B.Tech. Sem : III (IT) (MU)

Subject Name: Probability and Statistics

Marwadi University

Department of CE-AI

Unit no:- 4 & 5

Unit title:
Applied Statistics

Subject name :
PS(01AI1301)

Instructor:
Vipul Ladva

TERMS RELATED TO TESTS OF HYPOTHESIS

Population

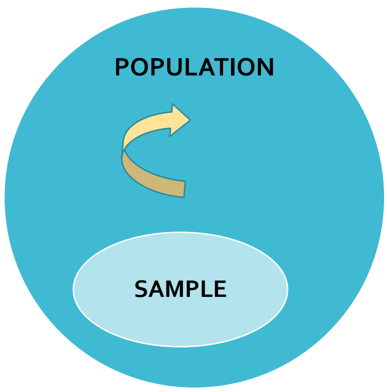
The group (say of size N) of units or items or individuals or observations or objects forming a subject matter of statistical investigation for their various characteristics is known as population.

The population may be finite or infinite. It may be real or hypothetical.

Sample

A finite subset of population is selected by some scientific method with a view of estimating population characteristics is known as sample.

The size of the sample is denoted by n. When $n \geq 30$, the sample size is said to be large and when $n < 30$, the sample size is said to be small.



TERMS RELATED TO TESTS OF HYPOTHESIS

Parameter (The statistical constants of population)	Mean(μ)
	Standard deviation(σ)
	Correlation coefficient(ρ)
	Population proportion (P)
Statistic (The statistical constants of Sample)	Mean($\bar{x}$)
	Standard deviation(s)
	Correlation coefficient(r)
	Sample proportion (p)

The values of parameters for a population does not change.

The values of statistics vary from sample to sample.

TERMS RELATED TO TESTS OF HYPOTHESIS

Hypothesis
A claim that we want to test or investigate with the help of sample data.

Null Hypothesis (H_0)
Currently accepted value of a parameter.

Alternative Hypothesis (Research Hypothesis) (H_1 or H_a)
It involves the claim to be tested.

Here note that Null and alternative hypothesis are two mutually exclusive statements for a population. They are called mathematical opposites of each other.

TERMS RELATED TO TESTS OF HYPOTHESIS

For example: From past data it is known that the average weight of MU students is 55kg. But the sports faculty claims that the average weight of MU students is no longer 55kg.

Here,

Null Hypothesis (H_0): $\mu = 55\text{kg}$

(Currently accepted value of mean)

Alternative Hypothesis (Research Hypothesis) (H_1): $\mu \neq 55\text{kg}$ ($H_1 > H_0$ or $H_1 < H_0$)

(The claim to be tested)

Here Possible outcomes of this test are:

1. Reject the Null hypothesis
2. Fail to reject the Null hypothesis

TERMS RELATED TO TESTS OF HYPOTHESIS

Two-sided and One-sided Hypothesis:

The alternative hypothesis can be Two-sided or One-sided.

Two-sided alternative hypothesis

- Also called non-directional hypothesis
- It is used to determine whether the population parameter is either greater than or less than the hypothesized value.
- In previous example H_1: $\mu \neq 55\text{kg}$ ($H_1 > H_0$ or $H_1 < H_0$) is a two-sided alternative hypothesis.

One-sided alternative hypothesis

- Also called directional hypothesis
- It is used to determine whether the population parameter differs from the hypothesized value in specific direction.
- In previous example H_1: $\mu > 55\text{kg}$ is a one-sided alternative hypothesis.

TERMS RELATED TO TESTS OF HYPOTHESIS

Type-I and type-II error:

No hypothesis test is 100% accurate. All the tests are based on probability so there are always chances of errors. For the hypothesis tests there are two types of errors.

Nature of null hypothesis H_o	Accept H_o	Reject H_o
H_o is true	Correct decision (probability = $1 - \alpha$)	Type I error
H_o is false	Type II error	Correct decision (probability = $1 - \beta$)

Where, α is called level of significance which we set for hypothesis test. α = 0.05, means that we are making type-I error 5 out of 100 times.

TERMS RELATED TO TESTS OF HYPOTHESIS

Possible outcomes of any test:

1. **Reject the Null hypothesis**
2. **Fail to reject the Null hypothesis**

Now, the question here is how to do it?

Test statistics

A test statistic is a random variable which is calculated from the sample data. It is used to determine whether to reject null hypothesis or not.

Different tests uses different test statistics based on probability distribution.

TERMS RELATED TO TESTS OF HYPOTHESIS

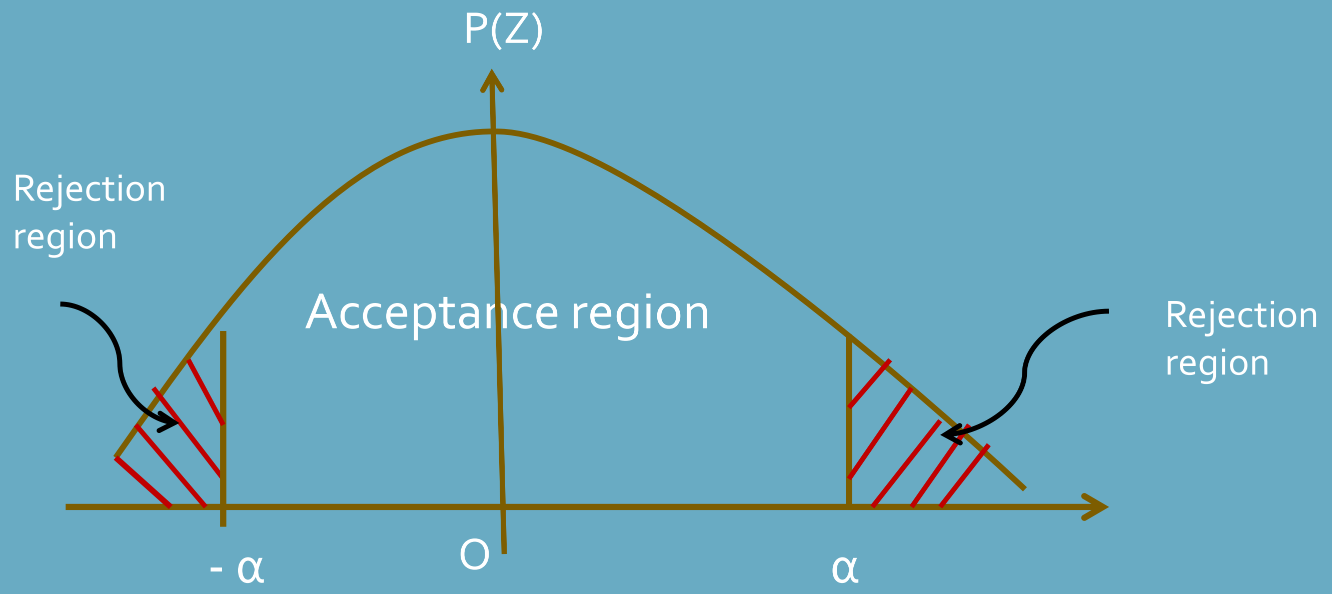
Critical value (Z_α)	Level of significance α		
	1 % (α=0.01)	5 % (α=0.05)	10 % (α=0.1)
Two tailed test	$\lvert Z_\propto \rvert$ = 2.58	$\lvert Z_\propto \rvert$ = 1.96	$\lvert Z_\propto \rvert$ = 1.645
Right tailed test	$Z_\propto$= 2.33	$Z_\propto$= 1.645	$Z_\propto$= 1.28
Left tailed test	$Z_\propto$= - 2.33	$Z_\propto$= - 1.645	$Z_\propto$=- 1.28

Confidence Limits

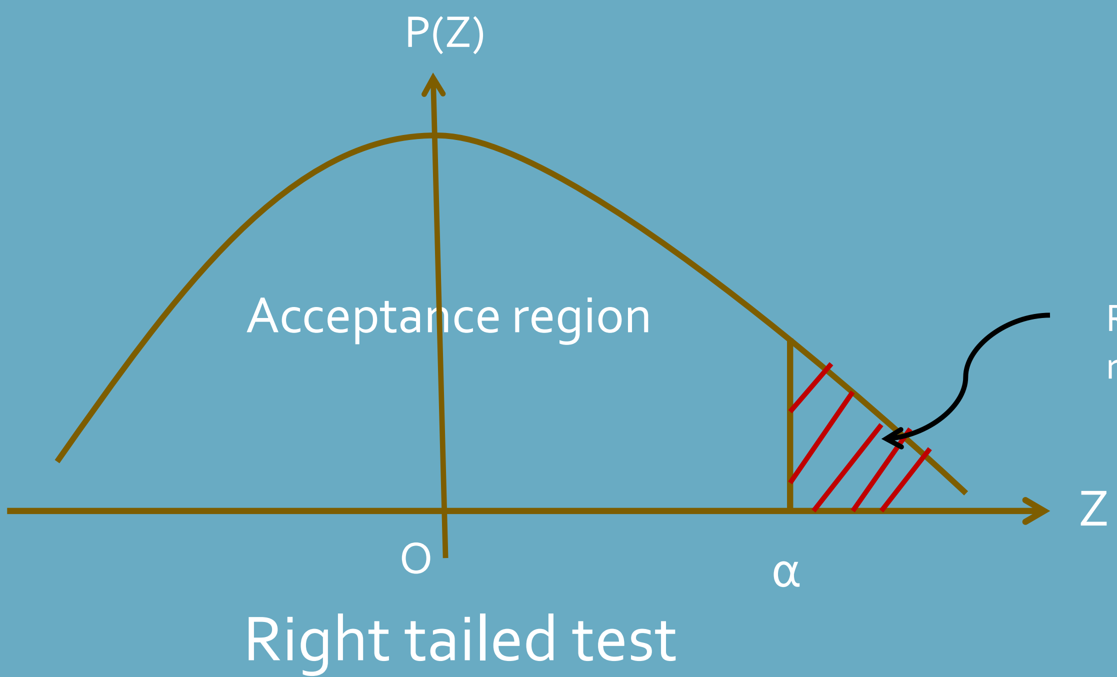
The limits within which a hypothesis should lie with specified probability are called confidence limits or fiducial limit.

Generally, the confidence limits are set up with 5% or 1% level of significance. If the sample value lies between the confidence limits, the hypothesis is accepted, if it does not, then the hypothesis is rejected at the specified level of significance.

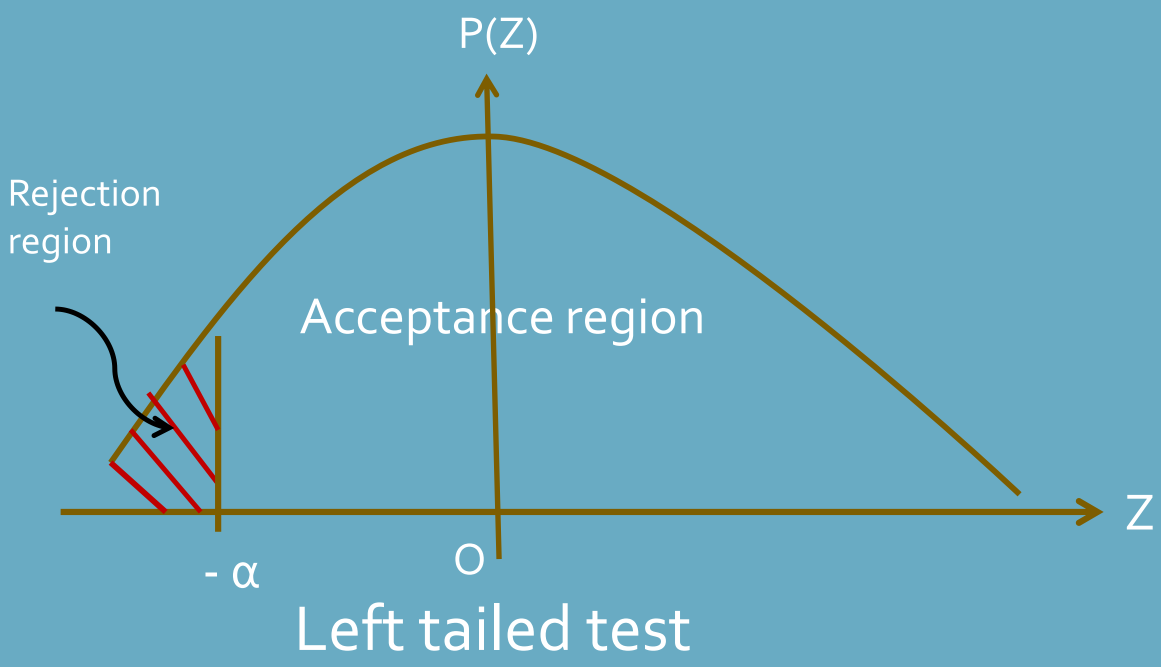
TERMS RELATED TO TESTS OF HYPOTHESIS



TERMS RELATED TO TESTS OF HYPOTHESIS



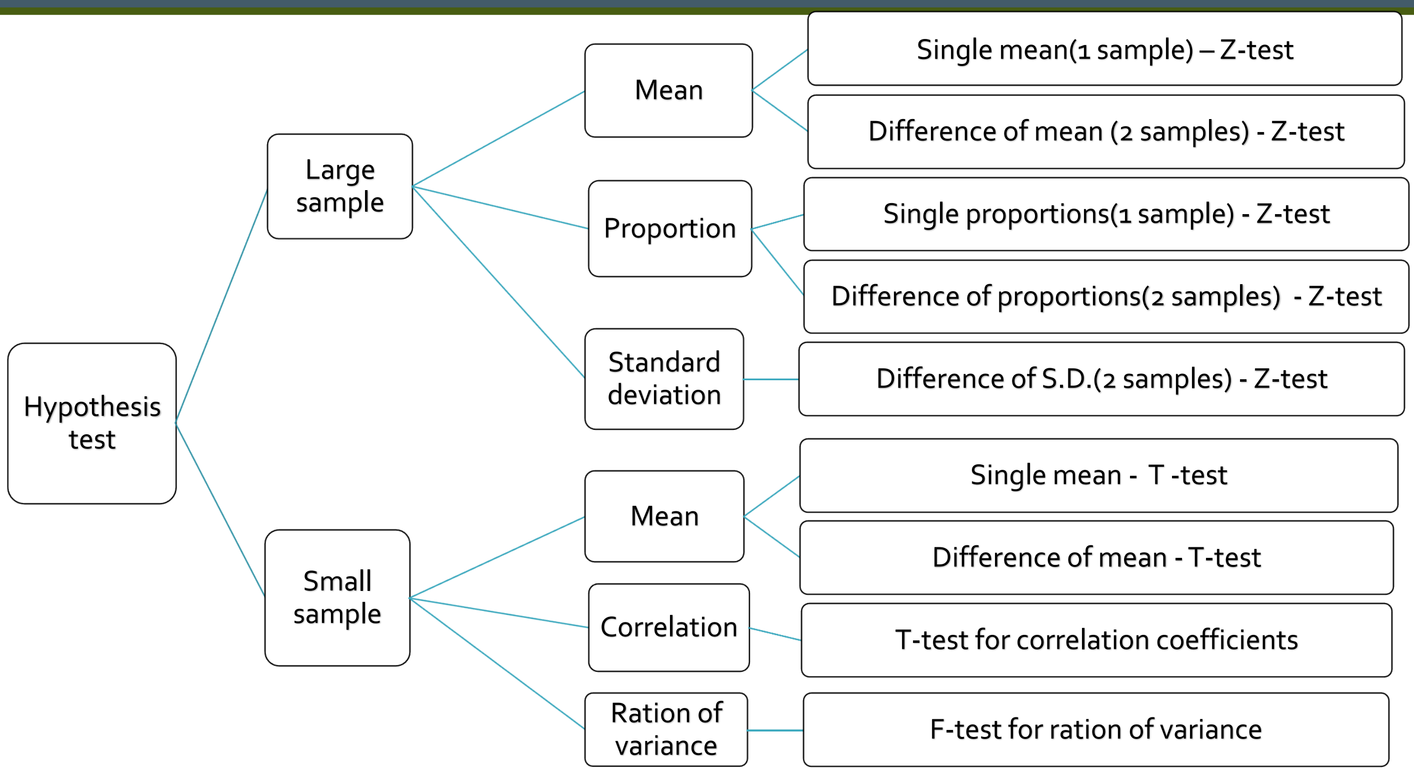
TERMS RELATED TO TESTS OF HYPOTHESIS



Which test to apply?

	Criterion	test
Parametric	If population is large ($n \geq 30$) **or** If S.D. is known	Z-test
	If population is small ($n < 30$) **or** If S.D. is not known	T-test
	Sample variance	F-test
Non-Parametric	To check compatibility of observed and expected frequency	Chi-Square test

VARIOUS TYPES OF PARAMETRIC HYPOTHESIS TESTS



PROCEDURE FOR TESTING OF HYPOTHESIS

1. Set up Null and Alternate hypothesis (H_0 and H_1)
2. Set up level of significance (α)
3. Decide which test to use (z-test, t-test, F-test, Chi-Square test)
4. Find the test statistics (Z)
5. Find out the table value of test statistics (Z_α)
6. Decision

- If computed value < Table value ($|Z| < Z_\alpha$) = Accept the Null hypothesis
- If computed value > Table value ($|Z| > Z_\alpha$) = Reject the Null hypothesis

TEST OF SIGNIFICANCE FOR SINGLE PROPORTION-LARGE SAMPLES

Let p be the sample proportion in a large random sample of size n drawn from a population having proportion P. Also, the population proportion P has a specified value P_0.

Working Rule

(i) **Null Hypothesis** $H_0 : P = P_0$, i.e. the population proportion P has a specified value P_0.

(ii) **Alternative Hypothesis** $H_1 : P \neq P_0 \ (i.e., P > P_0 \ or \ P < P_0)$
or $H_1 : P > P_0$
or $H_1 : P < P_0$

(iii) **Level of significance**: Select level of significance α

(iv) **Test statistic** : $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$, where $Q = 1 - P$

(v) **Critical Value**: Find the critical value (tabulated value) $Z_\propto$ of Z at the given level of significance.

TEST OF SIGNIFICANCE FOR SINGLE PROPORTION-LARGE SAMPLES

(vi) **Decision** : If $|Z| < Z_{\propto}$ at the level of significance $\propto$, the null hypothesis is accepted.
If $|Z| > Z_{\propto}$ at the level of significance $\propto$, the null hypothesis is rejected.

Note

1. Null Hypothesis H_0 is rejected when $|Z| > 3$ without mentioning any level of significance.
2. Confidence limits:

(i) 95% confidence limits $= p \pm 1.96\sqrt{\frac{PQ}{n}}$ i.e. $\left(p - 1.96\sqrt{\frac{PQ}{n}}, p + 1.96\sqrt{\frac{PQ}{n}}\right)$

(ii) 99% confidence limits $= p \pm 2.58\sqrt{\frac{PQ}{n}}$ i.e. $\left(p - 2.58\sqrt{\frac{PQ}{n}}, p + 2.58\sqrt{\frac{PQ}{n}}\right)$

If the population proportions P and Q are not known, p and q are used in equations.

Example 1

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

Solution

$$n = 600$$

$p =$ Sample proportion of smokers in city $= \frac{325}{600} = 0.542$

$P =$ Population proportion of smokers in city $= 0.5$

$Q = 1 - P = 1 - 0.5 = 0.5$

(i) **Null Hypothesis** $H_0 : P = 0.5$
i.e. the proportion of smokers in the city is 50%

(ii) **Alternative Hypothesis** $H_1 : P > 0.5$
(Right tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.542 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}} = 2.06$$

$|Z| = 2.06$

(v) **Critical value** $Z_{0.05} = 1.645$

(vi) **Decision**

Since $|Z| > Z_{0.05}$, the null hypothesis is rejected at 5% level of significance. i.e. , majority of men in this city are not smokers.

Example 2

In a study designed to investigate whether certain detonators used with explosives in a coal mining meet the requirement that at least 90% will ignite the explosive when charged. It is found that 174 of 200 detonators function properly. Test the null hypothesis $P \geq 0.9$ against the alternative hypothesis $P < 0.9$ at the 0.05 level of significance.

Solution

n=200

$p =$ Sample proportion of detonators function properly = $\frac{174}{200}$ = 0.87

$P =$ Population proportion of detonators functioning properly = 0.90

$Q = 1 - P = 1 - 0.90 = 0.10$

(i) **Null Hypothesis**

$H_0 : P \geq 0.90$

(ii) **Alternating Hypothesis**

$H_1 : P < 0.90$ (Left tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistic** $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.87-0.90}{\sqrt{\frac{(0.9)(0.1)}{200}}} = -1.41$

$|Z| = 1.41$

(v) Critical value $|Z_{0.05}| = 1.645$

(vi) **Decision** Since $|Z| < |Z_{0.05}|$, the null hypothesis is accepted at 5 % level of significance.

Example 3

A manufacturer claimed that at least 95% of the equipment which he supplied to a factory confirmed to specification. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance.

Solution

n=200

Number of pieces conforming to specification = 200-18 = 182

$p =$ Sample proportion of pieces confirming to specification = $\frac{182}{200}$ = 0.91

$P =$ Population proportion of pieces confirming to specification = 0.95

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

(i) **Null Hypothesis**

$H_0 : P \geq 0.95$ i.e. the proportion of pieces confirming to specification is at least 95%.

(ii) **Alternating Hypothesis**

$H_1 : P < 0.95$ (Left tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistic**

$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91-0.95}{\sqrt{\frac{(0.95)(0.05)}{200}}} = -2.59$

$|Z| = 2.59$

(v) **Critical value** $|Z_{0.05}| = 1.645$

(vi) **Decision** Since $|Z| > |Z_{0.05}|$, the null hypothesis is rejected at 5 % level of significance.
i.e., the manufacturer's claim is rejected.

Example 4

The fatality rate of covid-19 patients is believed to be 15 %. In a current year, 4000 patients suffering from covid-19 in Baroda city were treated in a private hospital and 540 patient died. Can you consider the hospital efficient at 1% level of significance?

Solution

$$n = 4000$$

$p =$ Sample proportion of covid-19 patients died $= \frac{540}{4000} = 0.135$

$P =$ Population proportion of covid-19 patients died $= 0.15$

$Q = 1 - P = 1 - 0.15 = 0.85$

(i) **Null Hypothesis** $H_0 : P = 0.15$
i.e. the hospital is efficient.

(ii) **Alternative Hypothesis** $H_1 : P < 0.15$
(One tailed test)

(iii) **Level of significance** $\propto = 0.01$

(iv) **Test statistics**

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.135 - 0.15}{\sqrt{\frac{(0.15)(0.85)}{4000}}} = \frac{-0.015}{0.005646} = -2.6567$$

$|Z| = 2.6567$

(v) **Critical value** $|Z_{0.01}| = 2.33$

(vi) **Decision**
Since $|Z| > |Z_{0.01}|$, the null hypothesis is rejected at 1% level of significance.
i.e. , the hospital is not efficient.

Example 5

In a random sample of 160 worker exposed to a certain amount of radiation, 24 experienced some ill effects. Construct a 95% confidence interval for the corresponding true percentage.

Solution

$$n = 160$$

$p =$ Sample proportion of workers exposed to radiation $= \frac{24}{160} = 0.15$

$q = 1 - p = 1 - 0.15 = 0.85$

Confidence interval at 95% level of significance is $\left(p - 1.96\sqrt{\frac{pq}{n}}, p + 1.96\sqrt{\frac{pq}{n}}\right)$

i.e., $\left(0.15 - 1.96\sqrt{\frac{(0.15)(0.85)}{160}}, 0.15 + 1.96\sqrt{\frac{(0.15)(0.85)}{160}}\right)$

$i.e.,$ $(0.0947, 0.2053)$

TEST OF SIGNIFICANCE FOR DIFFERENCE OF PROPORTION - LARGE SAMPLES

Let p_1 and p_2 be the sample proportions in two large samples of sizes n_1 and n_2 drawn from two population having proportions P_1 and P_2.

Working Rule

(i) **Null Hypothesis** $H_0 : P_1 = P_2$, i.e. there is no significant difference in two population proportions P_1 and P_2.

(ii) **Alternative Hypothesis** $H_1 : P_1 \neq P_2$

or $H_1 : P_1 > P_2$

or $H_1 : P_1 < P_2$

(iii) **Level of significance**: Select level of significance α

(iv) **Test statistic** : There are two cases

(a) When the population proportions P_1 and P_2 are known

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

(b) When the population proportions P_1 and P_2 are not known but sample

proportions p_1 and p_2 are known.

There are two methods to estimate P_1 and P_2.

Method of Substitution : In this method, sample proportion p_1 and p_2 are substituted for P_1 and P_2.

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Method of pooling : In this method, the estimated value of two population proportions is obtained by pooling the two sample proportions p_1 and p_2 into a single proportion p.

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

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

(v) **Critical Value**: Find the critical value (tabulated value) $Z_\propto$ of Z at the given level of significance.

(vi) **Decision** : If $|Z| < Z_\propto$ at the level of significance $\propto$, the null hypothesis is accepted.
If $|Z| > Z_\propto$ at the level of significance $\propto$, the null hypothesis is rejected.

Note

1. Null Hypothesis H_0 is rejected when $|Z| > 3$ without mentioning any level of significance.
2. Confidence limits:

(i) 95% confidence limits $= (p_1 - p_2) \pm 1.96 \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

(ii) 99% confidence limits $= (p_1 - p_2) \pm 2.58 \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

If the population proportions P_1 and P_2 are not known, p_1, p_2, q_1 and q_2 are used in equations.

Example 6

Random samples 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 at 5 % level of significance.

Solution $n_1 = 400, n_2 = 600$

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

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

$p = \frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{(400)(0.5)+(600)(0.541)}{400+600} = 0.525$

$q = 1 - p = 1 - 0.525 = 0.475$

(i) **Null Hypothesis** $H_0 : P_1 = P_2$, i.e. there is no significant difference in proportion of men and women in favour of the proposal.

(ii) **Alternative Hypothesis** $H_1 : P_1 \neq P_2$ (Two tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.5411}{\sqrt{(0.525)(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$= -1.28$

$|Z| = 1.28$

(v) **Critical value** $|Z_{0.05}| = 1.96$

(vi) **Decision**

Since $|Z| < |Z_{0.05}|$, the null hypothesis is accepted at 5% level of significance. i.e. , there is no significant of opinion between men and women in favour of the proposal.

Example 7

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

Solution $n_1 = 1000, n_2 = 1200$

$p_1 =$ Proportion before increasing duty $= \frac{800}{1000} = 0.8$

$p_2 =$ Proportion after increasing duty $= \frac{800}{1200} = 0.67$

$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(1000)(0.8)+(1200)(0..67)}{1000+1200} = 0.73$

$q = 1 - p = 1 - 0.73 = 0.27$

(i) **Null Hypothesis** $H_0 : P_1 = P_2$, i.e. there is no significant decrease in the consumption of tea after the increase in duty.

(ii) **Alternative Hypothesis** $H_1 : P_1 > P_2$ (Right tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.8 - 0.67}{\sqrt{(0.73)(0.27)\left(\frac{1}{1000} + \frac{1}{1200}\right)}}$$

$= 6.84$

$|Z| = 6.84$

(v) **Critical value** $Z_{0.05} = 1.645$

(vi) **Decision**

Since $|Z| > Z_{0.05}$, the null hypothesis is rejected at 5% level of significance. i.e. , there is significant decrease in the consumption of tea after the increase in duty.

Example 8

A random sample of 300 shoppers at a supermarket includes 204 who regularly uses cents off coupons. In another sample of 500 shoppers at a supermarket includes 75 who regularly uses cents off coupons. Obtain 95% confidence limits for the difference in the population proportions.

Solution $n_1 = 300, n_2 = 500$

$p_1 =$ Proportion of shoppers who uses cents of coupons in the first sample$= \frac{204}{300} = 0.68$

$q_1 = 1 - p_1 = 1 - 0.68 = 0.32$

$p_2 =$ Proportion of shoppers who uses cents of coupons in the second sample $= \frac{75}{500} = 0.15$

$q_2 = 1 - p_2 = 1 - 0.15 = 0.85$

SE$= \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = \sqrt{\frac{(0.68)(0.332)}{300} + \frac{(0.15)(0.85)}{500}} = 0.031$

95% confidence limits for the difference in population proportion is

$$\left((p_1 - p_2) - 1.96\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}, (p_1 - p_2) + 1.96\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$$

i.e., $\left((0.68 - 0.15) - 1.96(0.031), (0.68 - 0.15) + 1.96(0.031)\right)$

i.e., $(0.469,\ 0.591)$

TEST OF SIGNIFICANCE FOR SINGLE MEAN-LARGE SAMPLES

Let a random sample size n $(n > 30)$ has the sample mean $\bar{x}$ and population has the mean μ. Also, the population mean μ has a specified value μ_0.

Working Rule

(i) **Null Hypothesis** $H_0 : \mu = \mu_0$, i.e. the population mean μ has a specified value μ_0.

(ii) **Alternative Hypothesis** $H_1 : \mu \neq \mu_0$

(iii) **Level of significance**: Select level of significance α

(iv) **Test statistic** : There are two cases for calculating a test statistic Z.

(a) When the standard deviation σ of the population is known

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

(b) When the standard deviation σ of the population is not known

$$Z = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$ Where s is the sample SD.

TEST OF SIGNIFICANCE FOR SINGLE MEAN-LARGE SAMPLES

(v) **Critical Value**: Find the critical value (tabulated value) $Z_\propto$ of Z at the given level of significance.

(vi) **Decision** : If $|Z| < Z_\propto$ at the level of significance $\propto$, the null hypothesis is accepted. If $|Z| > Z_\propto$ at the level of significance $\propto$, the null hypothesis is rejected.

Note

1. Null Hypothesis H_0 is rejected when $|Z| > 3$ without mentioning any level of significance.
2. Confidence limits:

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

(ii) 99% confidence limits $= \bar{x} \pm 2.58\left(\frac{\sigma}{\sqrt{n}}\right)$

If standard deviation σ of population is not known, s is used in equations.

Example 9

A random sample of 100 Indians has an average life span of 71.8 years with standard deviation of 8.9 years. Can it be concluded that the average life span of an Indian is 70 years?

Solution

$$n = 100, \bar{x} = 71.8\ years,$$
$$\mu = 70\ years, s = 8.9\ years$$

(i) Null Hypothesis $H_0: \mu = 70\ years$, i.e. the average life span of an Indian is 70 years.

(ii) **Alternative Hypothesis** $H_1: \mu \neq 70\ years$ (Two tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{71.8 - 70}{\left(\frac{8.9}{\sqrt{100}}\right)} = 2.02$$

$$|Z| = 2.02$$

(v) **Critical value** $|Z_{0.05}| = 1.96$

(vi) **Decision**

Since $|Z| > |Z_{0.05}|$, the null hypothesis is rejected at 5% level of significance. i.e. , the average life span of an Indian is not 70 years.

Example 10

A random sample of 400 members is found to have a mean of 4.45 cm. Can it be reasonably regarded as a sample from a large population whose mean is 5 cm and variance is 4 cm?

Solution

$$n = 400, \bar{x} = 4.45\ cm,$$
$$\mu = 5\ cm, \sigma = \sqrt{4} = 2$$

(i) Null Hypothesis $H_0: \mu = 5\ cm$, i.e. the sample is drawn from a large population with mean 5 cm.

(ii) **Alternative Hypothesis** $H_1: \mu \neq 5\ cm$ (Two tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{4.45 - 5}{\left(\frac{2}{\sqrt{400}}\right)} = 5.55$$

$$|Z| = 5.55$$

(v) **Critical value** $|Z_{0.05}| = 1.96$

(vi) **Decision**

Since $|Z| > |Z_{0.05}|$, the null hypothesis is rejected at 5% level of significance. i.e. , the sample is not drawn from the large population with mean 5 cm.

Example 11

A tyre company claims that the lives of tyres have mean 42000 km with s.d. of 4000 km. A change in the production process is believed to result in better product. A test sample of 81 tyres has a mean life of 42500 km. Test at 5% level of significance that the new product is significantly better than the old one

Solution

$n = 81, \bar{x} = 42500\ km,$

$\mu = 42000\ km, \sigma = 4000\ km$

(i) Null Hypothesis $H_0 : \mu = 42000\ km$, the average lives of tyre are 42000 km

(ii) **Alternative Hypothesis** $H_1 : \mu > 42000 km$ (Right tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{42500 - 42000}{\left(\frac{4000}{\sqrt{81}}\right)} = 1.125$$

$$|Z| = 1.125$$

(v) **Critical value** $Z_{0.05} = 1.645$

(vi) **Decision**

Since $|Z| < Z_{0.05}$, the null hypothesis is accepted at 5% level of significance. i.e. , the new product is not significantly better than the old one.

Example 12

An ambulance service claims that it takes on the average at most 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 of significance.

Solution

$n = 36, \quad \bar{x} = 11\ minutes,$

$\mu = 10\ minutes, s = \sqrt{16} = 4$

(i) Null Hypothesis $H_0 : \mu \leq 10\ minutes$, i.e. ambulance service takes at most 10 minutes to reach the destination.

(ii) **Alternative Hypothesis** $H_1 : \mu > 10\ minutes$ (Right tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{11 - 10}{\left(\frac{4}{\sqrt{36}}\right)} = 1.5$$

$$|Z| = 1.5$$

(v) **Critical value** $Z_{0.05} = 1.645$

(vi) **Decision**

Since $|Z| < Z_{0.05}$, the null hypothesis is accepted at 5% level of significance. i.e. , the ambulance service takes at most 10 minutes to reach its destination.

TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS-LARGE SAMPLES

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample mean of two independent large random samples with sizes n_1 and n_2 $(n_1 > 30, n_2 > 30)$ drawn from two populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2.

Working Rule

(i) **Null Hypothesis** $H_0 : \mu_1 = \mu_2$, i.e. the two samples have been drawn from two different populations having the same means and equal standard deviations..

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

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

or $H_1 : \mu_1 > \mu_2$ (one tailed test)

(iii) **Level of significance**: Select level of significance α

(iv) **Test statistic** : There are two cases for calculating a test statistic Z.

(a) When the population standard deviations σ_1 and σ_2 are known

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

TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS-LARGE SAMPLES

(b) When the population standard deviations σ_1 and σ_2 are not known

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

Where s_1 and s_2 are sample standard deviations.

(v) **Critical Value**: Find the critical value (tabulated value) $Z_\propto$ of Z at the given level of significance.

(vi) **Decision** : If $|Z| < Z_\propto$ at the level of significance $\propto$, the null hypothesis is accepted.
If $|Z| > Z_\propto$ at the level of significance $\propto$, the null hypothesis is rejected.

Note

1. Null Hypothesis H_0 is rejected when $|Z| > 3$ without mentioning any level of significance.

TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS-LARGE SAMPLES

2. Confidence limits:

   (i) 95% confidence limits $= (\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

   (ii) 99% confidence limits $= (\bar{x}_1 - \bar{x}_2) \pm 2.58\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

If standard deviation σ_1 and σ_2 are not known, s_1 and s_2 are used in equations.

Example 14

Test the significance of the difference between the means of two normal population with the same standard deviation from the following data.

	Size	Mean	SD
Sample I	100	64	6
Sample II	200	67	8

Solution

$n_1 = 100, n_2 = 200, \bar{x}_1 = 64, \bar{x}_2 = 67,$
$s_1 = 6, s_2 = 8$

(i) Null Hypothesis $H_0 : \mu_1 = \mu_2$, i.e. there is no significant difference between two means.

(ii) **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$
(Two tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{64 - 67}{\sqrt{\frac{(6)^2}{100} + \frac{(8)^2}{200}}} = -3.64$$

$|Z| = 3.64$

(v) **Critical value** $|Z_{0.05}| = 1.96$

(vi) **Decision**

Since $|Z| > |Z_{0.05}|$, the null hypothesis is rejected at 5% level of significance. i.e. , the two population have not the same mean although they may have the same standard deviation.

Example 15

The means of samples of sizes 1000 and 2000 are 67.5 and 68 cm respectively. Can the samples be regarded as drawn from the same population of S.D. 2.5 cm.

Solution

$n_1 = 1000, n_2 = 2000$

$\bar{x}_1 = 67.5\ cm, \bar{x}_2 = 68\ cm, \sigma = 2.5\ cm$

(i) Null Hypothesis $H_0 : \mu_1 = \mu_2$, i.e. the samples have been drawn from the same population of S.D. 2.5 cm

(ii) **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$ (Two tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{67.5 - 68}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}} = -5.16$$

$|Z| = 5.16$

(v) **Critical value** $|Z_{0.05}| = 1.96$

(vi) **Decision**

Since $|Z| > |Z_{0.05}|$, the null hypothesis is rejected at 5% level of significance. i.e. , the samples cannot be regarded as drawn from the same population of SD 2.5 cm.

Example 16

In IPL, The average of runs scored by Chennai Super King in 50 matches is 175 with standard deviation 30 while that of Mumbai Indians in 60 matches is 165 with standard deviation 25. Test at 1% level of significance whether the Chennai Super King perform better than Mumbai Indians.

Solution

$n_1 = 50, n_2 = 60$

$\bar{x}_1 = 175\ , \bar{x}_2 = 165, s_1 = 30, s_2 = 25$

(i) Null Hypothesis $H_0 : \mu_1 = \mu_2$, i.e. there is no significant difference between the performance of Chennai Super King and Mumbai Indians.

(ii) **Alternative Hypothesis** $H_1 : \mu_1 > \mu_2$ (Right tailed test)

(iii) **Level of significance** $\propto = 0.01$

(iv) **Test statistics**

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{175 - 165}{\sqrt{\frac{(30)^2}{50} + \frac{(25)^2}{60}}} = \frac{10}{\sqrt{28.4167}}$$

$= 1.88$

$|Z| = 1.88$

(v) **Critical value** $Z_{0.01} = 2.33$

(vi) **Decision**

Since $|Z| < Z_{0.01}$, the null hypothesis is accepted at 1% level of significance, i.e., the Chennai Super King do not perform better than the Mumbai Indians.

Example 17

A sample of heights of 4000 Chinese soldiers has a mean of 168 cm and a standard deviation of 4 cm, while a sample of heights of 2000 Indian soldiers has a mean of 170 cm and a standard deviation of 8 cm. Do the data indicate that Indian soldiers are, on the average, taller than the Chinese soldiers at 1% level of significance?

Solution

$$n_1 = 2000, n_2 = 4000$$

$$\bar{x}_1 = 170\ cm, \bar{x}_2 = 168\ cm,$$

$$s_1 = 8\ cm, s_2 = 4\ cm$$

(i) Null Hypothesis $H_0 : \mu_1 = \mu_2$, i.e. there is no significant difference in heights of Indian soldiers and Chinese soldiers.

(ii) **Alternative Hypothesis** $H_1 : \mu_1 < \mu_2$ (Left tailed test)

(iii) **Level of significance** $\propto = 0.01$

(iv) **Test statistics**

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{170 - 168}{\sqrt{\frac{(8)^2}{2000} + \frac{(4)^2}{4000}}} = \frac{2}{\sqrt{0.0324}}$$

$$= 11.11$$

$$|Z| = 11.11$$

(v) **Critical value** $Z_{0.01} = -2.33$

(vi) **Decision**

Since $|Z| > |Z_{0.01}|$, the null hypothesis is rejected at 1% level of significance, i.e., Indian soldiers are, on average, taller than Chinese soldiers.

TEST OF SIGNIFICANCE FOR DIFFERENCE OF STANDARD DEVIATION-LARGE SAMPLES

Let s_1 and s_2 be the standard deviations of two independent large random samples with sizes n_1 and n_2 $(n_1 > 30, n_2 > 30)$ drawn from two populations with standard deviations σ_1 and σ_2.

Working Rule

(i) **Null Hypothesis** $H_0 : \sigma_1 = \sigma_2$, i.e. the two samples have been drawn from two different populations having the same standard deviations..

(ii) Alternative Hypothesis $H_1 : \sigma_1 \neq \sigma_2$ (two tailed test)

or $H_1 : \sigma_1 < \sigma_2$ (one tailed test)

or $H_1 : \sigma_1 > \sigma_2$ (one tailed test)

(iii) **Level of significance**: Select level of significance α

(iv) **Test statistic** : There are two cases for calculating a test statistic Z.

(a) When the population standard deviations σ_1 and σ_2 are known

TEST OF SIGNIFICANCE FOR DIFFERENCE OF STANDARD DEVIATION-LARGE SAMPLES

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

(b) When the population standard deviations σ_1 and σ_2 are not known

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

Where s_1 and s_2 are sample standard deviations.

(v) **Critical Value**: Find the critical value (tabulated value) $Z_\propto$ of Z at the given level of significance.

(vi) **Decision** : If $|Z| < Z_\propto$ at the level of significance $\propto$, the null hypothesis is accepted.
If $|Z| > Z_\propto$ at the level of significance $\propto$, the null hypothesis is rejected.

Example 18

The SD of a random sample of 1000 is found to be 2.6 and the SD of another random sample of 500 is 2.7. Assuming the samples to be independent, find whether the two samples could have come from populations with the same SD.

Solution

$$n_1 = 1000, n_2 = 500$$

$$s_1 = 2.6, s_2 = 2.7$$

(i) **Null Hypothesis** $H_0 : \sigma_1 = \sigma_2$, i.e. there is no significant difference between two standard deviations.

(ii) **Alternative Hypothesis** $H_1 : \sigma_1 \neq \sigma_2$
(Two tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} = \frac{2.6 - 2.7}{\sqrt{\frac{(2.6)^2}{2(1000)} + \frac{(2.7)^2}{2(500)}}} = -0.97$$

$$|Z| = 0.97$$

(v) **Critical value** $|Z_{0.05}| = 1.96$

(vi) **Decision**

Since $|Z| < |Z_{0.05}|$, the null hypothesis is accepted at 5% level of significance, i.e., there is no significance between two standard deviations and the two samples could have come from populations with the same SD.

Example 19

Examine whether the two samples for which the data are given in the following table could have been drawn from populations with the same SD.

	Size	SD
Sample I	100	5
Sample II	200	7

Solution

$$n_1 = 100, n_2 = 200$$

$$s_1 = 5, s_2 = 7$$

(i) Null Hypothesis $H_0: \sigma_1 = \sigma_2$, i.e. the two samples could have been drawn from populations with the same SD.

(ii) **Alternative Hypothesis** $H_1: \sigma_1 \neq \sigma_2$ (Two tailed test)

(iii) **Level of significance** $\propto = 0.05$

(iv) **Test statistics**

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} = \frac{5 - 7}{\sqrt{\frac{(5)^2}{2(100)} + \frac{(7)^2}{2(200)}}} = -4.02$$

$$|Z| = 4.02$$

(v) **Critical value** $|Z_{0.05}| = 1.96$

(vi) **Decision**

Since $|Z| > |Z_{0.05}|$, the null hypothesis is rejected at 5% level of significance, i.e., the two samples could not have been drawn from populations with the same SD.