## Goodness of fit test :-

The square of the standard normal variate is known as chi-square variate with 1 degrees of freedom (d.f.)

If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

$$\chi^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \text{ is a chi square variate with 1 d.f.}$$

In general, If $X_i \sim N(\mu_i, \sigma_i^2)$ then

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 \text{ is a chi-square variate with } n \text{ d.f.}$$

→ We have seen many times, the results obtained in samples donot always agree with the theoretical results.

E.g. If we toss a coin (fair coin) 100 times, then we expect 50 heads and 50 tails

$$\left(\because E(x) = np = 100 \times \frac{1}{2} = 50\right)$$

But It is not always true in practical situation.

→ To measure the significance of the discrepancy b/w observed ( ) and expected ( ) frequencies, we use $\chi^2$- goodness of fit test.

→ If $O_i$ $(i=1,2 \ldots n)$ be the observed frequencies and $E_i$ $(i=1,2\ldots n)$ be the expected frequencies

then $$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where $\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E$ follows chi-square distⁿ with $(n-1)$ d.f.

↳ If $\chi^2 = 0$ then observed and Expected frequencies agree exactly.

→ If $\chi^2 > 0$ then they do not agree exactly.

→ The larger the value of $\chi^2$, the greater is the discrepancy b/w the observed and the expected frequencies.

Condition for $\chi^2$-test

(1) The sample observations should be Independent.

(2) $\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i$

(3) The total frequency $\geq 50$. i.e. $\sum O_i \geq 50$

(4) No theoretical $(E_i)$ frequency should be less than 5.

Steps:- (1) Define the hypothesis $(H_0, H_1)$

(2) Calculate E and $\chi^2$

$E = Np$ where N – Total frequency
p – Prob.

(3) Conclude the result.

Que. The no. of scooter accidents per month in a Certain town were as follows:-

12 8 20 2 14 10 15 6 9 4

Are these frequencies in agreement with the belief that accident Conditions are same during this 10 month period.

Soln Step I

$H_0$:- accident Conditions are same during 10 month period.

$H_1$: accident Conditions are not same.

Step II $\quad \Sigma O = 100$

i.e. $N = 100$ — Total frequency.

Under $H_0$: accidents are uniformly distributed over the given period.

So $p = \frac{1}{10}$.

$\therefore E = Np = 100 \times \frac{1}{10} = 10$

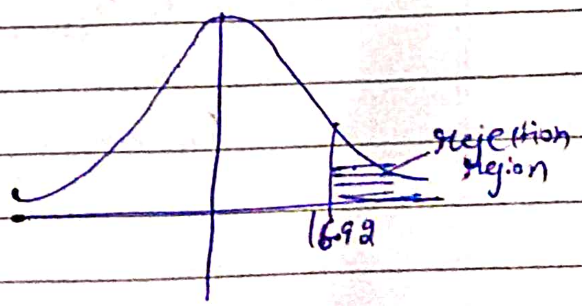
$\therefore \Sigma O = \Sigma E$

| Months | O | E | $(O-E)^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| 1 | 12 | 10 | 4 | 0.4 |
| 2 | 8 | 10 | 4 | 0.4 |
| 3 | 20 | 10 | 100 | 10 |
| 4 | 2 | 10 | 64 | 6.4 |
| 5 | 14 | 10 | 16 | 1.6 |
| 6 | 10 | 10 | 0 | 0 |
| 7 | 15 | 10 | 25 | 2.5 |
| 8 | 6 | 10 | 16 | 1.6 |
| 9 | 9 | 10 | 1 | 0.1 |
| 10 | 4 | 10 | 36 | 3.6 |
| | 100 | 100 | | 26.6 |

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 26.6$$

Step III $\quad d.f = n-1 = 9$

$\chi^2_{(0.05)}(9\ d.f.) = 16.92$




$\chi^2_{Stat} > \chi^2_{0.05}(9)$

$\Rightarrow$ reject $H_0$.

Date ______
Page ______

**Que** The demand for a particular spare part in a factory was found to be vary from day to day. In a sample study, following information was obtained.

| Days | M. | T | W | Th | F | S |
|---|---|---|---|---|---|---|
| No. of parts demanded | 1124 | 1125 | 1110 | 1120 | 1126 | 1115 |

Test the hyp. that the no. of parts demanded does not depend on the day of the week.

**Sol$^n$** Step I $H_0$: demand does not depend on the day of the week.

$H_1$: It depends on the day of the week.

Step II

$N = 6720$

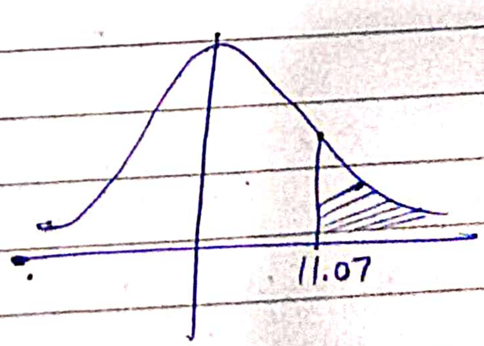
$p = \frac{1}{6}$

$E = Np = \frac{6720}{6} = 1120$

| Days | O | E | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|---|
| M | 1124 | 1120 | 16 | 0.014 |
| T | 1125 | 1120 | 25 | 0.022 |
| W | 1110 | 1120 | 100 | 0.089 |
| Th | 1120 | 1120 | 0 | 0 |
| F | 1126 | 1120 | 36 | 0.032 |
| S | 1115 | 1120 | 25 | 0.022 |
| Total | 6720 | 6720 | | 0.179 |

$\boxed{\chi^2_{Stat} = 0.179}$

$\chi^2_{(6-1)}(0.05) = 11.07$




$\chi^2_{Stat} < \chi^2_{(0.05)}(5)$

$\Rightarrow$ Do not reject $H_0$.

Que A Survey of 800 families with four children each revealed the following distⁿ

| | | | | | |
|---|---|---|---|---|---|
| no. of boys | 0 | 1 | 2 | 3 | 4 |
| no. of girls | 4 | 3 | 2 | 1 | 0 |
| no. of families | 32 | 178 | 290 | 236 | 64 |

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

Solⁿ → $H_0$: equal prob. of male and female births.

∴ Prob of male birth $p = \frac{1}{2}$

$E = N \cdot p$

Prob. of $x$ male births in family $= {}^4C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}$; $x = 0, 1, 2, 3, 4$

$P(X=x) = {}^4C_x \left(\frac{1}{2}\right)^4$

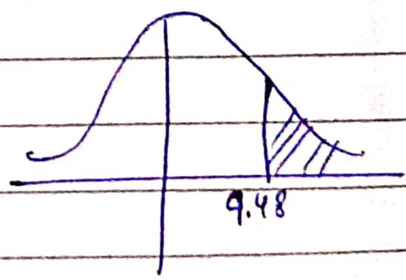
| No. of males | O | E | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|---|
| 0 | 32 | 50 | 324 | 6.48 |
| 1 | 178 | 200 | 484 | 2.42 |
| 2 | 290 | 300 | 100 | 0.33 |
| 3 | 236 | 200 | 1296 | 6.48 |
| 4 | 64 | 50 | 196 | 3.92 |
| | 800 | 800 | | 19.63 |

$E = N \cdot p = 800 \times P(X=x)$

$= 800 \times \left({}^4C_x\right) \left(\frac{1}{2}\right)^4 = 50 \times \left({}^4C_x\right)$

$\chi^2_{Stat} = 19.63$

$\chi^2_{0.05}(4) = 9.48$




∵ $\chi^2_{Stat} > \chi^2_{0.05}(4)$

⇒ Reject $H_0$.

⇒ Male and female births are not equally probable.