Advances and Applications in Statistics
Volume 28, Number 2, 2012, Pages 109-121
Published Online: September 2012
Available online at http://pphmj.com/journals/adas.htm
Published by Pushpa Publishing House, Allahabad, INDIA

MEASURE OF DEPARTURE FROM SYMMETRY MODEL FOR SQUARE CONTINGENCY TABLES

Bayo H. Lawal

Department of Statistics and Mathematical Sciences
Kwara State University
Malete, Nigeria
e-mail: bayo.lawal@kwasu.edu.ng

Abstract

We propose in this paper an alternative measure Ψ for measuring the degree of departure from symmetry (S) in square contingency tables. The proposed measure is based on the conditional difference asymmetry (CDAS) model and would be very useful to represent measures of departure from symmetry when the CDAS model holds. The measure is applied to previously employed sets of 3×3 and 4×4 data in Tomizawa et al. [13] and Tahata et al. [14]. Results obtained from the use of the proposed measure are very consistent with those presented in Tahata et al. and the measure is independent of both the dimension and sample size of a given table. The measure is much easier to compute and can be easily implemented in currently available Statistical Software.

1. Introduction

For an $R \times R$ square contingency table, with row variable denoted by R and column variable denoted by C, let π_{ij} denote the probability that an

2010 Mathematics Subject Classification: 62H17.

Keywords and phrases: asymmetry, conditional difference, measure, symmetry.

Received April 12, 2012

observation falls in cell (i, j), for $(i = 1, 2, ..., R, j = 1, 2, ..., R)$ and n_{ij} its corresponding observed frequency. For nominal classificatory variables in square contingency tables, interest usually centers on applying the symmetry or/and its associated decompositions (such as the quasi-symmetry etc.) models. Goodman [4] described the symmetry model (S) that has the following representation viz.:

$$\pi_{ij} = \pi_{ji}, \quad \text{for } (1 \le i < j \le R). \tag{1}$$

Most often, the symmetry model when applied to data does not often fit data. One would of course want to know why the model fails to fit our data. While numerous alternative models, ranging from asymmetry, non-independence and skew-symmetric models have been proposed, Lawal [6] discussed some of these models, however, Tomizawa et al. [13] proposed a measure of departure from symmetry defined by:

$$\Phi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2^{\lambda}-1} I^{(\lambda)}(\{\pi_{ij}^{*}\}; \{\pi_{ij}^{s}\}) \quad \text{for } \lambda > -1, \tag{2}$$

where

$$I^{(\lambda)}(.; .) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{R} \sum_{\substack{j=1 \\ j \ne i}}^{R} \pi_{ij}^{*} \left[\left(\frac{\pi_{ij}^{*}}{\pi_{ij}^{s}} \right)^{\lambda} - 1 \right] \tag{3}$$

and the value at $\lambda = 0$ is defined to be the limit as $\lambda \to 0$. Following Tomizawa et al. [13], if we assume that $\{\pi_{ij} + \pi_{ji}\}$ are all positive for $i \ne j$, then

$$\pi_{ij}^{*} = \pi_{ij}/\delta; \; \pi_{ij}^{s} = (\pi_{ij}^{*} + \pi_{ji}^{*})/2; \; \text{ and } \; \delta = \sum_{i \ne j}\sum \pi_{ij}.$$

They further indicated that, if we let $\pi_{ij}^{c} = \pi_{ij}/(\pi_{ij} + \pi_{ji})$ be the conditional probability that $(\mathrm{R, C}) = (i, j)$ on condition that $(\mathrm{R, C}) = (i, j)$ or (j, i) for $i \ne j$, then (2) and (3) can be expressed, respectively, as:

$$\Phi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2^{\lambda}-1}\sum_{i<j}\sum(\pi_{ij}^{*}+\pi_{ji}^{*})I_{ij}^{(\lambda)}\left(\{\pi_{ij}^{c},\pi_{ji}^{c}\};\left\{\frac{1}{2},\frac{1}{2}\right\}\right), \tag{4}$$

$$I_{ij}^{(\lambda)}(.;.) = \frac{1}{\lambda(\lambda+1)}\left\{\pi_{ij}^{c}\left[\left(\frac{\pi_{ij}^{c}}{1/2}\right)^{\lambda}-1\right]+\pi_{ji}^{c}\left[\left(\frac{\pi_{ji}^{c}}{1/2}\right)^{\lambda}-1\right]\right\} \tag{5}$$

with again the value at $\lambda = 0$ taken to be the limit as $\lambda \to 0$.

2. Measure of Departure from Symmetry

Tomizawa et al. [13] employed the $\Phi^{(\lambda)}$ above as a measure of departure from symmetry model and that the degree of departure from symmetry increases as $\Phi^{(\lambda)}$ increases. In this study, we would employ a measure based on the asymmetry model Tomizawa et al. [15] and Lawal [7] defined as:

$$\pi_{ij} = \begin{cases}\Delta_{ij}\psi_{ij}, & (i < j),\\ \psi_{ij}, & (i \geq j),\end{cases} \tag{6}$$

where $|\Delta_{ij}| = \Delta$ and $\psi_{ij} = \psi_{ji}$. With π_{ij}^{c} defined as above, then (6) can thus be expressed as

$$|\pi_{ij}^{c} - \pi_{ji}^{c}| = \Delta^{*} \quad (i \neq j), \tag{7}$$

where

$$\Delta^{*} = \frac{(\Delta-1)}{(\Delta+1)}.$$

The model described in (7) has been described as the *conditional difference asymmetry* (CDAS) *model*, with the following properties:

(a) $0 \leq \Delta^{*} < 1$.

(b) If $\Delta^{*} = 0$ and hence $\Delta = 0$, then the model in (7) reduces to the symmetry (S) model.

(c) If the model in (7) holds, then the degree of departure from symmetry increases as the parameter Δ^* approaches 1. Thus, Δ^* describes the degree of the structure of asymmetry of the table. The model is based on $(R-2)(R+1)/2$ d.f.

Thus, a good measure of departure from symmetry, proposed here is defined as:

$$\Psi = \frac{(\Delta - 1)}{(\Delta + 1)}, \tag{8}$$

where $\Delta = \exp(\delta)$. We see that $0 \leq \Psi \leq 1$ and the degree of departure from symmetry increases as Ψ approaches 1. This, thus should be a good alternative to $\Phi^{(\lambda)}$.

2.1. Asymptotic variance of Ψ

If we define $g(\Delta)$ to be

$$g(\Delta) = \frac{(\Delta - 1)}{(\Delta + 1)} = \frac{f(\Delta) - 1}{f(\Delta) + 1}.$$

Then the asymptotic variance of $g(\hat{\Delta})$ can be obtained using the *delta method*, Bishop et al. [2] and Lawal [5] as $[g'(\hat{\Delta})]^2 \operatorname{var}(\hat{\Delta})$. Hence, the asymptotic variance of $\hat{\Psi}$ can be obtained from the following expression:

$$\operatorname{var}(\hat{\Psi}) = \frac{4\hat{\Delta}^2}{(\hat{\Delta} + 1)^4} \hat{S}^2, \tag{9}$$

where $\hat{S}$ is the estimated standard error of $\log(\hat{\Delta})$.

3. Other Measures of Departure from Symmetry Model

A similar measure ϕ has been introduced by Tahata et al. [14] for ordinal categorical situation which utilizes the conditional symmetry model:

$$\pi_{ij} = \Delta \pi_{ji}, \quad \text{for } (1 \leq i < j \leq R). \tag{10}$$

Their proposed measure is defined as:

$$\phi = \frac{4}{\pi}\cos^{-1}\left(\frac{\Delta}{\sqrt{\Delta^2 + 1}}\right) - 1 \tag{11}$$

with corresponding asymptotic variance for ϕ defined as (Tahata et al. [14]):

$$\operatorname{var}(\phi) = \sum_{i<j}\sum(\pi_{ij}D_{ij}^2 + \pi_{ji}D_{ji}^2), \tag{12}$$

where for $i < j$,

$$D_{ij} = \frac{4}{\pi\delta}\left\{\cos^{-1}\left(\frac{\pi_{ij}}{\sqrt{\pi_{ij}^2 + \pi_{ji}^2}}\right) - \frac{\pi_{ij}(\pi_{ij} + \pi_{ji})}{\pi_{ij}^2 + \pi_{ji}^2}\right\} - \frac{\phi + 1}{\delta},$$

and

$$D_{ji} = \frac{4}{\pi\delta}\left\{\cos^{-1}\left(\frac{\pi_{ij}}{\sqrt{\pi_{ij}^2 + \pi_{ji}^2}}\right) + \frac{\pi_{ij}(\pi_{ij} + \pi_{ji})}{\pi_{ij}^2 + \pi_{ji}^2}\right\} - \frac{\phi + 1}{\delta}.$$

We propose here that the asymptotic variance for ϕ can similarly be obtained by employing the *delta method* again if we define $G(\Delta)$ as:

$$G(\Delta) = \frac{4}{\pi}\cos^{-1}\left(\frac{\Delta}{\sqrt{\Delta^2 + 1}}\right) - 1.$$

Then the expression for the asymptotic variance becomes:

$$\operatorname{var}(\phi) = \frac{4^2}{\pi^2}\left(\frac{-1}{\hat{\Delta}^2 + 1}\right)^2 \hat{\Delta}^2\hat{S}^2, \tag{13}$$

where

$$G' = \frac{dG}{d\Delta} = \frac{4}{\pi}\left[\cos^{-1}\left(\frac{\Delta}{\sqrt{\Delta^2 + 1}}\right)\right] = \frac{-1}{\hat{\Delta}^2 + 1}$$

and is evaluated at $\Delta = \hat{\Delta}$ and $\hat{S}$ is the estimated standard error of the parameter of the model in (10). We may also observe here that the functional form of our measure from departure from symmetry (Ψ) is exactly the same as Yule's [16], Q measure of association.

3.1. Implementing the CDAS model

Following Lawal [7], the model in (6) is implemented by employing a non-standard log-linear approach employed in Lawal and Sundheim [8]. The model can be implemented in SAS, STATA or SPSS, however, we employ SAS PROC GENMOD for our analysis in this paper. The model in (6) can be implemented with the log-linear model formulation:

$$l_{ij} = \mu + \lambda_{ij}^{\mathbf{S}} + \delta \mathrm{U}_{ij}, \tag{14}$$

where the factor variable **S** relates to symmetry factor variable (Lawal and Sundheim [8]) and the regression scalar variable which is generated as:

$$\mathrm{U}_{ij} = \begin{cases} 2(R - j) + 2, & \text{if } i < j, \\ 2(R - i) + 3, & \text{if } i > j, \\ 1, & \text{if } i = j \end{cases}$$

with additional restrictions,

$$\mathrm{U}_{12} = 2R - 1; \text{ and } \mathrm{U}_{21} = 2R - 2.$$

4. Examples

We present in Tables 1 and 2, the 3×3 sets of tables from Anderson [1] and employed in Tomizawa et al. [12]. The data are the results of three consecutive opinion polls held in August 1971, October 1971, and December 1973 relating to attitudes towards the European Economic Commission (EEC) from a random sample of 493 Danes.

Table 1. Attitudes towards the EEC in August 1971 and October 1971

	October 1971			
August 1971	Yes	No	Undecided	Total
Yes	176	33	40	249
No	21	94	32	147
Undecided	21	33	43	97
Total	218	160	115	493

Table 2. Attitudes towards the EEC in October 1971 and December 1973

	December 1973			
October 1971	Yes	No	Undecided	Total
Yes	167	36	15	218
No	19	131	10	160
Undecided	45	50	20	115
Total	231	217	45	493

The results of computing the measure of departure from symmetry are presented in Table 3(a), with their corresponding estimated 95% confidence intervals. The standard errors are obtained as suggested above. Clearly, the degree of departure is stronger in Table 2 than in Table 1 based on the magnitude of the $\hat{\Psi}$.

Table 3(a). Estimates of Ψ, estimated 95% confidence intervals and standard errors for Tables 1 and 2

Data in table	Estimated measure $\hat{\Psi}$	Standard error (se)	95% Confidence interval
Table 1	–0.0333	0.0745	(−0.1799, +0.1127)
Table 2	0.4971	0.0656	(+0.3685, +0.6257)

Table 3(b). GOF test statistics when both the S and CDAS models are applied to Tables 1 and 2

	Symmetry (S)			Asymmetry (CDAS)		
Table	d.f.	G^2	X^2	d.f.	G^2	X^2
Table 1	3	8.7221	8.6001	2	8.5221	8.4094
Table 2	3	50.1493	46.9212	2	4.9125	4.8745

The confidence interval for $\hat{\Psi}$ in Table 1 includes zero, and this would suggest that either there is a degree of symmetry structure in Table 1, or the degree of departure from symmetry is not too strong. The likelihood ratio test statistic and Pearson's test statistic computed under models S and CDAS are similarly presented in Table 3(b). Clearly, the S model does not fit Table 2 and the magnitude of the GOF's indicates there is very little structure of symmetry in this table. For Table 1 however, the G^2 values are not too large and are much closer than the theoretical value of $\chi_3^2 = 7.81$ for the model to fit. However, there is a dramatic improvement in the fit under the CDAS model for Table 2.

We also compare our results with those obtained by Tahata et al. [14] when they applied their measure ϕ to the set of 4×4 tables presented in Tables 4(a) to 4(c).

Table 4. Unaided distant vision for three sets of 4×4 tables

(a) Women in Britain

	Left Eye Grade				
Right Eye Grade	Best (1)	Second (2)	Third (3)	Worst (4)	Total
Best (1)	1520	266	124	66	1976
Second (2)	234	1512	432	78	2256
Third (3)	117	362	1772	205	2456
Worst (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

(b) Men in Britain

Right Eye Grade	Left Eye Grade				Total
	Best (1)	Second (2)	Third (3)	Worst (4)	
Best (1)	821	112	85	35	1053
Second (2)	116	494	145	27	782
Third (3)	72	151	583	87	893
Worst (4)	43	34	106	331	514
Total	1052	791	919	480	3242

(c) Students in Japan

Right Eye Grade	Left Eye Grade				Total
	Best (1)	Second (2)	Third (3)	Worst (4)	
Best (1)	1291	130	40	22	1483
Second (2)	149	221	114	23	507
Third (3)	64	124	660	185	1033
Worst (4)	20	25	249	1429	1723
Total	1524	500	1063	1659	4746

In Table 4, three are 4×4 unaided distance vision. Table 1, originally from Stuart [10] is that of 7474 women aged 30 to 39 employed in Royal Ordnance factories in Britain from 1943 to 1946. The data has received considerable attention by several authors. In Table 4(b) data are the unaided distance vision of 3242 men in Britain. The data was originally presented in Stuart [9]. Table 4(c) is the corresponding unaided distance vision of 4746 Japanese students aged between 18 and 25 and was originally presented in Tomizawa [11].

Except for the magnitudes of our measure $\hat{\Psi}$ relative to $\hat{\phi}$ in Tahata et al., our results generally agree, and we will undoubtedly arrive at about the same conclusions. For Table 4(a) for instance, $\hat{\Psi} = -0.0445$ with all values in the confidence interval negative, indicating as observed by Tahata et al., that the departure from symmetry from right to left eye tends to be towards the upper complete upper asymmetry. Thus, it can be assumed that the right eye is better than the left eye because the measure is negative. Results for this table in Table 5 indicate that the symmetry model fits the data poorly as does the CDAS.

Table 5. Estimated values of Ψ, and corresponding standard errors and approximate 95% confidence intervals applied to Tables 4(a), 4(b) and 4(c)

Data in table	Estimated measure $\hat{\Psi}$	Standard error (se)	95% Confidence interval
Table 4(a)	–0.0445	0.0214	(−0.0865, −0.0025)
Table 4(b)	0.0223	0.0314	(−0.0393, +0.0839)
Table 4(c)	0.0690	0.0295	(+0.0112, +0.1268)

Table 6. GOF statistics when both the S and CDAS models are applied to Tables 4(a), 4(b) and 4(c)

	Symmetry (S)			Asymmetry (CDAS)		
Table	d.f.	G^2	X^2	d.f.	G^2	X^2
Table 4(a)	6	19.2492	19.1065	5	14.9337	14.8218
Table 4(b)	6	4.7700	4.7625	5	4.2477	4.2425
Table 4(c)	6	16.9548	16.8689	5	11.4998	11.4729

For Table 4(b), the estimated measure is $\hat{\Psi} = 0.0223$. The estimated confidence interval here includes zero, indicating that there is at least a structure of symmetry in the data. Both the symmetry and CDAS models fit the data well.

In Table 4(c), $\hat{\Psi} = 0.0690$ and the confidence interval for Ψ are both positive. Thus, departure from symmetry here is towards the complete-lower asymmetry. This therefore indicates that the left eye is better than the right eye for all the students. Here too, both the symmetry and CDAS models fit the data poorly. The results obtained in this analysis therefore are very much consistent with those presented in Tahata et al. [14].

Further, we can simplify the computations of standard errors of $\hat{\phi}$ in Tahata et al. by using the expression in (13). For instance for Table 4(a), on implementing the conditional symmetry model, we have: $\log(\hat{\Delta}) = 0.1479$ with estimated standard error $S = 0.0429$, and thus, $\hat{\Delta} = 1.1594$. Hence, using the expression for the asymptotic variance in (13), we have:

$$\text{se}(\hat{\phi}) = \sqrt{\left\{\frac{4^2}{\pi^2}\left[\frac{-1}{(1.1594)^2 + 1}\right]^2 (1.1594)^2 (0.0429)^2\right\}} = 0.0270.$$

This is the exact estimated standard error reported in Tahata et al. Standard errors can also be similarly obtained for Tables 4(b) and 4(c). Our computations indicate that we get the same set of results reported in Tahata et al. [14].

For Table 4(a), when the CDAS model is applied, we have $\log(\hat{\Delta}) = -0.890$ with estimated standard error $S = 0.0429$, and thus, $\hat{\Delta} = 0.91485$. Hence, using the expression for the asymptotic variance in (9), we have:

$$\text{se}(\hat{\Psi}) = \sqrt{\left[\frac{4(0.91485)^2}{(0.91485 + 1)^4}(0.0429)^2\right]} = 0.0214.$$

The standard errors for the other tables are similarly computed.

5. Conclusions

Results obtained from the application of the proposed measure of departure from symmetry Ψ are very consistent with similar results obtained previously from the use of other measures of departure from symmetry. The advantage of this measure is that it is much easier to compute from available statistical software and does not require programming to estimate it. The measure can also be applied to contingency tables having ordered categories because this would have been well defined and one does not hope to interchange rows or columns during analysis. It can measure as similarly observed in Tahata et al. [14], two kinds of complete symmetry (the upper asymmetry and the lower asymmetry). This measure provides an alternative measure to those earlier proposed and results from application of this measure will be consistent with those in Tahata et al.

References

[1] E. B. Anderson, The Statistical Analysis of Categorical Data, 3rd ed., Springer-Verlag, Berlin, 1994.

[2] Y. M. M. Bishop, S. E. Fienberg and P. W. Holland, Discrete Multivariate Analysis: Theory and Practices, The MIT Press, Cambridge, Massachusetts, 1975.

[3] N. Cressie and T. R. C. Read, Multinomial goodness-of-fit tests, J. Roy. Statist. Soc. Ser. B 46 (1984), 440-464.

[4] L. A. Goodman, Multiplicative models for square contingency tables with ordered categories, Biometrika 66 (1979), 413-418.

[5] B. Lawal, Categorical Data Analysis with SAS and SPSS Applications, LEA, Mahwah, New Jersey, 2003.

[6] H. B. Lawal, Review of the non-independence, asymmetry, skew-symmetry and point-symmetry models in the analysis of social mobility data, Quality & Quantity 38 (2004), 259-289.

[7] H. B. Lawal, On fitting the conditional difference asymmetry models to square contingency tables with nominal categories, Quality & Quantity 42 (2008), 605-612.

[8] H. B. Lawal and R. Sundheim, Generating factor variables for asymmetry, non-independence and skew-symmetry models in square contingency tables with a SAS macro, Jour. Stat. Software 7 (2002), 1-23.

[9] A. Stuart, The estimation and comparison of strengths of association in Contingency tables, Biometrika 40 (1953), 105-110.

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

[11] S. Tomizawa, Three kinds of decompositions for the conditional symmetry model in a square contingency table, J. Japan Statist. Soc. 14 (1984), 35-42.

[12] S. Tomizawa, Two kinds of measures of departure from symmetry in square contingency tables having nominal categories, Statistica Sinica 4 (1994), 325-334.

[13] S. Tomizawa, T. Seo and H. Yamamoto, Power-divergence-type measure of departure from symmetry for square contingency tables that have nominal categories, J. Appl. Statist. 25 (1998), 387-398.

[14] K. Tahata, K. Yamamoto, N. Nagatani and S. Tomizawa, A measure of departure from average symmetry for square contingency tables with ordered categories, Australian Journal of Statistics 38 (2009), 101-108.

[15] S. Tomizawa, N. Miyamoto and R. Funato, Conditional difference asymmetry model for square contingency tables with nominal categories, J. Appl. Statist. 31 (2004), 271-277.

[16] G. U. Yule, On the association of attributes in statistics, Phil. Trans. 194 (1900), 257-319.