Course : Research statistics **Lecturer : Pr. Saliha CHELLI**

Level : Master

Lecture : 8

The chi-square test

Lecture objectives :Introducing the chi-square test (parametric test) and how to compute it

Introduction

As seen in lecture 7, tests can be parametric or non-parametric based on whether data are normally distributed or not and also the type of analysed data . Non-parametric tests should be used when the level of measurement of the variables is nominal or ordinal. This lecture introduces the chi-square non-parametric test ,also called distribution free test.

1. **Basic concepts**

- **Statistical significance** only relates to the amount of confidence we have that the findings we obtained were not the product of pure chance (Descombe, M. p.259).
- **The Significant level** or alpha level: conventionally, the level of significance is repoted as.01 or .05. A signicant level of .01 means that there is a 1% chance on the test of any null hypothesis (Salking Neil, J. p.163). 0.05 means that there is 5% chance on the test (significance generally used social sciences).
- **The Degrees of freedom (df)** refers to the values that have the freedom to vary in the data sample.
- **The Critical value** is a point on the test that is compared to the test statistic to determine whether to reject the null hypothesis.

2. **Chi-square test of independent samples**

The Chi-square statistic is a non-parametric (distribution free) tool designed to analyze group diff erences when the dependent variable is measured at a nominal level. The shi-Square ($X^{2)}$ is a statistical test used to determine whether your experimentally observed results are consistent with your hypothesis. This test statistics measures the agreement between actual counts and expected counts assuming the null hypothesis**.**

Example, Example, when we are interested in some **non-quantitative aspect of a subject's behaviour** like for example, does a subject '**think aloud' or not**? Does he consciously use mnemonics in a memory experiment, or doesn't he? Does he maintain eye contact with an interviewer, or does he look away? These measurements, in so far as they are measurements at all, are said to constitute ***nominal scaling*** *(Miller, 1984,p..72). Categorical data, examples of which could be* ***gender*** *(male or female) or university* ***degree classifications*** *(1, 2, 3, pass or fail) – or any other variable where each* ***participant falls into one category.***

** When the data we want to analyse is like this, a chi-square test, denoted* ***χ^2,*** *is usually the appropriate test to use*

The Chi-square test is used to test hypotheses a the distribution of observations in different categories.

- The **null hypothesis** (Ho) is that the **observed frequencies a**re the same as **the expected frequencies** (except for chance variation).
- If the observed and expected frequencies **are the same, then χ^2 = 0**.
- If the frequencies you observe **are different from expected frequencies,** the **value of χ^2 goes up**.
- The larger the value of χ^2, the more likely it is that the distributions are significantly different.

3. Chi-square test steps

A. Collect observed frequency data

B. Calculate expected frequency data (refer to the table below)

C. Determine Degrees of Freedom

D. Calculate the chi square

Result: If the chi square statistic exceeds the probability or critical table value (based upon a p-value of x and the degrees of freedom), the null hypothesis should be rejected**.**

Calculation of the expected frequency

Variable 2	Data type 1	Data type 2	totals
Category 1	a	b	a-b
Category 2	c	d	c-d
total	a-c	b-d	a-b-c-d

The formula used to calculate the expected frequency (E)

$$E = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

Example : Gender and Choice Preference

calculation of male's expected frequency (row total=50, column total= 66, Grand total= 105

	Male	Female	Total
Like	36	14	50
Dislike	30	25	55
Total	66	39	105

To find the expected frequencies, assume independence of the rows and columns. Multiply the row total to the column total and divide by grand total as show in the formula above.

E :(5O x 66) : 105= 31.43

Global expected frequencies (male and female's likes and dislikes)

	Male	Female	Total
Like	31.43	18.58	50.01
Dislike	34.58	20.43	55.01
Total	66.01	39.01	105.02

The number of degrees of freedom is calculated for an x-by-y table as (x-1) (y-1), so in this case (2-1) (2-1) = 1*1 = 1. The degrees of freedom is 1.

After having calculated the expected frequency, we calculate the chi-square test as follows

A contigency table is used in which we put:

1. Observed frequencies (O)
2. expected frequencies (E)
3. 0bserved minus the expected frequencies (0-E)
4. Obseved minus expected squared ($0-E)^2$
5. Obseved minus expected squared divided by the expected frequencies $(O-E)^2$: E

Finally, we sum $(O-E)^2$

Then, we use the following formula

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

χ^2 = the test statistic $\sum$ = the sum of

O = Observed frequencies E = Expected frequencies

O	E	O-E	$(O-E)^2$xE
36	31.43	4.57	**.67**

14	18.58	-4.58	**1.13**
30	34.58	-4.58	**.61**
25	20.43	4.57	**1.03**

Chi square observed statistic = 3.44

Or $\chi^2 = 3.44$

After having found the value of χ^2, the next step is to calculate the degree of freedom

The **degrees of freedom** for the **chi-square** are calculated using the following **formula**: **df** = **(r-1)(c**-1) where r is **the number of rows** and **c is the number of columns**.

df gender preferences

Df= (2-1) (2-1)= 1

If the observed value of **chi-square test** statistic (**χ^2) is greater than the critical value,** the **null hypothesis can be rejected**

The next step is to compare the value of **χ^2 with the critical value of the degree of freedom 1** using the following table

Df	0.5	0.10	0.05	0.02	0.01	0.001
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.837	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.51

The Chi Square (Observed statistic) = **3.44**

Probability Level (df=1 and .05) = **3.841** as shown in the table

So, Chi Square statistic < Probability Level (show in yellow in the table)

So, the null hypothesis hpothesis is accepted.**There is no significant difference between product choice and gender.**

If the observed value of **chi-square test** statistic (**χ^2) is greater than the critical value,** the **null hypothesis can be rejected. If it is lower, the null hypothesis is accepted.**

References

Miller, S. (1984). Experimental design and statistics (2nd ed.). London and New York: Routledge.

Salking, N.J. (2012): 100 Questions and Answers about Research Methods. University of Kansas: Sage Publications.

Descombe, M. (2003). The good research guide for small-scale social research projects (2nd ed). McGraw-Hill Education: Open University Press.