Chapter - 3

TESTING OF HYPOTHESIS

3.0 INTRODUCTION

Statistics is concerned with scientific methods of collection and analysis of data as well as with drawing valid conclusions and making reasonable decisions on the basis of such analysis. For many purposes data is only a sample from the population for which the inference is sought. If the sample is representative of the population, important conclusions about the population can be inferred from the analysis of the sample. Because such inference cannot be absolutely certain, the conclusions are stated in terms of probability.

An important aspect of sampling theory is to study the tests of significance or hypothesis testing which enable us to decide on the basis of sample results if the difference between the observed sample statistic and the hypothetical **population parameter** value is significant or not.

First we shall define certain technical terms which will be used in this section.

3.1 Sample Concepts

Definitions 3.1

1. **Population** : *The totality of any finite or infinite collection of individuals with which we are concerned, possessing a variable character, is called a* **population.**

2. **Sample** : *A* **sample** *is a finite subset of the population. The number of elements in the sample is called the size of the sample.*

3. Large and small samples : *The number 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.**

4. Random sample : *A* **random sample** *of size n is a sample that is chosen in such away that each subset of n elements of the population has the same probability of being selected.*

5. Parameters : *Statistical constants like mean μ, variance σ^2 etc., computed from a population are called* **parameters** *of the population. A population parameter is denoted by θ.*

6. 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.*

Note :

1. Since many random samples are possible from the same population, the sample statistic may vary from sample to sample. Hence a sample statistic is a random variable that depends on the observed sample.

2. Sample statistic t is an estimator of population parameter θ.

Example :

If the population mean μ is unknown and if sample mean $\bar{x}$ is known, we can estimate the value of μ using $\bar{x}$, so $\bar{x}$ is called an estimator of μ.

7. Sampling Distribution : *The probability distribution of a statistic t is called the sampling distribution of t.*

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

Similarly, we can have sampling distribution of sample variance s^2, sample median m etc.

TESTING OF HYPOTHESIS.

Notations.	Population	Sample
Size	N	n
Standard deviation	σ	s
Variance	σ^2	s^2
mean	$\bar{x}$	μ

Large Sample	$n \geq 30$.
Small Sample	$n < 30$.

Test for a Specified Mean:

A random sample of Size 'n' ($n \geq 30$) is drawn from a Population. We want to test that the Population mean has a Specified Value $\boxed{\mu_0}$

PROCEDURE FOR TESTING (Two Tail)

Null Hypothesis $H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

$$Z = \frac{\bar{x} - \mu}{S.E.}$$

$S.E = \frac{\sigma}{\sqrt{n}}$. [S.E = Standard Error]

$$Z = |Z_{cal}|$$

Level of Significance $\alpha = 0.05$ (5%) (or) 1%.

$\alpha = 0.05$	$Z_\alpha = 1.96$
$\alpha = 0.01$	$Z_\alpha = 2.58$

If $|Z_{cal}| > Z_\alpha$ H_0 Rejected.

$|Z_{cal}| < Z_\alpha$ H_0 accepted.

PROCEDURE FOR ONE TAIL TEST

$H_0 : \mu \geq \mu_0$

$H_1 : \mu < \mu_0$ [Left tail]

$H_0 : \mu \leq \mu_0$

$H_1 : \mu > \mu_0$ [Right tail]

Level of Significance $\alpha = 0.05, 0.01,$

$\alpha = 0.05$	$Z_\alpha = 1.645$
$\alpha = 0.01$	$Z_\alpha = 2.33$

If $|Z_{cal}| > Z_\alpha$ H_0 Rejected.

$|Z_{cal}| < Z_\alpha$ H_0 accepted.

Type I and Type II Error.

Decision	H_0 True	H_0 false
Accept H_0	Correct Decision	Type - II - error.
Reject H_0	Type - I - error.	Correct Decision

* **Type - I error :**

  Rejecting H_0 null Hypothesis when it is really true.

  Type - II error :

  Accepting H_0 null Hypothesis when it is false.

STEPS

Step 1 : Identify the test is about Large Sample (or) Small Sample

Step 2 : Identify Two tailed (or) One tailed.

Step 3 : Fix the null and Alternate Hypothesis H_0 and H_1.

Step 4 : Level of Significance and degrees of freedom

Step 5 : Write the test Statistic and find the Parameters [$|Z_{calc}|$ value].

Step 6 : Compare with table value Z_α

Step 7 : Write Conclusion.

LARGE SAMPLE TEST – NORMAL DISTRIBUTIONS. (Z-Test)

1. Single mean	$Z = \dfrac{\bar{x} - \mu}{S.E}$	$S.E = \sigma/\sqrt{n}$.
2. Test for two Sample means	$Z = \dfrac{\bar{x}_1 - \bar{x}_2}{S.E}$	$S.E = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
	If σ_1 & σ_2 are not known. $\sigma_1 \cong s_1$, $\sigma_2 \cong s_2$	$S.E = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

If $\sigma_1 = \sigma_2 = \sigma$ and, σ_1 and σ_2 are not known

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

$$S.E = \sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

1. Testing for single Proportion	$Z = \dfrac{p - P}{S.E}$	$S.E = \sqrt{\dfrac{PQ}{n}}$
2. Test for two Sample Proportions	$Z = \dfrac{p_1 - p_2}{S.E}$	$S.E = \sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$

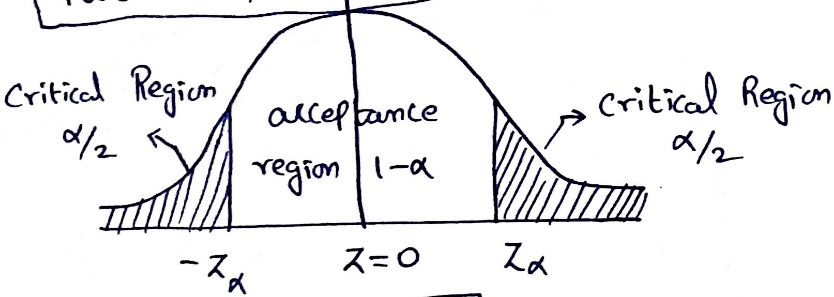
If P ~~is~~ not Known $\quad P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$.

level of Significance	5%	1%
Two tail test	$\lvert z\rvert_\alpha = 1.96$	$\lvert z\rvert_\alpha = 2.58$
Right tail	$z_\alpha = 1.645$	$z_\alpha = 2.33$
left tail	$z_\alpha = -1.645$	$z_\alpha = -2.33.$

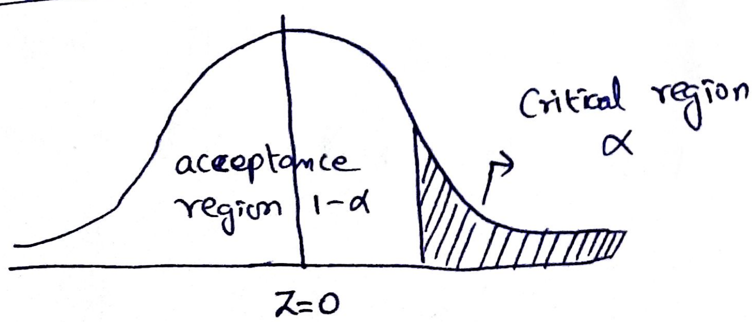
* Even if it left tail always go $\lvert Z_{cal}\rvert$ and Compare with $\lvert Z_\alpha\rvert$ values.

$\lvert Z_{calc}\rvert > \lvert Z_\alpha\rvert$	Reject H_0.
$\lvert Z_{calc}\rvert < \lvert Z_\alpha\rvert$	Accept H_0.

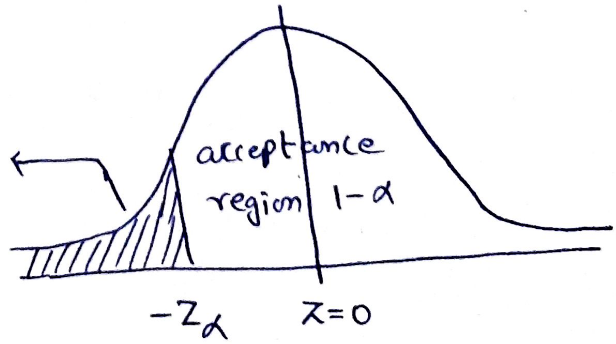
Two Tail Test.



Right Tail Test



Left Tail Test



EXACT SAMPLING DISTRIBUTIONS

(t, F, χ^2)

1) Student's t - distribution.

2) Snedecor's F - distribution.

3) χ^2 - (Chi- Square) distribution.

* When the Sample Size $n \leq 30$ Apply the above (distributions) Test.

Some Formula to remember

$$\bar{x} = A + \frac{\sum d}{n}$$, where take $\boxed{d = x - A}$

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

[* Choose 'A' Such that d Should be in Single digit.
It will make Calculation Part Very Easy]

$\bar{x}$ - Sample mean.

s^2 - Sample Variance.

* For each Test the Corresponding Tables are available to Calculate the Table Values.

SMALL SAMPLES TEST.

1. t- test for Single mean

$$t = \frac{\bar{x} - \mu}{S.E}$$

$$S.E = \frac{\sigma}{\sqrt{n-1}}$$

$$df = n-1$$

2. t - test for two Sample means

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S.E}$$

$$S.E = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$df = n_1 + n_2 - 2.$$

d.f = degree of freedom.

If $n_1 = n_2 = n$

$$t = \frac{\bar{x} - \mu}{S.E}$$

$$S.E = \sqrt{\frac{S_1^2 + S_2^2}{n-1}}$$

$$df = 2n-2.$$

3. F - test for Variance

$$F = \frac{S_1^2}{S_2^2} = \frac{\sigma_1^2}{\sigma_2^2}$$

$$\sigma_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = S_1^2$$

Either Capital S_1 (or) use σ_1 with $\sigma_1 > \sigma_2$ (or) $S_1 > S_2$

$$\sigma_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = S_2^2$$

$$\boxed{df_1 = n_1 - 1}$$

$$\boxed{df_2 = n_2 - 2.}$$

* The greater value should be always in the numbrator. Since [F>1.] table value & Calc value

If $\sigma_2 > \sigma_1$ then take $F = \frac{\sigma_2^2}{S_1^2}$

χ^2 - Test of goodness of fit.

Let $O_1, O_2, O_3, \ldots, O_n$ be the observed frequencies and $E_1, E_2, E_3, \ldots, E_n$ be the corresponding expected frequencies. We want to test the significance of the difference between observed and expected frequencies.

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right] \simeq \sum \frac{(O-E)^2}{E} \quad \text{[Simply].}$$

With degree of freedom $\boxed{df = n-1}$

* Procedure:

Make Table, $O, E, (O-E), (O-E)^2, \frac{(O-E)^2}{E}$.

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$

Example 1 : A sample of 400 students have a mean height 171.38 cms. Can it be reasonably regarded as a sample from a large population with mean height 171.17 cms and standard deviation 3.30 cms ?

Solution : Given, sample size $n = 400$

sample mean $\bar{x} = 171.38$ cms

population mean $\mu = 171.17$ cms

S.D of population $\sigma = 3.3$ cm

Here σ is known and sample is large we want to test the difference between sample mean and population mean.

$$H_o : \mu = 171.17 \text{ cm}$$

$$H_1 : \mu \neq 171.17 \quad (\therefore \text{ Two tailed test})$$

Under H_o, the test statistic is

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

$$= \frac{171.38 - 171.17}{\frac{3.30}{\sqrt{400}}}$$

$$= \frac{0.21 \times 20}{3.30} = \frac{0.42}{3.30} = 1.27$$

$$|Z| = 1.27$$

The table value of Z at 5% level is 1.96.

Inference :

Since $|Z| < 1.96$, H_o is accepted at 5% level of significance.

$\therefore$ The given sample can be regarded as a sample from the population with mean 171.17 cms.

Example 2 : 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. **[AU 2009]**

Solution : Given sample size $n = 100$

sample mean $\bar{x} = 71.8$ years

population mean $\mu = 70$ years

population s.d $\sigma = 8.9$

We want to test whether $\mu > 70$ years

$$H_o : \mu = 70$$

$$H_1 : \mu > 70$$

$\therefore$ Right tailed test.

Under H_o, the test statistic is

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

$$= \frac{71.8 - 70}{\frac{8.9}{\sqrt{100}}} = \frac{1.8 \times 10}{8.9} = 2.02$$

$$\therefore Z = 2.02$$

The table value of Z at 5% level is 1.645

Inference :

Since $Z > 1.645$, H_o is rejected at 5% level of significance.

$\therefore \mu > 70$ is acceptable.

ie. the mean life span today is greater than 70 years.

Example 3 : 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?

Solution : Given, $n = 100$

sample mean $\bar{x} = 28.8$

sample s.d. $s = 6.35$

population mean $\mu = 30.5$

σ is not known.

We want to test that the average article produced is less than 30.5

$H_o : \mu = 30.5$

$H_1 : \mu < 30.5$

$\therefore$ left tailed test

Under H_o, the test statistic is

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

$$= \frac{28.8 - 30.5}{\frac{6.35}{\sqrt{100}}} = \frac{-1.7 \times 10}{6.35} = -2.677 = -2.68$$

The table value of Z for left tailed test at 5% level is -1.645 and at 1% level is -2.33

Inference : Size $Z < -1.645$ and $Z < -2.33$, H_o is rejected at 5% and 1% level of significance.

$\therefore \mu < 30.5$ for all factories

$\therefore$ the claim is acceptable.

Example 4 : 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 cms at 5% level of significance.

Solution : Given, $n = 900$ **[AU 2008, 2010, 2016]**

$$\bar{x} = 34$$

$$s = 2.61$$

$$\mu = 3.25$$

σ is not known.

We have to test whether the sample be drawn from population with mean 3.25 cm.

$$H_o : \mu = 3.25$$

$$H_1 : \mu \neq 3.25$$

$\therefore$ two tailed test

under H_o, the test statistic is

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

$$= \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}}$$

$$= \frac{0.15 \times 30}{2.61} = 1.72$$

$$|Z| = 1.72$$

At 5% level, table value of Z is 1.96

Inference :

Since $|Z| < 1.96$, H_o is accepted at 5% level of significance.

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

Example 5 : The wages of a factory's workers are assumed to be normally distributed with mean μ and variance Rs. 25. A random sample of 36 workers gives the total wages equal to Rs. 1800. Test the hypothesis $\mu = 52$, against the alternative $\mu = 49$ at 1% level of significance.

Solution : Given, $n = 36$

Total wage of 36 workers = 1800

$$\therefore \text{Average wage } \bar{x} = \frac{1800}{36}$$

$$= 50$$

Mean of the population $\mu = 52$

population variance $\sigma^2 = 25$

$$\Rightarrow \sigma = 5$$

We want to test that average wage in the population $\mu = 52$

$$H_o : \mu = 52$$

$$H_1 : \mu = 49 < 52$$

$\therefore$ left tailed test

Under, H_o, the test statistic is

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

$$= \frac{50 - 52}{\frac{5}{\sqrt{36}}} = \frac{-2 \times 6}{5} = -2.4$$

$\therefore$ $Z = -2.4$

At 1% level, for left tailed test, the table value of $Z = -2.33$

Inference : the calculated value of $Z <$ the table value.

ie. $Z < -2.33$

Hence, H_o is rejected at 1% level of significance.

$\therefore$ the average $\mu = 49$ is accepted

3.5 Test for the significant difference between two means or Test for the equality of two means

Procedure :

Let two independent large samples of sizes n_1, n_2 be drawn from two normal populations N (μ_1, σ_1^2) and N (μ_2, σ_2^2). $\bar{x}_1$, $\bar{x}_2$ be the means of the two samples respectively. We want to test whether the means are equal.

Here $t = \bar{x}_1 - \bar{x}_2$. $E(t) = E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2)$

$$= \mu_1 - \mu_2$$

We want $E(t) = 0$

$\therefore$ assume $\mu_1 = \mu_2$

$$H_o : \mu_1 = \mu_2$$

$$H_2 : \mu_1 \neq \mu_2$$

ie. two tailed test

Case (i)

σ_1 and σ_2 are known.

$$\text{S.E of } \bar{x}_1 = \frac{\sigma_1}{\sqrt{n_1}}$$

$$\text{S.E of } \bar{x}_2 = \frac{\sigma_2}{\sqrt{n_2}}$$

$$\therefore \quad \text{S.E of } (\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\therefore \quad \text{Under } H_o \text{ the test statistic is } Z = \frac{\bar{x}_1 - \bar{x}_2}{S.E \text{ of } (\bar{x}_1 - \bar{x}_2)}$$

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

Inference

If $|Z| < 1.96$

$\therefore$ H_o is accepted at 5% level of significance.

If $|Z| > 1.96$, H_o is rejected at 5% level of significance.

If $|Z| < 2.58$, H_o is accepted at 1% level of significance.

$|Z| > 2.58$, H_o is rejected at 1% level of significance

Case 2 :

If σ_1 and σ_2 are not known.

Since n_1, n_2 are large, the best estimate of σ_1 is s_1 and the best estimate of σ_2 is s_2, where s_1 , s_2 are the standard deviation of the two samples.

$$\therefore \quad \text{S.E of } \bar{x}_1 = \frac{s_1}{\sqrt{n_1}}$$

$$\text{S.E of } \bar{x}_2 = \frac{s_2}{\sqrt{n_2}}$$

$$\therefore \quad \text{S.E of } (\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\therefore \text{ the test statistic is } Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

Inference as in case (i)

Case 3 :

If the two samples are drawn from same population or from two populations with same variance, then $\sigma_1^2 = \sigma_2^2 = \sigma^2$, say

$$\Rightarrow \sigma_1 = \sigma_2 = \sigma$$

Assume that σ is given

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

$$\Rightarrow Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

Inference : As in case (i)

Note : If $\mu_1 - \mu_2 = d$, then replace $\bar{x}_1 - \bar{x}_2$ by $\bar{x}_1 - \bar{x}_2 - d$

$$\therefore \quad Z = \frac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

WORKED EXAMPLES

Example 6 : The mean height of two samples of 1000 and 2000 members are respectively 67.5 and 68 inches. Can they be regarded as drawn from the same population with standard deviation 2.5 inches.

[B.D.U-95, AU 2012]

Solution : Given : **Sample I** **Sample II**

Size : $n_1 = 1000$ $\quad n_2 = 2000$

Mean : $\bar{x}_1 = 67.5''$ $\quad \bar{x}_2 = 68''$

Population standard deviation $\sigma = 2.5$

We want to test the significant difference between the two sample means

$$H_o : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

$\therefore$ Two tailed test and σ is known.

Under H_o, the test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

$$= \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$= \frac{-0.5 \times 10}{2.5\sqrt{\frac{1}{10} + \frac{1}{20}}}$$

$$= \frac{-5 \times \sqrt{20}}{2.5\sqrt{3}} = \frac{-22.36}{4.33} = -5.16$$

$$|Z| = 5.16$$

The table value of Z at 5% level is 1.96

Inference : Since $|Z| > 1.96$ and $|Z| > 2.58$, H_0 is rejected at 5% level as well as 1% level of significance.

The difference between means is highly significant. Hence the two samples cannot be regarded as drawn from the same population with $\sigma = 2.5$ inches.

Example 7 : 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 ?

Solution : Given : **Sample I** **Sample II**

Size : $n_1 = 400$ $n_2 = 500$

Mean : $\bar{x}_1 = 10.9$ $\bar{x}_2 = 11.5$

Population variance $\sigma^2 = 25$

σ is known $\therefore \sigma = 5$

We want to test that the two samples are drawn from a population with variance 25

$H_o : \mu_1 = \mu_2$ (No significant difference between the two population means)

$H_1 : \mu_1 \neq \mu_2$

∴ Two tailed test

Under H_o, the test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

$$= \frac{10.9 - 11.5}{5\sqrt{\frac{1}{400} + \frac{1}{500}}}$$

$$= \frac{-0.6 \times 10}{5\sqrt{\frac{1}{4} + \frac{1}{5}}}$$

$$= \frac{-0.6 \times 2 \times 10}{\sqrt{5 \times 9}} = \frac{-12}{\sqrt{45}}$$

$$= -1.78$$

∴ $|Z| = 1.78$

The table value of $|Z|$ at 5% level is 1.96

Inference

Since $|Z| < 1.96$,

H_o is accepted at 5% level of significance.

ie. the samples can be regarded as drawn from the same population with variance 25.

Example 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. **[M.U.96]**

Solution : Sample I

Size : $n_1 = 40$

Mean : $\bar{x}_1 = 117$

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

Sample II

Size $n_2 = 100$

Mean $\bar{x}_2 = 132$

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

$$= \frac{27244}{110} = 247.67$$

We want to test whether the hardness is affected or not.

$H_o : \mu_1 = \mu_2$ (hardness is not affected by weather)

$H_1 : \mu_1 \neq \mu_2$

$\therefore$ Two tailed test.

σ is unknown.

Under H_o, the test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

$$= \frac{117 - 132}{\sqrt{\frac{216.38}{40} + \frac{247.67}{110}}}$$

$$= \frac{-15}{\sqrt{5.4095 + 2.25}} = \frac{-15}{2.7676} = -5.42$$

$\therefore |Z| = 5.42$

The table value of $|Z|$ at 5% level in 1.96 and at 1% level is 2.58

Inference : $|Z| > 1.96$ and $|Z| > 2.58$

$\therefore H_o$ is rejected at 5% level and 1% level of significance.

$\therefore$ The hardness is affected by weathering.

Example 9 : A company produces two makes of bulbs A and B. 200 bulbs of each make were tested and it was found that the make A had mean life of 2560 hours and S.D 90 hours, whereas make B had 2650 hours mean life and S.D 75 hours. Is there a significant difference between the mean life of two makes ?

Solution : **Sample I** **Sample II**

Sample I	Sample II
$n_1 = 200$	$n_2 = 200$
$\bar{x}_1 = 2560$ hrs	$\bar{x}_2 = 2650$ hrs
$s_1 = 90$ hrs	$s_2 = 75$ hrs

σ is unknown.

We have to test whether mean life of the two makes are equal or not.

H_0 :Let us assume that there is no significant difference between the mean life of two makes.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$\therefore$ Two tailed test.

Under H_0, the test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

$$= \frac{2560 - 2650}{\sqrt{\frac{90^2}{200} + \frac{75^2}{200}}} = \frac{-90 \times 10\sqrt{2}}{\sqrt{8100 + 5625}} = \frac{-900 \times 1.414}{\sqrt{13725}}$$

$$= \frac{-1272.6}{117.1537} = -10.86$$

$$|Z| = 10.86$$

The table value of $|Z|$ at 5% level 1.96 and at 1% level 2.58

Inference

Since $|Z| > 1.96$ as well as $|Z| > 2.58$, H_0 is rejected at 5% level as well as 1% level of significance.

$\therefore$ difference between the mean life of the 2 makes is highly significant.

Example 10 : Intelligence test on two groups of boys and girls gave the following results.

	Mean	S.D	Sample size
Girls	75	15	150
Boys	70	20	250

Is there a significant difference in the mean scores obtained by boys and girls ?

Solution : **Sample I** **Sample II**

Girls Boys

$n_1 = 150$ $n_2 = 250$

$\bar{x}_1 = 75$ $\bar{x}_2 = 70$

$s_1 = 15$ $s_2 = 20$

σ is not known.

We want to test whether the mean scores obtained by boys an girls equivalent.

H_o : Let us assume that there is no significant difference between the mean scores obtained by boys and girls.

$H_o : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$\therefore$ Two tailed test

Under H_o, the test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

$$= \frac{75 - 70}{\sqrt{\frac{152}{150} + \frac{20^2}{250}}}$$

$$= \frac{5}{\sqrt{\frac{225}{150} + \frac{400}{250}}}$$

$$= \frac{5}{\sqrt{1.5 + 1.6}}$$

$$= \frac{5}{1.76} = 2.84$$

$$|Z| = 2.84$$

The table value of $|Z|$ at 5% level is 1.96 and at 1% level is 2.58.

Inference :

$|Z| > 1.96$ and $|Z| > 2.58$

$\therefore H_0$ is rejected at 5% level as well as 1% level of significance.

ie. the difference is highly significant.

Example 11 : Two types of new cars produced in our country are tested for petrol mileage. One group consisting of 36 cars averaged 14 km per litre. While the other group consisting of 72 cars averaged 12.5 km per litre.

(a) What test statistic is appropriate, if $\sigma_1^2 = 1.5$, $\sigma_2^2 = 2.0$?

(b) Test, whether there exists a significant difference in the petrol consumption of these two type cars.

Solution :

Sample I

$n_1 = 36$

$\bar{x}_1 = 14$ km per litre

I population variance $\sigma_1^2 = 1.5$

Sample II

$n_2 = 72$

$\bar{x}_2 = 12.5$ km per litre

II population variance $\sigma_2^2 = 2.0$

(a) The appropriate test statistic to be used is the test of difference between the means.

(b) $H_o : \mu_1 = \mu_2$ (ie. there is no difference between the means)

$H_1 : \mu_1 \neq \mu_2$

Under H_o the test statistic is

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

$$= \frac{14 - 12.5}{\sqrt{\frac{1.5}{36} + \frac{2.0}{72}}} = \frac{1.5}{\sqrt{0.0695}} = \frac{1.5}{0.264} = 5.68$$

$$|Z| = 5.68$$

The table value of $|Z|$ at 5% level is 1.96 and at 1% level is 2.58.

Inference :

$|Z| > 1.96$ and $|Z| > 2.58$

$\therefore$ H_o is rejected at 5% level as well as 1% level of significance.

ie. the difference of means is highly significant.

ie. The difference in the petrol consumption of the two type of cars is highly significant.

Example 12 : A sample of heights of 6400 Englishmen has a mean of 170 cms and a standard deviation of 6.4 cms, while a sample of heights of 1600 Australians has a mean of 172 cm and standard deviation of 6.3 cm. Do the data indicate that the Australians are on the average taller than the Englishmen. **[AU 2007]**

Solution :

Sample I	**Sample II**
Englishmen	Australians
$n_1 = 6400$	$n_2 = 1600$
$\bar{x}_1 = 170$ cm	$\bar{x}_2 = 172$ cm
$s_1 = 6.4$ cm	$s_2 = 6.3$ cm

Let μ_1 be the mean height of the population of Englishmen

μ_2 be the mean height of the population of Australians.

To test whether Australians are taller than Englishmen.

$H_o : \mu_1 = \mu_2$ [Englishmen and Australians have same mean height]

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

under H_o, the test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

$$= \frac{170 - 172}{\sqrt{\frac{6.4^2}{6400} + \frac{6.3^2}{1600}}} = \frac{-2 \times 10 \times 8}{\sqrt{40.98 + 4 \times 39.69}}$$

$$= \frac{-160}{\sqrt{199.74}} = \frac{-160}{14.13} = -11.3$$

$$\therefore Z = -11.3$$

For left tailed test at 5% level of significance table value is -1.645 and 1% level of significance, table value is -2.33.

Inference :

$Z < -1.645$ and $Z < -2.33$

$\therefore$ H_0 is rejected at 5% level as well as 1% level of significance.

ie. the difference is highly significant.

$\therefore$ Australians are taller than Englishmen.

Example 13 : A random sample of 100 bulb from a company A showed a mean life 1300 hrs and standard deviation 82 hours. Another random sample of 100 bulbs from company B showed a mean life 1248 hrs and standard deviation of 93 hrs. Are the bulbs of company A superior to bulbs of company B at 5% level of significance.

Solution :

	Sample I	**Sample II**
Given	$n_1 = 100$	$n_2 = 100$
	$\bar{x}_1 = 1300$	$\bar{x}_2 = 1248$
	$s_1 = 82$;	$s_2 = 93$

Population standard deviation σ is not given.

To test the bulbs of company A are superior to that of company B.

N.H : $H_0 : \mu_1 = \mu_2$

A.H : $H_1 : \mu_1 > \mu_2$

$\therefore$ right-tailed lost.

Under H_0, the test statistic is

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

$$= \frac{1300 - 1248}{\sqrt{\frac{82^2}{100} + \frac{93^2}{100}}}$$

$$= \frac{52 \times 10}{\sqrt{6724 + 849}}$$

$$= \frac{520}{\sqrt{15373}} = \frac{520}{123.99} = 4.19$$

For, right tailed test at 5% level, the table value of Z is 1.645

$\therefore \quad Z > 1.645$

Inference :

Since $Z > 1.645$, H_0 is rejected at 5% level of significance.

$\therefore H_1$ is accepted.

i.e., mean life of bulbs of company A is greater than that of company B.

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

3.8 Test for the significant difference between sample proportion and population proportion

Procedure

A large sample of size n be taken from a population. Let P be the proportion of some attribute in the population. Let p be the proportion of the same attribute in the sample. We have to test whether p and P differ significantly or not. Let P_0 be the hypothetical value of the proportion in the population.

Null hypothesis $H_o : P = P_0$

Alternative hypothesis $H_1 : P \neq P_0$ (two-tailed)

$$Q = 1 - P$$

Standard error of $p = \sqrt{\frac{PQ}{n}}$

Under H_o, for large n, the sampling distribution of p is approximately normal with mean P and variance $\sqrt{\frac{PQ}{n}}$

$\therefore$ the test statistic is

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0, 1)$$

Inference : If $|Z| < 1.96$, H_0 is accepted at 5% level of significance.

ie. difference is not significant.

$|Z| > 1.96$, H_0 is rejected at 5% level of significance.

ie. difference is significant.

$|Z| < 2.58$, H_0 is accepted 1% level of significance.

$|Z| > 2.58$, H_0 is rejected at 1% level of significance.

ie. difference is highly significant.

WORKED EXAMPLES

Example 1 : A die was thrown 400 times and 6 resulted 80 times. Do the data justify the hypothesis that the die is unbiased ?

Solution : Given $n = 400$

Let p be the proportion of getting 6 in the sample.

Let P be the proportion of getting 6 in the population.

Given die is thrown 400 times and 6 resulted in 80 times.

$$\therefore p = \frac{80}{400} = \frac{1}{5}$$

We want to test whether the die is unbiased or not.

H_0 : the die is unbiased

H_1 : the die is biased

$\therefore$ two tailed test.

Since the die is unbiased, $P = \frac{1}{6}$

$$\therefore \quad Q = 1 - P = 1 - \frac{1}{6} = \frac{5}{6}$$

Under H_o, the test statistic is

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0, 1)$$

$$= \frac{\frac{1}{5} - \frac{1}{6}}{\sqrt{\frac{1}{6} \times \frac{5}{6} \times \frac{1}{400}}} = \frac{(0.2 - 0.167) \times 6 \times 20}{\sqrt{5}}$$

$$= \frac{0.033 \times 20 \times 6}{2.23606} = \frac{3.996}{2.23606} = 1.79$$

$$|Z| = 1.79$$

The table value of Z at 5% level is 1.96

Inference : Since $|Z| < 1.96$, H_o is accepted at 5% level of significance, ie. the data justifies the hypothesis that the die is unbiased.

Example 2 : A die is thrown 400 times and a throw of 3 or 4 is observed 150 times. Test the hypothesis that the die is fair.

[AU 2012]

Solution : Given the die is thrown 400 times.

$$\therefore n = 400$$

Let P be the proportion of getting 3 or 4 in the sample.

Let P be the proportion of getting 3 or 4 in the population.

When the die is thrown 400 times 3 or 5 resulted in 150 times.

$$\therefore P = \frac{150}{400} = \frac{3}{8}$$

We want to test the hypothesis die is fair or not.

H_0 : the die is fair (unbiased)

H_1 : the die is not fair (biased)

$\therefore$ two tailed test.

Since die is unbiased $P = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

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

Under H_0, the test statistic is

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

$$= \frac{\frac{3}{8} - \frac{1}{3}}{\sqrt{\frac{1}{400} \cdot \frac{1}{3} \cdot \frac{2}{3}}}$$

$$= \frac{(9-8)\,3\sqrt{200}}{24}$$

$$= \frac{1}{8} \times 14.14 = 1.767 = 1.77$$

$$\therefore |Z| = 1.77$$

the table value of Z at 5% level is 1.96

$$\therefore |Z| < 1.96$$

Inference :

Since $|Z| < 1.96$, H_0 is accepted at 5% level of significance.
Hence the die is fair (unbiased).

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

Solution : Given, $n = 900$

Let p be the proportion of heads when a coin is tossed 900 times.

Since head appears 490 times,

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

$$\text{P} = \text{proportion of getting head in the population}$$

$$= \frac{1}{2}, \text{ if the coin is unbiased.}$$

We have to test whether the coin is unbiased or not.

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

$$\text{H}_0 : \text{P} = \frac{1}{2} \text{ (the coin is unbiased)}$$

$$\text{H}_1 : \text{P} \neq \frac{1}{2} \qquad \text{Two tailed test.}$$

Under H_0, the test statistic is

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

$$= \frac{0.544 - 0.5}{\sqrt{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{900}}} = 0.044 \times 2 \times 30 = 2.64$$

$$|\text{Z}| = 2.64$$

The table value of $|\text{Z}|$ at 5% level is 1.96 and 1% level is 2.58.

Inference : Since $|Z| > 1.96$ and $|Z| > 2.58$

H_o is rejected at 5% level and at 1% level of significance.

$\therefore$ the data does not support the hypothesis that the coin is unbiased.

Example 4 : A random sample of 400 mangoes was taken from a large consignment and 40 were found to be bad. Is this a sample from a consignment with proportion of bad mangoes 7.5% ?

Solution : Given $n = 400$

Let P be the proportion of bad mangoes in the consignment and p be the proportion of bad mangoes in the sample.

$$\therefore \quad P = 7.5\% = \frac{7.5}{100} = 0.075$$

$$\therefore \quad Q = 1 - 0.075 = 0.925$$

$$p = \frac{40}{400} = 0.1$$

We have to test whether P = 7.5% or not.

$H_o : P = 7.5\% = 0.075$ (ie. proportion of bad mangoes in the population is 7.5%)

$H_1 : P \neq 7.5\%$ (two tailed test)

under H_o, the test statistic is

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0, 1)$$

$$= \frac{0.1 - .075}{\sqrt{\frac{0.075 \times 0.925}{400}}} = \frac{0.025 \times 20}{0.2632} = 1.89$$

$$|Z| = 1.89$$

The table value of Z at 5% level is 1.96.

Inference : Since $|Z| < 1.96$, H_o is accepted at 5% level of significance.

$\therefore$ The proportion of bad mangoes in the consignment is 7.5%

Example 5 : In a sample of 400 parts produced by a factory, the number of defective parts was found to be 30. The company however claims that only 5% of their products is defective. Is the claim tenable?

Solution : Given $n = 400$

No. of defectives in the sample = 30

p = proportion of defective in the sample

$$\Rightarrow \quad p = \frac{30}{400} = 0.075$$

P = proportion of defective parts of the population

$$= 5\% = \frac{5}{100} = 0.05$$

$$Q = 1 - P = 1 - 0.05 = 0.95$$

We have to test whether the proportion in the population is P = 5% or not.

$$H_o : P = 5\%$$

$$H_1 : P \neq 5\%$$

Two tailed test

Under H_o, the test statistic is

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

$$= \frac{0.075 - 0.05}{\sqrt{\frac{0.05 \times 0.95}{400}}} = \frac{-0.025 \times 20}{\sqrt{0.0475}} = \frac{-0.5}{0.2179} = -2.29$$

$$|Z| = 2.29$$

The table value of $|Z|$ at 5% level is 1.96

Inference : Since $|Z| > 1.96$, H_o is rejected at 5% level of significance.

$\therefore$ The claim is not tenable.

Example 6 : In a sample of 900 articles produced in a factory 81 are found to be defective. The factory claims that atleast 95% of the article supplied are non-defective. Test the claim is true or not ?

Solution : Given $n = 900$

Number of defective articles = 81

Number of good articles = 900 − 81 = 819

p = proportion of good articles in the sample

$$= \frac{819}{900} = 0.91$$

P = Proportion of good articles produced in the factory

= 95%

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

$$Q = 1 - P = 1 - 0.95 = 0.05$$

We have to test whether $P \geq 95\%$ or not

$H_o : P \geq 95\% = 0.95$

$H_1 : P < 0.95$ (left tailed)

Under H_o, the test statistic is

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

$$= \frac{0.91 - 0.95}{\sqrt{\frac{0.95 \times 0.05}{900}}} = \frac{-0.04 \times 30}{0.0475} = \frac{-1.2}{0.2179}$$

$$= -5.507$$

$Z = -5.507$

The table value of Z at 5% level is -1.645

Inference : $Z < -1.645$ and $Z < -2.33$

H_0 is rejected at 5% level and 1% level of significance.

$\therefore$ The claim is not true.

Example 7 : A quality - control engineer suspects that the proportion of defective units among certain manufactured items has increased from the set limit of 0.01. To test the claim, he randomly selected 100 of these items and found that the proportion of defective units in the sample was 0.02. Test the engineer's hypothesis at the 0.05 level of significance.

[AU 2007]

Solution : Given Sample size $n = 100$

Let p = Proportion of defective units in the sample

$= 0.02$

P = Proportion of defective units in the population,

$= 0.01$

$Q = 1 - P = 1 - 0.01 = 0.99$

N.H : H_0 : $P = 0.01$

A.H : H_1 : $P > 0.01$ (Right tailed test)

Under H_0, the test statistic is $Z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$

$$= \frac{0.02 - 0.01}{\sqrt{\dfrac{0.01 \times 0.99}{100}}}$$

$$= \frac{0.01 \times 10}{\sqrt{0.0099}} = \frac{0.1}{0.0994} = 1.006$$

For one-tailed test, the value of Z at 5% level is 1.645

$\therefore Z < 1.645$

Inference : The calculated value of $Z < 1.645$

$\therefore$ H_0 is accepted at 5% level of significance.

The claim $P = 0.01$ is true.

3.9 Test for the significant difference between two proportions in two samples or test for the equality of two proportions

Procedure : Let two large samples of sizes n_1, n_2 be drawn from a population with proportion P for some attribute A. Let p_1, p_2 be the proportions of the same attribute A in the samples respectively. We want to test the significant difference between p_1 and p_2 and $Q = 1 - P$

$H_o : P_1 = P_2$ ie. there is no significant difference between sample proportions.

$H_1 : P_1 \neq P_2$

Two tailed test

Under H_o, the sampling distribution of $p_1 - p_2$ is approximately normal with mean P Variance $PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$

Case : P is known, ie. the population proportion is known. Under H_o, the test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

Inference : If $|Z| < 1.96$, H_o is accepted at 5% level of significance.

If $|Z| > 1.96$, H_o is rejected at 5% level of significance.

If $|Z| < 2.58$, H_o is accepted at level of significance.

If $|Z| > 2.58$, H_o is rejected at 1% level of significance.

Case 2 : P is not known, ie. population proportion is not known.

If P is not known, then the unbiased estimate of P using p_1 and p_2 is

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$\therefore \quad Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

Inference : As like case 1.

WORKED EXAMPLES

Example 8 : In a sample of 400, proportion of tea drinkers is 0.0125 and in another sample of 1200, 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.

Solution : P_1 is the proportion of tea drinkers in the I sample

P_2 is the proportion of tea drinkers in II sample

P is the proportion of tea drinkers in the population

$\therefore$ given, $n_1 = 400$, $n_2 = 1200$

$p_1 = 0.0125$, $p_2 = 0.0083$

$P = 0.01$, $Q = 1 - P = 1 - 0.01 = 0.99$

$\therefore$ P is known.

$H_o : p_1 = p_2$

$H_1 : p_1 \neq p_2$ $\therefore$ Two tailed test.

Under H_o, the test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

$$= \frac{0.0125 - 0.0083}{\sqrt{0.01 \times 0.99\left(\frac{1}{400} + \frac{1}{200}\right)}}$$

$$= \frac{0.0042}{\sqrt{0.0099 \times (0.0025 + 0.008)}} = \frac{0.0042}{0.0057} = 0.73$$

$|Z| = 0.73$

The table value of Z at 5% level is 1.96

Inference : Since $|Z| < 1.96$,

H_o is accepted at 5% level of significance.

ie. the difference is not significant.

$\therefore$ The samples are taken from a population with proportion 0.01.

Example 9 : Random samples of 400 men and 600 women were 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. **[B.D.U 95, A.U 2016]**

Solution : Let p_1, p_2 be the proportions favourable to have a fly over among men and women respectively.

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

p is not known. ∴ The best estimate of P $= \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

$$= \frac{400 \times \dfrac{200}{400} + 600 \times \dfrac{325}{600}}{400 + 600} = \frac{525}{100} = 0.525$$

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

$$H_o : p_1 = p_2$$

$$H_1 : p_1 \neq p_2 \text{ (two tailed test)}$$

Under H_o, the test statistic is

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

$$= \frac{0.5 - 0.542}{\sqrt{0.525 \times 0.475\left(\dfrac{1}{400} + \dfrac{1}{600}\right)}}$$

$$= \frac{-0.042}{\sqrt{0.2494 \times (0.0025 + 0.00167)}}$$

$$= \frac{-0.042}{0.0324} = -1.30$$

$$|Z| = 1.30$$

The table value of Z at 5% level is 1.96

Inference : Since $|Z| < 1.96$, H_o is accepted at 5% level of significance.

The difference is not significant ie. men and women are equally favourable for fly over near their residence.

Example 10 : 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 ?

Solution : Let p_1 and p_2 be the proportion of defective articles produced by the machine before and after overhauling.

$$\therefore \quad p_1 = \frac{20}{400}; \qquad p_2 = \frac{10}{300}$$

P is not known

$\therefore$ best estimate of P is

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{400 \times \frac{20}{400} + 300 \times \frac{10}{300}}{400 + 300}$$

$$= \frac{30}{700} = \frac{3}{70}$$

$$Q = 1 - P = 1 - \frac{3}{70} = \frac{67}{70}$$

$$H_o : p_1 = p_2$$

$$H_1 : p_2 > p_1$$

$\therefore$ one tail test.

Under H_o, the test statistic is

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

$$= \frac{0.05 - 0.0333}{\sqrt{\frac{3}{70} \times \frac{67}{70}\left(\frac{1}{400} + \frac{1}{300}\right)}}$$

$$= \frac{0.0167}{\sqrt{\frac{3}{70} \times \frac{67}{70}(+.0025 ++ 0.0033)}}$$

$$= \frac{0.0167}{0.0154} = 1.08$$

$$Z = 1.08$$

For one tailed test table value of Z at 5% levels 1.645.

Inference : Since $Z < 1.645$, H_o is accepted at 5% level of significance.

∴ The difference is not significant.

∴ The machine has not improved after overhauling.

3.10 Exact Sampling Distributions (t, F, χ^2)

SMALL SAMPLE TESTS

3.10.0 Introduction

We have seen that for large samples (i.e. $n \geq 30$) the sampling distribution of many sample statistics follow the normal distribution. So, the tests of significance for large samples, $Z = \frac{t - E(t)}{S.E(t)}$ followed the standard normal law. But, when the sample is small (i.e. $n < 30$) the sampling distribution of many statistics may not follow the normal law. Also we will not be justified in estimating the population parameters are equal to the corresponding sample statistics. So, a different theory known as small sampling theory was developed. **Small sampling theory is a study of sampling distributions of statistics of small samples.** It is also known as exact sampling theory, because the results obtained here hold for large samples also. In other words, the exact sample tests can be applied for large samples also, but the converse is not true.

In small sampling theory the basic assumption is that the parent population or populations is/are normally distributed. For small sample tests we will be using three distributions, mainly.

1. **Student's t - distribution**
2. **Snedecor's F - distribution**
3. **χ^2 - (chi-square) distribution**

WORKED EXAMPLES

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

Solution : Given, $n = 10$; $\bar{x} = 15.8$ kg ; $s = 0.50$ kg ; $\mu = 16$ kg

$$H_0 : \mu = 16 \text{ kg}$$

$$H_1 : \mu \neq 16 \text{ kg (two-tailed test)}$$

Under H_o, the test statistic is

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \sim \text{Student's '}t\text{' distribution with } \nu = n - 1 \text{ d.f.}$$

$$= \frac{15.8 - 16}{\frac{0.5}{\sqrt{10}}} = \frac{-0.2 \times \sqrt{9}}{0.5} = \frac{-0.2 \times 3}{0.5} = -1.2$$

$$|t| = 1.2$$

No. of degrees of freedom $\nu = n - 1 = 10 - 1 = 9$

For $\nu = 9$ degree of freedom, from the table of t, the value of t at 5% level is $t_{0.05} = 2.26$

$$\therefore |t| < 2.26$$

Inference : Since the calculated value of $|t|$ < the table value of t, H_0 is accepted at 5% level of significance.

$\therefore$ the difference is not significant.

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

Example 2 : 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.

Solution : For the given data we find mean and standard deviation

x	$d = x - A$, $A = 70$	d^2
66	−4	16
65	−5	25
69	−1	1
70	0	0
69	−1	1
71	1	1
70	0	0
63	−7	49
64	−6	36
68	−2	4
	−25	133

$$n = 10, \ \Sigma d = -25, \ \Sigma d^2 = 133$$

$$\therefore \bar{x} = A + \frac{\Sigma d}{n} = 70 - \frac{25}{10} = 70 - 2.5 = 67.5$$

$$s^2 = \frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2 = \frac{133}{10} - \left(\frac{-25}{10}\right)^2$$

$$= 13.3 - (2.5)^2 = 13.3 - 6.25 = 7.05$$

$$s^2 = 7.05$$

$$s = \sqrt{7.05} = 2.655$$

$\therefore n = 10, \bar{x} = 67.5, s = 2.655$

$H_0 : \mu = 65$ (The mean price of the share in the month is 65)

$H_1 : \mu \neq 65$ (The mean price of the share in the month is not 65)

Under H_0, the test statistic is

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

$$= \frac{6.75 - 65}{\frac{2.655}{\sqrt{10-1}}} = \frac{2.5 \times \sqrt{9}}{2.655} = \frac{2.5 \times 3}{2.655} = 2.82$$

$$|t| = 2.82$$

No.of degrees of freedom $\nu = n - 1 = 10 - 1 = 9$

For $\nu = 9$ d.f, from the table of t, the value of t at 5% level is

$t_{0.05} = 2.26$

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

Inference : Since the calculated value of $|t| <$ the table value of t, H_0 is rejected at 5% level of significance.

Hence the mean price of the shares in the month is not 65

Example 3 : A random sample of 9 items is taken from a certain measurement. From the data it is found that $\Sigma x = 108, \Sigma x^2 = 1584$. Test the hypothesis that the population mean is 8.

Solution : Given $n = 9, \Sigma x = 108, \Sigma x^2 = 1584$

$$\therefore \bar{x} = \frac{\Sigma x}{n} = \frac{108}{9} = 12$$

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

$$= \frac{1584}{9} - 12^2 = 176 - 144 = 32$$

$$\therefore s = \sqrt{32} = 5.66$$

We have $n = 9, \bar{x} = 12, s = 5.66, \mu = 8$

$H_0 : \mu = 8$ (i.e. the sample is drawn from a population with mean $\mu = 8$)

$H_1 : \mu \neq 8$

$\therefore$ Under H_0, the test statistic is

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \sim \text{Student's } t \text{ distribution with } \nu = n - 1 \text{ d.f.}$$

$$= \frac{12 - 8}{\frac{5.66}{\sqrt{9-1}}} = \frac{4 \times \sqrt{8}}{5.66} = \frac{4 \times 2.828}{5.66} = 1.99$$

Number of degrees of freedom $\nu = n - 1 = 9 - 1 = 8$

For $\nu = 8$ d.f, from the table of t, the value of t at 5% level is

$$t_{0.05} = 2.306$$

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

Inference : The calculated value of $|t|$ < the table value of t,

H_0 is accepted at 5% level of significant.

i.e. the population mean is 8.

Example 4 : 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. **[M.U.96]**

Solution : First we find mean and standard deviation from the given data.

x	$d = x - A$ $A = 163$	d^2
163	0	0
163	0	0
164	1	1
165	2	4
166	3	9
169	6	36
170	7	49
171	8	64
	27	163

$$n = 8,\ \Sigma d = 27,\ \Sigma d^2 = 163$$

$$\bar{x} = A + \frac{\Sigma d}{n} = 163 + \frac{27}{8} = 163 + 3.375 = 166.375$$

$$s^2 = \frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2 = \frac{163}{8} - \left(\frac{27}{8}\right)^2 = \frac{163}{8} - (3.375)^2$$

$$= 20.375 - 11.391 = 8.984$$

$$s = \sqrt{8.984} = 2.997$$

We have $n = 8,\ \bar{x} = 166.375,\ s = 2.997,\ \mu = 165$

$H_0 : \mu = 165$ (i.e. sample is taken from the population with mean ht. 165)

$H_1 : \mu \neq 165$ (sample is taken from the population with height not equal to 165)

Under H_0, the test statistic is

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

$$= \frac{166.375 - 165}{\frac{2.997}{\sqrt{8-1}}}$$

$$= \frac{1.375 \times \sqrt{7}}{2.997}$$

$$= \frac{1.375 \times 2.645}{2.997} = \frac{3.65}{2.997} = 1.214$$

Number of degrees of freedom $\nu = n - 1 = 8 - 1 = 7$

For $\nu = 7$ d.f, from the table of t, the value of t at 5% level of significance is $t_{0.05} = 2.365$.

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

Inference : Since the calculated value of $|t|$ < the table value of t, H_0 is accepted at 5% level of significance

$\therefore$ the mean height can be regarded as 165 cm.

Example 5 : 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 ?

[B.D.U. 94, 95] [A.U 2015, 2016]

Solution : First we compute mean $\bar{x}$ and standard deviation s.

x	$d = x - A$ $A = 100$	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

$$n = 10, \Sigma d = -28, \Sigma d^2 = 1912$$

$$\bar{x} = A + \frac{\Sigma d}{n} = 100 - \frac{28}{10} = 100 - 2.8 = 97.2$$

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

$$= \frac{1912}{10} - (2.8)^2$$

$$= 191.2 - 7.84 = 183.36$$

$$s = \sqrt{183.36} = 13.54$$

We want to test that mean IQ is 100

$H_0 : \mu = 100$ (mean IQ in the population is 100)

$H_1 : \mu \neq 100$ (mean IQ is the population is not 100)

$\therefore$ Under H_0, the test statistic is

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

$$= \frac{97.2 - 100}{\frac{13.54}{\sqrt{9}}} = \frac{2.8 \times 3}{13.54} = 0.62$$

$$|t| = 0.62$$

Number of degrees of freedom $\nu = n - 1 = 10 - 1 = 9$

For $\nu = 9$, from the table of t, the value of t at 5% level is

$t_{0.05} = 2.26$

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

Inference : Since the calculated value of $|t|$ < the table val of t, H_0 is accepted at 5% level of significance.

Therefore the mean IQ of the population can be 100.

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

Procedure

Let two independent small random samples of sizes n_1, n_2 be drawn from two normal populations with mean μ_1, μ_2. Let $\bar{x}_1, \bar{x}_2$ be the means of the samples and s_1, s_2 be their standard deviations. We assume that the two populations have equal unknown variances. So, the best estimate of the population variance using the sample standard deviations is obtained and the best estimate is given by

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

We have to test whether the two samples are taken from the same population or from the populations with same mean. Hence we assume,

$H_0 : \mu_1 = \mu_2$ [There is no significant difference between the means of the two populations]

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

The standard error of $(\bar{x}_1 - \bar{x}_2)$ is $\sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$

Since the samples are small, we use Student's t-test, to test the significance of the difference between the means.

$\therefore$ Under H_0, the test statistic is

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

which follows the Student's t distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom.

For $\nu = n_1 + n_2 - 2$, let $t_{0.05}$ be the table value of t at 5% level of significance.

Inference : If $|t| < t_{0.05}$, H_0 is accepted at 5% level of significance.

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

If $|t| > t_{0.05}$, H_0 is rejected at 5% level of significance.

$\therefore$ The difference between means is significant.

Similarly, if $|t| < t_{0.01}$, H_0 is accepted at 1% level of significance.

$\therefore$ the difference is not significant.

If $|t| > t_{0.01}$, H_0 is rejected at 1% level of significance.

$\therefore$ the difference is highly significant.

Note : If $n_1 = n_2 = n$ and the samples are independent, then

$$\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{n\,(s_1^2 + s_2^2)}{2n - 2} \times \frac{2}{n}$$

$$= \frac{s_1^2 + s_2^2}{n - 1}$$

$$\therefore\ t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2 + s_2^2}{n - 1}}}$$ follows Student's t-distribution with

$\nu = 2n - 2$ degrees of freedom.

Remark : Assumptions made to test the difference 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 i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$

3. The two samples are random samples and independent.

WORKED EXAMPLES

Example 1 : Two random samples gave the following results

Sample	Size	Sample mean	Sum of the squares of deviations from the mean
1	10	15	90
2	12	14	108

Examine whether the samples come from the same normal population. **[A.U.2006, 2012]**

Solution : The samples are small.

$\therefore$ We use Student's t test to test the significant difference.

Given $n_1 = 10, \quad n_2 = 12$

$\bar{x}_1 = 15, \quad \bar{x}_2 = 14$

$\Sigma(x_1 - \bar{x}_1)^2 = 90, \quad \Sigma(x_2 - \bar{x}_2)^2 = 108$

$$s_1^2 = \frac{1}{n_1}\Sigma(x_1 - \bar{x}_1)^2 = \frac{90}{10} = 9$$

$$s_2^2 = \frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2} = \frac{108}{12} = 9$$

Let μ_1, μ_2 be the means of the two populations.

We have to test whether the two samples come from the same normal population.

$H_0 : \mu_1 = \mu_2$ [There is no significant difference between the means of the two populations]

$H_1 : \mu_1 \neq \mu_2$ [There is significant difference between the means of the two populations]

$\therefore$ two tailed test.

Here $n_1 \neq n_2$

$\therefore$ Under H_0, the test statistic is

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

$$= \frac{15 - 14}{\sqrt{\frac{10 \times 9 + 12 \times 9}{10 + 12 - 2}\left(\frac{1}{10} + \frac{1}{12}\right)}}$$

$$= \frac{1}{3\sqrt{\frac{(10 + 12)}{20} \times \frac{(10 + 12)}{120}}}$$

$$= \frac{1}{3\sqrt{\frac{22 \times 22}{24 \times 100}}} = \frac{1}{3\sqrt{\frac{11 \times 11}{6 \times 100}}}$$

$$= \frac{10\sqrt{6}}{3 \times 11} = \frac{10 \times 2.4494}{33} = 0.74$$

$$|t| = 0.74$$

Number of degrees of freedom $\nu = n_1 + n_2 - 2 = 10 + 12 - 2 = 20$

For $\nu = 20$, d.f. from the table of t, the value of t at 5% level is $t_{0.05} = 2.086$

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

Inference : Since the calculated value of $|t|$ < the table value of t, H_0 is accepted at 5% level of significance.

i.e. the samples come from the same normal population.

Example 2 : Two independent samples from normal populations with equal variance gave the following.

Sample	Size	Mean	Standard deviation
1	16	23.4	2.5
2	12	24.9	2.8

Is the difference between the means significant. **[M.U.1996]**

Solution : The samples are small

$\therefore$ we use Student's t-test, to test the significant difference between means

Given : $n_1 = 16, \quad n_2 = 12$

$\bar{x}_1 = 23.4, \quad \bar{x}_2 = 24.9$

$s_1 = 2.5, \quad s_2 = 2.8$

Here $n_1 \neq n_2$

We have to test the significance of difference between the means of the two samples.

$H_0 : \mu_1 = \mu_2$ [There is no significant difference between means]

$H_1 : \mu_1 \neq \mu_2$ [There is significant difference between means]

$\therefore$ two tailed test.

Under H_0, the test statistic is

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

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

$$= \frac{1.5}{\sqrt{\frac{100 \times 94.08}{26} \times \left(\frac{3+4}{48}\right)}}$$

$$= \frac{1.5\sqrt{26 \times 48}}{\sqrt{194.08 \times 7}} = \frac{1.5 \times 35.33}{\sqrt{1358.56}} = \frac{52.995}{36.8586}$$

$$= 1.43779 = 1.44$$

Number of degree of freedom $\nu = n_1 + n_2 - 2 = 16 + 12 - 2 = 26$

For $\nu = 26$, from the table of t, the value of t at 5% level is $t_{0.05} = 2.06$

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

Inference : Since the calculated value of $|t|$ < table value of t,

$\therefore$ H_0 is accepted at 5% level of significance.

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

Example 3 : 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 significantly at 5% level of significance and 1% level of significance.

[M.U-90]

Solution : Let μ_1, μ_2 be the average marks by regular and part-time students respectively.

We want to test the significance of the difference between means.

The samples are small. So we use Student's t test.

$H_0 : \mu_1 = \mu_2$ [There is no significant difference between means]

$H_1 : \mu \neq \mu_2$ [There is significant difference between means]

$\therefore$ two tailed test.

Now we find means $\bar{x}_1, \bar{x}_2$ and variances s_1^2, s_2^2.

Regular

x_1	$d_1 = x_1 - A$, $A = 60$	d_1^2
56	−4	16
62	2	4
63	3	9
54	−6	36
60	0	0
51	−9	81
67	7	49
69	9	81
58	−2	4
	0	280

Part time

x_2	$d_2 = x_2 - B$, $B = 66$	d_2^2
62	−4	16
70	4	16
71	+5	25
62	−4	16
60	−6	36
56	−10	+100
75	+9	81
64	−2	4
72	+6	36
68	+2	4
66	0	0
	0	334

$$\bar{x}_1 = A + \frac{\Sigma d_1}{n_1} = 60 + 0 = 60$$

$$\bar{x}_2 = B + \frac{\Sigma d_2}{n_2} = 66 + 0 = 66$$

$$s_1^2 = \frac{\Sigma d_1^2}{n_1} - \left(\frac{\Sigma d_1}{n_2}\right)^2 = \frac{280}{9} - 0 = \frac{280}{9}$$

$$s_2^2 = \frac{\Sigma d_2^2}{n_2} - \left(\frac{\Sigma d_2}{n_2}\right)^2 = \frac{334}{11} - 0 = \frac{334}{11}$$

Here $n_1 \neq n_2$, $\bar{x}_1 = 60$, $\bar{x}_2 = 66$, $s_1^2 = \frac{280}{9}$, $s_2^2 = \frac{334}{11}$

Under H_0, the test statistic is

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

$$= \frac{60 - 66}{\sqrt{\frac{9 \times \frac{280}{9} + 11 \times \frac{334}{11}}{9 + 11 - 2}\left(\frac{1}{9} + \frac{1}{11}\right)}}$$

$$= \frac{-6}{\sqrt{\frac{280 + 334}{18}\left(\frac{11 + 9}{9 \times 11}\right)}} = \frac{-6}{\sqrt{34.11 \times 0.202}}$$

$$= \frac{-6}{\sqrt{6.89022}} = \frac{-6}{2.6249} = -2.28$$

$$\therefore \ |t| = 2.28$$

Number of degrees of freedom for $\nu = n_1 + n_2 = 9 + 11 - 2 = 18$

For $\nu = 18$, the table value of t at 5% level is $t_{0.05} = 2.101$

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

For $\nu = 18$, from the table of t, the value of t at 1% level is $t_{0.01} = 2.878$

$$\therefore |t| < t_{.01}$$

Inference : Since the calculated value of $|t|$ > the table value of t, H_0 is rejected at 5% level of significance.

∴ the marks obtained by regular students and part-time students differ significantly at 5% level

Since the calculated value of $|t|$ < table value of t, H_o is accepted at 1% level of significance.

∴ the difference is not significant at 1% level.

Example 4 : 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.

[A.U.2001]

Solution : Let μ_1, μ_2 be the means of the populations. Since the two samples are small,

we use Student's t test, to test the significance of difference between the means.

$H_0 : \mu_1 = \mu_2$ [No significant difference between means]

$H_1 : \mu_1 > \mu_2$ [B runs faster than A]

So one tail test-right tailed.

Now we shall find the sample means $\bar{x}_1, \bar{x}_2$ and their variances s_1^2, s_2^2 from the data.

Horse A

x_1	$d_1 = x_1 - A$, $A = 33$	d_1^2
28	−5	25
30	−3	9
32	−1	1
33	0	0
33	0	0
29	−4	16
34	1	1
	−12	52

$n_1 = 7$

$$\bar{x}_1 = A + \frac{\Sigma d_1}{n_1}$$

$$= 33 - \frac{12}{7}$$

$$= 33 - 1.714$$

$$= 31.285$$

$$= 31.29$$

Horse B

x_2	$d_2 = x_2 - B$, $B = 30$	d_2^2
29	−1	1
30	0	0
30	0	0
24	−6	36
27	−3	9
29	−1	1
	−11	47

$n_2 = 6$

$$\bar{x}_2 = B + \frac{\Sigma d_2}{n_2}$$

$$= 30 - \frac{11}{6}$$

$$= 30 - 1.83$$

$$= 30 - 1.83$$

$$= 28.17$$

$$s_1^2 = \frac{\Sigma d_1^2}{n_1} - \left(\frac{\Sigma d_1}{n_1}\right)^2$$

$$= \frac{52}{7} - (1.71)^2$$

$$= 7.42 - 2.92$$

$$= 4.50$$

$$s_2^2 = \frac{\Sigma d_2^2}{n_2} - \left(\frac{\Sigma d_2}{n_2}\right)^2$$

$$= \frac{47}{6} - (1.83)^2$$

$$= 7.83 - 3.35$$

$$= 4.48$$

Now we have $n_1 = 7$, $n_2 = 6$, $\bar{x}_1 = 31.29$, $\bar{x}_2 = 28.17$, $s_1^2 = 4.5$, $s_2^2 = 4.48$

Under H_0, the test statistic is

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

$$= \frac{31.29 - 28.17}{\sqrt{\frac{7 \times 4.5 + 6 \times 4.48}{7 + 6 - 2}\left(\frac{1}{7} + \frac{1}{6}\right)}}$$

$$= \frac{3.12}{\sqrt{5.307 \times 0.3095}}$$

$$= \frac{3.12}{1.2816} = 2.49$$

Number of degrees of freedom $\nu = n_1 + n_2 - 2$

$$= 7 + 6 - 2 = 11$$

For $\nu = 11$ d.f., from the table of t, the value of t at 5% level for one tailed test is

$$t_{0.05} = 1.796$$

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

Inference : Since the calculated value of $|t|$ is greater than table value of t, H_0 is rejected at 5% level of significance.

$\therefore$ difference is significant.

Hence B runs faster than A.

3.10.4 t-Test for paired Observations or Dependent Samples

So far, we considered two independent samples. But, there are situations where the pairs of values of x_1 and x_2 are associated in some way. Hence we cannot consider them as independent samples. So above formula cannot be used. For example, consider the IQ scores of 10 students before and after a training are obviously related pairs (infact correlated) and we want to test whether training has any impact on the IQ of students. In such cases we apply t-test after converting the data into a single sample type by finding their difference

$$d = x_1 - x_2. \qquad \therefore \bar{d} = \bar{x}_1 - \bar{x}_2.$$

We test the difference between $\bar{d}$ and 0 under the null hypothesis.

Let μ be the difference between the means of the population before and after training.

$H_0 : \mu = 0$ (i.e. there is no difference in the means)

$H_1 : \mu \neq 0$ (i.e. there is difference)

Under H_0, the test statistic is

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

where s^2 is variance of d and $\nu = n - 1$ is the degrees of freedom.

The table value of t for ν degrees of freedom at 5% level is $t_{0.05}$

Inference : If $|t| < t_{0.05}$, H_o is accepted at 5% level of significance.

If $|t| > t_{0.05}$, H_o is rejected at 5% level of significance.

WORKED EXAMPLES

Example 10 : 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 benefitted by the extra coaching.

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

Solution : Since the two values are associated, we use paired t-test for difference of means.

Let μ be the mean difference between marks of all the students before and after the coaching we want to test that the training is effective or not.

$H_0 : \mu = 0$ [The students are not benefitted by extra coaching]

$H_1 : \mu > 0$ [The students are benefitted by the coaching]

One tailed test

$\therefore$ Under H_0, the test statistic is $t = \dfrac{\bar{d}}{\dfrac{s}{\sqrt{n-1}}}$

where $d = y - x$ or $x - y$,

x denotes the marks in first test and y denotes marks in second test.

$\bar{d}$ the mean of d and s is the standard deviation of d.

We find $\bar{d}$ and s.

Marks in I test x	Marks in II test y	$d = y - x$	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	16
19	18	−1	1
		12	58

$$n = 11, \Sigma d = 12, \Sigma d^2 = 58$$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{12}{11} = 1.09$$

$$s^2 = \frac{\Sigma d^2}{n} - (\bar{d})^2 = \frac{58}{11} - (1.09)^2 = 5.273 - 1.188 = 4.085$$

$$s = \sqrt{4.085} = 2.02$$

Now $n = 11, \bar{d} = 1.09, s = 2.02$

$$\therefore t = \frac{\bar{d}}{\frac{s}{\sqrt{n-1}}} = \frac{1.09}{\frac{2.02}{\sqrt{10}}} = \frac{3.447}{2.02} = 1.71$$

$$t = 1.71$$

No. of degrees of freedom $v = n - 1 = 11 - 1 = 10$

For $v = 10$ d.f, from the table of t, the value of t at 5% level for one tailed test is $t_{0.05} = 1.812$

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

Inference : Since the calculated value of $t <$ the table value of t, H_0 is accepted at 5% level of significance.

$\therefore$ The students are not benefitted by extra coaching.

Example 11 : 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	121

Test whether there is change in IQ after the training.

Solution : Since the two pairs of values are associated, we use paired t test for the difference of means. We want to test the training is effective or not.

Let μ be the difference between mean IQ of all the persons before and after training.

$H_0 : \mu = 0$ [There is no change in IQ due to training]

$H_1 : \mu > 0$ [There is a change in IQ due to the training]

Under H_0, the test statistic is

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

We find $\bar{d}$ and s

x	y	$d = y - x$	d^2
110	120	10	100
120	118	−2	4
123	125	2	4
132	136	4	16
125	121	−4	16
		10	140

$$n = 5, \Sigma d = 10, \Sigma d^2 = 140$$

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

$$s^2 = \frac{\Sigma d^2}{n} - (\bar{d})^2 = \frac{140}{5} - 2^2 = 28 - 4 = 24$$

$$s = \sqrt{24} = 4.899$$

$$\therefore t = \frac{\bar{d}}{\frac{s}{\sqrt{n-1}}} = \frac{2}{\frac{4.899}{\sqrt{4}}} = \frac{4}{4.899} = 0.816$$

Number of degrees of freedom $\nu = n - 1 = 5 - 1 = 4$

For $\nu = 4$ d.f, from the table of t, the value of t for one tailed test at 5% level is $t_{0.05} = 2.132$

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

Inference : Since the calculated value of $t <$ the table value of t, H_0 is accepted at 5% level of significance.

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