Testing hypothesis

Population:

A collection of individuals is called a population (or) Universe

Sample:

A finite subset of the population is called a sample.

Sampling: The process of selection of such samples is called sampling

Size of the sample:

The number of elements in the sample is called the size of the sample.

Large Sample and Small sample

The no. of elements in a sample is greater than or equal to 30, then the sample is called a large sample and if it is less than 30, then the sample is called a small sample.

Parameters:

Statistical constants like mean μ, variance σ^2 etc., computed from a population are called parameters.

Statistics

Statistical constants like mean $\bar{x}$, variance s^2 etc, computed from a sample are called sample statistics. or Statistics.

A sample statistic is denoted by t.

Sampling distribution

The prob. distribution of a statistic t is called the sampling distribution of t.

For example, If we take k samples each of size $n_1, n_2, \ldots n_k$, we can find their means $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_k$. This set of sample means is called a sampling distribution of the sample statistic $\bar{x}$.

Standard error (S.E)

The S.D of the sampling distribution of a statistic t is called the standard error of t.

Statistical hypothesis:

In making statistical decisions, we make assumptions or gusses about populations involved. Such assumption which may be true or false are called statistical hypothesis.

Null Hypothesis:

A hypothesis which assumes that there is no significant difference between the sample statistic and the corresponding parameters. Such a hypothesis no difference is called null hypothesis and is denoted by H_0.

Alternative hypothesis

A hypothesis which is different from null hypothesis is called an alternative hypothesis

Errors in sampling:

After applying a test of significance a decision is to be taken to accept or reject the null hypothesis H_0. There is always some prob. of committing an error in taking a decision.

These errors are of two types.

(1) Type 1 error 2.) Type 2 error

Type 1 error:

The rejection of the null hypothesis H_0 when it is true is called Type 1 error.

Type 2 error:

The acceptance of the null hypothesis H_0 when it is false is called Type 2 error.

Level of significance:

The prob. of Type 1 error is called the level of significance of the test and is denoted by α.

We usually take either $\alpha = 5\%$ or 1%.

Critical region:

For a test statistic, the area under the prob. curve, which is normal, is divided into two regions namely the region of acceptance of H_0 and the region of rejection of H_0.

The region in which H_0 is rejected is called the critical region.

The region in which H_0 is accepted is called the acceptance region.

One tailed and two tailed test

Set $H_0 : \mu = \mu_0$

Then H_1 is any one of the following

$H_1 : \mu \neq \mu_0$, then the test is called two tailed test

$: \mu > \mu_0$ then the test is called right tailed test

$: \mu < \mu_0$ " " left

Table value of Z

~~Nature~~ Level of signifi	1 %	5 %
Two Tailed	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$
Right "	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$
left "	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$

Procedure for testing of hypothesis

(i) Set H_0

(ii) Decide the alternative hypothesis H_1

(iii) Choose the level of significance α ($\alpha = 5\%$ or 1%)

(iv) ~~find~~ Compute the test statistic

(v) Compare the computed value of $|z|$ with the table value of z_α and decide the acceptance or rejection of H_0

(vi) inference

If $|Z| < Z_\alpha$, H_0 is accepted.

Remark:

(1) 5% of level of significance means we are 95% confident that we have made the right decision.

ie, the hypothesis has a 0.05 prob. of being wrong.

Large Sample test

(1) test for single proportion
(2) test for two proportions
(3) test for single mean
4) " " two means.

Large Sample Space:

Test-I Test of significance of the single proportion

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

p = proportion of sample
P = proportion of Population

n = size, Q = 1-P

Level of significance: 5% or 1%

* At 5% level of significance, 95% confidence limits are

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

* At 1% level of significance, 99% confidence limits are

$$p - 2.58\sqrt{\frac{pq}{n}} \le P \le p + 2.58\sqrt{\frac{pq}{n}}$$

1.) Experience has shown that 20% of a manufactured product is of top quality. In one day's production of 400 articles, only 50 are of top quality. Show that either the production of the day chosen wasn't a represntative sample or the hypothesis of 20% was wrong. Based on the particular day's production, find also the 95% Confidence limits for the percentage of top quality product.

Sol:

P = proportion of top quality products in the Population

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

p = proportion of top quality products in the Sample

$= \frac{50}{400} = \frac{1}{8}$ and $n = 400$

$H_0 : P = \frac{1}{5}$

$H_1 : P \neq \frac{1}{5}$ (Two tailed Test)

Let us assume that LOS (Level of significance) is 5%. $\therefore Z_\alpha = 1.96$

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \quad \text{when} \quad Q = 1 - P$$

$$\Rightarrow z = \frac{1/8 - 1/5}{\sqrt{\frac{1}{400} \cdot \frac{1}{5} \cdot \frac{4}{5}}} = \frac{-3/40}{\sqrt{\frac{1}{2500}}} = \frac{-3/40}{1/50} = -3.75$$

$\Rightarrow |z| = 3.75$

Since $|z| > Z_\alpha$, Ho is rejected at 5% level of significance.

Hence 20% of products manufacture is not top quality.

* 95% confidence limits are $p - 1.96\sqrt{\frac{pq}{n}} \leq P \leq p + 1.96\sqrt{\frac{pq}{n}}$

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

$\Rightarrow 0.093 \le P \le 0.157$

$\therefore$ 95% confidence limits for top quality product are 9.3 and 15.7.

2) A diea is thrown 9000 times and a throw of 3 or 4 is observed 3,240 times. Show that the diea cann't be regarded as an unbiased one.

Sol: [Note: Unbiased: both the sides have the same prob.]

$n = 9000$

P = prob. of getting 3 or 4 [prob. of getting 3 or 4 in the popu.]

$= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

$Q = 1 - P = \frac{2}{3}$

p = proportion of getting 3 or 4 in the sample

$= \frac{3240}{9000} = 0.36$

$H_0: P = \frac{1}{3}$ $H_1: P \neq \frac{1}{3}$ (Two tailed)

LOS = 5%, $Z_\alpha = 1.96$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.36 - 0.333}{\sqrt{\frac{(0.333)(0.667)}{9000}}} = 5.37$$

Since $|Z| > Z_\alpha$, H_0 is rejected.

Hence the die is biased.

3) A Coin is tossed 900 times and head appears 490 times. Does this support the hypothesis that the coin is unbiased?

Sol:

$n = 900$

P = proportion of getting head in the population

$= \frac{1}{2}$

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

p = proportion of getting head in the sample

$= \frac{490}{900} = 0.544$

$H_0 : P = \frac{1}{2}$, $H_1 : P \neq \frac{1}{2}$ (Two tailed)

LOS = 5% , $Z_\alpha = 1.96$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.544 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} = 2.64$$

Since $|Z| > Z_\alpha$, H_0 is rejected.

∴ The data does not support the hypothesis that the coin is unbiased.

4) In a sample of 1000 people in Maharastra, 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: $n = 1000$

P = proportion of rice eaters in the population

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

p = proportion of " " sample

$= \frac{540}{1000} = 0.54$

$H_0 : P = 0.5$ (Both eaters are equally popular in maharastra)

$H_1 : P \neq 0.5$ (two tailed Test)

$LOS = 1\%$, $Z_\alpha = 2.58$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1000}}} = 2.532$$

Since $|Z| < Z_\alpha$, H_0 is accepted.

Hence the both rice and wheat are equally popular.

5.) A salesman in a departmental store claims that at most 60 percent of the shoppers entering the store leaves without making a purchase. A random sample of 50 shoppers showed that 35 of them left without making a purchase. Are these sample results consistent with the claim of the salesman? Use a level of significance of 0.05.

Sol: $n = 50$

P = proportions of shoppers not making a purchase in the population $= 60\% = 0.6$

$Q = 1 - P = 0.4$

p = proportions of " " in the sample

$= \frac{35}{50} = 0.7$

$H_0: p = P$, $H_1: p > P$ (one-tailed test for right)

LOS $= 5\%$, $Z_\alpha = 1.645$

$$z = \frac{0.7 - 0.6}{\sqrt{\frac{(0.6)(0.4)}{50}}} = 1.443$$

since $|z| < Z_\alpha$, H_0 is accepted.

Hence the sample results are consistent with the claim of the salesman.

6.) The fatality (death) rate of typoid patients is believed to be 17.26%. In a certain year 640 patients suffering from Typhoid were treated in metropolitan hospital and only 63 patients died. Can you consider the hospital efficient?

Sol: $n = 640$

P = death rate of typoid patients in the population

$= 17.26\% = 0.1726$

$Q = 1 - P = 0.8274$

p = death rate of the typhoid patients in the sample

$= \frac{63}{640} = 0.0984$

$H_0 : p = P$ (Hospital is not efficient)

$H_1 : p < P$ (Hospital is efficient)

one tailed test for left is to be used.

LOS = 5% , $Z_\alpha = -1.645$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.0984 - 0.1726}{\sqrt{\frac{(0.1726)(0.8274)}{640}}} = -5.013$$

Since $|Z| > |Z_\alpha|$, H_0 is rejected.

Hence the hospital is efficient.

Test : 2 Test of significance of the difference between the two sample proportions

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad \begin{array}{l} n_1 = \text{sample } 1 \\ n_2 = \text{sample } 2 \end{array}$$

p_1 = proportion of sample 1

p_2 = proportion of sample 2

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

If P is not known then $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

1) 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 between the proportion significant?

Sol: Given:

$n_1 = 900, \quad p_1 = 20\%, \quad n_2 = 1600, \quad p_2 = 18.5\%.$

$H_0 : p_1 = p_2, \quad H_1 : p_1 \neq p_2$ (Two tailed Test)

$LOS = 5\%, \quad Z_\alpha = 1.96$

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

To find P:

$$P = \frac{(900 \times 0.2) + (1600 \times 0.185)}{900 + 1600} = 0.1904$$

$$\Rightarrow Q = 1 - P = 0.8096$$

$$\therefore Z = \frac{0.2 - 0.185}{\sqrt{(0.1904 \times 0.8096)\left(\frac{1}{900} + \frac{1}{1600}\right)}} = 0.92$$

Since $|Z| < Z_\alpha$, H_0 is accepted.

Hence there is no significant difference between p_1 and p_2.

2.) 15.5% of a random sample of 1600 undergraduates were smokers, whereas 20% of a random sample

of 900 postgraduates were smokers in a state.
Can we Conclude that less number of undergraduates are Smokers than the postgraduates

Sol:

Given: $n_1 = 1600$ $p_1 = 15.5\% = 0.155$

$n_2 = 900$ $p_2 = 20\% = 0.2$

$H_0: p_1 = p_2$, $H_1: p_1 < p_2$ (one tailed test - left)

LOS = 5% , $z_\alpha = -1.645$

To find P:

$$P = \frac{(0.155 \times 1600) + (0.2 \times 900)}{\left(\frac{1}{1600} + \frac{1}{900}\right)} = 0.1712,$$

$Q = 1 - P = 0.8288$

$$\therefore z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.155 - 0.2}{\sqrt{(0.1712)(0.8288)\left(\frac{1}{1600} + \frac{1}{900}\right)}}$$

$= -2.87$

Since $|z| > |z_\alpha|$, H_0 is rejected.

Hence the less number of undergraduates are Smokers than the postgraduates.

3) In a sample of 400, proportion of tea drinkers is 0.0125 and in another sample of 1200,

Scanned by CamScanner

proportion of tea drinkers is 0.0083. Test whether the samples are taken from a population in which proportion of tea drinkers is 0.01.

Sol:

Given: $n_1 = 400$, $p_1 = 0.0125$, $n_2 = 1200$, $p_2 = 0.0083$

$P = 0.01$. $\therefore Q = 0.99$

$H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$ (Two tailed test)

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.0125 - 0.0083}{\sqrt{(0.01 \times 0.99)\left(\frac{1}{400}+\frac{1}{1200}\right)}} = 0.73$$

Let LOS 5%, $Z_\alpha = 1.96$

Since $|Z| \not> Z_\alpha$, H_0 is accepted.

Hence the sample are taken from a population with proportion 0.01.

4) Random samples of 400 men and 600 women are asked whether they would like to have a fly-over near their residence 200 men and 325 women were in favour of it. Test the equality of proportion of men and women in the proposal.

Sol:

Given: $n_1 = 400$, $n_2 = 600$, $p_1 = \frac{200}{400} = 0.5$, $p_2 = \frac{325}{600} = 0.542$

$H_0: p_1 = p_2$, $H_1: p_1 \neq p_2$ (Two tailed Test)

Let LOS 5% , $Z_\alpha = 1.96$

To find P

$$P = \frac{(400 \times 0.5) + (600 \times 0.542)}{400 + 600} = 0.525$$

$\Rightarrow Q = 1 - P = 0.475$

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.542}{\sqrt{(0.525 \times 0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$$= -1.30$$

Since $|Z| < Z_\alpha$, H_0 is accepted.

Hence the men and women in favour of the proposal are same.

5.) Before increase in excise duty on tea, 800 people out of a sample of 1000 persons were found to be tea drinkers. After an increase in duty, 800 people were tea drinkers out of 1200 persons. Find whether there is significant decrease in the consumption of tea after the increase in duty.

Sol:

Given $n_1 = 1000$, $n_2 = 1200$, $p_1 = \frac{800}{1000} = 4/5$ $p_2 = \frac{800}{1200} = 2/3$

$H_0: p_1 = p_2$, $H_1: p_1 > p_2$ (one tailed test - right)

Let $Los = 1\%$, $Z_\alpha = 2.33$

To find P:

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(1000 \times 0.8) + (1200 \times 0.667)}{1000 + 1200}$$

$$= 0.7273$$

$Q = 1 - P = 0.2727$

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.8 - 0.667}{\sqrt{(0.7273 \times 0.2727)\left(\frac{1}{1000} + \frac{1}{1200}\right)}} = 6.82$$

Since $|Z| > Z_\alpha$, H_0 is rejected.

Hence there is significant decrease in the consumption of tea after the increase in duty.

6.) A machine produced 20 defective articles in a batch of 400. After overhauling it produced 10 defectives in a batch of 300. Has the machine improved?

Sol: Given: $n_1 = 400$, $p_1 = \frac{20}{400}$, $n_2 = 300$, $p_2 = \frac{10}{300}$

$H_0 : p_1 = p_2$, $H_1 : p_1 < p_2$ (one tailed test - left.)

Let $LOS = 5\%$, $Z_\alpha = -1.645$

To find P

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{\left(400 \times \frac{20}{400}\right) + \left(300 \times \frac{10}{300}\right)}{400 + 300}$$

$$= \frac{3}{70} = 0.0429$$

$\therefore Q = 0.9571$

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.05 - 0.0333}{\sqrt{(0.0429 \times 0.9571)\left(\frac{1}{400}+\frac{1}{300}\right)}} = 1.08$$

Since $|Z| < |Z_\alpha|$, H_0 is accepted.

Hence the machine has not improved after overhauling.

7) In a year, there were 956 births in a town A of which 52.5% were males, while in towns A and B combined this proportion in a total of 1406 births was 0.496. Is there any significant difference in the proportion of male births in the two towns.

Sol:

Given $n_1 = 956$, $p_1 = 52.5\%$, $n_1 + n_2 = 1406$, $P = 0.496$

$\Rightarrow n_2 = 1406 - n_1 = 1406 - 956 = 450$, and $Q = 1 - P = 0.504$

Now $\frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = P \Rightarrow \frac{(956 \times 52.5\%) + (450 \times p_2)}{1406} = 0.496$

$\Rightarrow 501.9 + 450 p_2 = 697.376 \Rightarrow p_2 = \frac{195.476}{450} = 0.433$

$H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$

Let LOS $= 5\%$, $Z_\alpha = 1.96$

$$Z = \frac{0.525 - 0.433}{\sqrt{0.496 \times 0.504\left(\frac{1}{956}+\frac{1}{450}\right)}} = 3.40$$

Since $|Z| > Z_\alpha$, H_0 is rejected.

Hence there is significant difference in the proportion of male births in the two towns.

Test-3: Test of significance of difference between sample mean and population mean

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

where σ is S.D of population

If σ is not known, then $z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

where s is S.D of population.

* At 5% level of significance, 95% Confidence limits are $\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$, If σ is known

If σ is not known, then the 95% Confidence limits are $\bar{x} - 1.96\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{s}{\sqrt{n}}$

* At 1% level of significance, 99% Confidence limits are $\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}$, If σ is known.

If σ is not known, then the 99% Confidence limits are $\bar{x} - 2.58\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + 2.58\frac{s}{\sqrt{n}}$

1.) A sample of 900 members has a mean of 3.4 cm and S.D. 2.61 cm. Is the sample from a large population of mean 3.25 cm and S.D. 2.61 cm. Find the 95% Confidence limits

Sol:

Given $n = 900$, $\bar{x} = 3.4$, $s = 3.4$, $\mu = 3.25$, $\sigma = 2.61$

H_0: Assume that the sample has been drawn from the population with mean 3.25 cm. i.e., $\mu = 3.25$

H_1: $\mu \neq 3.25$ (Two tailed Test)

Let LOS = 5% $\quad z_\alpha = 1.96$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{600}}} = 1.724$$

Since $|z| < z_\alpha$, H_0 is accepted.

Hence the sample has been drawn from the population

95% Confidence limits are $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$

$$= 3.4 \pm 1.96\frac{(2.61)}{\sqrt{900}}$$

$$= 3.4 + \frac{(1.96 \times 2.61)}{\sqrt{900}} \text{ and } 3.4 - \frac{(1.96 \times 2.61)}{\sqrt{900}}$$

$$= 3.5 \text{ and } 3.2295$$

2.) A sample of 100 students is taken from a large population. The mean height of the students in this sample is 160 cm. Can it reasonably regarded that in the population, the mean height is 165 cm and S.D is 10 cm. at 1% LOS

Sol:

Given $n = 100$, $\bar{x} = 160$, $\mu = 165$, $\sigma = 10$

$H_0 : \mu = 165$, $H_1 : \mu \neq 165$ (Two tailed Test)

Let LOS = 1%, $Z_\alpha = 2.58$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{160 - 165}{10/\sqrt{100}} = -5$$

Now $|Z| > Z_\alpha$, H_0 is rejected.

$\therefore$ The difference between $\bar{x}$ and μ is significant.

3.) The mean breaking strength of cables is 1800 with a S.D 100. By a new technique in the manufacturing process, it is claimed that breaking strength of cables have increased. In order to test this claim a 50 cables is test and found the mean breaking strength is 1850. Can we support the claim at 1% level of significance.

Sol: Given: $\bar{x} = 1850$, $n = 50$, $\mu = 1800$, $\sigma = 100$

$H_0 : \mu = 1800$, $H_1 : \mu > 1800$ (one-tailed-right)

or $\bar{x} = \mu$: or $\bar{x} > \mu$

Let LOS = 1%, $Z_\alpha = 2.33$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{1850 - 1800}{100/\sqrt{50}} = 3.54$$

Now $|z| > z_\alpha$, $\therefore$ H_0 is rejected.

Hence based on the sample data, we may support the claim of increase in breaking strength.

4.) A sample of 100 people during the past year showed an average life span of 71.8 years. If the standard deviation of the population is 8.9 years, test whether the mean life span today is greater than 70 years.

Sol:

Given $n = 100$, $\bar{x} = 71.8$, $\mu = 70$, $\sigma = 8.9$

$H_0 : \mu = 70$, $H_1 : \mu > 70$ (one-tailed-right)

Let LOS = 5%, $Z_\alpha = 1.645$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{71.8 - 70}{8.9/\sqrt{100}} = 2.02$$

Since $|z| > Z_\alpha$, H_0 is rejected.

$\therefore$ the mean life span today is greater than 70 years

Scanned by CamScanner

5.) The average number of defective articles produced per day in a certain factory is claimed to be less than all the factories. The average of all the factories is 30.5. A random sample of 100 days production showed the mean defective as 28.8 and standard deviation 6.35. Is the average less than 30.5 for all the factories?

Sol:

Given $\mu = 30.5$, $n = 100$, $\bar{x} = 28.8$, $s = 6.35$

$H_0: \mu = 30.5$, $H_1: \mu < 30.5$ (one-tailed test - left)

Let LOS = 5%, $Z_\alpha = -2.33$

If σ is not known, then

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{28.8 - 30.5}{6.35/\sqrt{100}} = -2.68$$

Since $|Z| > |Z_\alpha|$, H_0 is rejected.

Hence the average is less than 30.5 for all factories.

6.) A sample of 900 items has mean 3.4 cms and standard deviation 2.61 cms. Can the sample be regarded as drawn from a population with mean 3.25 cm at 5% level of significance.

Sol: Given $n = 900$, $\bar{x} = 3.4$, $s = 2.61$, $\mu = 3.25$

$H_0 : \mu = 3.25$, $H_1 : \mu \neq 3.25$ (Two tailed test)

$LOS = 5\%$, $Z_\alpha = 1.96$

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{3.4 - 3.25}{2.61/\sqrt{900}} = 1.72$$

Since $|Z| < Z_\alpha$, H_0 is accepted

Hence the sample can be regarded as drawn from a population with mean 3.25 cm.

Test 4: Test of significance of the difference between the means of two samples

1) If $\sigma_1 = \sigma_2 = \sigma$ ie, If the samples are drawn from the sample population then

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

2) If σ_1 and σ_2 are not known and $\sigma_1 \neq \sigma_2$

then

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

3) If σ_1 and σ_2 are equal and not known

then

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}} \quad \text{or} \quad Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\sigma^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$

1) In a survey of buying habits, 400 women shoppers are chosen at random in supermarket A. Their average weekly food expenditure is Rs. 250 with a S.D of R.S. 40. For 400 women shoppers in supermarket B, their average weekly food expenditure is Rs. 220 with a S.D of Rs. 55. Test at 1% level of significance, whether the average weekly food expenditure of the two shopper are equal.

Sol:

Given $n_1 = 400$, $\bar{x}_1 = 250$, $s_1 = 40$

$n_2 = 400$, $\bar{x}_2 = 220$, $s_2 = 55$

$H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 \neq \bar{x}_2$ (Two tailed Test)

Let LOS = 1%, $Z_\alpha = 2.58$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{250 - 220}{\sqrt{\frac{(40)^2}{400} + \frac{(55)^2}{400}}} = 8.82$$

Since $|Z| > Z_\alpha$, Ho is rejected.

Hence the average weekly food expenditure of the two shopper are not equal.

2.) In a random sample of size 500, the mean is 20. In another independent sample of size 400 the mean is 15, Could the samples have been drawn from the same population with S.D. 4.

Sol:

Given $n_1 = 500$, $\bar{x}_1 = 20$, $n_2 = 400$, $\bar{x}_2 = 15$, $\sigma = 4$

$H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 \neq \bar{x}_2$ (Two tailed Test)

Let LOS = 5% , $Z_\alpha = 1.96$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{20-15}{4\sqrt{\frac{1}{500}+\frac{1}{400}}} = 18.6$$

Since $|Z| > Z_\alpha$, H_0 is rejected

Hence the samples Could not have been drawn from the sample population.

3.) A sample of height of 6400 English men has a mean of 170 cm and SD of 6.4 cm while a simple sample of heights of 1600 American has a mean of 17.2 cm and a S.D of 6.3 cm. Do the data indicate that Americans are on average taller than Englishmen.

Sol:

Given $n_1 = 6400$, $\bar{x}_1 = 170$, $s_1 = 6.4$

$n_2 = 1600, \bar{x}_2 = 172, S_2 = 6.3$

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 < x_2$ (one-tailed-left)

LOS = $\cancel{1}$ %, $Z_\alpha = -2.33$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{\overset{170-172}{\cancel{6400-1600}}}{\sqrt{\frac{(6.4)^2}{6400} + \frac{(6.3)^2}{1600}}} = -11.32$$

Since $|Z| > |Z_\alpha|$, H_0 is rejected.

Hence the americans are on the average taller than the Englishman.

4.) A random sample of 100 bulb from a Company A showed a mean life 1300hrs and S.D 82 hours. Another random sample of 100 bulbs from Company B showed a mean life 1248 hrs and S.D 93hrs. Are the bulbs of company A superior to bulbs of Company B at 5% level of significance.

Sol:

Given: $n_1 = 100, \bar{x}_1 = 1300, s_1 = 82$

$n_2 = 100, \bar{x}_2 = 1248, s_2 = 93$

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 > \bar{x}_2$ (one tailed Test-right)

LOS = 5%, $Z_\alpha = 1.645$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{1300 - 1248}{\sqrt{\frac{(82)^2}{100} + \frac{(93)^2}{100}}} = 4.19$$

Since $|Z| > Z_\alpha$, H_0 is rejected

Hence the bulbs of company A is superior to the bulbs of company B.

5 Two random samples of sizes 400 and 500 have mean 10.9 and 11.5 respectively. Can the samples be regarded as drawn from the same population with variance 25 ?

Sol:

Given $n_1 = 400$, $n_2 = 500$, $\bar{x}_1 = 10.9$, $\bar{x}_2 = 11.5$, $\sigma^2 = 25$

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 \neq \bar{x}_2$ (Two tailed test)

Let LOS = 5 %, $Z_\alpha = 1.96$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10.9 - 11.5}{5\sqrt{\frac{1}{400} + \frac{1}{500}}} = -1.78$$

Since $|Z| \not> Z_\alpha$, H_0 is accepted.

Hence the samples can be regarded as drawn from the ~~sample~~ same population with variance 25.

6 The average marks scored by 32 boys is 72 with a S.D of 8 while that for 36 girls is 70

with a S.D of 6. Test at 1% level of significance whether the boys perform better than girls.

Sol:

Given $n_1 = 32$, $\bar{x}_1 = 72$, $s_1 = 8$, $n_2 = 36$, $\bar{x}_2 = 70$, $s_2 = 6$

$H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 > x_2$ (one tailed - Right)

LOS = 1% , $Z_\alpha = 2.33$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{72 - 70}{\sqrt{\frac{8^2}{32} + \frac{6^2}{36}}} = 1.15$$

Since $|Z| < Z_\alpha$, H_0 is accepted.

Hence we cannot conclude that the boys perform better than girls.

7. Test the significance of the difference between the means of the samples, drawn from two normal populations with the same S.D from the following data:

	Size	Mean	S.D
Sample 1	100	61	4
sample 2	200	63	6

Sol:

Given $n_1 = 100$, $\bar{x}_1 = 61$, $s_1 = 4$, $n_2 = 200$, $\bar{x}_2 = 63$, $s_2 = 6$

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 \neq \bar{x}_2$ (Two tailed Test)

LOS = 5%, $Z_\alpha = 1.96$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_2} + \frac{S_2^2}{n_1}}}$$ [$\sigma_1 = \sigma_2$ but σ is not known]

$$= \frac{61-63}{\sqrt{\frac{4^2}{200} + \frac{6^2}{100}}} = -3.02$$

Since $|Z| > Z_\alpha$, H_0 is rejected.

Hence The two normal populations, from which the samples are drawn, may not have the same mean.

8.) The following table gives the data on the hardness of wood stored outside and inside the room.

	outside	inside
Sample size	40	110
Mean	117	132
Sum of squares of the deviation on the mean	8655	27244

Test whether the hardness is affected by weathering.

Sol: Given $n_1 = 40$, $n_2 = 110$, $\bar{x}_1 = 117$, $\bar{x}_2 = 132$,

$$\Sigma(x_1 - \bar{x}_1)^2 = 8655, \quad \Sigma(x_2 - \bar{x}_2)^2 = 27244$$

To find s_1^2 and s_2^2

$$s_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1} = \frac{8655}{40} = 216.38$$ and

$$s_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2} = \frac{27244}{110} = 247.67$$

$H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 \neq \bar{x}_2$ (Two tailed test)

Let LOS = 5%, $Z_\alpha = 1.96$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{117-132}{\sqrt{\frac{216.38}{40} + \frac{247.67}{110}}} = -5.42$$

Since $|z| > Z_\alpha$, H_0 is rejected.

Hence the hardness is affected by weathering.

t-test for single mean or Test the significant difference between sample mean and population mean:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}, \quad \text{degrees of freedom } (\nu) = n-1$$

95% confidence interval of μ is given by (two tailed test)

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

1) Ten oil tins are taken from an automatic filling machine. The mean weight of the tins is 15.8 kg and S.D. 0.50 kg. Does the sample mean differ significantly from the intended weight 16 kg?

Sol:

Given $n=10$, $\bar{x} = 15.8$ kg, $s = 0.5$ kg, $\mu = 16$ kg

$H_0: \mu = \bar{x}$, $H_1: \mu \neq \bar{x}$ (two tailed test)

Let LOS be 5%.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{15.8 - 16}{0.5/3} = -1.2 \quad \text{and} \quad \nu = n-1 = 9$$

From the t-table, for $\nu = 9$, $t_{0.05} = 2.26$

$\therefore |t| < t_{0.05}$

H_0 is accepted at 5% LOS.

Hence the difference between sample mean weight and the intended weight is not significant.

2. The machinist is expected to make engine parts with axle diameter of 1.75 cm. A random sample of 10 parts shows a mean diameter 1.85 cm with S.D. of 0.1 cm. On the basis of this sample, would you say that the work of the machinist is inferior?

Sol:

Given $\bar{x} = 1.85$, $s = 0.1$, $n = 10$, $\mu = 1.75$

$H_0: \bar{x} = \mu$, $H_1: \bar{x} \neq \mu$ (Two tailed test)

Let LOS be 5%.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{1.85 - 1.75}{0.1/\sqrt{9}} = 3 \text{ and } \nu = n-1 = 9$$

From the t-table, for $\nu = 9$, $t_{0.05} = 2.26$

$\therefore |t| > t_{0.05}$

$\therefore H_0$ is rejected.

The work of the machinist can be assumed to be inferior.

3.) The mean lifetime of a sample of 25 bulbs is found as 1550 hrs with a S.D. of 120 hrs. The company manufacturing the bulbs claims that the average life of their bulbs is 1600 hrs. Is the claim acceptable at 5% LOS.

Sol: Given $\bar{x} = 1550$, $s = 120$, $n = 25$, $\mu = 1600$

$H_0 : \bar{x} = \mu$, (The claim is acceptable) $H_1 : \bar{x} < \mu$ (one tailed test - left)

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{1550 - 1600}{120/\sqrt{24}} = -2.04 \quad \text{and} \quad \nu = 24$$

$\left.\begin{array}{l} t_{5\%} \text{ for one tailed test for } \nu = 24 \\ = t_{10\%} \text{ for two tailed test for } \nu = 24 \end{array}\right\} = 1.71$

$\therefore |t| > t_{0.01}$

$\therefore$ Ho is rejected.

ie, The claim of the Company cannot be accepted.

4.) A random sample of 10 boys has the following IQ's 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean IQ of 100?

Sol:

x	$d = x - A$ $= x - 100$	d^2
70	-30	900
120	+20	400
110	10	100
101	1	1
88	-12	144
83	-17	289
95	-5	25
98	-2	4
107	7	49
100	0	0
	-28	1912

$n = 10$

$$\bar{x} = A + \frac{\Sigma d}{n}$$

$$= 100 - \frac{28}{10} = 97.2$$

$$s^2 = \frac{\Sigma d^2}{n} - \bar{d}^2$$

$$= 183.36$$

$H_0: \mu = 100$, $H_1: \mu \neq 100$ (two tailed test)

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{97.2 - 100}{13.54/3} = -0.62$$

and $v = n - 1 = 9$

From t-table, $t_{5\%} = 2.26$

$\therefore$ H_0 is accepted.

ie, The mean IQ of the population can be 100.

Note:

$$\bar{x} = \frac{\Sigma x}{n}, \quad s^2 = \frac{\Sigma (x - \bar{x})^2}{n} \quad \text{or} \quad s^2 = \frac{1}{n}\Sigma x^2 - \bar{x}^2$$

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

Sol:

x	x^2
5	25
2	4
8	64
−1	1
3	9
0	0
6	36
−2	4
1	1
5	25
0	0
4	16
31	185

$n = 12$, $\bar{x} = \frac{\Sigma x}{n} = 2.58$

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

$$= \frac{185}{12} - (31)^2 = 8.76$$

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

$\Rightarrow \mu = 0$ Ia, the injection will not result in increase in B.P

$H_1 : \bar{x} > \mu$ (Right tailed test)

Let LOS be 5%.

Now $t_{5\%}$ for one-tailed test for $\nu = 11$

$= t_{10\%}$ for two-tailed test for $\nu = 11 = 1.80$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{2.58 - 0}{2.96/\sqrt{11}} = 2.89$$

$\therefore |t| > t_{10\%}$

$\therefore H_0$ is rejected. We may conclude that the injection is accompained by an increase in B.P.

6.) The height of ten males of a given locality are found to be 175, 168, 155, 170, 152, 175, 170, 160, 160 and 165 cms. Based on this sample, find the 95% confidence limits for the height of males in that locality.

Sol:

95% Confidence limits for μ are

$$\left(\bar{x} - t_{0.05} \cdot \frac{s}{\sqrt{n-1}}, \quad \bar{x} + t_{0.05} \cdot \frac{s}{\sqrt{n-1}}\right)$$

From the t-table, for $\nu = 9$, $t_{5\%} = 2.26$

$\therefore$ 95% confidence limits are

$$\left(\bar{x} - 2.26 \cdot \frac{s}{\sqrt{n-1}}, \quad \bar{x} + 2.26 \cdot \frac{s}{\sqrt{n-1}}\right)$$

x	$x-\bar{x}$	$(x-\bar{x})^2$
175	+10	100
168	+3	9
155	-10	100
170	5	25
152	-13	169
175	10	100
170	5	25
160	-5	25
160	-5	25
165	0	0
1650		578

Here $\bar{x} = \frac{\Sigma x}{n}$

$= \frac{1650}{10} = 165$

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

$= \frac{578}{10} = 57.8$

$\therefore \left(165 - 2.26\left(\frac{7.6}{3}\right),\ 165 + 2.26\left(\frac{7.6}{3}\right)\right) = (159.3, 170.7)$

Hence the heights of males in the locality are likely to lie within 159.3 cm and 170.7 cm

7.) Prices of shares (Rs) of a company on the different days in a month were found to be 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68. Test whether, the mean price of the shares in the month is 65. Ans: $t = 2.82$, $t_{5\%} = 2.31$
Ho is rejected.

8.) Eight individuals are chosen at random from a population and their heights are found to be in cms 163, 163, 164, 165, 166, 169, 170, 171. In the light of these data discuss the suggestion that the mean height in the universe is 165 cm. Ans: $t = 1.214$, $t_{5\%} = 2.37$
Ho is accepted

Test of significance of the difference between the means of the two samples.

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

Note: 1) If $n_1 = n_2 = n$ and If the samples are independent ie, the observations in the two samples are not at all related, then

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

2) Independent t samples: Compare scores on the same variable but for two different group of cases.

Remark: Assumptions made to test the diff. of means

1) Parent populations from which the samples have been drawn are normally distributed.

2) Variances of the two populations are equal and unknown. ie, $\sigma_1^2 = \sigma_2^2 = \sigma^2$

3) The two samples are random samples and independent.

1) Two independent samples from normal populations with equal variance gave the following

sample	size	mean	S.D
1	16	23.4	2.5
2	12	24.9	2.8

Is the difference between the means significant.

Sol:

Given $n_1 = 16$, $\bar{x}_1 = 23.4$, $s_1 = 2.5$

$n_2 = 12$, $\bar{x}_2 = 24.9$, $s_2 = 2.8$

$H_0: \bar{x}_1 = \bar{x}_2$. $H_1: \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

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

$$= \frac{23.4 - 24.9}{\sqrt{\left(\frac{[16 \times (2.5)^2] + [12 \times (2.8)^2]}{16 + 12 - 2}\right)\left(\frac{1}{16} + \frac{1}{12}\right)}} = -1.44$$

degrees of freedom $v = n_1 + n_2 - 2 = 16 + 12 - 2 = 26$

From t-table, $t_{5\%}(v = 26) = 2.06$

$\therefore |t| < t_{5\%}$. H_0 is accepted.

$\therefore$ the difference between the means is not significant.

2) Samples of two types of electric bulbs were tested for length of life and the following data were obtained.

	Size	Mean	S.D
Sample I	8	1234 hrs	36 hrs
Sample II	7	1036 hrs	40 hrs

Is the difference in the means sufficient to warrant that type I bulbs are superior to type II bulbs?

Sol:

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

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 > \bar{x}_2$ (one tailed test - Right)

Los 5%

$t_{5\%}$ $(\nu = 13)$ for one tailed test $= t_{10\%}$ $(\nu = 13)$ for two tailed test $= 1.77$.

$$t = \frac{1238 - 1036}{\sqrt{\left(\frac{(8 \times 36^2) + (7 \times 40^2)}{8+7-2}\right)\left(\frac{1}{8} + \frac{1}{7}\right)}} = 9.39$$

$\therefore |t| > t_{10\%}$, H_0 is rejected.

Type I bulbs may be regarded superior to type II bulbs.

3.) The mean height and the S.D height of eight randomly chosen soliders are 166.9 cm and 8.29 cm respectively. The corresponding values of six randomly chosen sailors are 170.3 cm and 8.50 cm respectively. Based on this data, can we conclude that soldiers are, in general, shorter than sailors?

Sol:

Given $n_1 = 8$, $\bar{x}_1 = 166.9$, $s_1 = 8.29$

$n_2 = 6$, $\bar{x}_2 = 170.3$, $s_2 = 8.5$

$H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 < \bar{x}_2$ (one tailed test - left)

LOS = 5%.

From t-table, $t_{5\%} (\nu = 12)$ for one tailed test = $t_{10\%} (\nu = 12)$ for two tailed test = 1.78

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -0.695$$

$\because |t| < t_{10\%}$, H_0 is accepted.

Based on the given data, we cannot conclude that soldiers are in general, shorter than sailors.

4) Two independent samples of sizes 8 and 7 contained the following values:

Sample I :	19	17	15	21	16	18	16	14
Sample II :	15	14	15	19	15	18	16	

Is the difference between the sample means significant?

Sol:

Sample I ($n_1 = 8$)

x_1	x_1^2
19	361
17	289
15	225
21	441
16	256
18	324
16	256
14	196
136	2348

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{136}{8} = 17$$

$$s_1^2 = \frac{1}{n_1}\Sigma x_1^2 - \bar{x}_1^2 = \frac{2348}{8} - (17)^2 = 4.5$$

Sample II ($n_2 = 7$)

x_2	x_2^2
15	225
14	196
15	225
19	361
15	225
18	324
16	256
112	1812

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{112}{7} = 16$$

$$s_2^2 = \frac{1}{n_2}\Sigma x_2^2 - \bar{x}_2^2 = \frac{1812}{7} - (16)^2 = 2.857$$

$H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

Let LOS be 5%.

From t-table, $t_{5\%}(\nu = 8+7-2 = 13) = 2.16$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{17-16}{\sqrt{\left(\frac{(8\times 4.5) + (7\times 2.857)}{8+7-2}\right)\left(\frac{1}{8}+\frac{1}{7}\right)}}$$

$= 0.93$

$\therefore |t| < t_{5\%}$, H_0 is accepted.

$\therefore$ The two sample means do not differ significantly.

Scanned by CamScanner

5.) The marks obtained by a group of 9 regular course students and another group of 11 part time course students in a test are given below.

Regular	56	62	63	54	60	51	67	69	58		
Part-time	62	70	71	62	60	56	75	64	72	68	66

Examine whether the marks obtained by regular students and part-time students differ siginificantly at 5% lelvel of significance and 1% of level of significance.

Sol:

Regular

x_1	x_1^2
56	3136
62	3844
63	3969
54	2916
60	3600
51	2601
67	4489
69	4761
58	3364
540	32680

$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{540}{9}$

$= 60$

$s_1^2 = \frac{1}{n_1}\Sigma x_1^2 - \bar{x}_1^2$

$= \frac{32680}{9} - (60)^2$

$= 31.11$

Part time

x_2	x_2^2
62	3844
70	4900
71	5041
62	3844
60	3600
56	3136
75	5625
64	4096
72	5184
68	4624
66	4356
726	48250

$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{726}{11}$

$= 66$

$s_2^2 = \frac{1}{n_2}\Sigma x_2^2 - \bar{x}_2^2$

$= \frac{48250}{11} - (66)^2$

$= 30.364$

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{60-66}{\sqrt{\left(\frac{(9\times 31.11)+(11\times 30.364)}{9+11-2}\right)\left(\frac{1}{9}+\frac{1}{11}\right)}}$$

$= -2.28$

degrees of freedom $(\nu) = n_1+n_2-2 = 9+11-2 = 18$

From t-table, $t_{5\%}\ (\nu=18) = 2.101$

and $t_{1\%}\ (\nu=18) = 2.878$

$\therefore |t| > t_{5\%}$ and $|t| < t_{1\%}$

Hence H_0 is rejected at 5% LOS and H_0 is accepted at 1% LOS.

$\therefore$ the marks obtained by regular students and part-time students differ significantly at 5% LOS. and the difference is not significant at 1% LOS.

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

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

Test whether horse A is running faster than B at 5% Level.

Sol:

Horse A			Horse B		
x_1	x_1^2	$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{219}{7}$	x_2	x_2^2	$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{169}{6}$
28	784	$= 31.29$	29	841	$= 28.17$
30	900	$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \bar{x}_1^2$	30	900	$s_2^2 = \frac{\Sigma x_2^2}{n_2} - \bar{x}_2^2$
32	1024	$= \frac{6883}{7} - (31.29)^2$	30	900	$= \frac{4787}{6} - (28.17)^2$
33	1089	$= 4.22$	24	576	$= 4.28$
33	1089		27	729	
29	841		29	841	
34	1156		169	4787	
219	6883				

$H_0 : \bar{x}_1 \neq \bar{x}_2$, $H_1 : \bar{x}_1 > \bar{x}_2$ (one-tailed test - right)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{31.29 - 28.17}{\sqrt{\left(\frac{(7 \times 4.22) + (6 \times 4.28)}{7+6-2}\right)\left(\frac{1}{7} + \frac{1}{6}\right)}}$$

$= 2.50$

LOS = 5% degrees of freedom $\nu = n_1 + n_2 - 2 = 11$

From t-table, $t_{5\%}$ for one tailed test for $\nu = 11$

$= t_{10\%}$ for two tailed test for $\nu = 11 = 1.796$

$\therefore |t| > t_{10\%}$, H_0 is rejected

Hence A runs faster than B.

7. A group of 10 rats fed on a diet A and another group of 8 rats fed on a different diet B, recorded the following increase in weight. Does it show the superiority of diet A over diet B.

Diet A	5	6	8	1	12	4	3	9	6	10
Diet B	2	3	6	8	1	10	2	8		

Sol:

Diet A

x_1	x_1^2
5	25
6	36
8	64
1	1
12	144
4	16
3	9
9	81
6	36
10	100
64	512

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{64}{10} = 6.4$$

$$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \bar{x}_1^2 = \frac{512}{10} - (6.4)^2 = 10.24$$

Diet B

x_2	x_2^2
2	4
3	9
6	36
8	64
1	1
10	100
2	4
8	64
40	282

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{40}{8} = 5$$

$$s_2^2 = \frac{\Sigma x_2^2}{n_2} - \bar{x}_2^2 = \frac{282}{8} - (5)^2 = 10.25$$

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 > \bar{x}_2$ (one-tailed - Right)

LOS = 5%, $\nu = n_1 + n_2 - 2 = 10 + 8 - 2 = 16$

$$t = \frac{6.4 - 5}{\sqrt{\left(\frac{(10 \times 10.24) + (8 \times 10.25)}{10+8-2}\right)\left(\frac{1}{10} + \frac{1}{8}\right)}} = 0.869$$

From t-table, $t_{5\%}$ for one tailed test for $(\nu=16)$

$= t_{10\%}$ for two tailed test for $\nu=16 = 1.746$

$\therefore |t| < t_{10\%}$, H_0 is accepted.

We cannot say diet A is superior to the diet B.

8) Two different types of drugs A and B were tried on certain patients for increasing weight 5 persons were given drug A and 7 persons were drug B. The increase in weight (in kgs) is given below

Drug A : 3.6 5.5 5.9 4.1 1.4

Drug B : 4.5 3.6 5.5 6.8 2.7 3.6 5.0

Do the two drugs differ significantly with regard to their effect in increasing weight.

Sol:

Drug A

x_1	x_1^2
3.6	12.96
5.5	30.25
5.9	34.81
4.1	16.81
1.4	1.96
20.5	96.79

$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{20.5}{5} = 4.1$

$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \bar{x}_1^2 = \frac{96.79}{5} - (4.1)^2$

$= 2.548$

Drug B

x_2	x_2^2
4.5	20.25
3.6	12.96
5.5	30.25
6.8	46.24
2.7	7.29
3.6	12.96
5.0	25
31.7	154.95

$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{31.7}{7} = 4.529$

$s_2^2 = \frac{154.95}{7} - (4.529)^2$

$= 1.624$

$H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

$$t = \frac{4.1 - 4.529}{\sqrt{\left(\frac{(5 \times 2.548) + (7 \times 1.624)}{5+7-2}\right)\left(\frac{1}{5} + \frac{1}{7}\right)}} = -0.4718$$

Los 5% , $t_{5\%}$ $(\nu = 5+7-2 = 10)$ for two tailed test $= 2.228$

$\therefore |t| < t_{5\%}$, H_0 is accepted

The two drugs donot differ significantly.

9. Two independent samples of size 15 are taken from two normal populations. The sample means are 26 and 10 and variations are 7 and 4 respectively. Is the difference between the means significant.

Sol:

Given $n = 15$, $\bar{x}_1 = 26$, $\bar{x}_2 = 10$, $s_1^2 = 7$, $s_2^2 = 4$

$H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n-1}}} = \frac{26-10}{\sqrt{\frac{7+4}{14}}} = 18.06$$

from t-table, $t_{5\%}$ $(\nu = 2n-2 = 28) = 2.05$

$\therefore |t| > t_{5\%}$, H_0 is rejected.

The difference between mean is significant.

10) The following data give's the biological values of protein from cow's milk and buffalo's milk at a certain level. Examine if the average values of protein in the two samples significantly differ.

Cow's 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

Sol:

Cow milk

x_1	x_1^2
1.82	3.312
2.02	4.080
1.88	3.534
1.61	2.592
1.81	3.276
1.54	2.372
10.68	19.166

$\frac{\Sigma x_1}{n} = \bar{x}_1$

$\bar{x}_1 = \frac{10.68}{6} = 1.78$

$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \bar{x}_1^2$

$= \frac{19.166}{6} - (1.78)^2$

$= 0.0259$

Buffalo milk

x_2	x_2^2
2.00	4
1.83	3.349
1.86	3.460
2.03	4.121
2.19	4.796
1.88	3.534
11.79	23.26

$\bar{x}_2 = \frac{\Sigma x_2}{n}$

$= \frac{11.79}{6} = 1.965$

$s_2^2 = \frac{\Sigma x_2^2}{n_2} - \bar{x}_2^2$

$= \frac{23.26}{6} - (1.965)^2$

$= 0.0154$

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n-1}}} = \frac{1.78 - 1.965}{\sqrt{\frac{0.0259 + 0.0154}{5}}} = -2.03$$

Degrees of freedom $\nu = 2n - 2 = 2(6) - 2 = 10$

From t-table, $t_{5\%}(\nu = 10) = 2.23$.

$\therefore |t| < t_{5\%}$, H_0 is accepted.

The difference between the average values of protein in the two samples is not significant.

11. Two independent groups of 10 children were tested to find how many digits they could repeat from memory after hearing them. The results are as follows

Group A:	8	6	5	7	6	8	7	4	5	6
Group B:	10	6	7	8	6	9	7	6	7	7

Is the difference between mean scores of the two groups significant?

Sol:

Group A

x_1	x_1^2
8	64
6	36
5	25
7	49
6	36
8	64
7	49
4	16
5	25
6	36
62	400

$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{62}{10} = 6.2$

$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \bar{x}^2$

$= \frac{400}{10} - (6.2)^2$

$= 1.56$

Group B

x_2	x_2^2
10	100
6	36
7	49
8	64
6	36
9	81
7	49
6	36
7	49
7	49
73	549

$\bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{73}{10} = 7.3$

$s_2^2 = \frac{\Sigma x_2^2}{n} - \bar{x}_2^2$

$= \frac{549}{10} - (7.3)^2$

$= 1.61$

$H_0: \bar{x}_1 = \bar{x}_2$, $H_1: \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n-1}}} = \frac{6.2 - 7.3}{\sqrt{\frac{1.56 + 1.61}{9}}} = -1.853$$

From t-table, $t_{5\%} (\nu = 2n-2 = 18) = 2.10$

$\because |t| < t_{5\%}$, H_0 is accepted.

The difference between mean scores of the two groups ~~are~~ is not significance.

12.) In a certain experiment t compare two types of pig foods A and B, the following results of increase in weights were observed in pigs

Pig number		1	2	3	4	5	6	7	8
increase in weight	Food A	49	53	51	52	47	50	52	53
	Food B	52	55	52	53	50	54	54	53

Can we conclude that food B is better than food A.

13) Type 3 t-Test for paired observations or Dependent samples:

$$t = \frac{\bar{d}}{s/\sqrt{n-1}}$$ where $d = x_1 - x_2$

degrees of freedom $\nu = n-1$

$H_0 : \bar{d} = 0$, $H_1 : \bar{d} < 0$

Remark: Compare scores on two different variables but for the same group of cases.

1.) The following data relate to the marks obtained by 11 students in two tests, one held at the beginning of a year and the other at the end of the year after intensive coaching. Do the data indicate that the students have benefited by coaching?

Test 1 : 19 23 16 24 17 18 20 18 21 19 20

Test 2 : 17 24 20 24 20 22 20 20 18 22 19

Sol:

$H_0 : \bar{d} = 0$, $H_1 : \bar{d} < 0$ $(\bar{x}_1 < \bar{x}_2)$

Let LOS be 5%. , $\nu = n-1 = 11-1 = 10$

$t_{5\%}(\nu = 10)$ for one tailed test $= t_{10\%}$ for two tailed test

$= 1.81$

$$t = \frac{\bar{d}}{s/\sqrt{n-1}}$$

Scanned by CamScanner

x_1	x_2	$d=x_1-x_2$	d^2
19	17	2	4
23	24	-1	1
16	20	-4	16
24	24	0	0
17	20	-3	9
18	22	-4	16
20	20	0	0
18	20	-2	4
21	18	3	9
19	22	-3	9
20	19	1	1
		$\Sigma d = -11$	$\Sigma d^2 = 69$

$\bar{d} = \frac{\Sigma d}{n} = \frac{-11}{-11} = -1$

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

$= \frac{69}{11} - (1)^2 = 5.27$

$t = \frac{-1}{\frac{2.296}{\sqrt{10}}} = 1.377$

$\therefore |t| < t_{10\%}$

H_0 is accepted

The students have not benefited by coaching

2) IQ tests were administered to 5 persons before and after they were trained. The results are given below

Candidates	1	2	3	4	5
IQ before training	110	120	123	132	125
IQ after training	120	118	125	136	125

~~IQ after training~~ Test whether there is change in after the training.

Sol:

x_1	x_2	$d = x_1 - x_2$	d^2
110	120	-10	100
120	118	2	4
123	125	-2	4
132	136	-4	16
125	121	4	16
		-10	140

$\bar{d} = \frac{\Sigma d}{n} = \frac{-10}{5} = -2$

$s^2 = \frac{140}{5} - (-2)^2 = 24$

$s = 4.899$

$H_0 : \bar{d} = 0, \quad H_1 : \bar{d} < 0$

$t = \frac{\bar{d}}{s/\sqrt{n-1}} = \frac{-2}{4.899/\sqrt{4}} = -0.816$, LOS = 5%.

From the t-table, $t_{5\%}(v=4)$ for one tailed test = $t_{10\%}(v=4)$ for two tailed test $t = 2.132$

$\therefore |t| < t_{10\%}$, H_0 is accepted.

$\therefore$ There is no change in IQ after the training.

3.) A company arranged an intensive training course for its team of salesman. A random sample of 10 salesmen was selected and the values (in 1000) of their sales made in the weeks immediately before and after the course are shown in the following tables:

Salesman	1	2	3	4	5	6	7	8	9	10
Sales before	12	23	5	18	10	21	19	15	8	14
Sales after	18	22	15	21	13	22	17	19	12	16

Test whether there is evidence of an increase in mean sales.

Sol:

x_1	x_2	$d = x_1 - x_2$	d^2
12	18	-6	36
23	22	+1	1
5	15	-10	100
18	21	-3	9
10	13	-3	9
21	22	-1	1
19	17	2	4
15	19	-4	16
8	12	-4	16
14	16	-2	4
		-30	196

$\bar{d} = \frac{\Sigma d}{n} = \frac{-30}{10} = -3$

$s^2 = \frac{196}{10} - (-3)^2 = 10.6$

$s = 3.256$

$H_0 : \bar{d} = 0, \quad H_1 : \bar{d} < 0$

$t = \frac{\bar{d}}{s/\sqrt{n-1}} = \frac{-3}{3.256/3} = -2.76, \quad v = n-1 = 9$

From t-table, $t_{5\%}$ for one tailed test

$= t_{10\%}$ for two tailed test $= 1.833$

$\therefore |t| > t_{10\%}$, H_0 is rejected.

$\therefore$ There is an evidence of increase in sales, after the training programme.

4) The following are the average weekly losses of working hours due to accidents in 10 industrial

Plants before and after an introduction of a safety program was put into operation.

Before	45	73	46	124	33	57	83	34	26	17
After	36	60	44	119	35	51	77	29	24	11

Use 0.05 LOS to test whether the safety program is effective.

Sol:

x_1	x_2	$d = x_1 - x_2$	d^2
45	36	-9	81
73	60	-13	169
46	44	-2	4
124	119	-5	25
33	35	2	4
57	51	-6	36
83	77	-6	36
34	29	-5	25
26	24	-2	4
17	11	-6	36
		-52	420

$\bar{d} = \frac{\Sigma d}{n} = \frac{-52}{10} = -5.2$

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

$= \frac{420}{10} - (-5.2)^2 = 14.96$

$s = 3.868$

$H_0 : \bar{d} = 0, \; H_1 : \bar{d} < 0$

$t = \frac{\bar{d}}{s/\sqrt{n-1}} = \frac{-5.2}{\frac{3.868}{3}} = -4.0331 \quad , \nu = 10 - 1 = 9$

From t-table, $t_{5\%}$ for one tailed test

$= t_{10\%}$ for two tailed test $= 1.83$

$\therefore |t| > t_{10\%}$

H_0 is rejected.

$\therefore$ the safety program is effective.

5.) ~~Even~~ Eleven school boys were given a test in Mathematics. They were given one month tution and the second test was held at the end of it. Do the marks provide evidence that the students were benefited by the extra coaching.

Marks in I test	23	20	19	21	18	20	18	17	23	16	9
marks in II test	24	19	22	18	20	22	20	20	23	20	18

Sol:

x_1	x_2	$d = x_1 - x_2$	d^2
23	24	-1	1
20	19	1	1
19	22	-3	9
21	18	3	9
18	20	-2	4
20	22	-2	4
18	20	-2	4
17	20	-3	9
23	23	0	0
16	20	-4	4
9	18	-1	1
		-12	58

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

$\Rightarrow \bar{d} = \frac{-12}{11} = -1.09$

$s^2 = \frac{\Sigma d^2}{n} - \bar{d}^2$

$= \frac{58}{11} - (-1.09)^2 = 4.085$

$s = 2.02$

$H_0: \bar{d} = 0, \ H_1: \bar{d} < 0$

$t = \frac{\bar{d}}{s/\sqrt{n-1}} = \frac{1.09}{2.02/3} = 1.71$

$v = 11 - 1 = 10$

From t-table, $t_{5\%}(v=10)$ for one tailed test

$= t_{10\%}(v=10)$ for two tailed test $= 1.81$

$\therefore |t| < t_{10\%}$ $\therefore$ H_0 is accepted. $\therefore$ The students are not benefit by extra coaching.