Unit - 3

Testing of Hypothesis

Population :- Every statistical investigation aims at collecting information about some collection of individuals or of there attributes in statistical language, such a collection is called population or universe.

A population is finite or infinite according ~~to~~ as the number of element is finite or infinite.

Sample :- A finite subset of population is called a sample and process of selection of such sample is called sampling.

Parameters & statistics :-
Statistical measures calculated on population is called parameters. Statistical measures calculated on sample is called statistics.

Statistical measures :-	Mean	Variance
Representation of Mean & variance of population →	μ	σ^2
Representation of Mean & variance of Sample →	$\bar{x}$	S

Sampling Distribution :-
If a number of sample each of size n are drawn from a population and if for each sample the value of some statistics (i-e mean), a set of statistics will be obtained.

Sampling distribution of a statistics :-

If we draw a sample of size n from a given finite population of size N, then total number of possible sample is

$$^{N}C_{n} = \frac{N!}{n!(N-n)!} = k \text{ (say).}$$

for each of these k samples, we can compute some statistics $t = t(x_1, x_2 - - x_n)$ in particular mean $\bar{x}$ & variance S^2.

Sample No.	Statistics.		
	t	$\bar{x}$	S^2
1	t_1	$\bar{x}_1$	S_1^2
2	t_2	$\bar{x}_2$	S_2^2
3	t_3	$\bar{x}_3$	S_3^2
⋮	⋮	⋮	⋮
n	t_n	$\bar{x}_n$	S_n^2

The set of value of statistic is obtained, one for each sample constitute what is called Sampling distribution of statistics. For example, the value $t_1, t_2, -- t_k$ determine the sampling distribution of statistics t, in other words, Statistics t may be regarded as a random variable which can take the value t_1, t_2 -- t_k and we can compute the various statistical constants like mean, variance, etc.

For example, Mean ~~of~~ & variance of Sampling distribution of statistics t is given by.

$$\bar{t} = \frac{1}{k}(t_1 + t_2 + \cdots + t_k) = \frac{1}{k}\sum_{i=1}^{k} t_i$$

$$Var(t) = \frac{1}{k}\left[(t_1-\bar{t})^2 + (t_2-\bar{t})^2 + \cdots\right] = \frac{1}{k}\sum_{i=1}^{k}(t_i-\bar{t})^2$$

Test of hypothesis & Test of significance :-

When we attempt to make decision about population on the basis of sample information, we have to make assumptions about the value of some parameters of population. Such assumptions may or may not be true are called statistical hypothesis.

Q. Define null & alternative hypothesis.

Ans→ We set up a hypothesis, which assumes that there is no significant difference is called null hypothesis. A hypothesis which is complementary to null hypothesis is called alternative hypothesis. ~~and~~

Null Hypothesis is denoted by H_0 and alternative hypothesis is denoted by H_1.

* A procedure of deciding whether to accept or to reject a null hypothesis is called test of hypothesis.

If θ_0 is parameter of population and θ is corresponding sample statistics, usually there are some difference between θ_0 & θ. Since θ is based on sample observation and is different for different, so, if difference is caused due to sampling fluctuations is called insignificant difference.

The difference that the sample has not be drawn from given population or from biased sample is called significant difference.

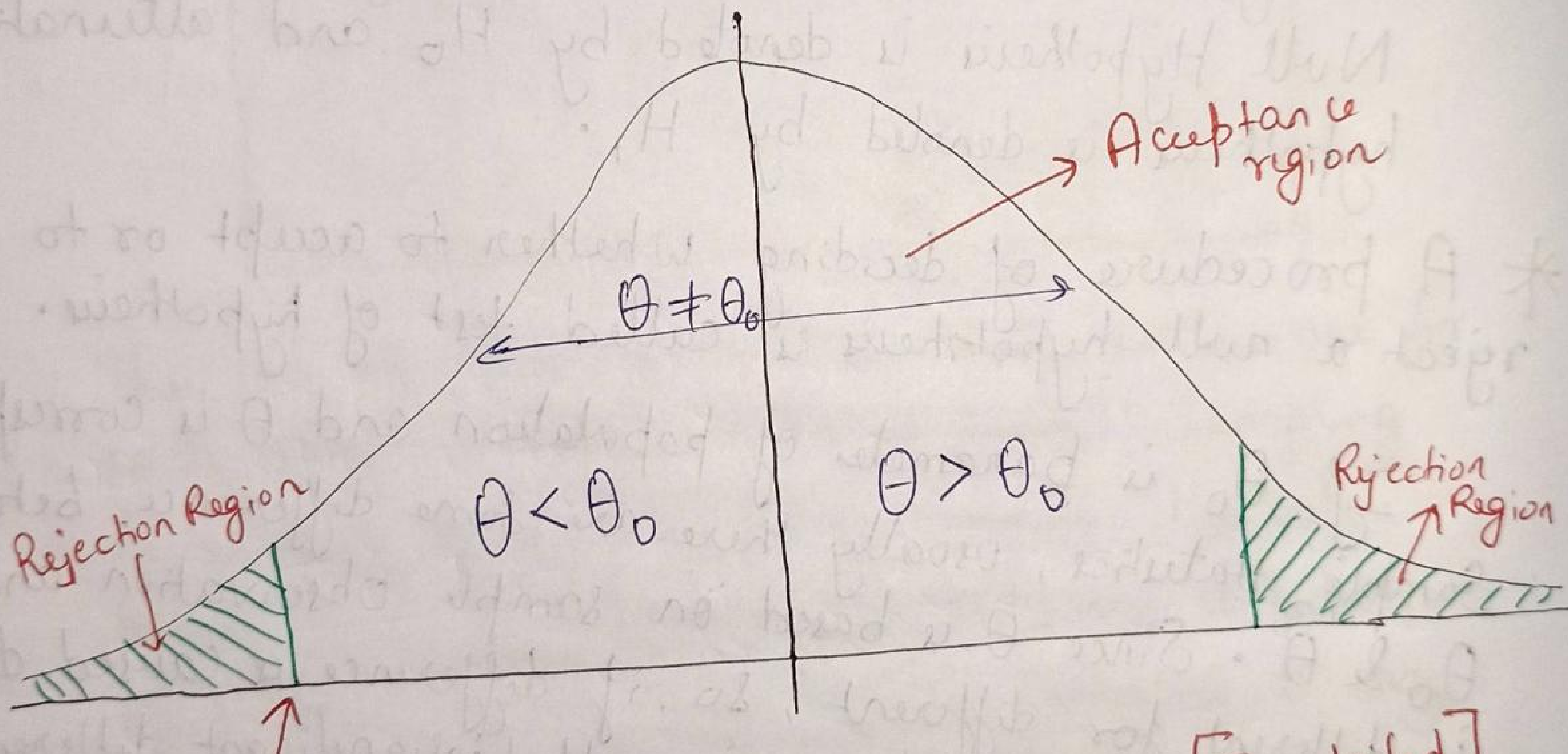
The process of testing whether the difference between θ_0 and θ is significant or not is called test of significant.

* Standard deviation of sampling distribution of a statistic is known as Standard error.

Critical region & Level of Significance (LOS) :-

If we are prepare to accept that the difference between a sample statistics & the corresponding parametric is significant & the sample statistics lie in a certain region, then the region is called critical region or region of rejection.

The region complementary to region of rejection is called region of acceptance. In case of large sample, the sampling distribution of many statistics tends to become normal distribution.



$H_1 :- \theta \neq \theta_0 \rightarrow$ Both side [Two tailed]

$H_1 :- \theta < \theta_0 \rightarrow$ left side [Left tailed]

$H_1 :- \theta > \theta_0 \rightarrow$ Right side [Right tailed]

value at point dividing the region of rejection & region of acceptance is called critical value.

also known as significant value denoted by Z_α

The Critical value for some standard LOS are given as follows :-

Nature of test	1%	5%
Two tailed	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$
Left tailed	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$
Right tailed	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$

*Imp

↓

Learn it Properly

Nature of test	2%	10%
Two tailed	$\lvert Z_\alpha \rvert = 2.33$	$\lvert Z_\alpha \rvert = 1.645$
Left tailed	$Z_\alpha = -2.055$	$Z_\alpha = -1.28$
Right tailed	$Z_\alpha = 2.055$	$Z_\alpha = 1.28$

→ * Not that much Imp.

$\lvert Z_\alpha \rvert = 1.96$ for 5%.

$\Rightarrow P(\lvert Z \rvert < 1.96) = P(-1.96 < Z < 1.96) = 95\%$

Procedure to Solve testing of hypothesis :-

1) Null hypothesis is defined

2) Alternative hypothesis is also defined and also the nature of test (one tail or two tail) is decided.

3) LOS α is fixed or function from the problem of specified Z_α is noted.

4) Test statistic $Z = \dfrac{t - E(t)}{SE(t)}$ is computed.

5) If $\lvert z \rvert < Z_\alpha$ H_0 is accepted, ~~or~~ else H_0 is rejected

(insignificant) (Significant)

Scanned by TapScanner

Q. Define two errors in Testing of hypothesis.

Ans→ Type 1 :- The error committed in rejecting H_0, when it is really true is called type 1 error. This is similar to a good product being rejected by consumer and hence type I error is known as producer's risk.

Type 2 :- The error committed in accepting H_0, when it is false is called type II error. As this error is similar to that of accepting a product of inferior quality, it is called as consumer's risk.

Test of significance of Large Sample :-

If sample size is greater than 30, we usually take sample as large sample.

Test 1 :- Test of significance of the difference between sample proportion and population proportion :-

Test statistic z is given by

$$\boxed{Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}} \quad , \quad Q = 1 - P$$

If $|z| \leq z_\alpha$, the difference between the sample proportion p and population proportion P is not significant at α% LDS.

Scanned by TapScanner

Note :-

1. If P is not known, we assume p is nearly equal to P.

Then, $Z = \dfrac{p-P}{\sqrt{\frac{pq}{n}}}$

2. 95% Confidence limit for P are given by

$$\frac{|P-p|}{\sqrt{\frac{pq}{n}}} \leq 1.96$$

$$\Rightarrow \quad -1.96 < \frac{P-p}{\sqrt{\frac{pq}{n}}} < 1.96$$

$$\Rightarrow \quad -1.96 \times \sqrt{\frac{pq}{n}} < P-p < 1.96\sqrt{\frac{pq}{n}}$$

$$\Rightarrow \quad p-1.96\sqrt{\frac{pq}{n}} < P < \cancel{1.96}\, p+1.96\sqrt{\frac{pq}{n}}$$

* If P is not known, the limit for the proportion in the population are $p \pm z_\alpha \sqrt{\frac{pq}{n}}$. The limit are called confidence limit at α% LOS.

Q. Experience has shown that 20% of manufactured product is of top quality. In one day production of 400 articles, only 50 are of top quality. Show that either the production of the day chosen was not a representative sample or the hypothesis of 20% was wrong based on particular day's production. Find the 95% confidence limit for the percentage of top quality product.

Sol → $P = 20\% = \frac{1}{5}$ [P is population proportion i.e. 20% of product many are good quality]

Let p be proportion of top quality product in the sample.

$\Rightarrow p = \frac{50}{400} = \frac{1}{8}$.

$H_0 : p = P$ (i.e no significant difference)

$H_1 : p \neq P$ (significant difference).

from H_1 we can say two-tailed test will be used.

Let LOS be 5%.

$\Rightarrow Z_\alpha = 1.96$ [from Table].

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{\frac{1}{8} - \frac{1}{5}}{\sqrt{\frac{\frac{1}{5} \times \frac{4}{5}}{400}}}$$

$$= \frac{\frac{1}{8} - \frac{1}{5}}{\sqrt{\frac{1}{5} \times \frac{4}{5} \times \frac{1}{400}}} = -3.75$$

$\Rightarrow |Z| > Z_\alpha$.

$\Rightarrow$ Difference between p & P is significant at 5% LOS and H_0 is rejected.

Hence, the production of choosen day was not a representative sample ~~and~~ or 20% hypothesis was wrong.

Scanned by TapScanner

Now, 95% Confidence Limit is given by.

$$p - 1.96\sqrt{\frac{pq}{n}} \le P \le p + 1.96\sqrt{\frac{pq}{n}}$$

$$\Rightarrow \frac{1}{8} - 1.96\sqrt{\frac{\frac{1}{8}\times\frac{7}{8}}{400}} \le P \le \frac{1}{8} + 1.96\sqrt{\frac{1/8 \times 7/8}{400}}$$

$$\Rightarrow 0.093 \le P \le 0.157 \quad \text{Ans}$$

Q The fatality rat of typhod patient is believed to be 17.26% in a certain year 840 patient suffering from typhoid were treated in a hospital and only 63 patient died. Can you consider the hospital efficient.

Sol → Here, $P = \frac{17.26}{100} = 0.1726$

$Q = 0.8274 \quad [1 - 0.1726 = 0.8274]$

p be sample proportion

$\Rightarrow p = \frac{67}{640} = 0.0984$

$H_0 : p = P$ [i.e. hospital is not efficient]

$H_1 : p < P$. [i.e. hospital is efficient]

$\Rightarrow$ One tailed test to be used in this.

Let LOS be 1%.

$\Rightarrow Z_\alpha = -2.33$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.0984 - 0.1726}{\sqrt{\frac{0.1726 \times 0.8274}{640}}} = -4.96$$

$\Rightarrow |z| > |Z_\alpha|$

$\Rightarrow$ Difference is significant & H_0 is rejected.

So, H_1 is accepted.

$\Rightarrow$ Hospital is efficient.

Test-2 :- Test of significance of the difference between two sample proportion.

Let p_1 & p_2 be proportion of two large sample and n_1 & n_2 be size of those two sample drawn from either same population or from two population of same proportion.

Then,

$$\boxed{Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}} \quad Q = 1 - P.$$

* If P is not known, then unbiased estimat of P based on both sample is taken.

$$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}.$$

If $|z| \le Z_\alpha$ the difference between the two sample is not significant at α% LOS. H_0 is accepted & H_1 is rejected.

Else the difference is significant & H_0 is rejected & H_1 is accepted.

Q. Random Samples of 400 men and 600 women were asked whether they would like to have a School near their ~~residen~~ residence. 200 men & 325 women were in favour of the proposal. Test the hypothesis that the porportion of the men & women in favour of proposal is same at 5% LOS.

Sol → Here, $n_1 = 400$, $n_2 = 600$

$p_1 = \frac{200}{400}$ $\quad p_2 = \frac{325}{600}$

$p_1 = 0.5$ $\quad p_2 = 0.54166$

$H_0 : p_1 = p_2$ [No significant difference b/w the men & women infavour of proposal]

$H_1 : p_1 \neq p_2$ [Significant difference b/w the men & women infavour of proposal]

⇒ We will use two tailed test.

LOS is 5%

$\Rightarrow Z_\alpha = 1.96$

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400 \times 0.5 + 600 \times 0.54166}{400 + 600}$$

$$= 0.525$$

$$Q = 1 - P = 1 - 0.525 = 0.475$$

$$Z = \frac{0.5 - 0.54166}{\sqrt{(0.525)(0.475)\left[\frac{1}{400} + \frac{1}{600}\right]}}$$

$$= -1.2924$$

$\Rightarrow |z| \leq Z_\alpha$

⇒ H_0 is accepted

⇒ There is no difference between men & women infavour of proposal.

Scanned by TapScanner

Q. In a large city A, 20% of a random sample of 900 school boys had a slight physical defect. In another large city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference b/w the proportion significant?

Sol → $n_1 = 900$, $n_2 = 1600$

$p_1 = 20\%$ $\quad p_2 = 18.5\%$

$p_1 = 0.2$ $\quad p_2 = 0.185$

$H_0 : p_1 = p_2$

$H_1 : p_1 \neq p_2$

Two tailed test is to be used.

Let LOS be 5%.

$\Rightarrow Z_\alpha = 1.96$

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{P} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{900 \times 0.2 + 1600 \times 0.185}{900 + 1600}$$

$= 0.1904$

$Q = 1 - 0.1904 = 0.8096$

$$Z = \frac{0.2 - 0.185}{\sqrt{(0.1904)(0.8096)\left(\frac{1}{900} + \frac{1}{1600}\right)}}$$

$= 0.92$

$\Rightarrow |Z| < Z_\alpha \Rightarrow H_0$ is accepted

$\Rightarrow$ There is no significant difference between two sample at 5% LOS.

Test-3 :- Test of significance of the difference between sample mean & population Mean.

Test statistics, $\boxed{Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$

* If $|z| \le Z_{\alpha}$, then the difference between the sample mean $\bar{x}$ & the population μ is not significant at $\alpha\%$ LOS. $\Rightarrow H_0$ is accepted.

Else, the difference is significant $\Rightarrow H_0$ is rejected.

Note :-

1) If σ is not known the sample space can be used in its place as s is nearly equal to σ.

2) 95% Confidence Level for μ is given by.

$$\left| \frac{\mu - \bar{x}}{\frac{\sigma}{\sqrt{n}}} \right| \le 1.96$$

$$\Rightarrow -1.96 \le \frac{\mu - \bar{x}}{\frac{\sigma}{\sqrt{n}}} \le 1.96$$

$$\Rightarrow -1.96 \frac{\sigma}{\sqrt{n}} \le \mu - \bar{x} \le 1.96 \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$$

* If σ is not known then 95% Confidence level

$$\Rightarrow \bar{x} - 1.96 \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + 1.96 \frac{s}{\sqrt{n}}$$

Q. A sample of 100 students is taken from a large population. The mean height of student in sample is 160 cm. Can it be reasonable regards that in the population? The mean height is 165 cm & SD is 10 cm.

Sol → Here, $\bar{x} = 160$, $n = 100$
$\mu = 165$ $\sigma = 10$

$H_0 : \bar{x} = \mu$ (The difference b/w $\bar{x}$ & μ is not significant)

$H_1 : \bar{x} \neq \mu$ (The difference b/w $\bar{x}$ & μ is significant)

⇒ Two tailed test is to be used.

Let LOS be 1%

$\therefore Z_\alpha = 2.58$

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{160 - 165}{\frac{10}{\sqrt{100}}}$$

$$= -5$$

$\Rightarrow |z| > Z_\alpha$.

∴ the difference between $\bar{x}$ & μ is significant at 1% & H_0 is ~~accep~~ rejected.

Q A sample of 100 student gave the mean weights of 58 kg and SD of 4 kg. find 95% Confidence limit of mean of population:

Sol → Here, $n = 100$, $\bar{x} = 58$, $s = 4$.

95% confidence level for μ -

$$\Rightarrow \bar{x} - 1.96 \frac{s}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{s}{\sqrt{n}}$$

$$\Rightarrow 58 - 1.96 \times \frac{4}{\sqrt{100}} < \mu < 58 + 1.96 \times \frac{4}{\sqrt{100}}$$

$$\Rightarrow 58 - 1.96 \times \frac{4}{10} < \mu < 58 + 1.96 \times \frac{4}{10}.$$

Test 4 :- Test the significance of difference between the mean of two sample.

Let $\bar{x}_1$ & $\bar{x}_2$ be mean of two large sample of size n_1 and n_2 which are drawn from 2 population of same mean μ and variances σ_1^2 and σ_2^2 respectively.

Then, Test statistics is given by

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

* If $|z| \leq z_\alpha$, then difference is insignificant & H_0 is accepted.

Else, difference is significant, H_0 is rejected.

Note :-

1) If the sample is drawn from same population i.e $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

2) If σ_1 & σ_2 is not known and $\sigma_1 \neq \sigma_2$,
σ_1 and σ_2 can be approximated by sample standard derivation S_1 and S_2.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

3) If σ_1 & σ_2 is not known & $\sigma_1 = \sigma_2 = \sigma$, then σ_1 & σ_2 is approximated by $\sigma^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$

$$\Rightarrow \quad Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Q. In a random sample of size 500, the mean found to be 20, in another independent sample of size 400, the mean is 15. Could the sample have been drawn from the ~~sample~~ same population with S.D. 4.

Sol→ $\bar{x}_1 = 20$, $n_1 = 500$
$\bar{x}_2 = 15$ $\quad n_2 = 400$

$H_0 : \bar{x}_1 = \bar{x}_2$ [the sample have been drawn from same population]

$H_1 : \bar{x}_1 \neq \bar{u}_2$ [the sample have been not drawn from same population]

⇒ Two tailed test is to be used.

Let LOS be 1%

$\Rightarrow Z_\alpha = 2.58$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{20 - 15}{4\sqrt{\frac{1}{500} + \frac{1}{400}}}$$

$= 18.6$

$\Rightarrow |z| > Z_\alpha$.

$\therefore$ Difference between the $\bar{x}_1$ & $\bar{x}_2$ is significant at 1% LOS.

$\Rightarrow H_0$ is rejected.

$\therefore$ Sample could not have been drawn from same population.

<u>Q</u>. Test that significance of the difference between the mean of sample drawn from two normal population with the same SD.

	Size	Mean	SD
Sample 1 →	100	61	4
Sample 2 →	100	63	6

Sol → Here, $n_1 = 100$ $\quad \bar{x}_1 = 61$

$n_2 = 100$ $\quad \bar{x}_2 = 63$.

$s_1 = 4$ $\quad s_2 = 6$.

$H_0 : \bar{x}_1 = \bar{x}_2$

$H_1 : \bar{x}_1 \neq \bar{x}_2$

Let LOS be 5%

$\Rightarrow Z_\alpha = 1.96$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{61 - 63}{\sqrt{\frac{16}{100} + \frac{36}{100}}}$$

$= -2.77$

$\Rightarrow |z| > Z_\alpha$

$\Rightarrow$ Difference is significant & H_0 is rejected

$\therefore$ Sample is not drawn from same SD populations.

SMALL Sample Test :-

Degree of freedom :-

The number of degree of freedom is given by $\nu = n-k$, where n is number of observations in the sample and k is the number of constraints imposed on them or k is the number of values that have been found out and specified by prior calculations.

Student's t-distribution :-

A random variable T is said to follow Student's t-distribution or sample t-distribution if its pdf is given by

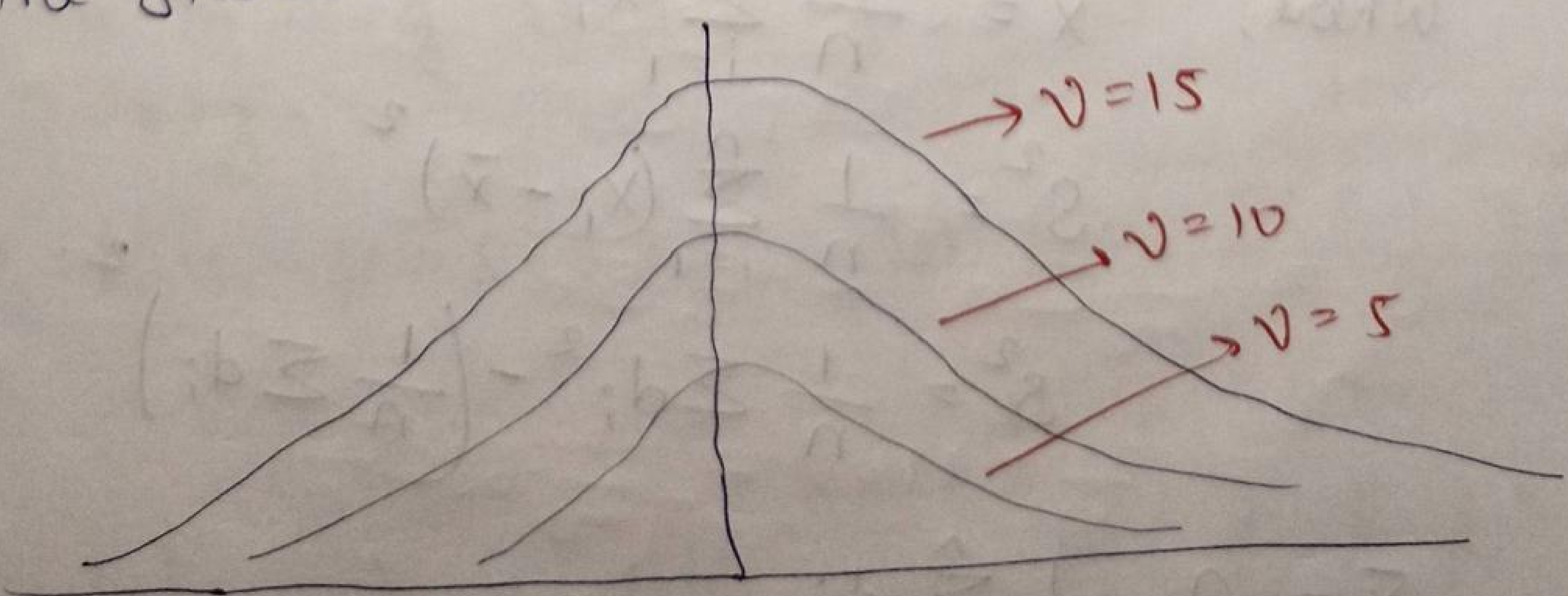
$$f(t) = \frac{1}{\sqrt{\nu}\,\beta\left(\frac{\nu}{2},\frac{1}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

$$-\infty < t < \infty$$

where ν denotes no. of degree of freedom.

Properties :-

1) The probability curve of t-distribution is similar to the standard normal curve.



2) For sufficiently large value of v, the t-distribution tends to the standard normal distribution.

3) The mean of t-distribution is 0.

4) The variance of t-distribution is $\frac{v}{v-2}$, $v \neq 2$. $v > 1$

If $v \to \infty$ then variance $\to 1$.

__Uses :-__

It is used to test significance of difference between

1) the mean of a small sample and the mean of population
2) The mean of two small sample.
3) The coefficient of corellation in the small sample and that in the population assumed zero.

__Test 1 :- Test of significance of the difference between sample mean and population mean.__

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}}$$

where, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

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

$$S^2 = \frac{1}{n} \sum d_i^2 - \left(\frac{1}{n} \sum d_i\right)^2.$$

Note $\rightarrow$ $\bar{x} = A + \frac{1}{n} \sum_{i=1}^{n} d_i$

$d_i = x_i - A$

We will get value of $t_\alpha(v)$ for LOS α and $v = n-1$ degree of freedom from t-table.

& If Calculated $|t| \le t_\alpha(v)$. hypothesis is accepted
Else hypothesis is rejected.

* 95 % confidence ~~int~~ limit for μ is given by.

$$\left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \right| \le t_{0.05}$$

$$\Rightarrow -t_{0.05} \le \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \le t_{0.05}$$

$$\Rightarrow \bar{x} - t_{0.05}\frac{s}{\sqrt{n-1}} \le \mu \le \bar{x} + t_{0.05}\frac{s}{\sqrt{n-1}}$$

where $t_{0.05}$ is 5% critical value of t.

Q. Tests made on the breaking strength of 10 piece of a metal gives following result :- 578, 572, 570, 568, 572, 570, 570, 572, 596, & 584 kg. Test if the mean breaking strength of wire can be assumed as 577 kg.

Sol → Here. $\mu = 577$ kg. Let $A = \frac{568 + 596}{2}$ (min, max)

$A = 582$

Lets make a table with x_i, d_i & d_i^2.

x_i	$d_i = x_i - A$	d_i^2
578	-4	16
572	-10	100
570	-12	144
568	-14	196
572	-10	100
570	-12	144
570	-12	144
592	+10	100
596	14	196
584	2	4
	-68	1144.

$$d_i = x_i - A$$

$$\Rightarrow x_i = d_i + A$$

$$\bar{x} = \frac{1}{n}\Sigma x_i$$

$$= \frac{1}{n}\Sigma d_i + A$$

$$= \frac{1}{n}\Sigma d_i + \frac{1}{n} \times An$$

$$= A + \frac{1}{n}\Sigma d_i$$

$$= 582 + \frac{1}{10}(-68)$$

$$= 582 - 6.8$$

$$= 575.2$$

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

$$= \frac{1}{10} \times 1144 - \left(\frac{1}{10} \cdot 68\right)^2$$

$$= 114.4 - (6.8)^2$$

$$s^2 = 68.16$$

$$s = 8.26$$

Let $H_0 : \bar{x} = \mu$.

$H_1 : \bar{x} \neq \mu$.

$\Rightarrow$ Two tailed test to be followed.

Let LOS be 5%

$$t_{0.05}(\nu = 9) = 2.26.$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{575.2 - 577}{\frac{8.26}{\sqrt{9}}}$$

$$= -0.65$$

$\Rightarrow |t| < t_{0.05}(\nu = 9)$

$\Rightarrow$ The difference is not significant. & H_0 is accepted.

So, the mean breaking strength can be assumed as 577 kg.

Q. The mean life time of a sample of 25 bulbs is found to be 1550 hr with SD of 120 hr. The company manufacturing bulbs claims that the average life of the bulbs is 1600 hr. Is the claim acceptable at 5% LOS.

Sol $\rightarrow$ Here, $\bar{x} = 1550$, $S = 120$

$n = 25$, $\mu = 1600$.

$H_0 : \bar{x} = \mu$ [No. significant difference]

$H_1 : \bar{x} < \mu$ [If mean of claim is more than sample then it is acceptable]

Let LOS be 5%

$t_{0.05}(\nu = 25 - 1) = t_{0.10}(\nu = 24) = 1.71$

for one tailed   for two tailed

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{1550 - 1600}{\frac{120}{\sqrt{24}}}$$

$$= -2.04$$

$\Rightarrow |t| > t_{0.10}(\nu = 24)$

$\Rightarrow$ H_0 is rejected.

$\Rightarrow$ The claim is acceptable at 5% LOS.

Test-2 :- Test of significance of difference between means of two small sample.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2 + n_1 s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\nu = n_1 + n_2 - 2$ ← degree of freedom

Note → 1) If $n_1 = n_2 = n$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n s_1^2 + n s_2^2}{n + n - 2}\right)\left(\frac{1}{n} + \frac{1}{n}\right)}}$$

$$\Rightarrow t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n(s_1^2 + s_2^2)}{2(n-1)} \times \frac{2}{n}}}$$

$$\Rightarrow t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n-1}}} \qquad \nu = 2n - 2$$

Scanned by TapScanner

27 If $n_1 = n_2 = n$ and if the pair of values x_1 and x_2 are associated in some way in this case, we assume that $H_0 : \bar{d} (= \bar{x} - \bar{y}) = 0$

& test of significance ~~bet~~ of difference between $\bar{d}$ and 0

Using the test statistics $t = \dfrac{\bar{d}}{\frac{s}{\sqrt{n-1}}}$ $\nu = n-1$

where, $d_i = x_i - y_i$

& $S = SD$ of $d = \frac{1}{n} \sum_{i=1}^{n} (d_i - \bar{d})^2$

Q. Two independent sample of size 8 and 7 contained following values :-

Sample 1 :- 19 17 15 21 16 18 16 14

Sample 2 :- 15 14 15 19 15 18 16

Is the difference between the sample means is significant.

Sol → Sample 1 :-

Let $A = 18$

x_1	$d_i = x_i - 18$	d_i^2
19	1	1
17	-1	1
15	-3	9
21	3	9
16	-2	4
18	0	0
16	-2	4
14	-4	16
Total →	-8	44

$$\bar{x}_1 = A + \frac{1}{n_1}\sum d_i$$
$$= 18 + \frac{1}{8}(-8)$$
$$= 17$$

$$S_1^2 = \frac{1}{n_1}\sum d_i^2 - \left(\frac{1}{n_1}\sum d_i\right)^2$$
$$= \frac{1}{8} \times 44 - \left[\frac{1}{8} \times (-8)\right]^2$$
$$= 5.5 - 1$$
$$S_1^2 = 4.5$$
$$S_1 = 2.12$$

<u>Sample 2 :-</u>

Let $A = 16$

x_1	$d_i = x_i - 16$	d_i^2
15	−1	1
14	2	4
15	−1	1
19	3	9
15	−1	1
18	2	4
16	0	0
Total →	0	20

$$\bar{x}_2 = A + \frac{1}{n_2}\sum d_i = 16 + 0 = 16$$

$$S_2^2 = \frac{1}{n_2}\sum d_i^2 - \left(\frac{1}{n_2}\sum d_i\right)^2$$
$$= \frac{1}{7} \cancel{\Sigma} \times 20 - 0$$
$$= 2.857$$
$$S_2 = 1.69$$

$H_0 : \bar{x}_1 = \bar{x}_2$

$H_1 : \bar{x}_1 \neq \bar{x}_2$

$\Rightarrow$ Two tailed test is to be used.

Let LOS be 5%

$\Rightarrow t_{0.05} (v = 8+7-2) = 2.16$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\Rightarrow t = \frac{17 - 16}{\sqrt{\frac{8 \times (2.12)^2 + 7 \times (1.69)^2}{8+7-2} \times \left(\frac{1}{8} + \frac{1}{7}\right)}}$$

$$t = 0.93$$

$\Rightarrow |t| < t_{0.05} (v = 13)$

$\Rightarrow H_0$ is accepted

So, the difference between two sample means is not significant.

Q. To verify whether in accounting improved performance similar test was given to 12 participant both before & after the course. Marks are :-

before	44	60	61	52	32	44	70	41	69	72	53	72
After	53	38	69	57	46	39	73	48	73	74	60	78

Sol $\rightarrow$ In this, t-test of paired value is to be used.

Let $d = x_1 - x_2$ where x_1 and x_2 denotes the marks in two tests.

Values of d are :-

-9, 22, -8, -5, -14, 5, -3, -7, -6, -2, -7, -6

$\Sigma d = -9+22-8-5-14+5-3-7-6-2-7-6$

$= -40$

$\Sigma d^2 = 81+484+64+25+196+25+9+49+36+4+49+36$

$= 1058$

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

$= \frac{1}{12} \times -40 = \frac{-40}{12}$

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

$s = \sqrt{\frac{1}{12} \times 1058 - \left(\frac{-40}{12}\right)^2}$

$s = 8.78$

$H_0 : \bar{d} = 0$ [student have not benefited from the course]

$H_1 : \bar{d} < 0$ [student get benefit of course]

Let LOS be 5%.

$t_{0.05}(v=13)$ for 1 tailed test

$= t_{0.10}(v=13)$ for 2 tailed test =

$t = \frac{\bar{d}}{\frac{s}{\sqrt{n-1}}} = \frac{40/12}{\frac{8.78}{\sqrt{12-1}}} = 1.003$

$\Rightarrow |t| < t_{0.10}(\nu = 11)$

$\Rightarrow H_0$ is accepted

$\Rightarrow$ Student have not benifited from course.

Q. The following data represent the biological values of protein from cow's milk and buffalo's milk at certain level.

Cow Milk	1.82	2.02	1.88	1.61	1.81	1.54
Buffalo Milk	2.00	1.83	1.86	2.03	2.19	1.88

Examine if the average value of protein in the two sample significantly differs.

Sol → Here, $n = 6$, $\bar{x}_1 = \frac{1}{6}[1.82 + 2.02 + 1.88 + 1.61 + 1.81 + 1.54]$

$= \frac{1}{6} \times 10.68 = 1.78$

$S_1^2 = \frac{1}{6}\Sigma x_1^2 - \left(\frac{1}{6}\Sigma x_1\right)^2$

$= 0.0261$

$\bar{x}_2 = \frac{1}{6}[2.00 + 1.83 + 1.86 + 20.03 + 2.19 + 1.88$

$= 1.965$

$S_2^2 = \frac{1}{6}\Sigma x_2^2 - \left(\frac{1}{6}\Sigma x_2\right)^2$

$= 0.0154$

Here, n is same. So.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2 + S_2^2}{n_2 - 1}}}$$

$$\Rightarrow t = \frac{1.78 - 1.965}{\sqrt{\frac{0.0261 + 0.0154}{6-1}}}$$

$$t = -2.03$$

Now, $H_0 : \bar{x}_1 = \bar{x}_2$

$H_1 : \bar{x}_1 \neq \bar{x}_2$

Let LOS be 5%.

$$t_{0.05}(\nu = 10) = 2.23$$

$$\Rightarrow |t| < t_{0.05}(\nu = 10)$$

$\Rightarrow$ H_0 is accepted

$\Rightarrow$ The difference between the mean protein value of the two varieties of milk is not significant at 5% Level.