IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 15, Issue 2 Ser. I (Mar – Apr 2019), PP 28-32
www.iosrjournals.org

Comparative Study of Pearson's Chi-Square and Some Ordinal Contingency Tables Models

Bushirat T. Bolarinwa
Department of Statistics Federal Polytechnic, P.M.B. 55, Bida, Nigeria
Corresponding Author: Bushirat T. Bolarinwa

Abstract: *The aim of this article was to compare Pearson's Chi-square to Uniform (U), Column (C), Row (R), and R+C ordinal contingency tables models. Data on gender, university attended for B.Sc., B.Sc. and M.Sc. grades of 116 M.Sc. graduates were collected from Department of Statistics, University of Ilorin, Nigeria. Model estimation was carried out by maximum likelihood method and goodness of fit was assessed by likelihood ratio statistic. Pearson's chi-square rejected the null hypothesis of independence in all cases; the U model rejected in 2 of 6 cases while R rejected in 4 cases. The C model rejected in 3 cases while R+C rejected in 5 out of 6 cases. Pearson's chi-square reached same conclusion with U model on 2 occasions, with C model on 3 occasions, with R on 4 and with R+C on 5 occasions. It reached same conclusion with U model on the only occasion the assumption of U was satisfied and reached same conclusion with R model on 3 out 5 occasions in which the assumption of the latter was satisfied. However, it reached a contrary conclusion with C and R+C models on the only occasion the assumptions of C and R+C models were met. While the four ordinal association models reached same conclusion of independence as did the Pearson's chi-square on the only occasion the assumption of the latter was met, Pearson's chi-square reached same conclusion only on 4 out of the 8 occasions on which the assumptions of ordinal association models were met, suggesting higher robustness of ordinal association models. It was found that Pearson's chi-square agreed mostly with R+C, followed by R, then C and lastly, U model when each ordinal model underlying assumption was not taken into cognizance However, when taken into consideration, it agreed mostly with R and then, U model. The need to conduct larger scale study was recommended.*

Keywords*: Contingency table, Chi-square, Row model, Colum model, Likelihood ratio*

Date of Submission: 11-03-2019 Date of acceptance: 27-03-2019

I. Introduction

Occasions often arise when one is interested in studying association in contingency tables involving two variables. That is, when interest lies in studying association between row and column variables. The practice has always been to use Pearson's chi-square statistic due to Pearson (1900). A shortcoming of the statistic is that it does not take into consideration the fact that either of the two variables involved may be ordinal, it simply takes variables as nominal. This led to the development of models that recognize ordering. Such models termed ordinal contingency tables models include Uniform (U), Row (R), Column (C), R+C, and RC association models. Each of these models is applicable under different circumstances. The Uniform association model relies on method of integer scoring that assumes that distance between any two adjacent categories is uniform across all values, hence, the name of the model. Integer scoring is imposed on both row and column variables. For the R association model, integer scoring is imposed on the column variable only while for the C association model it is imposed on the row variable only.

The R+C association model, also referred to as *Model I* by Goodman (1979) requires that the rows and the columns variables be correctly ordered, hence, the name. It is therefore, very suitable when we have doubly ordered categories with integer spacings of the categories known. Occasions arise when the row and column scores are unknown, this rules out the possibility of using any of R, C and R+C models; a model tagged *RC Association model*, proposed by Goodman (1979) becomes readily applicable. It is sometimes referred to as *model II* in comparison to the R+C model.

A lot of researches have been conducted on association models. Pecker and Clogg (1989) reviewed the general RC model, RC (M) and proposed alternative weighting systems for identifying interaction parameters. Gokhale and Klein (1995) proposed a way of assigning scores to category level based on the marginal frequency totals of the variable.

Takare (1987) proposed a method for handling contingency table based on ideal point discriminant analysis as an alternative to loglinear modeling and correspondence analysis. Ritov and Gilula (1991) derived order restricted maximum likelihood estimators for parametric scores assigned variable levels in RC model.

Eshima, Tubata, and Tsujitani (2001) derived property of the RC (M) model and a summary measure of association in contingency table. Aktas and Saracbasi (2003) compared uniform association and quassi-independence models.

Altunay and Saracbasi (2009) proposed symmetric disagreement plus uniform association model aimed at separating the association from the disagreement. Krampe, Kateri and Kuhnt (2011) proposed algebraic approach to modeling asymmetric models. Camminatielo, D'Ambra and Sarnacchiaro (2014) proposed a general framework for the analysis of the complete set of log-odds ratios generated by two-way contingency table.

The aim of this research is to compare inferences drawn from Pearson's chi-square to those of U, R, C and R+C association models using academic performance data. The article is organized as follows: Section 2 presents the Theoretical Framework; Section 3 presents the Methodology; Section 4 presents the Results and Discussion while the last section concludes the article.

II. Theoretical Framework

Discussion on uniform association model can be motivated from the linear-by-linear association model. For two-way tables with ordinal variables, X and Y, let us assign scores x_i and y_j to row and column categories respectively.

The *linear-by-linear association* model is

$$\ln(\hat{m}_{ij}) = \mu + \lambda_i^X + \lambda_j^Y + \phi x_i y_j \tag{1}$$

When ϕ = 0, the independence model results. The parameter ϕ specifies the direction and strength of association; when $\phi > 0$, the tendency is that as *X* increases, *Y* also increases and when $\phi < 0$, the tendency is for *Y* to decrease as *X* decreases. The term $\phi x_i y_j$ being the deviation of $\ln(\hat{m}_{ij})$ from independence is linear in the *Y* scores at fixed *X* and linear in the *X* scores at fixed *Y*; the model obtains its name from this linear property (Agresti, 2007). The *uniform association model* is a special case of the linear-by-linear model in which integer scoring is used.

The row association model is a consequence of relaxation of restrictions of the uniform association model. When the integer scoring is imposed on the columns variable, the resulting model is termed *row association model*. With one restriction removed from equation 1, the resulting row association model is of the form

$$\ln(\hat{m}_{ij}) = \mu + \lambda_i^X + \lambda_j^Y + \tau_i(v_j - \bar{v}) \tag{2}$$

where

$$\bar{v} = \frac{\sum_j v_j}{J} \; and \; \sum \lambda_i^X = \sum \lambda_j^Y = \sum \tau_i = 0 \tag{3}$$

When $\tau_i = 0$ for all *i*, model reduces to that of independence. Parameter τ_i is the deviation within a particular row of $\ln(\hat{m}_{ij})$ from row independence of a known function of the ordinal variable with slope, τ_i (Lawal, 2003).

The *column association model* is also a consequence of relaxation of restrictions of the uniform model. Unlike the row association model which requires correct integer ordering of the column variables, the column association model requires correct integer ordering of the row variables. Still leaning on Equation 1, the form of the column association model is

$$\ln(\hat{m}_{ij}) = \mu + \lambda_i^X + \lambda_j^Y + \rho_j(u_i - \bar{u}) \tag{4}$$

where

$$\bar{u} = \frac{\sum_i u_i}{I} \; and \; \sum \lambda_i^X = \sum \lambda_j^Y = \sum \rho_j = 0 \tag{5}$$

Equation 4 reduces to that of independence when $\rho_j = 0$ for all *j*. The column association Parameter ρ_j is the deviation within a particular column of $\ln(\hat{m}_{ij})$ from column independence of a known function of the ordinal variable with slope ρ_j (Lawal, 2003).

This model requires that the rows and the column variables be correctly ordered, hence, the name. It is also referred to as *Model I* by Goodman (1979). Since both the row and column variables are ordinal, any changes in the order of the row or the column change the structure of the model (Lawal, 2003). This means that the model is not invariant to possible changes in the categories of the row and column variables (Powers & Xie, 1999). It is based on (*I*-2) (*J*-2) degrees of freedom. The model is of the form

$$\ln(\hat{m}_{ij}) = \mu + \lambda_i^X + \lambda_j^Y + \tau_i(v_j - \bar{v}) + \rho_j(u_i - \bar{u}) \tag{6}$$

III. Methodology

This section presents data collection, model, model estimation, and goodness of fit tests.

Data

Data are gender, university attended for B.Sc., B.Sc. Grade, and M.Sc. Grade of 116 M.Sc. Statistics graduates of University of Ilorin, Nigeria.

Gender is classified as: Male and female. Male is coded 0 while Female is coded 1.

University attended is classified as follows: Group 1 for University of Ilorin, Group 2 for other universities. University of Ilorin is coded 1 while "other universities" is coded 0.

B.Sc. Grade is classified as: First Class, Second Class Upper, and Second Class Lower. Second Class Lower is coded 1, Second Class Upper is coded 2 while First Class is coded 3.

M.Sc. Grade is classified as: Terminal, M.Phil./Ph.D Grade and Ph.D. Grade. Terminal Grade is coded 1, M.Phil./Ph.D. Grade is coded 2 and Ph.D. Grade is coded 3.

Model

Pearson's chi-square statistic is defined

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{O_{ij}} \sim \chi^2_{(r-1)(c-1)} \tag{7}$$

where O_{ij} and E_{ij} are observed and expected frequencies for the (i, j)th class.

The ordinal models involved: U, C, R, and R+C models are discussed in Section 2.

Model Estimation

Mode was estimated by maximum likelihood estimation method.

Goodness of Fit Tests

The article utilized the likelihood ratio statistic due to (G^2) by Wilks (1938)

The G^2 statistic is defined

$$G^2 = 2\sum_i n_i \log\left(\frac{n_i}{m_i}\right) \tag{8}$$

where

n_i is the observed frequency and m_i is the expected frequency

G^2 is Chi-square distributed with degrees of freedom equal to number of cells in the table less number of independent parameters estimated.

IV. Results and Discussion

Table1 presents each model and its assumption violation status.

Table 1. Model assumption status

Combination	χ^2	U	R	C	R+C
Gender Vs B.Sc.	V	V	S	V	V
Gender Vs M.Sc.	V	V	S	V	V
University Vs B.Sc.	V	V	S	V	V
University Vs M.Sc.	V	V	S	V	V
B.Sc. Vs M.Sc.	V	S	S	S	S
University Vs Gender	S	V	V	V	V

Key: V- Assumption violated
S- Assumption satisfied

The Pearson's chi-square violates all but university-gender combination. The assumptions of U, C and R + C models are satisfied by all but B.Sc.- M.Sc. classification.

Table 2. Inferences of Models for various combinations

Combination	χ^2	U	R	C	R+C
Gender Vs B.Sc.	S	N	S	N	S
Gender Vs M.Sc.	S	N	S	N	S
University Vs B.Sc.	S	N	N	S	S
University Vs M.Sc.	S	N	N	S	S
B.Sc. Vs M.Sc.	S	S	S	N	N
University Vs Gender	S	S	S	S	S

Key: S- Significant
N- Not significant

Inferences drawn from each model application on the data are presented in Table 2. Pearson's chi-square rejected the null hypothesis of independence in all cases; the U model rejected in 2 of 6 cases while R rejected in 4 cases. The C model rejected in 3 cases while R+C rejected in 5 out of 6 cases. Pearson's chi-square hence, rejected most frequently, followed by R+C, then R, C and U that rejected least frequently. Pearson's chi-square reached same conclusion with U model on 2 occasions; it reached same conclusion with C model on 3 occasions, with R on 4 and with R+C on 5 occasions. This implies that Pearson's chi-square agreed mostly with R+C, followed by R, then C and lastly, U model.

Pearson's chi-square reached same conclusion with U model on the only occasion the assumption of U was satisfied. It reached same conclusion with R model on 3 out 5 occasions in which the assumption of the latter was satisfied; however, it reached a contrary conclusion with C and R+C models on the only occasion the assumption C and R+C models had their respective assumptions satisfied. The ordinal association models reached same conclusion of independence as did the Pearson's chi-square on the only occasion the assumption of the latter was met. This is unique and may be a pointer to higher robustness of ordinal association models than Pearson's chi-square, although larger scale work may be required to validate this. Generally, out of the 8 occasions in which assumptions of ordinal association models were satisfied, Pearson's chi-square reached same conclusion only 4 times. When satisfaction of each ordinal model assumption was taken into consideration, Pearson's chi-square agreed mostly with R and then, U model.

V. Conclusion

This article has compared inferences from Pearson's chi-square to those of some ordinal association models. It was found that Pearson's chi-square agreed mostly with R+C, followed by R, then C and lastly, U model when each model underlying assumption was not taken into cognizance. However, when taken into consideration, it agreed mostly with R, followed by U model. The ordinal association models demonstrated higher robustness than Pearson's chi-square. The need to conduct larger scale study is recommended.

References

[1]. Agresti, A. (2007). *An Introduction to Categorical Data Analysis (2nd ed.).* New Jersey: John Wiley.
[2]. Aktas, S. & Saracbasi, T. (2003). Analysis of triangular contingency tables. *Hacettepe J. of Math. & Stat*., *32*, 43-51.
[3]. Altunay, S.A. & Saracbasi, T. (2009). Estimation of symmetric disagreement using a uniform association model for ordinal agreement data. *ASTA Advances in Sta. Anal*., 93(3), 335-343.
[4]. Camminatiello, I., D'Ambra, A. & Sannacchiaro, A. (2014). The association in a two-way contingency table through log odds ratio analysis: the case of Sarno river pollution. Springer Plus, 3. DOI 10.1186/2193-1801-3-384.
[5]. Eshima, N., Tubata, M. & Tsujitani,, M. (2001). Property of the RC(M) association model and a summary measure of association in the contingency table. *J. Japan Stat. Soc*., *31*(1), 15-26.
[6]. Gokhale, D.V. & Klein, R. (1995). Analysis of contingency tables by marginal scores. *Brazillian J. of Prob. & Stat*., 25-42.
[7]. Goodman, L.A. (1979). Simple models for the analysis of association in cross-classification having ordered categories. *American Journal of Sociology, 84*, 804-829.
[8]. Krampe, A., Katari, M. & Kuhnt, S. (2011). Assymetry models for square contingency tables: exact tests via algebraic statistics. *Statistics and Computing*, 21(1), 55-67.
[9]. Lawal, B. (2003). *Categorical Data Analysis with SAS & SPSS Applications*. New Jersey: Lawrence Erlbaum. *System* (2nd ed.). SAS Institute.*Symposium on Mathematical Statistics and Probability*, 239-273.
[10]. Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philo. Mag., Series*, 5(50), 157-175.
[11]. Pecker, M.P. & Clogg, C.C. (1989). Analysis of sets of two-way contingency tables using asymmetric models. *JASA*, 84(405), 142-151.
[12]. Powers, D. & Xie, Y. (1999). *Statistical methods for categorical data analysis* (*2nd ed*.). Texas: Academic Press. production and hinders pollen performance in cucurbita texana. *Ecology*, 76, 437-443.

[13]. Ritov, Y. & Giluta, Z. (1991). The ordered-restricted RC model for ordered contingency tables: estimation and testing of fit. *The Annals of Statistics*, 19(4), 2090-2101.
[14]. Takare, Y. (1987). Analysis of contingency tables by ideal point discriminant analysis. *Psychometrics*, 52(4), 493-513.
[15]. Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist*., 9, 60-62.

Bushirat T. Bolarinwa. "Comparative Study of Pearson's Chi-Square and Some Ordinal Contingency Tables Models." IOSR Journal of Mathematics (IOSR-JM) 15.2 (2019): 28-32.