Optimal Data Analysis
Vol. 4 (April 24, 2015), 29-31

2155-0182/10/$3.00

UniODA *vs*. Bowker's Test for Symmetry: Region of Residence and Time

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

Bowker's test for symmetry is a generalization of McNemar's test for correlated proportions that is used for tables having more than two categories. The present study contrasts results achieved using Bowker's test versus an iterative UniODA-based procedure.

Bowker's test for symmetry—which is identical to McNemar's test for correlated proportions[1-3] for 2x2 tables—is used with square tables having more than two categories.[4] For both of these tests the null hypothesis is that the cell proportions are symmetric: that is, $p_{ij} = p_{ji}$ for all pairs of table cells. Both tests are inherently two-tailed (i.e., the alternative hypothesis is non-directional), and Bowker's test is chi-square asymptotic-based: it ignores empty cell pairs, and the minimum expectation assumption[5] must be satisfied for all cells in the table (however, an alternative exact methodology has been developed[6]).

The present exposition compares results achieved by using Bowker's test versus an iterative UniODA-based procedure previously demonstrated in analysis of turnover tables reflecting serial voting behavior, and in analysis of Markov transition models of geological soil sections.[7] The current application models both stability as well as change in region of residence (North East, Midwest, South, West) using data collected by the US Bureau of the Census in 1980 and 1985. These data were analyzed by Bowker's method using log-linear models that tested for independence, quasi independence, and quasi symmetry—all of which failed to achieve satisfactory fit and were thus found to be statistically untenable.[8]

Table 1: Data for Example[8]

Region of Residence				
	In 1985			
In 1980	North East	Midwest	East	South
North East	11,607	100	366	124
Midwest	87	13,677	515	302
East	172	225	17,819	270
South	63	176	286	10,192

The first step of the UniODA analysis tested the *a priori* hypothesis that region of residence was stable between 1980 (treated as the attribute) and 1985 (treated as the class variable).[9] This stability model specified that the

2155-0182/10/$3.00

data fell into the major diagonal of the table: it was statistically significant (exact p<0.0001), and had a very strong ESS statistic of 93.7—indicating that the model achieved 93.7% of the classification accuracy that is possible to attain above what is achieved by chance alone[7]). This model correctly classified 53,295 (95.2%) of the total of 55,981 observations in the table: the overwhelming majority of the residences did not change between regions from 1980 to 1985, and the resulting small minority of residences that did change regions will be troublesome to model via asymptotic-based statistical methods.[8]

In the second step of the UniODA analysis an exploratory model of regional residence changes (marginal dissymmetry) was sought: the directional hypothesis was eliminated and the table cells successfully modeled in the first step were set equal to zero in the UniODA data table.[10] The resulting model identified the movement of 172 residences from the East region to the North East, versus the 2.1-fold greater movement of 366 residences from the North East region to the East; as well as the movement of 176 residences from the South region to the Midwest, versus the 1.7-fold greater movement of 302 residences from the Midwest to the South. This result was statistically significant (exact p<0.0001), but represented a relatively weak effect[7] (ESS=21.1).

The third step of the analysis sought to identify a second marginal dissymmetry model of residence relocation: table cells successfully modeled in the first two steps were set equal to zero as seen in the UniODA data table.[11] The resulting model identified the movement of 63 residences from the South to the North East, versus the 2.0-fold greater movement of 124 residences from the North East to the South; as well as the movement of 225 residences from the East to the Midwest, versus the 2.3-fold greater movement of 535 residences from the Midwest to the East. This result was statistically significant (exact p<0.0001), and represented a moderate effect[7] (ESS=35.7).

A fourth UniODA statistical analysis is not conducted: for a categorical design with C class categories and C non-empty cells, classification accuracy and ESS will always be perfect.[7] A final exploratory model can however be identified without ascertaining Type I error or the ESS statistic: as seen in the UniODA data table, all cells already successfully modeled were set to zero.[12] The resulting model identified the movement of 87 residences from the Midwest to the Northeast, versus the comparable 1.1-fold greater movement of 100 residences from the North East to the Midwest; as well as the movement of 270 residences from the East to the South, versus the comparable 1.1-fold greater movement of 286 residences from the South to the East.

Together the four UniODA models correctly classified all of the data in the original table. The three UniODA models for which statistical significance was ascertained together correctly classified 55,238 (98.7%) of the total of 55,981 observations in the table, yielding an overall ESS statistic of 98.2. The initial *a priori* analysis showed that the overwhelming effect was that region of residence was stable between 1980 and 1985. Nevertheless two exploratory analyses identified eight specific statistically reliable instances of marginal dissymmetry. The final UniODA model revealed that the four residual cells in the table reflected marginal symmetry vis-à-vis comparable proportional changes. All of the analyses reported herein together required a total of 7 CPU seconds to solve running UniODA software[7] on a 3 GHz Intel Pentium D microcomputer.

References

[1]McNemar Q (1969). *Psychological statistics*. Hoboken, NJ, Wiley.

[2]Yarnold PR (2015). UniODA *vs*. McNemar's test for correlated proportions: Diagnosis of disease before *vs*. after treatment. *Optimal Data Analysis*, *4*, 24-26. URL: http://odajournal.com/2015/04/22/unioda-vs-mcnemars-test-for-correlated-proportions-diagnosis-of-disease-before-vs-after-treatment/

[3]Yarnold PR (2015). UniODA *vs*. McNemar's test: A small sample analysis. *Optimal Data Analysis*, *4*, 27-28. URL: http://odajournal.com/2015/04/22/unioda-vs-mcnemars-test-a-small-sample-analysis/

[4]Bowker A H (1948). Bowker's test for symmetry. *Journal of the American Statistical Association*, *43*, 572–574. URL: http://www.jstor.org/stable/2280710

[5]Yarnold JK (1970). The minimum expectation in χ^2 goodness of fit tests and the accuracy of approximations for the null distribution. *Journal of the American Statistical Association*, *65*, 864-886. URL: http://www.jstor.org/stable/2284594

[6]http://library.wolfram.com/infocenter/MathSource/7634/

[7]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*, Washington, DC, APA Books.

[8]Agresti A (1990). *Categorical data analysis*. Hoboken, NJ, Wiley (pp. 356-357, 360-361).

[9]The UniODA syntax used to conduct this analysis is:

```
OPEN DATA;
OUTPUT EXAMPLE.OUT;
CATEGORICAL ON;
TABLE 4;
CLASS COL;
DIRECTIONAL < 1 2 3 4;
MCARLO ITER 25000 TARGET .001 STOP 99.999;
DATA;
11607 100 366 124
87 13677 515 302
172 225 17819 270
63 176 286 10192
END DATA;
GO;
```

[10]The UniODA syntax used to conduct this analysis is:

```
OPEN DATA;
OUTPUT EXAMPLE.OUT;
CATEGORICAL ON;
TABLE 4;
CLASS COL;
MCARLO ITER 25000 TARGET .001 STOP 99.999;
DATA;
0 100 366 124
87 0 515 302
172 225 0 270
63 176 286 0
END DATA;
GO;
```

[11]The same UniODA syntax used to conduct the prior analysis was used for the present analysis, but with the following data table:

```
0 100 0 124
87 0 515 0
0 225 0 270
63 0 286 0
```

[12]The same UniODA syntax used to conduct the prior analysis was used for the final analysis, but the MCARLO command was eliminated and the following data table was used:

```
0 100 0 0
87 0 0 0
0 0 0 270
0 0 286 0
```

Author Notes

This study involved secondary data analysis of published de-identified data and was exempt from Institutional Review Board review.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., #163
Chicago, IL 60646
USA