**Practical 7**

**Chi Square (X2) Distribution & Chi Square Test**

The chi-square test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories. Do the number of individuals or objects that fall in each category differ significantly from the number you would expect? It is the most widely used test is test of independence. It has the following properties:

1. It is a non-parametric test used for Quantitative data e.g. marital status (single, married, widowed). (Deals with frequencies).
2. It is one of the most widely used tests in statistical application.
3. Derived from normal distribution, but it is a skewed distribution (not normal), started from zero and has only one tail (only positive values), between zero and + ∞, i.e. no negative values.
4. X2 relates to the frequencies of occurrence of individuals or events in categories of one or more variables.
5. X2 is used to test agreements between the observed frequencies with certain characteristics and expected frequenciesunder certain hypothesis.

 X2 distribution

GENERALLY: two criteria classification, when applied to the same set of entities, are independent (no association). In other words; if a sample of n size drawn from a population; the frequency of occurrence of entities are cross classified on the basis of the two variables of interest (X &Y). The corresponding cells are formed by the intersection of the rows and columns & constructed table is a contingency table as the adjacent cells are interrelated.

Ex:

|  |  |  |
| --- | --- | --- |
|  | **Smoking** | **Total** |
|  **Yes (Y1)** |  **No (Y2)** |
| **MI** | **Yes (X1)** | **a (Smoke & MI)** | **b (No smoke & MI)** | **a + b** | **Raw margin** |
| **No (X2)** | **c (Smoke & no MI)** | **d (No smoke & no MI)** | **c + d** |
| **Totals** | **a + c** | **b + d** | **a + b + c + d** |
| **Column margin** | **Grand total**  |

The question is whether the observed numbers in MI categories and smoking categories (frequencies in cells **a + b + c + d**) differ significantly from the expected one.

In other words, if there is no association between smoking and MI the observed numbers in each cell would be the same as expected.

So X2 test depends on:

* 1. Observed values (O); number of subjects in our sample that fall into the various categories of the variable of interest (data of the sample).
	2. Expected values (E); number of subjects that we would expect to observe in our sample if there is no association between the two variables.

To calculate the **Expected values**: **(E)** = [**Raw margin** X **Column margin**] / **Grand total**

**\* Always ∑ (O) = ∑ (E).**

 **X2 = (O-E) 2 /E**

\* df = (r-1)(c-1), and at ⍺=0.05 and from X2-distribution table we find the level of significance (X2 distribution curve is a single tail curve so α is not divided by 2). **d.f always equals to 1 in 2X2 table.**

 **Conclusion:** if calculated X2 > tabulated X2; then the association is significant and vice versa

**Analyze \_\_\_\_\_\_\_\_\_Descriptive Statistics \_\_\_\_\_Crosstabs \_\_\_\_\_\_Select first variable \_\_\_\_select second variable \_\_\_\_\_\_Statistics --------------click on chi sauare test \_\_\_\_\_\_ok**



****

****

**The result will be :**

****