UNIT-V

SMALL SAMPLES

If the size of the sample < 30 the sample is said to be small sample or exact sample.
There are three kinds of tests for small samples.
1. Students t test using t - distribution
2. χ^2 *test using* χ^2 *distribution*
3. F – test using F – distribution

DEGREES OF FREEDOM:

Degrees of freedom one number of values we can choose freely.
For example: If $x_1 + x_2 + x_3 = 50$ to find x_3 we have to assign any value to the two of the variables (say x_1, x_2). Similarly for four we require three values.
Degrees of freedom for 2 sample values = 2 – 1 = 1
Degrees of freedom for n sample = n – 1

SMALL SAMPLE TESTS :

(i) t-test:

Name of the test	Null HypothesisH_0	Level of significance(α)	Test statistic
1.Test for single mean	$\mu = \mu_0$	5% or 1% or 10%	$t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}}$
2.Test for difference of means	$\mu_1 = \mu_2$	5% or 1% or 10%	$t=\frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$
3. paired t-test	$\mu = 0$	5% or 1% or 10%	$=\frac{\bar{d}}{\frac{s}{\sqrt{n}}}$

(ii) F-test:

Test for difference of variance
$H_0: {\sigma_1}^2 = {\sigma_2}^2$
α : 5% or 10%.
Test statistic F = $\frac{{S_1}^2}{{S_2}^2} = \frac{Greater\ variance}{smaler\ variance}$

(iii) Chi-square (χ^2)test:

Test for goodness of fit and test for independence of attributes.
Test statistic $\chi^2 = \sum \frac{(O-E)^2}{E}$

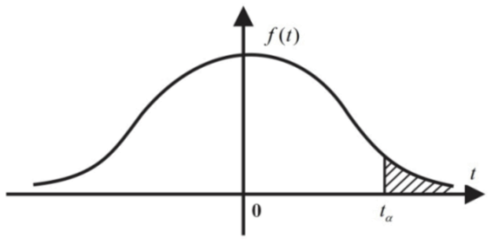
T – DISTRIBUTION:

T – Distribution is used when population standard deviation is unknown.
Let $\bar{x}$ be the mean of a random sample of size n, taken from a normal population having the mean μ and variance σ^2 and sample variance $S^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$ then test statistic is

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}\ where\ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ , n = sample\ size\ (n-1) = degrees\ of\ freedom$$

PROPERTIES OF T – DISTRIBUTION:

1. The shape of t – distribution is bell – shaped which is similar to that of a normal distribution.



2. The mean of standard normal distribution and as well as t – distribution is zero. Variance of t – distribution depends on degrees of freedom V.
3. The variance of t – distribution approaches 1 as n → ∞. In fact t – distribution with v – degrees of freedom approaches standard normal distribution as $V = (n-1) \to \infty$

APPLICATIONS OF THE T – DISTRIBUTION:

1. To test the significance of sample mean, when population variance is not given.
2. To test the significance of the mean of sample, i.e., to test if the sample mean differs significantly from the population mean.
3. To test the significance difference between two sample means or to compare two samples.
4. To test the significance of an observed sample correlation coefficient, sample regression coefficient.

2. CHI – SQUARED (χ^2) DISTRIBUTION:

Chi – squared distribution is a continuous probability distribution of continuous random variable X with probability density function is given by

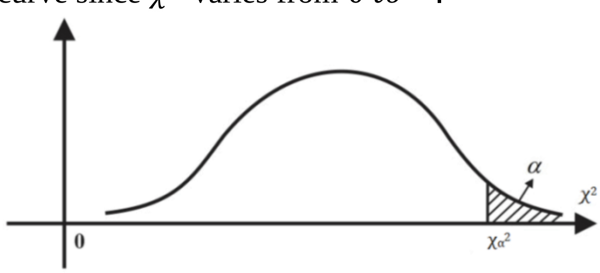
$$f(x) = \begin{cases} \frac{1}{2^{\frac{V}{2}}\Gamma\left(\frac{V}{2}\right)} \; x^{\frac{v}{2}-1}.e^{-\frac{x}{2}}, \; for\; x > 0 \\ 0, \; otherwise \end{cases}$$

$where\ V\ is\ the + ve\ integer\ also\ known\ as\ degrees\ of\ freedom.$

χ^2 distribution was used to measure of goodness of fit and to test independent attributes.

PROPERTIES OF χ^2 DISTRIBUTION:

1. χ^2 distribution curve is not symmetrical, lies entirely in the first quadrant. It is not a normal curve since χ^2 varies from $0\ to\ \infty$.



2. It depends only on the degrees of freedom v.
3. If $\chi_1^2\ and\ \chi_2^2$ are two independent distributions with $V_1\ and\ V_2$ degrees of freedom, the $\chi_1^2 + \chi_2^2$ will be χ^2- distribution with $(V_1 + V_2)$ degrees of freedom.
4. Here α denotes the area under the chi square distribution to the right of χ_α^2.

NOTE:

1. For various values of $\alpha\ and\ v$ the values of χ^2 are tabulated in tables.

2. Test of goodness of fit are used when we want to determine whether an actual sample distribution matches unknown theoretical distribution. This test is known as "$\chi^2 - test$" of goodness of fit.
   Test statistic for χ^2 – test is
   $\chi^2 = \sum\left(\frac{(O-E)^2}{E}\right)$ $where\ O = Observed\ frequency; E = Expected\ frequency$

3. i) If the data is given in series of 'n' numbers then degrees of freedom $v = n - 1$.
   ii) For binomial distribution d.f. $(v) = n - 1,\ for\ poisson\ distribution\ n - 2$
   iii) For normal distribution d.f. $(v) = n - 3$.

APPLICATIONS OF χ^2 – DISTRIBUTION:

1. To test the goodness of fit.
2. To test the independence of attributes.
3. To test the estimation of population variance.

F – DISTRIBUTION:

* F – distribution is sampling distribution of the ratio of the two sample variances.
* It is an important continuous probability distribution which plays an important role in connection with sampling from normal populations is the F – distribution.
* Let $S_1^2\ and\ S_2^2$ are two variances of two random samples of sizes $n_1\ and\ n_2$ respectively drawn from the normal population with variance $\sigma_1^2\ and\ \sigma_2^2$. To test whether the two samples come from two populations having equal variances. Test statistic is
  $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$
  $F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$
  Which follows F – distribution with $V_1 = n_1 - 1\ and\ V_2 = n_2 - 1$ degrees of freedom.
  If $\sigma_1^2 = \sigma_2^2\ then\ F = \frac{S_1^2}{S_2^2}$.

PROPERTIES OF F – DISTRIBUTION:

i) F – distribution is free from population parameter and depends upon degrees of freedom only.

ii) F – distribution curve lies entirely in first quadrant.

iii) The F – curve depends not only on the two parameters $V_1\ and\ V_2$ but also on the order in which they are stated.

iv) $F_{1-\alpha}(V_1,\ V_2) = \frac{1}{F_\alpha(V_1, V_2)}$

v) Mode of F – distribution is less than unity.

T – distribution is used when population S.D. is not known.

* For T – distribution
  $S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$ $\left(\bar{x} = \frac{\sum x}{n}\right)$
  $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ $where\ S = S.D.of\ sample$
  $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
  S.D. of population known - $t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n} - 1}$

1. For student t – test statistic when S.D. is t is defined as $t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n-1}}$

CONFIDENCE INTERVAL (LIMITS) FOR μ:

For 95% of confidence $\bar{X} \pm t_{0.05}\, {}^{S}/_{\sqrt{n}}$

For 99% of confidence $\bar{X} \pm t_{0.01}\, {}^{S}/_{\sqrt{n}}$

PROBLEMS ON T – DISTRIBUTION:

Find (a) $t_{0.05}\ when\ v = 16 \Rightarrow Ans : 1.746$

(b) $t_{-0.01}\ when\ v = 10 \Rightarrow Ans :\ -2.764$

(c) $t_{0.995}\ when\ v = 7$

$t_{0.995} = t_{1-0.005} = -t_{0.005} = -3.499 \quad (v = 7)$

T – Test for single mean:

2. A random sample of size 25 from a normal population has the mean $\bar{x} = 47.5\ and\ S.D.$ $S = 8.4$. Does this information tend to support or refuse the claim that the mean of the population is $\mu = 42.5$.

Sol: $n = 25, \bar{x} = 47.5,\ \mu = 42.5,\ S = 8.4$

1. $H_0: \mu = 42.5$
2. $H_1: \mu_1 \neq 42.5$ (Two tail)
3. $\alpha: 0.05$
4. Test statistic $: t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{47.5-42.5}{8.4/\sqrt{25}} = \frac{5\sqrt{25}}{8.4} = 2.98$
5. Conclusion: Value of t for 24 degrees of freedom at 0.025 is 2.064 (from t – table) $|t| > 2.067, \therefore Reject\ H_0$.

3. A mechanic is making engine parts with axle diameter of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a S.D. of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specifications.

Sol: $n = 10 < 30,\ \bar{x} = 0.742,$

$S.D.\ of\ population\ \sigma = 0.040$

$Mean\ of\ population\ \mu = 0.700$

1. $H_0: Product\ is\ confirming\ to\ specifications\ (or) \mu = 0.700$
2. $H_1: \mu_1 \neq 0.700$ (two tailed)
3. $\alpha: 0.05$
4. Test statistic $: t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n-1}} = \frac{0.742-0.700}{0.040/\sqrt{10-1}} = 3.15$
5. Conclusion: $|t| = 3.15 > t_{0.025} = 2.26\ for\ 9\ degrees\ of\ freedom, \therefore Reject\ H_0$.

4. The mean life of a sample of 25 fluorescent light bulbs produced by a company is computed to be 1570 hours. The company claims that the average life of the bulbs produced by a company is 1600 hours using the level of significance of 0.05. Is the claim acceptable.

Sol: Given sample size $n = 25,\ sample\ mean\ \bar{x} = 1570,$

$S.D.\ (\sigma) = 120$

$Mean\ of\ population\ \mu = 1600$

Degrees of freedom $= n - 1 = 24$

1. $H_0: The\ claim\ is\ acceptable\ \mu = 1600\ hrs$
2. $H_1: \mu_1 \neq 1600$ (Two tail)
3. $\alpha: 0.05$
4. Test statistic $: t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n-1}} = \frac{1570-1600}{120/24} = -1.22$

5. Conclusion : $|t| = 1.22 < t_{0.025} = 2.06\ for\ 24\ degrees\ of\ freedom,$
$\therefore Accept\ null\ hypothesis\ H_0$, i.e., the claim that the average life of the bulbs produced by the company is 1600 hrs is acceptable.

5. The average breaking strength of steel rods is specified to be 18.5 thousand pounds to test the sample of 14 rods were tested. The mean and S.D. obtained is 17.85 and 1.955. The result of experiment significance?

Sol: Given sample size $n = 14,\ sample\ mean\ \bar{x} = 17.85,$
$S.D.\ (S) = 1.955$
$Mean\ of\ population\ \mu = 18.5$
Degrees of freedom $= n - 1 = 13$
1. Null hypothesis H_0: $The\ result\ of\ the\ experiment\ is\ not\ significant\ \mu = 18.5$
2. Alternative hypothesis $H_1: \mu_1 \neq 18.5$ (Two tail)
3. Level of significance $\alpha: 0.05$
4. Test statistic $t: t = \frac{\bar{x}-\mu}{S/\sqrt{n-1}} = \frac{17.85-18.5}{1.955/\sqrt{13}} = \frac{-0.65}{0.542} = -1.199$
$|t| = 1.199 \Rightarrow$ Calculated t = 1.199
5. Conclusion: Tabulated value at 5% level of significance for 13 degrees of freedom for two tailed test = 2.16 , since calculated value < tabulated t we accept H_0 at 5% level of significance.

6. A random sample of six steel beams has a mean compressive strength of 58,392 (pounds per square inch), with S.D. of 648 p.s.i. Use this information and the level of significance α = 0.05 to test whether the true average compressive strength of the steel from which this sample came is 58,000 p.s.i.

Sol: Given sample size $n = 6,\ sample\ mean\ \bar{x} = 58392\ psi,$
$S.D.\ (S) = 648psi$
$Mean\ of\ population\ \mu = 58000$
Degrees of freedom $= n - 1 = 5$
1. Null hypothesis $H_0:\ \mu = 58000$
2. Alternative hypothesis $H_1: \mu_1 \neq 58000$ (Two tail)
3. Level of significance $\alpha: 0.05$
4. Test statistic $t: t = \frac{\bar{x}-\mu}{S/\sqrt{n-1}} = \frac{58392-58000}{648/\sqrt{5}} = 1.35$
$|t| = 1.35$
Calculated t = 1.35
5. Conclusion : Tabulated value = 3.365
Since calculated value < tabulated t we accept H_0.

PROBLEMS RELATED TO STUDENTS T – TEST
(WHEN S.D. OF SAMPLE IS NOT GIVEN DIRECTLY) :

1. A random sample of 10 boys has the following I.Qs 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. (a) Do these data support the assumption of population mean IQ of 100.
(b) Find reasonable range in which most of the mean IQ values of samples of 10 boys lie.

Sol: (a) Here S.D. and mean of the sample is not given directly.
Given data 70, 120, 110, 101, 88, 83, 95, 98, 107, 100.
Mean $\bar{x} = \frac{\sum x}{n} = \frac{972}{10} = 97.2$

x	$x - \bar{x}$	$(x - \bar{x})^2$
70	-27.2	739.84
120	22.8	519.84

110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
		1833.60

$S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} = \frac{1833.60}{9} = 203.73$

$S = \sqrt{203.73} = 14.27$

1. Null hypothesis H_0: $\mu = 100$
2. Alternative hypothesis H_1: $\mu_1 \neq 100$ (Two tail)
3. Level of significance α: 0.05
4. Test statistic t: $t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{97.2-100}{14.27/\sqrt{10}} = -0.62$

$|t| = 0.62 \Rightarrow$ Calculated t = 0.62

5. Conclusion : Tabulated value of t for 9 degrees of freedom at 5% level of significance is 2.26 (two tailed test) . Since calculated value < tabulated t we accept H_0, IQ μ = 100.

(b) Confidence limits

$\bar{x} \pm t_\alpha.\frac{S}{\sqrt{n}} = 97.2 \pm 2.26 \times 4.512 = (97.2 \pm 10.198) = (87, 107.4)$

2. The height of 10 males of given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches, Test at 5% significance level assuming that for 9 degrees of freedom.

Sol: Mean $\bar{x} = \frac{\sum x}{n} = \frac{660}{10} = 66$

x	$x - \bar{x}$	$(x - \bar{x})^2$
70	4	16
67	1	1
62	-4	16
68	2	4
61	-5	25
68	2	4
70	4	16
64	-2	4
64	2	4
66	0	0
660		**90**

$S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} = \frac{90}{9} = 10$

$S = \sqrt{10} = 3.1$

1. Null hypothesis H_0: $Average\ height\ is\ not\ greater\ than\ 64\ inches\ \mu = 64$
2. Alternative hypothesis H_1: $\mu > 64$ (One tail)
3. Level of significance α: 0.05
4. Test statistic t: $t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{66-64}{3.16/\sqrt{9}} = \frac{2}{1.05} = 1.9$

$|t| = 1.9$

Calculated t = 1.9

5. Conclusion : Tabulated value of t for 9 degrees of freedom at 5% level of significance for single test is 1.833. Since calculated value > tabulated t we reject H_0.

3. A random sample from a company's very extensive files shares that the order for certain kind of machinery were filled respectively in 10, 12, 19, 14, 15, 18, 11 and 13 days use the level of significance α = 0.01 to test claim that on the average such orders are filled in 10.5 days.

Sol: n = 8, Mean $\bar{x} = \frac{\sum x}{n} = \frac{112}{8} = 14$

$S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} = \frac{1}{7}((10-14)^2 + (12-14)^2 + (19-14)^2 + \cdots . + (13-14)^2)$

$= \frac{1}{7}(16 + 4 + 25 + \cdots . + 1 + 16) = \frac{1}{7}(72) = 10.286$

$S = \sqrt{10.286} = 3.207$

1. Null hypothesis H_0: $\mu = 10.5\ days$

2. Alternative hypothesis $H_1: \mu \neq 10.5$

3. Level of significance $\alpha: 0.01$

4. Test statistic $t: t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{14-10.5}{3.207/\sqrt{7}} = 3.087$

$|t| = 3.087$

Calculated t = 3.087

5. Conclusion : Tabulated value of t for 7 degrees of freedom at 1% level of significance is 3.499. Since calculated value < tabulated t we accept H_0.

STUDENTS T – TEST FOR DIFFERENCE OF MEANS:

To test the significant difference between the sample means $\bar{x}\ and\ \bar{y}$ of two independent samples of sizes $n_1\ and\ n_2$ with the same variance, we use statistic

$t = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\ (or) \frac{\bar{x}-\bar{y}}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$\bar{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i,\ \bar{y} = \frac{1}{n_2}\sum_{i=1}^{n_2} y_i$

$S^2 = \frac{1}{n_1+n_2-2}[\sum_i (x_i-\bar{x})^2 + \sum_i (y_i-\bar{y})^2]\ (or)\ S^2 = \frac{1}{n_1+n_2-2}[(n_1-1)S_1^2 + (n_2-1)S_2^2]$

T – distribution follows $(n_1 + n_2 - 2)$ degrees of freedom.

PROBLEMS:

1. Samples of two types of electric light bulbs were tested for length of life and following data were obtained.

Type I	Type II
Sample size $n_1 = 8$	$n_2 = 7$
Sample mean $\bar{x} = 1234\ hrs$	$\bar{y} = 1036\ hours$
Sample S.D. $S_1 = 36\ hrs$	$S_2 = 40\ hrs$

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life.

Sol: 1. $H_0: \mu_1 = \mu_2$

2. $H_1: \mu_1 > \mu_2$ (Two tail)

3. α: 5%

4. Test statistic t:

$S^2 = \frac{1}{n_1+n_2-2}[(n_1-1)S_1^2 + (n_2-1)S_2^2] = \frac{1}{8+7-2}[7(36)^2 + 6(40)^2] = 1436.31$

$t = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{1234-1036}{\sqrt{1436\left(\frac{1}{8}+\frac{1}{9}\right)}} = 10.09$

5. Conclusion: $D.F = 8 + 7 - 2 = 13; |t| > 1.77 \quad (0.05\ level - 1.77)\ from\ table$

2. Two independent samples have following values

Sample I	11	11	13	11	15	9	12	14
Sample II	9	11	10	13	9	8	10	-

Test the difference between the means.

Sol: $n_1 = 8,\ n_2 = 7,\ \bar{x} = \frac{96}{8} = 12,\ \bar{y} = \frac{70}{7} = 10$

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
11	-1	1	9	-1	1
11	-1	1	11	1	1
13	1	1	10	0	0
11	-1	1	13	3	9
15	3	9	9	-1	1
9	-3	9	8	-2	4
12	0	0	10	0	0
14	2	4			
		26			**16**

$S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i-\bar{x})^2 + \sum_i (y_i-\bar{y})^2\right] = \frac{1}{8+7-2}[26+16] = \frac{42}{13} = 3.23$

1. $H_0: \mu_1 = \mu_2$
2. $H_1: \mu_1 \neq \mu_2$
3. $\alpha = 0.05$
4. Test statistic : $t = \frac{\bar{x}-\bar{y}}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 2.15$
5. Conclusion: $At\ 13\ d.f.0.05\ level\ of\ significance\ 2.15 < 2.16$, therefore accept H_0.

3. Two horses A and B were tested according to the time (in seconds) to run a particular track with the following results

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether the two horses have to same running capacity.

Sol: $n_1 = 7,\ n_2 = 6,$

We first compute the sample means and S.D.

$\bar{x} = mean\ of\ first\ sample = \frac{1}{7}(219) = 31.286,$

$\bar{y} = mean\ of\ second\ sample = \frac{169}{6} = 28.16$

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
28	-3.286	10.8	29	0.84	0.7056
30	-1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	-4.16	17.3056
33	1.714	2.94	27	-1.16	1.3456
29	-2.286	5.226	29	0.84	0.7056
34	2.714	7.366			
219		**31.4358**	**169**		**26.8336**

$S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i-\bar{x})^2 + \sum_i (y_i-\bar{y})^2\right] = \frac{1}{11}[31.4358 + 26.8336]$

$= \frac{1}{11}(58.2694) = 5.23 \Rightarrow S = 2.3$

1. $H_0: \mu_1 = \mu_2$
2. $H_1: \mu_1 \neq \mu_2$
3. $\alpha = 0.05$
4. Test statistic : $t = \frac{\bar{x}-\bar{y}}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{31.286-28.16}{(2.3)\sqrt{\left(\frac{1}{7}+\frac{1}{6}\right)}} = 2.443$
5. Conclusion: Tabulated value of t for 7 + 6 – 2 = 11 d.f. at 5% level of significance is 2.201. 2.443 > 2.201 hence reject H_0.

4. To examine the hypothesis that the husbands are more intelligent than the wives. An investigator took a sample of 10 couples and administered them a test which measures the IQ the results are as follows:

Husbands	117	105	97	105	123	109	86	78	103	107
Wives	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a reasonable test at the level of significance of 0.05.

Sol: $n_1 = 10,\ n_2 = 10,$

We first compute the sample means and S.D. $\bar{x} = \frac{1030}{10} = 103,\ \bar{y} = \frac{958}{10} = 95.8$

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$y - \bar{y}$	$(y - \bar{y})^2$
117	14	196	106	10.2	104.04
105	2	4	98	2.2	4.84
97	-6	36	87	-8.8	77.84
105	2	4	104	8.2	67.24
123	20	400	116	20.2	408.04
109	6	36	95	-0.8	0.64
86	-17	289	90	-5.8	33.64
78	-25	625	69	-26.8	718.24
103	0	0	108	12.2	148.84
107	4	16	85	-10.8	116.64
1030		**1600**	**958**		**1679.6**

$S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i - \bar{x})^2 + \sum_i (y_i - \bar{y})^2\right] = \frac{1}{18}[1606 + 1679.6]$

$= \frac{1}{18}(3285.6) = 182.53 \Rightarrow S = 13.51$

1. $H_0: \mu_1 = \mu_2$
2. $H_1: \mu_1 > \mu_2 (husbands\ are\ more\ intelligent\ than\ wives)(one\ tailed)$
3. $\alpha = 0.05$
4. Test statistic : $t = \frac{\bar{x}-\bar{y}}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{103-95.8}{(13.51)\sqrt{\left(\frac{1}{10}+\frac{1}{10}\right)}} = 1.19168$
5. Conclusion : Tabulated value of t for 18 d.f. at 0.05 level of significance is 1.734. 1.19168 < 1.734 hence accept H_0. There is no difference in IQs .

5. To compare two kinds of bumper guards 6 of each kind were mounted on a car and the car was run into a concrete wall. The following are costs of repair.

Guard 1	107	148	123	165	102	119
Guard 2	134	115	112	151	133	129

Use 0.01 level of significance to test whether the difference between two sample means is significant.

Sol: $n_1 = 6,\ n_2 = 6,$

$\bar{x} = \frac{764}{6} = 127.33,\ \bar{y} = \frac{774}{6} = 129;\ \sum(\boldsymbol{x} - \bar{\boldsymbol{x}})^2 = 2989.34;\ \sum(\boldsymbol{y} - \bar{\boldsymbol{y}})^2 = 1010$

$S^2 = \frac{1}{n_1+n_2-2}[\sum_i(x_i - \bar{x})^2 + \sum_i(y_i - \bar{y})^2] = 399.39 \Rightarrow S = 19.999$

1. $H_0: \mu_1 = \mu_2$
2. $H_1: \mu_1 \neq \mu_2\ (two\ tail)$
3. $\alpha = 5\%$
4. Test statistic : $t = \frac{\bar{x}-\bar{y}}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 0.1446$
5. Conclusion : Tabulated value of t for 10 d.f. at 1% level of significance (two tailed test) is 3.169

$t_{cal} < t_{tab}$ hence accept H_0.

6. Find the maximum difference that we can expect with probability 0.95 between the means of samples of sizes 10 and 12 from a normal population if their S.D. are found to be 2 and 3.

Sol: $n_1 = 10, n_2 = 12, S_1 = 2, S_2 = 3$

$S^2 = \frac{1}{n_1+n_2-2}[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2] = \frac{1}{10+12-2}[9(2)^2 + 11(3)^2] = \frac{1}{20}(135) = 6.75$

$\Rightarrow S = \sqrt{6.75} = 2.6$

1. $H_0: \mu_1 = \mu_2$
2. $H_1: \mu_1 \neq \mu_2$
3. $\alpha = 0.05$
4. Test statistic : $t = \frac{\bar{x}-\bar{y}}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \Rightarrow \bar{x} - \bar{y} = |t| \times S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 2.08 \times 2.6 \times \sqrt{\frac{1}{2} + \frac{1}{3}}$

$2.08\ is\ the\ table\ value\ for\ 20\ d.f.at\ 5\%\ level\ of\ significance.$

$\bar{x} - \bar{y} = 2.32.$

Hence maximum difference between means = 2.32.

7. Measuring specimen of nylon yarn, taken from two machines, it was found that 8 specimens from first machine has a mean denier of 9.67 with S.D. of 1.81 while 10 specimens from second machine had a mean denier of 7.43 with S.D. of 1.48 assuming that the proportions are normal test the hypothesis $H_0: \mu_1 - \mu_2 = 1.5$ against $H_1: \mu_1 - \mu_2 > 1.5\ at\ 0.05$ level of significance.

Sol: $n_1 = 8, n_2 = 10, \bar{x} = 9.67,\ \bar{y} = 7.43,\ S_1 = 1.81, S_2 = 1.48$

$S^2 = \frac{1}{n_1+n_2-2}[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2] = \frac{42.64}{16} = 2.665$

$\Rightarrow S = \sqrt{2.665} = 1.63$

1. $H_0: \mu_1 - \mu_2 = 1.5(\delta)$
2. $H_1: \mu_1 \neq \mu_2 > 1.5(\delta)$
3. $\alpha = 0.05$
4. Test statistic : $t = \frac{\bar{x}-\bar{y}-\delta}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{(9.67-7.43)-(1.5)}{1.63\sqrt{\left(\frac{1}{8}+\frac{1}{10}\right)}} = \frac{0.74}{0.7732} = 0.96$
5. Conclusion: Table value for 16 d.f. at 5% level of significance is 1.746

$t_{cal} <\ t_{table}, Hence\ accept\ H_0.$

PAIRED SAMPLE T – TEST:

When the two samples are not independent paired t – test is applied for n paired observation by taking differences $d_1, d_2, \ldots \ldots d_n$ of the paired data. To test whether the differences d_1 from a random sample from a population with mean μ we use statistic

$t = \frac{\bar{d}-\mu}{S/\sqrt{n}}$ $where\ \bar{d} = \frac{1}{n}\sum d_i$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2\ or\ S^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right]$

Above statistic follows students t – distribution with n – 1 degrees of freedom.

NOTE:

Take mean of differences $\mu = 0\ as\ H_0\ and\ \mu > 0\ as\ H_1$

PROBLEMS:

1. Ten workers were given training program with a view to study their assembly time for a certain mechanism. The result of the time and motion studies before and after training programme are given below.

Workers	1	2	3	4	5	6	7	8	9	10
X_1	15	18	20	17	16	14	21	19	13	22
Y_1	14	16	21	10	15	18	17	16	14	20

X_1 = Time taken for assembling before training

Y_1 = Time taken for assembling after training

Test whether there is significant difference is assembly times before and after training.

Sol: Let μ be the mean of population of differences.

1. $H_0: \mu = 0\ (training\ is\ not\ useful)$
2. $H_1: \mu > 0$
3. $\alpha: 0.05$
4. Test statistic :

Computation Differences (before and after training)

1, 2, -1, 7, 1, -4, 4, 3, -1, 2

$\bar{d} = \frac{14}{10} = 1.4$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 = \frac{1}{9}(82.4) = 9.1555$

$S = 3.026$

$t = \frac{\bar{d}-\mu}{S/\sqrt{n}} = \frac{1.4-0}{3.026/\sqrt{10}}$

$\Rightarrow\ |t| = 1.46$

5. Conclusion : Table value with 9 d.f. at 0.05 is 1.833

$t_{cal} < t_{tab}\ \Rightarrow\ accept\ H_0$ i.e., training is not useful.

2. Scores obtained in a shooting competition by 10 soldiers before and after intensive training are given below:

Before	67	24	57	63	54	56	68	33	43
After	70	38	58	58	67	68	75	42	38

Test whether the intensive training is useful at 0.05 level of significance.

Sol: 1. $H_0: \mu = 0\ (training\ is\ not\ useful)$

2. $H_1: \mu > 0\ (training\ useful)$
3. $\alpha: 0.05$
4. Test statistic :

Computation Differences -3, -14, -1, -3, 7, -13, -12, -7, -9.5

$\bar{d} = \frac{1}{10}\sum_{i=1}^{10} d_i = -\frac{50}{10} = -5$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 = \frac{1}{9}(482) = 53.555$

$S = 7.32$

Test statistic $t = \frac{\bar{d}-\mu}{S/\sqrt{n}} = \frac{-5-0}{7.32/\sqrt{10}} = -2.16$

$\Rightarrow \ |t| = 2.16$

5. Conclusion: Table value with 9 d.f. at 0.05 is 1.83

$t_{cal} > \ t_{tab} \ \Rightarrow \ \ reject \ H_0$ i.e., intensive training is useful.

3. Blood pressure of 5 women before and after intake of a certain drug are given below.

Before	110	120	125	132	125
After	120	118	125	136	121

Test whether there is significant change in blood pressure at 1% level of significance.

Sol: Differences -10, 2, 0, -4, 4

$\bar{d} = 2$

$S^2 = 30 \Rightarrow S = \sqrt{30}$

1. $H_0: \mu = 0$
2. $H_1: \mu > 0 \ (one \ tail)$
3. $\alpha : 0.05$
4. Test statistic : $t = \frac{\bar{d}-\mu}{S/\sqrt{n}} = 0.82$
5. Conclusion: t – table value for 4 d.f. is 4.6

$t_{cal} \leq t_{table}$, accept H_0.

SNEDECOR'S F – TEST OF SIGNIFICANCE:

When testing the significance of the differences of the means of two samples came from same population or from populations with equal variances. If the variance of proportions are not equal a significant difference in the means may arise. Before we apply t – test for the significance of the difference of two means, we have to test for the equality of population variances using F – test of significance.

If $S_1^2 \ and \ S_2^2$ are the variances of two samples of sizes $n_1 \ and \ n_2$ respectively then population variances are given by

$n_1\sigma_1^2 = (n_1 - 1)S_1^2 \ \ and \ \ n_2\sigma_2^2 = (n_2 - 1)S_2^2$

$v_1 = n_1 - 1, v_2 = n_2 - 1$ degrees of freedom of these estimate.

We want to test if these estimates $S_1^2 \ and \ S_2^2$ are significantly different or if the samples may be regarded as drawn from the same populations or from two populations with same variance σ^2.

In this case set up null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$, i.e., population variances are same.

Test statistic is $F = \frac{S_1^2}{S_2^2} \ [S_1^2, S_2^2]$

$S_1^2 = \frac{\sum(x_i-\bar{x})^2}{n_1-1}, \ S_2^2 = \frac{\sum(y_i-\bar{y})^2}{n_2-1}$

$F = \frac{Greater \ variance}{Smaller \ variance}$

PROBLEMS:

1. In one sample of 8 observations the sum of the squares of deviations of the sample value from the sample mean was 84.4 and in other sample of 10 observations it was 102.6. Test whether this difference is significant at 5% level.

Sol: Here $n_1 = 8, n_2 = 10$

$\sum(x_i - \bar{x})^2 = 84.4; \ \sum(y_i - \bar{y})^2 = 102.6$

$S_1^2 = \frac{\sum(x_i-\bar{x})^2}{n_1-1} = \frac{84.4}{7} = 12.057$

$S_2^2 = \frac{\sum(y_i-\bar{y})^2}{n_2-1} = \frac{102.6}{9} = 11.4$

1. $H_0: S_1^2 = S_2^2$
2. $H_1: S_1^2 \neq S_2^2$
3. $\alpha : 5\%$

Now $F = \frac{S_1^2}{S_2^2} = \frac{12.057}{11.4} = 1.057$

i.e., calculated F = 1.057

Tabulated value of F at 5% level for (7, 9) degrees of freedom is 3.29

$F_{cal} < F_{tab}$ $i.e., accept\ H_0$.

2. Two random samples gave the following results

Sample	**Size**	**Sample mean**	**Sum of squares of deviation from the mean**
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population.

Sol: $n_1 = 10, n_2 = 12,\ \bar{x} = 15, \bar{y} = 14$

$\sum(x_i - \bar{x})^2 = 90;\ \sum(y_i - \bar{y})^2 = 108$

To test whether the samples drawn from same normal population.

i.e., $H_0: \mu_1 = \mu_2\ and\ \sigma_1^2 = \sigma_2^2$

Here we have to test

i) Test equality of variances by F - test

ii) Test equality of means by t – test

i) F – test (equality of variances)

1. $H_0: \sigma_1^2 = \sigma_2^2$
2. $H_1: \sigma_1^2 \neq \sigma_2^2$
3. $\alpha : 5\%$
4. Test statistic :

Given $n_1 = 10, n_2 = 12, \overline{X_1} = 15, \overline{Y_1} = 14, \sum(x_i - \bar{x})^2 = 90, \sum(y_i - \bar{y})^2 = 108$

$S_1^2 = \frac{\sum(x_i-\bar{x})^2}{n_1-1} = 10$

$S_2^2 = \frac{\sum(y_i-\bar{y})^2}{n_2-1} = 9.82$

$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.82} = 1.018$

i.e., calculated F = 1.018

5. Conclusion: Tabulated value of F at 5% level for (9, 11) degrees of freedom is 2.90

$F_{cal} < F_{tab}$ $i.e., accept\ H_0$. i.e., sample came from same normal population with same variance.

ii) t – test : (to test equality of means)

Given $n_1 = 10, n_2 = 12, \overline{X_2} = 15, \overline{Y_2} = 14, \sum(x_i - \bar{x})^2 = 90, \sum(y_i - \bar{y})^2 = 108$

$S^2 = \frac{1}{n_1+n_2-2}[\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2] = \frac{1}{20}[90 + 108] = 9.8 \Rightarrow S = 3.15$

1. $H_0: \mu_1 = \mu_2$
2. $H_1: \mu_1 \neq \mu_2$
3. $\alpha: 0.05$
4. Test statistic : $t = \frac{\bar{x}-\bar{y}}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{15-14}{315\sqrt{\frac{1}{10}+\frac{1}{12}}} = 0.74$
5. Conclusion : Tabulated for 20 d.f. at 5% level (two tailed) is 2.086

$t_{cal} < t_{tab} \Rightarrow accept\ H_0\ \ i.e., \mu_1 = \mu_2$

Given samples have been drawn from the same normal population.

3. Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins shown the sample S.D. of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal. Test the hypothesis that true variance are equal.

Sol: $n_1 = 11, n_2 = 9, S_1 = 0.8, S_2 = 0.5$

Sample S.D's are given

Population variances $\sigma_1^2\ and\ \sigma_2^2$ use the relation

$\sigma_1^2 = \frac{n_1 S_1^2}{n_1 - 1} = \frac{11\times(0.8)^2}{10} = 0.704;$

$\sigma_2^2 = \frac{n_2 S_2^2}{n_2 - 1} = \frac{9\times(0.5)^2}{8} = 0.281$

1. $H_0: \sigma_1^2 = \sigma_2^2$
2. $H_1: \sigma_1^2 \neq \sigma_2^2$
3. $\alpha = 0.05$
4. Test statistic : $F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{0.704}{0.281} = 2.5 \Rightarrow F = 2.5$
5. Conclusion : Tabulated value of F for (10, 8) d.f. at 5% level of significance is 3.35

$F_{cal} < F_{tab}\ i.e., accept\ H_0.$

4. Time taken by workers in performing a job by method I and method II is given below:

Method I	20	16	26	27	23	22	-
Method II	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly.

Sol: Calculation of sample variance

$n_1 = 6,\ n_2 = 7,$

$\bar{x} = \frac{\sum x}{n_1} = \frac{134}{6} = 22.3,\ \bar{y} = \frac{241}{7} = 34.4$

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$y - \bar{y}$	$(y - \bar{y})^2$
20	-2.3	5.29	27	-7.4	54.76
16	-6.3	39.69	33	-1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	-2.4	5.76
22	-0.3	0.09	34	-0.4	0.16
			38	3.6	12.96
		81.34			**133.72**

$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1} = \frac{81.34}{5} = 16.26$

$S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1} = \frac{133.72}{6} = 22.29$

1. $H_0: \sigma_1^2 = \sigma_2^2$
2. $H_1: \sigma_1^2 \neq \sigma_2^2$
3. $\alpha = 0.05$
4. Test statistic : $F = \frac{S_2^2}{S_1^2} = \frac{22.29}{16.26} = 1.3699$

$\Rightarrow F = 1.3699$

5. Conclusion: Tabulated value of F for (5, 6) d.f. at 5% level of significance is 4.95.

$F_{cal} < F_{tab}\ i.e., accept\ H_0.$

5. The measurements of the output of two units have given the following results assuming that both samples have been obtained from the normal population at 5% significant level test whether the two populations have same variances.

Unit A	14.1	10.1	14.7	13.7	14.0
Unit B	14.0	14.5	13.7	12.7	14.1

Sol: $n_1 = 5,\ n_2 = 5,$

$\bar{x} = \frac{\sum x}{n_1} = \frac{66.6}{5} = 13.32,\ \bar{y} = \frac{69}{5} = 13.8$

$\boldsymbol{x}$	$\boldsymbol{x - \bar{x}}$	$\boldsymbol{(x - \bar{x})^2}$	$\boldsymbol{y}$	$\boldsymbol{y - \bar{y}}$	$\boldsymbol{(y - \bar{y})^2}$
14.1	0.78	0.6084	14.0	0.2	0.04
10.1	-3.22	10.3684	14.5	0.7	0.49
14.7	1.38	1.9044	13.7	-0.1	0.01
13.7	0.38	0.1444	12.7	-1.1	1.21
14.0	0.68	0.4624	14.1	0.3	0.09
		13.488			**1.84**

$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1} = \frac{13.488}{4} = 3.372$

$S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1} = \frac{1.84}{4} = 0.46$

1. $H_0: \sigma_1^2 = \sigma_2^2$
2. $H_1: \sigma_1^2 \neq \sigma_2^2$
3. $\alpha = 0.05$
4. Test statistic : $F = \frac{S_1^2}{S_2^2} = \frac{3.372}{0.46} = 7.33$

$\Rightarrow F = 7.33$

5. Conclusion: Tabulated value of F for (4, 4) d.f. at 5% level of significance is 6.39.

$F_{cal} > F_{tab}\ i.e., reject\ null\ hypothesis\ H_0.$

6. The nicotine contents in milligrams in the samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	-
Sample B	27	30	28	31	22	36

Can it be said that two samples came from same normal population.

Sol: $n_1 = 5,\ n_2 = 6,$

$\bar{x} = \frac{\sum x}{n_1} = 24.6,\ \bar{y} = 29$

$\sum(\boldsymbol{x_i} - \boldsymbol{\bar{x}})^2 = 21.2;\ \sum(\boldsymbol{y_i} - \boldsymbol{\bar{y}})^2 = 108$

$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1} = \frac{21.2}{4} = 5.3$

$S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1} = 21.6$

F – test :

1. $H_0: \sigma_1^2 = \sigma_2^2$
2. $H_1: \sigma_1^2 \neq \sigma_2^2$
3. $\alpha = 5\%$
4. Test statistic : $F = \frac{S_2^2}{S_1^2} = \frac{21.6}{5.3}$

$F_{cal} = 4.075$

5. Conclusion: $F_{tab}\ at\ (5.4)$ d.f. is 6.2.

$F_{cal} \leq F_{table},\ Accept\ H_0.$

T – test for difference of mean :

1. $H_0: \mu_1 = \mu_2$
2. $H_1: \mu_1 \neq \mu_2$
3. $\alpha = 5\%$
4. Test statistic : $t = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{24.6-29}{3.78\sqrt{\frac{1}{5}+\frac{1}{6}}} = -1.92$
5. Conclusion : t – table value at 5 + 6 – 2 = 9 d.f. is 2.26; $|t| < t\ table,\ accept\ H_0$.

CHI – SQUARE (χ^2) TEST:

χ^2 – test is used to test the goodness of fit.

Test statistic for χ^2 – test

$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$ $with\ n-1\ degrees\ of\ freedom$

$O = Observed\ frequency$

$E = Expected\ frequency$

NOTE:

If the data is given in series of n numbers then degrees of freedom = n – 1
In case of Binomial distribution d.f. = n – 1
In case of Poisson distribution d.f. = n – 2
In case of Normal distribution d.f. = n – 3

χ^2 – TEST AS A TEST OF GOODNESS OF FIT:

1. The number of automobile accidents per week in a certain community are as follows 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

Sol: Expected frequency of accidents in each week $= \frac{100}{10} = 10$

1. H_0: Accident conditions were same during the 10 week period.
2. H_1 : Accident conditions were not same.
3. α : 5%

Observed frequency (O)	Expected frequency (E)	(O – E)	$\frac{(O-E)^2}{E}$
12	10	2	0.4
8	10	-2	0.4
20	10	10	10.0
2	10	-8	6.4
14	10	4	1.6
10	10	0	0.0
15	10	5	2.5
6	10	-4	1.6
9	10	-1	0.1
4	10	6	3.6
100	**100**		**26.6**

4. Test statistic $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 26.6$
5. Conclusion : Table value for 9 d.f. at 0.05 level of significance is 16.9
$\chi^2_{cal} > \chi^2_{tab},\ i.e., reject\ H_0$.

2. A sample analysis of examination result of 500 students was made. It was found that 220 students had failed, 170 had secured a third class, 90 were placed in second class and 20 got first class. Do these figures commensurate with the general examination result which is in the ratio 4 : 3 : 2 : 1 for the various categories respectively.

Sol: 1. H_0: The observed results commensurate with the general examination results.

2. H_1 : Observed results not equal to examination results.
3. $\alpha = 5\%$
4. Test statistic : Expected frequencies are in the ratio 4 : 3 : 2 : 1
Total frequency = 500
If we divide the total frequency 500 in the ratio 4 : 3 : 2 : 1, we get the frequencies as 200, 150, 100, 50.

Class	Observed frequency (O)	Expected frequency (E)	(O – E)	$\frac{(O-E)^2}{E}$
Failed	220	200	20	2.00
Third	170	150	20	2.667
Second	90	100	-10	1.000
First	20	50	-30	18.00
	500	**500**		**23.667**

Calculated $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 23.667$

5. Conclusion : Degrees of freedom n – 1 = 4 – 1 = 3
For $v = 3, \chi^2_{0.005} = 7.81$
$\chi^2_{cal} > \chi^2_{tab},\ i.e., reject\ H_0.$

3. The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally frequently in the directory.

Sol: 1. H_0: The digits occur equally frequently in the directory.
2. H_1 :The digits not occur equally frequently.
3. $\alpha = 5\%$
4. Test statistic : Under this null hypothesis the expected frequencies for each of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 is $\frac{10000}{10} = 1000$

Digits	Observed frequency (O)	Expected frequency (E)	$(O – E)^2$	$\frac{(O-E)^2}{E}$
0	1026	1000	676	0.676
1	1107	1000	11449	11.449
2	997	1000	9	0.009
3	966	1000	1156	1.156
4	1075	1000	5625	5.625
5	933	1000	4489	4.489
6	1107	1000	11449	11.449
7	972	1000	784	0.784
8	964	1000	1296	1.296
9	853	1000	21609	21.609
	10000	**10000**		**58.542**

Calculated $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 58.542$

5. Conclusion : Degrees of freedom n – 1 = 9 at 5% is 16.919
$\chi^2_{cal} < \chi^2_{tab},\ i.e., accept\ H_0.$

4. A dice is thrown 264 times with the following results. Show that the dice is biased (Given $\chi^2 = 11.07\ for\ 5\ d.f$)

No. appeared on the dice	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

Sol: 1. H_0: The dice is unbiased.

2. H_1 : The dice is biased
3. $\alpha = 5\%$
4. Test statistic : The expected frequency of each of the numbers 1, 2, 3, 4, 5, 6 is $\frac{264}{6} = 44$

Observed frequency (O)	Expected frequency (E)	(O – E)^2	$\frac{(O-E)^2}{E}$
40	44	16	0.3636
32	44	144	3.2727
28	44	256	5.8181
58	44	196	4.4545
54	44	100	2.2727
52	44	64	1.4545
264	**264**		**17.6362**

Calculated $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 17.6362$

5. Conclusion : Degrees of freedom n – 1 = 5 at 5% is 11.07; $\chi^2_{cal} > \chi^2_{tab},\ i.e., reject\ H_0.$

5. A pair of dice are thrown 360 times and the frequency of each sum is indicated below:

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the Chi Square test at 0.05 level of significance.

Sol: 1. H_0 : $dice\ are\ fair$
2. H_1: $dice\ is\ not\ fair$
3. $\alpha = 0.05$
4. Test statistic :

(x)	2	3	4	5	6	7	8	9	10	11	12
P(x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Sum	Observed frequency (O)	Expected frequency (E)	(O – E)^2	$\frac{(O-E)^2}{E}$
2	8	10	4	0.4
3	24	20	16	0.8
4	35	30	25	0.833
5	37	40	9	0.225
6	44	50	36	0.72
7	65	60	25	0.414
8	51	50	1	0.02
9	42	40	4	0.1
10	26	30	16	0.53
11	14	20	36	1.8
12	14	10	16	1.6
				7.445

Calculated $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 7.445$

5. Conclusion : Degrees of freedom n – 1 = 11 at 5% is 18.3
$\chi^2_{cal} < \chi^2_{tab},\ i.e., accept\ H_0.$

6. Fit a poisson distribution to the following data and test the goodness of the fit at 5% level of significance.

(x)	0	1	2	3	4
(f)	214	92	20	3	1

Sol: Given data:

(x)	0	1	2	3	4	
(f)	214	92	20	3	1	$\sum f$ =**330**

Mean $\mu = \frac{\sum fx}{\sum f} = \frac{0\times214+1\times92+2\times20+3\times3+4\times1}{330} = \frac{145}{330} = 0.439$

Mean $\mu = \lambda = 0.439$

$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$P(0) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda} = e^{-0.439} = 0.6446$

Recurrence formula for Poisson distribution

$P(x+1) = \frac{\lambda}{x+1}\ P(x)$

For $x = 0,\ P(1) = \frac{\lambda}{1}\ P(0) = 0.6446 \times 0.439 = 0.28297$

For $x = 1,\ P(2) = \frac{0.439}{2}\ P(1) = \frac{0.439}{2} \times 0.2829 = 0.06211$

For $x = 2,\ P(3) = \frac{0.439}{3}\ P(2) = \frac{0.439}{3} \times 0.06211 = 0.0091$

For $x = 3,\ P(4) = \frac{0.439}{4}\ P(3) = \frac{0.439}{4} \times 0.0091 = 0.0009$

N = 330

1. H_0 : Fitting is good
2. H_1 : Fitting is not good
3. $\alpha = 5\%$
4. Test statistic :

x	Observed frequency (O)	Expected frequency (E) N× P(x)	(O – E)	$\frac{(O-E)^2}{E}$
0	214	213	1	0.0046
1	92	93	-1	0.0107
2	20	20	0	0
3	3	3	0	0
4	1	1	0	0
				0.015

Calculated $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 0.00606$

5. Conclusion : χ^2 table value at 3 d.f. at 5% level is 7.815

$\chi^2_{cal} < \chi^2_{tab},\ i.e., accept\ H_0.$

CHI – SQUARED TEST FOR INDEPENDENCE OF ATTRIBUTES:

* Attributes means quality or characteristic examples of attributes are blindness, honesty, beauty etc.
* An attribute may be marked by its presence or absence in number of a given population
* Let us consider two attributes A and B. A is divided into two classes and B is divided into two classes.
* The various cell frequencies can be expressed in the following table.

A	a	B
B	c	D

a	B	a + b
c	D	c + d
a + c	b + d	N

Expected frequencies are given by

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

NOTE:

In this Chi – Square test we test if two attributes A and B under consideration are independent or not.

Null Hypothesis H_0 : $Attributes\ are\ independent$

$\chi^2\ distribution\ follows$

Degrees of freedom $(r-1)(s-1)$

$r = number\ of\ rows, s = number\ of\ columns.$

PROBLEMS:

1. On the basis of information given below about the treatment of 200 patients suffering from disease, state whether the new treatment is comparatively superior to the conventional treatment

	Favourable	**Not favourable**	**Total**
New	60	30	90
Conventional	40	70	110
Total	100	100	200

Sol: 1. H_0: $Null\ hypothesis$: No difference between new and conventional treatment (or) new and conventional treatment are.

2. H_1: New and conventional treatment are dependent

3. $\alpha = 0.05$

4. Computation : Expected frequencies are given in table:

$\frac{90 \times 100}{200} = 45$	$\frac{90 \times 100}{200} = 45$	90
$\frac{100 \times 110}{200} = 55$	$\frac{100 \times 110}{200} = 55$	110
100	100	200

Observed frequency (O)	Expected frequency (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
60	45	225	5
30	45	225	5
40	55	225	4.09
70	55	225	4.09
200	**200**		**18.18**

Calculated $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 18.18$

Degrees of freedom = (2 – 1)(2 – 1) = 1

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

$\chi^2_{cal} > \chi^2_{tab},\ i.e., reject\ H_0.$

2. Given following table for hair colour and eye colour. Find the value of χ^2. Is there good association between the two.

	Hair Colour				
		Fair	**Brown**	**Black**	**Total**

Eye Colour	**Blue**	15	5	20	40
	Grey	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

Sol: 1. $H_0: Null\ hypothesis$: Two attributes hair and eye colour are independent.

2. H_1: Alternate hypothesis : Two attributes hair and eye colour are dependent.

3. $\alpha = 0.05$

4. Computation : Expected frequencies are given in table:

$\frac{60 \times 40}{150} = 16$	$\frac{30 \times 40}{150} = 8$	$\frac{60 \times 40}{150} = 16$	40
$\frac{60 \times 50}{150} = 20$	$\frac{30 \times 60}{150} = 10$	$\frac{60 \times 50}{150} = 20$	50
$\frac{60 \times 60}{150} = 24$	$\frac{30 \times 60}{150} = 12$	$\frac{60 \times 60}{150} = 24$	60
60	30	60	150

Observed frequency (O)	**Expected frequency (E)**	**$(O - E)^2$**	$\frac{(O-E)^2}{E}$
15	16	1	0.0625
5	8	9	1.125
20	16	16	1.0
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.042
15	12	9	0.75
20	24	16	0.666
			3.6458

5. Test statistic $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 3.6458$

Degrees of freedom = (3 – 1)(3 – 1) = 4 d.f. at 5% level of significance is 9.488.

$\chi^2_{cal} < \chi^2_{tab},\ i.e., accept\ H_0.$

PRACTICE PROBLEM:

3. Two researchers adopted different sampling techniques while investigating some group of students to find the number of students falling into different intelligence level. The results are follows:

Researchers	**Below Average**	**Average**	**Above average**	**Genius**	**Total**
X	86	60	44	10	**200**
Y	40	33	25	2	**100**
Total	**126**	**93**	**69**	**12**	**200**

Would you say that sampling techniques adopted by the two researchers are significantly different.

Sol: $\chi^2\ value = 2.0971,\ d.f.(2-1)(4-1) = 3\ d.f. = 7.82\ at\ 5\%$

4. Four methods are under development for making discs of super conducting material. Fifty discs are made by each method and they are checked for super conductivity when cooled with liquid.

	I method	**II method**	**III method**	**IV method**	**Total**
Super conductors	31	42	22	25	**120**
Failures	19	8	28	25	**80**
Total	**50**	**50**	**50**	**50**	**200**

Test significant difference between the proportions of super conductivity at 0.05 level.

Sol: 1. $H_0: P_1 = P_2 = P_3 = P_4$

2. $H_1: P_1 \neq P_2 \neq P_3 \neq P_4$

3. $\alpha = 0.05$

4. Computation: Table of expected frequencies

$\frac{50 \times 120}{200} = 30$	$\frac{50 \times 120}{200} = 30$	$\frac{50 \times 120}{200} = 30$	$\frac{50 \times 120}{200} = 30$	120
$\frac{50 \times 80}{200} = 20$	$\frac{50 \times 80}{200} = 20$	$\frac{50 \times 80}{200} = 20$	$\frac{50 \times 80}{200} = 20$	80
50	50	50	50	200

Observed frequency (O)	**Expected frequency (E)**	**(O – E)²**	$\frac{(O-E)^2}{E}$
31	30	1	0.03
42	30	144	4.8
22	30	64	2.13
25	30	25	0.83
19	20	1	0.05
8	20	144	7.2
28	20	64	3.2
25	20	25	1.25
			19.5

5. Test statistic $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 19.5$

Degrees of freedom = (4 – 1)(2 – 1) = 3 d.f. at 5% level of significance is 7.815.

$\chi^2_{cal} > \chi^2_{tab},\ i.e., reject\ H_0.$

5. From the following data find whether there is any significant liking in the habit of taking soft drinks among categories of employees.

Soft Drinks	**Employees**			
	Clerk	**Teachers**	**Officers**	**Total**
Pepsi	10	25	65	**100**
Thumsup	15	30	65	**110**
Fanta	50	60	30	**140**
Total	**75**	**115**	**160**	**350**

Sol: 1. H_0: There is significant liking in the habit of taking soft drinks.

2. H_1: There is no significant liking in habit of taking soft drinks.

3. $\alpha = 0.05$

4. Computation: Table of expected frequencies

$\frac{75 \times 100}{350} = 21.4$	$\frac{115 \times 100}{350} = 32.9$	$\frac{160 \times 100}{350} = 45.71$
$\frac{75 \times 110}{350} = 23.6$	$\frac{115 \times 110}{350} = 36.1$	$\frac{160 \times 110}{350} = 50.3$
$\frac{75 \times 140}{350} = 30$	$\frac{115 \times 140}{350} = 46$	$\frac{160 \times 140}{350} = 64$

Observed frequency (O)	**Expected frequency (E)**	**(O – E)²**	$\frac{(O-E)^2}{E}$
10	21.4	129.96	6.073
25	32.9	62.41	1.897
65	45.7	372.49	8.151
15	23.6	73.96	3.134

30	36.1	37.21	1.031
65	50.3	216.09	4.3
50	30	400.0	13.333
60	46	196	4.261
30	64	1156.0	18.062
			60.2425

5. Test statistic $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 60.2425$

Degrees of freedom = (3 – 1)(3 – 1) = 4 d.f. at 5% level of significance is 9.488.

$\chi^2_{cal} > \chi^2_{tab},\ i.e., reject\ H_0.$

CHI – SQUARE TEST FOR POPULATION VARIANCE:

Suppose we want to test if a random sample X_i has been drawn from normal population with a specified variance σ^2. Under the null hypothesis H_0 that the population variance is σ^2 test statistic is

$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2}$

$\chi^2 = \frac{nS^2}{\sigma^2} \quad (n-1)d.f.$

EXAMPLE:

A firm manufacturing rivets wants to limit variations in their length as much as possible. The lengths in (cms) of 10 rivets manufactured by new process are

2.15	1.99	2.05	2.12	2.17
2.01	1.98	2.03	2.25	1.93

Examine whether the new process can be considered superior to the old if the old population has S.D. of 0.145 cm.

Sol: We have $n = 10,\ \bar{x} = \frac{\sum x_i}{N} = \frac{20.68}{10} = 2.068$

$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{0.09096}{9} = 0.01010$

$\sigma = 0.145$

1. $H_0 :\ \sigma^2 = \sigma_0^2\ (or) New\ process\ superior\ to\ old$

2. $H_1: \sigma^2 \neq \sigma_0^2$

3. α: 0.05

$\chi^2 = \frac{nS^2}{\sigma^2} = \frac{10(0.01010)}{(0.145)^2} = 4.8$

χ^2 table value of 9 d.f. is 16.916.

$\chi^2_{cal} < \chi^2_{tab},\ i.e., accept\ H_0.$

Key Points:

SMALL SAMPLE TESTS :

(i) t-test:

Name of the test	Null HypothesisH_0	Level of significance(α)	Test statistic
1.Test for single mean	$\mu = \mu_0$	5% or 1% or 10%	$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$
2.Test for difference of means	$\mu_1 = \mu_2$	5% or 1% or 10%	$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

3. paired t-test	μ =0	5% or 1% or 10%	$=\frac{\bar{d}}{\frac{s}{\sqrt{n}}}$

(ii) F-test:

Test for difference of variance

$H_0: {\sigma_1}^2 = {\sigma_2}^2$

α : 5% or 10%.

Test statistic F = $\frac{{S_1}^2}{{S_2}^2} = \frac{Greater\ variance}{smaler\ variance}$

(iii) Chi-square (χ^2)test:

Test for goodness of fit and test for independence of attributes.

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

PRACTICE PROBLEMS

t-TEST

1.Test for single mean

1. A mechanist id making engine parts with axle diameters of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a S.D of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specifications.
2. A Sample of 26bulbs give mean life of 990 hours with a S.D of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hour. Is the sample not up to the standard?
3. The average breaking strength of the steel rods is specified to be 18.5 thousand pounds. To test this sample of 14 rods were tested. The mean and S.D obtained were 17.85 and 1.955 respectively. Is the result of experiment significant?
4. A Random sample of six steel beams has a mean compressive strength of 58,392 p.s.i (pounds per square inch) with a standard deviation of 648p.s.i Use this information and the level of significance 0.05 to test the steel from which this sample came is 58,000p.s.i.

Problems related to student's t-test (When S.D of the sample is not given derectly)

1. A random sample of 10 boys had the following IQ 's : 70,120, 110 , 101 ; 88 ; 83 ,95 , 107 , and 100.
(a) do these data support the assumption of a population mean IQ of 100 ?
(b) Find a reasonable range in which most of them mean I Q values of samples of 10 boys ?
2. The heights of 10 males of a given locality are found to be 70, 67 , 62, 68 , 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches ? . Test at 5% level of siginificance that for 9 degrees of freedom (t = 1.833 at α = 0.05) .
3.Eight students were given a test in STASTICS and after one month coaching they were given another test of the similar nature . The following table gives the increase I their marks in second test over the first

Student nc	1	2	3	4	5	6	7	8
Increase of marks	4	_2	6	_8	12	5	_7	2

Do the marks indicate that the students have gained from the coaching .

4 . A random sample of 10 bags of pesticide are taken whose weights are 50, 49, 52, 44, 45, 48, 46, 45, 49, 45 (in kgs) .Test whether the average packing can be taken to be 50 kgs .

Student t _test for difference of means :

1. A group of 5 patients treated with medicine A weigh 42, 39, 48, 60 and 41 kgs . second group of 7 patients from the same hospital treated with medicine B weigh 38, 42, 56, 64, 68 , 69 and 62 kgs. Do you agree with the claim that medicine B increases the weigh siginificantly.

2. To examine the hypothesis that the husbands are more intelligent than the wives , an investigator took a sample of 10 couples and administered them a test which measures the I.Q. .the results are as follows .

Husband	117	105	97	105	123	109	86	78	103	107
Wives	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a reasonable test at the level of siginificance of 0.05 .

3 . two independent samples of 8 & 7 items respectively had the following values.

Sample I	11	11	13	11	15	9	12	14
Sample II	9	11	10	13	9	8	10	

Is the difference between the means of samples siginificant ?

4 . Two compare two kinds of bumper guards , 6 of each kind were mounted on a car and then the car was run into a concrete wall . The following are the costs of repairs

Guard 1	107	148	123	165	102	119
Guard 2	134	115	112	151	133	129

Use the 0.01 level of significance to test whether the difference between two sample means is significant.

5 . the table gives the biological values of protein from 6 cow's milk and 6 buffalo's milk. Examine the difference are significant.

cow's milk	1.8	2.0	1.9	1.6	1.8	1.5
buffalo's milk	2	1.8	1.8	2.0	2.1	1.9

Paired sample t test :

1. The blood pressure of 5 women before and after intake of a certain drug are given below :

before	110	120	125	132	125
after	120	118	125	136	121

2. Memory capacity of 10 students were tested before and after training . state whether the training was effective or not from the following scores

Before	12	14	11	8	7	10	3	0	5	6
after	15	16	10	7	5	12	10	2	3	8

3. The average losses of workers, before and after certain program are given below. Use 0.05 level of significance to test whether the program is effective (paired sample t-test). 40 and 35,70and 65,45, 42, 120 and 116, 35 and 55 and 50,77 and 73.

F-TEST☹Test for equality of two population variances:

1. In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5%level whether the populations have the same variance.

2. Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pupkins, show the sample standard deviation of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test hypothesis that the true variances are equal.

3. In one sample of 10 observations, the sum of the deviations of the sample values from sample mean was 120 and in the other sample of 12 observations it was 314. Test whether the difference is significant at 5% level.

4. Two random samples reveal the following results

Sample	Size	Sample mean	Sum of squares of deviations from the mean
1	10	15	90
2	12	14	108

Test whether the sample came from the same normal population.

5. the time taken by workers in performing a job by method I and method II is given below:

Method I	20	16	26	27	23	22	---

Method II	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?

CHI-SQUARE TEST FOR GOODNESS OF FIT:

1.. The number of automobile accidents per week in a certain community are as follows: 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 period week.

2.. A die is thrown 264 times with the following results. Show that the die is biased.(given chi-square 0.05 =11.07 for 5 d.f).

3. A pair of dice is thrown 360 tunes and the frequency of each sum is indicated below:

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the chi-square test at 0.05 level of significance.

4. A survey of 320 families with 5 children each revealed the following distribution.

No. of Boys	5	4	3	2	1	0
No. of Girls	0	1	2	3	4	5
No. of Families	14	56	110	88	40	12

CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

1.The following table gives the classification of 100 workers according to gender and nature of work. Test whether the nature of work is independent of the gender of the worker.

	Stable	Unstable	Total
Male	40	20	60
Female	10	30	40
Total	50	50	100

2. From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees.

Soft drinks	Clerks	Teachers	officers
Pepsi	10	25	65
Thumsup	15	30	65
Fanta	50	60	30

3. In an investigation on the machine performance, the following results are obtained.

	No.of units inspected	No.of defective
Machine1	375	17
Machine2	450	22

4. A survey of 240 families with 4 children each revealed the following distribution

Male Births	4	3	2	1	0
Female Births	10	55	105	58	12

Test whether the male and female births are equally popular.

****