Sol:- given,

$$\mu_0 = 72; \quad \sigma_0^2 = 5.76 \; ; \; n = 100$$

$$\bar{x} = 70, \quad \sigma = 8$$

i) $$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$= \frac{(100)(70)(5.76) + 72(8)^2}{100(5.76) + (8)^2}$$

$$= \frac{44928}{640} = \frac{351}{5} = 70.2$$

$$\sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}} = \sqrt{\frac{(5.76)(64)}{100(5.76) + 64}}$$

$$= 0.758$$

ii) $$\left[\mu_1 - z_{\alpha/2}\sigma_1 \; , \; \mu_1 + z_{\alpha/2}\sigma_1\right]$$

$$= \left[70 - (1.96)(0.758), \; 70 + (1.96)(0.758)\right]$$

$$= (70 - 1.48568 \; , 70 + 1.48568)$$

$$= (68.51432, \; 71.48568)$$

Maximum error of estimate:

→ The max. error is the statement of error which gives the difference between the estimate and the quantity to be estimated.

i.e $E_{max} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Where, α, E, σ are known the sample size

Sol:- given,

$$\mu_0 = 72; \quad \sigma_0^2 = 5.76 \; ; \; n = 100$$

$$\bar{x} = 70, \quad \sigma = 8$$

i) $$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$= \frac{(100)(70)(5.76) + 72(8)^2}{100(5.76) + (8)^2}$$

$$= \frac{44928}{640} = \frac{351}{5} = 70.2$$

$$\sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}} = \sqrt{\frac{(5.76)(64)}{100(5.76) + 64}}$$

$$= 0.758$$

ii) $$\left[\mu_1 - z_{\alpha/2}\sigma_1 \; , \; \mu_1 + z_{\alpha/2}\sigma_1\right]$$

$$= \left[70 - (1.96)(0.758), \; 70 + (1.96)(0.758)\right]$$

$$= (70 - 1.48568 \; , 70 + 1.48568)$$

$$= (68.51432, \; 71.48568)$$

Maximum error of estimate:

→ The max. error is the statement of error which gives the difference between the estimate and the quantity to be estimated.

i.e $$E_{max} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

When, α, E, σ are known the sample size

n is given by, $n = \left[\frac{z_{\alpha/2}\,\sigma}{E}\right]^2$

confidence interval :

→ A confidence interval displays the probabili-ty that a parameter will fall between a pair of values around the mean.

→ Confidence interval is used to measure uncert-ainity in a sample variable.

→ The confidence level represents the frequency of acceptable confidence intervals that contain the true value of unknown parameter.

→ The confidence interval for mean when σ is known is given by, $\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

→ The confidence interval for mean when σ is unknown is given by,

$$\left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

Confidence limits :

→ for 95%, confidence limit $\boxed{1.96}$

→ for 99%, confidence limit is $\boxed{2.58}$

→ for 90% confidence limit is $\boxed{1.645}$

→ for 98% confidence limit is $\boxed{2.33}$

(SAQ)

Q: What is the max. error one can expect to make with probability 0.90 when using the mean of a random sample of size $n = 64$

to estimate the mean of the population with $\sigma^2 = 2.56$

Sol:- $E_{max} = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$\sigma^2 = 2.56 \Rightarrow \sigma = 1.6$; $n = 64 \Rightarrow \sqrt{n} = 8$.

$Z_{\alpha/2} = 1.645$ at 90% [$\because$ probability $= 0.90$]

$E_{max} = \frac{(1.645)(1.6)}{8} = 0.329$

Q:- It is desired to estimate the mean no. of hours of continuous use until a certain computer will first require repair. If it can be assumed that $\sigma = 48$ hrs, how large a sample is needed so that, one will be able to assert with 90% confidence that the sample mean is of by atmost 10hrs.

Sol: $n = \left[\frac{Z_{\alpha/2}\,\sigma}{E}\right]^2 =$ size.

given, $\sigma = 48$hrs $E = 10$hrs

At 90% $\Rightarrow Z_{\alpha/2} = 1.645$

$n = \left[\frac{(1.645)(48)}{10}\right]^2$

$n = 62.34 \simeq 62$

Q:- The mean and S.D of a population are 11.795 and 14.054. What can one assert with 95% confidence about the max. error

if $\bar{x} = 11.795$ and $n = 50$. Construct 95% confidence interval for the mean.

sol:- confidence interval $= \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

At 95% $\Rightarrow z_{\alpha/2} = 1.96$.

$$= \left[11.795 - \frac{(1.96)(14.054)}{\sqrt{50}}, 11.795 + \frac{(1.96)(14.054)}{\sqrt{50}}\right]$$

$$= [7.89, 15.69]$$

Q:- find 95% confidence limit for a mean of a normally distribution population from which the following samples was taken.

15, 17, 10, 18, 16, 9, 7, 11, 13, 14

sol:- mean $(\bar{x}) = \frac{15+17+10+18+16+9+7+11+13+14}{10}$

$$= \frac{130}{10} = 13.$$

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

$$s^2 = \frac{(15-13)^2 + (17-13)^2 + (10-13)^2 + \cdots + (14-13)^2}{9}$$

$$s^2 = \frac{4+16+9+25+9+16+36+4+0+1}{9}$$

$$= \frac{120}{9} = 13.333 \Rightarrow s = \sqrt{13.333} = 3.651$$

At 95% $\Rightarrow z_{\alpha/2} = 1.96$; $\left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

$$= \left[13 - \frac{(1.96)(3.651)}{\sqrt{10}}, 13 + \frac{(1.96)(3.651)}{\sqrt{10}}\right] = [10.738, 15.262]$$

1Q:- A random sample of size 100 has a S.D 5. What can you say about the max. error with 95% confidence

2Q: assuming that $\sigma = 20$, how large a random sample is taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3 points.

3Q:- A random sample of size 81 was taken whose variance is 20.25 and mean is 32. Construct 98% confidence interval.

1sol:- $n = 100$, $\sigma = 5$

$$E_{max} = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

At 95% $\Rightarrow z_{\alpha/2} = 1.96$

$$E_{max} = 1.96 \times \frac{5}{\sqrt{100}} = 1.96 \times \frac{\cancel{5}}{\cancel{10}_2} = \frac{1.96}{2}$$

$$\boxed{E_{max} = 0.98}$$

2sol:- $\sigma = 20$, $E = 3$ points

At 95% $\Rightarrow z_{\alpha/2} = 1.96$

$$\text{size } (n) = \left[\frac{z_{\alpha/2}\, \sigma}{E} \right]^2$$

$$n = \left(\frac{1.96 \times 20}{3} \right)^2$$

$$n = 170.73 \simeq 171.$$

Sol:- $n = 81$, $\sigma^2 = 20.25$, $\bar{x} = 32$

$\sigma = \sqrt{20.25} = 4.5$

At 98% $\Rightarrow z_{\alpha/2} = 2.33$

confidence interval $= \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

$= \left[32 - \frac{(2.33)(4.5)}{\sqrt{81}}, 32 + \frac{(2.33)(4.5)}{\sqrt{81}}\right]$

$= \left[32 - \frac{10.485}{9}, 32 + \frac{10.485}{9}\right]$

$= [32 - 1.165, 32 + 1.165]$

$= [30.835, 33.165]$

Test of Hypothesis:

→ A statistical hypothesis is a statement about the parameter of one (or) more population....

→ Testing of hypothesis is a process for deciding whether to accept (or) reject the hypothesis.

→ There are 2 types of hypothesis.

1. Null hypothesis
2. Alternative hypothesis.

Null hypothesis:

→ A hypothesis of no difference is called Null hypothesis. It is denoted by 'H_0'.

Alternative hypothesis:

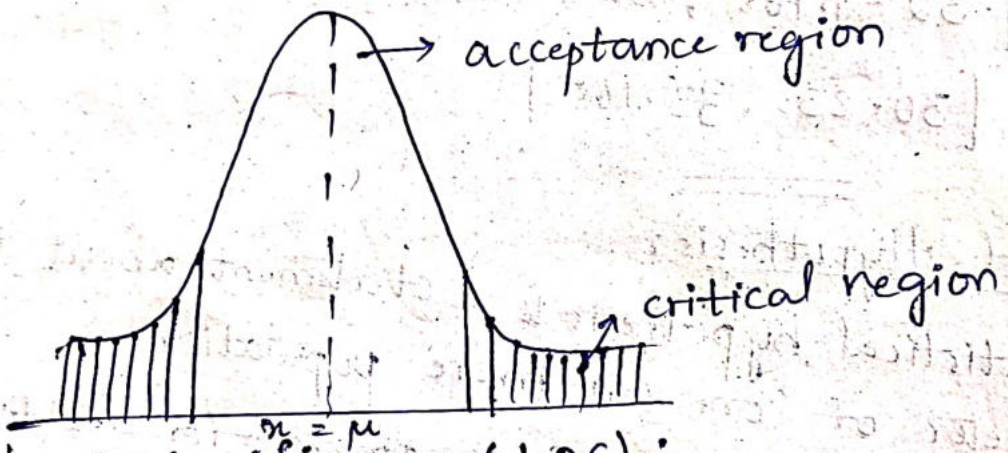
→ A hypothesis which contradicts the Null hypothesis is called Alternative hypothesis. It is denoted by 'H_1'.

Critical region:

→ A region corresponding to a statistic in the sample space which leads to reject Null hypothesis is called Critical region.

→ The region which leads to accept Null hypothesis is called 'Acceptance region'

→ The value of test statistic which separates rejection and acceptance region is called 'Critical value' (or) 'Significant value'.



level of significance (LOS):

→ The probability with which we will reject a Null hypothesis when it is true.

→ It is denoted by 'α' (alpha)

→ level of significance is the cut off value which determines whether to reject Null hypothesis (or) accept null hypothesis.

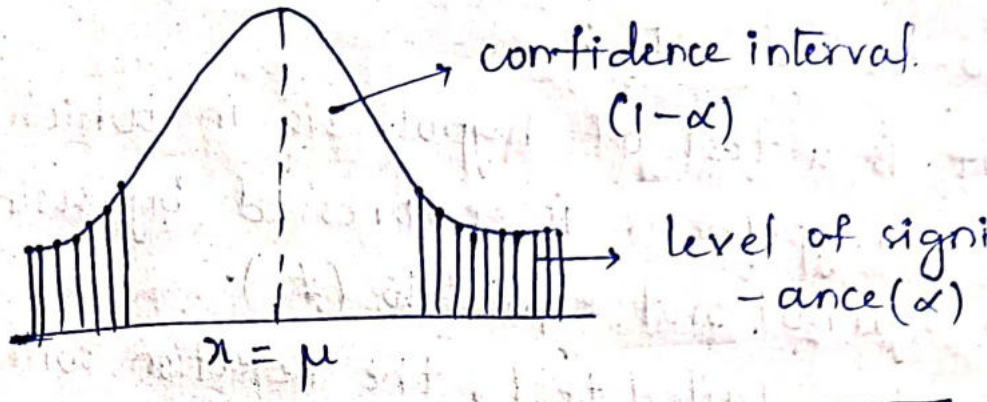
Confidence interval:

→ The probability with which we will accept a null hypothesis when it is true.

→ It is denoted by '$1-\alpha$'

→ The probability to accept a true null hypothesis

is denoted by $1-\alpha$.



LOS (α)		confidence interval $(1-\alpha)$	
5%	0.05	0.5%	0.95
1%	0.01	99%	0.99
10%	0.1	90%	0.90

*Errors of Sampling:

→ There are 2 types of errors.

1. Type-I error
2. Type-II error

Type-I error:

→ It is the error of rejecting null hypothesis (H_0) when it is true. The probability of making a Type-I error is denoted by 'α' i.e. level of significance

Type-II error:

→ It is the error of accepting the null hypothesis (H_0) when it is false. The Type-II error is denoted by 'β' i.e. P(reject H_0, when it is true)

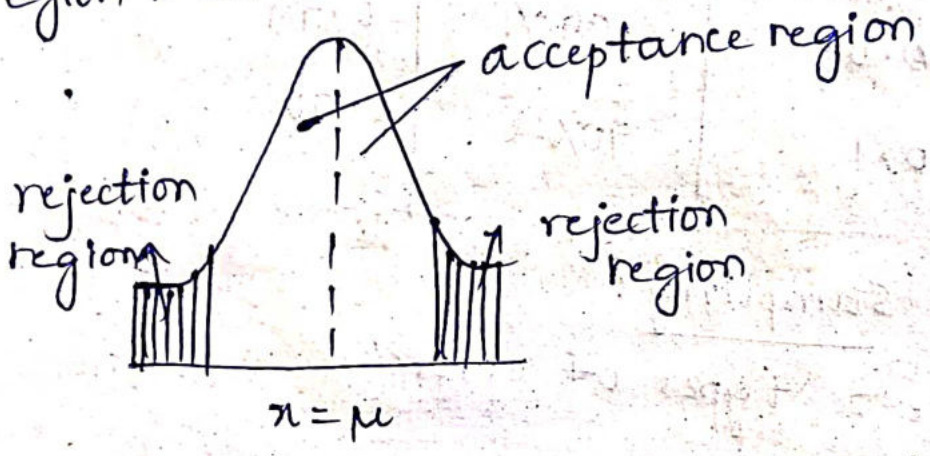
$P(\text{accept } H_0 \text{, when it is false}) = P(\text{type-II error})$
$= \beta$.

Two-tailed test:

→ It is a test of hypothesis in which alternative hypothesis is expressed by using the symbol not equal to ($\neq$).

→ In two-tailed test, the rejection will be on both the sides i.e. on the right and left side

→ In this test of hypothesis, the critical region is divided into 2 parts.



The critical region under the right tail
= The critical region under the left tail
= Half of the total area
$= \frac{1}{2}$ (probability of rejection)
$= \frac{1}{2}(\alpha)$
$= \frac{\alpha}{2}$ [Two-tailed test]

One-tailed test:

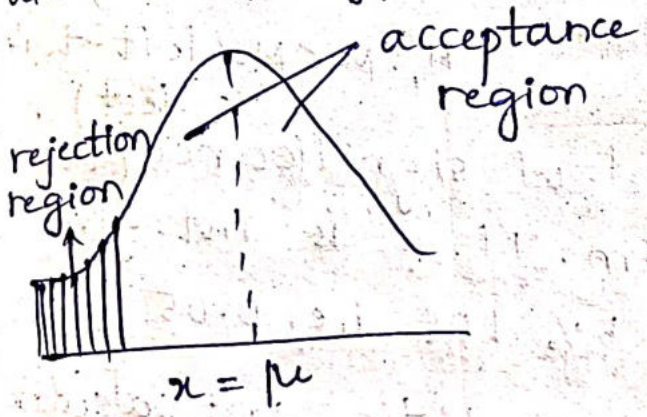
→ It is a test of hypothesis in which the alternative hypothesis is expressed by using

the symbol $< (\alpha) >$

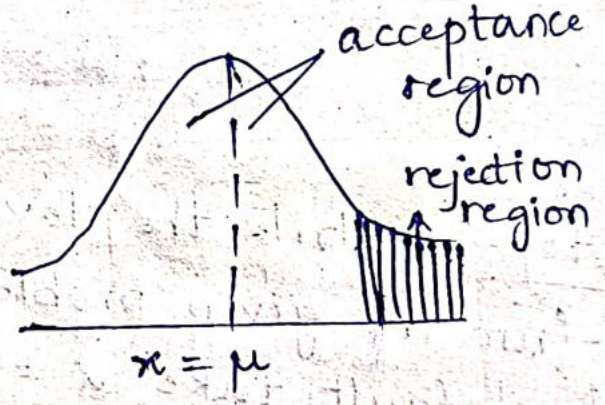
→ In this test of hypothesis, the entire critical region lies in the one tail of the distribution.

→ Critical region will lie at right tail if alternative hypothesis has $>$ symbol.

→ critical region will lie at left tail if alternative hypothesis has $<$ symbol.



left-tailed test



Right-tailed test

Note:

→ If the alternative hypothesis is two-tailed test it can be written as $\boxed{H_1 : \mu \neq \mu_0}$

→ If the alternative hypothesis is right-tailed test it can be written as $\boxed{H_1 : \mu > \mu_0}$

→ If the alternative hypothesis is left-tailed test it can be written as $\boxed{H_1 : \mu < \mu_0}$

	level of significance		
	1% (0.01)	5% (0.05)	10% (0.1)
Two-tailed test	$z_\alpha = 2.58$	$z_\alpha = 1.96$	$z_\alpha = 1.645$
right-tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
left-tailed test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

Procedure for testing of hypothesis:

Step-1: define (or) setup null hypothesis depending on the nature of the problem.

Step-2: define (or) setup alternative hypothesis inorder to decide whether to use 1 tailed test (or) 2 tailed test. i.e. $H_1 : \mu \neq \mu_0$ (2-tailed test)

$H_1 : \mu > \mu_0$ (right-tailed test)

$H_1 : \mu < \mu_0$ (left-tailed test).

Step-3: Write the level of significance (LOS) from the given problem. If 'α' is not given bydefault take $\alpha = 5\%$ LOS i.e. 0.05

Step-4: Write the test statistic and calculate 'Z' value i.e. Z calculated value.

Step-5: Now, compare $Z_{\text{calculated value}}$ and $Z_{\text{tabulated value}}$.

i) if $Z_{\text{calculated value}} < Z_{\text{tabulated}}$ then, we accept Null hypothesis.

ii) if $Z_{\text{calculated}} > Z_{\text{tabulated}}$ then, we reject Null hypothesis, i.e. We accept alternative hypothesis.

Methods of testing of hypothesis:

1. Test of significance for single mean
2. Test of significance for difference of 2 means.
3. Test of significare for single proportion

4. Test of significance for difference of 2 proportions

Method-1:

Test of significance for single mean

step-1: define null hypothesis depending on the nature of the problem.

step-2: define (or) setup alternative hypothesis. In order to decide whether to use 1 tailed test (or) 2 tailed test. i.e. $H_1 : \mu \neq \mu_0$ (2-tailed test)

$H_1 : \mu > \mu_0$ (right-tailed test)

$H_1 : \mu < \mu_0$ (left-tailed test)

step-3: Write LOS from the given problem. If 'α' is not given by default take $\alpha = 5\%$ LOS i.e 0.05

step-4: calculate the test statistic.

The test statistic for single mean is given by,

$$\boxed{Z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$$

where, μ = population mean

$\bar{x}$ = sample mean

σ = standard deviation

n = sample size

step-5: Compare $Z_{calculated}$ and $Z_{tabulated}$.

i) if $Z_{calculated} < Z_{tabulated} \Rightarrow$ accept null hypothesis.

ii) If $Z_{calculated} > Z_{tabulated} \Rightarrow$ reject null hypothesis, accept alternative hypothesis.

Q:- The mean life time of a sample of 100 lights produced by a company is found to be 1560 hrs with a population S.D of 90 hrs. Test the hypothesis for $\alpha = 0.05$ that the mean life

time of the lights produced by the company is 1580 hrs?

Sol: given, μ = population mean = 1580hrs

$\bar{x}$ = sample mean = 1560hrs

σ = 90hrs

n = 100

step-1: set up null hypothesis i.e.

$H_0 : \mu = \mu_0$,

$H_0 : \mu = 1580$hrs

i.e. The mean lifetime of the lights produced by the company is 1580 hrs.

step-2: set up alternative hypothesis.

i.e. $H_1 : \mu \neq \mu_0$ (2-tailed test)

$H_1 : \mu \neq 1580$hrs.

i.e. the mean lifetime of the lights produced by the company is not 1580 hrs.

step-3: given, LOS $\alpha = 0.05$ i.e. 5% LOS

Z_{tab} value $= 1.96$

step-4: The test statistic for single mean is,

$$|Z_{cal}| = \left|\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right| = \left|\frac{1560 - 1580}{\frac{90}{\sqrt{100}}}\right| = \left|\frac{-20}{\frac{90}{10}}\right|$$

$$= \left|\frac{-20}{9}\right| = 2.22$$

step-5: compare Z_{cal} and Z_{tab} values

$\Rightarrow Z_{cal} > Z_{tab}$

$2.22 > 1.96$

$\therefore$ We reject null hypothesis & accept alternative hypothesis.

Hence, The mean lifetime of the light produced by the company is not 1580 hrs.

Q:- In a random sample of 60 workers, the avg time taken by them to get into the work is 33.8 minutes with a S.D of 6.1 minutes. Can we reject the null hypothesis $\mu = 32.6$ minutes in favour of alternative hypothesis $\mu > 32.6$ min at $\alpha = 5\%$ LOS.

Sol: given, $\mu = 32.6$ min

$\bar{x} = 33.8$ min

$\sigma = 6.1$ min, $n = 60$

step-1: $H_0 : \mu = \mu_0 = 32.6$ min

step-2: $H_0 : \mu > 32.6$ (right-tailed test)

step-3: $\alpha = 5\% = 0.05$

$Z_{tab} = 1.645$

step-4: $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{33.8 - 32.6}{\frac{6.1}{\sqrt{60}}} = 1.523$

step-5: $Z_{tab} > Z_{cal} \Rightarrow Z_{cal} < Z_{tab}$

$\therefore$ We accept null hypothesis

An Ambulance service claims that it takes on the avg less than 10 min to reach its destination emergency calls. A sample of 36 calls has mean of 11 min and the variance 16 min. Test the claim at 0.05 LOS.

Sol: given, $n = 36$

$\mu = 10$ min

$\bar{x} = 11$ min

$\sigma^2 = 16$ min $\Rightarrow \sigma = 4$ min

Step-1: The ambulance service takes avg 10 min.

i.e. $H_0 : \mu = \mu_0 = 10$ min.

Step-2: The ambulance serivce takes less than 10 min. i.e. $H_1 : \mu < 10$ min (Left-tailed test)

Step-3: $\alpha = 0.05$

$|Z_{tab}| = |-1.645| = 1.645$

Step-4: $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{11-10}{\frac{4}{\sqrt{36}}} = \dfrac{1}{\frac{4}{6}} = \dfrac{\cancel{6}^{3}}{\cancel{4}_{2}}$

$= 1.5$

Step-5: $Z_{cal} < Z_{tab}$

$\therefore$ We accept null hypothesis

Hence, the ambulance service takes 10 min.

Q:- A sample of 400 items is taken from a population whose S.D is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also calculate 95% confidence interval for the population.

Sol:- given, $n = 400$

$\mu = 38$

$\bar{x} = 40$

$\sigma = 10$

Step-1: $H_0 : \mu = \mu_0 = 38$

Step-2: $H_1 : \mu \neq \mu_0 \Rightarrow \mu \neq 38$ (two-tailed test)

Step-3: $\alpha = 5\%$ LOS $= 0.05$

$Z_{tab} = 1.96$

step-4: $z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{40-38}{\frac{10}{\sqrt{400}}} = \frac{2}{\frac{10}{20}} = 4$

step-5: $z_{cal} > z_{tab}$

$\therefore$ We reject null hypothesis and accept alternative hypothesis.

95% confidence interval for single mean is

$$C.I = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$\Rightarrow \left[z_{\alpha/2} = z_{tab}\text{ value} = 1.96\right]$

$$C.I = \left[40 - \frac{(1.96)(10)}{\sqrt{400}}, 40 + \frac{(1.96)(10)}{\sqrt{400}}\right]$$

$$C.I = \left[40 - \frac{1.96}{2}, 40 + \frac{1.96}{2}\right]$$

$$= [40 - 0.98, 40 + 0.98]$$

$$C.I = [39.02, 40.98]$$

$$\begin{array}{r} 39.02 \\ 0.98 \\ \hline 39.02 \end{array}$$

Q:- A sample of 64 students have a mean weight 70kgs. Can this be regarded as a sample from a population with mean 56kgs and S.D 25kgs.

Sol: given, $n = 64$

$\mu = 56$ Kgs

$\bar{x} = 70$Kgs

$\sigma = 25$ Kgs

Step-1: $H_0 : \mu = \mu_0 = 56$ Kgs

Step-2: $H_1 : \mu \neq \mu_0 = 56$ Kgs (two-tailed test)

step-3: $\alpha = 5\%$ LOS $= 0.05$

$Z_{tab} = 1.96$

step-4: $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{70-56}{\frac{25}{\sqrt{64}}} = \dfrac{14}{\frac{25}{8}}$

$= \dfrac{14 \times 8}{25} = \dfrac{112}{25} = 4.48$

step-5: $Z_{cal} > Z_{tab}$

$\therefore$ We reject null hypothesis and accept alternative hypothesis.

Method-2:

Test of significance for difference of 2 means

step-1: set up null hypothesis i.e $H_0 : \mu_1 = \mu_2$

step-2: set up alternative hypothesis i.e

$H_1 : \mu_1 \neq \mu_2$ (2-tailed test)

$H_1 : \mu_1 > \mu_2$ (right tailed test)

$H_2 : \mu_1 < \mu_2$ (left tailed test)

step-3: Write LOS α value from the given problem. If not take $\alpha = 0.05$

step-4: Calculate the test statistic

The test statistic for difference of 2 means is given by $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $[\because \sigma_1^2 \neq \sigma_2^2]$

$z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $[\because \sigma_1^2 = \sigma_2^2]$

step-5: compare $Z_{calculated}$ and $Z_{tabulated}$

i) if $Z_{calculated} < Z_{tabulated}$ ⇒ accept null hypothesis.

ii) if $Z_{calculated} > Z_{tabulated}$ ⇒ reject null hypothesis, accept alternative hypothesis.

Q:- The research investigator is interested in studying whether there is a significant difference in the salaries of MBA graduates in two metropolitan cities. A sample of size 100 from mumbai earns an avg income of rupees 20,150/- Another random sample of size 60 from chennai earns an avg income of ₹. 20,250/-. If the variance of both the populations are given as $\sigma_1^2 = ₹.40,000/-$ and $\sigma_2^2 = ₹.32,400/-$.

Sol: given, $n_1 = 100$, $n_2 = 60$

$\bar{x}_1 = 20,150$, $\bar{x}_2 = 20,250$

$\sigma_1^2 = 40,000$, $\sigma_2^2 = 32,400$

step-1: set up null hypothesis. i.e there is no significant difference between 2 cities.

i.e $H_0 : \mu_1 = \mu_2$

Step-2: set up alternative hypothesis. i.e there is a significant difference between 2 cities.

i.e $H_1 : \mu_1 \neq \mu_2$ (two-tailed test)

Step-3: $\alpha = 5\% = 0.05$ LOS.

$Z_{tab} = 1.96$

Step-4: The test statistic for difference of 2 means is,

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \left| \frac{-100}{\sqrt{\frac{40,000}{100} + \frac{32,400}{60}}} \right|$$

$$= \frac{100}{\sqrt{400+540}} = \frac{100}{\sqrt{940}} = 3.26$$

$$Z_{cal} = 3.26$$

<u>step-5:</u> $Z_{cal} = 3.26$, $Z_{tab} = 1.96$

$$Z_{cal} > Z_{tab}$$

$\therefore$ we reject null hypothesis and accept alternative hypothesis.

Hence, there is a significant difference b/w 2 cities.

<u>Q:-</u> The mean of 2 large samples of size 1000 and 2000 members are 67.5 inches and 68 inches. Can the samples are regarded as drawn from the same population with S.D 2.5 inches.

<u>sol:</u> given, $n_1 = 1000$, $n_2 = 2000$

$\bar{x}_1 = 67.5$, $\bar{x}_2 = 68$

$\sigma = \sigma_1 = \sigma_2 = 2.5$

<u>step-1:</u> $H_0 : \mu_1 = \mu_2$

<u>step-2:</u> $H_1 : \mu_1 \neq \mu_2$ (two-tailed test)

<u>step-3:-</u> $\alpha = 5\% = 0.05 = LOS$

$$Z_{tab} = 1.96$$

<u>Step-4:</u> $$Z_{cal} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \right| = \frac{0.5}{2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}}$$

$$= \frac{\cancel{}}{\cancel{\sqrt{\frac{24}{2000}}}} =$$

$$= \frac{1}{5\sqrt{\frac{1}{1000}+\frac{1}{2000}}} = \frac{1}{5\times 0.038} = 5.26$$

Step-5: $z_{cal} > z_{tab}$

So, we reject null hypothesis and accept alternative hypothesis.

Q:- The avg marks got for 32 boys is 72 with S.D 8. While for 36 girls is 70, with a S.D of 6. Does this indicate that the boys perform better than girls at 5% LOS.

Sol: given, $n_1 = 32$, $n_2 = 36$

$\bar{x}_1 = 72$, $\bar{x}_2 = 70$

$\sigma_1 = 8$, $\sigma_2 = 6$

step 1: $H_0 : \mu_1 = \mu_2$

step 2: $H_1 : \mu_1 > \mu_2$ (right-tailed test)

step 3: $\alpha = 5\% = 0.05$ LOS

$z_{tab} = 1.645$

step-4: $$z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{2}{\sqrt{\frac{64}{32}+\frac{36}{36}}} = \frac{2}{\sqrt{3}}$$

$= 1.154$

step-5: $z_{cal} = 1.154$, $z_{tab} = 1.645$

$z_{cal} < z_{tab}$

∴ We accept null hypothesis

Hence, boys perform same as girls.

Q:- A sample of height of 6,400 englishmen has a mean 67.85 inches and a S.D of 2.56 inches. While a sample of height of 1600

australians has a mean of 68.55 inches with a S.D of 2.52 inches. Do the data indicate that the australians on the avg taller than the englishmen at 1% LOS.

Sol:- given, $n_1 = 6,400$, $n_2 = 1600$

$\bar{x}_1 = 67.85$, $\bar{x}_2 = 68.55$

$\sigma_1 = 2.56$, $\sigma_2 = 2.52$

step 1:- $H_0 : \mu_1 = \mu_2$

step 2: $H_1 : \mu_1 < \mu_2$ (left-tailed test)

step-3: $\alpha = 1\%$ LOS $= 0.01$

$Z_{tab} = |-2.33| = 2.33$

step-4: $Z_{cal} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{-0.7}{\sqrt{\dfrac{(2.56)^2}{6400} + \dfrac{(2.52)^2}{1600}}}$

$= \left|\dfrac{-0.7}{0.0706}\right| = 9.915$

step-5: $Z_{cal} > Z_{tab}$

So, we reject null hypothesis and accept alternative hypothesis.

1Q:- Samples of students are drawn from 2 universities and their weights are in Kgs. Mean and S.D are calculated and shown below. Make a large sample test to test the significance of difference between 2 means.

	mean	S.D	size
University A	55	10	400
University B	57	15	100

28:- The mean height of 50 male students who participated in sports is 68.2 inches with S.D 2.5. The mean height of another 50 male students who have not participated in the sports is 67.2 inches with SD 2.8. Test the hypothesis that the height of the students who participated in sports is more than the students who have not participated in the sports.

Sol: given, $n_1 = 400$, $n_2 = 100$

$\bar{x}_1 = 55$, $\bar{x}_2 = 57$

$\sigma_1 = 10$, $\sigma_2 = 15$

Step-1: $H_0 : \mu_1 = \mu_2$

Step-2: $H_1 : \mu_1 \neq \mu_2$ (two-tailed test)

Step-3: $\alpha = 5\% = 0.05$ LOS $\Rightarrow Z_{tab} = 1.96$

Step-4: $$Z_{cal} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| = \frac{2}{\sqrt{\frac{100}{400} + \frac{225}{100}}}$$

$$= \frac{2}{\sqrt{0.25 + 2.25}} = \frac{2}{\sqrt{2.50}}$$

$$Z_{cal} = 1.262$$

Step-5: $Z_{cal} < Z_{tab}$

$\therefore$ We accept null hypothesis.

29 Sol:- given, $n_1 = 50$, $n_2 = 50$

$\bar{x}_1 = 68.2$, $\bar{x}_2 = 67.2$

$\sigma_1 = 2.5$, $\sigma_2 = 2.8$

Step-1: $H_0 : \mu_1 = \mu_2$

Step-2: $H_1 : \mu_1 > \mu_2$ (right-tailed test)

Step-3: $\alpha = 5\%$ LOS $= 0.05$

$z_{tab} = 1.645$

Step-4: $z_{cal} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{1}{\sqrt{\dfrac{6.25+7.84}{50}}} = \dfrac{1}{\sqrt{\dfrac{14.09}{50}}}$

$= 1.883$

Step-5: $z_{cal} > z_{tab}$

So, we reject null hypothesis and accept alternative hypothesis.

∴ Height of the students who participated in sports is more than the students who have not participated in sports.

Method-3:

Testing of hypothesis for single proportion.

Step-1: setup null hypothesis

i.e. $H_0 : P = P_0$

Step-2: setup alternative hypothesis

i.e $H_1 : P \neq P_0$ (two-tailed test)

$H_1 : P > P_0$ (right-tailed test)

$H_1 : P < P_0$ (left-tailed test)

Step-3: Write LOS 'α' value from the given problem. If not, take $\alpha = 0.05$

Step-4: calculate the test statistic.

The test statistic for single proportion is given by,

$$\boxed{z_{cal} = \frac{p - P}{\sqrt{\dfrac{PQ}{n}}}}$$

where, P = population proportion, $Q = 1 - P$

$p = \dfrac{x}{n}$; x = no. of favourable cases

step-5: Compare Z_{cal} and Z_{tab} values

i) If $Z_{cal} < Z_{tab} \Rightarrow$ accept null hypothesis

ii) If $Z_{cal} > Z_{tab} \Rightarrow$ reject null hypothesis and accept alternative hypothesis

Q:- In a sample of 1000 people in Karnataka, 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% LOS.

Sol:- $n = 1000$

given, equally.

so, $P = \frac{1}{2}$, $Q = \frac{1}{2}$

$$p = \frac{x}{n} = \frac{54\cancel{0}}{100\cancel{0}} = \frac{54}{100} = 0.54$$

step-1: set up null hypothesis.

i.e. both rice and wheat eaters are equally popular in the state.

i.e. $H_0 : P = P_0 = 0.5$

step-2: set up alternative hypothesis. i.e both rice and wheat eaters are not equally popular in the state. i.e. $H_1 : P \neq P_0 = 0.5$ (two-tailed test)

step-3: given $\alpha = 1\% = 0.01$

$$Z_{tab} = 2.58$$

step-4: The test statistic for single proportion is,

$$Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1000}}}$$

$$= \frac{0.04}{\sqrt{\frac{0.25}{1000}}} = 2.529$$

Step-5: $Z_{cal} < Z_{tab}$

$\therefore$ We accept null hypothesis.

Hence, Both rice and wheat eaters are equally popular in the state.

Q:- In a big city, 325 men out of 600 men are found to be smokers. Does this information support the conclusion that majority of men in the city are smokers.

sol:- given, $n = 600$

$x = 325$

$P = \frac{1}{2}$, $Q = \frac{1}{2}$

$$p = \frac{x}{n} = \frac{\cancel{325}^{\cancel{65}\ 13}}{\cancel{600}_{\cancel{120}\ 24}} = \frac{13}{24} = 0.541$$

step-1: $H_0 : P = P_0 = 0.5$

step-2: $H_1 : P > P_0 = 0.5$ (right tailed test)

step-3: $\alpha = 0.05$

$Z_{tab} = 1.645$

step-4: $$Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.541 - 0.5}{\sqrt{\frac{0.25}{600}}}$$

$$= \frac{0.041}{\sqrt{\frac{0.25}{600}}} = 2.05$$

Step-5: $Z_{cal} > Z_{tab}$

$\therefore$ We reject null hypothesis and accept alternative hypothesis.

Hence, majority of men in the city are smokers.

Q:- Experience had shown that 20% of manufactured products are of top quality. In one day production of 400 articles only 50 are top quality. Test the hypothesis at 0.05 LOS.

Sol:- given, $n = 400$

$x = 50$

$P = 20\% = \frac{20}{100} = \frac{1}{5} = 0.2$

$Q = \frac{4}{5} = 0.8$

$p = \frac{50}{400} = \frac{1}{8} = 0.125$

step-1: $H_0 : P = P_0 = \frac{1}{5} = 0.2$

step-2: $H_1 : P \neq P_0 = 0.2$ (two-tailed test)

step-3: $\alpha = 0.05$

$Z_{tab} = 1.96$

step-4: $Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{\frac{1}{8} - \frac{1}{5}}{\sqrt{\frac{1}{5} \times \frac{4}{5} \times \frac{1}{400}}} = \frac{0.075}{\sqrt{\frac{(0.2)(0.8)}{400}}}$

$= 3.75$

step-5: $Z_{cal} > Z_{tab} \Rightarrow$ reject null hypothesis and accept alternative hypothesis.

Q:- In a hospital, 480 females and 520 male babies were born in a week. Do these figures confirm the hypothesis that males and females are born in equal number.

Sol:- $n = 480 + 520 = 1000$

females,

$x = 480$

$n = 1000$

$P = \frac{1}{2}$, $Q = \frac{1}{2}$

$p = \frac{x}{n} = \frac{480}{1000} = 0.48$

step 1:- $H_0 : P = P_0 = 0.5$

step-2: $H_1 : P \neq P_0 = 0.5$ (two-tailed test)

step-3: $\alpha = 0.05$

$Z_{tab} = 1.96$

step-4: $Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.48 - 0.5}{\sqrt{\frac{0.25}{1000}}}$

$= \frac{0.02}{\sqrt{\frac{0.25}{1000}}} = 1.264$

step-5: $Z_{cal} < Z_{tab}$

$\therefore$ We accept null hypothesis

Hence, males and females are born in equal number.

Note:

The confidence interval for single proportion is,

$$CI = \left[p - Z_{\alpha/2}\sqrt{\frac{PQ}{n}},\ p + Z_{\alpha/2}\sqrt{\frac{PQ}{n}} \right]$$

$Z_{\alpha/2} = LOS$

Q:- In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. Construct 99% confidence interval for the corresponding true

rtion.

given, $n = 160$

$x = 24$

$p = \frac{x}{n} = \frac{24}{160} = 0.15$

$q = 1 - 0.15 = 0.85$

% confidence interval is equal to 1% LOS.

$Z_{\alpha/2} \Rightarrow \alpha = 0.01 \Rightarrow Z_{tab} = 2.58$

$= \left[0.15 - 2.58\sqrt{\frac{(0.15)(0.85)}{160}},\ 0.15 + 2.58\sqrt{\frac{(0.15)(0.85)}{160}}\right]$

$= [0.15 - 0.072,\ 0.15 + 0.072]$

$\therefore I = [0.078,\ 0.222]$

A manufacturer claims that atleast 95% of equipment which he supplies to the factory good. An examination of a sample of 200 ces of equipment revealed that 18 are fective. Test the claim at 5% LOS.

given, $n = 200$

$x = 200 - 18 = 182$

$p = 95\% = \frac{95}{100} = 0.95$

$q = 0.05$

$p = \frac{x}{n} = \frac{182}{200} = 0.91$

step-1: $H_0 : P = P_0 = 0.95$

step-2: $H_1 : P \neq P_0 = 0.95$ (two-tailed test)

step-3: $\alpha = 0.05 \Rightarrow Z_{tab} = 1.96$

Step-4: $z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \left| \frac{0.91 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{200}}} \right|$

$= \frac{0.04}{0.0154} = 2.59$

Step-5: $z_{cal} > z_{tab}$

$\therefore$ We reject null hypothesis.

Method-4:

→ Test of significance for difference of 2 proportions

Step-1: set up NULL hypothesis i.e. H_0 there is no significant difference between 2 proportions.

i.e. $H_0 : P_1 = P_2$

Step-2: set up alternative hypothesis.

H_1; there is a significant difference between 2 proportions

$H_1 : P_1 \neq P_2$ (2-tailed test)

$H_1 : P_1 > P_2$ (right-tailed test)

$H_1 : P_1 < P_2$ (left-tailed test)

Step-3: Write LOS α from the given problem if not take $\alpha = 0.05$

Step-4: Calculate the test statistic.

The test statistic for difference of 2 proportions is,

$$z_{cal} = \frac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where, $\hat{p} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$

$\hat{q} = 1 - \hat{p}$.

$P_1 = \frac{x_1}{n_1}$, $P_2 = \frac{x_2}{n_2}$

step-5: compare z_{cal} and z_{tab}

If $z_{cal} < z_{tab} \Rightarrow$ accept null hypothesis

If $z_{cal} > z_{tab} \Rightarrow$ reject null hypothesis and accept alternative hypothesis.

Q:- Random samples of 400 men and 600 women are asked whether they would like to have a flyover near their residence. 200 men and 325 women are in the favour of the proposal. Test the hypothesis that proportions of men and women in the favour of the proposal are same at 5% LOS.

sol: given, $n_1 = 400$, $n_2 = 600$

$x_1 = 200$, $x_2 = 325$

$$P_1 = \frac{x_1}{n_1} = \frac{\cancel{200}}{\cancel{400}_{2}} = 0.5$$

$$P_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.5416$$

step-1: set up null hypothesis H_0; there is no significant difference between the opinions of men and women in favour of the proposal. i.e $H_0 : P_1 = P_2$

step-2: set up alternative hypothesis.

H_1 : There is a significant difference between men and women in favour of the proposal $H_1 : P_1 \neq P_2$ (two-tailed test)

Step-3: $\alpha = 0.05$

$Z_{tab} = 1.96$

Step-4: The test statistic for difference of proportions is, $Z_{cal} = \dfrac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$\hat{p} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{(400)(0.5) + 600(0.5416)}{400 + 600}$$

$$\frac{524.6}{100} = 0.5246$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{200 + 325}{400 + 600} = \frac{525}{1000}$$

$$\hat{p} = 0.525$$

$$\hat{q} = 1 - \hat{p} = 1 - 0.525 = 0.475$$

$$Z_{cal} = \left| \frac{0.5 - 0.5416}{\sqrt{(0.525)(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}} \right|$$

$$= \frac{0.041}{0.0322} = 1.27$$

Step-5: $Z_{cal} = 1.27$

$Z_{tab} = 1.96$

$Z_{cal} < Z_{tab}$

$\therefore$ We accept null hypothesis

$\therefore$ There is no significant difference b/w the opinions of men and women

---

Q:- In a sample of 600 students of a certain college, 400 are found to be using ball pens. In another college from a sample of 900 students 450 were found to use ball pens.

whether 2 colleges are significantly different
respect to the habit of using ball pens.

$n_1 = 600$, $n_2 = 900 \Rightarrow P_1 = \frac{x_1}{n_1} = 0.66$

$x_1 = 400$, $x_2 = 450 \Rightarrow P_2 = \frac{x_2}{n_2} = 0.5$

p1: $H_0 : P_1 = P_2$

p2: $H_1 : P_1 \neq P_2$ (2-tailed test).

p-3: $\alpha = 0.05$

$Z_{tab} = 1.96$

ep-4: $Z_{cal} = \dfrac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2} = \dfrac{400+450}{600+900} = \dfrac{850}{1500} = 0.566$

$\hat{q} = 1 - \hat{p} = 1 - 0.566 = 0.434$

$Z_{cal} = \dfrac{0.66 - 0.5}{\sqrt{(0.566)(0.434)\left(\frac{1}{600} + \frac{1}{900}\right)}} = 6.125$

ep-5: $Z_{cal} = 6.125$

$Z_{tab} = 1.96$

$Z_{cal} > Z_{tab}$

∴ We reject null-hypothesis.

. There is significant difference w.r.to the habit of using ball pens.

: A machine puts out 16 imperfect articles in sample of 500 articles, After the machine is overhauled (after the repair) it puts out 3 imperfect articles in a sample of 100 articles has the machine improved.

2Q: Before an ↑ on excise duty 500 people out of 900 found to have the habit of having tea after on ↑ an excise duty 250 are having the habit of having tea among 1100. Is there any ↓ in the consumption of the tea. Test at 5% level

Note: If only population proportions are given and sample proportions are not known then the test statistic is, $Z_{cal} = \dfrac{P_1 - P_2}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}}$

Q:- In 2 large populations there are 30% and 25% fair hair people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the 2 populations.

Sol:- given, $n_1 = 1200$, $n_2 = 900$

$$P_1 = 30\% = \frac{30}{100} = 0.3$$

$$P_2 = 25\% = \frac{25}{100} = 0.25$$

$$Q_1 = 1 - P_1 = 1 - 0.3 = 0.7$$

$$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$$

Step-1: $H_0: P_1 = P_2$

Step-2: $H_1: P_1 \neq P_2$ (2-tailed test)

Step-3: $\alpha = 0.05$

$$Z_{tab} = 1.96$$

Step-4: $Z_{cal} = \dfrac{0.3 - 0.25}{\sqrt{\dfrac{(0.3)(0.7)}{} + \dfrac{(0.25)(0.75)}{}}} = 2.55$

-5: $Z_{cal} = 2.55$

$Z_{tab} = 1.96$

$Z_{cal} > Z_{tab}$

ale reject null hypothesis

here is significant difference hidden in samples.

$n_1 = 500$, $n_2 = 100$ $\rightarrow$ $P_1 = \frac{x_1}{n_1} = \frac{16}{500} = 0.032$

$x_1 = 16$, $x_2 = 3$ $\Rightarrow$ $P_2 = \frac{x_2}{n_2} = \frac{3}{100} = 0.03$

p-1: $H_0 : P_1 = P_2$

ep-2: $H_1 : P_1 > P_2$ (right-tailed test)

ep-3: $\alpha = 0.05$

$Z_{tab} = 1.645$

tep-4: $Z_{cal} = \frac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{16+3}{500+100} = \frac{19}{600} = 0.031$

$\hat{q} = 1 - \hat{p} = 1 - 0.031 = 0.969$

$Z_{cal} = \frac{0.032 - 0.03}{\sqrt{(0.031)(0.969)\left(\frac{1}{500} + \frac{1}{100}\right)}}$

$= \frac{0.002}{\sqrt{(0.031)(0.969)\left(\frac{1}{500} + \frac{1}{100}\right)}} = 0.976$

Step-5: $Z_{cal} < Z_{tab}$ $\Rightarrow$ We accept null hypothesis.

$\therefore$ The machine doesn't improve

(HW)

2) $n_1 = 900,\ n_2 = 1100 \Rightarrow P_1 = \frac{x_1}{n_1} = \frac{500}{900} = 0.555$

$x_1 = 500,\ x_2 = 250 \Rightarrow P_2 = \frac{x_2}{n_2} = \frac{250}{1100} = 0.227$

<u>step-1</u>: $H_0 : P_1 = P_2$

<u>step-2</u>: $H_1 : P_1 < P_2$ (left-tailed test)

<u>step-3</u>: $\alpha = 0.05$

$Z_{tab} = -1.645$

<u>step-4</u>: $Z_{cal} = \frac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{500 + 250}{900 + 1100} = \frac{750}{2000} = \frac{75}{200} = 0.375$

$\hat{q} = 1 - \hat{p} = 1 - 0.375 = 0.625$

$Z_{cal} = \frac{0.555 - 0.227}{\sqrt{(0.375)(0.625)\left(\frac{1}{900} + \frac{1}{1100}\right)}}$

$= \frac{0.328}{\sqrt{0.0004734848}} = 15.07$

<u>step-5</u>: $Z_{cal} > Z_{tab}$

$\therefore$ We reject null hypothesis

$\therefore$ There is decrease in the consumption of tea.

Unit-4: Test of significance for small samples.

mall sample:
If the size of the sample $n \leq 30$ then the samples are said to be small samples.
Under small sample test, we have 3 types of tests.

. t-test (or) student t-test
. F-test
. chi-square test (χ^2-test)

lethod-1:
t-test (or) student t-test

Assumptions:
→ The size of the sample should be ≤ 30
→ The parent population from which the sample is drawn is normal.
The population S.D is unknown.
The sample observations are independent

ses:
→ to test for a specified mean.
→ to test for equality of 2 means of 2 independent samples drawn from 2 normal populations, S.D of population is unknown
→ to test the significance of difference between the means of paired data.

Properties:
→ The shape of t-distribution is 'bell shape' which is similar to normal distribution and is symmetrical about the mean.

→ It is also symmetrical about the line $t = 0$

→ The form of the probability curves changes with decrease of freedom i.e with sample size.

→ It is uni-model. i.e. $\boxed{\text{Mean} = \text{median} = \text{mode}}$

→ The mean of standard normal distribution and t-distribution is zero(0) but the variance of t-distribution depends up on the parameter ν which is called degrees of freedom.

Procedure:

Step-1: set up null hypothesis. i.e $H_0 : \mu = \mu_0$

Step-2: set up alternative hypothesis.

i.e. $H_1 : \mu \neq \mu_0$ (2-tailed test)

$H_1 : \mu > \mu_0$ (right tailed test)

$H_1 : \mu < \mu_0$ (left tailed test)

Step-3: Write LOS α value from the given problem if not, take $\alpha = 0.05$.

Step-4: calculate the test statistic.

The test statistic for t-distribution/test is,

$$t_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \quad [\text{If S.D is known}]$$

$$t_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad [\text{if SD is unknown}]$$

Step-5: find degrees of freedom (DOF)

The DOF for t-test is $\boxed{\nu = n-1}$

corresponding to α value from t-distribution table which gives t_{tab} value

Compare t_{cal} and t_{tab} values.

1) If $t_{cal} < t_{tab}$ then we accept null hypothesis.
2) If $t_{cal} > t_{tab}$ then we reject null hypothesis.

Note:
The t-distribution table values are given for only 1-tailed test. So, incase of 2-tailed test consider LOS as $\frac{\alpha}{2}$

Degrees of freedom:
The no. of independent values which can be assigned to a statistical distribution. It is denoted by 'V' and defined as $\boxed{V = n - k}$

Q:- A sample of 26 bulbs has a mean life of 990 hours with a standard deviation of 20hrs. The manufacturer claims that the mean life of the bulbs is 1000 hrs. Is the sample not up to the standard.

Sol:- given, $n = 26$
$\mu = 1000$hrs = population mean
$\bar{x} = 990$hrs = sample mean
$s = 20$hrs

Step-1: setup null hypothesis (H_0). The sample is up to the standard.
i.e. $H_0 : \mu = \mu_0 = 1000$hrs.

Step-2: setup alternative hypothesis
H_1: The sample is not up to the standard $\Rightarrow H_1 : \mu \neq \mu_0 \neq 1000$.. (2-tailed test)

Step-3: $\alpha = 0.05$
$\because$ it is a 2-tailed test
$\alpha = \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

Step-4: The test statistic for t-test is,

$$t_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \quad [\because S.D \text{ is given}].$$

$$= \left| \frac{990 - 1000}{\frac{20}{\sqrt{25}}} \right| = \frac{10}{\frac{20}{5}} = \frac{\cancel{10}^{5}}{\cancel{4}_{2}}$$

$$= 2.5$$

Step-5: find degrees of freedom

i.e. $V = n - 1 = 26 - 1 = 25$

$$\boxed{V = 25}$$

$\alpha = \frac{\alpha}{2} = 0.025$

Now, find the value from t-distribution table corresponding DOF V and α value

$$t_{tab} = t(V, \alpha) = t(25, 0.025)$$

$$= 2.060.$$

compare t_{cal} and t_{tab} values

i.e. $t_{cal} = 2.5$, $t_{tab} = 2.060$

$t_{cal} > t_{tab} \Rightarrow$ we reject null hypothesis

Hence, the sample is not up to the standard.

Q:- The average breaking strength of steel rods is specified to be 18.5 pounds. To test this sample of 14 rods with mean and S.D obtained as 17.85 and 1.955. Is the result of the experiment is significant

Sol: $n = 14$

$\mu = 18.5$

$\bar{x} = 17.85$

step-1: $H_0 : \mu = \mu_0 = 18.5$

step-2: $H_1 : \mu \neq \mu_0$ (2 tailed test)

step-3: $\alpha = \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$ ($\because$ 2 tailed test)

step-4: $t_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$ ($\because$ SD is given)

$$= \left| \frac{17.85 - 18.5}{\frac{1.955}{\sqrt{13}}} \right| = \frac{0.65}{0.5422} = 1.1988$$

step-5: DOF (V) $= n - 1 = 14 - 1 = 13$

$t_{tab} = t_{(13,\ 0.025)} = 2.160$

$t_{cal} < t_{tab} \Rightarrow$ we accept null hypothesis

Hence, the result of the experiment is significant.

Problems based on t-test when standard deviation of the sample is not given directly:

Q:- A random sample of 10 boys having the following IQ's. 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100.

1) Do this data support the assumption of the population mean IQ of 100.

2) find a reasonable range in which most of the mean IQ values of the sample of 10 boys lies.

Sol: Here, the mean and standard deviation of the sample is not given directly.

We have to determine the mean and S.D as follows.

$$\text{Mean}(\bar{x}) = \frac{70+120+\dots+100}{10}$$

$$= \frac{972}{100} = 97.2$$

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

$$= \frac{1}{9}\left[(70-97.2)^2 + (120-97.2)^2 + \dots + (100-97.2)^2\right]$$

$$= \frac{1}{9}(1833.6)$$

$$= 203.7333$$

$$s = \sigma^2 = \sqrt{s^2} = \sqrt{203.733} = 14.27$$

$\therefore n = 10$

$\bar{x} = 97.2$

$\mu = 100$

$s = 14.27$

<u>Step-1</u>: set up null hypothesis H_0.

The mean IQ of the students is 100

$$H_0 : \mu = \mu_0 = 100$$

<u>Step-2</u>: set up alternative hypothesis H_1.

The mean IQ of the students is not 100

$$H_1 : \mu \neq \mu_0$$

<u>Step-3</u>: $\alpha = 0.05$ (two-tailed test

$$\alpha = \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

<u>Step-4</u>: $t_{cal} = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

$$= \left| \frac{97.2 - 100}{\frac{14.27}{\sqrt{10}}} \right| = \frac{2.8}{4.515} = 0.6201$$

-5: $v = n - 1$

$v = 10 - 1 = 9$

$\alpha = \frac{\alpha}{2} = 0.025$

$t_{tab} = t_{(v, \alpha)} = t_{(9, 0.025)} = 2.262$

$t_{cal} < t_{tab} \Rightarrow$ We accept null hypothesis.

nce, the mean IQ of the students is 100.

The confidence interval for single mean is,

$$C.I = \left[\bar{x} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right), \bar{x} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) \right]$$

$$\left[t_{\alpha/2} = t_{tab} \right]$$

$$= \left[97.2 - 2.262\left(\frac{14.27}{\sqrt{10}}\right), 97.2 + 2.262\left(\frac{14.27}{\sqrt{10}}\right) \right]$$

$$= [86.997, 107.407]$$

$$C.I = [86.997, 107.407]$$

:- The life time of electric bulbs for a random sample of 10 from a factory are given below.

Items	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 avg

lifetime of electrical bulbs is 4000 hrs use 0.05 LOS.

Sol:- $n = 10$

$$\text{Mean} = \bar{x} = \frac{41}{10} = 4.1 \Rightarrow 4.1 \times 1000 = 4100$$

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

$$= \frac{1}{9}\left[(1.2-4.1)^2 + (4.6-4.1)^2 + \dots + (5.6-4.1)^2\right]$$

$$= \frac{1}{9}(12.42) = 1.38$$

$$\Rightarrow 1.38 \times 1000 = 1380$$

$$s = \sqrt{1380} = 37.14$$

Step-1: set up null hypothesis (H_0).

$$H_0 : \mu = \mu_0 = 4000$$

Step-2: set up alternative hypothesis

$$H_1 : \mu \neq \mu_0 \text{ (two-tailed test)}$$

Step-3: $\alpha = 0.05$

$$\alpha = \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

Step-4: $$t_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{4100 - 4000}{\frac{37.14}{\sqrt{10}}} = \frac{100}{\frac{37.14}{\sqrt{10}}}$$

$$= \frac{100}{\frac{37.14}{3.16}} = \frac{100}{11.75} = 8.508$$

Step-5: $v = n - 1 = 10 - 1 = 9$

$$t_{tab} = t_{(v,\alpha)} = t_{(9, 0.025)} = 2.262$$

$t_{cal} > t_{tab} \Rightarrow$ we reject null hypothesis

$\therefore$ The avg lifetime of electrical bulbs is not 4000 hrs.

8 students are given a test in statistics and after
onth of coaching they are given another test.
following table gives the increase in marks
2nd test over the 1st.

udent no.	1	2	3	4	5	6	7	8
ncrease of marks	4	-2	6	-8	12	5	-7	2

sume that the students are not benefited
the coaching it implies that the mean of
difference of 2 tests is zero.

Mean $\bar{x} = \frac{12}{8} = 1.5$

$$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{7}\left[(4-1.5)^2 + \cdots + (2-1.5)^2\right]$$

$$= \frac{324}{7} = 46.28$$

$$s = \sqrt{46.28} = 6.80$$

Step-1: $H_0 : \mu = \mu_0 = 0$

Step-2: $H_1 : \mu > \mu_0$ (right-tailed test)

Step-3: $\alpha = 0.05$

Step-4: $t_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{1.5}{\frac{6.80}{\sqrt{8}}} = \frac{1.5}{2.411} = 0.622$

Step-5: $v = n - 1 = 7$

$t_{tab} = t_{(7, 0.05)} = 1.895$

$t_{cal} < t_{tab} \Rightarrow$ We accept null hypothesis.

$\therefore$ The students are not benefited by coaching.

Q:- A random sample of 10 bags of pesticides are taken as 50, 49, 52, 44, 45, 48, 46, 45, 49, 45 Test whether the avg packing can be taken to 50 kgs.

Sol:- Mean $= \dfrac{50+49+52+44+\cdots+45}{10}$

$= \dfrac{473}{10} = 47.3$

Variance$(s^2) = \dfrac{1}{n-1}\sum(x_i-\bar{x})^2 = \dfrac{64.1}{9}$

$= 7.1222$

$s = \sqrt{7.1222} = 2.668$

Step-1: set up null hypothesis (H_0)

$H_0 : \mu = \mu_0 = 50$

Step-2: set up alternative hypothesis (H_1)

$H_1 : \mu \neq \mu_0 = 50$ (two-tailed test)

Step-3: $\alpha = 0.05$

$\alpha = \dfrac{\alpha}{2} = \dfrac{0.05}{2} = 0.025$

Step-4: $t_{cal} = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \left|\dfrac{47.3-50}{\frac{2.668}{\sqrt{10}}}\right| = \dfrac{+2.7}{0.8436}$

$= 3.204$

Step-5: $v = n-1 = 10-1 = 9$

$t_{tab} = t_{(v,\alpha)} = t_{(9,0.025)} = 2.262$

$t_{cal} > t_{tab} \Rightarrow$ We reject null hypothesis

$\therefore$ The avg packing cannot be taken to 50Kgs.

dent t-test for difference of means:

t, $\bar{x}$ and $\bar{y}$ be the means of 2 independent
mples of sizes n_1 and n_2 drawn from the
ulation having means μ_1 and μ_2. To test the
nificance for the difference of 2 means we
low the following steps.

-1: set up null hypothesis (H_0). $H_0 : \mu_1 = \mu_2$

p-2: set up alternative hypothesis

$H_1 : \mu_1 \neq \mu_2$ (2-tailed test)

$H_1 : \mu_1 > \mu_2$ (right tailed test)

$H_1 : \mu_1 < \mu_2$ (left tailed test)

tep-3: Write LOS α value from the given
problem if not take $\alpha = 0.05$

tep-4: calculate the test statistic.
The test statistic for difference of 2 means
is,

$$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where, $s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$ [If SD is known]

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

(If SD is unknown)

tep-5: find DOF i.e. $\boxed{v = n_1 + n_2 - 2}$ and find
the t_{tab} value from t-distribution table
corresponding DOF(v) and LOS (α) value

If $t_{cal} < t_{tab} \Rightarrow$ We accept null hypothesis

$t_{cal} > t_{tab} \Rightarrow$ we reject null hypothesis.

Q: Samples of 2 types of ... and the following data is obtained.

Type - I	Type - II
$n_1 = 8$	$n_2 = 7$
$\bar{x} = 1234$ hrs	$\bar{y} = 1036$ hrs
$s_1 = 36$ hrs	$s_2 = 40$ hrs

Is the difference in the means sufficient to warranty that type - I is superior to type - II regarding the lifetime of the bulbs.

Soln- step-1: setup null hypothesis
i.e There is no difference between 2 types of electric bulbs regarding the lifetime
i.e $H_0 : \mu_1 = \mu_2$

step-2: set up alternative hypothesis
i.e Type - I is superior to type - II regarding the life time. i.e. $H_1 : \mu_1 > \mu_2$ (right tailed test)

step-3: $\alpha = 0.05$ (LOS)

step-4: The test statistic for difference of 2 means is, $t_{cal} = \dfrac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

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

$s^2 = 1659.076$

$t_{cal} = \dfrac{1234 - 1036}{\sqrt{1659.076\left(\frac{1}{8} + \frac{1}{7}\right)}} = 9.392$

step-5: $V = n_1 + n_2 - 2 = 8 + 7 - 2 = 13$

$t_{tab} = t_{(v, \alpha)} = t_{(13, 0.05)} = 1.771$

$t_{cal} > t_{tab} \Rightarrow$ We reject null hypothesis.

Hence, type-I is superior to type-II regarding the lifetime.

Q:- 2 independent samples of 8 and 7 items respectively had the following values.

type-I sample-I	11	11	13	11	15	9	12	14
type-II sample-II	9	11	10	13	9	8	10	—

Sol: $\bar{x} = \frac{\text{sum of ob}}{\text{no. of ob}} = \frac{96}{8} = 12$; $n_1 = 8$

$\bar{y} = \frac{9 + 11 + \cdots + 10}{7} = \frac{70}{7} = 10$; $n_2 = 7$.

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

$$s^2 = \frac{1}{13}(26 + 16) = \frac{1}{13}(42) = 3.2307$$

$$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{12 - 10}{\sqrt{3.2307\left(\frac{1}{8} + \frac{1}{7}\right)}}$$

$= 2.1499 = 2.15$

Step-1: set up null hypothesis.

there is no difference b/w 2 samples

i.e. $H_0 : \mu_1 = \mu_2$

Step-2: set up alternate hypothesis

there is significant difference b/w 2 samples

$H_1 : \mu_1 \neq \mu_2$

step-3: $\alpha = \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

step-4: $t_{cal} = 2.15$

step-5: DOF $(v) = n_1 + n_2 - 2 = 8 + 7 - 2 = 13$

$\alpha = 0.025$

$t_{tab} = t_{(13, 0.025)} = 2.160$

$t_{cal} < t_{tab} \Rightarrow$ We accept null hypothesis

1Q:- The IQ (Intelligent Quotient) of 16 students of 1 area of a city has a mean 107 with S.D 10, while the IQ's of 14 students from another area of the city has a mean 112 with SD 8. Is there any significant difference between the IQ's of 2 groups at 0.05 LOS.

2Q:- To examine the hypothesis that the husbands are more intelligent than the wifes, an investigator took a sample of 10 couples and tested the IQ's. The result is as follows.

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

Test the hypothesis at 5% LOS.

1Sol: step-1: $H_0 : \mu_1 = \mu_2$

step-2: $H_0 : \mu_1 \neq \mu_2$ (2-tailed test)

step-3: $\alpha = 0.05 \Rightarrow \alpha = \frac{\alpha}{2} = 0.025$

-4: $n_1 = 16$, $n_2 = 14$

$\bar{x} = 107$, $\bar{y} = 112$

$s_1 = 10$, $s_2 = 8$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{16(10)^2 + 14(8)^2}{16+14-2} = \frac{2496}{28}$$

$$= 89.142$$

$$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \left|\frac{107 - 112}{\sqrt{89.142\left(\frac{1}{16} + \frac{1}{14}\right)}}\right| = 1.447$$

$$= 1.447.$$

-5: $v = n_1 + n_2 - 2 = 28$

$$t_{tab} = t_{(28, 0.025)} = 2.048$$

$t_{cal} < t_{tab} \Rightarrow$ We accept null hypothesis

ence, there is no significant difference between the IQ's of 2 groups.

ol: $$\bar{x} = \frac{117 + 105 + \cdots + 107}{10} = \frac{1030}{10} = 103$$

$$\bar{y} = \frac{106 + \cdots + 85}{10} = \frac{958}{10} = 95.8$$

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

$$= \frac{1}{18}(1606 + 1679.6) = \frac{3285.6}{18} = 182.53$$

step-1: $H_0 : \mu_1 = \mu_2$

step-2: $H_0 : \mu_1 > \mu_2$ (right-tailed test)

step-3: $\alpha = 0.05$

Step-4: $t_{tab} = \dfrac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$t_{tab} = \frac{108 - 95.8}{\sqrt{182.53\left(\frac{1}{10} + \frac{1}{10}\right)}} = \frac{7.2}{\sqrt{182.53\left(\frac{1}{5}\right)}}$$

$$= \frac{7.2}{\sqrt{36.506}} = 1.191$$

Step-5: $v = n_1 + n_2 - 2 = 18$

$$t_{tab} = t_{(v,\alpha)} = t_{(18, 0.05)} = 1.734$$

~~$t_{cal} > t_{tab}$ ⇒ We reject null hypothesis~~

$t_{cal} < t_{tab}$ ⇒ We accept null hypothesis

Hence, husbands and wifes are equally intelligent.

F - Distribution / F - test:

→ When testing the significance of difference of means of 2 samples, we assume that 2 samples came from same population (or) from populations with equal variances.

→ If the variance of the population are not equal, a significant difference in the means may occur. Hence before we apply t-test for the significance of difference of 2 means we have to test equality of population variances using F-Test.

Note:

→ t-test is used to test the significance of difference of 2 means.

-Test is used to test the significance of
fference of 2 variances.

dure:

-1: set up null hypothesis

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

-2: set up alternative hypothesis

i.e $H_1 : \sigma_1^2 \neq \sigma_2^2$ (2-tailed)

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

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

p-3: Write LOS α value from the given problem
if not, take $\alpha = 0.05$

p-4: Calculate the test statistic.

he test statistic for F test is,

$$\boxed{F_{cal} = \frac{\text{Greater variance}}{\text{Smaller variance}}}$$

If $s_1^2 > s_2^2$ $F_{cal} = \frac{s_1^2}{s_2^2}$

If $s_2^2 > s_1^2$ $F_{cal} = \frac{s_2^2}{s_1^2}$

Where,

$$\left.\begin{aligned} s_1^2 &= \frac{1}{n_1-1} \Sigma (x_i - \bar{x})^2 \\ s_2^2 &= \frac{1}{n_2-1} \Sigma (y_i - \bar{y})^2 \end{aligned}\right\}$$ If sum of squares of deviation is known.

$$\left.\begin{aligned} s_1^2 &= \frac{n_1 s_1^2}{n_1-1} \\ s_2^2 &= \frac{n_2 s_2^2}{n_2-1} \end{aligned}\right\}$$ If only s.D of sample is known.

Step-5: Find DOF i.e $v_1 = n_1 - 1$
$v_2 = n_2 - 1$

Find tabluated value from F-distribution table

corresponding to V_1, V_2 and α values.
Compare F_{cal} and F_{tab} values

i) If $F_{cal} < F_{tab}$ $\Rightarrow$ We accept null hypothesis

ii) If $F_{cal} > F_{tab}$ $\Rightarrow$ We reject null hypothesis.

Q:- In a sample of 8 observations from a normal population, the sum of squares of the sample values from sample mean is 84.4 and in another sample of 10 observations it was 102.6. At 5% LOS test whether the populations have same variance.

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

$$\Sigma(x_i - \bar{x})^2 = 84.4 \quad , \quad \Sigma(y_i - \bar{y})^2 = 102.6$$

Step-1: set up null hypothesis
i.e $H_0 : \sigma_1^2 = \sigma_2^2$ $\rightarrow$ The populations have the same variance

Step-2: set up alternative hypothesis
The two populations are not having same variance
i.e $H_1 : \sigma_1^2 \neq \sigma_2^2$ (2-tailed test)

Step-3:- $\alpha = 0.05$

Step-4:- The test statistic for F-test is,

$$F_{cal} = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$$

We have $s_1^2 = \frac{1}{n_1 - 1} \Sigma(x_i - \bar{x})^2$

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

$= \frac{1}{7}(84.4) = 12.057$

$= \frac{1}{9}(102.6) = 11.4$

ere, s_1^2 is greater than s_2^2

i.e. $F_{cal} = \frac{s_1^2}{s_2^2} = \frac{12.057}{11.4} = 1.057$.

find DOF,

p-5: $V_1 = n_1 - 1 = 8 - 1 = 7$

$V_2 = n_2 - 1 = 10 - 1 = 9$

$\alpha = 0.05$

$F_{tab} = F_{(V_1, V_2, \alpha)} = F_{(7,9,0.05)} = 3.29$

compare F_{cal} and F_{tab} values

. $F_{cal} < F_{tab} \Rightarrow$ We accept null hypothesis

ence, the populations have the same variance

Pumpkins are grown under 2 experimental conditi-
s. Two samples of 11 and 9 show that the
mple S.D as 0.8 and 0.5 respectively. Test
hypothesis that the variances are equal.

$n_1 = 11$, $n_2 = 9$

$s_1 = 0.8$, $s_2 = 0.5$

Step-1: $H_0 : \sigma_1^2 = \sigma_2^2$

step-2: $H_1 : \sigma_1^2 \neq \sigma_2^2$ (2-tailed test)

Step-3: $\alpha = 0.05$

Step-4: $s_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11(0.8)^2}{10} = \frac{7.04}{10} = 0.704$

$s_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{9(0.5)^2}{8} = \frac{2.25}{8} = 0.281$

$s_1^2 > s_2^2$

$$F_{cal} = \frac{s_1^2}{s_2^2} = \frac{0.704}{0.281} = 2.505$$

Step-5: $v_1 = n_1 - 1 = 11 - 1 = 10$

$v_2 = n_2 - 1 = 9 - 1 = 8$

$\alpha = 0.05$

$$F_{tab} = F_{(v_1, v_2, \alpha)} = F_{(10, 8, 0.05)} = 3.35$$

$F_{cal} < F_{tab} \Rightarrow$ We accept null hypothesis

Hence, the variances are equal.

Q:- Time taken by the workers in performing a job. Method-I and II are given below.

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

Test the hypothesis at 5% LOS.

Sol:- given, $n_1 = 6$, $n_2 = 7$

Step-1: set up null hypothesis. $H_0 : \sigma_1^2 = \sigma_2^2$

Step-2: set up alternative hypothesis.

$H_1 : \sigma_1^2 \neq \sigma_2^2$ (2-tailed test).

Step-3: $\alpha = 0.05$

Step-4: $$\bar{x} = \frac{20 + 16 + \cdots + 22}{6} = \frac{134}{6} = 22.33$$

$$\bar{y} = \frac{27 + 33 + \cdots + 38}{7} = \frac{241}{7} = 34.4$$

$$\Sigma(x_i - \bar{x})^2 = (20 - 22.3)^2 + \cdots + (22 - 22.3)^2 = 81.34$$

$(y_i - \bar{y})^2 = (27 - 34.4)^2 + \cdots + (38 - 34.4)^2 = 133.72$

$s_1^2 = \frac{1}{n_1 - 1} \Sigma (x_i - \bar{x})^2 = \frac{1}{5}(81.34) = 16.268$

$s_2^2 = \frac{1}{n_2 - 1} \Sigma (y_i - \bar{y})^2 = \frac{1}{6}(133.72) = 22.286$

$s_2^2 > s_1^2$

$F_{cal} = \frac{s_2^2}{s_1^2} = \frac{22.286}{16.268} = 1.369$

$\nu_1 = n_1 - 1 = 6 - 1 = 5 = \nu_2$

$\nu_2 = n_2 - 1 = 7 - 1 = 6 = \nu_1$

$\alpha = 0.05$

$F_{tab} = F_{(5, 6, 0.05)} = 4.95$

$F_{cal} < F_{tab} \Rightarrow$ We accept null hypothesis.

ence, The time taken by the workers in performing a job is same in 2 methods.

- In 1 sample of 10 observations from a normal population, the sum of squares of deviation of the sample values from sample mean is 102.4 and in another sample 12 observations from another normal population the sum of squares of deviation of the sample mean 120.5. Examine whether the 2 populations have same variance

Sol:- $n_1 = 10$, $n_2 = 12$

$\Sigma (x_i - \bar{x})^2 = 102.4$, $\Sigma (y_i - \bar{y})^2 = 120.5$

Step-1:- $H_0 : \sigma_1^2 = \sigma_2^2$

Step-2: $H_1 : \sigma_1^2 \neq \sigma_2^2$ (2-tailed test)

Step-3:- $\alpha = 0.05$

Step-4: $s_1^2 = \frac{1}{n_1 - 1} \Sigma (x_i - \bar{x})^2 = \frac{1}{9}(102.4) = 11.37$

$s_2^2 = \frac{1}{n_2 - 1} \Sigma (y_i - \bar{y})^2 = \frac{1}{11}(120.5) = 10.95$

$s_1^2 > s_2^2$

$F_{cal} = \frac{s_1^2}{s_2^2} = \frac{11.37}{10.95} = 1.038$

Step-5: $v_1 = n_1 - 1 = 10 - 1 = 9$

$v_2 = n_2 - 1 = 12 - 1 = 11$

$\alpha = 0.05$

$F_{tab} = F_{(9, 11, 0.05)} = 2.90$

$F_{cal} < F_{tab} \Rightarrow$ We accept null hypothesis

Hence, the 2 populations have same variance.

***

Q:- 2 random samples gives the following results.

Sample	size	sample mean	sum of squares of deviation
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population.

Sol:- If the problem is related to samples came from the same normal population, in this case apply both the methods i.e i) F-test, ii) t-test.

given, $n_1 = 10$, $n_2 = 12$

$\bar{x} = 15$, $\bar{y} = 14$

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

test:

Step-1: set up null hypothesis (H_0): 2 samples come from the same normal population.

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

Step-2: set up alternative hypothesis (H_1). The 2 samples haven't come from the same normal population $H_1 : \sigma_1^2 \neq \sigma_2^2$

Step-3:- $\alpha = 0.05$

Step-4:- $F_{cal} = \frac{\text{greater variance}}{\text{smaller variance}}$

$s_1^2 = \frac{1}{n_1 - 1} \Sigma(x_i - \bar{x})^2 = \frac{1}{9}(90) = 10$

$s_2^2 = \frac{1}{n_2 - 1} \Sigma(y_i - \bar{y})^2 = \frac{1}{11}(108) = 9.81$

$\because s_1^2 > s_2^2$

$F_{cal} = \frac{s_1^2}{s_2^2} = \frac{10}{9.81} = 1.01$

Step-5: find DOF

$\nu_1 = n_1 - 1 = 10 - 1 = 9$

$\nu_2 = n_2 - 1 = 12 - 1 = 11$

$\alpha = 0.05$

$F_{tab} = F_{(9, 11, 0.05)} = 2.90$

$F_{cal} < F_{tab} \Rightarrow$ We accept null hypothesis

Hence, 2 samples came from the same normal population.

T-Test:

step-1:- set up null hypothesis (H_0).

$$H_0 : \mu_1 = \mu_2$$

step-2: set up alternative hypothesis (H_1)

$$H_1 : \mu_1 \neq \mu_2 \text{ (2-tailed test)}$$

step-3:- $\alpha = \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

step-4: The test statistic for t-test is

$$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

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

$$s^2 = \frac{1}{20}(90 + 108) = \frac{198}{20} = 9.9$$

$$t_{cal} = \frac{15 - 14}{\sqrt{9.9\left(\frac{1}{10} + \frac{1}{12}\right)}} = \frac{1}{\sqrt{9.9\left(\frac{1}{10} + \frac{1}{12}\right)}} = 0.742$$

step-5:- find DOF

$$v = n_1 + n_2 - 2 = 10 + 12 - 2 = 20$$

$$\alpha = 0.025$$

$$t_{tab} = t_{(v, \alpha)} = t_{(20, 0.025)} = 2.086$$

$t_{cal} < t_{tab} \Rightarrow$ We accept null hypothesis.

$\because$ In both the tests, we accept null hypothesis

$\therefore$ We can conclude that the given samples have been drawn from the same normal population.

Q:- The reporting levels in, mg in 2 samples of tobacco are found to be as follows:

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

est whether the samples have come from the ame normal population.

$\because n_1 = 5, \; n_2 = 6$

$$\bar{x} = \frac{24 + \cdots + 25}{5} = 24.6$$

$$\bar{y} = \frac{27 + 30 + 28 + \cdots + 36}{6} = 29$$

$$\Sigma(x_i - \bar{x})^2 = 21.2$$

$$\Sigma(y_i - \bar{y})^2 = 108$$

When formula is $\frac{s_2^2}{s_1^2}$, v_1 becomes v_2 and v_2 becomes v_1

st:

step-1: $H_0 : \sigma_1^2 = \sigma_2^2$

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

step-3: $\alpha = 0.05$

step-4: $s_1^2 = \frac{1}{n_1 - 1} \Sigma(x_i - \bar{x})^2 = \frac{1}{4}(21.2) = 5.3$

$s_2^2 = \frac{1}{n_2 - 1} \Sigma(y_i - \bar{y})^2 = \frac{1}{5}(108) = 21.6$

$s_2^2 > s_1^2$

$$F_{cal} = \frac{s_2^2}{s_1^2} = \frac{21.6}{5.3} = 4.07$$

step-5: $v_1 = n_1 - 1 = 5 - 1 = 4 = v_2$

$v_2 = n_2 - 1 = 6 - 1 = 5 = v_1$

$\alpha = 0.05$

$F_{tab} = F_{(v_1, v_2, \alpha)} = t_{(4, 5, 0.05)} = 6.26$

$F_{cal} < F_{tab} \Rightarrow$ We accept null hypothesis

Hence, 2 samples have come from same population.

T- test:

step-1: $H_0 : \mu_1 = \mu_2$

step-2: $H_1 : \mu_1 \neq \mu_2$ (2-tailed test)

step-3: $\alpha = \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

step-4:- $t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

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

$$s^2 = \frac{1}{9}(21.2 + 108) = 14.35$$

$$t_{cal} = \left|\frac{24.6 - 29}{\sqrt{14.35\left(\frac{1}{5} + \frac{1}{6}\right)}}\right| = 1.91$$

step-5: $v = n_1 + n_2 - 2 = 5 + 6 - 2 = 11 - 2 = 9$

$$t_{tab} = t_{(v,\alpha)} = t_{(9, 0.025)} = 2.262$$

$t_{cal} < t_{tab}$ → We accept null hypothesis

∴ In both the test, we accept null hypothesis.

∴ We can conclude that the given samples have been drawn from the same normal population.

Chi-square test (χ^2- test) for independence of attributes:

Attribute: It means a quality (or) characteristic

ex: smoking, drinking, honesty etc.

→ An attribute may be marked by its presence (or) absence in a number of a given population.

→ let us consider, 2 attributes A and B. A is

divided into 2 classes and B is divided in to 2 classes. The various frequencies can be expressed in the following table.

Attribute			
A	a	b	a+b
B	c	d	c+d
	a+c	b+d	a+b+c+d = N

→ The above frequencies are known as observed frequencies.

→ for these observed frequencies find the corresponding expected frequencies by using following table.

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

Procedure:

Step-1: set up null hypothesis

Step-2: set up alternative hypothesis

Step-3: Write LOS α value from the given problem if not take $\alpha = 0.05$

Step-4: calculate the test statistic

The test statistic is given by, $\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$

Where, O_i = observed frequency

E_i = Expected frequency

Step-5: find the DOF i.e $\boxed{DOF = (\text{no. of rows} - 1) \times (\text{no. of columns} - 1)}$

corresponding to α value and DOF find the tabulated value from χ^2 distribution table which gives χ^2_{tab} value. compare χ^2_{cal} and χ^2_{tab} values.

i) If $\chi^2_{cal} < \chi^2_{tab}$ ⇒ We accept null hypothesis.

ii) If $\chi^2_{cal} > \chi^2_{tab}$ ⇒ We reject null hypothesis.

*(10M) *Q: On the basis of the information given below about the treatment of 200 patients suffering from a 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
	100	100	200 = N

Sol:- Step-1: set up null hypothesis (H_0).

There is no difference between New and conventional treatment

Step-2: set up alternative hypothesis (H_1)

The new treatment is comparitively superior to the conventional treatment

Step-3:- $\alpha = 0.05$ LOS.

Step-4: Calculate the test statistic i.e

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

The given values in the problem are called observed frequencies represented by O_i.

Corresponding to observed frequencies find the expected frequencies represented by E_i

from the table, N = 200

find the expected frequencies.

The Expected frequencies are,

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

O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
60	45	225	5
30	45	225	5
40	55	225	4.09
70	55	225	4.09

$5 + 5 + 4.09 + 4.09$

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

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

step-5: find DOF i.e (no. of rows − 1)(no. of columns − 1)

$= (2-1)(2-1) = 1 \times 1 = 1$

DOF = 1

$\alpha = 0.05$

$\chi^2_{tab} = \chi^2_{(1, 0.05)} = 3.841$

Compare χ^2_{cal} and χ^2_{tab} values i.e. $\chi^2_{cal} = 18.18$

$\chi^2_{tab} = 3.841$

$\chi^2_{cal} > \chi^2_{tab} \Rightarrow$ We reject null hypothesis.

Hence, the new treatment is comparitively superior to conventional treatment.

Q:- given the following for hair colour and eye colour. Find the value of χ^2. Is there any good association between the two.

		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	N = 150

step-1:- set up null hypothesis (H_0).

There is no significant difference between the two.

step-2: set up alternative hypothesis (H_1).

There is a significant difference between the two

step-3: $\alpha = 0.05$.

step-4: The test statistic for χ^2 test is,

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

The expected frequencies are,

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

O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
15	16	1	0.0625
5	8	9	1.125
20	16	16	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.041
15	12	9	0.75
20	24	16	0.666

$$\chi^2_{cal} = \frac{\Sigma(O_i - E_i)^2}{E_i} = 3.6445$$

step-5: DOF $= (3-1)(3-1) = 2 \times 2 = 4$

$$\chi^2_{tab} = \chi^2_{(4, 0.05)} = 9.488$$

$\chi^2_{cal} < \chi^2_{tab} \Rightarrow$ We accept null hypothesis. There is no significant difference b/w the two.

(HW) (1a) From the following data find whether there is any significant like in the habit of taking soft drinks among the categories of employees. Use χ^2 distribution with 0.05 LOS.

Employees

softdrinks.	clerks	teachers	officers
pepsi	10	25	65
Thums up	15	30	65
fanta	50	60	30

Q:- Two researchers adopted sampling techniques while investigating some group of students to find the no. of students falling in to different intelligent levels. The result is as follows:

Sol:

Researchers	below avg	avg	above avg	Genius	
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	N = 300

step-1:- set up null hypothesis

step-2: set up alternative hypothesis

step-3: $\alpha = 0.05$

Step-4: Expected frequencies,

$\frac{126 \times 200}{300} = 84$	$\frac{93 \times 200}{300} = 62$	$\frac{69 \times 200}{300} = 46$	$\frac{12 \times 200}{300} = 8$
$\frac{126 \times 100}{300} = 42$	$\frac{93 \times 100}{300} = 31$	$\frac{69 \times 100}{300} = 23$	$\frac{12 \times 100}{300} = 4$

O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
86	84	4	0.047
60	62	4	0.064
44	46	4	0.086
10	8	4	0.5
40	42	4	0.095
33	31	4	0.129
25	23	4	0.173
2	4	4	1

$$\chi^2_{cal} = 2.094$$

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 2.094$$

step-5: DOF = (no. of rows − 1)(no. of columns − 1)

$= (2-1)(4-1) = 1 \times 3 = 3$

$\chi^2_{tab} = \chi^2_{(3, 0.05)} = 7.815$

$\chi^2_{cal} < \chi^2_{tab} \Rightarrow$ We ~~reject~~ accept null hypothesis.

Hence, there is no significant difference between the no. of students falling in to different intelligent levels.

1501: (HW)

Employees

softdrinks	clerks	teachers	officers	Total
pepsi	10	25	65	100
thums up	15	30	65	110
fanta	50	60	30	140
total	75	115	160	350

step-1: set up null hypothesis (H_0)

step-2: set up alternative hypothesis (H_1).

step-3: LOS (α) = 0.05.

step-4: Expected frequencies,

$\frac{75 \times 100}{350} = 21.42$	$\frac{115 \times 100}{350} = 32.85$	$\frac{160 \times 100}{350} = 45.71$
$\frac{75 \times 110}{350} = 23.57$	$\frac{115 \times 110}{350} = 36.14$	$\frac{160 \times 110}{350} = 50.28$
$\frac{75 \times 140}{350} = 30$	$\frac{115 \times 140}{350} = 46$	$\frac{160 \times 140}{350} = 64$

O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
10	21.42	130.41	6.088
25	32.85	61.62	1.875
65	45.71	372.10	8.140
15	23.57	73.44	3.116
30	36.14	37.69	1.043
65	50.28	216.67	4.309
50	30	900	30
60	46	196	4.260
30	64	1156	18.062

$$\chi^2_{cal} = 77.163$$

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 77.163$$

<u>step-5:</u> DOF $= (\text{no. of rows} - 1)(\text{no. of columns} - 1)$

$= (3-1)(3-1) = 2 \times 2 = 4$

$$\chi^2_{tab} = \chi^2_{(4, 0.05)} = 9.488$$

$\chi^2_{cal} > \chi^2_{tab} \Rightarrow$ we reject null hypothesis

Hence, there is a significant difference between/among the categories of employees in the habit of taking soft drinks.