Sri SAI RAM ENGINEERING COLLEGE
INSTITUTE OF TECHNOLOGY
West Tambaram, Chennai - 44

SAIRAM DIGITAL RESOURCES

YEAR IV SEM VII

MA8391

PROBABILITY AND STATISTICS.

UNIT NO III

PROBABILITY AND RANDOM VARIABLES

3.6 CHI-SQUARE DISTRIBUTION-GOODNESS OF FIT

SCIENCE & HUMANITIES

CHI-SQUARE DISTRIBUTION

Definition:

Sum of squares of independent n standard normal variates is called Chi- square random variate follows Chi-square distribution with n degrees of freedom.

Definition: Degrees of freedom

Degrees of Freedom refers to the maximum number of logically independent values, which are values that have the **freedom** to vary, in the data sample. Calculating **Degrees of Freedom** is key when trying to understand the importance of a Chi-Square statistic and the validity of the null hypothesis.

Applications of Chi-square distribution.

1,To test the goodness of fit

2.To test the independence of attributes.

The formula for the goodness of fit of a random sample to a hypothetical distribution.

$$\chi^2 = \sum \frac{\{O_i - E_i\}^2}{E_i}$$

. E_i and O_i are the expected and the corresponding observed frequencies respectively with (n-1) degrees of freedom.

Independent of attributes

To test independent of attributes we calculate χ^2 value from The cell frequencies and compare with table value

χ^2 corresponding to degrees of freedom. For a2x2 contingency table, $\begin{matrix} a & b \\ c & d \end{matrix}$

$$\chi^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

Degree of freedom is (p-1)(q-1), Pxq contingency table.

1. Fitting a Bionamial distribution degrees of freedom =n-1.
2. Fitting a Poisson distribution degrees of freedom =n-2.
3. Fitting a normal distribution degrees of freedom =n-3.

The uses of chi-square test.

a) It is used to test the goodness of fit. i.e.., it is used to decide whether a given sample may be reasonably regarded as a simple sample from a certain hypothetical population.
b) It is used to test the independence of attributes. i.e., if a population has two attributes, then it is used to test whether the two attributes are associated or independent, based on a sample drawn from the population.

The conditions for the validity of chi-square test.

a) The number N of observations in the sample must be reasonably large, say ≥ 50.

b) Individual frequencies must not be too small, i.e. $O_i \geq 10$.

c) The number of classes n must be neither too small nor too large, say $4 \leq n \leq 16$.

1. The theory predicts that the proportion of beans in four given group should be $9:3:3:1$. In an examination with 1600 beans, the numbers in the four groups were 882, 313, 287 and 118. Does the experimental result support the theory?

Solution: We want to test that the proportion of beans in the four groups are in the ratio

$9:3:3:1$. We apply χ^2 - test.

H_0: Proportion of beans in the four groups is in the ratio $9:3:3:1$.

H_1: Proportion of beans in the four groups is not in the ratio $9:3:3:1$.

Given that the observed frequencies are respectively 882, 313, 287 and 118.

Total observed frequency = 882 + 313 + 287 + 118 = 1600

Under H_0, the expected frequencies are

$$\frac{9}{16}\times 1600, \frac{3}{16}\times 1600, \frac{3}{16}\times 1600, \frac{1}{16}\times 1600$$

i.e., 900, 300, 300, 100 respectively.

The test statistic is $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$\chi^2 = \frac{(882-900)^2}{900} + \frac{(313-300)^2}{300} + \frac{(287-300)^2}{300} + \frac{(118-100)^2}{100}$$

$$= 4.726$$

Number of degrees of freedom $r = n - 1 = 3$

From the χ^2 table, $\chi^2_{(0.05)}(r = 3) = 7.81$.

$\therefore \quad \chi^2 < \chi^2_{(0.05)}$

$\therefore \quad H_0$ is accepted at 5% level of significance.

$\therefore$ The proportion of beans in the four groups are in the ratio $9:3:3:1$.

2. A survey of 320 families with 5 children revealed the following distribution:

No. of boys:	0	1	2	3	4	5
No. of girls:	5	4	3	2	1	0
No. of families:	12	40	88	110	56	14.

Is this result consistent with the hypothesis that male and female births are equally probable?

H_0: Male and female births are equally probable. i.e., P(Male birth) = ½ .

Based on H_0, the probability that a family of 5 children has r male children is

Given by the formula $nC_r\left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^{n-r} = 5C_r\left(\frac{1}{2}\right)^5$ (by binomial law), $r = 0,1,2,\dots 5$

Expected No. of families with r male children = $320\ X 5C_r\left(\frac{1}{2}\right)^5 = 10\ X\ 5C_r$.

Thus,

E_i :	10	50	100	100	50	10
O_i :	12	40	88	110	56	14

$$\chi^2 = \sum\frac{(O_i - E_i)^2}{E_i} = \frac{4}{10} + \frac{100}{50} + \frac{144}{100} + \frac{100}{100} + \frac{36}{50} + \frac{16}{10} = 7.16.$$

We have used the sample data to get $\sum E_i$ only. The values of the probabilities of p and q are not found using the sample data.

Hence degrees of freedom $\nu = n - 1 = 6 - 1 = 5$.

$\chi^2_{0.05}$ at $\nu = 5$ is 11.07 from the table.

Since $\chi^2 < \chi^2_{0.05}$, H_0 is accepted.

Thus it is reasonable to accept that the male and female births are equally probable.

3. Fit a Poisson distribution for the following data and test the goodness of fit.

x	0	1	2	3	4	5	6
frequency	**56**	**156**	**132**	**92**	**37**	**22**	**5**

A: Let us fit a poisson distribution with parameter λ to the given data.

λ is the mean of the distribution

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{\lfloor x}, x = 0,1,2,3,4,5,6$$

To find the mean of the distribution $\frac{\sum fx}{\sum f}$

x	0	1	2	3	4	5	6
f	56	156	132	92	37	22	5
fx	0	156	264	276	148	110	30

$$\sum fx = 984\ ,\ \sum f = 500$$

$$\therefore \quad \lambda = \text{mean} = \frac{\sum fx}{\sum f} = \frac{984}{500} = 1.97$$

$$P(X=0) = \frac{e^{-\lambda}\lambda^0}{\lfloor 0} = e^{-1.97} = 0.1394$$

$$P(X=1) = \frac{e^{-\lambda}\lambda^1}{\lfloor 1} = e^{-1.97} \times 1.97 = 0.2746$$

$$P(X=2) = \frac{e^{-\lambda}\lambda^2}{\lfloor 2} = \frac{e^{-1.97} \times (1.97)^2}{2} = 0.2705$$

$$P(X=3) = \frac{e^{-\lambda}\lambda^3}{\lfloor 3} = \frac{e^{-1.97} \times (1.97)^3}{6} = 0.1776$$

$$P(X=4) = \frac{e^{-\lambda}\lambda^4}{\lfloor 4} = \frac{e^{-1.97} \times (1.97)^4}{24} = 0.0875$$

$$P(X=5)=\frac{e^{-\lambda}\lambda^{5}}{\lfloor 5}=\frac{e^{-1.97}\times(1.97)^{5}}{120}=0.0345$$

$$P(X=6)=\frac{e^{-\lambda}\lambda^{6}}{\lfloor 6}=\frac{e^{-1.97}\times(1.97)^{6}}{720}=0.0113$$

$\therefore$ the expected frequencies are

$$N\times P(X=0)=500\times 0.1394=69.7\ \square\ 70$$

$$N\times P(X=1)=500\times 0.2746=137.3\ \square\ 138$$

$$N\times P(X=2)=500\times 0.2705=135.25\ \square\ 135$$

$$N\times P(X=3)=500\times 0.1776=88.8\ \square\ 89$$

$$N\times P(X=4)=500\times 0.0875=43.75\ \square\ 44$$

$$N\times P(X=5)=500\times 0.0345=17.25\ \square\ 18$$

$$N\times P(X=6)=500\times 0.0113=5.65\ \square\ 6$$

Total = 500

H_0 : There is no significant difference between observed and expected frequencies.

H_1 : The difference between observed and expected frequencies is significant.

Under H_0, the test statistic is $\chi^2 = \sum \frac{(O-E)^2}{E}$. To find χ^2

x	O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	56	70	- 14	196	2.8
1	156	138	18	324	2.35
2	132	135	- 3	9	0.067
3	92	89	3	9	0.1011
4	37	44	- 7	49	1.11
5	22	18	3	9	0.375
6	5	6			
Total	**500**	**500**			$\chi^2 = 6.8$

Number of degrees of freedom $r = n - 2 = 6 - 2 = 4$

From the table value of χ^2 table for 4 degrees of freedom and 5% level of significance table value is 9.48

Calculated value is less than table value.

Hence null hypothesis is accepted.

Thus Poisson fit is good.