Small Samples

TEST OF SIGNIFICANCE

Small Sample:- If the sample size $n<30$, then that sample is considered as small sample.

Degrees of freedom (d.f):- The number of independent variates which make up the statistic is known as the degrees of freedom. and it is denoted by γ.

Test of significance for small samples:-

The following are some important tests for small samples.

(i) student's t-test (ii) F-test (iii) χ^2-test.

Assumptions for student's t-test:-

The following assumptions are made in student's t-test.

(i) sample size, $n<30$.

(ii) The parent population from which sample is drawn is normal.

(iii) The population standard deviation is unknown.

(iv) The sample observations are independent i.e., sample is random.

Uses of t-test:- This test is used

(i) to test for a specified mean.

(ii) to test for the equality of two means.

(iii) To test the significance of difference between the means of paired data.

Students "t" test for single mean:-

Let a random sample of size n ($n<30$) has a sample mean $\bar{x}$. To test the hypothesis that the population mean μ has a specified value μ_0 when population S.D σ is not known.

Let the null hypothesis be $H_0: \mu = \mu_0$.
Then the alternative hypothesis is $H_1: \mu \neq \mu_0$.
To test the above H_0, the test statistic t is given by

$$t = \frac{|\bar{x}-\mu|}{s/\sqrt{n-1}} \sim t_{n-1} \text{ d.f.}$$

If $t_{cal} \leq t_{tab}$ value, then we accept H_0 otherwise we reject it.

P) A mechanist is 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 specification at 0.05 level of significance.

Sol:- Here the sample size $n = 10 < 30$.
Hence the sample is small sample.
Also sample mean $\bar{x} = 0.742$ inches, the population mean $\mu = 0.700$ inches and $s.D = 0.040$ inches are given.
$\therefore$ We use student's t-test.

(i) Null hypothesis : $H_0: \mu = 0.7$
(ii) Alternative hypothesis : $H_1: \mu \neq 0.7$.
(iii) Level of significance : $\alpha = 0.05$.
(iv) Test statistic $t = \dfrac{|\bar{x}-\mu|}{s/\sqrt{n-1}} = \left|\dfrac{0.742-0.700}{\frac{0.040}{\sqrt{10-1}}}\right| = 3.15$

$t_{cal} = 3.15$
The tabulated value of t at 5% level with 9 degrees of freedom is $t_{0.05} = 2.26$.
Since $t_{cal} > t_{tab}$, Hence we reject H_0.
Hence conclude that $\mu \neq 0.7$.

P) A sample of 26 bulbs gives a mean life of 990 hours with a S.D of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the sample not up to the standard

Scanned with CamScanner

$\underline{\underline{Sol^n}}$:- Here Sample size $n = 26 < 30$.
$\therefore$ The Sample is Small Sample.
Also given, Sample mean $\bar{x} = 970$. and S.D, $s = 20$.
Population mean, $\mu = 1000$
Degrees of freedom $= n-1 = 26-1 = 25$.

(i) Null hypothesis : $H_0 : \mu = 1000$.
(ii) Alternative hypothesis : $H_1 : \mu < 1000$ (one-tailed test).
(iii) Level of significance : $\alpha = 0.05$
(iv) The test statistic : $t = \dfrac{|\bar{x}-\mu|}{s/\sqrt{n-1}} = \dfrac{|970-1000|}{20/\sqrt{25}} = 2.5$

t_{tab} (5% L.O.S, 25 d.f) $= 1.708$

(v) Conclusion:- Since $t_{cal} > t_{tab}$ value.
Hence we reject H_0
Hence we Conclude that $\mu < 1000$.

P) 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 648 P.S.I. Use this information and the level of significance $\alpha = 0.05$ to test whether the true average compressive strength of the steel from which this sample came is 58,000 P.S.I. Assume normality.

$\underline{\underline{Sol^n}}$:- We have, $n = 6 < 30$, $\therefore$ The sample is small.
$\bar{x}$ = sample mean = 58392 PSI
s = standard deviation of six beams = 648 PSI
degrees of freedom $= n-1 = 6-1 = 5$.

(i) Null hypothesis : $H_0 : \mu = 58000$.
(ii) Alternative hypothesis : $H_1 : \mu \neq 58000$.
(iii) Level of Significance : $\alpha = 0.05$.
(iv) Test statistic : $t = \dfrac{|\bar{x}-\mu|}{s/\sqrt{n-1}} = \dfrac{58392-58000}{648/\sqrt{5}} = 1.353$.

t_{tab} value at 5 d.o.f at 5% L.O.S = ~~3.365~~ 2.571
(2.5%) L.O.S)

$t_{cal} < t_{tab}$ value
Hence we accept H_0
Hence we Conclude that $\mu = 58000$

Scanned with CamScanner

P) A random sample from a Company's very extensive files show that the orders for a certain kind of machinery were filled respectively in 10, 12, 19, 14, 15, 18, 11 and 13 days. Use the level of significance $\alpha = 0.01$ to test the claim that on the average such orders are filled in 10.5 days. choose the alternative hypothesis so that rejection of null hypothesis $\mu = 10.5$ days implies that it takes longer than indicated.

Sol^n:- We have

$n = 8$, $\bar{x} = \frac{1}{8}(10+12+19+14+15+18+11+13) = \frac{112}{8} = 14.$

and sample S.D, S is given by.

$$s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$$

$$= \frac{1}{7}\left[(10-14)^2 + (12-14)^2 + (19-14)^2 + \cdots + (13-14)^2\right]$$

$$= \frac{1}{7}(16+4+25+0+1+16+9+1)$$

$$= \frac{1}{7}(72) = 10.286$$

$$s = \sqrt{10.286} = 3.207.$$

(i) Null hypothesis $H_0: \mu = 10.5$ days.

(ii) Alternative hypothesis $H_1: \mu > 10.5$ days

(iii) Level of significance: $\alpha = 0.01$.

(iv) Test statistic $t = \frac{|\bar{x} - \mu|}{s/\sqrt{n}} = \frac{|14 - 10.5|}{3.207/\sqrt{7}} = 3.087.$

t_{tab} value for $(8-1)$ df at $\cancel{0.05}$ 0.01 L.OS = 2.998

(v) Conclusion:- Since $t_{cal} > t_{tab}$ value

Hence we reject H_0.

Hence We conclude that $\mu > 10.5$ days.

P) 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 1000 hrs	1.2	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 bulbs is 4000 hours?

Here $n = 10$

Null hypothesis $H_0 : \mu = 4000$

Alternative hypothesis $H_1 : \mu \neq 4000$

Level of significance : $\alpha = 0.05$

For the given data, $\bar{x}$ = average life time of bulbs

$$= \frac{1.2 + 4.6 + 3.9 + 4.1 + 5.2 + 7.8 + 3.9 + 4.3 + 4.4 + 5.6}{10}$$

$$= 4.1$$

$$s^2 = \sum_{i=1}^{10} \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$$

$$= \frac{1}{9} \Big[(1.2-4.1)^2 + (4.6-4.1)^2 + (3.9-4.1)^2 + (4.1-4.1)^2 + (4.4-4.1)^2$$

$$+ (5.2-4.1)^2 + (3.8-4.1)^2 + (3.9-4.1)^2 + (4.3-4.1)^2 + (4.4-4.1)^2$$

$$+ (5.6-4.1)^2 \Big]$$

$$= \frac{1}{9} [8.41 + 0.25 + 0.04 + 1.21 + 0.09 + 0.04 + 0.03 + 0.09 + 2.25]$$

$$= \frac{12.42}{9} = 1.38$$

$$\text{Hence } t = \frac{|\bar{x} - \mu|}{s/\sqrt{n}} = \frac{4.1 - 1.38 \ \cancel{4400 - 4000}}{\sqrt{1.38}/\sqrt{10}} = 8.512$$

No of degrees of freedom $= 9$

$\therefore t_{\alpha/2} = 2.262$

Since $t_{cal} > t_{tab}$, Hence H_0 is rejected.

P) Producer of "gutkha", claims that the nicotine content in his "gutkha" on the average of is 1.83 mg. Can this claim accepted if a random sample of 8 "gutkha" of this type have the nicotine contents of 2.0, 1.7, 2.1, 1.9, 2.2, 2.1, 2.0, 1.6 mg? Use a 0.05 level of significance.

Solⁿ:- Given, $n = 8$ and $\mu = 1.83$ mg.

1. Null hypothesis $H_0 : \mu = 1.83$
2. Alternative hypothesis : $H_1 : \mu \neq 1.83$
3. Level of significance $\alpha = 0.05$
4. The test statistic is $t = \frac{|\bar{x} - \mu|}{s/\sqrt{n}} \sim t_{n-1}$.

where $\bar{x}$ and s^2 are to be computed from the sample values of Sutkha

x	$x - \bar{x}$	$(x - \bar{x})^2$
2.0	0.05	0.0025
1.7	−0.85	0.0625
2.1	0.15	0.0225
1.9	−0.05	0.0025
2.2	0.25	0.0625
2.1	0.15	0.0225
2.0	0.05	0.0025
1.6	−0.35	0.1225
Total 15.6		0.3

$\bar{x} = \frac{15.6}{8} = 1.95$ and $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{0.3}{7} = \cancel{0.21}$ 0.045

$s = \sqrt{0.0439} = 0.207$

$\therefore t = \frac{1.95 - 1.83}{0.21/\sqrt{8}} = 1.62$

tabulated $t_{0.05}$ for $(8-1)$ i.e., 7 df is $\cancel{1.895}$ 2.365

$\alpha/2 = 0.025$

since $t_{cal} < t_{tab}$, Hence we accepted H_0.

Scanned with CamScanner

P) Eight students were given a test in STATISTICS and after one month coaching they were given another test of the similar nature. The following table gives the increase of their marks in the second test over the first

Student no	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 coaching?

Sol:- We compute the mean and S.D of the increase of marks as follows

$$\bar{x} = \frac{\Sigma x}{n} = \frac{4-2+6-8+12+5-7+2}{8} = \frac{12}{8} = 1.5$$

We have $s^2 = \frac{1}{n-1} \Sigma(x_i - \bar{x})^2$

$$= \frac{1}{7}\left[(4-1.5)^2 + (-2-1.5)^2 + (6-1.5)^2 + (-8-1.5)^2 + (12-1.5)^2 + (5-1.5)^2 + (-7-1.5)^2 + (2-1.5)^2\right]$$

$$= \frac{1}{7}(6.25 + 12.25 + 20.25 + 90.25 + 110.25 + 12.25 + 72.25 + 0.25)$$

$$= \frac{324}{7} = 46.2857$$

$\therefore$ S.D $= s = \sqrt{46.2857} = 6.8$

Assuming that the students have not been benefited by coaching, it implies that the mean of the difference of the two tests is zero i.e., $\mu = 0$.

Then $t = \frac{|\bar{x} - \mu|}{s/\sqrt{n-1}} = \frac{|1.5 - 0|}{(6.8)/\sqrt{7}} = 0.5836$

No of degrees of freedom $= 8 - 1 = 7$.

$\therefore$ Tabulated $t_{0.05} = 2.36$

$\therefore$ $t_{cal} < t_{tab}$, Hence we accept H_0 at 5% 2.05

i.e., the test provides no evidence that the students have been benefitted by coaching.

Student's t-test for difference of means:-

Let $\bar{x}$ and $\bar{y}$ be the means of two independent samples of sizes n_1 and n_2 $(n_1<30, n_2<30)$ drawn from two normal populations having means μ_1 and μ_2. To test whether the two population means are equal, we write the H_0 as

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

To test the above H_0, we write the test statistic as

$$t = \frac{|\bar{x}-\bar{y}|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{(n_1+n_2-2)} \text{ d.f}$$

where $S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2+\Sigma(y_i-\bar{y})^2\right]$ where S_1 and S_2 are not given

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1+n_2-2}$$, where S_1 and S_2 are given

If $t_{cal} < t_{(n_1+n_2-2)}$ d.f, then we accept H_0 at given L.O.S otherwise we reject it.

P) Two horses A and B are 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 the same running capacity?

Sol:-

Given, $n_1 = 7$, $n_2 = 6$

We first compute the sample means and standard deviations.

$\bar{x}$ = Mean of the first sample $= \frac{1}{7}(28+30+32+33+29+34) =$

$\frac{1}{7}(219) = 31.86$

$\bar{y}$ = Mean of the second sample $= \frac{1}{6}(29+30+30+24+27+29)$

$= \frac{1}{6}(169) = 28.16$

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	−1.26	1.8538	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

Now $s^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2\right]$

$= \frac{1}{11}\left[31.4358 + 26.8336\right] = \frac{1}{11}(58.26294) = 5.23$

$\therefore s = \sqrt{5.23} = 2.3$

Null hypothesis : $H_0: \mu_1 = \mu_2$

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2$

Level of Significance: $\alpha = 0.05$

Test statistic $t = \frac{|\bar{x}-\bar{y}|}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{31.286 - 28.16}{(2.3)\sqrt{\frac{1}{7}+\frac{1}{6}}} = 2.443.$

Tabulated value of t for $7+6-2 = 11$ at 5% LOS $= 2.2$

Since $t_{cal} > t_{tab}$, we reject H_0 and we conclude that $\mu_1 \neq \mu_2$ (or) both horses A and B do not have the same running capacity.

$\therefore$ Now $s^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2\right]$

$= \frac{1}{18}\left[1606 + 10~~7~~9.6\right] = 182.53$

$s = 13.51$

Null hypothesis $H_0: \mu_1 = \mu_2$.

Alternative hypothesis $H_1: \mu_1 > \mu_2$.

Level of significance : $\alpha = 0.05$

Test statistic $t = \frac{|\bar{x}-\bar{y}|}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{103-95.8}{(13.51)\sqrt{\frac{1}{10}+\frac{1}{10}}} = 1.19168$

Since $t_{cal} = 1.19168 < t_{tab} = 1.734$.

Hence we accept H_0

Hence we Conclude that $\mu_1 = \mu_2$ (there is no difference in I Q S)

P) H.W.

To 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 L.O.S to test whether the difference between two sample means is significant?

$s = 19.999$, $t = 0.1446$

~~tcal~~ $t_{tab} = t_{6+6-2} = 10$ d.f at 1% L.O.S $= 3.169$

we accept H_0

Scanned with CamScanner

7)   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:

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 L.O.S 0.05)

Sol:- We have

$n_1 = 10, \; n_2 = 10$ and

$\bar{x} = \frac{1}{10}(117 + 105 + \cdots + 103 + 107) = 103.$

$\bar{y} = \frac{1}{10}(106 + 98 + \cdots + 108 + 85) = 95.8$

Now we compute the standard deviation of both sample

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	77.44
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		1606	958		1679.6

8. P) The I.Q's of 16 students from one area of a city showed a mean of 107 with a standard deviation 10, while the IQ's of 14 students from another area of the city showed a mean of 112 with a s.d of 8. Is there a significant difference between the IQ's of the two groups at a 0.05 l.o.s?

Sol:- Let μ_1 and μ_2 be the means of the two populations

Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative hypothesis $H_1: \mu_1 \neq \mu_2$.

we have $n_1 = 16$, $\bar{x} = 107$, $S_1 = 10$ and $n_2 = 14$, $\bar{y} = 112$,

$S_2 = 8$.

$$\therefore S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{(16 \times 100) + (14 \times 64)}{16 + 14 - 2} = \frac{2496}{28} = 89.14$$

$$\therefore S = \sqrt{89.14} = 9.44.$$

$$\therefore t = \frac{|\bar{x} - \bar{y}|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|107 - 112|}{9.44\sqrt{\frac{1}{16} + \frac{1}{14}}} = 1.447$$

$d.f = 16 + 14 - 2 = 28$

t_{tab} value at 28 d.f = $\cancel{1.701}$ 2.05 at 5% LOS

$t_{cal} < t_{tab}$

Hence we accept H_0.

Hence we conclude that $\mu_1 = \mu_2$ (i.e. (there is a significant difference between the IQ's of the two groups)

VIP

Paired Sample t-test :- If $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ be the pairs of data.

Let $d_i = x_i - y_i$, $i = 1, 2, \ldots n$

Then $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

To test the above H_0, we write the t-test statistic

as $t = \frac{\bar{d}}{s/\sqrt{n}} \sim t_{(n-1)}$, where $\bar{d} = \frac{1}{n}\Sigma d_i$

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$.

If $t_{cal} \leq t_{tab}$, then we accept H_0 otherwise we reject it.

Problems

P) Ten workers were given a training programme with a view to study their assembly time for a certain mechanism. The results of the time and motion study before and after the training programme are given below.

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

x_i = Time taken for assembling before training

y_i = Time taken for assembling after training

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

Sol: From the given paired data, we see that we are given paired t-test.

Let μ be the mean of population of differences.

1. Null hypothesis : $H_0: \mu_1 = \mu_2$
2. Alternative hypothesis : $H_1: \mu_1 \neq \mu_2$
3. Level of Significance : $\alpha = 0.05$
4. Differences d_i's (before and after training) are

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

$\bar{d}$ = Mean of the sample data $= \frac{\sum d}{n} = \frac{14}{10} = 1.4$.

$$s^2 = \frac{1}{n-1} \sum (d_i - \bar{d})^2$$

$$= \frac{1}{9}\left[(1-1.4)^2 + (2-1.4)^2 + (-1-1.4)^2 + (7-1.4)^2 + (1-1.4)^2 + (-4-1.4)^2 + (4-1.4)^2 + (3-1.4)^2 + (-1-1.4)^2 + (2-1.4)^2\right]$$

$$= \frac{1}{9}\left[0.16 + 0.36 + 5.76 + 31.36 + 0.16 + 29.16 + 6.76 + 2.56 + 5.76 + 0.36\right]$$

$$= \frac{82.4}{9} = 9.1555$$

$s = 3.026$.

4. The test statistic : $t = \frac{|\bar{d}|}{s/\sqrt{n}} = \frac{1.4}{3.026/\sqrt{10}} = 1.46$.

$t_{cal} = 1.46$

t_{tab} value at $10-1 = 9$ degrees of freedom is 1.833.

$t_{cal} < t_{tab}$

Hence we accept H_0

Hence we Conclude that $\mu_1 = \mu_2$.

H.W

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

Before	67	24	57	55	63	54	56	68	33	43
After	70	38	58	58	56	67	68	75	42	38

Test whether the intensive training is useful at 0.05 L.O.S?

3) 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

Test whether there is significant change in Blood pressure at 1% Level of Significance?

Sol^n:- Difference d_i's (before and after intake of drugs) are

10, −2, 2, 4, −4.

$$\therefore \bar{d} = \frac{\Sigma d_i}{n} = \frac{10}{5} = 2.$$

$$s^2 = \frac{1}{n-1}\Sigma(d_i-\bar{d})^2$$

$$= \frac{1}{4}\left[(10-2)^2+(-2-2)^2+(2-2)^2+(4-2)^2+(-4-2)^2\right]$$

$$= 30$$

$$\therefore s = \sqrt{30}$$

(1) Null hypothesis : $H_0: \mu_1 = \mu_2$

(2) Alternative hypothesis : $H_1: \mu_1 < \mu_2$

(3) Level of Significance : $\alpha = 0.001$

(4) Test statistic : $t = \dfrac{\bar{d}}{S/\sqrt{n}}$

$$= \frac{2}{\sqrt{30}/\sqrt{5}} = 0.82.$$

t_{tab} value at 0.01 L.O.S with $n-1=5-1=4$ d.f $= 4.6$

$t_{cal} < t_{tab}$

(5) Conclusion :- $t_{cal} < t_{tab}$

Hence we accept Ho.

Hence we Conclude that $\mu_1 = \mu_2$.

Hence we Conclude that there is no significant change in Blood Pressure after intake of a certain drug.

Snedecor's F-test of significance :-

We use F-test, to test for the equality of two population variances.

Test for Equality of Two Population Variances :-

Let two independent random samples of sizes n_1 and n_2 are drawn from two normal populations.

Here we write the Ho as.

$H_0 : \sigma_1^2 = \sigma_2^2$.

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

Then, the estimates of σ_1^2 and σ_2^2 are given by.

$$S_1^2 = \frac{1}{n_1-1}\Sigma(x_i-\bar{x})^2 \text{ and } S_2^2 = \frac{1}{n_2-1}\Sigma(y_i-\bar{y})^2.$$

To test the above H_0, we use the F-test statistic as

$$F = \frac{S_1^2}{S_2^2} \text{ if } S_1^2 > S_2^2 \quad (\text{or}) \sim F_{(n_1-1, n_2-1)} \text{ d.f.}$$

$$F = \frac{S_2^2}{S_1^2} \text{ if } S_2^2 > S_1^2. \sim F_{(n_2-1, n_1-1)}$$

If $F_{cal} \leq F_{tab}$ value, then we accept H_0 otherwise we reject it.

Problems

P) 1) 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 10% significant level, test whether the two population have the same variance.

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:- Given $n_1 = 5$, $n_2 = 5$

Now, $\bar{x} = \frac{\Sigma x_i}{n} = \frac{1}{5}(14.1 + 10.1 + 14.7 + 13.7 + 14.0)$

$= \frac{66.6}{5} = 13.32$

and $\bar{y} = \frac{\Sigma y}{5} = \frac{1}{5}(14.0 + 14.5 + 13.7 + 12.7 + 14.1)$

$= \frac{69}{5} = 13.8.$

Computing standard deviation of the samples.

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
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
66.6		13.4188	69		1.84

If s_1^2 and s_2^2 be the estimate of σ_1^2 and σ_2^2, then

$$s_1^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2 = \frac{13.4888}{4} = 3.372$$

$$s_2^2 = \frac{1}{n_2-1}\Sigma(y_i-\bar{y})^2 = \frac{1.84}{4} = 0.46$$

(1) Null ~~Here S.~~ hypothesis $H_0: \sigma_1^2 = \sigma_2^2$

(2) Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$.

(3) Level of significance: $\alpha = \cancel{0.02}\ 0.10$

(4) Test statistic: $F = \frac{s_1^2}{s_2^2}$ $\left(\because s_1^2 > s_2^2\right) \sim F_{(n_1-1, n_2-1)}$ df

$$F = \frac{3.372\ \cancel{15750}}{\cancel{10920}\ 0.46} = \cancel{1.44}\ 7.33.$$

$$F_{cal} = \cancel{1.44}\ 7.33.$$

(5) Conclusion:- F_{tab} value $= F_{(n_1-1, n_2-1)}$ df $= F_{(5-1,5-1)} = F_{(4,4)} = 6.39.$

Here $F_{cal} > F_{tab}$.

Hence we reject H_0

Hence we conclude that $\sigma_1^2 \neq \sigma_2^2$.

The following random samples are measurements of the heat-producing capacity (in millions of calories per ton) of specimens of coal from two mines.

Mine 1	8260	8130	8350	8070	8340	—
Mine 2	7950	7890	7900	8140	7920	7840

Use the 0.02 level of significance to test whether it is reasonable to assume that the variances of the two population samples are equal?

<u>Solⁿ</u>:- Here $n_1 = 5$, $n_2 = 6$

Mine 1			Mine 2		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
8260	30	900	7950	10	100
8130	−100	10000	7890	−50	2500
8350	120	14400	7900	−40	1600
8070	−160	25600	8140	200	40000
8340	110	12100	7920	−20	400
			7840	−100	10000
41150		63000	47640		54600

$$\therefore \bar{x} = \frac{\Sigma x_i}{n} = \frac{41150}{5} = 8230, \quad \bar{y} = \frac{\Sigma y}{n_2} = \frac{47640}{6} = 7940$$

$\Sigma(x-\bar{x})^2 = 63000$ and $\Sigma(y-\bar{y})^2 = 54600$.

$$\therefore S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1} = \frac{63000}{4} = 15750$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1} = \frac{54600}{5} = 10920$$

(1) Null hypothesis : $H_0: S_1^2 = S_2^2$.

(2) Alternative hypothesis : $H_1: S_1^2 \neq S_2^2$.

(3) Level of significance : $\alpha = 0.02$

(4) Test statistic : $F = \frac{S_1^2}{S_2^2}$ $(\because S_1^2 > S_2^2)$

i.e., $F = \frac{15750}{10920} = 1.44$.

$F_{cal} = 1.44$

(5) Conclusion:- $F_{(4,5)df}$ at 5% L.O.S = 5.19

$F_{cal} = 1.44$.

$\therefore F_{cal} < F_{tab}$ value

Hence we accept H_0.

Hence we conclude that $S_1^2 = S_2^2$

CHI - SQUARE (χ^2) TEST :-

If O_i ($i=1,2,\dots n$) is a set of observed (experimental) frequencies and E_i ($i=1,2,\dots n$) is the corresponding set of expected frequencies, then χ^2 is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \text{ with } (n-1) \text{ d.f}$$

Uses (or) Applications of χ^2-test :-

The χ^2-test is used to test

(i) the goodness of fit,

(ii) the independence of attributes

(iii) if the population has a specified value of the variance σ^2.

(ii) Conditions of χ^2-test :-

Following are the conditions which should be satisfied before χ^2 test can be applied.

(i) The sample observations should be independent.

(ii) N, the total frequency is large, i.e, > 50

(iii) The constraints on the cell frequencies, if any, are linear

(iv) No ~~expe~~ theoretical (or expected) frequency should be less than 10. If small theoretical frequencies occur, the difficulty is overcome by grouping 2 & more classes together before calculating $(O-E)$

Problems

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 frequencies of accident each week $= \frac{100}{10} = 10$.

$H_0: O_i = E_i$

$H_1: O_i \neq E_i$

Observed frequency	Expected frequency	$(O_i - E_i)$	$\frac{(O_i-E_i)^2}{E_i}$
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.	26.6
100	100		

Now $\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 26.6$ i.e., $\chi^2_{cal} = 26.6$

χ^2_9 at 0.05 LOS = 16.9

$\chi^2_{cal} > \chi^2_{tab}$

Hence we reject H_0.

χ^2 test for goodness of fit:-

Let $O_1, O_2, \dots O_n$ be a set of observed frequencies and $E_1, E_2, \dots E_n$ be a set of expected frequencies. Then χ^2 test statistic is given by

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

Here we write H_0 as $H_0: O_i = E_i$

$\chi^2_{cal} \leq \chi^2_{n-1}$ then we accept H_0 otherwise we reject it.

2) A die is thrown 264 times with the following results. Show that the die is biase (Given $\chi^2_{0.05} = 11.07$ for 5 d.f.)

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

Sol:-

$H_0: O_i = E_i$ (or) The die is unbiased

$H_1: O_i \neq E_i$ (or) The die is biased

The expected frequency of each of the numbers 1, 2, 3, 4, 5, 6 is $\frac{264}{6} = 44$.

Scanned with CamScanner

Observed frequency (O_i)	Expected frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
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

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 17.6362$$

χ^2_{n-1} at 5% LOS = 11.07

$\chi^2_{cal} > \chi^2_{tab}$, we reject Ho

$\therefore$ The die is biased. $(O_i \neq E_i)$

Chi-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES :-

Def :- Literally an attribute means a quality or characteristic, examples of attributes are drinking, smoking, blindness, honesty, beauty etc.

In this test, we write Ho as

Ho : $O_i = E_i$ (or) The given attributes are independent

In this test, the expected frequency (E_i) for any cell is

given by $= \dfrac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(r-1)(s-1)} \text{ d.f}$$

If $\chi^2_{cal} \leq \chi^2_{tab}$ value, then we accept H_0 otherwise we reject it.

Problems

P) The following table gives the classification of 100 workers according to sex and nature of work. Test whether the nature of work is independent of the sex of the workers.

(May 2013)

Worker	Stable	Unstable	total
Males	40	20	60
Females	10	30	40
Total	50	50	100.

Soln:- H_0: $O_i = E_i$ (i) The nature of work is independent of the sex of the workers

Expected frequency are given

$\frac{50\times60}{100}=30$	$\frac{50\times60}{100}=30$	60
$\frac{50\times40}{100}=20$	$\frac{50\times40}{100}=20$	40
50	50	100

Calculation of χ^2

O_i	E_i	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
40	30	100	3.333
20	30	100	3.333
10	20	100	5.000
30	20	100	5.000
100	100		16.66

$$\chi^2 = \Sigma \frac{(O_i-E_i)^2}{E_i} = 16.66$$

$$\chi^2_{cal} = 16.66$$

$$\chi^2_{(2-1)(2-1)} \text{ at } 5\% \text{ LOS} = 3.84$$

$$\chi^2_{cal} > \chi^2_{tab}$$

Hence we reject Ho

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	Employees Clerks	Teachers	Officers
Pepsi	10	25	65
Thumsup	15	30	65
Fanta	50	60	30

Sol^n:- $H_0: O_i = E_i$ (i) There is no significant liking

$H_1: O_i \neq E_i$

LOS : $\alpha = 0.05$

Soft drinks	clerks	Teachers	officers	total
Pepsi	10	25	65	100
Thumsup	15	30	65	110
Fanta	50	60	30	140
total	75	115	160	350

~~Calculation for~~ X^2

Observed frequency ~~E_i~~ Table of Expected frequencies

$\frac{75\times100}{350} = 21.4$ $\frac{115\times100}{350} = 32.9$ $\frac{160\times100}{350} = 45.71$

$\frac{75\times110}{350} = 23.6$ $\frac{115\times110}{350} = 36.1$ $\frac{160\times110}{350} = 50.3$

$\frac{75\times140}{350} = 30$ $\frac{115\times140}{350} = 46$ $\frac{160\times140}{350} = 64$

Calculation of X^2

observed frequency (O_i)	Expected frequency (E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
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	13.333
60	46	196	4.261
30	64	1156	18.062
			60.2425

$X^2_{cal} = 602425$

$X^2_{tab} = X^2_{(3-1)(3-1)}$ at 5% L.O.S = 9.488

$X^2_{cal} > X^2_{tab}$ Hence we reject H_0

χ^2-Square test for Population Variance :-

Suppose that a random sample $x_i\ (i = 1, 2, \ldots, n)$ is drawn from a normal population with mean μ and variance σ^2. To test the hypothesis that the population variance σ^2 has a specified value σ_0^2.

Set the null hypothesis $H_0 : \sigma^2 = \sigma_0^2$

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

$$\chi^2 = \frac{ns^2}{\sigma_0^2}.$$

where s^2 = sample variance = $\frac{\Sigma(x_i - \bar{x})^2}{n}$

If $\chi^2_{cal} \leq \chi^2_{tab}$, then we accept H_0 otherwise we reject it

P) A firm manufacturing rivets wants to limit variations in their length as much as possible. The lengths (in cm) of 10 rivets manufactured by a 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 0.145 cm

Sol^n :- we have

$n = 10$, $\bar{x} = \frac{\Sigma x}{n} = \frac{20.69}{10} = 2.068$

$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{0.09096}{10} = 0.0091$

and $\sigma_0 = 0.145$

1. Null hypothesis $H_0: \sigma^2 = \sigma_0^2$

2. Alternative hypothesis $H_1: \sigma^2 > \sigma_0^2$

3. Level of Significance : $\alpha = 0.05$

4. Test statistic $\chi^2 = \frac{nS^2}{\sigma_0^2} = \frac{0.09096}{(0.1455)^2} = 4.326$

$$\chi^2_{cal} = 4.8$$

$$\chi^2_{n-1} = \chi^2_{10-1} = \chi^2_9 \text{ at } 5\% \text{ LOS} = 16.919$$

$$\chi^2_{cal} < \chi^2_{tab}.$$

Hence we accept H_0