41-The ability

By Jaka Nugraha
The Ability The Chi Squares Statistics To Rejecting The Null Hypothesis on Contingency Tables 2x2

Jaka Nugraha
Statistics Department, Islamic University of Indonesia, Yogyakarta
jugalha@ui.ac.id

Abstract: The Chi-Square statistic is the primary statistic used for testing the statistical significance of the cross-tabulation table. The Chi-Square test is based on an approximation that works best when the expected frequencies are fairly large. There is still a lack of consensus on the optimum method most texts recommend the use of the chi squared test and there is disagreement on the boundary between 'large' and 'small' sample sizes. So in this paper discussed the limitations and requirements for proper use, especially related to the minimum number of sample requirements for table 2x2.

The study focused on four versions of the Chi-Squared test the tests were: Pearson’s Chi-Squared test, Yates’s Chi-Squared test, Likelihood Ratio Chi-Square test and ‘N - 1’ Chi-Squared test. From each of the test statistic equations, composed functions that can explain the relationship between the size of the sample (m), the probability of each category (k) and the amount of difference can be detected (d). Concluded that it takes a larger sample when the value of k nearest 0.5 and smaller when approaching a value of 0 or 1. The four test statistics that have the same pattern, but the statistics Yate’s need a sample of the largest while the three others the sample sizes are nearly equal.

Keywords: Pearson Chi-Square; Likelihood Ratio; Normal Distribution; 'N - 1' Chi-Squared

1. Introduction

The contingency table, also known as cross-tabulation, is a joint frequency distribution of cases based on two or more categorical variables. Contingency table analysis is a common method of analyzing the association between two categorical variables. There are two separate sampling strategies lead to the contingency table analysis. First, Test of Independence. A single random sample of observations is selected from the population of interest, and the data are categorized on the basis of the two variables of interest. Second, Test for Homogeneity [1]. Separate random samples are taken from each of two or more populations to determine whether the responses related to a single categorical variable are consistent across populations.

The Chi-Square statistic is the primary statistic used for testing the statistical significance of the cross-tabulation table. The two-way table is set up the same way regardless of the sampling strategy, and the chi-square test is conducted in exactly the same way. Chi-Square tests whether or not the two variables are independent. The chi-square test is based on an approximation that works best when the expected frequencies are fairly large. No expected frequency should be less than 1 and no more than 20% of the expected frequencies should be less than 5. For tables larger than 2x2, the chi-square distribution with the appropriate degrees of freedom provides a good approximation to the sampling distribution of Pearson's chi-square and the Likelihood Ratio Chi-Square. However, the Chi-Square statistic is only approximated by the chi-square distribution, and that approximation worsens with small expected frequencies. When we have very small expected frequencies, the possible values of the chi-square statistic are quite discrete. The general rule is that the smallest expected frequency should be at least five. However Cochran (1952), who is generally considered the source of this rule, acknowledged that the number “5” seems to be chosen arbitrarily [3].

Statistical test of 2 × 2 tables have been under discussion for a hundred years and dozens of research papers have been devoted to them. However, there is still a lack of consensus on the optimum method most texts recommend the use of the chi squared test for large sample sizes and the Fisher-Irwin test for small sample sizes, but there is disagreement on the boundary between ‘large’ and ‘small’ sample sizes [4]. Because the Pearson Chi-Square statistics and Likelihood Ratio Chi-Square statistics are based on a normal distribution approach, so in this paper discussed the limitations and requirements for proper use, especially related to the minimum number of sample requirements for table 2x2. Two distinct research designs can give rise to 2 × 2 tables: Model I (Row totals fixed, column totals free to vary), Model II (Both row & column totals free to vary). Model I
Nugraha, Jaka, “The Ability The Chi Squares Statistics To Rejecting The Null Hypothesis On Contingency Tables 2x2”

where we wish to test the hypothesis that two proportions are equal. Model II where we wish to test the hypothesis of independence of two variables of classification.

The validity of the chi-square test depends on both the sample size and the number of cells. Several rules of thumb have been suggested to indicate whether the chi-square approximation is satisfactory. One such rule suggested by Cochran (1954) says that the approximation is adequate if no expected cell frequencies are less than one and no more than 20% are less than five. Because of the expected cell frequency criterion in the second sampling strategy, it may be necessary to combine similar categories to lessen the number of categories in your table or to examine the data by subcategories.

2. Contingency Table 2x2 Analysis

A number of experiments involve binary outcomes (i.e., 1 and 0, yes and no). Typically, these occur when you are observing the presence or absence of a characteristic such as a disease, flaw, mechanical breakdown, death, failure, and so on. The analysis of the relationship between two bivariate categorical variables results in a $2 \times 2$ crosstabulation table of counts. There are $2 \times 2$ possible combinations of responses for these two variables. The $2 \times 2$ crosstabulation or contingency table has 2 rows and 2 columns consisting of $2 \times 2$ cells containing the observed counts (frequencies) for each of the $2 \times 2$ combinations. The observed frequencies are presented in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>B^c</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>b</td>
<td>m_1</td>
</tr>
<tr>
<td>A^c</td>
<td>b</td>
<td>d</td>
<td>m_2</td>
</tr>
<tr>
<td>Total</td>
<td>m_1</td>
<td>m_2</td>
<td>N</td>
</tr>
</tbody>
</table>

This type of analysis is called a contingency table analysis and is usually accomplished using a chi-square statistic that compares the observed counts with those that would be expected if there were no association between the two variables. Two separate sampling strategies lead to the chi-square contingency table analysis:

1. **Test for Homogeneity (Model I).** Separate random samples are taken from each of two or more populations to determine whether the responses related to a single categorical variable are consistent across populations. In this setting, you have a categorical variable collected separately from two or more populations. The hypotheses are as follows:

   - $H_0$: The distribution of the categorical variable is the same across the populations.
   - $H_a$: The distribution of the categorical variable differs across the populations.

2. **Test of Independence (Model II).** A single random sample of observations is selected from the population of interest, and the data are categorized on the basis of the two variables of interest. In this case, you have two variables and are interested in testing whether there is an association between the two variables. Specifically, the hypotheses to be tested are the following:

   - $H_0$: There is no association between the two variables.
   - $H_a$: The two variables are associated.

The two-way table is set up the same way regardless of the sampling strategy, and the chi-square test is conducted in exactly the same way. The only real difference in the analysis is in the statement of the hypotheses and conclusions. The chi squared test were (1) Pearson’s Chi-Squared test (2) Yates’s Chi-Squared test (3) Likelihood Ratio Chi-Square test (4) The ‘N - 1’ Chi-Squared test.

The original chi-square test, often known as Pearson's chi-square, dates from papers by Karl Pearson in the earlier 1900s. The test serves both as a "goodness-of-fit" test, where the data are categorized along one dimension, and as a test for the more common "contingency table", in which categorization is across two or more dimensions. The standard Pearson chi-square statistic for 2x2 Contingency table is defined as

$$
\chi^2 = \frac{(ad - bc)^2}{m_1 m_2 n_1 n_2}
$$

Pearson's chi-square statistic is not the only chi-square test that we have.

The likelihood ratio chi-square builds on the likelihood of the data under the null hypothesis relative to the maximum likelihood. It is defined as

$$
\chi^2 = 2(a \log(aN/m_1n_1) + b \log(bN/m_1n_2) + c \log(cN/m_2n_1) + d \log(dN/m_2n_2))
$$
Nugraha, Jaka, “The Ability The Chi Squares Statistics To Rejecting The Null Hypothesis On Contingency Tables 2x2”

If an expected frequency is lower than five, you have alternatives: Yates correction (Yates’s Chi-Squared test) and the N - 1 chi-square test. [2]

Yates’ correction (Yates, 1934) is equivalent to Pearson’s chi-square but with a continuity correction. In cases where an expected frequency is below 5, Yates’ correction brings the result more in line with the true probability.

\[
\chi^2 = \frac{(ad - bc)^2}{N} / m_1 m_2 n_1 n_2
\]

The N - 1 chi-square test is another option. Campbell (2007) carried out a very large sampling study on 2x2 tables comparing different chi-square statistics under different sample sizes and different underlying designs. He found that across all sampling designs, a statistic suggested by Karl Pearson’s worked best in most situations. The statistic is defined as

\[
\chi^2 = (ad - bc) / (N - 1) / m_1 m_2 n_1 n_2
\]

3. Calculation of Minimum Samples Size

Assuming a sample size of 2m by each population of m, observational data can be presented in Table 2. P(A | B) = k and P(Ac | B) = k with 0 < k < 1 and 0 < x < 1.

Table 2. Table of contingency for the sample P(A | B) = k

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Bm</td>
<td>m</td>
</tr>
<tr>
<td>A</td>
<td>Xm</td>
<td>m</td>
</tr>
<tr>
<td>Total</td>
<td>(k + x)m</td>
<td>2m</td>
</tr>
</tbody>
</table>

Based on chi square statistic formulas in equation (1) and (4), going to look for the minimum number of samples that can detect the difference between the proportion of population A and population of A’. The hypothesis is

\[ H_0: P(A | B) = P(A | B) \text{ vs } H_1: P(A | B) \neq P(A | B) \]

The minimum sample size to be able to reject \( H_0 \) with significance level \( \alpha \) for each test statistic occurs as follows:

1. Pearson’s Statistics

\[
\frac{(km(1-x) - xm(1-k)m)^2}{m(mk + x)m(2k - x)m} \geq \chi^2
\]

\[
\Rightarrow m \geq \frac{\chi^2}{2(k-x)^2} (k-x)^2
\]

\[
F(k; \alpha; x) = \frac{\chi^2}{2(k-x)^2}
\]

(5)

\( F(k; \alpha; x) \) is a function that shows the minimum sample required to be able to reject \( H_0 \).

2. Yate’s Statistics

\[
\frac{(km(1-x) - xm(1-k)m - m)^2}{m(mk + x)m(2k - x)m} \geq \chi^2
\]

\[
\Rightarrow 2(k-x)^2m^2 - (4|k-x| + (k+x)(2k-x)\chi^2)m + 2 \geq 0
\]

\[
\Rightarrow m \geq \frac{4(k-x)[(k+x)(2k-x)\chi^2]^{1/2} + \left((4|k-x| + (k+x)(2k-x)\chi^2)^2 - 16(k-x)^2\right)^{1/2}}{2(k-x)^2}
\]

(6)
3. **Campbell’s Statistics**

\[
\frac{(km(1-x)m - xm(1-k)m)^2}{m(m(k+x)m(2-k-x)m} \geq x^2_{\alpha}
\]

\[
\Longleftrightarrow m \geq \frac{x^2_{\alpha}(k+x)(2-k-x))}{2(k-x)^2} + \frac{1}{2}
\]

\[
F(k; \alpha; x) = \frac{x^2_{\alpha}(k+x)(2-k-x))}{2(k-x)^2} + \frac{1}{2}
\]

(7)

4. **Likelihood’s Statistics**

\[
\ln \left( \frac{2km}{(k+x)m} \right)^{2km} \left( \frac{2(1-k)m}{(2-k-x)m} \right)^{2(1-k)m} \left( \frac{2xm}{(k+x)m} \right)^{2x} \left( \frac{2(1-x)m}{(2-k-x)m} \right)^{2(1-x)m} \right) \geq x^2_{\alpha}
\]

\[
\Longleftrightarrow 2m \ln \left( \frac{2k}{(k+x)} \right)^{k} \left( \frac{2(1-k)}{(2-k-x)} \right)^{(1-k)} \left( \frac{2x}{(k+x)} \right)^{x} \left( \frac{2(1-x)}{(2-k-x)} \right)^{(1-x)} \right) \geq x^2_{\alpha}
\]

\[
\Longleftrightarrow m \geq \frac{2m \ln \left( \frac{2k}{(k+x)} \right)^{k} \left( \frac{2(1-k)}{(2-k-x)} \right)^{(1-k)} \left( \frac{2x}{(k+x)} \right)^{x} \left( \frac{2(1-x)}{(2-k-x)} \right)^{(1-x)} \right)}{\chi^2_{\alpha} - 1}
\]

\[
F(k; \alpha; x) = \frac{\chi^2_{\alpha}}{2m \ln \left( \frac{2k}{(k+x)} \right)^{k} \left( \frac{2(1-k)}{(2-k-x)} \right)^{(1-k)} \left( \frac{2x}{(k+x)} \right)^{x} \left( \frac{2(1-x)}{(2-k-x)} \right)^{(1-x)} \right) - 1}
\]

(8)

4. **Properties of Functions** \(F(K; \alpha; X)\)

The properties of the function \(F(k; \alpha; x)\) illustrated using Figure 1 and Figure 2 by taking the value of \(\alpha=0.05\). Figure 1 and Figure 2 describes the effects of changes in the value of \(k\), namely \(P(A \cap B)\) and the difference between the value of \(P(A\mid B)\) and \(P(A\mid B)\).

\(\chi^2 = 3.8415\) and \(d=|k-x|\)

**Figure 1.** Graph \(F(k; \alpha; x)\) on somevaluek  
**Figure 2.** Graph \(F(k; \alpha; x)\) on someevaluated

As in Figure 1, the greater the ability to detect a difference (d smaller) the number of samples required is also greater. Figure 2 illustrates that the sample also needs greater when k approaches 0.5. Furthermore, a comparison of several test statistics can be explained in Table 3 and Figure 3 to Figure 6. Table 3 is an example in the case of k = 0.5 and k = 0.4 which shows that (a) The number of
samples required at Pearson’s statistics, Likelihood’s statistics and Campbell’s statistics relatively similar (b) Yate’s statistics require larger samples than the three other statistics. (c) the greater the ability to detect a difference (d smaller) the number of samples required also getting bigger.  

Table 3. Table contingency $k=0.4$ and $k=0.5$

<table>
<thead>
<tr>
<th>Nilai d</th>
<th>$k=0.4$</th>
<th>$k=0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson</td>
<td>Yate</td>
</tr>
<tr>
<td>0.01</td>
<td>18514</td>
<td>19454</td>
</tr>
<tr>
<td>0.02</td>
<td>4647</td>
<td>5117</td>
</tr>
<tr>
<td>0.03</td>
<td>2073</td>
<td>2387</td>
</tr>
<tr>
<td>0.04</td>
<td>1170</td>
<td>1406</td>
</tr>
<tr>
<td>0.05</td>
<td>752</td>
<td>941</td>
</tr>
</tbody>
</table>

In general, to be able to detect the differences of the particular, Yate’s statistics require larger samples than the likelihood’s statistics, Campbell’s statistics and Pearson’s statistics. As in Figure 3 and Figure 4, likelihood’s statistics, Campbell’s statistics and Pearson’s statistics have relatively similar properties. The four statistic has the same pattern (as a function of d or k).

**Figure 3.** The sample size and $d=0.05$

**Figure 4.** The sample size $k=0.1$

In more detail, to determine the statistical differences, likelihood’s statistics, Campbell’s statistics and Pearson’s statistics taken short intervals. From Figure 5 and Figure 6, it appears that the order from the smallest sample size respectively: likelihood’s statistics, and Pearson’s statistics and Campbell’s statistics.
Nugraha, Jaka., “The Ability The Chi Squares Statistics To Rejecting The Null Hypothesis On Contingency Tables 2x2”

**Figure 5.** The sample size at $k \in (0.01, 0.04)$

**Figure 6.** The sample size at $k \in (0.5, 0.53)$

5. Conclusion

From the above discussion it can be concluded that (a) The higher the sample size, the ability to detect differences (rejecting $H_0$), the better. (b) The ability of Yate’s statistics to be lower than Pearson’s statistics, Likelihood’s statistics, and Campbell’s statistics. (c) Ability Likelihood’s statistics relatively similar to the Pearson’s statistics and better than the Campbell’s statistics should give a summary of:

Acknowledgement. The author would like to thank to Islamic University of Indonesia for supporting this research.

References

# Originality Report

## Similarity Index

<table>
<thead>
<tr>
<th>#</th>
<th>Source</th>
<th>Words</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>International Encyclopedia of Statistical Science, 2011.</td>
<td>213</td>
<td>8%</td>
</tr>
<tr>
<td>2</td>
<td><a href="http://www.iancampbell.co.uk">www.iancampbell.co.uk</a></td>
<td>87</td>
<td>3%</td>
</tr>
<tr>
<td>3</td>
<td>imt-n-section-d-2103-15.blogspot.com</td>
<td>57</td>
<td>2%</td>
</tr>
<tr>
<td>4</td>
<td>partnershungary.hu</td>
<td>38</td>
<td>1%</td>
</tr>
<tr>
<td>5</td>
<td>sites.google.com</td>
<td>29</td>
<td>1%</td>
</tr>
</tbody>
</table>

- **Exclude Quotes**: On
- **Exclude Bibliography**: On
- **Exclude Matches**: < 15 Words
Berita Acara Hasil Pengecekan Keaslian Karya Ilmiah
Atas Nama Jaka Nugraha, Dr., S.Si., M.Si. Untuk kenaikan Jabatan
Dari Lektor Kepala (400 AK) ke Guru Besar (850 AK)

Pada tanggal 23 Oktober 2020 telah dilakukan pengecekan *originality* atau *similarity* terhadap karya Ilmiah Dosen Tetap Universitas Islam Indonesia:

<table>
<thead>
<tr>
<th>Nama</th>
<th>Jaka Nugraha, Dr., S.Si., M.Si.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIP/NIDN/NIK</td>
<td>-/0524027101/956110102</td>
</tr>
<tr>
<td>Prodi</td>
<td>Statistika (S1)</td>
</tr>
<tr>
<td>Fakultas</td>
<td>Matematika dan Ilmu Pengetahuan Alam</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NO</th>
<th>KARYA</th>
<th>REPORT ORIGINALITY</th>
<th>KETERANGAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Menghasilkan karya ilmiah berupa jurnal internasional terindeks bereputasi dengan impact factor pada Sustainable Chemistry and Pharmacy, Scopus &amp; Scimagojr, SJR=0.62 pada bulan Juni 2018, dengan judul 'Preparation, characterization, and modelling activity of potassium fluoride modified hydrotalcite for microwave assisted biodiesel conversion'. ISSN: 2352-5541. Volume 8; issue 43252; pages 63-70; jumlah halaman 8</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. Menghasilkan karya ilmiah berupa jurnal internasional terindeks pada The Turkish Online Journal of Design Art and Communication, Copernicus-DOAJ pada bulan September 2018, dengan judul 'Brown's Weighted Exponential Moving Average (B-Wema) with Levenberg-Marquardt Optimization to Forecasting Rate of Return'. ISSN: 2146-5193. Volume Special Edition; issue ; pages 1744-1749 ; jumlah halaman 6</td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8. Menghasilkan karya ilmiah berupa jurnal internasional terindeks pada International Journal of Engineering and Technology, ProQuest-DOAJ pada bulan September 2018, dengan judul 'Sentiment Analysis on Mobile Banking Application Using Naive Bayes Classifier and Association Methods'. ISSN: 2227-524X. Volume 7; issue 4.15 (Special Issue); pages 244-247; jumlah halaman 4</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11. Menghasilkan karya ilmiah berupa prosiding seminar internasional terindeks pada 2nd International Conference on</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Judul</td>
<td>Persentase</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----------------------------------------------------------------------</td>
<td>------------</td>
<td></td>
</tr>
</tbody>
</table>

Pengecekan di atas menggunakan alat Ithenticate dengan meniadakan (exclude) beberapa hal dengan ketentuan sebagai berikut:

1. Meniadakan (exclude) hasil cek kesamaan karya yang kurang dari 2 persen.
2. Meniadakan (exclude) hasil cek kesamaan karya yang disitas oleh pihak lain.
3. Meniadakan (exclude) hasil cek kesamaan karya yang terindikasi plagiasi kepada karya ilmiah yang bersangkutan.

4. Meniadakan (exclude) hasil cek kesamaan karya yang menunjukan url atau laman karya ilmiah yang bersangkutan.

5. Meniadakan (exclude) hasil cek kesamaan karya yang diupload dalam bentuk yang berbeda (online pribadi) yang terdeteksi merupakan karya sendiri bukan merupakan laman publikasi Jurnal resmi hanya untuk kepentingan sharing (seperti https://www.researchgate.net facebook.com dll) sehingga bukan termasuk auto-plagisasi/self plagiarism.

Berdasarkan hasil pengecekan di atas, maka karya ilmiah tersebut di atas dapat diteruskan usulannya ke Lembaga Layanan Dikti Wilayah V.

Yogyakarta, 23 Oktober 2020
Rektor

[Signature]

Prof. Fathul Wahid, S.T., M.Sc., Ph.D.