Small Samples

When the value of $n < 30$, then the sample is called small sample. The values given by the sample data are sufficiently close to the population values.

t - Test or

Student's t - Distribution

This t-test is used when sample size ≤ 30 & the population standard deviation is unknown.

t-statistic is defined as $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1\ d.f.)$

where $s = \sqrt{\dfrac{\Sigma (X - \bar{X})^2}{n-1}}$, d.f $\rightarrow$ degree of freedom.

Test 1:- t- Test of significance of the Mean of a Random Sample.

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

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}, \quad \bar{X} \rightarrow \text{mean of sample}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}, \quad (n-1) \rightarrow \text{degree of freedom}$$

* Instead of calculating s, we calculate S for the sample.

Since $s^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ $\therefore$ $S^2 = \dfrac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

$$\Rightarrow \boxed{s^2 = \frac{n}{n-1} S^2}$$

Ques17:- A random sample of size 16 has 53 as a 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 a mean? obtain 95% & 99% confidence limits of the mean of the population.

Soln:- Given: $n = 16$, $\bar{X} = 53$, $\mu = 56$

$\Sigma (X - \bar{X})^2 = 135.$

$H_0 : \mu = 56$

$H_1 : \mu \neq 56$ (Two tailed test).

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

$$s = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n-1}} = \sqrt{\frac{135}{15}} = 3$$

$$\Rightarrow t = \frac{53 - 56}{3 / \sqrt{16}} = \frac{-3 \times 4}{3} = -4$$

$|t| = 4$, $d.f = 16 - 1 = 15$

Conclusion:- $t_{0.05} = 2.13$ i.e. The calculated value of t is more than the tabular value. Ho is rejected. Hence the sample mean has not come from a population having 56 as mean.

95% confidence limits of the population mean

$$\bar{X} \pm \frac{s}{\sqrt{n}} t_{0.05} = 53 \pm \frac{3}{\sqrt{16}} (2.13) = 53 \pm 1.5975 = (51.4025, 54.5975)$$

99% confidence limits of the population mean.

$$\bar{X} \pm \frac{s}{\sqrt{n}} t_{0.01} = 53 \pm \frac{3}{\sqrt{16}} (2.95) = 53 \pm 2.2125 = (50.7875, 55.2125)$$

Q27 The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data:

Item	1	2	3	4	5	6	7	8	9	10
life in '000hrs	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average life time of bulb is 4000 hrs?

Soln $n = 10$, $\mu = \frac{4000}{1000}$ hrs. $\Rightarrow \mu = 4$ hrs

$H_0: \mu = 4$ i.e. No significant diff. in sample & population mean.

$H_1: \mu \neq 4$.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} , \quad \bar{x} = \frac{\Sigma x}{n} = \frac{44}{10} = 4.4.$$

x :	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6
$(x-\bar{x})^2$:	0.04	0.04	0.25	0.09	0.64	0.36	0.25	0.01	0	1.44.

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{3.12}{9}} = 0.589.$$

$$t = \frac{4.4 - 4}{0.589/\sqrt{10}} = 2.123$$

From table $t_{0.05} = 2.26$ for 9 d.f.

Conclusion:- since the calculated value of $t < 2.26$. So, the H_0 is accepted i.e. $\mu = 4000$ hrs.

$\therefore$ The average life time of bulbs could be 4000 hours.

Q3] A sample of 20 items has mean 42 units & S.D. 5 units. Test the hypothesis that it is a random sample from a normal population with mean 45 units.

Sol^n $\mu = 45$ units, $n = 20$, $\bar{x} = 42$, $\nu = 19$ d.f., $S = 5$.

$H_0: \mu = 45$ units

$H_1: \mu \neq 45$ units [two tailed test].

$$s^2 = \frac{n}{n-1} S^2 = \left[\frac{20}{19}\right](5)^2 = 26.31 \Rightarrow \boxed{s = 5.129}$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{42-45}{5.129/\sqrt{20}} = -2.615 \Rightarrow |t| = 2.615$$

From the table $t_{0.05} = 2.09$ with 19 df.

Conclusion:- Since $|t| > t_{0.05}$, H_0 is rejected.

i.e. there is no significant difference b/w the sample mean & population mean.

Q4]:- The 9 items of a sample have following values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these values differ significantly from the assumed mean 47.5

Sol^n $\mu = 47.5$, $\bar{x} = 49.11$, $s = 2.619$, $\nu = 8$ df.

$t = 1.7279$, $t_{0.05} = 2.31$ for $\nu = 8$ df.

Conclusion: H_0 is accepted.

Test II :- t - Test for difference of means of two small samples.

This test is used to test whether the two samples, $x_1, x_2, \ldots, x_{n_1}$; $y_1, y_2, \ldots, y_{n_2}$ of sizes n_1, n_2 have been drawn from two normal populations with mean μ_1 & μ_2 resp. under the assumption that the population variance are equal ($\sigma_1 = \sigma_2 = \sigma$).

$$t = \frac{\bar{X} - \bar{Y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2\ d.f.)$$

under the H_0: $\mu_1 = \mu_2$. & H_1: $\mu_1 \neq \mu_2$

* If the S.D of two samples s_1, s_2 are given then we have $s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$.

* If $n_1 = n_2 = n$, $t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2 + s_2^2}{n-1}}}$.

* If the pairs of the values are in some way correlated, we can't use the test statistic as given above. In this case we find the differences of the correlated pairs of values & apply for single mean.

ie $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$, with $n-1$ d.f.

The test statistic is $t = \frac{\bar{d}}{s/\sqrt{n}}$ or $t = \frac{\bar{d}}{s/\sqrt{n-1}}$; where $\bar{d}$ is mean of paired difference

ie $d_i = x_i - y_i$ & $\bar{d_i} = \frac{\Sigma d_i}{n}$, where (x_i, y_i) are paired data $i = 1, 2, 3, \ldots, n$.

Scanned by CamScanner

Quesn- Eleven school boys were given a test in drawing. They were given a month's further tuition & a second test of equal difficulty was held at the end of it. Do the marks give evidence that the students have benefitted by extra coaching?

Soln.

Boys:	1	2	3	4	5	6	7	8	9	10	11
Marks I test:	23	20	19	21	18	20	18	17	23	16	19
Marks II test:	24	19	22	18	20	22	20	20	23	20	17

Soln:- $n = 11$, $\nu = 11 - 1 = 10$.

H_0: There is no significant difference b/w Marks of I^{st} test & II^{nd} test ie $\mu_1 = \mu_2$

H_1: There is a significant difference ie. $\mu_1 \neq \mu_2$

$$t = \frac{\bar{d}}{s/\sqrt{n}}, \quad s = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}}, \quad \bar{d} = \frac{\Sigma d}{n} = \frac{11}{11} = 1.$$

Boys	x_1	x_2	$d = x_2 - x_1$	$d - \bar{d}$	$(d-\bar{d})^2$
1.	23	24	1	0	0
2.	20	19	-1	-2	4
3.	19	22	3	2	4
4.	21	18	-3	-4	16
5.	18	20	2	1	1
6.	20	22	2	1	1
7.	18	20	2	1	1
8.	17	20	3	2	4
9.	23	23	—	-1	1
10.	16	20	4	3	9
11.	19	17	-2	-3	9
			$\Sigma d = 11$		$\Sigma(d-\bar{d})^2 = 50$

$$s = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = \sqrt{\frac{50}{10}} = 2.24$$

$$t = \frac{\bar{d}-\mu}{s/\sqrt{n}} \quad \text{i.e.} \quad \frac{\bar{d}}{s/\sqrt{n}} = \frac{1}{2.24/\sqrt{11}} = 1.48$$

From table $t_{0.05} = 2.23$ for $\nu = 10$

Conclusion:- The value of $t < t_{0.05}$. So, H_0 is accepted. i.e. There is no significant diff. b/w marks of two tests. i.e. The marks give no evidence that the students have benefitted by extra coaching.

Ques 27:- A group of boys & girls were given an intelligence test. The mean score, S.D's & numbers in each group are as follows:

	Mean	S.D	n
Boys	124	12	18
Girls	121	10	14

Is the mean score of boys significantly differ from that of girls?

Ans: $n_1 = 18$, $n_2 = 14$, $\bar{X_1} = 124$, $\bar{X_2} = 121$

$s_1 = 12$, $s_2 = 10$, $d.f = n_1+n_2-2 = 30$

$H_0 : \mu_1 = \mu_2$ i.e. No significant difference b/w mean scores of boys & girls.

$H_1 : \mu_1 \neq \mu_2$

$$s^2 = \frac{n_2 s_2^2 + n_1 s_1^2}{n_1+n_2-2} = \frac{18\times(12)^2 + 14\times(10)^2}{18+14-2} = \frac{2592+1400}{30}$$

$$= 133.0667$$

$\therefore s = 11.5355$

The t-statistic, $t = \dfrac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \dfrac{124-121}{11.5355\sqrt{\frac{1}{18}+\frac{1}{14}}} = \dfrac{3}{11.5355 \times 0.12698}$

$\Rightarrow t = \dfrac{3}{1.4648} = 2.0481$

$t_{0.05}$ at d.f. 30 = 2.04

Conclusion: As the calculated value of $t > t_{0.05}$ at d.f is 30. So, Ho is rejected i.e. mean score of boys is significantly differ from girls.

Ques 37) Two samples of sodium vapour bulbs were tested for length of life & the following results were got:

	size	Sample mean	Sample S.D.
Type I	8	1234 hrs	36 hrs
Type II	7	1036 hrs	40 hrs

Is the difference in the means significant to generalise the type I is superior to type II regarding length of life?

Soln: $n_1 = 8$, $n_2 = 7$, $\bar{X}_1 = 1234$, $\bar{X}_2 = 1036$, $s_1 = 36$, $s_2 = 40$.

$H_0: \mu_1 = \mu_2$ i.e. two types of bulbs have same lifetime.

$H_1: \mu_1 > \mu_2$ i.e. type I is superior to type II

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

$\therefore s = 40.7318$

The t-statistic i.e. $t = \dfrac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \dfrac{1234-1036}{40.7318\sqrt{\frac{1}{8}+\frac{1}{7}}} = \dfrac{198}{21.081} = 9.39$

$t_{0.05}$ at d.f 13 is 1.77 [10% level of significant of one tail test]

Conclusion:- Since calculated $|t| > t_{0.05}$. So, H_0 is rejected.
ie Type I is definitely superior to type II

Ques47:- Samples of size 10 & 14 were taken from two normal populations with S.D 3.5 & 5.2. The sample means were found to be 20.3 & 18.6. Test whether the means of two populations are the same at 5% level.

Soln Do yourself.

$S = 4.772$; $|t| = 0.8604$
$t_{0.05} = 2.07$ for 22 df.
Conclusion :- H_0 is accepted.

Ques5] The height of 6 randomly chosen sailors are in inches are 63, 65, 68, 69, 71 & 72. Those of 9 randomly chosen soldiers are 61, 62, 65, 66, 69, 70, 71, 72 & 73. Test whether the sailors are on the average taller than soldiers.

Soln $n_1 = 6$, $n_2 = 9$, Let X_1 & X_2 be two samples denoting the heights of sailors & soldiers.

$H_0: \mu_1 = \mu_2$ ie. mean of both the population are same.

$H_1: \mu_1 > \mu_2$ (one tail test)

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{63+65+68+69+71+72}{6} = 68$$

$$\bar{X}_2 = \frac{\Sigma X_2}{n_2} = 67.667$$

Scanned by CamScanner

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
72	4	16	70	2.34	5.4756
			71	3.34	11.1556
			72	4.34	18.8356
			73	5.34	28.5156

$\Sigma(X_1-\overline{X_1})^2 = 60$ $\qquad$ $\Sigma(X_2-\overline{X_2})^2 = 152.0002$

$$s = \sqrt{\frac{\Sigma(x_1-\overline{x_1})^2+\Sigma(x_2-\overline{x_2})^2}{n_1+n_2-2}} = \sqrt{\frac{60+152.0002}{6+9-2}} = 4.038$$

$$t = \frac{\overline{x_1}-\overline{x_2}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{68-67.667}{4.038\sqrt{\frac{1}{6}+\frac{1}{9}}} = 0.1565$$

$t_{0.10}$ ~~for~~ d.f 13 is 1.77.

Conclusion :- The calculated $|t| < t_{0.10}$. So, H_0 is accepted i.e. there is no significance diff. b/w their average.

Ques 6) Memory capacity of 9 students was tested before & after a course of meditation for a month. State whether the course was effective or not from the data below:-

Before	10	15	9	3	7	12	16	17	4
After	12	17	8	5	6	11	18	20	3

Sol^n Do yourself.

$\bar{d} = -0.778$, $s = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = 1.6177$

$|t| = 1.359$, $t_{0.05}$ for 8 d.f $= 2.31$

H_0 is accepted.

Ques 7) The following figures refer to observations in live independent samples.

Sample I :	25	30	28	34	24	20	13	32	22	38
Sample II :	40	34	22	20	31	40	30	23	36	17

Analyse whether the samples have been drawn from the population of equal means.

Sol^n:- Do yourself

$\bar{X}_1 = 26.6$, $\bar{X}_2 = 29.3$

$\Sigma(X_1 - \bar{X}_1)^2 = 486.4$, $\Sigma(X_2 - \bar{X}_2)^2 = 630.08$

$s = 7.875$, $|t| = 0.7666$

$t_{0.05}$ for 18 d.f is 2.10

F-Test or Snedecor's Variance ratio test

The F-test is to discover whether the two independent estimates of population variance differ significantly or whether the two samples may be regarded as drawn from the normal populations having the same variance. Hence before applying t-test for the significance of the difference of two means, we have to test for the equality of population variance by using F-test.

Let n_1 & n_2 be sizes of two samples with variances S_1^2 & S_2^2. The estimates of the population variance based on these samples are $s_1^2 = \frac{n_1 S_1^2}{n_1 - 1}$ & $s_2^2 = \frac{n_2 S_2^2}{n_2 - 1}$.

The degree of freedom of these estimates are $\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$.

We set up the null hypothesis $H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$.

To carry out the test of significance of difference of variances, we calculate $F = \frac{s_1^2}{s_2^2}$, $s_1^2 > s_2^2$.

<u>Applications</u>:
1) whether two independent samples have been drawn from normal populations with the same variance σ^2.
2) whether two independent estimates of population variance are homogeneous or not.

Q1]: In two independent samples of sizes 8 & 10 the sum of squares of deviations of the sample values from respective sample means were 84.4 & 102.6. Test whether the difference of variances of the populations is significant or not.

Soln $n_1 = 8$, $n_2 = 10$, $\Sigma(x_1-\bar{x}_1)^2 = 84.4$, $\Sigma(x_2-\bar{x}_2)^2 = 102.6$

H_0: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ i.e, there is no significant difference b/w population variance.

H_1: $\sigma_1^2 \neq \sigma_2^2$.

$$F = \frac{s_1^2}{s_2^2} \sim F(v_1, v_2 \text{ df})$$

$v_1 = n_1 - 1 = 8 - 1 = 7$, $v_2 = n_2 - 1 = 10 - 1 = 9$.

$$s_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n-1} = \frac{84.4}{7} = 12.057, \quad s_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1} = \frac{102.6}{9} = 11.4$$

$$F = \frac{s_1^2}{s_2^2} = \frac{12.057}{11.4} = 1.0576 \quad [\because s_1^2 > s_2^2]$$

Conclusion: $F_{0.05}$ for $(7, 9)$ d.f is 3.29.

Q27: Two random samples are drawn from 2 normal populations are as follows

A	17	27	18	25	27	29	13	17
B	16	16	20	27	26	25	21	

Test whether the samples are drawn from the same normal populations.

Sol^n To test if two independent samples have been drawn from the same population we have to test

(i) equality of means by applying t-test.

(ii) equality of population variance by applying F-test.

Since t-test assumes that the samples variances are equal, we shall first apply the F-test.

F-test: $H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$ i.e. population variance do not differ significantly.

$H_1: \sigma_1^2 \neq \sigma_2^2$

$n_1 = 8$, $n_2 = 7$, $\bar{X}_1 = \frac{\Sigma X_1}{n_1} = 21.625$, $\bar{X}_2 = \frac{\Sigma X_2}{n_2} = 18.714$ 21.55

$F = \frac{s_1^2}{s_2^2}$, (if $s_1^2 > s_2^2$)

X_1	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	X_2	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
17	-4.625	21.39	16	-2.714	7.365
27	5.735	28.89	16	-2.714	7.365
18	-3.625	13.14	20	1.286	1.653
25	3.375	11.39	27	8.286	68.657
27	5.735	28.89	26	7.286	53.085
29	7.735	54.39	25	6.286	39.513
13	-8.625	74.39	21	2.286	5.226
17	-4.625	21.39			

$\Sigma(X_1 - \bar{X}_1)^2 = 253.87$ $\Sigma(X_2 - \bar{X}_2)^2 = 182.859$

$$s_1^2 = \frac{\Sigma(X_1-\bar{X}_1)^2}{n_1-1} = \frac{253.87}{7} = 36.267$$

$$s_2^2 = \frac{\Sigma(X_2-\bar{X}_2)^2}{n_2-1} = \frac{182.859}{6} = 30.477$$ 2061

$$F = \frac{s_1^2}{s_2^2} = \frac{36.267}{30.477} = 1.190$$

Conclusion:- $F_{0.05}$ for $(7,6)$ d.f is 4.21.

The calculated $F < F_{0.05}$. $\therefore$ H_0 is accepted. Hence the variance of the two populations is same.

t-test:- $H_0: \mu_1 = \mu_2$ i.e the population means are equal.

$H_1: \mu_1 \neq \mu_2$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad ; \quad s = \sqrt{\frac{\Sigma(X_1-\bar{X}_1)^2 + \Sigma(X_2-\bar{X}_2)^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{253.87+182.859}{8+7-2}} = 5.796$$

$$\Rightarrow t = \frac{21.625 - 18.714}{5.796\sqrt{\frac{1}{8}+\frac{1}{7}}} = 0.9704$$

Conclusion: $t_{0.05}$ for 13 d.f is 2.16.

The calculated $t < t_{0.05}$. H_0 is accepted i.e. the population means are equal i.e. $\mu_1 = \mu_2$.

$\therefore$ we conclude that the two samples have been drawn from the same normal population.

Q3) Two independent samples of sizes 7 & 6 had following values:-

Sample A	28	30	32	33	31	29	34
Sample B	29	30	30	24	27	28	

Examine whether the samples have been drawn from normal populations having the same variance.

F-test.

Sol^n Do yourself.

$\bar{x}_1 = 31$, $\bar{x}_2 = 28$, $\Sigma(x_1-\bar{x}_1)^2 = 28$,
$\Sigma(x_2-\bar{x}_2)^2 = 26$, $s_1^2 = 4.666$, $s_2^2 = 5.2$
$F = 1.1158$, F for $(5,6)$ at 5% $= 4.39$

Q4) Two random samples reveal the following data:-

Sample no.	size	Mean	variance
I	16	440	40
II	25	460	42

Test whether the samples come from the same normal population.

Sol^n Do yourself

Apply F-test then t-test

$$F = \frac{s_1^2}{s_2^2} = \frac{\frac{n_1 s_1^2}{n_1-1}}{\frac{n_2 s_2^2}{n_2-1}} = \frac{16\times 40}{15}\times\frac{24}{25\times 42} = 0.9752$$

$F_{0.05}$ for $(15,24)$ df $= 2.11$

$|t| = 9.490$, $s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}}$, $t_{0.05}$ for 39 df $= 1.96$

Q5) Two random samples drawn from two normal populations have the variable values as below:

Sample I	19	17	16	28	22	23	19	24	26			
Sample II	28	32	40	37	30	35	40	28	41	45	30	36

Obtain estimate variance of population & test two population have same variance.

Sol^n Do yourself

$\bar{x}_1 = 21.55$, $\bar{x}_2 = 35.166$, $\Sigma(x-\bar{x}_1)^2 = 134.2$, $\Sigma(x_2-\bar{x}_2)^2 = 347.664$
$s_1^2 = 16.775$, $s_2^2 = 31.606$, $F = \frac{s_2^2}{s_1^2} = 1.884$.
F for $(11,8)$ df $= 3.315$.

CHI-SQUARE (χ^2) Test

χ^2 affords a measure of the correspondence b/w theory & observation.

If O_i $(i=1,2,--,n)$ is a set of observations (experimental) frequencies & E_i $(i=1,2,-,n)$ is the corresponding set of expected (theoretical or hypothetical) frequencies, then χ^2 is defined as

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

where $\Sigma O_i = \Sigma E_i = N$ (total frequency) & $d.f = n-1$.

Note: (i) If $\chi^2 = 0$, the observed & theoretical frequencies agree exactly.

(ii) If $\chi^2 > 0$, they do not agree exactly.

Conditions for Applying χ^2-Test

(a) N should be atleast 50.

(b) No theoretical cell-frequency should be small. If small theoretical frequency occur (i.e. <10), the difficulty is overcome by grouping two or more classes together before calculating $(O-E)$.

In the case of Binomial distribution $d.f = n-1$

" " " " Poisson " $d.f = n-2$

" " " " Normal " $d.f = n-3$

Q1] The following table gives the no. of accidents that took place in an industry during various days of the week. Test if accidents are uniformly distributed over the week.

Day :	Mon	Tue	Wed	Thu	Fri	Sat
No. of accidents :	14	18	12	11	15	14

Sol^n H_0: The accidents are uniformly distributed over the week.

Under this H_0, the expected frequencies of the accidents on each of these days $= \frac{84}{6} = 14$

Observed frequency O_i	14	18	12	11	15	14
Expected frequency E_i	14	14	14	14	14	14
$(O_i - E_i)^2$	0	16	4	9	1	0

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i} = \frac{30}{14} = 2.1428$$

Conclusion:- From table χ^2 at 5% for 5 d.f is 11.09.

Since the calculated value of χ^2 < tabulated value; H_0 is accepted i.e. accidents are uniformly distributed over the week.

Q2] A die is thrown 270 times & the results of these throws are given below:-

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

Test whether the die is biased or not.

Sol^n H_0: Die is unbiased.

The expected frequencies 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	$\Sigma(O_i-E_i)^2 = 980$

$$\chi^2 = \frac{\Sigma(O_i-E_i)^2}{E_i} = \frac{980}{46} = 21.30$$

Conclusion: From table $\chi^2_{0.05}$ for 5 df is 11.09.

Since calculated $\chi^2 > \chi^2_{0.05}$ $\therefore$ H_0 is rejected.

i.e. Die is not unbiased.

Q3] The following table shows the distribution of digits in numbers chosen at random from a telephone directory:

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

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

Sol^n Do yourself

$E_i = \frac{10,000}{10} = 1000$, $\Sigma(O_i-E_i)^2 = 58542$

$\chi^2 = 58.542$.

Q4) Records taken of the no. 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 the binomial law holds & the chance of male birth is equal to that of female birth namely $p = q = 1/2$.

Sol^n H_0: The data are consistent with the hypothesis that the ~~binomial~~ equal probability for male & female births i.e $p = q = 1/2$

To calculate theoretical frequency, we use binomial distribution given by :-

$N(r) = N \times {}^nC_r\, p^r q^{n-r}$; $N(r) \rightarrow$ no. of families with r male children

$N(0) = 800 \times {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 = 50$

$N(1) = 800 \times {}^4C_1 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)^3 = 200$

$N(2) = 800 \times {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = 300$

$N(3) = 800 \times {}^4C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right) = 200$

$N(4) = 800 \times {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0 = 50$

O_i	32	178	290	236	94
E_i	50	200	300	200	50
$(O_i - E_i)^2$	324	484	100	1296	1936
$\frac{(O_i - E_i)^2}{E_i}$	6.48	2.42	0.333	6.48	38.72

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i}\right] = 54.433$$

Conclusion :- from table $\chi^2_{0.05}$ for 4 d.f = 9.49

since calculated $\chi^2 > \chi^2_{0.05}$ $\therefore$ H_0 is rejected

i.e the data are not consistent with the hypothesis that the binomial law holds & $p \neq q \neq 1/2$

Q5) Verify whether Poisson distribution can be assumed from the data given below:- (11)

No. of defects : 0 1 2 3 4 5

Frequency : 6 13 13 8 4 3

Sol^n H_0 : Poisson fit is a good fit data

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{94}{47} = 2 \quad \text{ie } \lambda = 2.$$

By Poisson distribution the freq. of ~~r~~ success is

$$N(r) = N \times \frac{e^{-\lambda}\lambda^r}{r!}, \quad N \to \text{total frequency.}$$

$$N(0) = 47 \times \frac{e^{-2}(2)^0}{0!} = 6.36, \quad N(1) = \frac{47 \times e^{-2}(2)^1}{1!} = 12.72$$

$$N(2) = \frac{47 \times e^{-2} \times (2)^2}{2!} = 12.72, \quad N(3) = \frac{47 \times e^{-2}(2)^3}{3!} = 8.48$$

$$N(4) = \frac{47 \times e^{-2} \times (2)^4}{4!} = 4.24, \quad N(5) = \frac{47 \times e^{-2} \times (2)^5}{5!} = 1.696$$

O_i :	6	13	13	8	4	3
E_i :	6.36	12.72	12.72	8.48	4.24	1.696
$(O_i - E_i)^2$:	0.1296	0.0784	0.0784	0.2304	0.0576	1.7004
$\frac{(O_i - E_i)^2}{E_i}$:	0.0204	0.0062	0.0062	0.0272	0.0135	1.0026

$$\chi^2 = \sum\left[\frac{(O_i - E_i)^2}{E_i}\right] = 1.0761$$

Conclusion: From table $\chi^2_{0.05}$ for $4(6-2)$ df is 9.49

Since the calculated $\chi^2 < \chi^2_{0.05}$ $\therefore$ H_0 is accepted.

i.e. Poisson distribution is a good fit to the data.

Q6/ The theory predicts the proportion of beans in 4 groups G_1, G_2, G_3, G_4 should be in the ratio 9 : 3 : 3 : 1. In an experiment with 1600 beans the numbers in the four groups were 882, 313, 287 & 118. Does the experimental results support the theory.

Sol? H_0: The experimental result support the theory.

The theoretical frequency can be calculated as follows :-

$E(G_1) = \frac{1600 \times 9}{16} = 900$, $E(G_2) = \frac{1600 \times 3}{16} = 300$

$E(G_3) = \frac{1600 \times 3}{16} = 300$, $E(G_4) = \frac{1600 \times 1}{16} = 100$

O_i	:	882	313	287	118
E_i	:	900	300	300	100
$\frac{(O_i - E_i)^2}{E_i}$	:	0.36	0.5633	0.5633	3.24

$$\chi^2 = \Sigma \left[\frac{(O_i - E_i)^2}{E_i}\right] = 4.7266$$

Conclusion :- from table $\chi^2_{0.05}$ for $(n-1) = 3$ d.f is 7.815

Since the Calculated $\chi^2 < \chi^2_{0.05}$ ∴ H_0 is accepted.

i.e. the experimental result support the theory.

Scanned by CamScanner

Table IX : 5% and 1% points of F.

ν_1 / ν_2	1	2	3	4	5	6	8	12	24	∞
2	18·51	19·00	19·16	19·25	19·30	19·32	19·37	19·41	19·45	19·50
	98·49	99·00	99·17	99·25	99·30	99·33	99·36	99·42	99·46	99·50
3	10·13	9·55	9·28	9·12	9·01	8·94	8·84	8·74	8·64	8·53
	34·12	30·82	29·46	28·71	28·24	27·91	27·49	27·05	26·60	26·12
4	7·71	6·94	6·59	6·39	6·26	6·16	6·04	5·91	5·77	5·63
	21·20	18·00	16·69	15·98	15·52	15·21	14·80	14·37	13·93	13·46
5	6·61	5·79	5·41	5·19	5·05	4·95	4·82	4·68	4·53	4·36
	16·26	13·27	12·06	11·39	10·97	10·67	10·27	9·89	9·47	9·02
6	5·99	5·14	4·76	4·53	4·39	4·28	4·15	4·00	3·84	3·67
	13·74	10·92	9·78	9·15	8·75	8·47	8·10	7·72	7·31	6·88
7	5·59	4·74	4·35	4·12	3·97	3·87	3·73	3·57	3·41	3·23
	12·25	9·55	8·45	7·85	7·46	7·19	6·84	6·47	6·07	5·65
8	5·32	4·46	4·07	3·84	3·69	3·58	3·44	3·28	3·12	2·93
	11·26	8·65	7·59	7·01	6·63	6·37	6·03	5·67	5·28	4·86
9	5·12	4·26	3·86	3·63	3·48	3·37	3·23	3·07	2·90	2·71
	10·56	8·02	6·99	6·42	6·06	5·80	5·47	5·11	4·73	4·31
10	4·96	4·10	3·71	3·48	3·33	3·22	3·07	2·91	2·74	2·54
	10·04	7·56	6·55	5·99	5·64	5·39	5·06	4·71	4·33	3·91
12	4·75	3·88	3·49	3·26	3·11	3·00	2·85	2·69	2·50	2·30
	9·33	6·93	5·95	5·41	5·06	4·82	4·50	4·16	3·78	3·36
14	4·60	3·74	3·34	3·11	2·96	2·85	2·70	2·53	2·35	2·13
	8·86	6·51	5·56	5·03	4·69	4·46	4·14	3·80	3·43	3·00
16	4·49	3·63	3·24	3·01	2·85	2·74	2·59	2·42	2·24	2·01
	8·53	6·23	5·29	4·77	4·44	4·20	3·89	3·55	3·18	2·75
18	4·41	3·55	3·16	2·93	2·77	2·66	2·51	2·34	2·15	1·92
	8·28	6·01	5·09	4·58	4·25	4·01	3·71	3·37	3·01	2·57
20	4·35	3·49	3·10	2·87	2·71	2·60	2·45	2·28	2·08	1·84
	8·10	5·85	4·94	4·43	4·10	3·87	3·56	3·23	2·86	2·42
25	4·24	3·38	2·99	2·76	2·60	2·49	2·34	2·16	1·96	1·71
	7·77	5·57	4·68	4·18	3·86	3·63	3·32	2·99	2·62	2·17
30	4·17	3·32	2·92	2·69	2·53	2·42	2·27	2·09	1·89	1·62
	7·56	5·39	4·51	4·02	3·70	3·47	3·17	2·84	2·47	2·01
40	4·08	3·23	2·84	2·61	2·45	2·34	2·18	2·00	1·79	1·51
	7·31	5·18	4·31	3·83	3·51	3·29	2·99	2·66	2·29	1·81
60	4·00	3·15	2·76	2·52	2·37	2·25	2·10	1·92	1·70	1·39
	7·08	4·98	4·13	3·65	3·34	3·12	2·82	2·50	2·12	1·60