2016

Applications of Madansky's Q

D J. Best
University of Wollongong

J. C. W Rayner
University of Wollongong

Applications of Madansky's Q

Abstract

In this expository note we draw the reader's attention to Madansky's Q statistic which tests marginal homogeneity of repeated measures categorical data. Some classic voting intention data and some sensory evaluation data are used as illustrations.

National Institute for Applied Statistics Research Australia

University of Wollongong

Working Paper

05-16

Applications of Madansky's Q

D. J. BEST and J. C. W. RAYNER

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

Applications of Madansky's Q

D. J. BEST[1] and J. C. W. RAYNER[2,1]
John.Best@newcastle.edu.au and John.Rayner@newcastle.edu.au

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

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

Abstract: In this expository note we draw the reader’s attention to Madansky’s Q statistic which tests marginal homogeneity of repeated measures categorical data. Some classic voting intention data and some sensory evaluation data are used as illustrations.

AMS Subject Classification: 62G07, 62G10

Keywords: Correlated counts; Marginal homogeneity; Nonparametric statistics; Repeated measures categorical response data; Sensory evaluation.

1. Introduction

The McNemar (1947), Cochran (1950) and Stuart (1955) nonparametric statistics are special cases of the Madansky (1963) ‘interchangeability’ statistic, Q, which tests marginal homogeneity of repeated measures categorical response data. We remind the reader that repeated measures data implies the same subject makes repeated responses and the traditional Pearson X^2 test is not appropriate as not all the observations are independent. Categorical repeated measures data can be presented as square contingency tables. In general suppose responses for t treatments are categorised into k categories. McNemar’s statistic can be used to analyse 2×2 square tables. Stuart’s statistic can be used to analyse $k \times k$ square tables for $k \geq 2$. There are many applications of these two nonparametric tests that compare marginal homogeneity of two treatments. Cochran's test compares $t > 2$ treatments when $k = 2$. In our experience marginal homogeneity is often what a scientist with categorical data and treatments to compare is interested in. Here we consider a nonparametric test for general t and k.

A common parametric option for ordinal repeated measures categorical data is to ignore the categorical nature of the data and use repeated measures or randomised block ANOVA. Ignoring the categorical nature of the data is an assumption that Madansky analysis does not make. Other more sophisticated parametric options such as generalised estimating equations or weighted least squares make further assumptions.

Section 2 below introduces some classic voting data and some sensory evaluation data. Section 3 defines Q with notation changes that we hope will aid the reader and section 4 gives some concluding remarks. An Appendix gives some computing details. Our first application involves larger repeated measures data where it is convenient to present the data as square contingency tables.

2. Two Examples

2.1 Voting Intentions

Madansky (1963) gives a three-way table of counts. His counts are related to voting intentions of 450 subjects during the American election campaign of 1940. The data were collected by the Bureau of Applied Social Research, Columbia University and are shown in Table 1. The responses here are nominal.

Table 1. Voting intentions during three months (Jun, Jul, Aug) where 1 means Republican, 2 means Democrat and 3 means Undecided. The voters are from Erie County

	AUG 1			AUG 2			AUG 3		
JUL	1	2	3	1	2	3	1	2	3
JUN 1	120	6	20	2	1	1	2	1	2
JUN 2	1	2	3	2	103	6	1	4	1
JUN 3	8	1	31	1	5	30	7	8	81
Sum JUL	129	9	54	5	109	37	10	13	84
Sum AUG	192			151			107		

2.2 Sensory Evaluation Data

Suppose four varieties of tomato (Momataro, Floradade, Summit and Rutgers) are compared by 30 panellists. A seven point hedonic scale in which 1 means extreme dislike and 7 means extreme like was used. Table 2 shows the results.

For this ordinal data set we could do a randomized block ANOVA on the category code responses in Table 2. However suppose we do not know that the category codes provide sensible scores or have concerns about ANOVA assumptions. We proceed to use an analysis based on Q. For such a small data set, it serves no purpose to present the data as square contingency tables as they would contain many zero counts.

3. Madansky's *Q*

If there are t products to compare given k categories with data obtained from c subjects or judges then define $X_{uvw} = 1$ if product u is assessed by judge v as being in category w and $X_{uvw} = 0$ otherwise. Then

Table 2 Tomato flavour categories

Judge	Momotaro	Floradade	Summit	Rutgers
1	6	2	5	5
2	7	4	6	5
3	5	7	4	4
4	6	4	5	4
5	4	5	4	4
6	1	1	1	1
7	5	7	5	5
8	3	3	3	3
9	6	2	5	5
10	4	1	3	3
11	6	4	6	6
12	3	2	4	5
13	7	1	7	7
14	4	4	4	4
15	2	6	3	3
16	6	2	5	5
17	5	3	4	4
18	4	3	4	3
19	7	7	7	7
20	5	2	5	4
21	2	2	2	2
22	6	3	5	5
23	5	4	5	5
24	7	1	7	7
25	5	5	5	5
26	4	3	5	3
27	5	4	6	3
28	6	2	4	4
29	2	1	3	2
30	6	6	6	6

$$Q = \frac{t-1}{t}\sum_{u=1}^{t} q_u^T V^{-1} q_u \ ,$$

in which, for u = 1, 2, ..., t, the vector q_u has w^{th} element $X_{u \bullet w} - X_{\bullet\bullet w} / t$, for w = 1, …, k – 1 and the $(w, z)^{\text{th}}$ element of matrix V is, for w, z = 1, …, k – 1,

$$v_{wz} = -\sum_{v=1}^{c} X_{\bullet vw} X_{\bullet vz} / t^2 \text{ for } w \neq z \text{ and}$$

$$v_{ww} = X_{\bullet\bullet w} / t - \sum_{v=1}^{c} X_{\bullet vw}^2 / t^2 \text{ for } w = z.$$

A dot subscript indicates summation over that subscript. The test statistic Q can be shown to have an approximate $\chi^2_{(k-1)(t-1)}$ distribution. Notice that this formula for Q is simpler than previous formulae such as that of Somes (1986).

All the quantities needed to calculate Q can be derived from Table 1. Appendix (a) describes how to use an R computer routine to calculate Q for data given as a three way contingency table. Appendix (b) gives details for an R routine when the data are given as c lines each of t responses. Using the routine described in Appendix (a) we find for the Madansky data of Section 2.1 that Q = 70.76 on four degrees of freedom, so there is highly significant marginal heterogeneity. Madansky gave Q as 70.77 with the $(1, 1)^{\text{th}}$ element of V as 20.45 whereas we obtained 20.44 for this element. Table 3 gives the marginal counts.

Table 3. Marginal Voting counts

	1	2	3
JUN	155	123	172
JUL	144	131	175
AUG	192	151	107

From Table 3 we see that by August many of the undecided voters had decided for whom they would vote. The Republicans picked up about twice as many of these voters as did the Democrats.

For the sensory evaluation data of section 2.2 we find using the routine described in Appendix (b) that Q = 44.87 with an approximate chi-squared p-value of 0.004 based on 18 degrees of freedom. The marginal counts were as in Table 4. The distribution of counts and the 0.004 p-value indicates variety Momataro has the best flavour and Floradade the worst.

Table 4. Marginal flavour counts

Variety	1	2	3	4	5	6	7
Momataro	1	3	2	5	7	8	4
Floradade	5	7	5	6	2	2	3
Summit	1	1	4	7	10	4	3
Rutgers	1	2	6	7	9	2	3

4. Conclusion

We noted that the McNemar, Cochran and Stuart statistics are special cases of Q. Moreover the test based on Q is nonparametric and so avoids assumptions that alternative tests make, and Q applies for k, $t > 1$. Also we note that Rayner and Best (2001) give an alternative derivation of Q to that of Madansky (1963). For other recent references to data where Q could be applied see, for example, Agresti (2013, Table 13.10) and Stokes et al. (2012, Table 6.14). For a sensory evaluation example using a repeated measures categorical JAR scale and following an alternative approach of Best and Rayner (2001) to the calculation of Q see Bi (2015). The Q statistic copes with both nominal and ordinal data.

References

Agresti, A. (2013). *Categorical Data Analysis* (3rd edition). New York: Wiley.

Best, D.J. and Rayner, J.C.W. (2001). Application of the Stuart test to sensory evaluation data. *Food Quality and Preference*, 6, 221-225.

Bi, J. (2015). *Sensory Discrimination Tests and Measurements - Sensometrics in Sensory Evaluation* (2nd edition). New York: Wiley.

Cochran, W.G. (1950). The comparison of percentages in matched samples. *Biometrika*, 39, 256-266.

Madansky, A. (1963).Tests of homogeneity of correlated samples. *Journal of the American Statistical Association*, 58, 97-119.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika*, 12, 153-157.

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

Somes, G.W. (1986). The generalized Mantel-Haenszel statistic. *The American Statistician*, 40, 106-108.

Stokes, M.E., Davis, C.S. and Koch, G.G. (2012). *Categorical Data Analysis Using SAS* (3rd edition). Cary, NC: SAS Institute Inc.

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

Appendix

(a) Calculation of *Q* when the data are given as a multidimensional square contingency table

The following R commands can be used. The first two lines are not R commands but are included to show which version of R was used. The command `install.packages` is only needed the first time package "`coin`" is used. This command will ask for an R mirror site - choose USA (CA 1) Berkeley. It may take a little time to do this.

```
#R version 3.2.5 (2016-04-14)

#Platform: x86_64-w64-mingw32/x64 (64-bit)


>install.packages("coin")
>library(coin)
## Madansky (1963, pp. 107-108)
>vote <- array(
c(120, 1, 8, 2, 2, 1, 2, 1, 7,
6, 2, 1, 1, 103, 5, 1, 4, 8,
20, 3, 31, 1, 6, 30, 2, 1, 81),
dim = c(3, 3, 3),
dimnames = NULL)
>mh_test(as.table(vote))
```

The following output is obtained.

```
        Asymptotic Marginal Homogeneity Test

data:  response by
         conditions (Var1, Var2, Var3)
         stratified by block
chi-squared = 70.763, df = 4, p-value = 1.565e-14
```

(b) Calculation of *Q* when the data are given as *c* lines each of *t* responses

The R package '`crblocks`' can be installed from a CRAN repository using the following commands in R

```
>install.packages("crblocks")
```

Thereafter the package is accessed by entering

```
>library(crblocks)
```

To perform the analysis, for example using a file called `"tom.dat"` in directory `C:/Users/user` and using 1000 permutations as a check on the chi-squared approximation use the function `"catrandpvaluepermute"`:

```
>catrandpvaluepermute("C:/Users/user/tom.dat",1000)


        Statistic      df     Value      Chi2P      Sim P
            S          18     44.87      0.004      0.003
```

The `tom.dat` file should be the 30 rows of four category codes given in Table 2. Note that $Q = S$.