Modeling HIVAIDS Variables, A Case Of Contingency Analysis

O. C. Asogwa.

Department of Mathematics/ Statistics/ Computer Science & Informatics,
Federal University Ndufu Alike-Ikwo, Ebonyi State.
qackasoo@yahoo.com

Abstract

Hypothesis testing may be performed on contingency tables in order to decide whether or not effects are present. Effects in a contingency table are defined as relationships or interactions between variables of interest. However, considering the number of adolescence patients diagnosed of HIV/AIDS within a short frame of time in BSH, Nsukka, the researcher was moved to carry out a study to actually confirm whether patients' diagnosed of the aliment have any association based on their gender, years and age of diagnoses. The observation from the hypotheses carried out on this study using χ^2 *statistic, was that none of the variables of interest considered were dependent of each other. Gender* (X) *and age* (Y) *of patients diagnosed of HIV/AIDS in the Bishop Shanahan Missionary Hospital, Nsukka* (B.S.H) *are independent of the years* (Z) *of affection. Also the conditional independence of the pair-wise variables of interest existed. Age* (Y) *is conditionally independent of both gender* (X) *of the affected patient and the year* (Z) *of the affection. Moreover, gender* (X) *is conditionally unrelated to both the year* (Z) *of affection and the age* (Y) *of the affected patients and lastly, year* (Z) *of affection has no conditional interaction with both the gender* (X) *and age* (Y) *of the affected patients.*

Keywords: Modeling, Contingency, Independency, Chi-square, HIV/AIDS

1.0 Introduction

In statistics, a contingency table (also referred to as cross tabulation or crosstab) is a type of table in a matrix format that displays the (multivariate) frequency distribution of the variables. They are heavily used in survey research, business intelligence, engineering and scientific research. They provide a basic picture of the interrelation between two variables and can help find interactions or associations between them. The term contingency table was first used by [1] in his research "On the Theory of Contingency and Its Relation to Association and Normal Correlation.

Moreover, according to [23], the analysis of contingency tables with multi-way classifications originates from the historical development of statistical inference with 2 × 2 tables. In the initial extension to the case of 2 × 2 × *K* tables, [1] discussed testing for three-way interaction and derived an estimate of the common odds ration suggested by R. A. Fisher. [20] and [27] supplied interpretations of varied interactions which led to the well-known Simpson's paradox [3]. [26] showed that Bartlett's procedure is an implicit maximum likelihood estimation (MLE), conditioned upon the fixed margins of each 2×2 table. The celebrated analysis of variance (ANOVA, [8]) inspired discussions of partitioning chi-squares within the contingency tables, notably by [16], [19], and [4], among others. In related research in biostatistics, [5], [32] and [18] developed chi-square tests for no association between two variables across levels of the third variable. These early studies led to further analyses of three-way tables, which include estimating the common odds ratio, testing zero interaction and testing no association across strata, for examples, [15], [10], [6], [17], [24], [2[a]],[2[b]], and [11].

The classical method of partitioning the chi-squares for the three-way $(I \times J \times K)$ Contingency table does not provide a convenient test of the null hypothesis that the three-way interaction is zero [16]. The null distribution of Lancaster's test statistic need not be asymptotically Chi-square distributed [24]. A similar remark can also be applied to the methods of partitioning Chi-squares by [15] and [4]. Since partitions of chi-squares are closely related to the likelihood ratio tests [31] maximum likelihood estimation of association and interaction significantly influenced the early studies of multi-way contingency tables [26]. In particular, it led to the development of a likelihood ratio test for zero interaction in a multi-way contingency table [6]. [30] used chi square test to determine if there were pre-post difference among a group of students in an ethnic literature classroom. Also in a study by [28] a Chi-square test was used to identify the difference between categories of data as applied on public administration research. For further literatures and brief history see ([29], [25], [14], [9]).

Hoverer, the goal of this study is to verify whether the diagnosed HIV/AIDS positive adolescence patients of the case study has a direct or conditional interactions among the years of affection, gender of the patients and the ages of the patients.

2.0 **Material And Methods**

Contingency Analysis otherwise known as Categorical Data Analysis constituted the framework of this research and the statistical tools adopted by [22] – Chi-squared test (χ^2) aided in the analysis of this study. The table comprised of the adolescence aged patients who were diagnosed positive to Human Immune Virus/Acquired Immune Deficiency Syndrome (HIV/AIDS), from 2011 to 2014. Moreover, the data which were collected from the laboratory section of Bishop Shanahan Missionary Hospital Nsukka, were partitioned into two main adolescence age categories (12 - 14 & 15 – 18), and each of the years considered in this study, data collection were from January to December. Also data were collected on both genders. The data were compiled in a categorical table otherwise contingency table in other to capturing the study objectives

2. 1 **Layout of the Study Design (Multi-way (2×2×4) Contingency Table)**

Table 1: Layout of the study design

	Age (Y)										
	y_1 (12-14)					y_2 (15-18)					
	Year (Z)										
Gender(X)	2011(z_1)	2012(z_2)	2013(z_3)	2014(z_4)	n_{+j+}	2011(z_1)	2012(z_2)	2013(z_3)	2014(z_4)	n_{+j+}	n_{i++}
Male(x_1)	$n_{111}(\pi_{111})$	$n_{112}(\pi_{112})$	$n_{113}(\pi_{113})$	$n_{114}(\pi_{114})$	$n_{11+}(\pi_{11+})$	$n_{121}(\pi_{121})$	$n_{122}(\pi_{122})$	$n_{123}(\pi_{123})$	$n_{124}(\pi_{124})$	$n_{12+}(\pi_{12+})$	$n_{1++}(\pi_{1++})$
Female(x_2)	$n_{211}(\pi_{211})$	$n_{212}(\pi_{212})$	$n_{213}(\pi_{213})$	$n_{214}(\pi_{214})$	$n_{21+}(\pi_{21+})$	$n_{221}(\pi_{221})$	$n_{222}(\pi_{222})$	$n_{223}(\pi_{223})$	$n_{224}(\pi_{224})$	$n_{22+}(\pi_{22+})$	$n_{2++}(\pi_{2++})$
n_{++k}	$n_{+11}(\pi_{+11})$	$n_{+12}(\pi_{+12})$	$n_{+13}(\pi_{+13})$	$n_{+14}(\pi_{+14})$	$n_{+1+}(\pi_{+1+})$	$n_{+21}(\pi_{+21})$	$n_{+22}(\pi_{+22})$	$n_{+23}(\pi_{+23})$	$n_{+24}(\pi_{+24})$	$n_{+2+}(\pi_{+2+})$	$n = n_{+++}$

From the table 1 above, n = $n_{+++} = \sum_{i=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{u} n_{ijk}$; $i = 1,2;\ j = 1,2;\ k = 1,2,3,4.$

$i = 1,\dots,r;\ j = 1,\dots,m;\ k = 1,\dots,u$.

$n_{i++} = \sum_{j}^{m}\sum_{k}^{u} n_{ijk} \Rightarrow n_{1++} = (n_{11+} + n_{12+}),\ n_{2++} = (n_{21+} + n_{22+})$

$n_{+j+} = \sum_{i}^{r}\sum_{k}^{u} n_{ijk} \Rightarrow n_{+1+} = (n_{11+} + n_{21+}),\ n_{+2+} = (n_{12+} + n_{22+})$

$n_{++k} = \sum_{i}^{r}\sum_{j}^{m} n_{ijk} \Rightarrow n_{++1} = (n_{+11} + n_{+21}),\ n_{++2} = (n_{+12} + n_{+22}),\ n_{++3} = (n_{+13} + n_{+23}),\ n_{++4} = (n_{+14} + n_{+24})$

2. 2 The Chi-Square Distribution

Chi-square distribution which was first discovered by the German Statistician Friedrich Robert Helmert in papers of 1875 - 6 ([12], [13]) is a theoretical or mathematical distribution which has wide applicability in statistical work. The term 'Chi-square' (pronounced with a hard `Ch') is used because the Greek letter χ is used to define this distribution. It was observed that the elements on which this distribution is based are squared, so that the symbol χ^2 is used to denote the distribution.

However, in probability theory, as well as in Statistics, the Chi-squared distribution with degree of freedom k, denoted by χ_k^2 is the distributed of sum of squares of k independent standard normal random variables. The Chi-square distribution is used in the common chi square tests for goodness of fit of observed distribution to a theoretical one, the independence of two criteria of classification of quantitative data, and in confidence interval estimation for a population standard deviation of a normal distribution from a sample standard deviation. Some other statistical tests also use this distribution, like Friedman's analysis of variance by rank. Also Chi-square distribution is very important because many test statistics are approximately distributed as Chi-square. Two of the more common tests using the Chi-square distribution are tests of deviations of differences between theoretically expected and observation frequencies (multi-way tables) and the relationship between categorical variable (contingency tables).

Moreover, the distribution was independently rediscovered by the English Mathematician Karl Pearson in the context of goodness of fit, for which he developed his Pearson's Chi-square test, published in 1900, with computed table of values published in [7]. If $Z_1, Z_2, Z_3, \dots, Z_k$ are independent, standard normal random variables, then the sum of their squares is distributed according to the Chi-squared distribution with k degrees of freedom. This is usually represented as, $Q = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$, bearing in mind that the chi-squared distribution has one parameter; k: a positive integer that specifies the number of degrees of freedom

.

.

2.3 Derivation of the Distribution

In mathematical terms, the χ^2 variable is the sum of the squares of a set of normally distributed variables. Given a standardized normal distribution for a variable *Q* with mean 0 and standard deviation, and suppose that a particular value Q_1 is randomly selected from this distribution. Then suppose another value Q_2 is selected from the same standardized normal distribution. If there are *t* degrees of freedom, then let this process continue until *t* different *Q* values are selected from this distribution. The χ^2 variable is defined as the sum of the squares of these *Q values*. That is

$$\theta = Q_1^2 + Q_2^2 + \dots + Q_t^2 \sim \chi_t^2$$

Suppose this process of independent selection of t different values is repeated many times. The variable θ will vary, because each random selection from the normal distribution will be different. Note that this variable is a continuous variable since each of the Q values is continuous. This sum of squares of t normally distributed variables has a distribution which is called the θ distribution with t degrees of freedom. It can be shown mathematically that the mean of this θ distribution is t and the standard deviation is $2t$. An intuitive idea of the general shape of the distribution can also be obtained by considering this sum of squares. Since θ is the sum of a set of squared values, it can never be negative. The minimum Chi-squared value would be obtained if each $Q = 0$ so that θ would also be 0. There is no upper limit to the θ value. If all the Q values were quite large, then θ would also be large. But note that this is not too likely to happen. Since large Q values are distant from the mean of a normal distribution, these large θ values have relatively low probabilities of occurrence, also implying that the probability of obtaining a large θ value is low. Also note that as there are more degrees of freedom, there are more squared Q values, and this means larger θ values. Further, since most values under the normal curve are quite close to the center of the distribution, within 1 or 2 standard deviations of center, the values of θ also tend to be concentrated around the mean of the θ distribution

2.4 **Chi-Square Test for Independence**

The Chi-square test for independence of two variables is a test which uses a cross classification table to examine the nature of the relationship between these variables. These tables are sometimes referred to as contingency tables. These tables show the manner in which two variables are either related or are not related to each other. The test for independence examines whether the observed pattern between the variables in the table is strong enough to show that the two variables are dependent on each other or not. While the Chi-square statistic and distribution are used in this test, the test is quite distinct from the test of goodness of fit. The goodness of fit test examines only one variable, while the test of independence is concerned with the relationship between two variables.

Like the goodness of fit test, the Chi-square test of independence is very general, and can be used with variables measured on any type of scale, nominal, ordinal, interval or ratio. The only limitation on the use of this test is that the sample sizes must be sufficiently large to ensure that the expected number of cases in each category is positive or more. This rule can be modified somewhat, but as with all approximations, larger sample sizes are preferable to smaller sample sizes. There are no other limitations on the use of the test, and the Chi-square statistic can be used to test any contingency or cross classification table for independence of the two variables. The Chi-square test for independence is conducted by assuming that there is no relationship between the two variables being examined. The alternative hypothesis is that there is some relationship between the variables

2.5 **The Chi-Square Statistic**

The χ^2 Statistic appears quite different from the other Statistics which have been used in the previous hypotheses tests. For both the goodness of fit test and the test of independence, the Chi- square statistic is the same. For both of these tests, all the categories into which the data have been divided are used. The data obtained from the sample are referred to as the observed numbers of cases. These are the frequencies of occurrence for each category into which the data have been grouped.

In the Chi-square tests, the null hypothesis makes a statement concerning how many cases are to be expected in each category if this hypothesis is correct. The chi square test is based on the difference between the observed and the expected values for each category.

The Chi-square Statistic is defined as

$$\chi^2 = \sum_i \left(\frac{n_{ij} - f_{ij}}{f_{ij}}\right)^2$$

where n_{ij} is the observed number of cases in category i and f_{ij} is the expected number of cases in category j. This Chi-square Statistic is obtained by calculating the difference between the observed number of cases and the expected number of cases in each category. This difference is squared and divided by the expected number of cases in that category. These values are then added for all the categories, and the total is referred to as the Chi-squared value.

The null hypothesis is a particular claim concerning how the data is distributed. The null and alternative hypotheses for each Chi-square test can be stated as

H_0 : $n_{ij} = f_{ij}$ (the observed and the expected values are independent or the same)

H_1 : $n_{ij} \neq f_{ij}$ (the observed and the expected values are dependent or not the same)

However, if the claim made in the null hypothesis is true, the observed and the expected values are close to each other and $n_{ij} - f_{ij}$,is small for each category. When the observed data does not conform to what has been expected on the basis of the null hypothesis, the difference between the observed and expected values, $n_{ij} - f_{ij}$, is large. The Chi-square Statistic is thus small when the null hypothesis is true and large when the null hypothesis is not true. Exactly how large the χ^2 value must be in order to be considered large enough to reject the null hypothesis can be determined from the level of significance and the Chi-square table in any statistical textbook. A general formula for determining the degrees of freedom is not given at this stage, because this differs for the two types of chi square tests. In each type of test though, the degrees of freedom is based on the number of categories which are used in the calculation of the statistic.

The Chi-square Statistic, along with the Chi-square distribution, allows the researcher to determine whether the data is distributed as claimed. If the Chi-square statistic is large enough to reject H_0, then the sample provides evidence that the distribution is not as claimed in H_0. If the chi square statistic is not so large, then the researcher may have insufficient evidence to reject the claim made in the null hypothesis. It is equally used to investigate whether distributions of categorical variables differ from one another. The chi-square statistic compares the tallies or counts of categorical responses between two or more independent groups

.

2.0 Layout Presentation of Data

Table 2: Table of the observed and the expected values of the HIV positive persons as categorized

	Age(Y)										
	y_1(12 – 14)					y_2(15 – 18)					
	Year(Z)										
Gender(X)	(z_1)	(z_2)	(z_3)	(z_4)	n_{+j+}	(z_1)	(z_2)	(z_3)	(z_4)	n_{+j+}	n_{i++}
Male(x1)	20(16)	58(61)	63(68)	38 (42)	179	48 (38)	144 (149)	172 (165)	99 (103)	463	642
Female(x2)	33(30)	115(120)	127(133)	100(83)	375	57(74)	306(292)	328(324)	193(202)	884	1,259
n_{++k}	53	173	190	138	554	105	450	500	292	1,347	**1,901**

X = Gender : Male (x_1); Female (x_2)
Y = Age : 12 – 14 (y_1); 15 – 18(y_2)
Z = Year : 2011 (z_1); 2012 (z_2); 2013 (z_3); 2014 (z_4)
$n = n_{+++} = \sum_{i=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{u} n_{ijk} = 1{,}901$

3.1. Hypotheses to be tested

(a)Test of interactions among gender (X) age(Y) and year (Z)

(b)Test of conditional independence of gender (X) on both age (Y) and year (Z)

(c)To access the conditional interaction of age (Y) on both gender (X) and year (Z)

(d)To test for the conditional independence of year (Z) given both gender (X) and age (Y)

3.1.1. Hypothesis (a)

$H_{o(a)}$: There is no interaction among gender (X), age (Y) and year (Z)

$H_{1(a)}$: There is interaction among gender (X), age (Y) and year (Z)

at level of significance $\alpha = 0.01$

Theoretically, from above, the hypothesis implies;

$H_{o(a)} : \pi_{ijk} = P(\pi_{i++} \cap \pi_{+j+} \cap \pi_{++k}) = P(\pi_{i++}).P(\pi_{+j+}).P(\pi_{++k})$

$H_{1(a)} : \pi_{ijk} \neq P(\pi_{i++} \cap \pi_{+j+} \cap \pi_{++k}) \neq P(\pi_{i++}).P(\pi_{+j+}).P(\pi_{++k})$

where $\pi_{ijk} = P(\pi_{i++}).P(\pi_{+j+}).P(\pi_{++k}) = \left(\frac{n_{i++}}{n}\right).\left(\frac{n_{+j+}}{n}\right).\left(\frac{n_{++k}}{n}\right)$

also, $\hat{f}_{ijk} = n.\pi_{ijk}$,

$$\hat{f}_{ijk} = n\left(\frac{n_{i++}}{n}\right).\left(\frac{n_{+j+}}{n}\right).\left(\frac{n_{++k}}{n}\right) = \frac{n}{n^3}(n_{i++}).(n_{+j+}).(n_{++k})$$

Test Statistic: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{u}\frac{(n_{ijk}-\hat{f}_{ijk})^2}{\hat{f}_{ijk}} \sim \chi^{2\ (\alpha)}_{(ijk)-(i+j+k)+2}$

where $i = 1,2,\ldots,r\,; j = 1,2,\ldots,m;\ k = 1,2,\ldots,u$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{4}\frac{(n_{ijk}-\hat{f}_{ijk})^2}{\hat{f}_{ijk}} = \left(\frac{20-16}{16}\right)^2 + \left(\frac{58-61}{61}\right)^2 + \ldots + \left(\frac{193-202}{202}\right)^2$$

$\chi^2 = 1.0000 + 0.1475 + \cdots + 0.4010 = 14.4357$

therefore, $\chi^2_{cal} = 14.44$

$\chi^{2\ (\alpha)}_{(ijk)-(i+j+k)+2} = \chi^{2\ (0.01)}_{(2\times2\times4)-(2+2+4)+2} = \chi^2_{10,\ 0.01} = 23.21$

therefore, $\chi^2_{tab} = 23.21$

Decision Rule: Reject $H_{o(a)}$ if $\chi^2_{cal} > \chi^2_{tab}$, accept if otherwise.

Conclusion: Since $\chi^2_{cal} = 14.44 < \chi^2_{tab} = 23.21$, $H_{o(a)}$ was accepted at $\alpha = 0.01$

3.1.2. Hypothesis (b)

$H_{o(a)}$: Gender (X) is conditionally independent of both age (Y) and year (Z)

$H_{1(a)}$: Gender (X) is conditionally dependent of both age (Y) and year (Z)

at level of significance $\alpha = 0.01$

Theoretically, from above, the hypothesis implies;

$H_{o(b)} : \pi_{ijk} = P(\pi_{i++}).P(\pi_{+j+} \cap \pi_{++k})$

$H_{1(b)} : \pi_{ijk} \neq P(\pi_{i++}).P(\pi_{+j+} \cap \pi_{++k})$

where $P(\pi_{+j+} \cap \pi_{++k}) = P(\pi_{+j+}).P(\pi_{++k}/\pi_{+j+})$

$\pi_{ijk} = P(\pi_{i++}).P(\pi_{+j+} \cap \pi_{++k}) = \left(\frac{n_{i++}}{n}\right).\left(\frac{n_{+jk}}{n}\right)$

also, $\hat{f}_{ijk} = n.\pi_{ijk}$,

$\hat{f}_{ijk} = n\left(\frac{n_{i++}}{n}\right).\left(\frac{n_{+jk}}{n}\right) = \frac{n}{n^2}(n_{i++}).(n_{+jk})$

Test Statistic: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{u}\frac{(n_{ijk}-\hat{f}_{ijk})^2}{\hat{f}_{ijk}} \sim \chi^{2\ (\alpha)}_{(i-1)(jk-1)}$

where $i = 1,2,\dots,r\,; j = 1,2,\dots,m;\ k = 1,2,\dots,u$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{4}\frac{(n_{ijk}-\hat{f}_{ijk})^2}{\hat{f}_{ijk}} = \left(\frac{20-16}{16}\right)^2 + \left(\frac{58-61}{61}\right)^2 + \dots + \left(\frac{193-202}{202}\right)^2$$

$\chi^2 = 1.0000 + 0.1475 + \cdots + 0.4010 = 14.4357$

therefore, $\chi^2_{cal} = 14.44$

$\chi^{2\ (\alpha)}_{(i-1)(jk-1)} = \chi^{2\ (0.01)}_{(2-1)-((2\times 4)-1)} = \chi^2_{7,\ 0.01} = 18.48$

therefore, $\chi^2_{tab} = 18.48$

Decision Rule: Reject $H_{o(b)}$ if $\chi^2_{cal} > \chi^2_{tab}$, accept if otherwise.

Conclusion: Since $\chi^2_{cal} = 14.44 < \chi^2_{tab} = 18.48$, $H_{o(b)}$ was accepted at $\alpha = 0.01$

3.1.3. Hypothesis (c)

$H_{o(c)}$: Age (Y) is conditionally not interacted with both gender (X) and year (Z)

$H_{1(c)}$: Age (Y) is conditionally interacted with both gender (X) and year (Z)

at level of significance $\alpha = 0.01$

Theoretically, from above, the hypothesis implies;

$H_{o(c)} : \pi_{ijk} = P(\pi_{+j+}).P(\pi_{i++} \cap \pi_{++k})$

$H_{1(c)} : \pi_{ijk} \neq P(\pi_{+j+}).P(\pi_{i++} \cap \pi_{++k})$

where $P(\pi_{i++} \cap \pi_{++k}) = P(\pi_{i++}).P(\pi_{++k}/\pi_{i++})$

$\pi_{ijk} = P(\pi_{+j+}).P(\pi_{i++} \cap \pi_{++k}) = \left(\frac{n_{+j+}}{n}\right).\left(\frac{n_{i+k}}{n}\right)$

also, $\hat{f}_{ijk} = n.\pi_{ijk}$,

$\hat{f}_{ijk} = n\left(\frac{n_{+j+}}{n}\right).\left(\frac{n_{i+k}}{n}\right) = \frac{n}{n^2}(n_{+j+}).(n_{i+k})$

Test Statistic: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{u}\frac{(n_{ijk}-\hat{f}_{ijk})^2}{\hat{f}_{ijk}} \sim \chi^{2\ (\alpha)}_{(j-1)(ik-1)}$

where $i = 1,2,\dots,r\ ; j = 1,2,\dots,m;\ k = 1,2,\dots,u$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{4}\frac{(n_{ijk}-\hat{f}_{ijk})^2}{\hat{f}_{ijk}} = \left(\frac{20-16}{16}\right)^2 + \left(\frac{58-61}{61}\right)^2 + \dots + \left(\frac{193-202}{202}\right)^2$$

$\chi^2 = 1.0000 + 0.1475 + \cdots + 0.4010 = 14.4357$

therefore, $\chi^2_{cal} = 14.44$

$\chi^{2\ (\alpha)}_{(j-1)(ik-1)} = \chi^{2\ (0.01)}_{(2-1)-((2\times4)-1)} = \chi^2_{7,\ 0.01} = 18.48$

therefore, $\chi^2_{tab} = 18.48$

Decision Rule: Reject $H_{o(c)}$ if $\chi^2_{cal} > \chi^2_{tab}$, accept if otherwise.

Conclusion: Since $\chi^2_{cal} = 14.44 < \chi^2_{tab} = 18.48$, $H_{o(c)}$ was accepted at $\alpha = 0.01$

3.1.4. **Hypothesis (d)**

$H_{o(d)}$: Year (Z) is conditionally independent of both gender (X) and age (Y)

$H_{1(d)}$: Year (Z) is conditionally independent of both gender (X) and age (Y)

at level of significance $\alpha = 0.01$

Theoretically, from above, the hypothesis implies;

$H_{o(d)} : \pi_{ijk} = P(\pi_{++k}).P(\pi_{i++} \cap \pi_{+j+})$

$H_{1(d)} : \pi_{ijk} \neq P(\pi_{++k}).P(\pi_{i++} \cap \pi_{+j+})$

where $P(\pi_{i++} \cap \pi_{+j+}) = P(\pi_{i++}).P(\pi_{+j+}/\pi_{i++})$

$\pi_{ijk} = P(\pi_{++k}).P(\pi_{i++} \cap \pi_{+j+}) = \left(\frac{n_{++k}}{n}\right).\left(\frac{n_{ij+}}{n}\right)$

also, $\hat{f}_{ijk} = n.\pi_{ijk}$,

$\hat{f}_{ijk} = n\left(\frac{n_{++k}}{n}\right).\left(\frac{n_{ij+}}{n}\right) = \frac{n}{n^2}(n_{++k}).(n_{ij+})$

Test Statistic: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{u}\frac{(n_{ijk}-\hat{f}_{ijk})^2}{\hat{f}_{ijk}} \sim \chi^{2\ (\alpha)}_{(k-1)(ij-1)}$

where $i = 1,2,\dots,r\ ; j = 1,2,\dots,m;\ k = 1,2,\dots,u$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{4}\frac{(n_{ijk}-\hat{f}_{ijk})^2}{\hat{f}_{ijk}} = \left(\frac{20-16}{16}\right)^2 + \left(\frac{58-61}{61}\right)^2 + \dots + \left(\frac{193-202}{202}\right)^2$$

$\chi^2 = 1.0000 + 0.1475 + \cdots + 0.4010 = 14.4357$

therefore, $\chi^2_{cal} = 14.44$

$\chi^{2\ (\alpha)}_{(k-1)(ij-1)} = \chi^{2\ (0.01)}_{(4-1)-((2\times2)-1)} = \chi^2_{9,\ 0.01} = 21.67$

therefore, $\chi^2_{tab} = 21.67$

Decision Rule: Reject $H_{o(d)}$ if $\chi^2_{cal} > \chi^2_{tab}$, accept if otherwise.

Conclusion: Since $\chi^2_{cal} = 14.44 < \chi^2_{tab} = 21.67$, $H_{o(d)}$ was accepted at $\alpha = 0.01$

4.0. Abridged Table of Results

Table 3: Summarize table showing the tested hypotheses and evaluated $\chi^2\ at\ \alpha = 0.01$, where (X) stands for Gender, Age for (Y) and (Z) stands for Year

Hypotheses	χ^2 – Evaluations	*Remarks*
(a) Test of interaction among X, Y and Z	$\chi^2_{cal} = 14.44 < \chi^2_{tab} = 23.21$	Accept $H_{0(a)}$
(b) Test of conditional independence of X given Y and Z	$\chi^2_{cal} = 14.44 < \chi^2_{tab} = 18.48$	Accept $H_{0(b)}$
(c) Test of conditional independence of Y given X and Z	$\chi^2_{cal} = 14.44 < \chi^2_{tab} = 18.48$	Accept $H_{0(c)}$
(d) Test of conditional independence of Z given X and Y	$\chi^2_{cal} = 14.44 < \chi^2_{tab} = 21.67$	Accept $H_{0(d)}$

.

5.0. Conclusion

From table 3, it could be verified that the study hypotheses were accomplished and none of the hypotheses tested yielded an interaction effects among each of the interaction or association effects considered. There were no interaction among years of affection, gender of the affected and the age of the patients of adolescence persons diagnosed of the aliment of the case study. Also the pair-wise conditional interactive effects were not met in this study. Therefore from the final remark, the test of dependence effect of the variables considered in this study were not met, the pair-wise interactions also failed. So the gender, year and age of the affected adolescence patients of the case study were independent of one another and their pair-wise conditional effect were equally independent of each other.

ACKNOWLEDGEMENT

The author is grateful to the Bishop Shanahan Missionary Hospital (BSH) Nsukka, Enugu State, Nigeria, for making the data for this research available.

References

[1] Bartlett, M. S. (1935). Contingency table interactions. *Supplement to the journal of the Royal Statistical* 2 (**2**), 248–252.

[2] Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society* 25, 220–233.

[3] Blyth, C. R. (1972). On Simpson's paradox and the sure thing principle. *Journal of American Statistics Association.* 67, 364-366.

[4] Claringbold, P. J. (1961). The use of orthogonal polynomials in the partition of chi-square. *Australia Journal of Statistics.* 3, 48-63.

[5] Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* 10, 417–451.

[6] Darroch, J. N (1962). Interaction in multi-factor contingency tables. *Journal of the Royal Statistical Society* 24, 251–263.

[7] Elderton, William Palin (1902). "Tables for testing the goodness of fit of theory to observation" Biometrika 1(**2**): 155-163.

[8] Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables and the calculation of p. *Journal of the Royal Statistical Society* 85(**1**), 87–94.

[9] Galton, F. (1892). *Finger Prints*. London: Macmillan.

[10] Gart, J. J. (1962). On the combination of relative risks. *Biometrics* 18, 601-610.

[11] Goodman, L. A. (1964). Simultaneous confidence limits for cross-products ratios in tables. *Journal of the Royal Statistical Society* 26, 86–102.
[12] Hald, A. (1998). A history of Mathematical Statistics from 1750–1930. New York Wiley. ISBN 0–471–17912–4.
[13] Helmert, F. R. (1876). Ueber die Wahrscheinlichk der potenzsummen der Beobachtungsfehler and ubereinige damit im zusammenhange stehende fragen, zeitschrift fur mathamatik und physic 21, 1876, S. 102–219.
[14] Heyde, C. C. and Seneta, E. (1977). I. J. Bienayme; *Statistical theory Anticipated*. New York: Springer–Verley.
[15] Kullback, S. (1959). *Information Theory and Statistics.* Wiley.
[16] Lancaster, H. O. (1951). Complex contingency tables treated by the partition of chi-square. *Journal of Royal Statistical Society, Series, B* 13, 242-249.
[17] Lewis, B. N. (1962). On the Analysis of Interaction in Multi-Dimensional contingency. *Journal of the Royal Statistical Society Series.*
[18] Mantel, N. and Hacnszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22, 719–748.
[19] Mood, A. M. (1950). Introduction to the theory of *Statistics*. New York: McGraw Hill.
[20] Norton, H. W. (1945). Calculation of chi – square for complex contingency tables. *Journal of the American Statistical Association* 40(**230**), 251–258.
[21] Pearson, K. (1904). Mathematical contributions to the theory of evolution, XIII: On the theory of contingency and its relation to association and normal correlation. Draper's Co. *Research memoirs in biometry series*. I, 1–35.
[22] Pearson, K. (1990). On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical magazine* 50, 59-73.
[23] Philip E. Cheng, Jiun W. Liou, Michelle Liou and John A. D. Aston. (2006). Data Information in Contingency Tables: A Fallacy of Hierarchical Log linear model. *Journal of Data Sciences* 4, 387–398.
[24] Placket, R. L. (1963). "Karl Pearson and the Chi-square Test". *International Statistical Review* (International Statistical Institute ((ISI)) 51(**1**): 59–72. JSTOR.
[25] Quetelet, M. (1849). Letter Addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha on the theory of probabilities as Applied to the moral and political sciences (translated from the French by Olinthus Gregory Downs). London: Charles and Edwin Layton.
[26] Roy, S. N. and Kastenbaum, M. A. (1956). On the hypothesis of no "interaction in a multi-way contingency tables. *Annals of Mathematical Statistics* 27(**3**), 749–757.
[27] Simpson, E. H (1951). The interpretation of interaction in Contingency tables. *Journal of the Royal Statistical Society*. JSTOR.
[28] Stallings, R. A and Ferris, J. M. (1988). Public Administration Research: work in PAR; 1940 – 1984, *Public Review*, Vol. 48 (**1**), pp. 580–587.
[29] Stigler, S. (2002). The missing early history of contingency tables. *Annales de la Faculte' des sciences de Toulouse* II (**4**), 563–573.
[30] Trzyna, T. and Abbott, M. (1991). "Grieving in the Ethnic Literature Classroom," *College Literature* 18(**3**), 1–2.
[31] Wilks, S. S. (1935). The likelihood test of independence in contingency tables. *Annals of Mathematical Statistics* 6, 190–196.
[32] Woolf, B. (1955). On estimating the relation between blood group and disease. *Annals of Human Genetics*, Vol. 19(**4**), 251–253.