THE KARL PEARSON'S CHI-SQUARE: A MEDICAL RESEARCH LIBERO, AND A VERSATILE TEST STATISTIC: AN EDITORIAL.

K.I Egbuchulem

Division of Paediatric Surgery, Department of Surgery, University College Hospital, Ibadan

INTRODUCTION

The logic of hypothesis testing was first scripted by Karl Pearson (1857–1936), a renaissance scientist, in Victorian London, and in 1900 he published his manuscript in the philosophical magazine to elaborate the invention of Chi square distribution and goodness of fit test.[1]

Pearson's Chi square distribution and the Chi square test, also known as test for goodness of fit, and test of independence, are his most important contribution to the modern theory of statistics since the early 20th century.[1]

The Chi-square test of independence (also known as the Karl Pearson Chi-square test, or simply the Chi-square) is one of the most useful statistics for testing hypotheses when the variables are nominal, as it often happens in clinical research.[1]

The Chi-square test is based on the Chi-square (χ^2) distribution. which is roughly skewed, and it is a normal approximation of a binomial distribution as against the Fischer's exact test, which is more exact.[2] Chi square distribution tables are available from a variety of sources they can easily be found online or in scientific documents, and statistics textbooks.[2]

Also, the concept of degree of freedom is also employed in calculating χ^2. The Chi-square test is a significance test statistic, and should be followed with a strength statistic such as the Confidence interval, Cramer's V when a significant Chi-square result has been obtained.[4]

The Chi-square is a non-parametric test statistic, also called a distribution-free test designed to analyse group differences when the dependent variable is measured at a nominal level. Like all non-parametric statistics, the Chi-square is robust with respect to the distribution of the data and the parametric equivalent is the Z-test, which is suited for smaller sample size less than 30. The Chi-square test does not require equality of variances among the study groups as seen in analysis of variance (ANOVA).

Unlike most statistics, the Chi-square (χ^2) permits evaluation of both dichotomous independent variables, and multiple group studies and can provide considerable information about how each of the groups performed in the study. It also gives information not only on the significance of any observed differences, but also provides detailed information on exactly which categories account for any differences found.[3] Thus, the amount and richness of the detailed information this statistic can provide renders it one of the most useful tools in the researcher's array of available analytic tools, hence *the libero or midfielder role Chi-square test plays in biomedical research.*

Non-parametric tests like χ^2 are used when any one of the following conditions are met: [4]

1. The level of measurement of all the variables are nominal or ordinal.
2. The sample sizes of the study groups are unequal; for the χ^2, the groups may be of equal size or unequal size whereas some parametric tests require groups of approximately equal size.
3. The original data is measured at an interval or ratio level, but violate any of the following assumptions of a parametric test:
a) The distribution of the data was seriously skewed or kurtotic (parametric tests assume approximately normal distribution of the dependent variable).
b) The data violate the assumptions of equal variance.
c) When the continuous data were collapsed into a small number of categories, and thus the data are no longer interval or ratio.

Applications

1. Chi-square test and its application in the logic of hypothesis testing of no association between two or more groups, population or criteria (that is to check independence or association between two variables).
2. To test how likely the observed distribution of data fits with the distribution that is expected (i.e., to test the goodness of fit also called *single-sample goodness of fit test* or *Pearson's Chi-square of goodness of fit test*). It is used to analyse categorical data (e.g. young or elderly patients, smokers and non-smokers, etc.), but it is not meant to analyse parametric or continuous data (e.g., height measured in meters or weight measured in kg, etc.).

3. Homogeneity tests of Karl Pearson. The only difference between the independence test and homogeneity test is the specification of null hypothesis. The homogeneity test tests the null hypothesis that claims homogeneity or equality based on some attributes. Chi-square test of homogeneity is used to determine if two or more independent sample vary by distributions on a single variable. A common use of this test is to compare two or more groups or conditions on a categorical outcome variable.

Assumptions of the Chi-square Test[5]

As with any statistic, there are requirements for its appropriate use, which are called "assumptions" of the statistic. As with parametric tests, the non-parametric tests, including the χ^2, assume the data were obtained through random selection. However, it is not uncommon to find inferential statistics used when data are from convenience samples rather than random samples. To have confidence in the results when the random sampling assumption is violated, several replication studies should be performed with essentially the same result obtained. The assumptions of the Chi-square test include:

1) The data in the cells should be frequencies or counts of cases rather than percentages or some other transformation of the data.

2) The levels (or categories) of the variables are mutually exclusive. That is, a particular subject fits into one and only one level of each of the variables.

3) Each subject may contribute data to one and only one cell in the χ^2. If, for example, the same subjects are tested over time such that the comparisons are the same, then χ^2 may not be used.

4) The study groups must be independent. This means that a different test must be used if the two groups are related. For example, a different test must be used if the researcher's data consists of paired samples, such as in studies in which a parent is paired with his or her child. McNemar Chi-square test is employed in paired dichotomous dependent variables.

5) There are two variables, and both are measured as categories, usually at the nominal level. However, the data may be ordinal. Interval or ratio data that have been collapsed into ordinal categories may also be used. While Chi-square has no rule about limiting the number of cells (by limiting the number of categories for each variable), a very large number of cells (over 20) can make it difficult to meet assumption 6, and to interpret the meaning of the results.

6) The value of the cells expected should be 5 or more in at least 80% of the cells, and no cell should have a value of less than one. This assumption is most likely to be met if the sample size equals at least the number of cells multiplied by 5. Essentially, this assumption specifies the number of cases (sample size) needed to use the χ^2 for any number of cells in that χ^2. This requirement will be fully explained in the case study.

Case presentation

To illustrate the calculation and interpretation of the χ^2 statistic, the following case example will be used: Example: in a class of 100 students, 10 of the 30 who prefers to participate in board games are males while there are a total of 40 females in the class. What is the difference between the genders in their preference for either board games or field games?

Solution
a. Traditional calculation method

The first step in calculating a χ^2 is to calculate the sum of each row, and the sum of each column.

These sums are called the "marginal values" and there are row marginal values and column marginal values. The marginal values for the case study data are presented in Table 1.

The second step is to calculate the expected values for each cell. In the Chi-square statistic, the "expected" values represent an estimate of how the genders would be distributed with relation to their choice of games. Expected values must reflect both the preference of games in each category and the gender distribution as in Table 2.

Once the cell χ^2 values have been calculated, they are summed to obtain the χ^2 statistic for the table. In this case below, the χ^2 is 12.70 (rounded up). The Chi-square table requires the use of the table's degrees of freedom (df) in order to determine the significance level of the statistic. The degrees of freedom are the maximum number of logically independent values, which may vary in a data sample. The degrees of freedom for a χ^2 table are calculated with the formula: (Number of rows - 1) x (Number of columns - 1).

For example, a 2 x 2 table has 1 df. (2 -1) x (2 -1) = 1. Assuming a χ^2 value of 12.70 with each of the df levels of 1, the significance levels from the table of χ^2 values are: df = 1, $P < 0.05$. Note, as degrees of freedom increases, the P value level becomes less significant, until the χ^2 value of 12.70 is no longer statistically significant at the 0.05 level, because P was greater than 0.05.

This is illustrated below, in a 2 x 2 cross table, placing the observed values of the independent variables on the rows and dependent variables on the columns. The cross table or contingency table is essentially a display

format used to analyse and record the relationship between two or more categorical variables. It is the categorical equivalent of the scatterplot used to analyse the relationship between two continuous variables.

For cell a = 60 X 30/100 = 18

Table 1: Observed values.

Gender\ Games	Board games	Field games	Total
Males	a-10	b-50	60
Females	c-20	d-20	40
Total	30	70	100

The expected values are then calculated as

$$= \frac{\text{Row total x Column total}}{\text{Grand total}}$$

For cell b = 60 X 70/100 = 42
For cell c = 40 X 30/100 = 12
For cell d = 40 X 70/100 = 28
The contingency table for the expected values

That is

Table 2: Expected values.

Gender\ Games	Board games	Field games	Total
Males	a-18	b-42	60
Females	c-12	d-28	40
Total	30	70	100

The calculated Chi-square value for each cell χ^2 is

$$= \frac{(\text{Observed} - \text{Expected})^2}{\text{Exected}}$$

Cell a = $(10\text{-}18)^2/18$ = 3.56
Cell b = $(50\text{-}42)^2/42$ = 1.52
Cell c = $(20\text{-}12)^2/12$ = 5.33
Cell d = $(20\text{-}28)^2/28$ = 2.29
The total Chi-Square value =
$\sum\chi^2 = (O\text{-}E)^2/E$
= 3.56 + 1.52 + 5.33 + 2.29 = 12.77 = Calculated Chi-square.
The degree of freedom (df) = (row - 1) x (column - 1)
For a 2 x 2 table
= (2 - 1) x (2 - 1)
= 1 x 1 = 1.
Using the four figure table at df of 1, = 3.841 = Critical Chi-square value.
Because the calculated Chi-square is greater than the critical Chi-square, it suggests $p < 0.05$, and the null hypothesis will be rejected.
Therefore, there is a statistical difference between gender and preference of games beyond what is attributable to chance.

a. Another way of resolving this is by using a set formula for 2 x 2 tables only
Here the table of observed values are used
= 12.70 = Calculated Chi-square.
The degree of freedom (df) = (row-1) x (column-1)
For a 2 x 2 contingency table
= (2 - 1) x (2 - 1)
=1 x 1 = 1.
Using the statistical tables at df of 1 where it intersects with p value of 0.05, = 3.841 = Critical Chi-square value.

Since the calculated Chi-square value is greater than the critical Chi-square, it suggests $p < 0.05$, and the null hypothesis will be rejected. Each time the null hypothesis is rejected, it is an opportunity to accept the alternate hypothesis.
Therefore, there is a statistical difference between gender and preference of games beyond what is attributable to chance.
Note that there may be real or apparent clinical difference which may or may not be statistically significant.
Statistical significance is calculated or computed as seen above,
A clinically significant, statistically significant difference may as well not have a public health significance or difference.

b. Chi-square can be calculated online, [6] or using tools such as SPSS, STATA, R, et cetera
The Chi-square statistics and its associated *P* value can be calculated through online calculators also which are easily available on the internet. For user-friendly online calculator, you may visit this uniform resource locator www.socscistatistics.com/tests/chisquare/default2.aspx. Many more online calculators are available on the World Wide Web. The basic step for using an online calculator is to correctly fill in your data into it.

The step by step procedure of using an online calculator is described below:
• Step 1: For our example of finding an association between smoking and lung disease, we have to fill in the observed values in the cells of an online calculator.
• Step 2: Click on the next button and another screen will pop up.
• Step 3: Click on the Calculate $\div^2$ button and you are done with your calculation; the output of the Chi-square test will be shown.

The Advantages of the Chi-square test include [6]
1. Its robustness with respect to distribution of the data,

2. Its ease of computation and analysis,
3. The detailed information that can be derived from the test,
4. Its use in studies for which parametric assumptions cannot be met, and
5. Its flexibility in handling data from both two groups and multiple group studies.

Its limitations include
1. Its sample size requirements,
2. The difficulty of interpretation when there are large numbers of categories (20 or more) in the independent or dependent variables.

Chi-square test and closely related non-parametric **tests** [4]
Nominal variables require the use of non-parametric tests, and there are three commonly used significance tests that can be used for this type of data.
1. The first and most commonly used is the Chi-square. Note, the McNemar Chi-square; it is a special form of Chi-square used to test only paired 2 x 2 dependent nominal data and checks the marginal homogeneity of the two dichotomous variables.
2. The second is the Fisher's exact test, which is a bit more precise than the Chi-square, and more suited for 2 x 2 Tables. It is suitable for tables with greater than 20% of cells being less than or equal to 5 or the row or column total less than or equal to 20.
3. The third test is the maximum likelihood ratio Chi-square test which is most often used when the data set is too small to meet the sample size assumption of the Chi-square test.

CONCLUSIONS
The Chi-square is a valuable analytic tool that provides considerable information about the nature of the research data. It is a powerful statistic that enables researchers to test hypotheses about variables measured at the nominal level. The Chi-square is also an excellent tool to use when violations of assumptions of equal variances are present and parametric statistics such as the t-test and ANOVA cannot provide reliable results. One of three outcomes are possible when both the appropriate and inappropriate tests statistics are applied on the same data:
1. First, the appropriate and the inappropriate test may give the same results.
2. Second, the appropriate test may produce a significant result while the inappropriate test provides a result that is not statistically significant, which is a Type II error.
3. Third, the appropriate test may provide a non-significant result while the inappropriate test may provide a significant result, which is a Type I error.

REFERENCES
1. **Pearson K**. 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. Philos Mag Ser 1900; 50:157 175.
2. **Miller R**, Siegmund D. Maximally selected Chi-square statistics. Biometrics 1982; 38:1101-1106.
3. **Streiner D**. Chapter 3: Breaking up is hard to do: The heartbreak of dichotomizing continuous data. In Streiner, D. A Guide for the Statistically Perplexed. Buff alo, NY: University of Toronto Press 2013.
4. **McHugh ML**. The Chi-square test of independence. Biochemia Medica 2013; 23(2):143–149.
5. **Scott M**, Flaherty D, Currall J. Statistics: Dealing with categorical data. J Small Anim Pract 2013; 54:3-8.
6. **Rakesh KR**. Chi-square test and its application in hypothesis testing. Journal of the Practice of Cardiovascular Sciences. 2015; 1(1):69-71.