 1, \ldots, c$, in which μ and σ are are the mean and standard deviation of the distribution defined by $P(X = 1) = b/n_{..}$ and $P(X = 2) = a/n_{..}$. It follows that $\mu = 1 + a/n_{..}$ and $\sigma^2 = ab/n_{..}^2$. Now

$$\begin{aligned} \sigma\sqrt{n_{i.}}V_{1i} &= \sum_{j=1}^{2} N_{ij}g_{1j} = N_{i1}(-a/n_{..}) + (n_{i.} - N_{i1})(1 - a/n_{..}) \\ &= (n_{i.} - an_{i.}/n_{..}) - N_{i1} = n_{i.}b/n_{..} - N_{i1}. \end{aligned}$$

It follows that, as required, $\boldsymbol{V}_1^T \boldsymbol{V}_1 = T$.

References

1. D.J. Best and J.C.W. Rayner. Nonparametric analysis for doubly ordered two-way contingency tables. *Biometrics*, 52:1153-1156, 1996.
2. W.J. Conover. *Practical Nonparametric Statistics* (2nd ed.). Wiley, New York, 1980.
3. P.L. Emerson. Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics*, 24:695-701, 1968.
4. H.O. Lancaster. *The Chi-squared Distribution*. Wiley, New York, 1969.
5. J.C.W. Rayner and D.J. Best. *Smooth Tests of Goodness of Fit*. Oxford University Press, New York, 1989.
6. J.C.W. Rayner and D.J. Best. Smooth extensions of Pearson's produce moment correlation and Spearman's rho. *Statistics and Probability Letters*, 30(2):171-177, 1996a.

7. J.C.W. Rayner and D.J. Best. Extensions to some important nonparametric tests. In *Proceedings of the A.C. Aitken Centenary Conference, Dunedin 1995*, (Kavalieris, L. et al. ed.), University of Otago Press, Dunedin, 257-266, 1996b.

8. S.N. Roy and S.K. Mitra. An introduction to some non-parametric generalisations of analysis of variance and multivariate analysis. *Biometrika*, 43:361-376, 1956.