Chi-square Procedures

The symbol X^2 is often used rather than the term Chi-square. The Greek letter ***x*** is pronounced **Chi**. Like the t distribution.

GOODNESS-OF-FIT TEST

In many situations, each element of a population is assigned to one and only one of **k** categories or classes. The goodness-of-Fit test statistic is given in formula:.

$$X^2 = \sum \frac{(o-e)^2}{e}$$

Where ***O*** represents an observed frequency and ***e*** represents an expected frequency, and the sum is over all **K** categories.

The steps for performing a Goodness- of-Fit test are :.

Step 1:. State the null and alternative (research) hypothesis concerning the hypothesized distribution for the **K** categories.

Step 2:. Use the X^2 table and the level of significance , α , to determine the rejection region.

Step 3:. Compute the value of the test statistic as follows:.

$$X^2 = \sum \frac{(o-e)^2}{e}$$

Step 4:. State your conclusion. The null hypothesis is rejected if the computed value of the test statistic falls in the rejection region. Otherwise, the null hypothesis is not rejected.

- df.= k-1 in goodness-of-Fit test.

Example:.

Is the number of G^+ bacteria is equal to G^- bacteria in the following table data. Tested this at α: 0.05.

Bacteria type	The number
G^+	70
G^-	90

Answer:.

H_0:. The number of G^+ = The number of G^-

The expected value = 160/2 = 80

$$X^2 = \sum \frac{(o-e)^2}{e}$$

$$= \frac{(70-80)^2}{80} + \frac{(90-80)^2}{80}$$

= 2.5

X^2 table at α: 0.05 and df. =2-1 =1 is 3.84

X^2 cal. < X^2 tab.

Therefore , the number of G^+ and G^- is equal and without any significance difference , and this difference belong to chance factor.

TEST OF INDEPENDENCE

The " e " is called *estimated frequency* , the frequencies we expected to have under the null hypothesis of independence. The have the same marginal totals as those of the data observed. In this problem we do not compare proportions (because we have only one sample), what we really want to see is if the two factors or variables X_1 and X_2 are ***related*** ; the task we perform is a ***test for independence***. We achieve that by comparing the observed frequencies ***O***, versus those expected under the null hypothesis of independence , the expected frequencies ***e*** . This needed comparison is done through Pearson's Chi-square statistic :.

$$X^2 = \sum \frac{(o-e)^2}{e}$$

$$\boldsymbol{e} = \frac{(row\ total)(column\ total)}{sample\ size}$$

For large sample , X^2 has approximately a chi – square distribution with degrees of freedom under the null hypothesis of ***independence*** ,

df. = (r-1)(c-1)

with greater values lead to a rejection of H_0.

Example:.

A study was undertaken to investigate the roles of blood borne environmental exposures on ovarian cancer from assessment of consumption of coffee , tobacco and alcohol. Study subjects consist of 188 women in Baghdad with epithelial ovarian cancers , and 539 control women.

Coffee Drinkers	**Cases**	**Hospital Controls**	**Population Controls**	***Total***
Yes	177	249	233	***659***
No	11	31	26	***68***
Total	***188***	***280***	***259***	***727***

***Answer*:.** In this example, we want to compare the three proportions of coffee drinkers, but we still can apply the same chi – square test:.

$$e_{11} = \frac{659 \times 188}{727} = 170.42$$

$$e_{12} = \frac{659 \times 280}{727} = 253.81$$

$$e_{13} = \frac{659 \times 259}{727} = 234.77$$

$$e_{21} = \frac{68 \times 188}{727} = 17.58$$

$$e_{22} = \frac{68 \times 280}{727} = 26.19$$

$$e_{23} = \frac{68 \times 259}{727} = 24.23$$

$$X^2 = \sum \frac{(o-e)^2}{e}$$

$$= \frac{(177-170.42)^2}{170.42} + \frac{(249-253.81)^2}{253.81} + \frac{(233-234.77)^2}{234.77} + \frac{(11-17.58)^2}{17.58} + \frac{(31-26.19)^2}{26.19} + \frac{(26-24.23)^2}{24.23}$$

$$= 3.83$$

X^2 table at α:0.05 and df.=(2-1)(3-1) = 2 is 5.99

X^2cal < X^2 tab

The result indicates that the difference between the three groups is not significant at the 0.05 level. In other words, there is enough evidence to implicate coffee consumption in this study of epithelial ovarian cancer.

Example:. In a study to determine the relationship between the blood groups and the infection strength on the malaria disease. Data are calculated from 1500 persons as bellow. Is the two cases are related. Tested this at α: 0.05 .

Blood groups / Infection	A	B	AB	O	*Total*
None infection	543	211	90	476	***1320***
Infection median	44	22	8	31	***105***
Infection strong	28	9	7	31	***75***
Total	***615***	***242***	***105***	***538***	***1500***

Answer:.

H_0:. The two characteristics are independence.

H_a:. The two characteristics are related.

The expected values:.

$e_{11}= \frac{1320\times615}{1500} = 541.2$

$e_{12} = \frac{1320\times242}{1500} = 212.96$

$e_{13} = \frac{1320\times105}{1500} = 92.4$

$e_{14} = \frac{1320\times538}{1500} = 473.44$

$e_{21} = \frac{105\times615}{1500} = 43.05$

$e_{22} = \frac{105\times242}{1500} = 16.94$

$e_{23} = \frac{105\times105}{1500} = 7.35$

$e_{24} = \frac{105\times538}{1500} = 37.66$

$e_{31} = \frac{75\times615}{1500} = 30.75$

$e_{32} = \frac{75\times242}{1500} = 12.10$

$e_{33} = \frac{75\times105}{1500} = 5.25$

$e_{34} = \frac{75\times538}{1500} = 26.9$

$$X^2 = \Sigma \frac{(o-e)^2}{e}$$

$$= \frac{(543-541.2)^2}{541.2} + \ldots\ldots\ldots\ldots\ldots + \frac{(31-26.9)^2}{26.9}$$

$= 5.12$

df.= (r-1)(c-1)

= (3-1)(4-1)

= 6

X^2 table at α: 0.05 and df. 6 is 12.59.

X^2 cal. < X^2 tab.

Therefore , we should be accept the null hypothesis and rejected the alternative hypothesis. In other words, the blood group and the infection strength of malaria disease are independence.