Chi Square (χ^2) Test

Basics, formula & conditions

- Various significance tests such as Z-test, t-test or F-test were based on the assumption that the samples were drawn from the normally distributed population.

Parametric Tests :- Testing procedure requires the assumption about type or parameters of the population.

e.g: Z-test, t-test, F-test etc.

Non-parametric Tests:- It is applied, when no exact information is available about the population; whether population distribution is binomial, poission or normal.

It is also considered as "distribution free testing"

- χ^2 test is commonly used non-parametric test.

χ^2 - Test:

- The quantity χ^2 describes the magnitude of the descripancy between theoretical & observed values.

FORMULA :-

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

∴ O = Observed frequency.
E = Expected "

$$E = \frac{RT \times CT}{N}$$

∴ RT = The row total for all rows
CT = The column total
N = Total no. of observations.

$\nu = (r-1)(c-1)$

ν = degree of freedom
r = Rows & c = No. of Colu

Conditions for applying χ^2 test:

- Each cell should contain at least 5 observations (Generally preffered 10 observations), because if it is less than 5 then χ^2 will be overestimated which leads to the rejection of Null hypothesis (H_0)
- All individual observations should be independent & completely random.
- The total sample size should be atleast 50 observations i.e., $N \geq 50$
- The data should be expressed in original units. It should not be expressed in %age or ratio.

APPLICATIONS

1:- Test for independence of Attributes:

- With the help of χ^2 test, we can find out whether two or more attributes are associated or not.

2: χ^2 test as goodness of fit:

- On various Ocassions, the decision maker needs to understand whether an actual sample distribution matches or coincides with a known probability distribution such as poisson, binomial or normal.

- The χ^2 test for goodness of fit enable us to determine the extent to which the theoretical probability distributions coincides with empirical sample distribution.

③ χ^2 test for Yate's correction for continuity:

- The distribution of χ^2 test statistics is continuous but the data under the test is categorical which is discrete.

- It causes error due to the discrete data & if it is a 2x2 contingency table then we can apply yate's correction for continuity.

④ χ^2 - test for population variance:

- This is considered as a parametric test.
- The assumption underlying the χ^2 test is that the population from which the samples are drawn is normally distributed.

$$\chi^2 = \frac{\delta_s^2}{\delta_p^2} \times (n-1)$$

$\therefore$ δ_s = Variance of the sample.
δ_p = Variance of the normally distributed population.
n = Sample size.

$\nu = n-1$.

⑤ Test for homogeneity:

- This is useful in a case when we intend to verify whether several populations are homogenous with respect to some characteristics of interest.

Scanned with CamScanner

QUESTION No: 01

Sol: H_0: Quinine = Not effective in testing for malaria.

H_a: " ≠ Not " " "

Given data

	Treatment	Fever	Expected Value 1	No fever	Expected Value 2	Total
1	Quinine	20	30	480	470	500
2	No Quinine	100	90	1400	1410	1500
	Total	120		1880		2000

$$\chi^2 = \frac{\sum (O-E)^2}{E} \; ; \quad E = \frac{RT \times CT}{N} \; ; \quad N = 2000$$

$$E_{11} = \frac{500 \times 120}{2000} = 30 \; ; \quad E_{12} = \frac{1500 \times 120}{2000} = 90$$

$$E_{21} = \frac{500 \times 1880}{2000} = 470 \; ; \quad E_{22} = \frac{1500 \times 1880}{2000} = 1410$$

- Calculation of χ^2

O	E	(O-E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
20	30	-10	100	3.33
100	90	10	100	.
480	470	10	100	0.21
1400	1410	-10	100	0.07

$$\frac{\sum (O-E)^2}{E} = 4.72$$

$$\boxed{\chi^2 = 4.72}$$ Calculated.

$\nu = (r-1)(c-1)$

$\nu = (2-1)(2-1) = 1$ (Degree of Freedom)

Level of significance = 5%

$\chi^2_{0.05} = 3.84$

H_0 is failed & rejected, Hence, Quinine is useful in testing for malaria.

QUESTION NO: 02

Sol:

H_0: Drug = Placebo

H_a: Drug $\neq$ Placebo

Treatment	Helped	E_1	Reaction	E_2	No effect	E_3	Total
Drug x	150	140	30	35	70	75	250
Placebo	130	140	40	35	80	75	250
Total	280		70		150		500

$$\chi^2 = \frac{\sum (O-E)^2}{E} \; ; \quad E = \frac{RT \times CT}{N}$$

$$E_{11} = \frac{25\cancel{0} \times 28\cancel{0}}{5\cancel{00}} = 140 \; ; \quad E_{21} = \frac{25\cancel{0} \times 7\cancel{0}}{5\cancel{00}} = 35$$

$$E_{31} = \frac{25\cancel{0} \times 15\cancel{0}}{50\cancel{0}} = 75$$

Calculation of χ^2

O	E	$(O-E)$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
150	140	10	100	0.714
130	140	-10	100	0.714
30	35	-5	25	0.714
40	35	5	25	0.714
70	75	-5	25	0.333
80	75	5	25	0.333

$\Sigma = 3.522$

$$\chi^2 = \frac{\Sigma (O-E)^2}{E} = 3.522$$

$$\nu = (r-1)(c-1) = (2-1)(3-1) = 2$$

$$\chi^2_{0.05} = 5.99$$

H_0 is passed and accepted. Hence, no significant difference in the effect of drug x & placebo.

χ^2-TEST

- **KEY POINTS**

1. Chi-square test is a statistical test used to determine if there is a significant difference between the expected frequencies & the observed frequencies in one or more categories.

2. It is commonly used for categorical data such as comparing the distribution of a certain trait within a population to a known distribution.

3. It is a non-parametric test.

4. It is used to test hypotheses about independence homogeneity, and goodness of fit.

5. The p-value is used to determine the level of significance, which tells us the probability of observing the test statistic by chance if the null hypothesis is true.

6. Lower p-value implies stronger evidence against the null hypothesis.

7. The test is sensitive to the sample size, the larger the sample size, the smaller the p-value.