National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

02-19

A Marginal Homogeneity Test for Blocked Categorical Data

D.J. Best, J.C.W. Rayner

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia Phone +61 2 4221 5076, Fax +61 2 4221 4998.
Email: karink@uow.edu.au

A MARGINAL HOMOGENEITY TEST FOR BLOCKED CATEGORICAL DATA

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

University of Wollongong and University of Newcastle

Summary

For blocked categorical data the Cochran-Mantel-Haenszel general association statistic has a simpler form than that used generally. The reader is stepped through this formula which is applied to several examples, principally from market research and sensory evaluation of food products.

Keywords: Cochran-Mantel-Haenszel (CMH) general association test; cross-classified data; market research data; randomized blocks; sensory evaluation data.

AMS SUBJECT CLASSIFICATION CODES: 62-07; 62G10

1. Introduction

This note is concerned with categorical data in blocks – the usual chi-squared test of independence in two-way contingency tables does not apply to data in blocks. The analysis here will not assign scores to the categories. Suppose the blocked data involves t products or treatments, c subjects or blocks and k categories. Let u denote a product, so that u =1, 2, ..., t, v denote a subject, so that v =1, 2, ..., c, and w denote a category, so that w = 1, 2, ..., k. We propose using a statistic based on a products by categories counts matrix, $\mathbf{R} = (r_{uw})$, with r_{uw} being the number of times product u is classified into category w, and a subject by categories counts matrix, $\mathbf{S} = (s_{vw})$ with s_{vw} being the number of times subject v is classified into category w. However, in both **S** and **T** the final outcome category, being redundant, is truncated, so **R** is $t \times (k – 1)$ and **S** is $c \times (k – 1)$. The matrix **S** is large when the number of subjects or blocks is large. We use the notation + to denote summation over a subscript, so that, for example, r_{+w} =

* Author to whom correspondence should be addressed.

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

[2] National Institute for Applied Statistics Research Australia, University of Wollongong, NSW 2522, Australia

Email: John.Rayner@newcastle.edu.au

$\sum_u r_{uv}$. Thus $r_{+w} = s_{+w}$, the number of times products over all blocks are categorised into the wth outcome category.

Our proposed statistic is a special case of the Cochran-Mantel-Haenszel (CMH) general association statistic. Define $Q = \{(t-1)/t\}\sum_{u=1}^{t} d_u^t \mathbf{V}^{-1} d_u$ where d_u has wth element $(r_{uw} - r_{+w}/t)$ for $u = 1, \ldots, t$ and $w = 1, \ldots, (k-1)$. Further write $\mathbf{T}$ for the $(t-1) \times (t-1)$ diagonal matrix with elements s_{+w} for $w = 1, \ldots, (k-1)$ and $\mathbf{V} = \mathbf{T}/t - \mathbf{S}^{\mathrm{T}}\mathbf{S}/t^2$. It can be shown that Q has an asymptotic chi-squared distribution with $(k-1)(t-1)$ degrees of freedom: $\chi^2_{(k-1)(t-1)}$. That the CMH general association statistic can be given in this form is shown in Rayner and Best (2018).

The convenient matrix formula we give for $\mathbf{V}$ is new. Note that $\mathbf{V}$ is a $(k-1) \times (k-1)$ matrix and not a larger $(k-1)(t-1) \times (k-1)(t-1)$ matrix that the traditional CMH formula requires inverting. With our formulation Q does not use category scores which are sometimes subjectively chosen. A common approach to the analysis of blocked categorical data is to use a repeated measures analysis of variance which, unlike Q, relies on category scores and the assumption of normal residuals.

We begin by illustrating that the Stuart (1955) test and the Cochran (1950) binary data statistic are special cases of Q. With one exception, all the data given subsequently relate to market research or sensory evaluation for food products.

2. The Stuart marginal homogeneity statistic

Table 1(a) gives cross-classified counts based on categorical responses made by 15 subjects for two different prices of the same food product. Table 1(b) gives the same data in a $t \times c$ layout. In Table 1(b) we use code 1 for buy, code 2 for undecided and code 3 for not buy. Here $c = 15$, $k = 3$, and $t = 2$.

The two products are price 1 and price 2. The first row of $\mathbf{R}$ reflects that nine times price 1 is given outcome 1, which is to buy, then four times it is given outcome 2, undecided, and twice it is given outcome 3, don't buy. The second row reflects the outcomes for the second price. The matrix $\mathbf{S}$ has 15 rows, corresponding to the subjects. The first six subjects all give two outcomes buy and no outcomes undecided. The next two give one outcome buy and one undecided, and so on. The vector d_1 has first element $r_{11} - r_{+1}/2 = 9 - 15/2$ and second element $r_{12} - r_{+2}/2 = 4 - 10/2$. Note that for $t = 2$, $d_1 + d_2$ always equals zero, so that $d_2^{\mathrm{T}} = -d_1^{\mathrm{T}}$.

We find

$$\mathbf{R} = \begin{pmatrix} 9 & 4 & 2 \\ 6 & 6 & 3 \end{pmatrix}, \mathbf{S}^{\mathrm{T}} = \begin{pmatrix} 2 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 2 & 2 & 2 & 2 & 0 & 0 \end{pmatrix},$$

$$d_1^{\mathrm{T}} = (3/2, -1), \mathbf{T} = \begin{pmatrix} 15 & 0 \\ 0 & 10 \end{pmatrix} \text{ and } \mathbf{V} = \begin{pmatrix} 3/4 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}$$

giving $Q = d_1^{\mathrm{T}}\mathbf{V}^{-1}d_1 = 3.00$ and χ_2^2 p-value 0.22. These results for Q and the p-value are just those obtained by use of the formula given by Stuart (1955).

Table 1. Counts of categorical responses made by 15 subjects for two different prices of the same food product.

(a) Cross-classified counts

	Price 2		
Price 1	Buy	Undecided	Not Buy
Buy	6	2	1
Undecided	0	4	0
Not Buy	0	0	2

(b) Coded data

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Price 1	1	1	1	1	1	1	1	1	1	2	2	2	2	3	3
Price 2	1	1	1	1	1	1	2	2	3	2	2	2	2	3	3

3. Cochran's binary response statistic

Gacula et al. (2009) in their Exercise 9.9 give responses of 12 subjects to four food prototypes to see whether the prototypes have similar and acceptable sensory characteristics compared to a control sample. Our Table 2 here reproduces this table where 1 codes acceptable and 0 codes unacceptable. Here c = 12, k = 2 and t = 4.
We find

$$\mathbf{R} = \begin{pmatrix} 10 & 2 \\ 8 & 4 \\ 3 & 9 \\ 8 & 4 \end{pmatrix}, \quad \mathbf{S}^{\mathrm{T}} = (3\ 3\ 2\ 3\ 2\ 3\ 2\ 0\ 3\ 4\ 2\ 2),$$

$$d_1 = 2.75,\ d_2 = 0.75,\ d_3 = -4.25 \text{ and } d_4 = 0.75 \text{ and } \mathbf{V} = 2.1875.$$

As is the case for all complete block binary data the d_u are all scalars, as is $\mathbf{V}$. Thus Q = 9.17 with χ_3^2 p-value 0.027. These results are identical with those obtained using the Q statistic of Cochran (1950) which, although defined differently, is the same as our Q of section 1 above.

We note that the ranking analysis of binary data in section 6.2 of Rayner et al. (2005) is incorrect.

Table 2. Responses of 12 subjects to four food prototypes

	Prototype			
Subject	1	2	3	4
1	1	1	0	1
2	1	1	0	1
3	1	0	0	1
4	0	1	1	1
5	1	1	0	0
6	1	0	1	1
7	1	1	0	0
8	0	0	0	0
9	1	1	0	1
10	1	1	1	1
11	1	0	0	1
12	1	1	0	0

4. Beyond *t* = 2 or *k* = 2

Our statistic Q can be used to test marginal homogeneity when $t > 2$ or $k > 2$. Some other statistics which do this, including some CMH implementations, need inversion of a $(k - 1)(t - 1) \times (k - 1)(t - 1)$ matrix while our Q only needs inversion of a $(k - 1) \times (k - 1)$ matrix.

Table 3 below gives five category evaluations by 20 subjects on how likely they were to purchase four cat foods. A five point likely to purchase category scale was used. The data values in Table 3 are codes for the five categories – they are not numerical values. The data are from Rayner et al. (2005, Table 3.27).

Table 3. Catfood Data

Subject	A	B	C	D	Subject	A	B	C	D
1	2	1	2	2	11	3	5	2	2
2	3	4	1	4	12	3	1	3	5
3	3	2	2	3	13	4	5	1	5
4	2	1	1	3	14	1	5	1	5
5	3	2	5	5	15	3	2	3	4
6	2	2	2	4	16	3	4	3	4
7	1	3	2	3	17	3	1	2	2
8	3	3	3	4	18	1	3	1	3
9	3	4	1	4	19	1	3	1	3
10	4	4	2	5	20	1	5	3	2

As part of the category evaluations subjects were asked to observe how keen their cat was when eating the pet foods. Table 4 gives counts for the data in Table 3. The Table 4 counts are the elements of **R**.

Table 4. Counts for Table 3 data

	Category				
Catfood	1	2	3	4	5
A	5	3	10	2	0
B	4	4	4	4	4
C	7	7	5	0	1
D	0	4	5	6	5

We find

$$\mathbf{S}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & 0 & 2 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 2 & 0 & 0 & 1 & 2 & 2 & 1 \\ 3 & 0 & 2 & 1 & 1 & 3 & 1 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 & 1 & 0 & 2 & 3 & 1 & 0 & 1 & 0 & 2 & 0 & 2 & 2 & 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 2 & 0 & 1 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Our statistic Q compares the four cat food distributions in Table 4. Product D has the most 4's and 5's, product C has most 1's and 2's while perhaps there is some market segmentation for product B.

We also find

$$d_1^{\mathrm{T}} = (1, -1.5, 4, -1),\ d_2^{\mathrm{T}} = (0, -0.5, -2, -1),\ d_3^{\mathrm{T}} = (3, 2.5, -1, -3),$$

$$d_4^{\mathrm{T}} = (-4, -0.5, -1, 3) \text{ and } \mathbf{V} = \begin{pmatrix} 2.5 & -0.5625 & -1.125 & -0.3125 \\ -0.5625 & 2.25 & -0.9375 & -0.375 \\ -1.125 & -0.9375 & 3.25 & -0.8125 \\ -0.3125 & -0.375 & -0.8125 & 1.75 \end{pmatrix},$$

giving $Q = 27.36$ with χ^2_{12} p-value 0.01. This suggests the observations made following Table 4 are statistically significant.

Table 5 gives data on cooking oils for hot chips taken from Rayner et al. (2005, Table 6.14). On p.140 of this text there is an error where $Q = 14.9$ is given instead of the correct $Q = 12.2$ using

$$d_1^{\mathrm{T}} = (-6.0, 14/3, 10/3, 5/3),\ d_2^{\mathrm{T}} = (-1.0, 2/3, -2/3, 4/3),\ d_3^{\mathrm{T}} = (7.0, 16/3, -8/3, 1/3)$$

$$\text{and } \mathbf{V} = \begin{pmatrix} 7.78 & -2.11 & -2.89 & -2.56 \\ -2.11 & 5.78 & -2.22 & -1.11 \\ -2.89 & -2.22 & 7.33 & -1.89 \\ -2.56 & -1.11 & -1.89 & 6.00 \end{pmatrix}.$$

Table 5. Cooking oils for hot chip data

Oil A	Oil B	Oil C	Oil A	Oil B	Oil C	Oil A	Oil B	Oil C
3	4	4	2	3	3	2	3	1
3	4	2	2	2	4	3	2	3
3	4	1	3	4	1	2	2	5
4	1	4	3	3	1	4	4	1
3	3	3	1	1	4	3	3	4
2	5	3	1	1	4	2	1	1
2	2	3	4	3	5	2	3	4
1	2	1	4	1	1	2	1	2
3	1	3	2	1	1	4	1	1
3	4	2	1	4	1	3	1	3
3	3	4	2	1	4	2	4	1
5	4	5	3	4	2	3	4	1
1	1	1	3	1	3	4	1	1
1	1	3	4	2	4	3	2	1
4	5	1	1	3	1	4	4	3
5	3	1	2	3	2	2	1	1
2	3	1	1	3	1	3	2	1
1	2	3	1	3	4	3	2	1
1	2	3						

All our examples except the following relate to sensory evaluation or market research for food products. Of course, the statistic Q applies to many areas. In the following scenario three pathologists, E, A and C, use a four item scale to rate diagnoses of carcinoma in situ of the uterine cervix. The diagnoses are rated 1, 2, 3 and 4 where code 1 means negative, code 2 means atypical squamous hyperplasia, code 3 means carcinoma in situ, and code 4 means squamous carcinoma. Kateri and Dellaportas (2012) give data for 118 patients in a 4 × 4 × 4 contingency table whereas in our Table 6 we give the ratings in 118 blocks of size three. Interest is in whether the three pathologists give similar diagnoses.

We find

$$\mathbf{R} = \begin{pmatrix} 16 & 30 & 54 & 18 \\ 26 & 26 & 38 & 28 \\ 31 & 42 & 37 & 8 \end{pmatrix}, \mathbf{V} = \begin{pmatrix} 6.44 & -5.67 & -0.33 \\ -5.67 & 12.00 & -5.67 \\ -0.33 & -5.67 & 11.33 \end{pmatrix},$$

$$d_1^{\mathrm{T}} = (-8.33, -2.67, 11.00),\ d_2^{\mathrm{T}} = (1.67, -6.67, -5.00) \text{ and } d_3^{\mathrm{T}} = (6.67, 9.33, -6.00).$$

This leads to Q = 61.4 with a χ_6^2 p-value less than 0.001. Pathologist E gives most ratings of 3, pathologist A gives a fairly even set of ratings while pathologist C gives very few 4 ratings.

Table 6. Ratings by three pathologists of carcinoma diagnoses

E	A	C	E	A	C	E	A	C
3	4	4	2	3	2	4	4	4
1	1	1	2	1	1	3	3	2
3	3	3	3	2	1	3	4	3
3	4	3	3	4	4	3	4	3
3	3	3	3	3	3	4	2	2
1	2	2	2	3	2	4	3	3
2	1	1	2	2	2	3	3	2
2	3	2	3	3	3	4	4	3
3	2	2	3	4	3	3	3	2
2	1	1	4	3	2	2	1	2
4	4	4	3	3	3	4	3	3
2	1	1	2	2	2	2	4	1
3	3	3	3	2	2	4	4	3
1	2	2	1	1	1	2	1	2
3	4	3	3	3	3	4	3	3
3	3	2	1	1	2	4	4	3
3	2	2	2	1	2	3	3	2
2	2	2	3	4	3	1	1	1
2	2	2	2	1	2	4	2	2
1	1	2	3	2	2	3	3	3
3	4	3	3	4	3	3	2	1
2	1	2	3	3	3	3	3	2
1	1	1	1	1	1	3	3	2
2	2	2	3	2	2	2	2	1
4	4	4	3	3	2	2	1	1
3	3	3	1	1	1	2	3	2
3	3	3	3	4	3	2	3	2
1	1	1	3	3	3	2	2	1
3	4	3	3	3	3	3	3	2
3	3	3	3	4	1	2	1	1
1	1	1	1	1	1	3	3	3
3	3	3	2	2	1	3	3	2
3	2	2	3	2	2	2	1	1
3	3	2	1	2	1	1	1	1
4	4	3	4	4	3	2	2	1
2	2	1	1	1	1	3	4	4
3	3	2	4	4	3	4	4	4
3	3	3	4	4	1	2	1	1
4	4	4	2	2	2	2	2	1
3	4	3						

5. Balanced Incomplete Blocks

So far we have concentrated on complete blocks. Here we propose using our Q statistic for balanced incomplete blocks (BIBs).

The rational for this is that the formula given for Q for complete blocks was derived in Rayner and Best (2018) using Wald-type test statistics. Such statistics depend on a likelihood, which, for cell probabilities $\{p_{uvw}\}$ and cell counts $\{N_{uvw}\}$, is of the form $\prod_{u,v,w} p_{uvw}^{n_{uvw}}$, where the product is over all tck cells in the table of counts. If the missing cells in an incomplete block are assumed to have structural zeros, the likelihoods for the complete and incomplete blocks designs will be same, because $p^0 = 1$. We now give an example of applying the formula to a data set using a balanced incomplete block (BIB) design.

Best et al. (2006) consider off flavour in $t = 6$ ice-creams using a $k = 7$ point category scale and a BIB design with $c = 15$ subjects. Each subject tasted four ice-creams. Table 7 gives the data. The values in Table 7 are codes for the seven categories – they are not numbers.

We find

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 1 & 2 & 2 & 2 & 2 \\ 1 & 5 & 1 & 2 & 1 & 0 & 0 \\ 3 & 3 & 2 & 2 & 0 & 0 & 0 \\ 5 & 4 & 0 & 0 & 1 & 0 & 0 \\ 3 & 2 & 3 & 0 & 2 & 0 & 0 \\ 4 & 3 & 3 & 0 & 0 & 0 & 0 \end{pmatrix} \text{ and}$$

$$\mathbf{S}^{\mathrm{T}} = \begin{pmatrix} 2 & 1 & 0 & 0 & 2 & 2 & 0 & 3 & 0 & 0 & 2 & 1 & 2 & 1 & 0 \\ 1 & 0 & 2 & 2 & 0 & 1 & 0 & 1 & 3 & 2 & 0 & 2 & 1 & 2 & 1 \\ 0 & 2 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

From **R** and **S** we have

$$d_1^{\mathrm{T}} = (-8/3, -2, -2/3, 1, 1, 5/3),\ d_2^{\mathrm{T}} = (-5/3, 2, -2/3, 1, 0, -1/3),$$
$$d_3^{\mathrm{T}} = (1/3, 0, 1/3, 1, -1, -1/3),\ d_4^{\mathrm{T}} = (7/3, 1, -5/3, -1, 0, -1/3),$$
$$d_5^{\mathrm{T}} = (1/3, -1, 4/3, -1, 1, -1/3),\ d_6^{\mathrm{T}} = (4/3, 0, 4/3, -1, -1, -1/3) \text{ and}$$

$$\mathbf{V} = \begin{pmatrix} 1.778 & -0.361 & -0.250 & -0.056 & -0.139 & -0.083 \\ -0.361 & 2.056 & -0.250 & -0.167 & -0.194 & -0.028 \\ -0.250 & -0.250 & 1.333 & -0.083 & -0.083 & -0.056 \\ -0.056 & -0.167 & -0.083 & 0.778 & -0.083 & 0.000 \\ -0.139 & -0.194 & -0.083 & -0.083 & 0.833 & 0.000 \\ -0.083 & -0.028 & -0.056 & 0.000 & 0.000 & 0.278 \end{pmatrix}.$$

Table 7. Off flavour ratings for six ice creams

Subject/Ice cream	A	B	C	D	E	F
1	6	1	1	2	-	-
2	6	-	-	1	3	3
3	-	4	2	-	5	2
4	7	2	3	-	2	-
5	3	5	-	1	-	1
6	-	-	1	1	3	2
7	7	4	4	-	-	3
8	2	-	1	1	1	-
9	-	2	-	2	2	3
10	4	2	-	2	5	-
11	5	-	3	-	1	1
12	-	3	2	1	-	2
13	4	2	-	-	1	1
14	5	-	2	2	-	1
15	-	2	4	5	3	-

Hence Q = 34.01 with a χ^2_{30} p-value 0.28. Best et al. (2006) obtained a similar value for Q using a more complicated alternative method involving setting up a $t \times k$ matrix of 0's and 1's for each subject.

6. Conclusion

We have given a marginal homogeneity test statistic for categorical data in blocks. Both randomized block data and balanced incomplete block data were considered. Our test statistic does not require scores to be assigned to the categories, does not require a $t \times k$ matrix of 0's and 1's to be set up for each subject and only involves inversion of a $(k-1) \times (k-1)$ categories by categories matrix rather than a $(k-1)(t-1) \times (k-1)(t-1)$ matrix. The simpler formula we give for Q could assist with teaching explanations of the application of the CMH general association methodology to blocked categorical data.

References

Best, D.J., Rayner, J.C.W. and Thas, O. (2006). Nonparametric analysis of blocked ordered categories data: Some examples revisited. *Journal of Applied Mathematics and Decision Sciences*, volume 2006, id=31089, pages 1-9.

Cochran, W.G. (1950). The comparison of percentages in matched samples. *Biometrika*, **37**, 256-266.

Gacula, M., Singh, J., Bi, J. and Altan, S. (2009). *Statistical Methods in Food and Consumer Research.* 2nd edn: New York: Wiley.

Kateri, M. and Dellaportas, P. (2012). Conditional symmetry models for three-way contingency tables. *Journal of Statistical Planning and Inference*, **142**, 2430-2439.

Rayner, J.C.W. and Best, D.J. (2018). Conditional and unconditional tests for the extended Stuart-Maxwell design. Submitted.

Rayner, J.C.W., Best, D.J., Brockhoff, P.B. and Rayner, G.D. (2005). *Nonparametrics for Sensory Science – A More Informative Approach*. Ames, Iowa: Blackwell.

Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, **42,** 412-416.