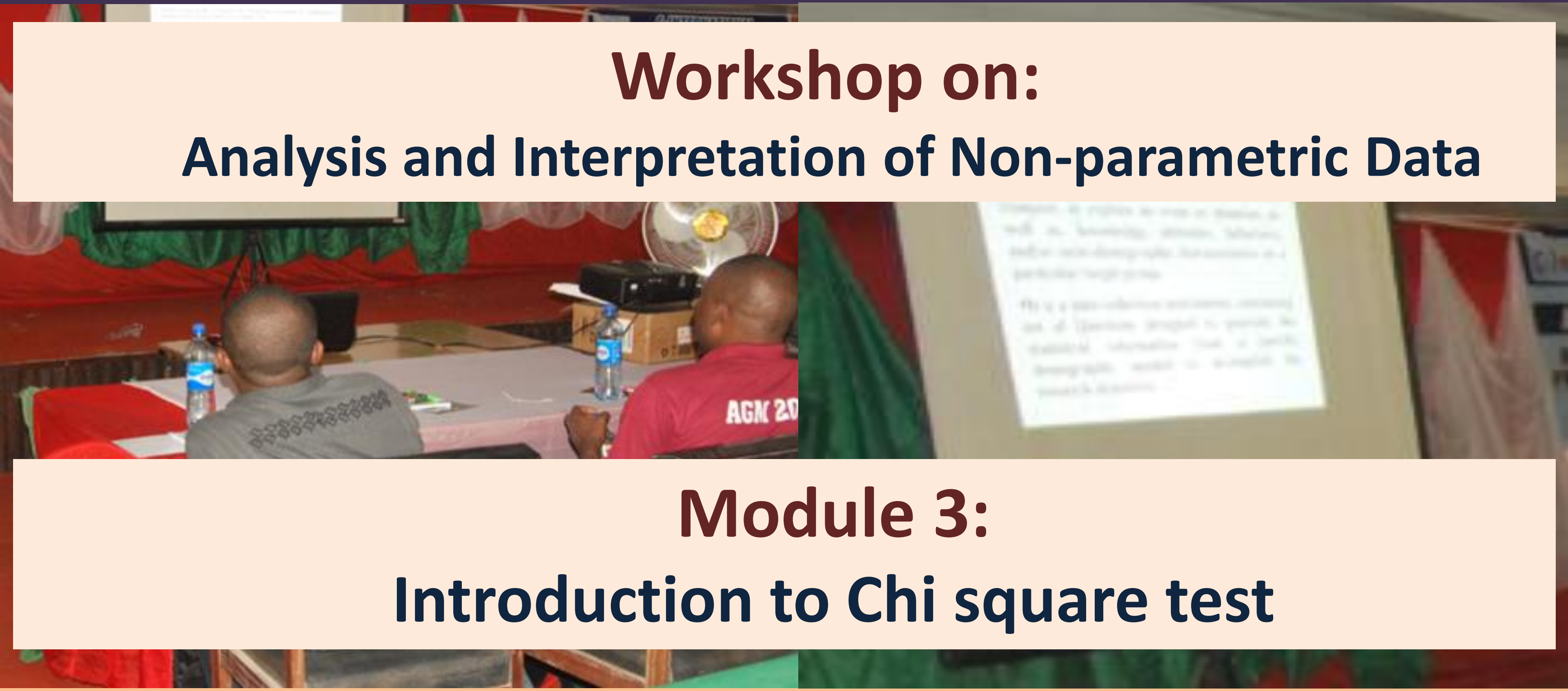
DOI: 10.33515/cint.sld/2023/002

C-International Tutors

Committed to bringing knowledge and skills to your doorsteps



Workshop on:

Analysis and Interpretation of Non-parametric Data

Module 3:

Introduction to Chi square test

https://cintarch.com/tutorials-videos/

https://www.facebook.com/cintarch.tutors/

cintarch.tutors@gmail.com

https://www.instagram.com/cintarchtutors/?hl=en

https://twitter.com/CInternational6

https://www.youtube.com/channel/UCPEhxtU4B3Tu0PtPyMGBU8Q?view_as=subscriber

Introduction to Chi square test

By

Prof. Awosan K.J.

Learning outcomes

After this session, you will know:

- The type of data the chi square test is used to analyze
- The types of chi square tests
- The conditions (or assumptions) for using the chi square test
- The types of comparisons that are done in Pearson's chi square test
- Measures of effect size in chi square
- Post-hoc pairwise comparison in chi square

Chi square test

- The chi square (χ^2) test is a non-parametric test that measures how a model compares to actual observed data. It is used in hypothesis testing for categorical data

❑ Types of chi square tests

The common types of chi square tests includes:

- **Pearson's chi square test** (for independent groups)
- **McNemar's chi square test** (for paired observations)
- **Fisher's exact test** (used in situations where the assumptions for Pearson's chi-square test are not met in a 2 x 2 table)
- **Likelihood ratio chi-square test** (used in situations where the assumptions for Pearson's chi square test are not met in an R x C table)

Chi square test contd.

❖ Pearson's chi square test

- The Pearson's chi square test (χ^2) is the most commonly performed chi-square test. It is used to determine whether there is a statistically significance difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table
- **The conditions to be fulfilled for Pearson's chi square test include**;

- Chi square only works when frequencies are used in the table (proportions, means and physical measurements are not valid)
- The observed frequencies were obtained through a simple random sample
- The variable under study is categorical (i.e., nominal or ordinal)

Chi square test contd.

❖ Pearson's chi square test contd.

- **The conditions to be fulfilled for Pearson's chi square test include contd.**;
  - All of the expected frequencies should be more than 1
  - At least 80% of the expected frequencies should be more than 5
- The Pearson's chi square test is used for three types of comparisons, these are:
  - Test of goodness of fit
  - Test of independence
  - Test of homogeneity

Chi square test contd.

❖ McNemar's chi square test

- The McNemar's chi square test ($\chi^2_{McNemar}$) is used to determine if there is a statistically significant difference in proportions between paired data
- It is a type of 2 x 2 chi square test that is specifically used for comparisons of variables from matched pairs, and uses information only from discordant pairs
- The McNemar's test is more appropriate for testing matched pairs than the simple Chi square test, because it tends to decrease the likelihood of a type I error (i.e. the null hypothesis is less often incorrectly rejected)

Chi square test contd.

❖ McNemar's chi square test contd.

- In a McNemar's chi square test there is only 1 degree of freedom because, with the marginal values constant, one of the values in the discordant pair automatically determines the other value

- **Application and characteristics of the McNemar's χ^2 test**

  - Except that the variables being compared are from matched pairs and are not independent, the uses and restrictions of the McNemar's χ^2 test are the same as those of 2 x 2 chi-squared test
  - The null hypothesis tested is that the expected frequencies for the discordant pairs are equal
  - The assumptions are basically the same as those for Chi square test

Chi square test contd.

❖ Fisher's exact test

- The Fisher's exact test is used if the conditions for the Pearson's chi square test couldn't be met in a 2 x 2 contingency table (especially, when more than 20% of cells have expected frequencies less than 5)
- Unlike the Pearson's chi square test which relies on approximation, the Fisher's exact test is one of exact test
- Also, unlike the Pearson's chi square test, the Fisher's exact test has no test statistic to report (only the p value is reported, but the name of the test being reported should be specified)

Chi square test contd.

❖ Likelihood ratio chi square test

- The likelihood ratio chi square statistic (G^2) is based on the ratio of the observed to the expected frequencies, unlike in the case of the Pearson’ chi square statistic (χ^2) which involves the squared difference between the observed and the expected frequencies
- It is used when the sample size is small (in which case the conditions for using the Pearson’s chi square test could not be fulfilled) in analysis involving large tables (i.e., tables with more than 2 rows or columns)

Steps in hypothesis testing

- The steps in hypothesis testing are as follows:
- **Step 1**:Extract the relevant data and state the assumptions
- **Step 2**: State the hypotheses (i.e., the null and alternative hypotheses)
- **Step 3**: Set the criteria for a decision (i.e., the significance level)
- **Step 4**: Compute the test statistic (depending on the type of variables, the distribution and the number of groups concerned)
- **Step 5**: Make a decision and conclude

Vital Issues in Hypothesis testing

- While performing hypothesis testing the vital issues to consider are as follows:
- The research hypothesis (in terms of establishing associations, making comparisons for significant differences, making predictions, or identifying predictors)
- The type of test (i.e., one-tailed or two-tailed tests)
- The corresponding p-value of the derived test statistic
- The significance level
- Type I (**α**) and type II (**β**) errors

Vital Issues in Hypothesis testing contd.

❑ P value

- After conducting a hypothesis testing, the value of the test statistic and the corresponding p-value are obtained.
- The p-value is the probability of obtaining an effect (i.e., a value of the test statistic) as extreme as, or more extreme than that observed if the null hypothesis (Ho) is true (i.e., the probability of obtaining it by chance)
- The p-value of the test statistic shows how extreme that statistic is for a sample data

Vital Issues in Hypothesis testing contd.

❑ The significance level (α)

- The significance level (α) of a statistical hypothesis test is the probability of wrongly rejecting the null hypothesis (Ho), if it is in fact true
- It is the probability of a type I error (or alpha) and it is set by the investigator in relation to the consequences of such an error
- The significance level is made as small as possible in order to protect the null hypothesis and to prevent as far as possible inadvertently making false claims

Vital Issues in Hypothesis testing contd.

❑ The significance level (α) contd.

- The significance level (α) is the threshold value that we measure the corresponding p-value of the computed test statistic against (to make a decision on the null hypothesis); it is usually set at $p \leq 0.05$)

- It tells us how extreme observed results must be in order to reject the null hypothesis of a test of significance

Vital Issues in Hypothesis testing contd.

❑ The significance level (α) contd.

- If the p-value is less than or equal to alpha (e.g., $p \leq 0.05$), the null hypothesis is rejected (in favor of the alternative hypothesis), and it is concluded that the result is statistically significant. This means that one is reasonably sure that the observed effect is due to something besides chance alone

- If the p-value is greater than alpha (e.g., $p > 0.05$), the null hypothesis is not rejected, and it is concluded that the result is not statistically significant. This means that one is reasonably sure that the observed effect can be explained by chance alone

Measuring the Effect Size in Chi square analysis

- Chi square test only shows whether there is, or there is no significant relationship (i.e., association) between the dependent and independent variables, but it does not measure the strength of the association (i.e., the effect size). The measures of effect size in chi square analysis include:

The dr

For a 2 x 2 contingency table:

- **Odds ratio**, or
- **Phi** (φ): It is computed as: $\varphi = \sqrt{\frac{\chi^2}{n}}$

Where:
φ = Phi
χ^2 = Chi square
n = Sample size
df = degree of freedom

For larger tables:

- **Cramer's V** (V): It is computed as: $V = \sqrt{\frac{\chi^2}{n-df}}$

Measuring the Effect Size in Chi square analysis contd.

❖ Interpretation of effect size using Odds Ratio (OR)

- The effect size using the Odds Ratio is interpreted as follows:

Measure	Interpretation
OR > 1	There is positive or direct association between the exposure and the outcome (e.g., there is a direct association between optimal lecture attendance by students and their examinations scores)
OR = 1	There is no association between the exposure and the outcome (e.g., there is no association between students admission numbers and their examination score)
OR < 1	There is negative or inverse association between the exposure and the outcome (i.e., the exposure protects against the outcome; e.g., full immunization protects children against childhood diseases)

Measuring the Effect Size in Chi square analysis contd.

❖ Grading of Phi and Cramer’s V effect sizes

- The effect size using Phi and Cramer’s V is graded as follows:

Degree of freedom	Effect size		
	Small	Medium	Large
1	0.10	0.30	0.50
2	0.07	0.21	0.35
3	0.06	0.17	0.29
4	0.05	0.15	0.25
5	0.04	0.13	0.22

- The **degrees of freedom** (df) is the number of independent pieces of information on a statistic. The degrees of freedom for a table is (number of rows – 1), multiplied by (number of columns – 1).
- It is usually presented as (**r-1**) x (**c-1**)

Post-hoc pairwise comparison in chi square test

- **Determining the actual source(s) of association in chi-square analysis**

- The Chi square test generally shows whether or not there is, association between two variables

- If there are three or more levels in either variable, a post-hoc pairwise comparison (which uses the Bonferroni correction) is needed to know the levels that are actually associated

Chi square analysis in SPSS

- Chi square analysis in SPSS shall be covered in Module 4
- The datasets can be accessed through the links below:

Cintarch Data_Chi square test 1_Pearson's chi square test_Cross-sectional study:
https://drive.google.com/file/d/1imquVozXQv6ysRttzFb5K62SwrgdIM6P/view?usp=sharing

Cintarch Data_Chi square test 2_Pearson's chi square test_Experimental study:
https://drive.google.com/file/d/1Ppd6xpBQof6UY43chT51YyCpW0UVWqJM/view?usp=sharing

To access the videos, please visit: https://cintarch.com/tutorials-videos/

Chi square analysis in SPSS contd.

- The remaining dataset can be accessed through the link below:

Cintarch Data_Chi square test 3_McNemar's chi square test_Experimental study:
https://drive.google.com/file/d/1TK96NdZtEjtXEvQQIakLskhQnVjT89AU/view?usp=sharing

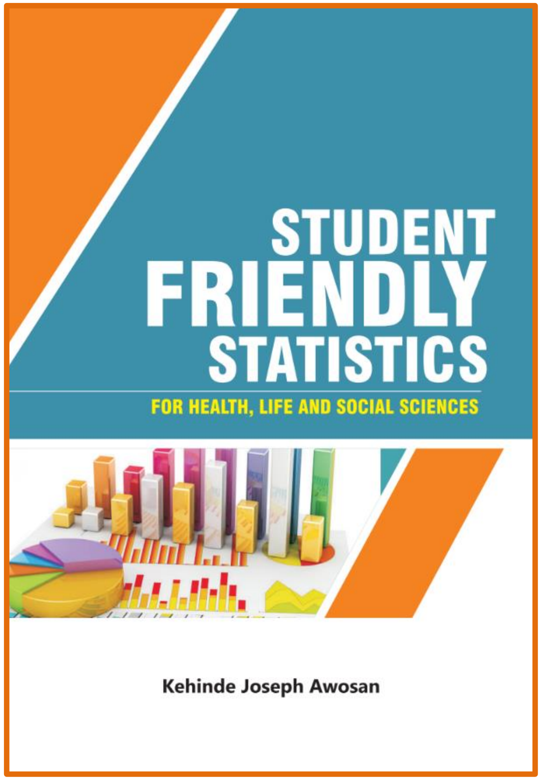
To access the videos, please visit: https://cintarch.com/tutorials-videos/

Further reading

Awosan KJ (2020). Student Friendly Statistics for Health, Life and Social Sciences. Ikeja, Lagos: Somerest Ventures

Now On Sale

Student Friendly Statistics for Health, Life and Social Sciences



Please click on the link below to access the excerpts from the book:

https://cintarch.org/wp-content/uploads/erf_uploads/2022/02/CB_Excerpts-from-Student-Friendly-Statistics-for-Health-Life-and-Social-Sciences.pdf

To buy a copy or copies of the book online, please click on the link below to access the "ORDER FORM":

https://forms.gle/3J8dWn6Ng6oMKd2e8

C-International Archives' Journals

Please find the List of our Journals at:

https://cintarch.com/journals/

Please find our Recently Published Articles at:

https://cintarch.com/about-us/

Please Submit your Manuscripts at:

https://cintarch.com/submit-manuscripts/