

# An Overview of New Results in Cochran–Mantel–Haenszel Testing

J.C.W. RAYNER\*

*National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong, Australia and*

*School of Mathematical and Physical Sciences, University of Newcastle, Newcastle, Australia*

Email: [John.Rayner@newcastle.edu.au](mailto:John.Rayner@newcastle.edu.au)

and

Paul RIPPON

*School of Mathematical and Physical Sciences, University of Newcastle, Newcastle, Australia*

Email: [Paul.Rippon@newcastle.edu.au](mailto:Paul.Rippon@newcastle.edu.au)

---

**Abstract:** The Cochran–Mantel–Haenszel (CMH) methodology is a suite of tests applicable to particular tables of count data. Inference is *conditional* on the treatment and outcome totals on each stratum being known before sighting the data. The CMH tests are important for analyzing randomised blocks data when the responses are categorical rather than continuous.

This overview will first review the traditional CMH tests and then explore new alternative presentations of the *ordinal* CMH tests. Next the ordinal CMH tests will be extended so they can be used to test for higher moment effects. Finally unconditional analogues of the extended CMH tests will be developed.

**KEYWORDS:** Generalised correlations, orthonormal polynomials, randomised block designs, umbrella effects

**AMS SUBJECT CLASSIFICATION CODES:** 62G10; 62F03

---

## 1. Introduction

The Cochran–Mantel–Haenszel (CMH) methodology is a suite of tests applicable to particular tables of count data. The data are counts of the number of times the  $i$ th of  $t$  treatments results in an outcome in the  $h$ th of  $c$  categories on the  $j$ th of  $b$  blocks or strata. Inference is *conditional* on the treatment and outcome totals on each stratum being known before sighting the data. The CMH tests are applicable to more complex designs than randomised blocks, but the analysis of randomised blocks data when the responses are categorical rather than continuous is certainly an important application of the CMH tests.

\* Corresponding author

This overview will first review the traditional CMH tests and then explore new alternative presentations of the *ordinal* CMH tests. These could be used to calculate the test statistics but their main use will perhaps be to develop intuition about CMH testing. Next the ordinal CMH tests will be extended so they can be used to test for higher moment effects. One rationale for developing these extensions was to enable a comparison between the nonparametric ANOVA tests introduced in Rayner and Best (2013), Rayner et al. (2015) and Rayner (2017). These tests permit univariate moment assessments beyond the mean, and bivariate assessments beyond the (order (1, 1) ) correlation. The nonparametric ANOVA tests are therefore briefly reviewed before considering the CMH extensions. Finally unconditional analogues of the extended CMH tests will be developed.

## 2. The CMH tests

The Cochran–Mantel–Haenszel (CMH) tests are a class of nonparametric tests used to nonparametrically analyze tables  $\{N_{ihj}\}$  of count data of a particular structure. Specifically  $N_{ihj}$  counts the number of times treatment  $i$  is classified into outcome category  $h$  in the  $j$ th stratum,  $i = 1, \dots, t$ ,  $h = 1, \dots, c$  and  $j = 1, \dots, b$ . Strata are independent, and the treatments present in each stratum are fixed by design.

Dot notation is used to reflect summation over a subscript. In the traditional CMH tests (Landis et al., 1978, 1979, Davis, 2002)

- the strata totals  $\sum_{i,h} N_{ihj} = n_{\bullet\bullet j}$ ,
- the treatment totals within strata,  $\sum_h N_{ihj} = n_{i\bullet j}$ , and
- the outcome totals within strata,  $\sum_i N_{ihj} = n_{\bullet hj}$

are all assumed to be known prior to sighting the data: they are not random variables. Inference is *conditional*.

The traditional CMH tests assess

- overall partial association (OPA) ( $\chi^2_{b(c-1)(t-1)}$ )
- general association (GA) ( $\chi^2_{(c-1)(t-1)}$ )
- mean scores (MS) ( $\chi^2_{t-1}$ ) and
- correlation (C) ( $\chi^2_1$ ).

In parentheses are the asymptotic null distribution of the test statistics. The test statistics are quadratic forms with vectors using the table counts and some of the covariance matrices involving Kronecker products. They will be denoted by  $S_{OPA}$ ,  $S_{GA}$ ,  $S_{MS}$  and  $S_C$ .

This set-up is appropriate for categorical randomised block data. While the CMH methodology is appropriate more generally, it won't accommodate more complex designs such as Latin square, multifactor and many other designs. However it is an extremely important analysis tool for randomised blocks when the responses are categorical rather than continuous. In consumer studies with *just about right* (JAR) responses such as the *Jams Example* following, when  $c$  is small the randomised block F test will often be invalid in spite of the well-known ANOVA robustness.

Two examples will now be introduced. They will be reconsidered throughout this article.

*Homosexual Marriage Example.* Scores of 1, 2 and 3 are assigned to the responses agree, neutral and disagree respectively to the proposition “Homosexuals should be able to marry” and scores of 1, 2 and 3 are assigned for the religions fundamentalist, moderate and liberal respectively. Agresti (2002) finds  $S_{GA}$  takes the value 19.76 with  $\chi_4^2$  p-value 0.0006,  $S_{MS}$  takes the value 17.94 with  $\chi_2^2$  p-value 0.0001, and  $S_C$  takes the value 16.83 with  $\chi_1^2$  p-value less than 0.0001. From these tests we conclude there is strong evidence of an association between the proposition responses and religion. In particular there is evidence of mean differences in the responses and of a (linear -linear) correlation between responses and religion.

TABLE 1

*Opinions on homosexual marriage by religious beliefs and education levels for ages 18 to 25*

		Homosexuals should be able to marry		
Education	Religion	Agree	Neutral	Disagree
School	Fundamentalist	6	2	10
	Moderate	8	3	9
	Liberal	11	5	6
College	Fundamentalist	4	2	11
	Moderate	21	3	5
	Liberal	22	4	1

TABLE 2

*JAR codes for sweetness of three plum jams*

Judge	Jam		
	A	B	C
1	3	2	3
2	4	5	4
3	3	2	3
4	1	4	2
5	2	4	2
6	1	3	3
7	2	5	4
8	2	5	2

*Jams Example.* Three plum jams, A, B and C are given JAR sweetness codes by eight judges as in the Table 2. Here 1 denotes *not sweet enough*, 2 *not quite sweet enough*, 3 *just about right*, 4 *a little too sweet* and 5 *too sweet*. The treatment sums for jams A, B and C are 18, 30 and 23 respectively. We find  $S_{MS} = 9.6177$  with  $\chi_2^2$  p-value 0.0082 and  $S_C = 1.1029$  with  $\chi_1^2$  p-value 0.2936. There is evidence of a mean effect: on average the jams are different. However there is

no evidence of a correlation effect: as we pass from jam A to B and then to C there is no evidence of an increasing (or decreasing) response.

### 3. The Nominal CMH tests

The CMH MS test assumes that response categories are ordered while the CMH C test assumes that both treatment categories and response categories are ordered. It is therefore appropriate to label these as ordinal CMH tests, while the OPA and GA CMH tests, that make no assumption about ordering, can be called nominal CMH tests. Scores for the ordered categories are needed to apply the ordinal tests, while no scores are required for the nominal tests.

It is important to note that the traditional CMH tests are conditional tests, conditional on the treatment totals within strata,  $\sum_h N_{ihj} = n_{i\bullet j}$ , and the outcome totals within strata,  $\sum_i N_{ihj} = n_{\bullet hj}$  being known prior to collecting or sighting the data. For randomised block data in which the responses are untied ranks the marginal totals are known before collecting the data (in every block each treatment is applied once and each response is a rank that is assigned once), while that is not the case for data such as in the *Homosexual Marriage Example*.

For conditional testing the distribution theory involved in determining means and covariance matrices uses a product extended hypergeometric distribution. For each stratum since the row and column totals are known, the counts  $N_{ihj}$  follow an extended hypergeometric distribution. Moreover since the strata are mutually independent, a product distribution is appropriate for the collection of these strata counts.

To define the CMH GA test statistic  $S_{GA}$ , first define the vector of counts on the  $j$ th stratum  $U_j = (N_{11j}, \dots, N_{1cj}, \dots, N_{t1j}, \dots, N_{tcj})^T$ . Summing over strata gives

$$U_{\bullet} = (N_{11\bullet}, \dots, N_{1c\bullet}, \dots, N_{t1\bullet}, \dots, N_{tc\bullet})^T.$$

Using the extended hypergeometric distribution, it may be shown, as in Landis et al. (1978), that  $U_{\bullet}$  has mean

$$E[U_{\bullet}] = \left( \sum_{j=1}^b \frac{n_{i\bullet j} n_{\bullet hj}}{n_{\bullet\bullet j}} \right)$$

and covariance matrix

$$\text{cov}(U_{\bullet}) = \sum_{j=1}^b \text{cov}(U_j) = \sum_{j=1}^b \frac{n_{\bullet\bullet j}^2}{(n_{\bullet\bullet j} - 1)} V_{Tj} \otimes V_{Cj}$$

in which, after writing  $p_{i\bullet j} = n_{i\bullet j} / n_{\bullet\bullet j}$  and  $p_{\bullet hj} = n_{\bullet hj} / n_{\bullet\bullet j}$ ,

$$V_{Tj} = \{\text{diag}(p_{i\bullet j}) - (p_{i\bullet j})(p_{i\bullet j})^T\} \text{ and } V_{Cj} = \{\text{diag}(p_{\bullet hj}) - (p_{\bullet hj})(p_{\bullet hj})^T\}.$$

Here  $\otimes$  is the Kronecker or direct product.

The CMH GA test statistic is a quadratic form using  $U_{\bullet} - E[U_{\bullet}]$  and an inverse of  $\text{cov}(U_{\bullet})$ , namely  $S_{GA} = (U_{\bullet} - E[U_{\bullet}])^T \text{cov}^{-1}(U_{\bullet})(U_{\bullet} - E[U_{\bullet}])$ . As  $\text{cov}(U_{\bullet})$  is not of full rank either a generalised inverse  $\text{cov}^{-1}(U_{\bullet})$  can be used, or dependent variables can be chosen and omitted to produce a covariance matrix of full rank. All unknown parameters are estimated using maximum likelihood under the null hypothesis. Asymptotically, as  $n_{\bullet\bullet\bullet}$  becomes large,  $S_{GA}$  has the  $\chi_{(c-1)(t-1)}^2$  distribution. The test statistic is symmetric in the treatments and outcome categories, and independent of the choice of the dependent variables.

The test statistic  $S_{GA}$  is too complicated for routine hand calculation; it is almost always applied using software in packages such as R.

To calculate  $S_{OPAS}$  the vector of the quadratic form involves the aggregation of the  $U_j$  via  $U = (U_1^T, \dots, U_b^T)^T$ . Again the covariance matrix is calculated using the product extended hypergeometric distribution. The overall partial association statistic so derived is given by

$$S_{OPAS} = \sum_{j=1}^b \frac{n_{\bullet\bullet j} - 1}{n_{\bullet\bullet j}} \sum_{i=1}^t \sum_{h=1}^c \frac{(N_{ihj} - n_{i\bullet j} n_{\bullet hj} / n_{\bullet\bullet j})^2}{n_{i\bullet j} n_{\bullet hj} / n_{\bullet\bullet j}}.$$

Asymptotically this has the  $\chi_{b(c-1)(t-1)}^2$  distribution.

The difference between the two tests is that the CMH GA test is seeking to detect *average* partial association while the CMH OPA test is seeking to detect *overall* partial association. The former is more *focused*, with  $(c-1)(t-1)$  degrees of freedom versus  $b$  times that for CMH OPA statistic. The degrees of freedom are the dimension of the alternative hypothesis space. The CMH OPA test is seeking to detect very general alternatives to the null hypothesis and will have relatively low power for these alternatives. The CMH GA test seeks to detect fewer alternatives and will have more power than the CMH OPA test for these alternatives. However, CMH OPA test will have some power for alternatives to which the CMH GA is insensitive. In other contexts more focused tests have been constructed using components of an omnibus test statistic. This idea will be taken up subsequently.

An alternative test for overall partial association is the Pearson test, with statistic

$$T_{OPA} = \sum_{j=1}^b \sum_{i=1}^t \sum_{h=1}^c \frac{(N_{ihj} - n_{i\bullet j} n_{\bullet hj} / n_{\bullet\bullet j})^2}{n_{i\bullet j} n_{\bullet hj} / n_{\bullet\bullet j}}.$$

This is an unconditional test that does not assume all treatment and outcome categories are known before sighting the data. The difference between  $S_{OPA}$  and  $T_{OPA}$  is merely the factor

$(n_{\bullet\bullet j} - 1) / n_{\bullet\bullet j}$  applied to each stratum. For large stratum counts this will make little difference in the values of  $S_{OPA}$  and  $T_{OPA}$ .

Likewise general association can be tested for using a Pearson test for the two-way table of counts  $\{N_{ih\bullet}\}$ ,  $T_{GA}$  say. The Pearson test statistics  $T_{OPA}$  and  $T_{GA}$  will have the same asymptotic distributions as the corresponding conditional tests. Most users will have more familiarity with the unconditional tests and most packages will have routines for their calculation even if they don't have routines for CMH tests.

Subsequently the main focus will be the CMH ordinal tests.

## 4. The Ordinal CMH tests

### 4.1. CMH mean scores test

Suppose, as before, that  $N_{ihj}$  counts the number of times treatment  $i$  is classified into outcome category  $h$  in the  $j$ th stratum,  $i = 1, \dots, t$ ,  $h = 1, \dots, c$  and  $j = 1, \dots, b$ . Assume that outcomes are ordinal and assign the *score*  $b_{hj}$  to the  $h$ th response on the  $j$ th stratum. All marginal totals are assumed to be known, so the product extended hypergeometric model is assumed.

The score sum for treatment  $i$  in stratum  $j$  is  $M_{ij} = \sum_{h=1}^c b_{hj} N_{ihj}$ . Taking expectations  $E[M_{ij}] = n_{i\bullet j} \sum_{h=1}^c b_{hj} n_{\bullet hj} / n_{\bullet\bullet j}$  since  $E[N_{ihj}] = n_{i\bullet j} n_{\bullet hj} / n_{\bullet\bullet j}$ . If  $M_j = (M_{ij})$  then inference is based on  $M = \sum_{j=1}^b M_j$  through the quadratic form

$$S_M = (M - E[M])^T \text{cov}^{-1}(M)(M - E[M])$$

in which all unknown parameters are estimated by maximum likelihood under the null hypothesis. Now  $E[M] = \sum_{j=1}^b E[M_j]$  and as strata are independent  $\text{cov}(M) = \sum_{j=1}^b \text{cov}(M_j)$ . To give  $\text{cov}(M_j)$  first define

$$S_j^2 = \frac{n_{\bullet\bullet j} \sum_{h=1}^c b_{hj}^2 n_{\bullet hj} - \left(\sum_{h=1}^c b_{hj} n_{\bullet hj}\right)^2}{(n_{\bullet\bullet j} - 1)}.$$

Then it may be shown that  $\text{cov}(M_j) = S_j^2 V_{T_j}$ . Since  $\text{cov}(M_j)$  is not of full rank the usual approaches, such as dropping appropriate treatment and/or outcome categories, or using a generalised inverse, can be used. Software exists to calculate  $S_M$ .

*Note.* The mean scores statistic depends on the scores assigned to the response categories. Thus the statistic could be written  $S_M(\{b_{hj}\})$  to emphasise this dependence.

*Aside.* The derivation of  $\text{cov}(M_j)$  requires routine but tedious algebra. If  $\delta_{uv}$  is the Kronecker delta, = 1 if  $u = v$  and zero otherwise, using standard distribution theory for the product extended hypergeometric distribution,  $E[N_{ihj}] = n_{i..}n_{.hj} / n_{..j}$  and the covariance between  $N_{ihj}$  and  $N_{i'h'j}$  is

$$n_{i..}n_{.hj}(\delta_{ii'}n_{..j} - n_{i'.j})(\delta_{hh'}n_{..j} - n_{.h'j}) / \{n_{..j}^2(n_{..j} - 1)\}.$$

It follows that

$$\text{var}(\sum_h b_{hj}N_{ihj}) = n_{i..}(n_{..j} - n_{i'.j})S_j^2 / \{n_{..j}^2(n_{..j} - 1)\} \text{ and}$$

$$\text{cov}(\sum_h b_{hj}N_{ihj}, \sum_{h'} b_{h'j}N_{i'h'j}) = -n_{i..}n_{i'.j}S_j^2 / \{n_{..j}^2(n_{..j} - 1)\}.$$

These lead to the stated covariance matrix.

Under the null hypothesis of no treatment effects the distribution of  $S_M$  can be shown to be asymptotically  $\chi_{t-1}^2$ ; see Landis et al. (1978).

#### 4.2. The CMH correlation test

The CMH correlation tests assume that the treatment and response variables are both measured on either the ordinal or the interval scale, and that for the  $i$ th treatment the scores are  $a_{hi}$ ,  $i = 1, \dots, t$ , and on the  $j$ th stratum the response scores are  $b_{hj}$ ,  $j = 1, \dots, b$ .

The null hypothesis of no association between the treatment and response variables having adjusted for the  $b$  strata, is tested against the alternative that across strata there is a consistent association, positive or negative, between the treatment scores and response scores.

Take

- $C_j = \sum_i \sum_h a_{hi}b_{hj} \{N_{ihj} - E[N_{ihj}]\}$  and
- $C = \sum_j C_j$ .

The CMH correlation (CMH  $C$ ) statistic is  $C^2/\text{var}(C) = S_C$  say. The derivation of  $\text{var}(C)$  is relatively complex if scalars are used, but is routine using Kronecker products.

To derive  $\text{var}(C)$  first define  $a_j = (a_{1j}, \dots, a_{tj})^T$ ,  $b_j = (b_{1j}, \dots, b_{cj})^T$  and  $N_j = (N_{11j}, \dots, N_{1cj}, \dots, N_{t1j}, \dots, N_{tcj})^T$ . Then  $C_j = (a_j \otimes b_j)^T (N_j - E[N_j])$  and

$$\begin{aligned} \text{var}(C_j) &= E[(a_j \otimes b_j)^T (N_j - E[N_j]) (N_j - E[N_j])^T (a_j \otimes b_j)] = \\ &= (a_j \otimes b_j)^T E[(N_j - E[N_j]) (N_j - E[N_j])^T] (a_j \otimes b_j) = \\ &= (a_j \otimes b_j)^T \text{cov}(N_j) (a_j \otimes b_j). \end{aligned}$$

Recall from section 2 that with  $p_{i\cdot j} = n_{i\cdot j} / n_{\dots}$  and  $p_{\cdot hj} = n_{\cdot hj} / n_{\dots}$ , we have  $V_{Tj} = \text{diag}(p_{i\cdot j}) - (p_{i\cdot j})(p_{i\cdot j})^T$  and  $V_{Cj} = \text{diag}(p_{\cdot hj}) - (p_{\cdot hj})(p_{\cdot hj})^T$ . Now from Landis et al. (1978)

$$\text{cov}(N_j) = \frac{n_{\dots j}^2}{n_{\dots j} - 1} (V_{Tj} \otimes V_{Cj}).$$

Hence

$$\text{var}(C_j) = \frac{n_{\dots j}^2}{n_{\dots j} - 1} (a_j^T V_{Tj} a_j \otimes b_j^T V_{Cj} b_j) = \frac{n_{\dots j}^2}{n_{\dots j} - 1} (a_j^T V_{Tj} a_j)(b_j^T V_{Cj} b_j)$$

because both factors in the Kronecker product are scalars. Finally  $\text{var}(C) = \sum_j \text{cov}(C_j)$  because counts in different strata are mutually independent. The CMH correlation statistic  $S_C$  is now fully specified.

The Central Limit Theorem assures the asymptotic normality of  $C$ , so as the total sample size  $n_{\dots} = n_{\dots 1} + \dots + n_{\dots s}$  approaches infinity  $S_C$  has asymptotic distribution  $\chi_1^2$ . Again see Landis et al. (1978).

## 5. Alternative presentations of the ordinal CMH test statistics

In this section the focus is on alternative expressions for the ordinal CMH test statistics. In the case of randomised block data simple expressions are given for the ready calculation of the test statistics. The expression for the correlation statistic is more general and quite insightful, so that will be considered first.

### 5.1 The CMH correlation statistic

Using the definitions previously given for  $V_{Tj}$  and  $V_{Cj}$  we have

$$\begin{aligned} n_{\dots j} a_j^T V_{Tj} a_j &= \sum_i a_{ij}^2 n_{i\cdot j} - (\sum_i a_{ij} n_{i\cdot j})^2 / n_{\dots j} \text{ and} \\ n_{\dots j} b_j^T V_{Cj} b_j &= \sum_h b_{hj}^2 n_{\cdot hj} - (\sum_i b_{ij} n_{\cdot hj})^2 / n_{\dots j}. \end{aligned}$$

On stratum  $j$  now define

$$\begin{aligned} S_{XXj} &= \sum_i a_{ij}^2 n_{i\cdot j} - (\sum_i a_{ij} n_{i\cdot j})^2 / n_{\dots j}, \\ S_{XYj} &= \sum_i \sum_h a_{ij} b_{hj} N_{ihj} - (\sum_i a_{ij} n_{i\cdot j})(\sum_h b_{hj} n_{\cdot hj}) / n_{\dots j} \text{ and} \\ S_{YYj} &= \sum_h b_{hj}^2 n_{\cdot hj} - (\sum_h b_{hj} n_{\cdot hj})^2 / n_{\dots j}. \end{aligned}$$



With appropriate divisors the  $S_{XXj}$  etc. give unbiased estimators of the stratum variances and covariances. With these definitions

$$C_j = S_{XYj}, \quad n_{..j} a_j^T V_{Tj} a_j = S_{XXj} \text{ and } n_{..j} b_j^T V_{Cj} b_j = S_{YYj}.$$

Finally since  $C = \sum_j C_j$  and  $\text{var}(C) = \sum_j \text{cov}(C_j)$

$$S_C = C^2 / \text{var}(C) = \frac{\{\sum_j S_{XYj}\}^2}{\sum_j \{S_{XXj}^2 S_{YYj}^2 / (n_{..j} - 1)\}}.$$

The expression here uses the  $S_{XYj}$  etc., familiar in formulae for regression coefficients.

Three special cases will be considered:

- (1) the data consists of one stratum only;
- (2) the treatment scores are independent of the strata:  $a_{ij} = a_i$  for all  $i$  and  $j$ ; and
- (3) the randomised block design (in section 5.2).

In the second case  $S_{XX}$  is constant over strata. This gives a slight simplification of the  $S_C$  formula. See the *Jams Example* below. If the data come from a randomised block design a considerable simplification is possible if the same treatment and response scores are used on each stratum or block. See section 5.2.

In the first case the CMH correlation statistic simplifies to

$$S_C = \frac{(n_1 - 1) S_{XY1}^2}{S_{XX1}^2 S_{YY1}^2} = (n_1 - 1) r_p^2,$$

in which  $r_p$  is the Pearson correlation coefficient. This is well known. See, for example, Davis (2002, p.253).

If we now write  $r_{pj}$  for the Pearson correlation in the  $j$ th stratum, it follows that since  $S_{XYj} = r_{pj} S_{XXj} S_{YYj}$

$$S_C = \frac{\{\sum_j r_{pj} S_{XXj} S_{YYj}\}^2}{\sum_j \{S_{XXj}^2 S_{YYj}^2 / (n_{..j} - 1)\}}.$$

$S_C$  is proportional to the square of a linear combination of the Pearson correlations in each stratum. The proportionality factor ensures  $S_C$  has the  $\chi_1^2$  distribution. This formula demonstrates how the Pearson correlations in each stratum contribute to the overall correlation measure.

*Homosexual Marriage Example.* These data were considered in section 1; they are given in Table 1. Noting that stratum 1 is school and stratum 2 is college, we find  $S_{XX1} = 50$ ,  $S_{XY1} = -9$  and  $S_{YY1} = 39.7333$ ,  $r_{p1} = -0.2019$  and the CMH C statistic for school takes the value 2.4055

with  $\chi^2_1$  p-value 0.1209. Similarly  $S_{XX2} = 51.6712$ ,  $S_{XY2} = -23.8904$ ,  $S_{YY2} = 42.6301$ ,  $r_{P2} = -0.5090$  and the CMH C statistic for college takes the value 18.6558 with  $\chi^2_1$  p-value 0.0000. From these the value of the CMH C statistic and its  $\chi^2$  p-value are confirmed: it was previously noted that  $S_C$  takes the value 16.8328 with  $\chi^2_1$  p-value 0.0000. Clearly there is an insignificant Pearson correlation for schools and a highly significant Pearson correlation for college. The latter dominates the former, so that overall there is strong evidence of a correlation effect: as religion becomes increasingly liberal there is greater agreement with the proposition that homosexuals should be able to marry. This is due mainly to the stratum college.

*Whiskey Example.* O'Mahony (1986, p.363) gave the data in Table 3, which were analysed in Rayner and Best (2001). They use mid-rank scores and find the Spearman correlation, which takes the value 0.73. In testing if this is zero against two-sided alternatives they give a Monte Carlo p-value of 0.09 and an asymptotic p-value of 0.04.

TABLE 3  
Cross classification of age and whiskey

Years of maturing	Grade			Total
	First	Second	Third	
One	0	0	2	2
Five	1	1	1	3
Seven	2	1	0	3
Total	3	2	3	8

Using scores 1, 2 and 3 for grade and 1, 5 and 7 for years of maturity, we find  $S_{XX} = 6$ ,  $S_{XY} = -12$  and  $S_{YY} = 43.5$ . It follows that the Pearson correlation is  $-0.7428$  and the CMH C statistic takes the value 3.8621 with  $\chi^2_1$  p-value 0.0494. There is some evidence that as maturity increases so does the grade of the whiskey.

*Jams Example.* Here as there are eight strata hand calculation is possible if a little tedious. Because  $S_{XX}$  is constant over strata it is not too much extra work to calculate  $S_{XY}$ ,  $S_{YY}$ , the Pearson correlation, the CMH correlation statistic and its p-value on each stratum.

TABLE 4  
*Analysis of Jams data*

	Stratum							
	1	2	3	4	5	6	7	8
$S_{XY}$	0	0	0	1	0	2	2	0
$S_{YY}$	0.6667	0.6667	0.6667	4.6667	2.6667	2.6667	4.6667	6
$r_p$	0	0	0	0.3273	0	0.8660	0.6547	0
$S_C$	0	0	0	0.2143	0	1.5	0.8571	0
p-value	1	1	1	0.6434	1	0.2207	0.3545	1

We find  $S_{XX} = 2$  on all strata and  $S_C = 1.1029$  with  $\chi_1^2$  p-value 0.2936. There is no significant correlation effect, which would, if present, indicate that as we pass from jam A to B to C there is increasing (or decreasing) sweetness. It could be that overall there is no correlation effect with the contrary being the case in a minority of strata. That is not the case here; no stratum shows any evidence of a slight correlation effect. Here this is hardly surprising; with only three observations in each stratum there can be little power in testing for a correlation.

If there is interest in the contributions to the correlation from individual strata this is a reasonable approach. However, if there is not, then for the randomised blocks design with the same treatment scores in each stratum a considerable simplification is possible. This is now considered.

## 5.2. The randomised blocks design

As developed the CMH methodology doesn't apply to Latin square, multifactor and many other designs. However it is an extremely important analysis tool for randomised blocks when the responses are categorical rather than continuous. See, for example, Davis (2002, Chapter 8) who uses CMH methods to analyse repeated measures categorical data.

In the case of the randomised block design a considerable simplification of the CMH MS statistic is possible. In this case  $n_{i\bullet j} = 1$  for all  $i$  and  $j$ . Consequently  $n_{\bullet\bullet j} = t$ ,  $n_{i\bullet\bullet} = b$ ,  $n_{\bullet\bullet\bullet} = bt$  and  $p_{i\bullet j} = n_{i\bullet\bullet} / n_{\bullet\bullet\bullet} = 1/t$ . Substituting in the definitions given in Section 4 ultimately gives

$$S_M = \frac{b(t-1)F}{(b-1+F)}$$

in which  $F$  is the ANOVA  $F$  test statistic. A derivation of this relationship is given in the supplementary material. The significance of this result is that the CMH MS and ANOVA  $F$  tests are formally equivalent.

Coincidentally the given relationship is the same as that relating the Friedman test statistic and the  $F$  test statistic using the ranks as data.

P-values may be obtained by referring  $S_M$  to its  $\chi_{t-1}^2$  distribution, or  $F$  to its  $F_{(t-1), (b-1)(t-1)}$  distribution, or otherwise. An empirical study would be required to assess which is the more reliable in the sense of closeness to, say, permutation test p-values. In the examples we have analysed there has been little difference in these methods.

For the CMH C test statistic with the same scores in every stratum

$$S_C = \frac{(t-1)C^2}{S_{XX} \sum_j S_{YYj}}$$

in which  $C$  is the sum of the products of the treatment sums and the centred treatment scores,  $S_{XX}$  is the sum of the squares of the centred treatment scores and  $\sum_j S_{YYj}$  may be read from the output for a one-way ANOVA. Again the derivation is given in the supplementary material.

*Jams Example.* For the jams data  $a_i = 1, 2$  and  $3$ , so  $S_{XX}^2 = 2$ . From a one-way ANOVA with judges as treatments,  $\sum_j S_{Yj}^2 = 22 \frac{2}{3}$ . This can be confirmed by summing across the  $S_{YY}$  row in Table 4. By summing the columns in Table 2 the treatment score sums are found to be 18, 30 and 23, and since the centred treatment scores are  $-1, 0$  and  $1$ ,  $C = -18 + 0 + 23 = 5$ . Substituting gives  $S_C = 2 \cdot 25 / (2 \cdot 68 / 3) = 75 / 68 = 1.1029$  with  $\chi_1^2$  p-value 0.2936, as previously.

## 6. Nonparametric ANOVA

A suite of tests that can analyse the same data as the CMH tests is the nonparametric ANOVA tests of Rayner and Best (2013), Rayner et al. (2015) and Rayner (2017). However, these tests are applicable for a wider range of designs and are more flexible than the CMH tests, in that they assess both moment effects and generalised correlation effects of all orders, although only lower order tests are applied in practice. The extension of the CMH tests considered in this section makes the CMH tests equally as flexible as the nonparametric ANOVA tests.

The nonparametric ANOVA tests are nonparametric in that they are based on weak multinomial models incorporating smooth alternatives. The sampling distributions of the test statistics are not known, but the usual F distributions provide approximations that agree very well with permutation test results, and, of course, are very convenient. Responses for both the CMH and nonparametric ANOVA are categorical, although use of the term ANOVA suggests continuous responses. In fact, the issue of whether the data are categorical or continuous is of no practical importance. The nonparametric ANOVA tests are robust to even highly categorical data, and while the sampling distributions for the CMH tests are asymptotic, they are quite good even for small sample sizes. A simulation study in Rayner & Best (2018) showed that for randomised blocks data both suites of tests showed good agreement between their nominal significance levels and the test sizes achieved.

It is interesting to note that if the data are ranks then the first order nonparametric ANOVA F tests are equivalent to the Kruskal-Wallis when the design is the completely randomised design and the Friedman for the randomised blocks design. See, for example, Rayner (2016, section 2.5).

Nonparametric ANOVA may be applied to designs consistent with the general linear model while of course, the design for the CMH extensions is much more restrictive. There are two types of nonparametric ANOVA: ordered and unordered. For the unordered analysis orthonormal polynomials on the response variable up to a given level, typically three, are constructed. For any given ANOVA the analysis with the responses transformed by the order  $r$  orthonormal polynomial is called the order  $r$  analysis. Analyses of different orders are uncorrelated, so significance or not at any given order doesn't affect the significance or not at any other order.

For a given design there will usually be several effects, such as main effects, interactions and so on. The order  $r$  analysis tests null hypotheses that the responses transformed by the  $r$ th

orthonormal polynomial are consistent across levels of these effects. Suppose, for example, that one effect is the application of treatment A. Analyses of successive orders seek to assess whether or not the responses transformed by the first, second etc. order orthonormal polynomials are consistent across levels of treatment A.

The ordered analysis assumes that at least one of the independent variables is ordered in the sense that the levels of the effects corresponding to that variable are ordered. Orthonormal polynomials are constructed on the response variable and on each ordered independent variable. The responses are transformed by an orthonormal polynomial of particular order, and each independent variable is also transformed by an orthonormal polynomial of particular order. Then the product of the transformed variables is taken. A new, reduced design is formed from the original, without the independent variables that are ordered and with a new response, the product of orthonormal polynomials. ANOVAs of interest on the reduced design are then performed. These analyses assess generalised correlations, for which see Rayner and Beh (2009). As with the unordered analysis the conclusions from one ANOVA do not affect the conclusions from other ANOVAs. Of most interest are the usual order (1, 1) correlation that assesses linear-linear effects, and order (1, 2) and (2, 1) correlations that assess umbrella effects.

Both the extended CMH tests and the unordered and ordered nonparametric ANOVAs assume the existence of the orthonormal polynomials. An orthonormal polynomial of order  $r$  requires the existence of moments up to order  $2r$  while the corresponding orthogonal polynomial requires the existence of moments up to order  $r + 1$ . However, as responses are classified into, say,  $c$  classes, it is only possible to calculate moments of the responses up to order  $c - 1$ . Thus moments, orthonormal polynomials and hence tests of certain orders are not available for some parameter choices.

## 7. Extensions of the CMH mean scores and correlation tests

To construct extensions to  $S_M$  consider the scores  $b_{hj}$  on the  $j$ th stratum. A common choice of scores to give a ‘mean’ assessment would be the natural scores  $1, 2, \dots, c$  on all strata. However a ‘dispersion’ assessment could be given by choosing scores  $1^2, 2^2, \dots, c^2$  and similarly higher order powers might be of interest:  $b_{hj} = h^r$ . One problem with using the scores  $b_{hj} = h^r$  is that the test statistics are correlated. Thus the significance or not of the test for any order may affect the significance or not of tests at other orders. We now look at using more general scores with the objective of having uncorrelated test statistics. To this end now denote order  $r$  scores that are *not* stratum specific by  $b_h^r, h = 1, \dots, c$ . Define the order  $r$  score sum for treatment  $i$  by  $\sum_{j=1}^b \sum_{h=1}^c b_h^r N_{ihj} = \sum_{h=1}^c b_h^r N_{ih\bullet} = M_i^r$ . Now suppose that  $\{b_h^r\}$  are orthonormal using the weight function  $\{n_{\bullet h\bullet} / n_{\bullet\bullet\bullet}\}$ . Then for  $r \neq s = 1, 2, \dots, t$  it can be shown that  $\text{cov}(M_i^r, M_i^s) = 0$ : the  $i$ th score sums of different orders  $M_i^r$  are uncorrelated. In this sense the information provided by the scores sums of different orders for the same treatment is not related.

To construct extensions to  $S_C$  suppose that instead of a single set of outcome scores  $\{b_{hj}\}$ , we consider  $c$  sets of scores,  $\{b_{hj}^{(s)}\}$  for  $s = 0, 1, \dots, c - 1$ . Moreover suppose the scores

are orthonormal in the sense that  $\sum_h b_{hj}^{(r)} b_{hj}^{(s)} n_{\cdot hj} / n_{\cdot j} = \delta_{rs}$  with  $r, s = 0, 1, \dots, c - 1$  and with  $b_{hj}^{(0)} = 1$  for  $h = 1, \dots, c$ . Similarly instead of a single set of treatment scores  $\{a_{ij}\}$ , consider  $t$  sets of scores,  $\{a_{ij}^{(r)}\}$  for  $r = 0, 1, \dots, t - 1$  and suppose these scores are orthonormal in the sense that  $\sum_i a_{ij}^{(r)} a_{ij}^{(s)} n_{i \cdot j} / n_{\cdot j} = \delta_{rs}$  with  $r, s = 0, 1, \dots, t - 1$  and with  $a_{ij}^{(0)} = 1$  for  $i = 1, \dots, t$ . Both sets of scores may be stratum specific. Define  $S_{Cr_s}$  as before, but using  $a_{ij}^{(r)}$  and  $b_{hj}^{(s)}$ . It can be shown that the  $S_{Cr_s}$  are uncorrelated. See Rayner and Best (2018) for details.

*Jams Example.* For the jams data the CMH MS and unordered nonparametric ANOVA p-values are exactly the same. This is because for a randomised block design the CMH MS tests using the ANOVA F p-value use the same orthonormal functions, based on the category weights, as the unordered nonparametric ANOVA.

For the jams data the p-values using the F distribution are 0.0278, 0.2435 and 0.5554, and those using the  $\chi_{t-1}^2$  p-distribution are 0.0082, 0.1116 and 0.3802 respectively. At the 0.05 level there is evidence of mean differences in the scores for jams, but not of higher moment effects.

We may test for generalised correlation effects using both CMH GC and nonparametric ANOVA. In the Table 5 the nonparametric ANOVA cells give p-values, first using the t-test and second using the Wilcoxon signed ranks test.

Only the (1, 2) p-value using the t-test is less than 0.05. The treatment sums for jams A, B and C are 18, 30 and 23 respectively. It seems that in passing from jam A to B and then C there is evidence that the sweetness is assessed to increase and then decrease.

Here and in other examples we find that the CMH extensions and the corresponding nonparametric ANOVA tests give fairly similar conclusions.

TABLE 5

*CMH and nonparametric ANOVA GC p-values when testing for zero generalised correlations for the jam data*

	<i>CMH GC tests</i>		<i>Nonparametric ANOVA</i>	
	Treatment Order		Treatment Order	
Category Order	1	2	1	2
1	0.2936	0.0212	0.2249 0.2403	0.0333 0.0853
2	0.1862	0.2780	0.1410 0.2058	0.2640 0.1558
3	0.6098	0.3104	0.6041 0.7229	0.3069 0.2601

*Homosexual Marriage Example.* For the homosexual marriage data, the CMH MS extensions of first and second order yield p-values of 0.0000147 and 0.2732 respectively.

The CMH C extensions of orders (1, 1), (1, 2), (2, 1) and (2, 2) were calculated. An intermediate step gave the generalised correlations of these orders for each stratum. Again, the (1, 1) correlation gave very strong evidence of being non-zero. The others were not significant at the commonly used levels of significance. See Table 6.

TABLE 6

*CMH p-values for each stratum and overall when testing for zero generalised correlations for the jam data*

	CMH C extensions			
	(1, 1)	(1, 2)	(2, 1)	(2, 2)
School	0.1209	0.8352	0.3246	0.8523
College	0.0866	0.8183	0.2766	0.8371
Overall	0.0000	0.2328	0.1272	0.8777

The unordered nonparametric ANOVA gave p-values of less than 0.0001 for first order and 0.3719 for second order. As with the CMH MS extensions there is strong evidence of a mean difference between religions but no evidence of a second order effect.

The ordered nonparametric ANOVA may be used to test for non-zero generalised correlations. For orders (1, 1), (1, 2), (2, 1) and (2, 2) the Wilcoxon signed rank test two-tailed p-values were 0.0003, 0.0634, 0.0710 and 0.2540. In all cases the t-test was invalid due to non-normality. Nonparametric ANOVA confirms that there is evidence that the order (1, 1) generalised correlation is non-zero, while possibly reserving judgement on the generalised correlations of order (1, 2) and (2, 1).

## 8. Development of unconditional CMH tests

It was noted in section 3 that the nominal CMH test statistics  $S_{OPAS}$  and  $S_{GA}$  had analogues denoted by  $T_{OPAS}$  and  $T_{GA}$ . These were based on Pearson tests. The nominal CMH tests were identified as conditional tests while the analogues, since they are not based on the product extended hypergeometric distribution, are unconditional tests. Since the unconditional Pearson tests are very familiar, they would be the tests of preference for most users. It is of interest to develop unconditional analogues of the ordinal CMH tests. This was done in Rayner and Best (2017a).

For singly ordered two-way tables an unconditional model is explored in Rayner and Best (2001, Chapter 4). On the  $j$ th stratum define orthonormal polynomials  $\{\omega_{ij}(h)\}$  on  $(\hat{p}_{\bullet 1j}, \dots, \hat{p}_{\bullet cj})$ :

$$\sum_{h=1}^c \omega_{ij}(h) \omega_{vj}(h) \hat{p}_{\bullet hj} = \delta_{uv}.$$

Next define

$$V_{uij} = \sum_{h=1}^c N_{ihj} \omega_{uj}(h) / \sqrt{n_{i\bullet j}};$$

$V_{uij}$  reflects the contribution of treatment  $i$  to the order  $u$  effect on stratum  $j$ . Rayner and Best (2001, Chapter 4) show that

$$\sum_{h=1}^{c-1} \sum_{i=1}^t V_{uij}^2 = X_{pj}^2,$$

the Pearson statistic on the  $j$ th stratum. The statistic  $V_{u1j}^2 + \dots + V_{utj}^2$  gives an order  $u$  assessment of the consistency of treatments in the  $j$ th stratum. It has asymptotic distribution  $\chi_{t-1}^2$ . Summing over strata  $\sum_{i,j} V_{uij}^2$  gives an order  $u$  assessment of the consistency of treatments over all  $b$  strata and has asymptotic distribution  $\chi_{b(t-1)}^2$ .

In particular  $\sum_{i,j} V_{1ij}^2$  gives an overall mean assessment of the consistency of treatments,  $\sum_{i,j} V_{2ij}^2$  gives an overall dispersion assessment of the consistency of treatments, and so on. The  $\sum_{i,j} V_{uij}^2$  partition  $\sum_j X_{pj}^2 = T_{OPAS}$ .

Average assessments may be obtained by summing over strata. Put  $f_j = (\sqrt{n_{1\bullet j}}, \dots, \sqrt{n_{t\bullet j}})^T$  and  $\sum_j (I_t - f_j f_j^T / n_{\bullet\bullet j}) = \Sigma$  say. An *average* order  $u$  assessment of the consistency of treatments can be based on the quadratic form

$$(V_{u1\bullet}, \dots, V_{ut\bullet}) \Sigma^- (V_{u1\bullet}, \dots, V_{ut\bullet})^T = T_{Mu} \text{ say}$$

in which  $\Sigma^-$  is a generalised inverse of  $\Sigma$ . Of course  $T_{M1}$ , gives an unconditional assessment of mean score differences in treatments in contrast with  $S_M$  that gives a conditional assessment of mean score differences in treatments. The  $T_{Mu}$  are all asymptotically  $\chi_{t-1}^2$  distributed.

Note that the tests for overall and even average association may have quite large degrees of freedom. This means they are seeking to detect very general alternatives to the null hypothesis of no association. The test for average partial association may well be more focused than that for overall association, but may well still have low power because  $(t-1)(c-1)$  is quite large: the alternative to the null is too general. The moment tests here, and the generalised correlation tests following, offer focused tests for alternatives that are important and easy to understand.

For correlation tests results analogous to those in Rayner and Best (2001, section 8.2) for two-way tables are needed. There a single multinomial  $\{N_{ij}\}$  that follows a multinomial



distribution with total count  $n = N_{\bullet\bullet}$  and cell probabilities  $\{p_{ij}\}$  with  $p_{\bullet\bullet} = 1$  is assumed. Instead we need to assume  $\{N_{ij}\}$  follows a multinomial distribution with total count  $n_{i\bullet}$  and cell probabilities  $\{p_{ij}\}$  with  $p_{i\bullet} = 1$ . Nevertheless it may be shown that similar results apply.

In particular if  $\{\pi_r(i)\}$  are orthonormal on the  $\{p_{i\bullet}\}$  with  $\pi_r(i) = 1$  for all  $i$ , if  $\{\omega_s(j)\}$  are orthonormal on the  $\{p_{\bullet j}\}$  with  $\omega_s(j) = 1$  for all  $j$  and if  $V_{rs} = \sum_i \sum_j N_{ij} \pi_r(i) \omega_s(j) / \sqrt{n}$ , then the  $\{V_{rs}\}$  numerically partition the Pearson statistic and are asymptotically standard normal score test statistics. In particular  $V_{11}/\sqrt{n}$  is the Pearson product moment correlation for grouped data, and, if the scores are ranks,  $V_{11}/\sqrt{n}$  is the Spearman correlation.

To use the results immediately above in the context here first define, on the  $j$ th stratum, orthonormal polynomials  $\{\pi_{rj}(i)\}$  on  $(\hat{p}_{1\cdot j}, \dots, \hat{p}_{c\cdot j})$  with  $\pi_{rj}(i) = 1$  for all  $i$ , and  $\{\omega_{sj}(h)\}$  on  $(\hat{p}_{\cdot 1 j}, \dots, \hat{p}_{\cdot c j})$  with  $\omega_{sj}(h) = 1$  for all  $j$ . Next define, for  $r = 1, \dots, t-1$  and  $s = 1, \dots, c-1$

$$V_{rsj} = \sum_{i=1}^t \sum_{h=1}^c N_{ihj} \pi_{rj}(i) \omega_{sj}(h) / \sqrt{n_{\bullet\bullet j}}.$$

These assess order  $(r, s)$  generalised correlations on stratum  $j$ . Results in Rayner and Best (2001, section 8.2) (translated to the current context) show that

$$\sum_{r=1}^{t-1} \sum_{s=1}^{c-1} V_{rsj}^2 = X_{Pj}^2,$$

the Pearson statistic on the  $j$ th stratum. The  $V_{rsj}^2$  may be summed over strata to give an *overall* order  $(r, s)$  correlation assessment. Because strata are independent  $\sum_j V_{rsj}^2$  will have an asymptotic  $\chi_b^2$  distribution. In particular  $\sum_j V_{11j}^2$  will give an overall unconditional assessment of the usual (that is, order  $(1, 1)$  generalised) correlation. Since the sum of the strata Pearson statistics is the overall Pearson statistic, the  $\sum_j V_{rsj}^2$  numerically partition the overall Pearson statistic.

For a comparison of the conditional and unconditional analyses for the *Homosexual Marriage Example*, see Table 7.

Rayner and Best (2017a) note that care must be exercised in the application of these tests. Any response category that included no observations would normally be deleted. For the jams data see Table 2 to confirm that overall there are no null categories; however that is not the case within each stratum. For example, the first judge gives scores of 2, 3 and 3. The first, fourth and fifth categories include no responses. That means  $\sum_i N_{ihj} = n_{\cdot hj} = 0$  for  $h = 1, 4$  and  $5$  when  $j = 1$ . Thus  $X_{P1}^2$  is not defined, as is the case for both  $S_{OPAS}$  and  $T_{OPAS}$ . It is some consolation that  $T_{GA}$  is defined.

TABLE 7  
*Homosexual marriage opinions p-values*

Conditional			Unconditional		
Statistic	Value	p-value	Statistic	Value	p-value
$S_{GA}$	19.76	0.0006	$T_{GA}$	20.68	0.0004
$S_{OPA}$	26.71	0.0008	$T_{OPA}$	27.09	0.0007
$S_{MS}$	17.94	0.0001			
$S_{M1}$	22.32	0.0000	$T_{M1}$	23.71	0.0000
$S_{M2}$	2.59	0.2732	$T_{M2}$	2.44	0.2954
$S_{C11}$	17.98	0.0000	$T_C = V_{11\bullet}^2/2$	17.48	0.0000
$S_{C12}$	1.42	0.2328	$V_{12\bullet}^2/2$	2.35	0.1254
$S_{C21}$	2.33	0.1272	$V_{21\bullet}^2/2$	1.28	0.2570
$S_{C22}$	0.02	0.8777	$V_{22\bullet}^2/2$	0.03	0.8726

What is happening with  $T_{OPAS}$  is that in order to sum the Pearson statistics from each stratum the same treatment and responses need to be included in all statistics and with sparse data that need not be the case. The same applies to  $S_{OPAS}$  which is using a weighted sum of Pearson statistics.

The sparseness of the data also affects the definition of the unconditional moment and correlation tests. Both require the definition of orthonormal functions on each stratum. If there are three distinct responses then the orthonormal functions up to order two can be defined; but that for the jams data this only happens on strata 4 and 7. All other strata have two distinct responses and so only orthonormal functions of order zero and one can be defined. It is feasible that a judge not be able to distinguish between the treatments, and would therefore give the same score to all treatments. There is then no information about differences between the treatments, and such a judge should be excluded from the analysis. For these data there are no non-informative judges to be removed, but only first order orthonormal functions can be constructed on all strata. Second order orthonormal functions are not defined on all strata, so no second order analysis can be given.

It seems necessary to accept that sparse data yields sparse information. Here the data gives information about the usual (order (1, 1) generalised) correlation but not enough information about, for example, the order (1, 2) generalised correlation, to be able to say anything further.

## References

- Agresti, A. (2002). *Categorical Data Analysis*. 2<sup>nd</sup> edn: New York: Wiley.
- Davis, C.S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. New York: Springer.
- Landis, J.R., Heyman, E.R. and Koch, G.G. (1978). Average partial association in three-way contingency tables: a review and discussion of alternative tests. *International Statistical Review*, 46, 237-254.

- Landis, J.R., Cooper, M.M., Kennedy, T., Koch, G.G. (1979). A computer program for testing average partial association in three-way contingency tables (PARCAT). *Computer Programs in Biomedicine*, 9, 223-246.
- O'Mahony, M. (1986). *Sensory Evaluation of Food - Statistical Methods and Procedures*. New York: Marcel Dekker.
- Rayner, J.C.W. (2016). *Introductory Nonparametrics*. Copenhagen: Bookboon.
- Rayner, J.C.W. and Beh, Eric J. (2009). Towards a Better Understanding of Correlation. *Statistica Neerlandica*. 63(3), 324-333.
- Rayner, J.C.W. and Best, D.J. (2001). *A Contingency Table Approach to Nonparametric Testing*. Boca Raton FL: Chapman & Hall/CRC.
- Rayner, J.C.W. and Best, D.J. (2013). Extended ANOVA and rank transform procedures. *Australian & NZ Journal of Statistics*, 55(3), 305-319.
- Rayner, J.C.W. (2017). Extended ANOVA. *Journal of Statistical Theory and Practice*, 11, 1, 208-219.
- Rayner, J.C.W. and Best, D.J. (2017a). Unconditional analogues of Cochran–Mantel–Haenszel tests. *Australian & NZ Journal of Statistics*, 59(4), 485-494.
- Rayner, J.C.W. & Best, D.J. (2018). Extensions to the Cochran–Mantel–Haenszel mean scores and correlation tests, *Journal of Statistical Theory and Practice*, DOI: 10.1080/15598608.2018.1439424.
- Rayner, J.C.W., Best, D.J. and Thas, O. (2015). Extended analysis of at least partially ordered multi-factor ANOVA. *Australian & NZ Journal of Statistics*, 57(2), 211-224.

## Supplementary Material

### 1. Derivation of the relationship between $S_M$ and $F$

In the case of the randomised block design a considerable simplification of the CMH MS statistic is possible. In this case  $n_{i\bullet j} = 1$  for all  $i$  and  $j$ . Consequently  $n_{\bullet\bullet j} = t$ ,  $n_{i\bullet\bullet} = b$ ,  $n_{\bullet\bullet\bullet} = bt$  and  $p_{i\bullet j} = n_{i\bullet\bullet} / n_{\bullet\bullet\bullet} = 1/t$ . Substituting in the definitions given in Section 4 leads to

$$(t-1)S_j^2 = t \sum_h b_{hj}^2 n_{\bullet hj} - \left( \sum_h b_{hj} n_{\bullet hj} \right)^2,$$

$$\text{cov}(M_j) = S_j^2 \{I_t - 1_t 1_t^T / t\} / t \text{ and}$$

$$\text{cov}(M) = \sum_{j=1}^b \text{cov}(M_j) = \{I_t - 1_t 1_t^T / t\} \sum_j S_j^2 / t.$$

Now  $1_t 1_t^T / t$  is idempotent, as is  $I_t - 1_t 1_t^T / t$ , and as any idempotent matrix is its own Moore-Penrose inverse

$$\text{cov}^{-}(M) = t\{I_t - 1_t 1_t^T / t\} / (\sum_j S_j^2).$$

In the definition of  $S_M$ ,

$$M - E[M] = (\sum_j M_{ij}) - (\sum_{h,j} b_{hj} n_{\bullet hj}) / t.$$

Put  $V = (V_i) = M - E[M]$  so that  $V_i = M_{i\bullet} - \sum_{h,j} b_{hj} n_{\bullet hj} / t$ . The  $V_i$  are differences between the sum of the treatment scores over all strata and the mean of these. Now

$$1_t^T V = \sum_i (M_{ij} - E[M_{ij}]) = \sum_{h=1}^c b_{hj} n_{\bullet hj} - n_{\bullet\bullet} \sum_{h=1}^c b_{hj} n_{\bullet hj} / n_{\bullet\bullet} = 0.$$

It follows that

$$S_M = V^T \text{cov}^{-}(M) V = t\{\sum_i V_i^2\} / \{\sum_j S_j^2\}.$$

To simplify this first put  $x_{ij} = \sum_h b_{hj} N_{ihj}$ . Note that  $x_{\bullet j} = \sum_h b_{hj} n_{\bullet hj}$ . On the  $j$ th block there is only one observation of treatment  $i$ , so for only one value of  $h$  is  $n_{ihj} = 1$ ; otherwise  $n_{ihj} = 0$ . Thus, for example,  $x_{ij}^2 = \sum_h b_{hj}^2 N_{ihj}$ . Consider the data set  $\{x_{ij}\}$ . For the two-way ANOVA of these data write  $TSS$  for the treatments sum of squares and  $ESS$  for the error sum of squares.

If the data  $\{x_{ij}\}$  are analysed as a one-way ANOVA or completely randomised design with blocks as treatments then the error sum of squares is  $TSS + ESS$  and is given by

$$\begin{aligned} TSS + ESS &= \sum_{i,j} (x_{ij} - x_{\bullet j} / t)^2 = \sum_{i,j} x_{ij}^2 - \sum_j x_{\bullet j}^2 / t = \sum_j \{\sum_i x_{ij}^2 - x_{\bullet j}^2 / t\} = \\ &= \sum_j \{\sum_i (\sum_h b_{hj}^2 N_{ihj}) - (\sum_h b_{hj} n_{\bullet hj})^2 / t\} = \sum_j \{\sum_h b_{hj}^2 n_{\bullet hj} - (\sum_h b_{hj} n_{\bullet hj})^2 / t\} = (t-1) \sum_j S_j^2 / t. \end{aligned}$$

Further, for the two-way ANOVA the treatment sum of squares is (note that  $\bar{x}$  means the mean of  $\{x_i\}$  while  $x_{\bullet}$  means the sum of the  $\{x_i\}$ )

$$\begin{aligned} TSS &= \sum_{i,j} (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2 = b \sum_i \{x_{i\bullet} / b - x_{\bullet\bullet} / (bt)\}^2 \\ &= \frac{1}{b} \sum_i \{\sum_{h,j} b_{hj} N_{ihj} - \sum_{h,j} b_{hj} n_{\bullet hj} / t\}^2 = \frac{1}{b} \sum_i V_i^2. \end{aligned}$$

It follows that

$$S_M = t\{\sum_i V_i^2\} / \{\sum_j S_j^2\} = \frac{bt TSS}{t\{TSS + ESS\} / (t-1)} = \frac{b(t-1)F}{(b-1+F)}$$

on using  $TMS = TSS/(t-1)$ ,  $EMS = ESS/\{(b-1)(t-1)\}$  and  $F = TMS/EMS$ , the F test statistic for treatments in the two-way ANOVA.

## 2. Derivation of the CMH C test statistic with the same scores in every stratum

Suppose that the same treatment and response scores are used on each stratum or block. Then  $a_{ij} = a_i$  say, for all  $i$  and  $j$ , and  $b_{hj} = b_h$  say, for all  $h$  and  $j$ . As  $n_{i\bullet} = 1$  for all  $i$  and  $j$ , and as a consequence  $n_{\bullet j} = t$  for all  $j$   $S_{XXj} = \sum_i a_{ij}^2 n_{i\bullet j} - (\sum_i a_{ij} n_{i\bullet j})^2 / n_{\bullet j} = \sum_i a_i^2 - (\sum_i a_i)^2 / t$ , which is independent of  $j$ , so we may write  $S_{XXj} = S_{XX}$ . It is *not* true that  $S_{YYj}$  is independent of  $j$ .

Next we focus on  $S_{YYj}$ . Put  $x_{ij} = \sum_h b_{hj} N_{ihj}$ . Then  $x_{ij}^2 = \sum_h b_{hj}^2 N_{ihj}$ . Consider the data set  $\{x_{ij}\}$ . If these are analysed as a one-way ANOVA or completely randomised design with blocks as treatments then the error sum of squares is

$$\begin{aligned} \sum_{i,j} (x_{ij} - \bar{x}_{\bullet j})^2 &= \sum_{i,j} x_{ij}^2 - t \sum_j \bar{x}_{\bullet j}^2 = \sum_j \{ \sum_i x_{ij}^2 - x_{\bullet j}^2 / t \} = \\ &= \sum_j \{ \sum_i (\sum_h b_{hj}^2 N_{ihj}) - (\sum_h b_{hj} n_{\bullet hj})^2 / t \} = \sum_j \{ \sum_h b_{hj}^2 n_{\bullet hj} - (\sum_h b_{hj} n_{\bullet hj})^2 / t \} = \sum_j S_j^2. \end{aligned}$$

Finally we focus on  $S_{XYj}$ . Write  $\bar{a} = \sum_i a_i / t$ . Then

$$\sum_j C_j = \sum_{i,h,j} a_i b_{hj} \{N_{ihj} - E[N_{ihj}]\} = \sum_{i,h,j} a_i b_{hj} \{N_{ihj} - n_{\bullet hj} / t\}.$$

Now

$$\sum_{i,h,j} \{a_i - \bar{a}\} b_{hj} \{N_{ihj} - n_{\bullet hj} / t\} = \sum_{i,h,j} \{a_i - \bar{a}\} b_h N_{ihj} - \sum_{i,h,j} \{a_i - \bar{a}\} b_{hj} n_{\bullet hj} / t.$$

Since

$$\sum_{i,h,j} \{a_i - \bar{a}\} b_{hj} n_{\bullet hj} / t = \sum_i \{a_i - \bar{a}\} \sum_h b_h \sum_j n_{\bullet hj} / t = \{ \sum_h n_{\bullet h} b_h \} \sum_i \{a_i - \bar{a}\} / t = 0$$

$$C = \sum_j C_j = \sum_i \{a_i - \bar{a}\} \sum_{h,j} b_h N_{ihj}.$$

Note that  $\sum_{h,j} b_h N_{ihj}$  is the sum of the  $i$ th treatment scores over responses and blocks, and these are usually easy to calculate directly or are readily available in most packaged analyses. The  $\{a_i - \bar{a}\}$  are the centred treatment scores. Thus

$$S_C = \frac{(t-1)C^2}{S_{XX}^2 \sum_j S_{YYj}^2}$$

in which  $C$  is the sum of the products of the treatment sums and the centred treatment scores,  $S_{XX}^2$  is the sum of the squares of the centred treatment scores and  $\sum_j S_{YYj}^2$  may be read from the output for a one-way ANOVA.