

Day :- 6 :
*

Page No.

Date

* Chi Square Test *

Chi Square Test, this test claims population proportions.

Chi Square test is a non parametric test performed on categorical (normal or Ordinal data).
(Rank) \swarrow (Gaussian) \searrow

Q. In the 2000 U.S Census, the ages of Indians in a small town were found to be the following

✓	✓	✓
< 18.	18 - 35	> 35
20%	30%	50%

In 2010, ages of $n = 500$ individuals were sampled, Below are the results.

< 18	18 - 35	> 35
121	288	91

Using $\alpha = 0.05$, would you conclude that the population distribution of ages has changed in last 10 years?

		< 18	18 - 35	> 35
→	Expected	20%	30%	50%

$n = 500$ $\alpha = 0.05$		< 18	18 - 35	> 35	
	observed	121	288	91	} Huge difference.
	Expected	100	150	250	

i) null hypothesis

H_0 : The data meets the expected distribution

H_1 : Data don't meet the expected distribution

ii) $\alpha = 0.05$

C.I. = 95%

iii) Degree of freedom

$$df : n - 1$$

$$\therefore 3 - 1$$

$$\therefore \underline{2} \text{ (categories)}$$

($n = \text{categories}$)

iv) Decision Boundary \Rightarrow

Chi-Square Table
 $\underline{2} = \underline{5.991}$

v) Chi Square Test Stats.

$\chi^2 > 5.991$
Reject null

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

$$= \frac{(121 - 100)^2}{100} + \frac{(288 - 150)^2}{150} + \frac{(91 - 250)^2}{250}$$

O : observed
 E : expected

$$\chi^2 = \underline{\underline{232.494}}$$

$\chi^2 > 5.99$ (Reject the H_0)

Conclusion :

The data does not meet the distribution.

Q]

500 elementary School boys and girls are asked which is their favourite color : blue, green or pink. Result are shown below.

$$\underline{\underline{n-1}} = (3-1) = 2$$

$$\underline{\underline{n=1}} \\ \underline{\underline{(2-1)}} \\ \underline{\underline{=1}}$$

	Blue	Green	Pink
Boys	200	150	20
Girls	20	30	180
	120	180	200

Using $\alpha = 0.05$, conclude that there is relationship between gender and favourite color.

→ H_0 : The gender and favourite color are related.

H_1 : The gender and favourite color are not related.

$$\begin{aligned}
 Df &= (\text{rows} - 1) (\text{columns} - 1) \\
 &= (2 - 1) (3 - 1) \\
 &= (1) \times (2) \\
 &= \underline{\underline{2}}
 \end{aligned}$$

$$\text{Test Stats} = \underline{\underline{5.991}}$$

Formula
Chi-Square Test

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Observed	Blue	Green	Pink	
Boys	100	150	20	= 270
Girls	20	30	180	= 230
	120	180	200	= 500

Expected :- ~~Observed~~

Expected	Blue	Green	Pink	
Girls	55.2	92.8	92	230
Boys	64.8	97.2	108	270
	120	180	200	500

$$\begin{aligned}
 \text{Boy, Blue} &= \frac{(120 \times 270)}{500} = 64.8 \\
 \text{Green} &= \frac{(180 \times 270)}{500} = 97.2 \\
 \text{Pink} &= \frac{200 \times 270}{500} = 108
 \end{aligned}$$

$$J_e = \frac{J_e}{J_n} J_e \} \text{ expected formula.}$$

$$\therefore \chi^2 = \sum \frac{(J_o - J_e)^2}{J_e}$$

$$\therefore \frac{(100 - 64.8)^2}{64.8} + \frac{(150 - 97.2)^2}{97.2} + \frac{(20 - 108)^2}{108} +$$

$$\frac{(20 - 55.2)^2}{55.2} + \frac{(30 - 82.8)^2}{82.8} + \frac{(180 - 92)^2}{92}$$

$$= 259.77$$

$$\chi^2 = 259.77 > 5.991$$

Reject null hypothesis.
(The Gender and Jananite Color are not related).