Population :– An aggregate of objects under study is called population or universe.

Finite Universe [containing finite no. of individuals]

Infinite Universe. [having infinite no. of members.]

SAMPLING :–

A finite sub set of a universe is called a sample. The number of individuals in a sample is called sample size.

The process of selecting a sample from a universe is called sampling.

The theory of sampling is a study of relationship existing between a population and samples drawn from the population.

The fundamental object of sampling is to get as much information as possible of the whole universe by examining only a part of it.

Sampling is quite open used in our day to day practical life. For example, in a shop we asses the quality of sugar, rice or any other commodity by taking only a handful of it from the bag and then decide whether to purchase it or not.

* Parameters of statistics :—

The statistical constants such as mean, variance etc. are known as parameters.

The statistical concepts of the sample to estimate the parameters of the population from which the sample has been drawn is known as statistic.

Population mean and variance are denoted by μ and σ^2.

Sample mean and variance are denoted by $\bar{x}$ and s^2.

* Standard Error :—

The standard deviation of the sampling distribution of a statistic is known as the standard error.

* Test of Significance :—

An important aspect of sampling theory is to study the test of significance which will enable us to decide, on the basis of the results of the sample, whether

(i) the deviation between the observed sample statistic and the hypothetical parameter value

OR

(ii) the deviation b/w two sample statistics is significant or might be attributed due to chance or the fluctuations of the sampling.

TESTING OF STATISTICAL HYPOTHESIS :—

STEP 1. NULL HYPOTHESIS :—

Null hypothesis is a definite statement about the population parameter, it is denoted by H_0.

It is the hypothesis, which is tested for possible rejection under the assumption that it is true.

STEP 2. ALTERNATIVE HYPOTHESIS :—

Any hypothesis which is complementary to the null hypothesis (H_0) is called an alternate hypothesis. It is denoted by H_1.

For Example, we want to test the null hypothesis that population has a specified mean μ_0,

i.e $\qquad H_0 : \mu = \mu_0$

Then possible cases for alternate hypothesis will be

(i) $H_1 : \mu \neq \mu_0$ [Two tailed alternate hypo.

(ii) $H_1 : \mu > \mu_0$ [Right tailed alter. hypo.] Single tailed.

(iii) $H_1 : \mu < \mu_0$ [Left tailed alter. hypo.]

STEP 3. LEVEL OF SIGNIFICANCE :—

A region corresponding to a statistic t which amounts to rejection of null hypothesis H_0 is called as critical region [or region of rejection] while the region which amounts to acceptance of H_0 is called acceptance region.

The probability α that a random ~~variable~~ value of statistic t belongs to critical region is known as the level of Singnificance.

④

STEP 4: TEST STATISTIC [OR TEST CRITERION]

We find the test statistic z under the null Hypothesis.

For large sample, the variable $\boxed{z = \frac{t - E(t)}{S.E.(t)}}$ is normally distributed with mean 0 and variance 1.

STEP 5: CONCLUSION :—

We compare the computed value of z with the critical value z_α at level of significance (α).

⇒ The critical value of z_α of the test statistic at level of significance α for a two tailed test is given by

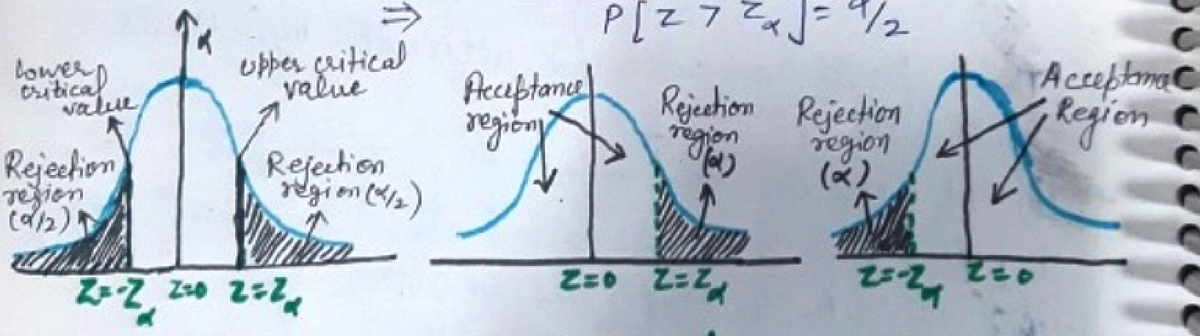
$$P[\,|z| > z_\alpha] = \alpha \quad \text{—①}$$

Since the normal curve is symmetrical, then

$$P[z > z_\alpha] + P[z < -z_\alpha] = \alpha$$

$$\Rightarrow 2P[z > z_\alpha] = \alpha$$

$$\Rightarrow P[z > z_\alpha] = \alpha/2$$



Two tailed Test | Right tailed Test | Left tailed Test

The critical value of z for a single tailed test at level of significance α is same as the critical value of z for two tailed test at level of significance 2α.

Using the normal tables, the critical value of z at different levels of significance (α) are listed below –

Level of Significance

	1% (0.01)	5% (0.05)	10% (0.1)
Two tailed Test	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 1.96$	$\lvert z_\alpha \rvert = 1.645$
Right tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Left tailed test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

If $|z| > z_\alpha$, we reject H_0 and conclude that there is [redacted] significant difference.
If $|z| < z_\alpha$, we accept H_0 and conclude that there is no significant difference.

Errors in Sampling:—

In Sampling we may commit the following two types of errors –

(i) Type I Error :–

When H_0 is true, we may reject it

$$P[\text{Reject } H_0, \text{ when it is true}] = \alpha$$

α is known as size of the error and also reffered as producer's risk.

(ii) Type II Error :–

When H_0 is wrong, we may accept it.

$$P[\text{Accept } H_0, \text{ when it is wrong}] = \beta$$

β is known as size of type II error, also reffered to as consumer's risk.

Testing of Significance for Single Proportion :—

This test is used to find significant difference between proportion of the sample and the population.

let X = No. of Successes in n independent trials with probab. of success P

Then $E(X) = nP$

$V(X) = nPQ$, $Q = 1 - P$ = Probability of failure

let $p = \frac{X}{n}$ be the observed proportion of Succs.

$$E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{nP}{n} = P$$

$$V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{nPQ}{n^2} = \frac{PQ}{n}$$

Then, Standard error $= S.E.(p) = \sqrt{V(p)}$

$$= \sqrt{\frac{PQ}{n}}$$

Then

$$z = \frac{p - E(p)}{S.E.(p)} = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

this z is called test statistic which is used to test the significant difference of sample and population proportion.

⑦

Example :— A coin was tossed 400 times and the head turned up 216 times. Then test the hypothesis that coin is unbiased.

Solution :— Null Hypothesis :

H_0 : The coin is unbiased i.e. $P = \frac{1}{2}$

Alternate Hypothesis :

H_1 : The coin is biased i.e. $P \neq \frac{1}{2}$

Hence we use here two tailed test

Given that, $n = 400$

X = no. of success = 260

$\therefore p$ = proportion of success in the sample

$= \frac{X}{n} = \frac{216}{400} = 0.54$

P = population proportion = 0.5

$Q = 1 - P = 0.5$

Test Statistic :—

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \Rightarrow |z| = \left| \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} \right|$$

$$= 1.6$$

Conclusion :—

Since $|z| = 1.6 < 1.96$

i.e $|z| < z_\alpha$, where z_α is the significant value of z at 5% level of significance.

Hence, H_0 is accepted and we conclude that the coin is unbiased.

Example:– A manufacturer claims that only 4% of his products supplied by him are defective. A random sample of 600 products contains 36 defectives. Test the claim of manufacturer.

Solution:– $n = 600$, $X = 36$

p = proportion of defective in the sample

$$= \frac{X}{n} = \frac{36}{600} = 0.06$$

P = proportion of defectives in the population

$= 0.04$ [As claimed by the Manufacturer]

$$Q = 1 - P$$
$$= 1 - 0.04 = 0.96$$

Now, Null Hypothesis:

H_0: $P = 0.04$ is true

i.e. the claim is accepted.

Alternate Hypothesis:

H_1: $P \neq 0.04$

Hence, we use two tailed test.

Test statistic:

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.06 - 0.04}{\sqrt{\frac{(0.04) \times (0.96)}{600}}}$$

$$Z = 2.5$$

Conclusion:– Since $|Z| = 2.5 > 1.96$

$\Rightarrow$ Null hypothesis H_0 is rejected at 5% level of significance.

$\Rightarrow$ Hence Manufacturer's claim is not acceptable.

Example:- A machine is producing bolts of which a certain fraction is defective. A random sample of 400 is taken from a large batch and is found to contain 30 defective bolts. Does this indicate that the proportion of defectives is larger than that claimed by the manufacturer where the manufacturer claims that only 5% of his products are defective. Find 95% confidence limits of the proportion of defective bolts in batch.

Solution :- $n=400, \quad x=30$

$p = \frac{x}{n} = \frac{30}{400} = 0.075$

$P = 0.05$

$Q = 0.95$

Null hypothesis: $H_0: P = 0.05$ i.e. Manufacturer's claim is accepted

Alternate hypothesis:

$H_1: P > 0.05$

Hence we use right tailed test.

Test statistic: $z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.075-0.05}{\sqrt{\frac{0.05 \times 0.95}{400}}} = 2.29$

Conclusion: $|z| < Z_\alpha = 1.645$

$\Rightarrow$ H_0 is rejected at 5% level of significance.

$\Rightarrow$ Manufacturer's claim is not accepted.

To find 95% Confidence limits of the proportion –

The probable limit in the population are

$$\boxed{P \pm Z_\alpha \sqrt{\frac{PQ}{n}}}$$

$\Rightarrow \quad 0.05 \pm 1.645 \sqrt{\frac{0.05 \times 0.95}{400}}$

$\Rightarrow 0.06089, \; 0.03910$

Testing of Significance for difference of Proportions:—

Consider two samples X_1 and X_2 of sizes n_1 and n_2 resp. taken from two different populations. To test the significance of the difference b/w the sample proportions p_1 and p_2, the test statistic is

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \text{ where } P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

and $Q = 1-P$

Example:— A machine produced 16 defectives articles in a batch of 500. After overhauling it produced 3 defectives in a batch of 100. Has the machine improved?

solution:— $n_1 = 500$, $p_1 = \frac{X_1}{n_1} = \frac{16}{500} = 0.032$

$n_2 = 100$, $p_2 = \frac{X_2}{n_2} = \frac{3}{100} = 0.03$

$$P = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{19}{600} = 0.032$$

$Q = 1 - P$

Null Hypothesis:—

H_0: The machine has not improved due to overhauling

i.e. $p_1 = p_2$

Alternate Hypothesis:— H_1: $p_1 > p_2$

Hence we use right tailed test.

Test statistic: $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 0.104$

Conclusion :—

$|Z| = 0.104 < Z_\alpha = 1.645$, at 5% level of significance.

$\Rightarrow$ Null hypothesis accepted.

$\Rightarrow$ Machine has not improved due to overhauling.

Example :— Before an increase in excise duty on tea, 800 people out of a sample of 1000 persons were found to be tea drinkers. After an increase in the duty, 800 persons were known to be tea drinkers in a sample of 1200 people. Do you think that there has been a significant difference [decrease] in the consumption of tea after the increase in excise duty ?

Sol :— $n_1 = 1000$ $\quad$ $n_2 = 1200$

$x_1 = 800$ $\quad$ $x_2 = 800$

$p_1 = \frac{x_1}{n_1} = \frac{800}{1000} = \frac{4}{5}$, $\quad$ $p_2 = \frac{x_2}{n_2} = \frac{800}{1200} = \frac{2}{3}$

$$P = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{8}{11}, \quad Q = 1 - P = \frac{3}{11}$$

Null hypothesis : $H_0 : p_1 = p_2$

i.e. No significant diff in the consumption of tea

Alternate Hypothesis: $H_1 : p_1 > p_2$, $\quad$ Right tailed test

Test statistic :

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 6.842$$

Conclusion :— $|z| = 6.842 > Z_\alpha = 1.645$ at 5% level of significance, Hence H_0 is rejected.

$\Rightarrow$ there is a significant decrease in the consumption of tea due to increase in excise duty.

TESTING OF SIGNIFICANCE FOR SINGLE MEAN

Sample mean $= \bar{x}$
variance $= s^2$

Population
Mean $= \mu$
Variance $= \sigma^2$

let $x_1, x_2 \ldots x_n$ be a random sample of size n from a large population $X_1, X_2 \ldots X_N$ of size N, with mean μ and variance σ^2.

then the test statistic for testing of significance of single mean :—

$$\boxed{Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

Note :- If σ is not given, then we use

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

, where s is the standard deviation of the sample.

Ex :- The average marks in mathematics of a sample of 100 students was 51 with a standard deviation of 6 marks. Could this have been a random sample from a population with averge marks 50.?

Solution — $n = 100$, $\bar{x} = 51$, $s = 6$ ⇒ information about sample

$\mu = 50$, σ = not given ← Population

Null Hypo: H_0 : Sample is drawn from a population with mean 50 i.e. $\mu = 50$

Alternate Hypo: H_1 : $\boxed{\mu \neq 50}$ Two tailed test

Test Statistic: —
$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{51-50}{6/\sqrt{100}} = 1.666$$

Conclusion: — Since $|z| = 1.666 < z_\alpha = 1.96$

z_α is significant value at 5% level of significance.

⇒ H_0 is Accepted.

⇒ Sample is drawn from a population with mean 50.

Example: — A normal population has a mean 6.8 and S.D. of 1.5. A sample of 400 gave a mean of 6.75. Is this difference significant?

Solution: — $n = 400$, $\bar{x} = 6.75$ — Sample Information.

$\mu = 6.8$, $\sigma = 1.5$ ⇒ Population info

H_0: there is no significant diff. i.e $\boxed{\bar{x} = \mu}$

H_1 : $\boxed{\bar{x} \neq \mu}$ i.e. there is significant diff b/w means.

Test statistic: —
$$|z| = \left|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right| = \left|\frac{6.75 - 6.8}{1.5\sqrt{400}}\right| = 0.67$$

Conclusion: — $|z| = 0.67 < z_\alpha = 1.96$, at 5% level of significance

⇒ Null Hypo. accepted

⇒ there is no significant diff b/w $\bar{x}$ & μ.

Test of Significance for difference of means of two large samples

Let $\bar{x}_1$ be the mean of a sample of size n_1 from a population with mean μ_1 & variance σ_1^2.
And let $\bar{x}_1$ be the mean of a sample of size n_2 from another population with mean μ_2 & variance σ_2^2.
Then test statistic is given by

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

Note :- 1 :- If the samples are drawn from the same population where $\sigma_1 = \sigma_2 = \sigma$, then Test statistic is given by

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

(2) If σ_1 & σ_2 are not given & $\sigma_1 \neq \sigma_2$ then

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}$$

Where s_1 and s_2 are S.D. of samples.

(3) If σ_1 & σ_2 are not given but $\sigma_1 = \sigma_2$, then

$$\boxed{\sigma = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}$$

Example:- A random sample of 200 villages from Mathura district gives the mean population per village at 485 with a S.D. of 50. Another random sample of the same size from the same district gives the mean population per village at 510 with a S.D of 40. Is the diff. b/w the mean values statistically significant?

Solution :-

$n_1 = 200$ $\qquad\qquad$ $n_2 = 200$

$\bar{x}_1 = 485$ $\qquad\qquad$ $\bar{x}_2 = 510$

$s_1 = 50$ $\qquad\qquad$ $s_2 = 40$

Null Hypo: H_0: there is no significant diff. b/w mean values

i.e, $\bar{x}_1 = \bar{x}_2$

Alternate Hypothesis: H_1 : $\bar{x}_1 \neq \bar{x}_2$

$\Rightarrow$ two tailed test

Test statistic -

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{485 - 510}{\sqrt{\frac{(50)^2}{200} + \frac{(40)^2}{200}}}$$

$z = -5.52$

$|z| = 5.52$

Conclusion :- $|z| = 5.52 > z_\alpha = 1.96$ at 5% level of significance, H_0 is rejected

$\Rightarrow$ there is significant difference b/w the mean values of the two samples.

TEST OF SIGNIFICANCE OF SMALL SAMPLES

When the size of the sample is less than 30, then the sample is called small sample.

t-Test of Significance of the mean of a random sample :-

To test whether there is a significant difference between the sample mean $\bar{x}$ & the population mean μ,

we use the test statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$ with degree of freedom $(n-1)$.

$\Rightarrow$ At given level of significance α and degree of freedom $(n-1)$, we refer to t-table (t_α).

If calculated value of $|t| < t_\alpha$ then Null hypothesis is Accepted.

And if $|t| > t_\alpha$, H_0 is rejected.

Example:- A random sample of size 16 has 53 as mean. The sum of squares of the deviation from mean is 135. Can this sample be regarded as taken from the population having 56 as mean.

Solution

$$n = 16 \quad , \quad \mu = 56$$
$$\bar{x} = 53 \quad , \quad \Sigma(x-\bar{x})^2 = 135$$

Null Hypothesis:- No significant diff b/w sample mean & population mean

i.e. $H_0 : \mu = 56$

Alternate Hypothesis:- $H_1 : \mu \neq 56$ (Two tailed test)

Test statistic:-

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

$$t = \frac{53-56}{3/\sqrt{16}} \quad , \quad s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$= -4 \qquad\qquad = \sqrt{\frac{135}{15}} = 3$$

$|t| = 4$, with degree of freedom 15

Conclusion:- Since $|t| = 4 > t_{0.05} = 2.13$,

$\Rightarrow$ Null Hypothesis rejected.

$\Rightarrow$ Sample is not taken from the population with mean 56.

t-Test for difference of means of two small samples:—

Population I: Size $= N_1$, Mean $= \mu_1$, $S.D = \sigma_1 = \sigma$

$\Downarrow$

Sample I: Size $= n_1$, mean $= \bar{x}$, $S.D = s_1$

Population II: Size $= N_2$, Mean $= \mu_2$, $S.D = \sigma_2 = \sigma$

$\Downarrow$

Sample II: Size $= n_2$, Mean $= \bar{y}$, $S.D = s_2$

Then the test statistic is given by

$$\boxed{t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

with degree of freedom $(n_1 + n_2 - 2)$

where $s = \sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$

Example :- Two Samples of sodium bulbs were tested for length of life and the following results were obtained—

	Size	Sample Mean	Sample S.D
Sample I	8	1234 Hrs	36 Hrs
Sample II	7	1036 Hrs	40 Hrs

Is the difference in the means significant to generalise that Sample I is superior to Sample II, regarding length of life.

Solution:-

$n_1 = 8$ $\quad$ $n_2 = 7$

$\bar{x} = 1234$ $\quad$ $\bar{y} = 1036$

$s_1 = 36$ $\quad$ $s_2 = 40$

thus $s = \sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = 40.73$

& degree of freedom $= n_1 + n_2 - 2 = 13$.

Null Hypothesis:-

$H_0 : \mu_1 = \mu_2$ i.e. both type bulbs have same life time

Alternate Hypo: $H_1 : \mu_1 > \mu_2$ [Single tailed]

i.e. Sample I is superior to Sample II

Test statistic:

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1234 - 1036}{40.73\sqrt{\frac{1}{8} + \frac{1}{7}}}$$

$$= 18.14$$

but $t_{0.05}$ at degree of freedom $13 = 1.77$

Conclusion:- Since $|t| = 18.14 > t_{0.05} = 1.77$

$\Rightarrow$ Null hypo. H_0 rejected.

$\Rightarrow$ Sample I is superior than Sample II.

NOTE:- The significant value of t at level of significance α, for a single tailed test can be found from those of two tailed test by referring to the value at 2α.

CHI-SQUARE TEST [χ^2-test]

If we toss a coin 400 times, then theoretically we may expect 200 heads and 200 tails. But in practice, these results are rarely achieved.

Thus, the magnitude of discrepency between theory and observation is described by the quantity χ^2 [A greek letter, pronounced as chi-square]

Hence, the greater the discrepency between the observed and expected frequencies, the greater is the value of χ^2.

Thus χ^2 describes a measure of the correspondence between theory & observation.

If O_i = observed frequencies

E_i = corresponding set of expected freq.

$i = 1, 2, 3 --- n$

Then

$$\boxed{\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]}$$

where $\Sigma O_i = \Sigma E_i = N$ [Total frequency]

& degree of freedom $= (n-1)$

Conditions for applying χ^2 Test —

(i) N, the total numbers of frequencies should be large. As an arbitrary figure, we may say that N should be at least 50.

(ii) No theoretical cell frequency should be small.

(iii) The constraints on the cell frequencies, if any, should be linear.

χ^2 Test as a test of Goodness of FIT :—

If calculated value of χ^2 is less than the table value at a specified level of significance, the the fit is considered to be good.

If the calculated value is greater than the table value, the fit is considered to be poor.

Example:- In experiment on pea breading, the following frequencies of seeds were obtained:

Round & Yelow	Wrinkled & yellow	Round & Green	Wrinkled & Green	Total
315	101	108	32.	556

Theory predicts that the frequencies should be in proportions 9:3:3:1. Examine the correspondence between theory & Experiment.

Solution:-

Null Hypothesis: Ho: there is no significant difference between the observed & theoretical frequency.

The theoretical freq. can be calculated as

$$E_1 = \frac{556 \times 9}{16} = 312.75$$

$$E_2 = \frac{556 \times 3}{16} = 104.25, \quad E_3 = 104.25$$

$$E_4 = \frac{556 \times 1}{16} = 34.75$$

To calculate χ^2:

O_i	315	101	108	32
E_i	312.75	104.25	104.25	34.75
$\frac{(O_i - E_i)^2}{E_i}$	0.016187	0.101319	0.13489	0.217626

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 0.470024$$

Conclusion:-

Tabular value of χ^2 at 5% level of significance for 3 degree of freedom is 7·815.

i.e $\chi^2_{0.05} = 7{\cdot}815$

Now [Calculated value] < [Tabulated value]

$\Rightarrow$ Null Hypothesis accepted

$\Rightarrow$ Experimental results support the theory.

Example:-

Records taken of the numbers of male & female births in 800 families having four children are as follows:

No. of Male births	0	1	2	3	4
No. of female births	4	3	2	1	0
No. of families	32	178	290	236	94

Test whether the data are consistent with the hypothesis that binomial law holds and the chance of male birth is equal to that of female birth, namely $p = q = 1/2$.

Sol:- Null Hypothesis H_0:

~~that~~ the data are consistent with the hypo of equal probab. for male & female birth i.e. $p = q = 1/2$

we shall use Binomial distribution to calulate theoretical frequency given by

$$N(r) = N \cdot P(X = r)$$
$$= N \cdot {}^{n}C_{r}\, p^{r} (q)^{n-r}$$

where N = total frequency

$N(r)$ = No. of families with r male children

p = Probab. of male

q = ——— female

n = No. of children.

$$N(0) = 800 \times {}^{4}C_{0} \left(\tfrac{1}{2}\right)^{4} = 50$$

$$N(1) = 800 \times {}^{4}C_{1} \left(\tfrac{1}{2}\right)^{4} = 200, \quad N(2) = 300$$

$$N(3) = 200, \quad N(4) = 50$$

O_i	32	178	290	236	94
E_i	50	200	300	200	50
$\frac{(O_i - E_i)^2}{E_i}$	6.48	2.42	0.33	6.48	38.72

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right] = 54.48$$

Conclusion:

$\chi^2_{0.05} = 9.4$ with 4 degree of freedom

Hence $\chi^2_{\text{calulated}} > \chi^2_{\text{tabulated}}$

$\Rightarrow$ H_0 is Rejected.

$\Rightarrow$ chance of male birth is not equal to that of a female birth.

χ^2-Test as a Test of Independence:—

With the help of χ^2, we can decide whether the two attributes are associated or not.
If we assume that two attributes are independent as Null hypothesis. Then if the calculated value of χ^2 is less than the table value, then hypothesis holds good.
i.e. Attributes are independent.
But χ^2 > (greater than) the table value, then hypothesis is rejected.
i.e. the results do not support hypothesis.

The sample data set out into two-way table is called contingency table.

Let us consider two attributes A and B divided into r classes A_1, A_2, A_3 ------ A_r and B divided into s classes B_1, B_2, B_3 ------ B_s.
If $\sum_{i=1}^{r} A_i = \sum_{j=1}^{s} B_j = N$ (Total frequency)
The contingency table for $r \times s$ is given below:

B \ A	A_1	A_2	----	A_r	Total
B_1	A_1B_1	A_2B_1			B_1
B_2	A_1B_2	A_2B_2			B_2
⋮					⋮
B_s					B_s
Total	A_1	A_2	---	A_r	N

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \left[\frac{\left[(A_i B_j) - (A_i B_j)_0\right]^2}{(A_i B_j)_0} \right]$$

where $(A_i B_j)_0 = \dfrac{(A_i)(B_j)}{N}$

with $(r-1)(s-1)$ degree of freedom.

Ex. 1:- What are the expected frequency 2×2 contingency table:

(i)

a	b
c	d

(ii)

2	10
6	6

Solution:- Expec. freq. in each cell = $\dfrac{\text{Product of column total \& row total}}{\text{whole total}}$

(i) Observed frequencies:

a	b	a+b	B_1
c	d	c+d	B_2
a+c	b+d	a+b+c+d = N	
A_1	A_2		

Expected frequencies:

$\frac{A_1 B_1}{N}$	$\frac{A_2 B_1}{N}$
$\frac{A_1 B_2}{N}$	$\frac{A_2 B_2}{N}$

$\Rightarrow$

$\frac{(a+c)(a+b)}{a+b+c+d}$	$\frac{(b+d)(a+b)}{a+b+c+d}$
$\frac{(a+c)(c+d)}{a+b+c+d}$	$\frac{(b+d)(c+d)}{a+b+c+d}$

Example:- From the following table regarding the colour of eyes of father and son, test if the colour of Son's eye is associated with that of father.

Eye colour of Father \ Eye colour of son.	Light	Not light	
Light	471	51	522
Not light	148	230	378
	619	281	

Sol:- Null Hypothesis:- The colour of son's eye is not associated with that of father.

Expected frequencies:-

Eye colour of SON / Eye colour of Father	Light	Not light	Total
Light	$\frac{619 \times 522}{900} = 359.02$	$\frac{281 \times 522}{900} = 167.62$	522
Not light	$\frac{619 \times 378}{900} = 259.98$	$\frac{281 \times 378}{900} = 121.38$	378
Total	619	281	