Level of significance			
	1% (0.01)	5% (0.05)	10% (0.1)
Two tailed test	$\|z_\alpha\| = 2.58$	$\|z_\alpha\| = 1.96$	$\|z_\alpha\| = 1.645$
Right tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Left tailed test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

5.3 Student's t-Distribution (t- Test)

t- distribution is used when sample size ≤ 30. t- statics is defined as

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}, \text{ where } S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

$\bar{x}$ is the mean of the sample, μ is population mean, S is standard deviation of population and n is sample size.
If the standard deviation of the sample $'s'$ is given then t-statics is defined as

$$t = \frac{\bar{x}-\mu}{s/\sqrt{(n-1)}}$$

Application of t-distribution:

1. To test if the sample mean $(\bar{x})$ differ significantly from the hypothetical value μ of the population mean.
2. To test significance difference between two sample means.
3. To test the significance of observed partial and multiple correlation coefficients.

Test I: To test whether the mean of a sample drawn from normal population deviates significantly from a stated value when variance of population is unknown.

Working rule

- H_0 – There is no significant difference between sample mean $\bar{x}$ and population mean μ.
- Calculate t-statics:

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}, \text{ where } S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

with degree of freedom $(n-1)$.

- If calculated value is such that $|t| < t_\alpha$, the null hypothesis is accepted and if $|t| > t_\alpha$, the null hypothesis is rejected.

Note:

1. 95% confidence limits (level of significance 5%) are $\bar{x} \pm t_{0.05} S/\sqrt{n}$
2. 99% confidence limits (level of significance 1%) are $\bar{x} \pm t_{0.01} S/\sqrt{n}$

Example 1: A random sample size 16 and 53 as mean. The sum of the squares of the deviation from the mean is 135. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% and 99% confidence limits of the mean of the population.
Solution: Null hypothesis H_0: there is no significant difference between the sample mean and hypothetical population mean i.e. $\mu = 56$.
Alternative hypothesis $H_1: \mu \neq 56$ (two tailed test).
We know that

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

Given $\bar{x} = 53, \mu = 56, n = 16, \sum(x-\bar{x})^2 = 135$.

Now we have $S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{135}{15}} = 3$

$$\Rightarrow t = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} = \frac{53-56}{\frac{3}{\sqrt{16}}} = -4$$

$$\Rightarrow |t| = 4$$

Alternative hypothesis $H_1: \mu_1 > \mu_2$ i. e. Type I is superior than Type II.

Hence we use right tailed test.

We know that

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \qquad (1)$$

where

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{8(36)^2 + 7(40)^2}{8 + 7 - 2} = 1659.076$$

$$\therefore S = 40.7317$$

From eq. (1)

$$t = \frac{1234 - 1036}{40.7317\sqrt{\frac{1}{8} + \frac{1}{7}}} = 18.1480$$

$$d.f. = n_1 + n_2 - 2 = 13$$

$\therefore t_{0.05}$ at 13 $d.f.$ is 1.77.

Since calculated value of $|t| >$ tabulated value at 5% level of significance.

$\therefore H_0$ is rejected. Hence Type I is superior than Type II.

$\Longrightarrow$ There are discriminate between two horses at 5% level of significance.

Example 6: The height of 6 randomly chosen sailors in inches are 63, 65, 68, 69, 71, and 72. Those of 9 randomly chosen soldiers are 61, 62, 65, 66, 69, 70, 71, 72 and 73. Test whether the sailors are on the average taller than soldiers.

Solutions: Let x_1 and x_2 be the two samples denoting the heights of sailors and soldiers.

$$n_1 = 6, \qquad n_2 = 9$$

Null hypothesis $H_0: \mu_1 = \mu_2$ i.e. the mean of the population are the same.

Alternative hypothesis $H_1: \mu_1 > \mu_2$ (one tailed test)

We know that $\quad t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, where $S^2 = \frac{\sum(x_1 - \overline{x_1})^2 + \sum(x_2 - \overline{x_2})^2}{n_1 + n_2 - 2}$

$$\overline{x_1} = \frac{\sum x_1}{n_1} = 68, \qquad \overline{x_2} = \frac{\sum x_2}{n_2} = 67.6$$

Calculation of two sample means-

x_1	$x_1 - \overline{x_1}$	$(x_1 - \overline{x_1})^2$	x_2	$x_2 - \overline{x_2}$	$(x_2 - \overline{x_2})^2$
63	-5	25	61	-6.66	44.36
65	-3	9	62	-5.66	32.035
68	0	0	65	-2.66	7.0756
69	1	1	66	1.66	2.7556
71	3	9	69	1.34	1.7956

The tabulated value of t at 5% level of significance for 13 d.f. is 2.16. The calculated value of t less than tabulated value. H_0 is accepted. There is no significant difference between population mean i.e. $\mu_1 = \mu_2$.

Therefore two sample have been drawn from the same population.

Practice Questions

1. In laboratory experiment, two sample gave the following results:

Sample	Size	Sample mean	Sum of squares of deviation from mean
I	10	15	90
II	12	14	108

Test whether the sample come from same normal population.

2. Daily wages in Rupees of skilled workers in two cities are as follows:

	Size of sample of workers	S.D. of wages in the sample
City A	16	25
City B	13	32

Test whether the sample come from same normal population.

3. Two independent samples of size 8 and 9 had the following values of the variables:

Sample I	20	30	23	25	21	22	23	24	
Sample II	30	31	32	34	35	29	28	27	26

Do the estimates of the population variance differ significantly?

4. The standard deviation calculated from two random samples of sizes 9 and 13 are 2:1 and 1:8 respectively. Can the samples be regarded as drawn from normal populations with the same standard deviation?

5.5 Chi-Square χ^2 Test

The Chi-square test is very powerful test for testing the significance of the discrepancy between actual (or observed) frequencies and theoretical (or expected) frequencies. If O_i $(i = 1, 2, \dots . n)$ is the set of observed frequencies and E_i $(i = 1,2, \dots n)$ is the corresponding set of expected frequencies, then χ^2 is defined as

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

where $\sum O_i = \sum E_i = N$ (total frequency)

Application of Chi-square test

(1) **Test of independents of attributes –** With the help of chi-square test we can find out whether two or more attributes associated or not.

(2) **Test of goodness of fit-** On several occasions the decision makers need to understand whether an actual sample distribution matches with known probability distribution. Such as poison, binomial or normal.

(3) Compare a number of frequency distributions.
(4) Test for specified standard deviation i.e. it may be used to test of population variance.

Conditions for applying χ^2 Test

(1) Each cell should contain at least 5 observation.
(2) The members of sample should be independent.
(3) Constrains on the cell frequencies should be linear.
(4) Total frequencies N should be reasonably large, say greater than 50.

Working Rule

Step 1- Consider the null hypothesis and alternative hypothesis.

Step 2- Calculate the expected frequency E_i corresponding to each cell.

Step 3- Calculate χ^2 by the formula

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

and calculate dof.

Step 4- See the value of χ^2 from the table i.e. value of χ^2 at $\alpha\%$ level of significance and for dof v, as calculated in step 3.

Step 5- Decision:

(i) If calculated value of $\chi^2 <$ tabulated value of $\chi^2_{\alpha,\nu}$. Then accept the null hypothesis.

(ii) If calculated value of $\chi^2 >$ tabulated value of $\chi^2_{\alpha,\nu}$. Then reject the null hypothesis i.e. accept the alternative hypothesis.

χ^2 Test as a test of goodness of fit

Example 1: A die is thrown 276 times and the results of these throws are given bellow:

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	29	59	57	59

Test whether the die is biased or not.

Solution: Null hypothesis H_0: Die is unbiased.

The expected frequency for each digit is $\frac{276}{6} = 46$.

O_i	40	32	29	59	57	59
E_i	46	46	46	46	46	46
$(O_i - E_i)^2$	36	196	289	169	121	169

Since $\chi^2_{\text{cal}} < \chi^2_{\text{tab}}$ at 5% level of significance.∴ H_0 is accepted. The production of bad part is independent of the shift on which they were produced.

	Yes	No	Total
Yes	56	31	87
No	18	6	24
Total	74	37	111

Example 8: In a sample survey of public opinion, answers to questions (i) Do you drink? (ii) Are you in favor of local option on sale of liquor? are tabulated below:

Can you infer whether or not the local option on the sale of liquor is dependent on individual drink ? (Given that the value of $\chi^2$for 1 dof at 5% level of significance is 3.841)

Solution: Null hypothesis H_0: The option on sale of liquor is not dependent with individual drinking.

Calculation of expected frequencies:

$$E_{11} = \frac{74 \times 87}{111} = 58$$

$$E_{12} = \frac{37 \times 87}{111} = 29$$

$$E_{21} = \frac{74 \times 24}{111} = 16$$

$$E_{21} = \frac{37 \times 24}{111} = 8$$

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

We have

O_{ij}	E_{ij}	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
56	58	4/58
31	29	4/29
18	16	4/6
6	8	4/8
Total		0.957

Hence the value of $\chi^2 = 0.957$

Also the degree of freedom $\nu = (2-1)(2-1) = 1$

The tabulated value of χ^2 at 5% level of significance for 1 d.f. is 3.841.

$\because \chi^2_{\text{cal}} < \chi^2_{\text{tab}} \Longrightarrow H_0$ is accepted.