African Journal of Mathematics and Statistics Studies
Volume 2, Issue 1, 2019 (pp. 37-43)

www.abjournals.org

ANALYZING STUDENTS' ACADEMIC PERFORMANCE PREFERENCE USING CONTINGENCY TABLE

Okeke Charles C.

Department of Statistics, Nnamdi Azikiwe University, Awka,

ABSTRACT: *This study investigates academic performance preference among students using contingency table by comparing between two profile variables (male and female). The comparison is done by investigating students' interest of either off school hostel or school hostel by determining if the latter is preferred over the former due to near for (Male and female) and far. Chi square test, likelihood ratio chi square test, Mantel-Haenszel chi-square and the phi coefficient was applied to determine if the above factors influence students' academic performance. The computed results indicated that the null hypothesis is rejected in favor of the alternate hypothesis. This implies that (male & female) students prefer school hostel over off school hostel due to the closeness to their classes, safe time and reduced cost of transportation. The phi coefficient based on 0-1 scale indicated that the dependency of these profile variables is moderate.*

KEYWORDS: Contingency Table, (male & female) School Hostel, Off School Hostel, Chi Square, Likelihood Ratio Chi Square.

INTRODUCTION

The term contingency table was first used by Pearson in 1904; the emphasis was based on the theory of association

(Pearson 1904). The contingency table is also called the frequency table, a table of count (Barak 2007). Contingency table is applied to organize categorical variables and testing hypothesis using the chi-square test to determine independence. In general, the contingency table describes the relationship between two categorical variables. Contingency table helps to determine the effectiveness of a system under study or if the effectiveness of the system is based on certain profile variables (Albrecht 2013).

However, chi-square test is used to analyse frequency data. This analysis can be used for the following tests: goodness of fit test; test of independence and the tests of homogeneity. Nduka (2015). The goodness of fit test is a statistical procedure used to test if an assumed distribution is correct. Hsiao-Mei (2009). When the experimenter is concerned with association between variables then the test of independence is carried out. The test of homogeneity is employed to check if two or more well-defined populations are equal. Nduka (2015).

Pearson's chi-squared test (Pearson, 1900) is easily generalizable to analysis of contingency tables. There are three different but equivalent ways of conceptualizing the test. One is as a test of homogeneity, which refers to whether two groups are different on a binary response (e.g., whether males and females differ in the choice between two candidates). This conceptualization assumes one variable is an explanatory variable and one is a response, but

the chi-square does not require such a designation. The goodness-of-fit conceptualization concerns the degree to which the observed frequencies fit the frequencies that would be expected do to chance.

Howel (2009), pointed out that the chi-square statistic is only an approximation of the chi-square distribution and the approximation only worsens with small expected frequencies. He noted that with small expected values the chi-square statistic is quite discrete rather than continuous. He also referenced Cochran as stating an arbitrary minimum value of 5 as a general guide to the implementation of the chi-square test.

Marie (2008) in a lecture on Use of the Chi-Square Statistic in a Test of Association Between a Risk Factor and a Disease at the John Hopkins Bloomberg School of Public Health designed a 2 x 2 contingency table with categories exposure (yes and no) and diseased (yes and no) found out that the chi-square obtained was 29.1 with a degree of freedom of 1. From the chi-square table, in order to obtain discrepancies between observed and expected frequencies of this magnitude less than 0.001 that is unlikely to occur by chance. Thus, the findings are unlikely to occur if there is no association between exposure and disease, hence there appears to be an association.

METHODOLOGY

Some statistical tools used to analysis students' academic performance preference using contingency table. They include: chi square test, likelihood ratio chi square test, Mantel-Haenszel chi-square and the phi coefficient will be employed.

The Contingency Table

A contingency table, though depending on the number of rows and columns can be denoted and defined in a matrix form, say

Table 1. General Form of Contingency Table.

A/B	I	C	Σ
1	N_{11}	N_{1C}	$N_{1.}$
		nd	
R	N_{R1}	N_{RC}	$N_{R.}$
Σ	$N.I$	$N.C$	N

The values of N_{ij} denotes the observed values, where $N_{i.}$ and $N_{.j}$ denotes the row and column sums or simply called the marginal counts (Andel 2002; Lauritzen 2002; Gomez and Perez, 2005). The probabilities of the observed values based on each cell is computed and denoted by

$$p_{ij} = N_{ij} / N, N = \sum_{ij=1}^{k} N_{ij.} \qquad (1)$$

Equation (2.1.3) implies that all the cells in Table 1, say all the N objects are stochastically independent. Equation (2.1.3) satisfies the following conditions:

1. $P_{ij} \geq 0$

2. $\sum_{ij} P_{ij} = 1.$

From the above, Equation (1) is unrestricted and as such,

$P_{ij} = P_{i.} * P_{.j}$

Is independent (Lauritzen 2002). Conventionally, this procedure describes relationships or non-relationships between factors being considered.

The Likelihood Ratio Chi Square

In this section, the objective is to compute and compare the Likelihood Ratio Chi square value with the Chi square table value in order to infer more information. The likelihood ratio Chi square is defined as

$$G^2 = 2\sum[O_{ij} \log(O_{ij} / E_{ij})]. \qquad (2)$$

Though, this formula is applicable for a large dimensional table which can be decomposed into smaller components (Howell, 2012), this allows its application to this data set values. The likelihood Ratio Chi square can be applied to investigate if there is a significant difference between observed and the expected frequency (Olmus and Erbas 2012). Using the formula above, the G^2 is equal to 156.3. This implies that $G^2 > \chi^2$. Suppose, we are to compare the likelihood ratio chi square value with the chi square table value at 0.05 level of significant. The conclusion will corroborate with the rejection of the null hypothesis. This result revealed that both techniques converge as the sample sizes increases. It is evident that the larger the G^2 value the higher the probability of rejecting the null hypothesis. For large sample size, G^2 has a chi square null distribution with $k-1$ degree of freedom. If the null hypothesis is true, both the Pearson chi square χ^2 and the likelihood Ratio chi square have asymptotic chi square distributions with $k-1$ degree of freedom (Agresti 2002).

Mantel-Haenszel chi-square

The Mantel-Haenszel chi-square statistic tests the alternative hypothesis that there is a linear association between the row variable and the column variable. Both variables must lie on an ordinal scale. The statistic is computed as

$$\chi^2_{MH} = \frac{\left(\sum_{i=1}^{q} a_i - E(a_i)\right)^2}{\sum_{i=1}^{q} \mathrm{var}(a_i)} \square \chi_1^2 \qquad (3)$$

where $E(a_i)$ is the expected value between the row variable and the column variable.

$Var(a_i)$ is the variance value between the row variable and the column variable.

Phi Coefficient

The measure of association, Phi is measured which adjusts the Chi square statistic by the sample size. The symbol for Phi is the Greek letter phi, written φ and usually pronounced 'fye' when used in statistics. Phi is most easily defined as

$$\varphi = \sqrt{\frac{\chi^2}{n}} \qquad (4)$$

Where χ^2 is the chi square

where n is the total number of observations.

The test hypothesis is designed as follows;

H_0: students' performance in Nnamdi Azikiwe University, Awka because (school hostel) is nearer to their classes of Male and Female.

H_1: students' performance in Nnamdi Azikiwe university, Awka because (off - school hostel) is far from their classes of Male and Female.

Therefore, the comparative acceptance of either of the above, lies strictly on computed Chi square value. In this discussion we use 5% level of significance to determine the acceptability or otherwise of the hypothesis.

DATA PRESENTATION

The data generated for this study will be analysed using chi square test, likelihood ratio chi square test, Mantel-Haenszel chi-square and the phi coefficient will be employed.

2 x 2 Contingency Table for Students' Data

This paper investigates students' academic performance preference using the school hostel and the off - school hostel as a case study in Nnamdi Azikiwe University, Awka Anambra State. This survey was carried out for the period of three months (December to March, 2019). The objective is to investigate the reason while some categories of students prefer school hostel over the off – school hostel and also to determine otherwise. The data reported in this paper are based on extensive questionnaire survey in Nnamdi Azikiwe University, Awka, Anambra State. Precisely, about 1535 students were selected for this survey of which 664 prefer school hostel and 871 prefer off school hostel. In this survey, the profile variables used are male and female. This implies that students may prefer either school hostel because it is near due to

closeness to their classes or off – school hostel because it is far from their classes. Based on the information of the survey, we apply 2 by 2 contingency table to determine if students' academic performance (school hostel or off school hostel) and the profile variables influence students' academic performance. As noted in (Ingersoll G.M. 2010) a 2 x 2 contingency table is used to conceptualize, organize and report data. The null hypothesis tested with a chi-square test based on a 2 x 2 contingency table is considered as test of independence (Ingersoll G.M. 2010). It is a well-established fact that the expected frequency should be at least five to enable the application of the chi-square meaningful otherwise other techniques are applied. Detail of the survey is reported in Table 2 below.

Table 2. Observed Value for the Students' Academic Performance Data.

	Male	**Female**	**Total**
School hostel	212	452	664
Off school hostel	556	315	871
Total	**768**	**767**	**1535**

In this consideration, we consider the test of independence by using the above data set. In this case, the Chi square distribution involves using the sample data to test for independence of the three variables. In this case we consider whether preference for Students' depends on Male or Female. For the 2x2 contingency table, the Chi square test value based on the observed and expected frequencies is computed based on the following formula

$$\chi^2 = \sum_i \sum_j (OV - EF_{ij})^2 / EF_{ij} \qquad (5)$$

Where OV denote the observed value and EF denote the expected frequency. Note that Equation (5) is approximately distributed as chi square on degrees of freedom, where and denotes the rows and columns, respectively. Observe that if the value of the observed value is greater than the expected frequency then the chi square value will be small. On the contrary, if the expected frequency is greater than the observed value the chi square value tends to be large. In this case, since the expected frequency is greater than five for all cases, it is straight forward to apply the Chi square test statistic.

Table 3. Expected Frequency for the Students' Academic Performance Data.

	Male	**Female**	**Total**
School hostel	332.2	331.8	664
Off - school hostel	435.8	435.2	871
Total	**768**	**767**	**1535**

The chi square table value at 5% level of significant with (n -1) (k -1) = (2 -1) (2 -1) = 1 degrees of freedom is equal to 3.84. Since the computed value is greater than the table value, we reject the null hypothesis and conclude that student prefer school hostel over off – school hostel because school hostel is near due to closeness to their classes.

Table 4. Computed Chi Square Value for Students' Academic Performance Data.

Observed Value (OV)	Expected Frequency (EF)	D=(OV-EF)	D^2/EF
212	332.2	- 120.2	43.492
452	331.8	120.2	43.544
556	435.8	120.2	33.153
315	435.2	- 120.2	33.199
			153.4

RESULTS AND DISCUSSION

Although the null hypothesis was rejected indicating that there is a relationship between students' academic performance and the profile variables, the chi square value of 153.4 will produce moderate Phi coefficient to indicate a moderate relationship between these two profile variables. The Phi coefficient for this data set is 0.31612. The Mantel-Haenszel chi-square value for this data set is 153.3 is equivalent to the Pearson chi square value and the likelihood ratio chi square value, respectively. The moderate Phi coefficient revealed that there is moderate association between students' academic performance and the profile variables. In addition to concluding that the students' academic performance and the profile variables have a relationship, we can also infer on the differences in the proportions in the 2 x 2 table (Utts 2006; Walpole 2011). They are: Observation 1 (School hostel): (male & female) students prefer school hostel to off school hostel because it is near due to closeness to their classes. Observation 2 (Off school hostel): most (male & female) students prefer off school hostel to school hostel because it is far from their classes. The analysis revealed (male & female) students' preference in terms of academic performance which depends strictly on the profile variables. Based on this analysis, the school management can infer quality information on how to improve the (male & female) students' academic performance in Awka, Anambra State. Nigeria.

CONCLUSIONS

Results based on the Chi Square, the Likelihood Ratio Chi square, and the Mantel-Haenszel Chi-square value revealed that the null hypothesis is rejected which implies that academic performance is dependent on the profile variables. The Chi square value of 153.4 and Mantel-Haenszel Chi square value of 153.3 will produce moderate phi coefficient to indicate a moderate relationship between these two profile variables. The phi coefficient, however, on a scale of zero to one indicated that the dependency of these variables is moderate. Conclusively,

the analysis revealed that (male & female) students prefer School hostel over off school hostel due to the closeness to their classes, safe time and reduced cost of transportation.

REFERENCES

Agresti, A., 2002, Categorical data analysis (John Wiley & Sons, INC., Hoboken, New Jersey, Hoboken).

Albrecht, C., 2013, Contingency table and the chi square statistics: Interpreting computer printouts and constructing tables, *http://extension.usu.edu/evaluation/files /uploads/Start%20Your%20Engine/Study%20the%20Route/Analyze%20the%20Data /Interpreting_Chi_Square_Printouts.pdf*.

Andel, J., 2002, Zaklady matematicke statistiky, *MFF UK Praha*, 283-317.

Barak, B., Chaudhuri, K., Dwork, C.,Kale, S., McSherry, F.,and Talwar, K., 2007, Privacy, accuracy and consistency too. A holistic solution to contingency table release (http://cseweb.ucsd.edu/~kamalika/pubs/bcdkmt07.pdf, ACM 978-1-59593-685-01/07/0006, Beijing, China).

Gomez and Perez, 2005, Bayesian analysis of contingency tables, *Communication in Statistics - Theory and Methods* 34, 1743-1754.

Howell, D. C., "chi-square test- analysis of contingency tables, "women, vol 35, no 3, 2009, pp 28-83

Hsiao-Mei, wang (2009) *Comparison of the Goodness-of-Fit Tests: The Pearson Chi-square and Kolmogorov - Smirnov Tests*. Available at: http://joqm.ctu.edu.tw/Download/joqm/QM6-1/QM-0601-05paper_proof.pdf (Accessed: 27 November 2016).

Ingersoll, G. M., 2010. "Analysis of 2x2 contingency tables in educational research and evaluation."

International Journal for Research in Education, vol. 27, pp. 1-14. Available: http://www.cedu.uaeu.ac.ae/journal/issue27/ch5_27en.pdf

Lauritzen, S. L., 2002, Lectures on contingency tables, *http://www.stats.ox.ac.uk/~steffen/papers/cont.pdf*.

Marie D. W. (2008). Use of Chi-Square Statistic, Retrieved from http://ocw.jhsph.edu/courses/FundEpiII/PDFs/Lecture17.pdf(Accessed: 27 November 2016)

Nduka, E.C. (2015) *Statistics Concepts and Methods*. 3rd edn. Nigeria: Clara's Prints. In-line Citation:(Nduka, 2015)

Olmus and Erbas 2012, Analysis of traffic accidents caused by drivers by using log-linear model, *Promet- Traffic and Transportation* 24, 495-504.

Pearson, K. (1900). On the 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. *Philosophical Magazine Series 5, 50* (302), 157–175.

Pearson, K., 1904, On the theory of contingency and its relation to association and normal correlation, the analysis of contingency table. Forward to Draper company research memoirs, *Biomteric*.

Utts, J. M., Heckard, R.F., 2006, Mind on Statistics (Cengage Learning, Belmont, CA.

Walpole, R. E., Myers, R.H., Myers, S.L., Ye, K.E., 2011, Probability and Statistics for Engineers and Scientists (Pearson, New York)