STAT
LAB-08
Test for Goodness of Fit

Q1) For example, we collected wild tulips and found that the 71 were red, 40 were yellow and 29 were white.

1. Are these colors equally common? Find the expected values and residuals.

2. Suppose that, in the region where you collected the data, the ratio of red, yellow and white tulip is 4:3:1. Find the expected values and residuals?..

→1) Given Data:

O_1 (Red) = 71

O_2 (Yellow) = 40

O_3 (White) = 29

∴ Total (N) = 71 + 40 + 29 = 140

Now,

H_0: The colors are equally common (Proportions are $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$).

H_a: The colors are not equally common

If the colors are equally common, each color has an expected proportion of $\frac{1}{3}$. The expected frequencies are -

$E_1 = \frac{1}{3} \times 140 = 46.67$ (Red)

$E_2 = \frac{1}{3} \times 140 = 46.67$ (Yellow)

$E_3 = \frac{1}{3} \times 140 = 46.67$ (White)

The residuals are the differences between observed and expected values:

$$\text{Residual} = O - E$$

~~Residuals = [71 - 46.67, 44]~~

$$\text{Residuals} = [71 - 46.67, 40 - 46.67, 29 - 46.67]$$

$$= [24.33, -6.67, -17.67]$$

$\therefore$ The formula for chi-square statistic is

$$\chi^2 = \sum_1 \frac{(O-E)^2}{E}$$

Now,

$$\chi^2_1 = \frac{(71 - 46.67)^2}{46.67} = \frac{(24.33)^2}{46.67} = 12.69$$

$$\chi^2_2 = \frac{(40 - 46.67)^2}{46.67} = \frac{(-6.67)^2}{46.67} = 0.95$$

$$\chi^2 = \frac{(29 - 46.67)^2}{46.67} = \frac{(-17.67)^2}{46.67} = 6.69$$

$$\text{Total } (\chi^2) = 12.69 + 0.95 + 6.69 = 20.33$$

2) H_0: The proportions follow the 4:3:1 ratio.

H_a: The proportions do not follow the 4:3:1 ratio.

Now,

In a 4:3:1 ratio, the total number of parts

$= 4 + 3 + 1 = 8.$

∴ The expected proportions are:

$$Red = \frac{4}{8} = 0.5$$
$$Yellow = \frac{3}{8} = 0.375$$
$$White = \frac{1}{8} = 0.125$$

with $N = 140$, the expected values are:

$$E_1 = 0.5 \times 140 = 70$$
$$E_2 = 0.375 \times 140 = 52.5$$
$$E_3 = 0.125 \times 140 = 17.5$$

∴ Residuals $= [71-70, 40-52.5, 29-17.5]$
$= [1.0, -12.5, 11.5]$

Now,
For chi-square statistic (χ^2)

$$\chi^2_1 = \frac{(71-70)^2}{70} = \frac{1^2}{70} = 0.014$$
$$\chi^2_2 = \frac{(40-52.5)^2}{52.5} = \frac{(-12.5)^2}{52.5} = 2.976$$
$$\chi^2_3 = \frac{(29-17.5)^2}{17.5} = \frac{(11.5)^2}{17.5} = 7.558$$

Total $(\chi^2) = 0.014 + 2.976 + 7.558$
$= 10.55$

Now,
The degrees of freedom are calculated as:
$df = (\text{No. of categories}) - 1$
$= 3 - 1 = 2$

1) Ans → $\chi^2 = 20.33$

Residuals = [24.33, -6.67, -17.67]

Expected values = [46.67, 46.67, 46.67]

2) Ans →

$\chi^2 = 10.55$

Residuals = [1.0, -12.5, 11.5]

Expected values = [70.0, 52.5, 17.5]

Q2> Aumi is testing an octahedral die to see if it is biased. The observed results are given:

observed	7	10	11	9	12	10	14	7

Test the hypothesis that the die is fair.

→ H_0: The die is fair, meaning each face has an equal probability of being rolled ($p = \frac{1}{8}$)

H_a: The die is not fair, meaning atleast one face does not follow the expected probability.

Observed values(O) = [7, 10, 11, 9, 12, 10, 14, 7]

Total Rolls(N) = 7+10+11+9+12+10+14+7

= 80

Expected value (E) for each face (if the die is fair).

$$\cancel{\boxed{E \times \frac{1}{8}}} \quad E = \frac{1}{8} \times N$$

$$= \frac{1}{8} \times 80 = 10$$

So, the expected values for all faces are:

$$E = [10, 10, 10, 10, 10, 10, 10, 10]$$

Now,

For chi-square statistic (χ^2)

The formula for the chi-square statistic is

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

For each face, we calculate the contribution to χ^2.

Face 1,

$$\frac{(7-10)^2}{10} = \frac{(-3)^2}{10} = \frac{9}{10} = 0.9$$

Face 2,

$$\frac{(10-10)^2}{10} = 0$$

Face 3,

$$\frac{(11-10)^2}{10} = 0.1$$

Face 4,

$$\frac{(9-10)^2}{10} = 0.1$$

Face 5,

$$\frac{(12-10)^2}{10} = 0.4$$

Case 6,

$$\frac{(10-10)^2}{10} = 0$$

Case 7,

$$\frac{(14-10)^2}{10} = 1.6$$

Case 8,

$$\frac{(7-10)^2}{10} = 0.9$$

Now,
Adding all the cases, we get,

$$\chi^2 = 0.9 + 0 + 0.1 + 0.4 + 0 + 1.6 + 0.9$$
$$= 4.0$$

The degrees of freedom (df) are = No. of categories - 1
$$= 8 - 1$$
$$= 7$$

Now,
Using a chi-square table $\alpha = 0.05$ and $df = 7$, the critical value is approximately

$$\chi^2_{critical} = 14.07$$

Comparing χ^2 with $\chi^2_{critical}$ we get,

$$\chi^2 = 4.0 < 14.07$$

Since χ^2 is less than the critical value, we fail to reject the null hypothesis (H_0) at the 0.05 significance level.

Hence the data do not provide sufficient evidence to conclude that die is biased. It appears to be fair.

Q2> When randomly selecting a card from a deck with replacement, are we equally likely to select a heart, diamond, spade and club? I randomly selected a card from a deck 40 times with replacement. I pulled 13 hearts, 8 diamonds, 8 spades and 11 clubs.

-> Ho: The cards are equally likely to be hearts, diamonds, spades or clubs. The expected proportion for each suit is $\frac{1}{4}$.

Ha: The cards are not equally likely to be hearts diamonds spades or clubs

observed values (O): [13 hearts, 8 diamonds, 8 spades, 11 clubs]

Total No. of Cards Drawn (N) = 13 + 8 + 8 + 11
= 40

: Expected Value (E) for each suit (if equally likely):

$$E = \frac{1}{4} \times N = \frac{1}{4} \times 40 = 10$$

So, the expected values for all suits are (E) = [10, 10, 10, 10]

: The formula for chi-square statistic is

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

Now, we calculate the contribution to χ^2 for each suit

$$\text{Hearts} = \frac{(13-10)^2}{10} = \frac{(3)^2}{10} = 0.9$$

$$\text{Diamonds} = \frac{(8-10)^2}{10} = \frac{4}{10} = 0.4$$

$$\text{Spades} = \frac{(8-10)^2}{10} = \frac{4}{10} = 0.4$$

$$\text{clubs} = \frac{(11-10)^2}{10} = \frac{1}{10} = 0.1$$

Adding these contributions, we get,

$$\chi^2 = 0.9 + 0.4 + 0.4 + 0.1$$
$$= 1.8$$

$\therefore df =$ No. of categories $-1 = 4-1 = 3$

Now,
Using a Chi-square table for $\alpha = 0.05$ and $df = 3$, the critical value is approximately

$$\chi^2_{critical} = 7.815$$

Comparing χ^2 with $\chi^2_{critical}$ we get

$$\chi^2 = 1.8 < 7.815$$

Since, χ^2 is less than the critical value, we fail to reject the null hypothesis (H_0) at the 0.05 significance level.