UNIT – IV

TEST OF HYPOTHESIS

Population: Population (or Universe) is the aggregate or totality of statistical data forming a subject of investigation. For example

(i) The population of the heights of Indians.

(ii) The population of nationalized Banks in Indian, etc.

The number of observations in the population is defined to be the size of the population. It may be finite or infinite. Size of population is denoted by N.

Sample: Most of times, study of entire population may not be possible to carry out and hence a part alone is selected from the given population. A portion of the population which with a view to determining the population characteristics is called sample. i.e., a sample is a subject of population and the number of objects in the sample is called size of the sample. Size of sample is denoted by n.

Classification of Samples: Sample are classified in two ways

(i) **Large sample:** If the size of the sample $(n) \geq 30$, the sample is said to be large sample.

(ii) **Small sample:** If the size of the sample $(n) < 30$, the sample is said to be small sample.

Estimation: Estimation is a statement made to find an unknown population parameter. There are two kinds of estimations to determine the statistic of the population parameters namely (i) Point estimation (ii) Interval estimation.

Point Estimation: If an estimate of a population parameter is given by a single value, then the estimate is called a point estimation of the parameter.

Interval Estimation: If an estimate of a population parameter is given by two different values between which the parameter may be considered to lie, then the estimate is called an interval estimation of the parameter.

Parameters and statistics: Parameter is a statistical measure based on all the units (observations) of a population. Statistics is a statistical measure based only all the units selected in a sample. Thus parameters refer to population while statistics refer to sample. The statistical constants of the population namely μ, variance σ^2 are usually referred to a parameters, statistical measures computed from the sample observations alone e.g., sample mean ($\bar{x}$), sample variance (s^2) etc are usually referred to as statistics. In other words, mean, median, mode, variance and standard deviation measures of the population are called parameters and measures obtained from the sample of the population are called statistics.

Sampling Distribution:

Sampling theory is the study of relationships between a population and samples drawn from the population, and it is applicable to random samples only. We need to obtain maximum information about the population with the help of the sample, i.e., to determine the true value of the population parameters (population mean, S.D, proportions etc,) by using the sample statistics like a sample mean , S.D, proportions etc., and to find the limits of accuracy of estimates based on the samples. It helps us to determine whether the differences between two samples are actually due to change variations or whether they are really significance. Sampling theory also useful in testing of hypothesis and significance.

Sampling Distribution of Statistic:

Sampling distribution of statistic is the frequency distribution which is formed with varies values of statistics computed from different samples of the same size drawn from the same population. We can draw a large number of samples of same size from a population of fixed size, each sample containing different population members. Any statistic like mean, medium, variance, etc. may be computed for each of these samples. As a result a series of various values of that statistic may be obtained. These various values can be arranged into a frequency distribution table, which is known as the sampling distribution of statistic.

Central limit Theorem: If $\bar{x}$ be the mean of a sample size n drawn from a population having mean μ and S.D σ the standardized sample mean $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is a random variable whose distribution function approaches that of standard normal distribution.

Standard Error (S.E) of Statistic:

Standard Error (S.E) of Statistic t is the S.D of sampling distribution of the statistic. Thus Standard Error of sample mean $\bar{x}$ is of the sampling distribution of sample mean.

Formulae for Standard Error (S.E)

1. The S.E of sample mean $\bar{x} = \frac{\sigma}{\sqrt{n}}$
2. The S.E of sample proportion $p = \sqrt{\frac{P\,Q}{n}}$ Q = 1 – P
3. The S.E of sample S.D (s) $= \frac{\sigma}{\sqrt{2\,n}}$
4. The S.E of the difference of two sample means $\bar{x}_1$ *and* $\bar{x}_2$

   i.e., S.E of $(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
5. The S.E of the difference of two sample proportion p_1 *and* p_2

   i.e., S.E of $(p_1 - p_2) = \sqrt{\frac{P_1\,Q_1}{n_1} + \frac{P_2\,Q_2}{n_2}}$

   Where p_1 *and* p_2 are the proportions of two random samples of sizes n_1 *and* n_2 drawn from the two populations with proportions P_1 *and* P_2 respectively.

6. The S.E of the difference of S.D's s_1 and s_2

i.e., S.E of $(s_1 - s_2) = \sqrt{\frac{\sigma_1^2}{2\,n_1} + \frac{\sigma_2^2}{2\,n_2}}$

Test of Hypothesis: We need to decide whether to accept or reject a statement about the parameter. This statement is called a hypothesis and the decision making procedure about the hypothesis is called Test of Hypothesis.

Statistical Hypothesis: In many circumstances, to arrive at decisions about the population on the basis of sample information, we make assumptions (or guesses) about the population parameters involved. Such an assumption (or statement) is called a statistical hypothesis which may or may not true. The procedure which enables us to decide on the basis of sample results whether a hypothesis is true or not, is called the Test of Hypothesis or Test of significance. The Test of Hypothesis are two types

(i) Null Hypothesis (ii) Alternative Hypothesis

Null Hypothesis: For applying the tests of significance, we first set up a hypothesis – a definite statement about the population parameter. Such a hypothesis is namely a hypothesis of no – difference, called Null Hypothesis. It is denoted by H_0 Or

A null hypothesis is the hypothesis which asserts that there is no significance difference between statistic and the population parameter. It is denoted by H_0. $H_0: \mu = \mu_0$

Alternative Hypothesis: Any hypothesis which contradicts the null hypothesis is called Alternative Hypothesis, usually it is denoted by H_1. The two Hypothesis H_0 and H_1 are such that if one true, the other is false and vice versa. For example, if we want test the null hypothesis that the population has specified mean μ_0 (say) i.e., $H_0: \mu = \mu_0$, the alternative hypothesis would be

(i) $H_1: \mu \neq \mu_0$ (ii) $H_1: \mu > \mu_0$ (iii) $H_1: \mu < \mu_0$

The alternative hypothesis (i) is known as a two – tailed alternative and alternative hypothesis in (ii) is known as right – tailed and in (iii) is known as left – tailed test.

Errors of Sampling:

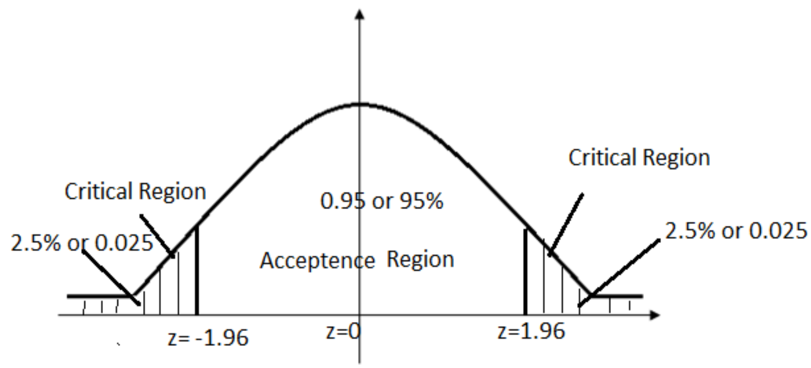
(i) **Type – I error**: Reject H_0 when it is true.

If the Null hypothesis H_0 is true but it is rejected by test procedure then the error is called *Type – I error.*

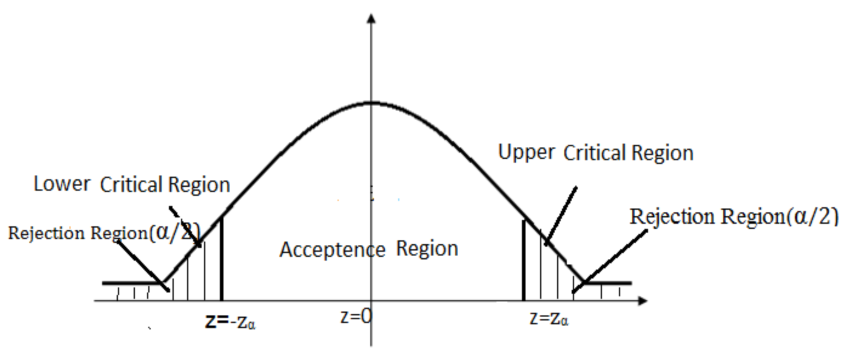
(ii) **Type – II error:** Accept H_0 when it is false.

If the Null hypothesis H_0 is false but it is accepted by test procedure then the error is called *Type – II error*.

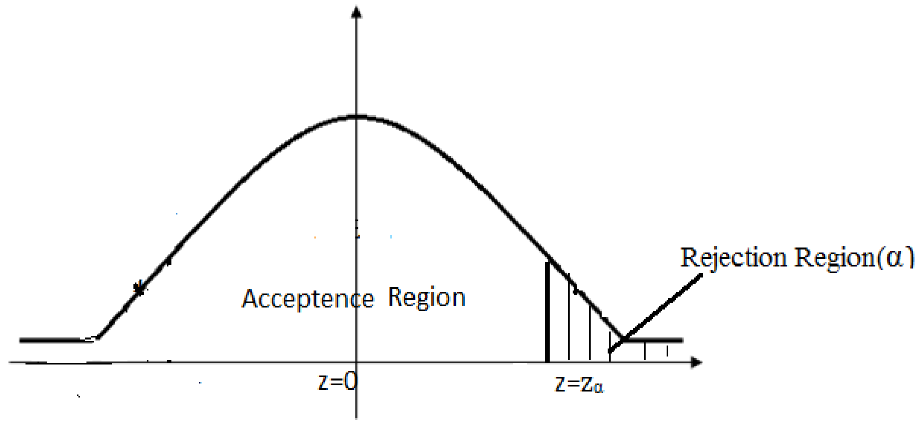
Critical Region: A region corresponding to a statistic '*t*' in the sample space S which leads to the rejection H_0 *critical region* or *rejection region*. Those region which leads to the acceptance H_0 gives us a region called *Acceptance Region.*



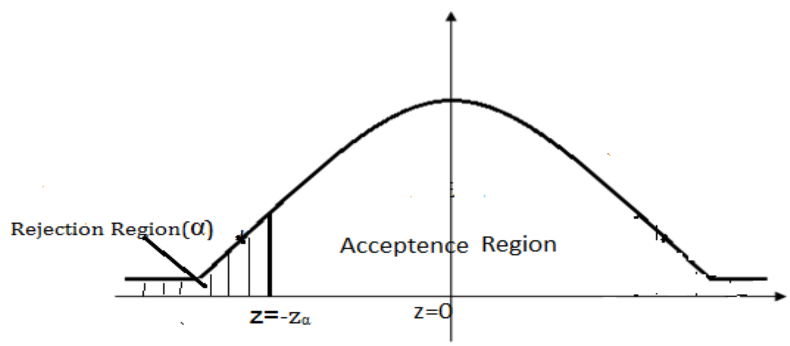
Two – tailed test at level of significance 'α'



Right – tailed test at level of significance 'α'



Left – tailed test at level of significance ‘α'



Critical Values (or) Significance Values:

The value of the test statistic, which separates the critical region (rejection region) and the acceptance region, is called *critical value or significance value.*

Procedure for Testing of Hypothesis:

Various steps involved in Testing of Hypothesis are given bellow

Step 1 : Null Hypothesis : Define or set up a Null Hypothesis H_0 taking into consideration the nature of the problem and data involved.

Step 2 : Alternative Hypothesis : Set up the alternative hypothesis H_1 so that we could decide whether we should use One – tailed test or Two – tailed test.

Step 3 : Level of Significance: Select appropriate Level of significance (α) depending on the reliability of the estimates and permissible risk. That is a suitable α is selected in advance. (If it is not given in the problem usually we choose 5% Level of Significance)

Step 4 : Test Statistic : Compute the Test Statistic

$$Z = \frac{t - E(t)}{S.E\ of\ t}$$ under the Null Hypothesis

Here t is the sample statistic and S.E is Standard Error of t.

Step 5 : Conclusion : We compare the computed value of the Test Statistic Z with the critical value Z_α at given Level of significance.

If $|Z| < Z_\alpha$ we Accept H_0 . If $|Z| > Z_\alpha$ we Reject H_0.

For two – tailed test:

(i) If $|Z| < 1.96$ accept H_0 at 5% Level of significance

(ii) If $|Z| > 1.96$ reject H_0 at 5% Level of significance

(iii) If $|Z| < 2.58$ accept H_0 at 1% Level of significance

(iv) If $|Z| > 2.58$ reject H_0 at 1% Level of significance

For Single – tailed test (right – tailed test or left – tailed test)

(i) If $|Z| < 1.645$ accept H_0 at 5% Level of significance

(ii) If $|Z| > 1.645$ reject H_0 at 5% Level of significance

(iii) If $|Z| < 2.33$ accept H_0 at 1% Level of significance

(iv) If $|Z| > 2.33$ reject t H_0 at 1% Level of significance

Test of Significance for Large Samples:

If the sample size n ≥ 30, then we consider such a samples as large samples. If n is large, the distributions such as Binomial Distribution , Poisson Distribution, Chi–Square Distribution etc. are closely approximated by Normal Distribution. Therefore, the large samples, the sampling distribution of a statistic is approximately Normal Distribution.

Suppose we wish to test the hypothesis that the probability of success in such trials is p. Assuming it to be true , the mean μ and the standard deviation σ of the sampling distribution of number of success are

np & $\sqrt{npq}$ respectively.

If x be the observed number of success in the sample and Z is the standard normal variate then $\boldsymbol{Z = \frac{x-\mu}{\sigma}}$.

Examples:

1. A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use 5% level of significance.

Solution: Here n = 960, p= probability of getting head = ½

∴ q = 1 – p = ½ $\quad \mu = np \;\; \& \;\; \sigma = \sqrt{npq}$

$$\therefore \mu = np = 960\left(\frac{1}{2}\right) = 480$$

$$\therefore \sigma = \sqrt{npq} = \sqrt{960\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{240} = 15.49$$

x = the number of successes = 183

1. **Null Hypothesis** $\boldsymbol{H_0}$ **:** The coin is unbiased.
2. **Alternative Hypothesis** $\boldsymbol{H_1}$ **:** The coin is biased.
3. **Level of Significance** **:** α = 0.05
4. **The test statistic is** $Z = \frac{x-\mu}{\sigma} = \frac{183-480}{15.49} = \frac{-297}{15.49} = -19.17$

∴ | Z | = 19.17

∴ | Z | > 1.96

∴ The Null Hypothesis H_0 is rejected

5. **Conclusion:** The coin is biased.

2. A coin was tossed 400 times and returned heads 216 times. Test the hypothesis that the coin is unbiased. Use 5% level of significance.

Solution: Here n = 400, p= probability of getting head = ½

∴ q = 1 – p = ½ $\mu = np$ & $\sigma = \sqrt{npq}$

$$\therefore \mu = np = 400\left(\frac{1}{2}\right) = 200$$

$$\therefore \sigma = \sqrt{npq} = \sqrt{400\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{100} = 10$$

x = the number of successes = 216

1. **Null Hypothesis** H_0 **:** The coin is unbiased.
2. **Alternative Hypothesis** H_1 **:** The coin is biased.
3. **Level of Significance** **:** α = 0.05
4. **The test statistic is** $Z = \frac{x-\mu}{\sigma} = \frac{216-200}{10} = \frac{16}{10} = 1.6$

∴ | Z | = 1.6

∴ | Z | < 1.96

∴ The Null Hypothesis H_0 is accepted

5. **Conclusion:** The coin is unbiased.

3. A die is tossed 960 times and it falls with 5 upwards 184 times. If the die unbiased at 1% level of significance.

Solution:

Here n = 960, p= probability of throwing 5 with one die $=\left(\frac{1}{6}\right)$

$\therefore$ q = 1 – p = $\left(\frac{5}{6}\right)$ $\quad \mu = np \;\; \& \;\; \sigma = \sqrt{npq}$

$$\therefore \mu = np = 960 \left(\frac{1}{6}\right) = 160$$

$$\therefore \sigma = \sqrt{npq} = \sqrt{960\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)} = \sqrt{133.33} = 11.55$$

x = the number of successes = 184

1. **Null Hypothesis** $\boldsymbol{H_0}$ **:** The die is unbiased.
2. **Alternative Hypothesis** $\boldsymbol{H_1}$ **:** The die is biased.
3. **Level of Significance** **:** $\alpha = 0.01$
4. **The test statistic is** $Z = \frac{x-\mu}{\sigma} = \frac{184-160}{11.55} = \frac{24}{11.55} = 2.08$

$$\therefore |Z| = 2.08$$

$$\therefore |Z| < 2.58$$

$\therefore$ The Null Hypothesis $\boldsymbol{H_0}$ is accept

5. **Conclusion:** The die is unbiased.

4. A die is tossed 256 times and it turns up with an even digit 156 times. Test the hypothesis that the die is unbiased at 5% level of significance.

Solution:

Here n = 256, p= probability of getting an even digit (2 or 4 or 6)

$=\left(\frac{3}{6}\right) = \left(\frac{1}{2}\right)$

$\therefore$ q = 1 – p = $\left(\frac{1}{2}\right)$ $\quad \mu = np \ \ \& \ \sigma = \sqrt{npq}$

$$\therefore \mu = np = 256\left(\frac{1}{2}\right) = 128$$

$$\therefore \sigma = \sqrt{npq} = \sqrt{256\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{64} = 8$$

x = the number of successes = 156

1. **Null Hypothesis** $\boldsymbol{H_0}$ **:** The die is unbiased.
2. **Alternative Hypothesis** $\boldsymbol{H_1}$ **:** The die is biased.
3. **Level of Significance** **:** α = 0.05
4. **The test statistic is** $Z = \frac{x-\mu}{\sigma} = \frac{156-128}{8} = \frac{28}{8} = 3.75$

∴ | Z | = 3.75

∴ | Z | > 1.96

∴ The Null Hypothesis $\boldsymbol{H_0}$ is rejected

5. **Conclusion:** The die is biased.

Large Sample tests:

The following impartment tests to test the significance:

1. **Testing of significance for single proportion**
2. **Testing of significance for difference of proportions**
3. **Testing of significance for single mean**
4. **Testing of significance for difference of means**
5. **Testing of significance for difference of standard deviation's**

Testing of significance for single proportion

(Testing of Hypothesis for single proportion):

Suppose a large random sample of size *n* has a sample proportion *p* of members possessing a certain attribute (proportion of successes). To test the hypothesis that the proportion *P* in the population has a specified value P_0.

1. **Let us set the Null Hypothesis** $\boldsymbol{H_0}: P = P_0$ **.**
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: P \neq P_0$ $(P > P_0 \; or \; P < P_0)$
3. **Level of Significance** **:** α
4. **The test statistic is** $Z = \frac{p - P_0}{S.E \; of \; p} = \frac{p - P_0}{\sqrt{PQ/n}}$ where Q = 1 – P
5. **Conclusion:** If |Z| < Z_α we Accept H_0 .

   If |Z| > Z_α we Reject H_0 at α level.

Limits for proportion P is given by $p \pm 3\sqrt{pq/n}$

Confidence interval for proportion P for large sample at level of significance is given by $P \pm Z_{\alpha/2}\sqrt{PQ/n}$ where Q = 1 – P

$Z_{\alpha/2} = 1.96(95\%)$, $Z_{\alpha/2} = 2.33(98\%)$ and $Z_{\alpha/2} = 2.58(99\%)$

Examples:

1. **In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level significance?**

Solution: Given that n = 1000.

p = sample proportion of rice esters = $\frac{540}{1000} = 0.54$

P = population proportion of rice esters = ½ = 0.5

Q = 1 – P = 0.5.

1. **Null Hypothesis** $\boldsymbol{H_0}$: both rice and wheat are equally in the state.
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: P \neq 0.5$
3. **Level of Significance** : 0.01
4. **The test statistic is** $Z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.54-0.5}{\sqrt{(0.5)(0.5)/1000}} = 2.532$

∴ the calculated value of |Z| =2.532< 2.58

∴ H_0 is accepted at 1% level of significance.

∴ both rice and wheat are equally popular in this state at 1% LOS.

2. **A manufacturer claimed that atleast 95% of the equipment which be supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance.**

Solution: Given that sample size, n = 200.

Number of pieces conforming to specifications =200 – 18= 182

p = Proportion of pieces conforming to specifications = $\frac{182}{200} = 0.91$

P = population proportion = 95% = 0.95

Q = 1 – P = 1 – 0.95= 0.05

1. **Null Hypothesis** $\boldsymbol{H_0}$: The proportion of pieces conforming to Specifications $P = 0.95$
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: P < 0.95$ (Left – tailed test)
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.91-0.95}{\sqrt{(0.95)(0.05)/200}} = \frac{-0.04}{0.0154} = -2.59$

∴ The calculated value of |Z| =2.59 > 1.645

∴ H_0 is rejected at 5% level of significance.

∴ Manufacturer's claim is rejected at 5% level of significance.

3. **In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?**

Solution: Given that n = 600.

Number of smokers =325

p = Sample Proportion of smokers = $\frac{325}{600} = 0.5417$

P = Population proportion of smokers in this city = ½ = 0.5

Q = 1 – P = 1 – 0.5= 0.5

1. **Null Hypothesis** $\boldsymbol{H_0}$: The number of smokers and non – smokers are equal in the city
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: P > 0.5$ (right – tailed test)
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.5417-0.5}{\sqrt{(0.5)(0.5)/600}} = -2.04$

∴ The calculated value of |Z| =2.04 > 1.645

∴ H_0 is rejected at 5% level of significance.

∴ The majority of smokers in the city are non – smokers.

4. **In a hospital 480 females and 520 male babies were born in a week. Do these figures confirm the hypothesis that males and females are born in equal number?**

Solution: Given that n = Total number of birth = 480+520=1000.

p = Proportion of females born = $\frac{480}{1000} = 0.48$

P = Population proportion = ½ = 0.5

Q = 1 – P = 1 – 0.5= 0.5

1. **Null Hypothesis** $\boldsymbol{H_0}$: *P = 0.5*
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: P \neq 0.5$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.48-0.5}{\sqrt{(0.5)(0.5)/1000}} = -1.265$

∴ The calculated value of |Z| =1.265 < 1.96

∴ H_0 is accepted at 5% level of significance.

∴ The males and females are born in equal proportions

5. **20 peoples were attacked by a disease and 18 survived. Will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of the hypothesis that is more at 5% level.**

Solution: Given that n = 20. The number of survived people =18

p = Proportion of survived people = $\frac{18}{20} = 0.9$

P = Population proportion = 0.85

Q = 1 – P = 1 – 0.85= 0.15

1. **Null Hypothesis** $\boldsymbol{H_0}$: *P = 0.85*
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: P > 0.85$ (right – tailed test)
3. **Level of Significance** **:** 0.05
4. **The test statistic is** $Z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.9-0.85}{\sqrt{(0.85)(0.15)/20}} = \frac{0.5}{0.8} = 0.625$

∴ The calculated value of |Z| =0.625 < 1.645

∴ H_0 is accepted at 5% level of significance.

∴ The proportion of the survived people is 0.85

6. Experience had shown that 20% of manufactured product is of the top quality. In one day's production of 400 articles only 50 are of top quality. Test the hypothesis at 0.05 level.

Solution: Given that n = 400, x = 50, and p = $\frac{x}{n} = \frac{50}{400} = 0.125$

P = 20% = 0.2 ⇒ Q = 1 – P = 1 – 0.2 = 0.8

1. **Null Hypothesis** $\boldsymbol{H_0}$: *P = 0.2*
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: P \neq 0.2$
3. **Level of Significance** **:** 0.05
4. **The test statistic is** $Z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.125-0.2}{\sqrt{(0.2)(0.8)/400}} = \frac{-0.075}{0.02} = -3.75$

∴ The calculated value of |Z| = 3.75 >1.96

∴ H_0 is rejected at 5% level of significance.

∴ P = 20% is not correct.

7. A die was thrown 9000 times and of these 3220 yielded a 3 or 4. Is this consistent with the hypothesis that the die was unbiased?

Solution: Given that n = 9000, x = 3220, and p = $\frac{x}{n} = \frac{3220}{9000} = 0.3578$

P = Population proportion of success =P (3 or 4 success) = 2/6 =1/3

∴P = 1/3 ⇒ Q = 1 – P = 2/3

1. **Null Hypothesis** $\boldsymbol{H_0}$: $P = 1/3$
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: P \neq 1/3$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.3578-0.3333}{\sqrt{\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)/9000}} = 4.94$

∴ The calculated value of |Z| = 4.94 >1.96

∴ H_0 is rejected at 5% level of significance.

∴ **The die is biased**

8. In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. Construct a 99% Confidence interval for the corresponding true percentage.

Solution: Given that n = 160, x = 24, and P = $\frac{x}{n} = \frac{24}{160} = 0.15$

∴P = 0.15 ⇒ Q = 1 – 0.15 = 0.85

Now $\sqrt{PQ/n}$ = $\sqrt{(0.15)(0.85)/160}$ = 0.028

∴Confidence interval at 99% level of significance

$$\left(P - 2.58\sqrt{\frac{PQ}{n}},\ P + 2.58\sqrt{\frac{PQ}{n}}\right) = (0.15 - 3(0.028),\ 0.15 +$$

3(0.028)

$$\therefore (0.077\ ,\ 0.222).$$

Confidence interval for the corresponding true percentage

$$\therefore (7.7\ ,\ 22.2).$$

9. Among 900 people in a state 90 are found to be chapatti eaters. Construct a 99% confidence interval for the true proportion.

Solution: Given that n = 900, x = 90, and P = $\frac{x}{n} = \frac{90}{900} = 0.1$

$\therefore$P = 0.1 $\Rightarrow$ Q = 1 – 0.1 = 0.9

Now $\sqrt{PQ/n}$ = $\sqrt{(0.1)(0.9)/900}$ = 0.01

$\therefore$Confidence interval at 99% level of significance

$$\left(P - 3\sqrt{\frac{PQ}{n}},\ P + 3\sqrt{\frac{PQ}{n}}\right) = (0.1 - 3(0.01),\ 0.1 + 3(0.01))$$

$$\therefore (0.07\ ,\ 0.13).$$

10. In a random sample of 500 apples was taken from a large consignment of 60 were found to be bad, obtain the 98% Confidence limits for the percentage of number of bad apples in the consignment

Solution: Given that n = sample size =500, x = 60, and P = $\frac{x}{n} = \frac{60}{500} = 0.12$

∴P = 0.12 ⇒ Q = 1 – 0.12 = 0.88

Now $\sqrt{PQ/n}$ = $\sqrt{(0.12)(0.88)/500}$ = 0.014

∴Confidence interval at 98% level of significance

$$\left(P - 2.33\sqrt{\frac{PQ}{n}},\ P + 2.33\sqrt{\frac{PQ}{n}}\right) = (0.12 - 2.33(0.014), 0.12 + 3(0.014))$$

∴ (0.087 , 0.153).

Thus confidence limits for the percentage of bad apples ∴ (8.7 , 15.3).

11. In a random sample of 100 packages shipped by air fright 13 had some damage. Construct 95% confidence interval for the true proportion of damage packages.

Solution: Given that n = sample size =100, x = 13.

p = Sample proportion of damage packages $= \frac{x}{n} = \frac{13}{100} = 0.13$

∴p = 0.13 ⇒ q = 1 – 0.13 = 0.87

Now $\sqrt{pq/n}$ = $\sqrt{(0.13)(0.87)/100}$ = 0.034

∴Confidence interval at 95% level of significance

$$\left(p - 1.96\sqrt{\frac{pq}{n}},\ p + 1.96\sqrt{\frac{pq}{n}}\right) = (0.13 - 1.96(0.034), 0.13 + 1.96(0.034))$$

∴ (0.063 , 0.197).

Hence the 95% confidence limits for the true proportion of damage packages

∴ (0.063 , 0.197).

Testing of significance for difference of proportions:

(Testing of Hypothesis for difference of proportions):

Let $p_1 \, and \, p_2$ be the sample in two large random samples of sizes $n_1 \, and \, n_2$ proportions drawn from the two populations having proportions $P_1 \, and \, P_2$. To test whether the two population proportions $P_1 \, and \, P_2$ are equal.

1. **Let us set the Null Hypothesis** $\boldsymbol{H_0}: P_1 = P_2$ **.**
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: P_1 \neq P_2$
3. **Level of Significance :** α

Let assume that H_0 is true. Then $P_1 = P_2 = P(say)$ and the sampling distribution of $p_1 - p_2$ is approximately normal with mean = 0 and the standard deviation $= \sqrt{p\,q\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$

4. **The test statistic is** $Z = \frac{p_1 - p_2}{S.E \; of \; p} = \frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where q = 1 – p
5. **Conclusion:** If $|Z| < Z_\alpha$ we Accept H_0 .

   If $|Z| > Z_\alpha$ we Reject H_0 at α level.

Examples:

1. **In a city A, 20% of a random sample of 900 school boys has a certain slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys has the same defect. Is the difference between the proportions significant at 5% level?**

Solution: Given that n_1= size of the first sample = 900

n_2= size of the second sample = 1600

x_1 = 20% of 900 = 180 $\quad\quad$ x_2 = 18.5% of 1600 = 296

$$p_1 = \frac{x_1}{n_1} = \frac{180}{900} = 0.2 \qquad p_2 = \frac{x_2}{n_2} = \frac{296}{1600} = 0.185$$

$$\text{Then } p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{180+296}{900+1600} = \frac{476}{2500} = 0.19$$

q = 1 – p = 1 – 0.19 = 0.81

1. **Null Hypothesis** $H_0: p_1 = p_2$.
2. **Alternative Hypothesis is** $H_1: p_1 \neq p_2$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.2-0.185}{\sqrt{(0.19)(0.81)\left(\frac{1}{900}+\frac{1}{1600}\right)}} = 0.918$

∴ The calculated value of |Z| = 0.918 <1.96

∴ H_0 is accepted at 5% level of significance.

∴ There no significant difference between the proportions.

2. In a random sample of 1000 persons from town A, 400 are found to consumers of wheat. In sample of 800 from town B, 400 are found to consumers of wheat. Does the data reveal a significant difference between town A and town B, so for as the proportion wheat consumers are concerned.

Solution: Given that n_1= sample size of town A= *1000*

n_2= sample size of town B = *800*

x_1 = Number of consumers of wheat from town A= *400*

x_2 = Number of consumers of wheat from town B= *400*

p_1 = Proportion of consumers of wheat in town $A = \frac{x_1}{n_1} = \frac{400}{1000} = 0.40$

p_2 = Proportion of consumers of wheat in town $B = \frac{x_2}{n_2} = \frac{400}{800} = 0.50$

Then $p = \frac{n_1 p_1 + n_2 p_2}{n_1+n_2} = \frac{x_1 + x_2}{n_1+n_2} = \frac{400+400}{1000+800} = \frac{800}{1800} = \frac{4}{9}$ and q = 1 – p = 1 – $\frac{4}{9} = \frac{5}{9}$

1. **Null Hypothesis** $\boldsymbol{H_0}: p_1 = p_2$ **.**
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: p_1 \neq p_2$
3. **Level of Significance** **:** 0.05
4. **The test statistic is** $Z = \frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.4-0.5}{\sqrt{\left(\frac{4}{9}\right)\left(\frac{5}{9}\right)\left(\frac{1}{1000}+\frac{1}{800}\right)}} = -4.247$

∴ The calculated value of |Z| = 4.242 >1.96

∴ H_0 is rejected at 5% level of significance.

∴ There is a significant difference between town *A* and town *B.*

3. Random sample of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the Hypothesis that proportions of men and women in favour of the proposal are same as 5% level.

Solution: Given sample sizes, n_1=*400* and n_2= *600*, x_1=*200* , x_2=*325*

p_1 = Proportion of men = $\frac{x_1}{n_1} = \frac{200}{400} = 0.5$

p_2 = Proportion of women = $\frac{x_2}{n_2} = \frac{325}{600} = 0.541$

Then $p = \frac{n_1 p_1 + n_2 p_2}{n_1+n_2} = \frac{x_1 + x_2}{n_1+n_2} = \frac{200+325}{400+600} = \frac{525}{1000} = 0.525$

q = 1 – p =1 – 0.525= 0.475

1. **Null Hypothesis** $\boldsymbol{H_0}: p_1 = p_2$ **.**
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: p_1 \neq p_2$
3. **Level of Significance :** 0.05
4. **The test statistic is** $Z = \frac{p_1 - p_2}{\sqrt{p\, q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.5-0.541}{\sqrt{(0.525)(0.475)\left(\frac{1}{400}+\frac{1}{600}\right)}} = -1.28$

∴ The calculated value of |Z| = 1.28 <1.96

∴ H_0 is accepted at 5% level of significance.

∴ There no significant difference between men and women as far as proposal of flyover is concerned.

4. On the basis of their total scores, 200 candidates of a civil service examination are divided in to two groups, the upper 30% and the remaining 70%. Consider the first question of the examination. Among the first group, 40 had the correct answer, where as among the second group, 80 had the correct answer. On the basis of these results, can one conclude that the first question is not good at discriminating ability of the type being examined here?

Solution: Given that $n_1 = 60$, $n_2 = 140$, $x_1 = 40$ and $x_2 = 80$

$$p_1 = \frac{x_1}{n_1} = \frac{40}{60} = 0.667 \qquad p_2 = \frac{x_2}{n_2} = \frac{80}{140} = 0.571$$

$$\text{Then } p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{40+80}{60+140} = \frac{120}{200} = 0.6$$

$$q = 1 - p = 1 - 0.6 = 0.4$$

1. **Null Hypothesis** $H_0 : p_1 = p_2$.
2. **Alternative Hypothesis is** $H_1 : p_1 \neq p_2$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.667-0.571}{\sqrt{(0.6)(0.4)\left(\frac{1}{60}+\frac{1}{140}\right)}} = 1.27$

∴ The calculated value of |Z| = 1.27 <1.96

∴ H_0 is accepted at 5% level of significance.

∴The first question is good enough in discriminating ability of the students of both groups.

5. In a sample of 600 students of a certain college 400 are found to use ball pens. In another college from a sample of 900 students 450 were found to use ball pens. Test whether two colleges are significantly different with respect to the habit of using ball pens.

Solution: Given sample sizes, n_1=*600* and n_2= *900*, x_1=*400* , x_2=*450*

p_1 = Proportion of students who use ball pens in first college = $\frac{x_1}{n_1} = \frac{400}{600} = \frac{2}{3}$

p_2 = Proportion of students who use ball pens in second college = $\frac{x_2}{n_2} = \frac{450}{900} = \frac{1}{2}$

Then $p = \frac{n_1p_1 + n_2p_2}{n_1+n_2} = \frac{x_1 + x_2}{n_1+n_2} = \frac{400+450}{600+900} = \frac{850}{1500} = 0.57$

$$q = 1 - p = 1 - 0.57 = 0.43$$

1. **Null Hypothesis** $\boldsymbol{H_0}: p_1 = p_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: p_1 \neq p_2$
3. **Level of Significance** : 0.05
4. **The test statistic** $Z = \frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.667 - 0.5}{\sqrt{(0.57)(0.43)\left(\frac{1}{600}+\frac{1}{900}\right)}} = 6.39$

∴ The calculated value of |Z| = 6.39 >1.96

∴ H_0 is rejected at 5% level of significance.

∴ There is a significant difference between two colleges as far as using ball pens habit is concerned.

6. 100 articles from a factory are examined and 10 are found to be defective, 500 similar articles from second factory are found to be 15 defective. Test the significance between differences of two proportions at 5% level.

Solution: Given sample sizes, n_1=*100* and n_2= *500*, x_1=*10* , x_2=*15*

p_1 = Proportion of defective articles in first factory $= \frac{x_1}{n_1} = \frac{10}{100} = 0.1$

p_2 = Proportion of defective articles in first factory $= \frac{x_2}{n_2} = \frac{15}{500} = 0.03$

Then $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{10+15}{100+500} = \frac{25}{600} = 0.042$

q = 1 – p =1 – 0.042= 0.958

1. **Null Hypothesis** $\boldsymbol{H_0}: p_1 = p_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: p_1 \neq p_2$
3. **Level of Significance :** 0.05
4. **The test statistic** $Z = \frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.1-0.03}{\sqrt{(0.042)(0.958)\left(\frac{1}{100}+\frac{1}{500}\right)}} = 3.18$

∴ The calculated value of |Z| = 3.18 >1.96

∴ H_0 is rejected at 5% level of significance.

∴ There is a significant difference between the two proportions.

7. A machine produced 20 defective articles in batch of 400. After overhauling it produced 10 defective in a batch of 300. Has the machine being improved after overhauling?

Solution: Given sample sizes, n_1=*400* and n_2= *300,* x_1=*20* , x_2=*10*

p_1 = Proportion of defective articles in the first batch $= \frac{x_1}{n_1} = \frac{20}{400} = 0.05$

p_2 = Proportion of defective articles in the second batch $= \frac{x_2}{n_2} = \frac{10}{300} = 0.033$

Then $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{20+10}{400+300} = \frac{30}{700} = 0.043$

$$q = 1 - p = 1 - 0.043 = 0.957$$

1. **Null Hypothesis** $H_0: p_1 = p_2$.
2. **Alternative Hypothesis is** $H_1: p_1 > p_2$ (right – tailed test)
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.05-0.033}{\sqrt{(0.043)(0.957)\left(\frac{1}{400}+\frac{1}{300}\right)}} = 1.0844$

∴ The calculated value of |Z| = 1.0844 <1.645

∴ H_0 is accepted at 5% level of significance.

∴ There is no significant difference. Hence, the machine has not improved after overhauling.

8. **Before an increase on excise duty on tea 500 people out of a sample 900 found to have the habit of having tea. After an increase on excise duty 250 are have the habit of having tea among 1100. Is there any decrease in the consumption of tea? Test the hypothesis at 5% level.**

Solution: Here n_1=*900* and n_2= *1100*, x_1=*500* , x_2=*250*

p_1 = Proportion of people who have the habit of tea before an increase on excise duty = $\frac{x_1}{n_1} = \frac{500}{900} = 0.556$

p_2 = Proportion of people who have the habit of tea after an increase on excise duty = $\frac{x_2}{n_2} = \frac{250}{1100} = 0.227$

Then $p = \frac{n_1p_1 + n_2p_2}{n_1+n_2} = \frac{x_1+x_2}{n_1+n_2} = \frac{500+250}{900+1100} = \frac{750}{2000} = 0.375$ and q = 1 – p =0.625

1. **Null Hypothesis** $\boldsymbol{H_0}: p_1 = p_2$.
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: p_1 > p_2$ (right – tailed test)
3. **Level of Significance** **:** 0.05
4. **The test statistic is** $Z = \frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.556-0.227}{\sqrt{(0.375)(0.625)\left(\frac{1}{900}+\frac{1}{1100}\right)}} = 15.30$

∴ The calculated value of |Z| = 15.30 >1.645

∴ H_0 is rejected at 5% level of significance.

∴ There is a significant difference in the consumption of tea before and after increase on excise duty on tea.

9. In an investigation on the machine performance the following results are obtained.

	No. of units inspected	No. of units defective
Machine -1	375	17
Machine -2	450	22

Test whether there is any significant performance of two machines at 5% level.

Solution: Here n_1=*375* and n_2= *450,* x_1= *17,* x_2=*22*

p_1 = Proportion of defective units in the machine 1 $= \frac{x_1}{n_1} = \frac{17}{375} = 0.045$

p_2 = Proportion of defective units in the machine 2 $= \frac{x_2}{n_2} = \frac{22}{450} = 0.049$

Then $p = \frac{n_1p_1 + n_2p_2}{n_1+n_2} = \frac{x_1 + x_2}{n_1+n_2} = \frac{17+22}{375+450} = \frac{39}{825} = 0.047$

q = 1 – p =1 – 0.047= 0.953

1. **Null Hypothesis** $\boldsymbol{H_0}: p_1 = p_2$.
2. **Alternative Hypothesis is** $\boldsymbol{H_1}: p_1 > p_2$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{p_1 - p_2}{\sqrt{p\, q\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.045-0.049}{\sqrt{(0.047)(0.953)\left(\frac{1}{375}+\frac{1}{450}\right)}} = -0.267$

∴ The calculated value of |Z| = 0.267 <1.645

∴ H_0 is accepted at 5% level of significance.

∴ There is no significant difference in performance of two machines.

10. A sample poll of 300 voters from district A and 200voters from district B showed that 56% and 48% respectively, were in favour of a given candidates. At 5% level of significance, test the hypothesis that there is a difference in the district.

Solution: Here n_1=*300* and n_2=*200,*

P_1 = Proportion of district A = 56% = 0.56 $\Rightarrow Q_1 = 1 - P_1 = 1 - 0.56 = 0.44$

P_2 = Proportion of district B = 48% = 0.48 $\Rightarrow Q_2 = 1 - P_2 = 1 - 0.48 = 0.52$

1. **Null Hypothesis** $\boldsymbol{H_0}: P_1 = P_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: P_1 \neq P_2$
3. **Level of Significance** : 0.05
4. **The test statistic** $Z = \frac{P_1 - P_2}{\sqrt{\left(\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_1}{n_2}\right)}} = \frac{0.56 - 0.48}{\sqrt{\left(\frac{(0.56)(0.44)}{300} + \frac{(0.48)(0.52}{200}\right)}} = 1.758$

∴ The calculated value of |Z| = 1.758 <1.96

∴ H_0 is accepted at 5% level of significance.

∴ There is no significant difference between the two districts towards the favour of a given candidate.

11. In two large populations, there are 30% and 25% respectively of fair haired peoples. Is this deference likely to be hidden in samples of 1200 and 900 respectively from the two populations?

Solution: Here n_1=*1200* and n_2=*900,*

P_1 = Proportion of fair haired peoples *in the first population* = 30% = 0.3

$\Rightarrow \quad Q_1 = 1 - P_1 = 1 - 0.3 = 0.7$

P_2 = Proportion of fair haired peoples *in the second population* = 25% = 0.25

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

1. **Null Hypothesis** $\boldsymbol{H_0}: P_1 = P_2$ **.**
2. **Alternative Hypothesis** $\boldsymbol{H_1}: P_1 \neq P_2$
3. **Level of Significance** **:** 0.05
4. **The test statistic** $Z = \frac{P_1 - P_2}{\sqrt{\left(\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_1}{n_2}\right)}} = \frac{0.3 - 0.25}{\sqrt{\left(\frac{(0.3)(0.7)}{1200} + \frac{(0.25)(0.75)}{900}\right)}} = 2.56$

∴ The calculated value of |Z| = 2.56 >1.96

∴ H_0 is rejected at 5% level of significance.

∴ There is a significant difference between the two populations.

Testing of Significance for a Single Mean: Large Sample

(Testing of Hypothesis for a Single Mean): Large Samples

Let a random sample of size n (≥ 30) has the sample mean $\bar{x}$, test the hypothesis that the population mean μ has a specified value μ_0.

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu = \mu_0$.
2. **The Alternative Hypothesis** $\boldsymbol{H_1}: \mu \neq \mu_0$ $(\mu > \mu_0 \; or \; \mu < \mu_0)$
3. **Level of Significance** : α
4. **The test statistic** $Z = \frac{\bar{x}-\mu}{S.E \; of \; \bar{x}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ (population S.D σ is known)

   $Z = \frac{\bar{x}-\mu}{S.E \; of \; \bar{x}} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ (Sample S.D s is known)
5. **Conclusion:** If |Z| < Z_α we Accept H_0 .

   If |Z| > Z_α we reject H_0 at α level.

For two – tailed test:

i) If |Z| < 1.96 accept H_0 at 5% Level of significance

ii) If |Z| > 1.96 reject H_0 at 5% Level of significance

iii) If |Z| < 2.33 accept H_0 at 2% Level of significance

iv) If |Z| > 2.33 reject H_0 at 2% Level of significance

v) If |Z| < 2.58 accept H_0 at 1% Level of significance

vi) If |Z| > 2.58 reject H_0 at 1% Level of significance

vii) If |Z| < 1.645 accept H_0 at 10% Level of significance

viii) If |Z| > 1.645 reject H_0 at 10% Level of significance

For Single – tailed test (right – tailed test or left – tailed test)

i) If |Z| < 1.645 accept H_0 at 5% Level of significance

ii) If |Z| > 1.645 reject H_0 at 5% Level of significance

iii) If |Z| < 2.33 accept H_0 at 1% Level of significance

iv) If |Z| > 2.33 reject t H_0 at 1% Level of significance

v) If |Z| < 1.28 accept H_0 at 10% Level of significance

vi) If |Z| > 1.28 reject t H_0 at 10% Level of significance

confidence limits for a single Mean:

i) 95% confidence limits = $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

ii) 98% confidence limits = $\bar{x} \pm 2.33 \frac{\sigma}{\sqrt{n}}$

iii) 99% confidence limits = $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$

Examples:

1. **A Sample of 64 students have a mean weight of 70kgs. Can this be regarded as sample from a population with mean weight 56 kgs and the standard deviation 25 kgs.**

Solution: Given that $\bar{x}$ = mean of the sample = 70 kgs.

μ = mean of the population = 56 kgs.

σ = standard deviation of the population = 25 kgs.

n = size of the sample = 64

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu = 56$ **.**
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: \mu \neq 56$
3. **Level of Significance** **:** 0.05
4. **The test statistic is** $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{70-56}{25/\sqrt{64}} = 4.48$

   ∴ The calculated value of |Z| = 4.48 >1.96

   ∴ H_0 is rejected at 5% level of significance.

∴ the sample is not from the population whose mean weight is 56 kgs .

2. **A Sample of 400 items is taken from a population whose standard deviation 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also find 95% confidence interval for the population.**

Solution: Given that $\bar{x} = \text{mean of the sample} = 40$

$\mu = \text{mean of the population} = 38.\ n = \text{size of the sample} = 400$

$\sigma = \text{standard deviation of the population} = 10$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu = 38$ **.**
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: \mu \neq 38$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{40-38}{10/\sqrt{400}} = 4$

∴ the calculated value of |Z| = 4 >1.96

∴ H_0 is rejected at 5% level of significance.

∴ the sample is not from the population whose mean is 38.

95% confidence limits = $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 40 \pm 1.96\left(\frac{10}{\sqrt{400}}\right) = 40 \pm 0.98$

95% confidence interval = $(39.02, 40.98)$.

3. **In a random Sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis μ = 32.6 minutes in favour of alternative hypothesis μ > 32.6 at α = 0.005 level of significance.**

Solution: Given that $\bar{x} =$ mean of the sample $= 33.8$

$\mu =$ mean of the population $= 32.6.$

$n =$ size of the sample $= 60$

$\sigma =$ standard deviation of the population $= 6.1$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu = 32.6$ **.**
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: \mu > 32.6$
3. **Level of Significance** **:** 0.025
4. **The test statistic is** $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{33.8-32.6}{6.1/\sqrt{60}} = 1.5238$

∴ the calculated value of |Z| = 1.5238 < 2.58

∴ H_0 is accepted at 1% level of significance.

∴ the null hypothesis H_0 is accepted.

4. The mean life time of a random Sample of 100 light tubes produced by a company is found to be 1560 hrs with a population standard deviation of 90 hrs. Test the hypothesis for $\alpha = 0.05$ that the mean life time of the tubes produced by the company is 1580 hrs.

Solution: Given that $\bar{x} =$ mean of the sample $= 1560$ hrs

$\mu =$ mean of the population $= 1580$ hrs

$n =$ size of the sample $= 100$

$\sigma =$ standard deviation of the population $= 90$ hrs

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu = 1580$
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: \mu \neq 1580$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{1560-1580}{90/\sqrt{100}} = -2.22$

$\therefore$ the calculated value of $|Z| = 2.22 > 1.96$

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

$\therefore$ the null hypothesis H_0 is rejected.

5. **An ambulance service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the claim at 0.05 level significance.**

Solution: Given that $\bar{x} =$ mean of the sample $= 11$ minutes

$\mu = 10$ minutes. $n =$ size of the sample $= 36$

Variance $\sigma^2 = 16$ $\Rightarrow$ standard deviation $\sigma = 4$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu = 10$
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: \mu < 10$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{11-10}{4/\sqrt{36}} = 1.5$

$\therefore$ the calculated value of |Z| = 1.5 < 1.645

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

$\therefore$ the null hypothesis H_0 is accepted.

6. **It is claimed that a random sample of 49 types has a mean life of 15200 km. This sample was drawn from a population whose mean is 15150 km and a standard deviation of 1200km. Test the significance at 0.05 level.**

Solution: Given that $\bar{x} =$ mean of the sample $= 15200$

$\mu =$ mean of the population $= 15150$

$n =$ size of the sample $= 49$

$\sigma =$ standard deviation of the population $= 1200$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu = 15150$
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: \mu \neq 15150$
3. **Level of Significance** : 0.05
4. **The test statistic is** $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{15200-15150}{1200/\sqrt{49}} = 0.2917$

∴ the calculated value of |Z| = 0.2917 < 1.96

∴ H_0 is accepted at 5% level of significance.

∴ the null hypothesis H_0 is accepted.

7. **A sample of 900 members has a mean of 3.4 cms and standard deviation 2.61 cms. Is this sample has be taken from a large population of mean 3.25 cms and standard deviation 2.61 cms. If the population is normal and its mean is unknown find the 95% confidence limits of true mean.**

Solution: Given that $\bar{x} =$ mean of the sample $= 3.4$ cms.

$\mu =$ mean of the population $=3.25$ cms $\quad n =$ size of the sample $= 900$

$\sigma =$ standard deviation of the population$= 2.61$

s $=$ standard deviation of the sample $= 2.61$

1. **N ull Hypothesis** $\boldsymbol{H_0}: \mu = 3.25$
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: \mu \neq 3.25$
3. **Level of Significance** **:** 0.05
4. **The test statistic is** $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{3.4-3.25}{2.61/\sqrt{900}} = 1.724$

$\therefore$ the calculated value of |Z| = 1.724 < 1.96

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

$\therefore$ the null hypothesis H_0 is accepted.

95% confidence limits = $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96\left(\frac{2.61}{\sqrt{900}}\right) = 3.4 \pm 0.1705$

95% confidence interval = $(3.2295, 3.57)$.

8. An insurance agent has claimed that the average age of policy holders who issue through him is less than the average for all agents which is 30.5 years. A random sample of 100 policy holders who had issued through him gave the following age distribution.

age	16 – 20	21 – 25	26 – 30	31 – 35	36 – 40
No. of persons	12	22	20	30	16

Calculate the mean and standard deviation of the distribution and use these values to test his claim at 5% level of significance.

Solution: Let d = (x –A) / h

C.I	16 – 20	21 – 25	26 – 30	31 – 35	36 – 40
C.I	15.5 – 20.5	20.5 – 25.5	25.5 – 30.5	30.5– 35.5	35.5 – 40.5
f	12	22	20	30	16
x	18	23	28(A)	33	38
d	–2	–1	0	1	2
fd	–24	–22	0	30	32
fd^2	48	22	0	30	64

$$\bar{x} = A + \frac{\sum fd}{\sum f} \times c = 28 + \left(\frac{16}{100}\right) \times 5 = 28 + 0.8 = 28.8$$

$$s^2 = \frac{\sum fd^2}{\sum f} \times c^2 - \left(\frac{\sum fd}{\sum f}\right)^2 c^2 = 25\left[\left(\frac{164}{100}\right) - (\frac{16}{100})^2\right] = 40.36$$

$s = \sqrt{40.36} = 6.35$

Given that $\bar{x} =$ mean of the sample = 28.8

$\mu =$ mean of the population =30.5 cms $n =$ size of the sample = 100

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu = 30.5$
2. **The Alternative Hypothesis is** $\boldsymbol{H_1}: \mu < 30.5$ (left tailed test)
3. **Level of Significance** **:** 0.05
4. **The test statistic is** $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{28.8-30.5}{6.35/\sqrt{100}} = -2.677$

∴ the calculated value of |Z| = 2.677 > 1.645

∴ H_0 is rejected at 5% level of significance.

∴ the null hypothesis H_0 is accepted.

∴ the sample is not drawn from a population with mean 30.5 years.

Test of significance for Difference of means of two large samples:

(Test for equality of two means of two large samples)

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means of two independent large sample sizes n_1 and n_2 drawn from two populations having means μ_1 and μ_2 and standard deviations σ_1 and σ_2. To test whether the two population means are equal.

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$ **.**
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 \neq \mu_2$
3. **Level of Significance** **:** α
4. **The test statistic is** $Z = \frac{\bar{x}_1 - \bar{x}_2}{S.E\ of\ (\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

If samples have been drawn from the population with common standard deviation σ, then $\sigma_1^2 = \sigma_2^2 = \sigma$. Hence

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If σ is not known we can use an estimate of σ^2 given by $\sigma^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$

If the two samples are drawn from two population with unknown variances σ_1^2 and σ_2^2 , then σ_1^2 and σ_2^2 can be replaced by sample variances S_1^2 and S_2^2 provided both sample sizes n_1 and n_2 are large.

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

5. **Conclusion:** If |Z| < Z_α we Accept H_0 .

If |Z| > Z_α we reject H_0 at α level.

(i) If |Z| < 1.96 Accept H_0 at 5% level of significance

(ii) If |Z| > 1.96 Reject H_0 at 5% level of significance

(iii) If |Z| < 2.58 Accept H_0 at 1% level of significance

(iv) If |Z| > 2.58 Reject H_0 at 1% level of significance

(v) If |Z| < 2.33 Accept H_0 at 2% level of significance

(vi) If |Z| > 2.33 Reject H_0 at 2% level of significance

(vii) If |Z| < 1.645 Accept H_0 at 10% level of significance

(viii) If |Z| > 1.645 Reject H_0 at 10% level of significance

If |Z| > 3 then either the samples have not been drawn from the same population.

Examples:

1. **The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation of 2.5 inches?**

Solution: Let μ_1 and μ_2 be the means of the two populations.

Given that $n_1 = 1000$, $n_2 = 2000$, $\bar{x}_1 = 67.5$, $\bar{x}_2$=68.0 and $\sigma = 2.5$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$ **.**
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 \neq \mu_2$
3. **Level of Significance** **:** 0.05
4. **The test statistic** $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{67.5 - 68}{\sqrt{(2.5)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}} = -5.16$

$\therefore$ the calculated value of |Z| = 5.16 > 1.96

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

$\therefore$ The samples are not drawn from the same population of S.D 2.5 inches.

2. **The average mark scored by 32 boys is 72 with standard deviation of 8. While that for 36 girls is 70 with standard deviation of 6. Does this indicate that the boys performance better than girls at level of significance 0.05?**

Solution: Let μ_1 and μ_2 be the means of the two populations.

Given that $n_1 = 32, n_2 = 36, \bar{x}_1 = 72, \bar{x}_2$=70, $\sigma_1 = s_1 = 8$ and $\sigma_2 = s_2 = 6$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 \neq \mu_2$
3. **Level of Significance** : 0.05
4. **The test statistic** $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \frac{72-70}{\sqrt{\left(\frac{64}{32} + \frac{36}{36}\right)}} = 1.1547$

∴ the calculated value of |Z| = 1.1547 < 1.96

∴ H_0 is accepted at 5% level of significance.

∴ The performance of boys and girls is the same.

3. **Two types of new cars produced in U.S.A are tested for petrol mileage, one sample is consisting of 42 cars averaged 15 kmpl while the other sample consisting of 80 cars averaged 11.5 kmpl with population variances as $\boldsymbol{\sigma_1^2 = 2.0}$ and $\boldsymbol{\sigma_2^2 = 1.5}$ respectively. Test weather there any significance difference in the petrol consumption of these two types cars at 1% level.**

Solution: Let μ_1 and μ_2 be the means of the two populations.

Given that $n_1 = 42, n_2 = 80, \bar{x}_1 = 15, \bar{x}_2$=11.5, $\sigma_1^2 = 2.0$ and $\sigma_2^2 = 1.5$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 \neq \mu_2$
3. **Level of Significance** : 0.01
4. **The test statistic** $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} = \frac{15-11.5}{\sqrt{\left(\frac{2}{42} + \frac{1.5}{80}\right)}} = 13.587$

$\therefore$ the calculated value of |Z| = 13.587 > 2.58

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

$\therefore$ There is a significance difference in the petrol consumption.

4. A researcher wants to know the intelligence of students in a school. He selected two groups of students. In the first group there are 150 students having mean IQ of 75 with standard deviation of 15 in the second group there are 250 students having mean IQ of 70 with standard deviation of 20.

Solution: Let μ_1 and μ_2 be the means of the two populations.

Here $n_1 = 150, n_2 = 250, \bar{x}_1 = 75, \bar{x}_2=70, \sigma_1 = s_1 = 15$ and $\sigma_2 = s_2 = 20$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 \neq \mu_2$
3. **Level of Significance** : 0.05
4. **The test statistic** $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \frac{75-70}{\sqrt{\left(\frac{225}{150} + \frac{400}{250}\right)}} = 2.839$

$\therefore$ the calculated value of |Z| = 2.839 > 1.96

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

$\therefore$ The groups have not been taken from the same population.

5. **The research investigator is interested in studying whether there is a significance difference in the salaries of MBA grades in two metropolitan cities. A random sample of size 100 from Mumbai yielded on average income of Rs. 20,150. Another random sample of 60 from Chennai results in an average income of Rs. 20,250. If the variances of both the populations are given as $\sigma_1^2 = Rs40,000$ and $\sigma_2^2 = Rs\ 32,400$ respectively.**

Solution: Let μ_1 and μ_2 be the means of the two populations.

Given that $n_1 = 100, n_2 = 60$, $\bar{x}_1 = 20{,}150$, $\bar{x}_2 = 20{,}250$,

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

1. **Null Hypothesis** $H_0: \mu_1 = \mu_2$.
2. **Alternative Hypothesis** $H_1: \mu_1 \neq \mu_2$
3. **Level of Significance** : 0.05
4. **The test statistic** $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} = \frac{20150 - 20250}{\sqrt{\left(\frac{40000}{100} + \frac{32400}{60}\right)}} = -3.26$

∴ the calculated value of |Z| = 3.26 > 1.96

∴ H_0 is rejected at 5% level of significance.

∴ There is a significance difference in the salaries of MBA grades in two metropolitan cities.

6. The mean height of 50 male students who participated in sports is 68.2 inches with a standard deviation of 2.5. The mean height of 50 male students who have not participated in sports is 67.2 inches with a standard deviation of 2.8. Test the hypothesis that the height of students who participated in sports is more than the students who have not participated in sports.

Solution: Let μ_1 and μ_2 be the means of the two populations.

Here $n_1 = 50, n_2 = 50, \bar{x}_1 = 68.2, \bar{x}_2$=67.2, $\sigma_1 = s_1 = 2.5$ and $\sigma_2 = s_2 = 2.8$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 > \mu_2$
3. **Level of Significance** : 0.05
4. **The test statistic** $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \frac{68.2-67.2}{\sqrt{\left(\frac{6.5}{50} + \frac{7.84}{50}\right)}} = 1.88$

∴ the calculated value of |Z| = 1.88 >1.645

∴ H_0 is rejected at 5% level of significance.

∴ There is a significant difference in the height.

7. A simple sample of the heights of 6400 Englishman has a mean of 67.85 inches and a standard deviation of 2.56 inches while a simple sample of the heights of 1600 Australians has a mean of 68.55 inches and a standard deviation of 2.52 inches. Do the data indicate the Australians are on the average taller than the Englishman at 5% level of significance?

Solution: Let μ_1 and μ_2 be the means of the two populations.

Here $n_1 = 6400,\ n_2 = 1600,\ \bar{x}_1 = 67.85,\ \bar{x}_2 = 68.55,$

$\sigma_1 = s_1 = 2.56$ and $\sigma_2 = s_2 = 2.52$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 > \mu_2$
3. **Level of Significance** : 0.05
4. **The test statistic** $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \dfrac{67.85 - 68.55}{\sqrt{\left(\frac{6.5536}{6400} + \frac{6.35}{1600}\right)}} = -9.9$

∴ the calculated value of $|Z| = 9.9 > 1.645$

∴ H_0 is rejected at 5% level of significance.

∴ Australians are taller than the Englishman

8. Sample of students were drawn from two universities and from their weights in kilograms, mean and standard deviations are calculated and show bellow. Make a large sample test to test the significance of the difference between the means and standard deviations.

	Mean	S.D	Size of the sample
University - A	55	10	400
University - B	57	15	100

Solution: Let μ_1 and μ_2 be the means of the two populations.

Here $n_1 = 400,\ n_2 = 100,\ \bar{x}_1 = 55,\ \bar{x}_2 = 57,\ s_1 = 10$ and $s_2 = 15$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$ **.**
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 \neq \mu_2$
3. **Level of Significance** **:** 0.05
4. **The test statistic** $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \dfrac{55 - 57}{\sqrt{\left(\frac{100}{400} + \frac{225}{100}\right)}} = -1.26$

$\therefore$ The calculated value of |Z| = 1.26 < 1.96

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

There is no significance of the difference between the means

To test the significance of the difference between the S.D's:

1. **Null Hypothesis** $\boldsymbol{H_0}: \sigma_1 = \sigma_2$.
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \sigma_1 \neq \sigma_2$
3. **Level of Significance** : 0.05
4. **The test statistic** $Z = \frac{s_1 - s_2}{\sqrt{\left(\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}\right)}} = \frac{10-15}{\sqrt{\left(\frac{100}{800} + \frac{225}{200}\right)}} = -4.4721$

$\therefore$ The calculated value of |Z| = 4.4721 > 1.96

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

There is a significance of the difference between the S.D's

9. Random samples drawn from two countries gave the following data relating to the heights of adult males.

	Mean	S.D	Size of the sample
Country – A	67.42	2.58	1000
Country – B	67.25	2.50	1200

(i) Is the difference between the means significant?

(ii) Is the difference between the standard deviations significant?

Solution:

(i) **To test the significance of the difference between means**

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

Here $n_1 = 1000$, $n_2 = 1200$, $\bar{x}_1 = 67.42$, $\bar{x}_2$=67.25, $s_1 = 2.58$ and $s_2 = 2.5$

1. **Null Hypothesis** $\boldsymbol{H_0}: \mu_1 = \mu_2$ **.**
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \mu_1 \neq \mu_2$
3. **Level of Significance** **:** 0.05
4. **The test statistic** $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \frac{67.42 - 67.25}{\sqrt{\left(\frac{6.6564}{1000} + \frac{6.25}{1200}\right)}}$ =1.56

∴ The calculated value of |Z| = 1.56 < 1.96

∴ H_0 is accepted at 5% level of significance.

There is no significance of the difference between the means

To test the significance of the difference between the S.D's:

1. **Null Hypothesis** $\boldsymbol{H_0}: \sigma_1 = \sigma_2$ **.**
2. **Alternative Hypothesis** $\boldsymbol{H_1}: \sigma_1 \neq \sigma_2$
3. **Level of Significance** **:** 0.05
4. **The test statistic** $Z = \frac{s_1 - s_2}{\sqrt{\left(\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}\right)}} = \frac{2.58 - 2.50}{\sqrt{\left(\frac{6.6564}{2000} + \frac{6.25}{2400}\right)}} = 1.0387$

∴ The calculated value of |Z| = 1.0387 < 1.96

∴ H_0 is accepted at 5% level of significance.

There is no significance of the difference between the S.D's