MODULE-4

Statistical Inference 2

Sampling variables, central limit theorem and confidences limit for unknown mean. Test of Significance for means of two small samples, student's-'t' distribution, Chi-square distribution as a test of goodness of fit. F-Distribution.

Sampling of variables: Each member of the population gives a value of variable and the population is a frequency distribution of variables.

Thus a random sample of size n from the population is same as selecting n values of variables from those of the distribution.

Sampling distribution of the mean: If a population is distributed normally with mean μ and standard deviation σ, then the means of all positive random samples of size n, are also distributed normally with mean μ and standard error $\frac{\sigma}{\sqrt{n}}$.

Central limit theorem: If the variable X has a non-normal distribution with mean μ and standard deviation σ,

then the limiting distribution of $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ as $n \longrightarrow \infty$, is the standard normal distribution.

This theorem holds good for a sample of 25 or more which is regarded as large.

For unknown mean if $\bar{x}$ is not significant at 5% level of probability then $\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$.
Similarly for 99% confidence limit $\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}$.

Examples:

1. A sample of 900 members is found to have a mean of $3.4cm$. Can it be reasonably regarded as truly random sample from a large population with mean $3.25cm$ and standard deviation 1.61cm?

Solution: Given that $\bar{x} = 3.4cm, \quad n = 900, \;\; \mu = 3.25 \;\; and \;\; \sigma = 1.61cm.$

$$\therefore \;\; z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3.4-3.25}{\frac{1.61}{30}} = 2.795 > 1.96.$$

Deviation of sample mean from the mean is significant, and hence it cannot be regarded as random sample.

2. The mean of a certain normal population is equal to the standard error of the mean of samples of 100 from that distribution. Find the probability that the mean of the sample of 25 from the distribution will be negative.

Solution: Let μ be the mean and σ standard deviation of the distribution.

Given that for $n = 100$, μ = standard error = $\frac{\sigma}{\sqrt{n}}$

$$\Longrightarrow \mu = \frac{\sigma}{10}.$$

For $n = 25$, we have $z = \frac{\bar{x}-\mu}{\frac{\sigma}{5}} = \frac{5\bar{x}}{\sigma} - 0.5$

Since $\bar{x} < 0, \quad z < -0.5.$

Therefore $P(\bar{x} < 0) = P(Z < -0.5) = 0.5 - A(0.5) = 0.5 - 0.1915 = 0.3085.$

3. An unbiased coin is thrown n times. It is desired that the relative frequency of the appearance of heads should lie between 0.49 and 0.51. Find the smallest value of n that will ensure this result with 90% confidence.

Solution: Standard error of proportion of heads= $\sqrt{\frac{pq}{n}} = \sqrt{\frac{1}{4n}} = \frac{1}{2\sqrt{n}}$.

90% confidence means $A(z) = 0.45$ in standard normal variate. Therefore, $z = 1.645$.

Since relative frequency of the appearance of heads should lie between 0.49 and 0.51,

$$p - \frac{1.645}{2\sqrt{n}} < p_1 < p + \frac{1.645}{2\sqrt{n}} \quad \Longrightarrow \quad 0.5 - \frac{1.645}{2\sqrt{n}} < p_1 < 0.5 + \frac{1.645}{2\sqrt{n}} \ .$$

Given that, $0.49 < p_1 < 0.51$.

$$0.49 = 0.5 - \frac{1.645}{2\sqrt{n}} \ \ and \ \ 0.51 = 0.5 + \frac{1.645}{2\sqrt{n}} \ .$$

$$\Longrightarrow \frac{1.645}{2\sqrt{n}} = 0.01 \quad \Longrightarrow \quad \sqrt{n} = \frac{1.645}{0.02} = \frac{329}{4} \quad \Longrightarrow n \approx 6765.$$

Test of significance for means of two large samples:

Standard errors of the means of the two samples of same population is $e = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ and $z = \frac{\overline{x_1} \sim \overline{x_2}}{e}$.

If $z > 3$, then the sampling is not simple or samples are not drawn from the same population.
If $z > 1.96$, then the difference is significant at 5% level of significance.

Standard errors of the means of the two samples of different populations is $e = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ and $z = \frac{\overline{x_1} - \overline{x_2}}{e}$.

Examples:

1. The means of simple samples of sizes 1000 and 2000 are 67.5 and 68.0 cm respectively. Can the samples are drawn from the same population of S.D. 2.5cm?

Solution: Let samples are drawn from the same population of S.D. 2.5cm.

Given that $\overline{x_1} = 67.5$, $\overline{x_2} = 68.0$, $n_1 = 1000$, $n_2 = 2000$ and $\sigma = 2.5$.

$$e = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.0968 \ ,$$

$$z = \frac{\overline{x_1} \sim \overline{x_2}}{e} = \frac{0.5}{0.0968} = 5.16 > 2.58.$$ Difference is significant, hypothesis is rejected.

Hence the samples are not drawn from the same population.

2. A sample of height of 6400 soldiers has a mean of 67.85 inches and S.D. 2.56 inches while a sample of heights of 1600 sailors has a mean of 68.55 inches and S.D. 2.52 inches. Do the data indicate that sailors are on the average taller than the soldiers?

Solution: H_1: Sailors are on the average taller than the soldiers.
H_0: Sailors and soldiers are on the average same height.

Given that $\overline{x_1} = 67.85$, $\overline{x_2} = 68.55$, $n_1 = 6400$, $n_2 = 1600$, $\sigma_1 = 2.56$, $\sigma_2 = 2.52$.

$$e = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 0.0707 \ .$$

and $z = \frac{\overline{x_2} - \overline{x_1}}{e} = \frac{0.7}{0.0707} = 9.9010 > 2.325$. (1% level of significance in one-tailed test).

This is highly significant. H_0 is rejected and hence alternate hypothesis H_1 is accepted.

Therefore sailors are on the average taller than the soldiers.

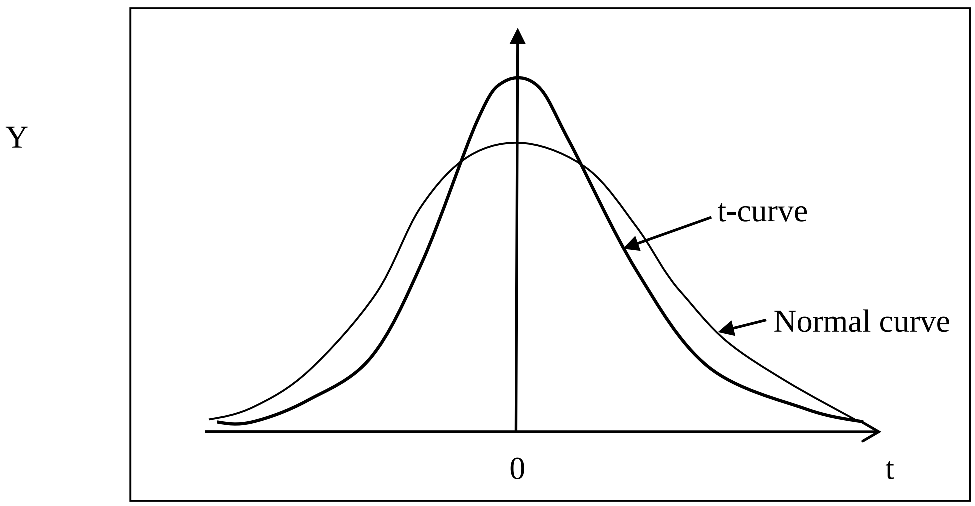
Student's t − Distribution: Consider a small sample of size n, drawn from a normal population with mean μ and S.D. σ . If $\overline{x}$ *and* σ_s be the sample mean and S.D. Then the statistic, t is defined as

$$t = \frac{\overline{x} - \mu}{\sigma_s}\sqrt{n-1}, \qquad \text{where } \nu = n - 1 \text{ denotes the degree of freedom of } t.$$

Sampling distribution for t is called Student's t − Distribution is given by $y = \frac{y_0}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}$.

The probability P that the value of t will exceed t_0 is $P = \int_{t_0}^{\infty} y \, dt$

Where y_0 is a constant such that the area under the curve is unity.



Significance test of a sample mean: Given a random sample $x_1, x_2, x_3, \cdots \cdots x_n$ from a normal population, we have to test the hypothesis that the mean of the population is μ.

For this, we first calculate $t = \frac{\overline{x} - \mu}{\sigma_s}\sqrt{n-1}$

Where, $\overline{x} = \frac{\sum_1^n x_i}{n}$, $\sigma_s{}^2 = \frac{1}{n-1} \sum_1^n (x_i - \overline{x})^2$.

Then find the value of P for the given d.f. from the table.

If the calculated $t > t_{0.05}$, the difference between $\overline{x}$ and μ is said to be significant at 5% level of significance.

If $t > t_{0.01}$, the difference between $\overline{x}$ and μ is said to be significant at 1% level of significance.

If $t < t_{0.05}$, the data is said to be consistent with the hypothesis.

Critical values of t, for two tail tests.

Degrees of freedom *(df)*	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	4.165	6.314	12.706	25.452	63.657	127.321	636.619	1273.239
2	1.886	2.282	2.920	4.303	6.205	9.925	14.089	31.599	44.705
3	1.638	1.924	2.353	3.182	4.177	5.841	7.453	12.924	16.326
4	1.533	1.778	2.132	2.776	3.495	4.604	5.598	8.610	10.306
5	1.476	1.699	2.015	2.571	3.163	4.032	4.773	6.869	7.976
6	1.440	1.650	1.943	2.447	2.969	3.707	4.317	5.959	6.788
7	1.415	1.617	1.895	2.365	2.841	3.499	4.029	5.408	6.082
8	1.397	1.592	1.860	2.306	2.752	3.355	3.833	5.041	5.617
9	1.383	1.574	1.833	2.262	2.685	3.250	3.690	4.781	5.291
10	1.372	1.559	1.812	2.228	2.634	3.169	3.581	4.587	5.049
11	1.363	1.548	1.796	2.201	2.593	3.106	3.497	4.437	4.863
12	1.356	1.538	1.782	2.179	2.560	3.055	3.428	4.318	4.716
13	1.350	1.530	1.771	2.160	2.533	3.012	3.372	4.221	4.597
14	1.345	1.523	1.761	2.145	2.510	2.977	3.326	4.140	4.499
15	1.341	1.517	1.753	2.131	2.490	2.947	3.286	4.073	4.417
16	1.337	1.512	1.746	2.120	2.473	2.921	3.252	4.015	4.346
17	1.333	1.508	1.740	2.110	2.458	2.898	3.222	3.965	4.286
18	1.330	1.504	1.734	2.101	2.445	2.878	3.197	3.922	4.233
19	1.328	1.500	1.729	2.093	2.433	2.861	3.174	3.883	4.187
20	1.325	1.497	1.725	2.086	2.423	2.845	3.153	3.850	4.146
21	1.323	1.494	1.721	2.080	2.414	2.831	3.135	3.819	4.110
22	1.321	1.492	1.717	2.074	2.405	2.819	3.119	3.792	4.077
23	1.319	1.489	1.714	2.069	2.398	2.807	3.104	3.768	4.047
24	1.318	1.487	1.711	2.064	2.391	2.797	3.091	3.745	4.021
25	1.316	1.485	1.708	2.060	2.385	2.787	3.078	3.725	3.996
26	1.315	1.483	1.706	2.056	2.379	2.779	3.067	3.707	3.974
27	1.314	1.482	1.703	2.052	2.373	2.771	3.057	3.690	3.954
28	1.313	1.480	1.701	2.048	2.368	2.763	3.047	3.674	3.935
29	1.311	1.479	1.699	2.045	2.364	2.756	3.038	3.659	3.918
30	1.310	1.477	1.697	2.042	2.360	2.750	3.030	3.646	3.902
40	1.303	1.468	1.684	2.021	2.329	2.704	2.971	3.551	3.788
50	1.299	1.462	1.676	2.009	2.311	2.678	2.937	3.496	3.723
60	1.296	1.458	1.671	2.000	2.299	2.660	2.915	3.460	3.681
70	1.294	1.456	1.667	1.994	2.291	2.648	2.899	3.435	3.651
80	1.292	1.453	1.664	1.990	2.284	2.639	2.887	3.416	3.629
100	1.290	1.451	1.660	1.984	2.276	2.626	2.871	3.390	3.598
1000	1.282	1.441	1.646	1.962	2.245	2.581	2.813	3.300	3.492
Infinite	1.282	1.440	1.645	1.960	2.241	2.576	2.807	3.291	3.481

Examples:

1. A certain stimulus administered to each of 12 patients resulted in the following increases of blood pressure: 5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4, 6. Can it be concluded that the stimulus will in general be accompanied by an increase in blood pressure.

Solution: Let us assume that the stimulus does not change the B.P.

Taking the population to be normal with mean 0 and S.D. σ .

$\overline{x} = \frac{\sum_1^n x_i}{n} = 2.5833,$

$\sigma_s{}^2 = \frac{1}{n-1} \sum_1^n (x_i - \overline{x})^2$

$= \frac{1}{11}[5.8404 + 0.3402 + 29.3406 + 12.84 + 0.1736 + 21.0066 + 2.5068 + 5.8404 + 2.0070 + 11.6738]$

$= 8.8554$

$\therefore \sigma_s = 2.9758$

Now $t = \frac{\overline{x}-\mu}{\sigma_s}\sqrt{n-1} = \frac{2.5833-0}{2.9758}\sqrt{11} = 2.8792.$

For ν =11, from the table $t_{0.05} = 1.8.$ for single-tailed test.

Since, $t > t_{0.05}$, the difference between $\overline{x}$ and μ is said to be significant at 5% level of significance. Therefore hypothesis is rejected, that is the stimulus will increase the B.P.

2. The nine items of a sample have the following values: 45, 47, 50, 52, 48, 47, 49, 53 and 51. Does the mean of these differ significantly from the assumed mean of 47.5?

Solution: $\overline{x} = \frac{\sum_1^n x_i}{n} = 49.1111,$

$\sigma_s{}^2 = \frac{1}{n-1} \sum_1^n (x_i - \overline{x})^2$

$= \frac{1}{8}[16.9011 + 4.4567 + 0.7901 + 8.3457 + 1.2345 + 4.4567 + 0.0123 + 15.1235 + 3.5679]$

$= 6.8611$

$\therefore \sigma_s = 2.6194,$ given that $\mu = 47.5$.

$$t = \frac{\overline{x}-\mu}{\sigma_x}\sqrt{n-1} = \frac{49.1111-47.5}{2.6196}\sqrt{8} = 1.7395.$$

For $\nu = 8,$ $t_{0.05} = 2.31$, since $t < t_{0.05}$, the value of t is not significant at 5% level of significance. Thus the test provides no evidence against the population mean being 47.5 .

3. Ten individuals are chosen at random from a population and their heights in inches are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71, 71 . Test the hypothesis that the mean height of the universe is 66 inches. (For d.f. 9, $t_{0.05} = 2.262$)

Solution: $\overline{x} = \frac{\sum_1^n x_i}{n} = 67.8,$

$\sigma_s{}^2 = \frac{1}{n-1} \sum_1^n (x_i - \overline{x})^2$

$= \frac{1}{9}[23.04 + 23.04 + 3.24 + 0.64 + 0.04 + 1.44 + 4.84 + 4.84 + 10.24 + 10.24]$

$= 9.0667$

$\therefore \sigma_s = 3.0111,$ given that $\mu = 66$.

$t = \frac{\overline{x}-\mu}{\sigma_s}\sqrt{n-1} = \frac{67.8-66}{3.0111}\sqrt{9} = 1.7934.$

For $\nu = 9,\ t_{0.05} = 2.262$, since $t < t_{0.05}$, the value of t is not significant at 5% level of significance. Thus the test provides no evidence against the population mean being 66 inches.

4. A machinist is making engine parts with axle diameter of 0.7 inch. A random sample of 10 parts shows mean diameter 0.742 inch with a S.D 0.04 inch. On the basis of this sample, would you say that the work is inferior?

Solution: Assuming that the work is not inferior i.e. there is no significant difference between $\overline{x}$ & μ .

Given that, $\overline{x} = 0.742,\quad \mu = 0.7,\quad \sigma_s = 0.04,\quad n = 10.$

$$t = \frac{\overline{x}-\mu}{\sigma_s}\sqrt{n-1} = \frac{0.742-0.7}{0.04}\sqrt{9} = 3.15.$$

For $\nu = 9,\ t_{0.05} = 2.262$, since $t > t_{0.05}$, the value of t is significant at 5% level of significance.

This implies that $\overline{x}$ differs significantly from μ and null hypothesis is rejected. Hence the work is inferior.

5. For a random sample of 16 values with mean 41inches, and the sum of the squares of the deviations from the mean is 135 $inches^2$. Estimate the 95% confident limits for the mean of the population.($t_{0.05} = 2.13$ for$\nu = 15$)

Solution: Given that, $\overline{x} = 41,\quad \sum_1^n (x_i - \overline{x})^2 = 135,\ n = 16\ .\ \therefore \nu = 15.$

$$\sigma_s{}^2 = \frac{1}{n-1}\sum_1^n (x_i - \overline{x})^2 = \frac{135}{15} = 9. \qquad \therefore \sigma_s = 3.$$

$$|t| < 2.13 \Longrightarrow |\overline{x} - \mu| < 2.13\frac{\sigma_s}{\sqrt{n-1}}$$

$$\Longrightarrow |\overline{x} - \mu| < 1.65,\ \text{approximately.}$$

$$\Longrightarrow 41 - 1.65 < \mu < 41 + 1.65$$

Or $39.35 < \mu < 42.65.$

Significance test of difference between sample means: $t = \dfrac{\overline{x}\sim\overline{y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

Where $\overline{x} = \dfrac{\sum_1^{n_1} x_i}{n_1},\quad \overline{y} = \dfrac{\sum_1^{n_2} y_i}{n_2}\ ,\quad \sigma_s{}^2 = \dfrac{1}{n_1+n_2-2}\left\{\sum_1^{n_1}(x_i - \overline{x})^2 + \sum_1^{n_2}(y_i - \overline{y})^2\right\}.$

For the different standard deviation, $\sigma_s{}^2 = \dfrac{1}{n_1+n_2-2}\left\{(n_1 - 1)\sigma_x{}^2 + (n_2 - 1)\sigma_y{}^2\right\}$

If the two samples are of same size, then $\sigma^2 = \dfrac{1}{n-1}\sum_1^n \left(d_i - \overline{d}\right)^2$ and $t = \dfrac{\overline{d}-0}{\sigma}\sqrt{n}$, where $d = x - y$.

1. Two horses A and B were tested according to the time (in seconds) to run a particular race gives the following results

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

Test whether you can discriminate between the two horses. (For, $\nu = 11,\ t_{0.05} = 2.2,\ t_{0.02} = 2.72$)

Solution: $\overline{x} = \dfrac{\sum_1^{n_1} x_i}{n_1} = 31.8,\quad \overline{y} = \dfrac{\sum_1^{n_2} y_i}{n_2} = 28.2$,

$$\sigma^2 = \frac{1}{n_1+n_2-2}\left\{\sum_1^{n_1}(x_i - \overline{x})^2 + \sum_1^{n_2}(y_i - \overline{y})^2\right\} = 5.9273. \qquad \sigma = 2.3016.$$

Assume that there is no discriminate between the two horses.

$$t = \frac{\bar{x} \sim \bar{y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 2.42. \qquad \Rightarrow t_{0.05} < t < t_{0.02}\,. \qquad (\nu = n_1 + n_2 - 2 = 11)$$

Therefore the discrimination between the horses is significant at 5% level but not at 2 level.

2. A group of boys and girls were given an intelligence test. The mean score, S.D.s and numbers in each group are as follows.

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

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

Solution: Let x be the boy's score, y be girl's core.

Let mean score of boys are not significantly different from that of girls

Given that , $\bar{x} = 124,\ \bar{y} = 121\,,\ \sigma_x = 12,\ \sigma_y = 10,\ n_1 = 18\ ,n_2 = 14\,.$

$$\sigma^2 = \frac{1}{n_1+n_2-2}\{(n_1-1){\sigma_x}^2 + (n_2-1){\sigma_y}^2\}$$

$$= \frac{1}{30}\{17 \times 144 + 13 \times 100\} = 124.9333.$$

$$\therefore \sigma = 11.1774.$$

$$t = \frac{\bar{x} \sim \bar{y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{124-121}{11.1774\times\sqrt{\frac{1}{18}+\frac{1}{14}}} = 0.7532 < 2.04 = t_{0.05} \quad . \qquad (\text{for } \nu = 30).$$

Null hypothesis is accepted. Mean score of boys are not significantly different from that of girls.

3. Eleven school boys were given a test in drawing. Further they were given a month's tuition and a second test of equal difficulty was held at the end of it. Do the marks give the evidence that students have benefitted by extra coaching? For d.f. ν =10, $t_{0.05} = 2.228$.

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

Solution:

x_2	x_1	$d = x_2 - x_1$	$d - \bar{d}$	$(d-\bar{d})^2$
24	23	1	0	0
19	20	-1	-2	4
22	19	3	2	4
18	21	-3	-4	16
20	18	2	1	1
22	20	2	1	1
20	18	2	1	1
20	17	3	2	4
23	23	0	-1	1
20	16	4	3	9
17	19	-2	-3	9
		$\sum d = 11$		$\sum(d-\bar{d})^2$ =50

$$\bar{d} = \frac{\sum d}{n} = 1$$

$${\sigma_s}^2 = \frac{1}{n-1}\left\{\sum_1^n (d_i - \bar{d})^2\right\} = 5\,.$$

$$\therefore \sigma_s = 2.2361\,.$$

Assuming that the students have not been benefitted by extra coaching.

Then the mean of the difference between the marks $\mu = 0$.

$$t = \frac{\bar{d}-\mu}{\sigma_s}\sqrt{n} = \frac{\sqrt{11}}{2.2361} = 1.4832 < t_{0.05}.$$

Hence difference is not significant.

There is no evidence that the students have benefitted by extra coaching.

CHI-SQUARE (χ^2) **TEST**: If O_i and E_i are observed and expected frequencies for $i = 1,2 \cdots n$.

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

The equation of χ^2 curve is $y = y_0 e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{\nu - 1}{2}}$, where $\nu = n - 1$.

Goodness of fit: The value of χ^2 is used to test whether the deviations of the observed frequencies from theoretical frequencies are significant or not.

Table: Chi-Square Probabilities

The areas given across the top are the areas to the right of the critical value. To look up an area on the left, subtract it from one, a left is 0.95 on the right)

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Examples:

1. In experiments on pea breeding, the following frequencies of seeds were obtained.

Round and yellow	Wrinkled and yellow	Round and green	Wrinkled and 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 and experiment.

Theoretical frequencies are $\frac{9}{16}\times 556,\ \frac{3}{16}\times 556,\ \frac{3}{16}\times 556,\ \frac{1}{16}\times 556$.
i.e. 313, 104, 104, 35.

$$\chi^2 = \sum\frac{(O_i-E_i)^2}{E_i} = \frac{4}{313}+\frac{9}{104}+\frac{16}{104}+\frac{9}{35} = 0.5103\ .$$

For d.f. $\nu = n-1 = 3$, $\chi^2_{0.05} = 7.815$. Since calculated value of χ^2 is much less than $\chi^2_{0.05}$, there is a very high degree of agreement between theory and experiment.

2. A set of five similar coins is tossed 320 times and the result is

No. of heads	0	1	2	3	4	5
f	6	27	72	112	71	32

Test the hypothesis that the data follow a binomial distribution.

Solution: In binomial distribution $P(x) = {}^nC_x p^x q^{n-x} = \frac{{}^5C_x}{32}$.

No. of heads	0	1	2	3	4	5
O_x	6	27	72	112	71	32
$E_x = \frac{{}^5C_x}{32}\times 320$	10	50	100	100	50	10

$$\chi^2 = \sum\frac{(O_i-E_i)^2}{E_i} = \frac{16}{10}+\frac{529}{50}+\frac{784}{100}+\frac{144}{100}+\frac{441}{50}+\frac{484}{10} = 78.68\ .$$

For d.f. $\nu = n-1 = 5$, $\chi^2_{0.05} = 11.07$. Since calculated value of χ^2 is much greater than $\chi^2_{0.05}$, the hypothesis that the data follow the binomial distribution is rejected.

3. The following table gives the number of aircraft accidents that occurred during the various days of the week. Find whether the accidents are uniformly distributed over the week.

Days	Sun	Mon	Tue	Wed	Thru	Fri	Sat	Total
No. of accidents	14	16	8	12	11	9	14	84

For d.f. $\nu = 6$, $\chi^2_{0.05} = 12.59$

Solution: If O_i and E_i are observed and expected frequencies for $i = 1,2 \cdots 7$. Clearly $E_i = 12\ for\ each\ i$

Then $\chi^2 = \sum\frac{(O_i-E_i)^2}{E_i} = \frac{4+16+16+0+1+9+4}{12} = 4.1667 < 12.59$.

Since calculated value of χ^2 is much less than $\chi^2_{0.05}$, the accidents are uniformly distributed over the week.

F-Distribution: Let $x_1,\ x_2,\ x_3 \cdots x_{n_1}$ and $y_1,\ y_2,\ y_3 \cdots y_{n_1}$ are two independent random samples of a normal populations with equal standard deviation σ . Let $\overline{x}$ and $\overline{y}$ are sample mean,

$$s_1{}^2 = \frac{1}{n_1 - 1}\left\{\sum_1^{n_1}(x_i - \overline{x})^2\right\} \text{ and } s_2{}^2 = \frac{1}{n_2 - 1}\left\{\sum_1^{n_2}(y_i - \overline{y})^2\right\} \text{ be the sample variance.}$$

Then define $F = \frac{s_1{}^2}{s_2{}^2}$ or $F = \frac{s_2{}^2}{s_1{}^2}$ depending on either $s_1{}^2 > s_2{}^2$ or $s_2{}^2 > s_1{}^2$ respectively.

This gives *F-distribution* (or variance ratio distribution). Clearly *F-distribution* depends only on ν_1 and ν_2.

$F_\alpha(\nu_1, \nu_2)$ is the value of F for ν_1 and ν_2 such that area to the right of F_α is α.

Snedecor's F-table for 5% and 1% significance levels are given below.

F-table of Critical Values of α = 0.05 for F(df1, df2)

	DF1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
DF2=1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

F-table of Critical Values of α = 0.01 for F(df1, df2)

	DF1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
DF2=1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.29	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6365.86
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.51	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.93	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.67	3.59	3.51	3.43	3.34	3.26	3.17
14	8.86	6.52	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.90	3.81	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.85	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.84	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.02	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.93	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.70	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.77	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.82	2.66	2.59	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.05	3.73	3.50	3.33	3.20	3.09	3.01	2.87	2.73	2.57	2.50	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.67	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.81
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.04	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.19	2.04	1.88	1.79	1.70	1.59	1.47	1.33	1.00

F-distribution is useful for testing the equality of population means by comparing the sample variances.

Examples:

1. Two samples of sizes 9 and 8 give the sum of squares of deviations from their respective means equal to 160 and 91 respectively. Can these be regarded as drawn from the same normal population?

Solution: Given that $\sum_1^9 (x_i - \overline{x})^2 = 160$ and $\sum_1^8 (y_i - \overline{y})^2 = 91$

$$s_1{}^2 = \frac{1}{n_1 - 1}\left\{\sum_1^{n_1} (x_i - \overline{x})^2\right\} = \frac{160}{8} = 20\,, \quad s_2{}^2 = \frac{1}{n_2 - 1}\left\{\sum_1^{n_2} (y_i - \overline{y})^2\right\} = \frac{91}{7} = 13.$$

$$F = \frac{s_1{}^2}{s_2{}^2} = \frac{20}{13} = 1.5385.$$

From the table, $F_{0.05}(8,7) = 3.73$. Since the calculated value of $F < F_{0.05}$, populations variances are not

Significantly different. Hence, the samples can be regarded as drawn from the same normal population.

2. Two independent samples of size 7 and 6 have the following values.

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

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

Given that $F_{0.05}(6,5) = 4.95$ and $F_{0.05}(5,6) = 4.39$.

Solution: Let the samples have been drawn from normal populations having the same variance.

$$\overline{x} = \frac{\sum_1^{n_1} x_i}{n_1} = \frac{28+30+32+33+33+29+34}{7} = \frac{219}{7} = 31.2857$$

$$s_1{}^2 = \frac{1}{n_1-1}\left\{\sum_1^{n_1}(x_i - \overline{x})^2\right\} = \frac{1}{6}[10.7959 + 1.6530 + 0.5102 + 2.9388 + 2.9388 + 5.2244 + 7.3674]$$

$$= 5.2381$$

$$\overline{y} = \frac{\sum_1^{n_2} y_i}{n_2} = \frac{29+30+30+24+27+29}{6} = \frac{169}{6} = 28.1667$$

$$s_2{}^2 = \frac{1}{n_2-1}\left\{\sum_1^{n_2}(y_i - \overline{y})^2\right\} = \frac{1}{5}[0.6944 + 3.3610 + 3.3610 + 17.3614 + 1.3612 + 0.6944]$$

$$= 5.3667$$

$$F = \frac{s_2{}^2}{s_1{}^2} = \frac{5.3667}{5.2381} = 1.0245.$$

Given that, $F_{0.05}(5,6) = 4.39$. Clearly the calculated value of $F < F_{0.05}$, the samples can be regarded as drawn from the same normal population.