

UNIT - III

TESTING OF HYPOTHESIS

A statistical hypothesis is an assumption about any aspect of a population. It could be the parameters of a distribution like mean of normal distribution, describing the population, the parameters of two or more populations, correlation or association between two or more characteristics of a population like age and height etc.

Hypothesis is an integral part of any research or investigation. Many a time, initially experiments or investigations are carried out to test a hypothesis, and the ultimate decisions are taken on the basis of the collected information and the result of the test.

To test any statistical hypothesis on the basis of a random sample of n observations, we divide the n - dimensional sample space into two regions. If the observed sample point $X(x_1, x_2, \dots, x_n)$, falls into a region called critical or rejection region, the hypothesis is rejected, but if the sample falls into complementary region, the hypothesis is accepted. The complementary region is called acceptance region.

We can respect the various possibilities in decision making about a population from a sample with the help of the following figure.

Decision	Hypothesis		
	Accept	TRUE	FALSE
		Right Decision	Type-II Error
	Reject	Type-II Error	Right Decision

Null hypothesis

A type of hypothesis used in statistics that proposes that no statistical significance exists in a set of given observations. The null hypothesis attempts to show that no variation exists between variables, or that a single variable is no different than zero. It is presumed to be true until statistical evidence nullifies it for an alternative hypothesis.

Steps for conducting tests of Significance for Mean

- (I). Set up the Null Hypothesis

It is in the form $H_0: m = m_0$
 m_0 is the value which is assumed or claimed for the population characteristic. It is reference point against which the alternative hypothesis.

(II). Set up the Alternative Hypothesis

It is in one of the following forms

$$H_1: m \neq m_0$$

$$H_1: m > m_0$$

$$H_1: m < m_0$$

We can choose from the above three forms depending on the situation posed.

(III). Decide the Level of Significance

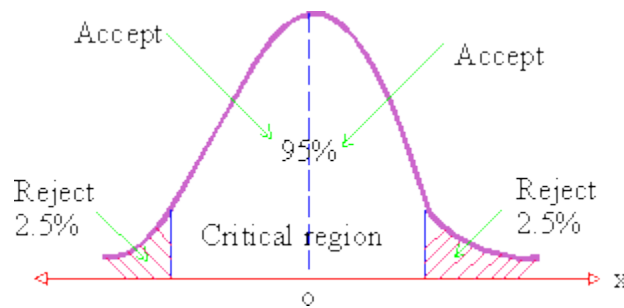
Usually, it is fixed as 5%, or sometimes 1%; if one wants to decrease the chance of rejecting when it is true. However, other values of the level of significance like 2%, 3% etc are also possible.

(IV). Decide the appropriate Statistics like z or t etc.

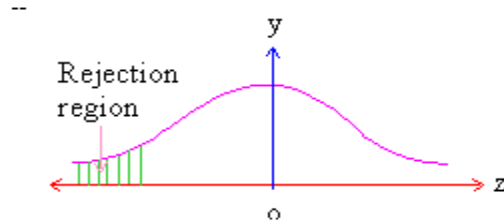
(V). Indicate the Critical Region

The critical region is formed based on the following factors:

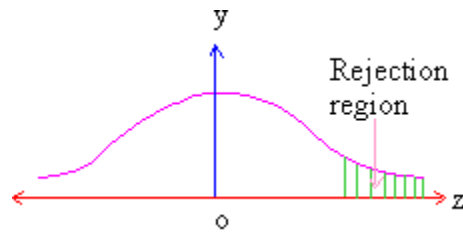
- (1). Distribution of the statistic i.e, whether the statistic follows the normal, 't', 'F' distribution.
- (2). Form the alternative hypothesis. If the form has \neq sign (e.g $m \neq m_0$) , the critical region is divided equally in the left & right tails/sides of the distribution.



If the form of alternative hypothesis has $<$ sign (e.g. $m < m_0$), the entire critical region is taken in the left tail of the distribution.



If the form of alternative hypothesis has $>$ sign (e.g. $m > m_0$), the entire critical region is taken on the right side of the distribution.



(IV). Ascertain Tabulated Values

Find out the tabulated values of the statistic based on the value of the level of significance and indicate the critical region-it can be one-sided, i.e, the entire region is on one side or it can be both sided.

(VII). Calculate the value of the statistic from the given data

A statistic is always calculated on the assumption that the null hypothesis is true.

(VIII). Accept or Reject the Null Hypothesis

If the calculated value of the statistic falls in the critical region, reject the null hypothesis; otherwise accept the null hypothesis.

TESTING OF HYPOTHESIS ABOUT A POPULATION PROPORTION

The test statistic for the population proportion is given by

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \quad \text{app } N(0,1)$$

Example: 1

In a sample of 1000 people in Mumbai, 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 of significance?

Solution:

Given that $n = 1000$ and $x = 540$

$$\text{Sample proportion } p = \frac{x}{n} = \frac{540}{1000} = 0.54$$

Here the population proportion is not given, so we choose $P = Q = 0.5$

Null Hypothesis: $H_0 : P = 0.5$

Alternative Hypothesis: $H_1 : P \neq 0.5$

Hence $p = 0.54$, $n = 1000$, $P = 0.5$ and $Q = 0.5$

$$\text{The test statistic we used as } z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$\therefore z = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 * 0.5}{1000}}} = \frac{0.04}{\sqrt{0.00025}} = 2.529$$

Calculated value $z = 2.529$

At 1% of significance the tabulated value $z_{\alpha} = 2.58$

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|2.529| \leq 2.58 \Rightarrow \text{Accept } H_0$$

Conclusion : We accept the hypothesis. That is we conclude that both rice and wheat are equally popular in this state.

Example: 2

40 people are attacked by disease and only 36 survived. Will you reject the hypothesis that the survival rate, if attacked by this disease, is 85% in favour of hypothesis that it is more, at 5% level of significance?

Solution:

Given that $n = 40$ and $x = 36$

$$\text{Sample proportion } p = \frac{x}{n} = \frac{36}{40} = 0.90$$

Population proportion $P = 85\%$ or $P = 0.85$ and $Q = 0.15$

Null Hypothesis: $H_0 : P = 0.85$

Alternative Hypothesis: $H_1 : P > 0.85$

The test statistic we used as
$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$\therefore z = \frac{0.9 - 0.85}{\sqrt{\frac{0.85 * 0.15}{40}}} = \frac{0.05}{0.05646} = 0.886$$

Calculated value $z = 0.886$

At 5% of significance the tabulated value $z_{\alpha} = 1.645$

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|0.886| \leq 1.645 \Rightarrow \text{Accept } H_0$$

Conclusion: We accept the hypothesis. That is we conclude that the survival rate may be taken as 85%.

Example: 3

A producer confesses that 22% of the items manufactured by him will be defective. To test his claim a random sample of 80 items were selected and 20 items were noted to be defective. Test the validity of the producer's claim at 1% level of significance.

Solution:

Given that $n = 80$ and $x = 20$

$$\text{Sample proportion } p = \frac{x}{n} = \frac{20}{80} = 0.25$$

Population proportion $P = 22\%$ or $P = 0.22$ and $Q = 0.78$

Null Hypothesis: $H_0 : P = 0.22$

Alternative Hypothesis: $H_1 : P > 0.22$ (Right tailed test)

The test statistic we used as
$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$\therefore z = \frac{0.25 - 0.22}{\sqrt{\frac{0.22 * 0.78}{80}}} = \frac{-0.03}{0.0463} = -0.6479$$

Calculated value $z = -0.6479$

At 1% of significance the tabulated value $z_{\alpha} = 2.33$

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|-0.6479| \leq 2.33 \Rightarrow \text{Accept } H_0$$

Conclusion: We accept the null hypothesis.

Example: 4

A die is thrown 9000 times and throw of 3 or 4 is observed 3240 times. Show that the die cannot be regarded as an unbiased one and find the limits between which the probability of a throw of 3 or 4 lies.

Solution:

Given that $n = 9000$ and $x = 3240$

$$\text{Sample proportion } p = \frac{x}{n} = \frac{3240}{9000} = 0.36$$

$$\text{Population proportion } P = \text{Probability of success} = \frac{1}{3}$$

$$\left[\text{getting 3 or 4 is success i.e., } P = \frac{2}{6} = \frac{1}{3} \right]$$

$$P = \frac{1}{3} = 0.3333 \quad \text{and} \quad Q = 1 - \frac{1}{3} = \frac{2}{3} = 0.6666$$

Null Hypothesis: $H_0 : P = \frac{1}{3}$ the die is unbiased

Alternative Hypothesis: $H_1 : P \neq \frac{1}{3}$ (Two tailed test)

$$\text{The test statistic we used as } z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$\therefore z = \frac{0.36 - 0.3333}{\sqrt{\frac{0.3333 * 0.6666}{9000}}} = \frac{0.0267}{0.004969} = 5.37$$

$$\text{Calculated value } z = 5.37$$

Since the level of significance is not given, so we choose $\alpha = 5\%$. That is at 5% of significance the tabulated value $z_{\alpha} = 1.96$.

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|5.37| > 1.96 \Rightarrow \text{Reject } H_0$$

Conclusion: We reject the null hypothesis. That is the die is a biased.

Example: 5

A manufacturer of light bulbs claims that an average 2% of the bulbs manufactured by his firm are defective. A random sample of 400 bulbs contained 13 defective bulbs. On the basis of this sample, can you support the manufacturer's claim at 5% level of significance?

Solution:

Given that $n = 400$ and $x = 13$

$$\text{Sample proportion } p = \frac{x}{n} = \frac{13}{400} = 0.0325$$

Population proportion $P = 2\%$ or $P = 0.02$ and $Q = 1 - P = 1 - 0.02 = 0.98$

Null Hypothesis: $H_0 : P = 0.02$

Alternative Hypothesis: $H_1 : P > 0.02$ (Right tailed test)

$$\text{The test statistic we used as } z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$\therefore z = \frac{0.0325 - 0.02}{\sqrt{\frac{0.02 * 0.98}{400}}} = \frac{0.0125}{0.007} = 1.7857$$

Calculated value $z = 1.7857$

At 5% level of significance the tabulated value $z_{\alpha} = 1.645$

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|1.7857| > 1.645 \Rightarrow \text{Reject } H_0$$

Conclusion: We reject the null hypothesis. That is the manufactures cannot be supported.

Example: 6

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

Solution:

Given that $n = 100$

Sample proportion $p = 0.02$ (given)

Population proportion $P = 0.01$ (given) and $Q = 1 - P = 1 - 0.01 = 0.99$

Null Hypothesis: $H_0 : P = 0.01$

Alternative Hypothesis: $H_1 : P > 0.01$ (Right tailed test)

The test statistic we used as $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$$\therefore z = \frac{0.02 - 0.01}{\sqrt{\frac{0.01 * 0.99}{100}}} = \frac{0.01}{0.009949} = 1.005$$

Calculated value $z = 1.005$

At 0.05 level or 5% level of significance the tabulated value $z_{\alpha} = 1.645$

$$|calculated\ value| \leq tabulated\ value \Rightarrow Accept\ H_0$$

$$|1.005| \leq 1.645 \Rightarrow Accept\ H_0$$

Conclusion: We accept the null hypothesis. That is the proportion of defective units has not increased.

Example: 7

A coin is tossed 256 times and 132 heads are obtained. Would you conclude that the coin is a biased one?

Solution:

Given that $n = 256$ and $x = 132$ no. of heads

$$Sample\ proportion\ p = \frac{x}{n} = \frac{132}{256} = 0.516$$

Since the population proportion is not given, so we choose

Population proportion $P = 0.5$ and $Q = 0.5$

Null Hypothesis: $H_0 : P = 0.5$

Alternative Hypothesis: $H_1 : P \neq 0.5$ (Two tailed test)

The test statistic we used as $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$$\therefore z = \frac{0.516 - 0.5}{\sqrt{\frac{0.5 * 0.5}{256}}} = \frac{0.016}{0.03125} = 0.512$$

Calculated value $z = 0.512$

Since the level of significance is not given, so we choose $\alpha = 5\%$. That is at 5% of significance the tabulated value $z_{\alpha} = 1.96$.

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|0.512| \leq 1.96 \Rightarrow \text{Accept } H_0$$

Conclusion: We accept the null hypothesis. That is the coin is unbiased.

Testing of hypothesis about the difference between two proportions

The test statistic for the difference between two proportions is given by

$$z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{app } N(0,1)$$

$$\text{where } p_1 = \frac{x_1}{n_1} \quad \text{and} \quad q_1 = 1 - p_1$$

$$\text{and } p_2 = \frac{x_2}{n_2} \quad \text{and} \quad q_2 = 1 - p_2$$

If P is not known, we use

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \quad \text{is used in place of } P \quad \text{and} \quad Q = 1 - P$$

Example: 8

A machine puts out 16 imperfect articles in a sample of 500. After the machine is overhauled, it puts out 3 imperfect articles in a batch of 100. Has the machine improved?

Solution:

$$\text{Given that } n_1 = 500, \quad x_1 = 16 \quad \text{and} \quad n_2 = 100, \quad x_2 = 3$$

$$p_1 = \text{Prop before service} \quad p_1 = \frac{x_1}{n_1} = \frac{16}{500} = 0.032 \quad \& \quad q_1 = 1 - p_1 = 1 - 0.032 = 0.968$$

$$p_2 = \text{Prop after service} \quad p_2 = \frac{x_2}{n_2} = \frac{3}{100} = 0.03 \quad \& \quad q_2 = 1 - p_2 = 1 - 0.03 = 0.97$$

Null Hypothesis:

$$H_0 : P_1 = P_2 \quad \text{i.e.,} \quad \text{the machine has not improved}$$

Alternative hypothesis:

$$H_1 : P_1 > P_2 \quad (\text{Right Tailed}) \quad \text{i.e.,} \quad \text{the machine has improved}$$

Here Population Proportion P is not known.

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

$$P = \frac{500 * 0.032 + 100 * 0.03}{500 + 100} = 0.03167$$

$$\therefore P = 0.032 \text{ and } Q = 1 - P = 1 - 0.032 = 0.968$$

$$\text{The test statistic is given by } z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$z = \frac{0.032 - 0.03}{\sqrt{[0.032 * 0.968] \left(\frac{1}{500} + \frac{1}{100} \right)}} = \frac{0.002}{\sqrt{[0.030976](0.012)}} = \frac{0.002}{0.0193}$$

$$z = 0.1037$$

At 5% level of significance, the table value for $z_\alpha = 1.645$.

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|0.1037| \leq 1.645 \Rightarrow \text{Accept } H_0$$

Conclusion: We accept the null hypothesis. That is the machine has improved after service.

Example: 8a

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

Solution:

$$\text{Given that } n_1 = 1000, x_1 = 800 \text{ and } n_2 = 1200, x_2 = 800$$

$$p_1 = \text{Prop of tea drinkers before excise} \quad p_1 = \frac{x_1}{n_1} = \frac{800}{1000} = 0.80 \quad \& \quad q_1 = 0.20$$

$$p_2 = \text{Prop of tea drinkers after excise} \quad p_2 = \frac{x_2}{n_2} = \frac{800}{1200} = 0.677 \quad \& \quad q_2 = 0.33$$

Null Hypothesis:

$$H_0 : P_1 = P_2 \quad \text{i.e.,} \quad \text{there is no difference before \& after excise}$$

Alternative hypothesis:

$$H_1 : P_1 > P_2 \quad (\text{Right Tailed})$$

Here Population Proportion P is not known.

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

$$P = \frac{0.8 * 1000 + 0.67 * 1200}{1000 + 1200} = 0.729$$

$$\therefore P = 0.729 \text{ and } Q = 0.271$$

$$\text{The test statistic is given by } z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$z = \frac{0.8 - 0.67}{\sqrt{[0.729 * 0.271] \left(\frac{1}{1000} + \frac{1}{1200} \right)}} = \frac{0.13}{\sqrt{[0.198](0.00183)}} = \frac{0.13}{0.0191}$$

$$z = 6.8 \text{ app}$$

At 5% level of significance, the table value for $z_{\alpha} = 1.645$.

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|6.8| > 1.645 \Rightarrow \text{Reject } H_0$$

Conclusion: We reject the null hypothesis. That is there is a significant difference in the consumption of tea before and after inverse in excise duty.

Example: 9

In two large populations, there are 30 and 25 percent respectively of blue-eyed people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?

Solution:

$$\text{Given that } n_1 = 1200 \text{ and } n_2 = 900$$

$$\text{Population proportion } P_1 = 30\% \text{ and } P_2 = 25\%$$

$$\text{Hence } P_1 = 0.30, \quad Q_1 = 0.70 \text{ and } P_2 = 0.25, \quad Q_2 = 0.75$$

Null Hypothesis:

$$H_0 : p_1 = p_2 \quad i.e., \text{ there is no difference before \& after excise}$$

Alternative hypothesis:

$$H_1 : p_1 \neq p_2 \text{ (Two Tailed)}$$

$$\text{The test statistic is given by } z = \frac{|P_1 - P_2|}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

$$z = \frac{|0.30 - 0.25|}{\sqrt{\frac{0.30 * 0.70}{1200} + \frac{0.25 * 0.75}{900}}} = \frac{0.05}{\sqrt{0.000383}} = \frac{0.05}{0.0195}$$

$$z = 2.55 \text{ app}$$

At 5% level of significance, the table value for $z_{\alpha} = 1.96$.

$$|calculated\ value| \leq tabulated\ value \Rightarrow Accept\ H_0$$

$$But\ |2.55| > 1.96 \Rightarrow Reject\ H_0$$

Conclusion: We reject the null hypothesis. We conclude that the difference in population proportions is unlikely to be hidden in sampling.

Example: 10

In a random sample of 100 men are taken from a village A, 60 were found to be consuming alcohol. In other sample of 200 men are taken from village B, 100 were found to be consuming alcohol. Do the two villages differ significantly in respect of the proportion of men who consume alcohol?

Solution:

$$Given\ that\ n_1 = 100, x_1 = 60 \quad and \quad n_2 = 200, x_2 = 100$$

$$p_1 = \text{Sample prop of consume alcohol in A} \quad p_1 = \frac{x_1}{n_1} = \frac{60}{100} = 0.6 \quad \& \quad q_1 = 0.4$$

$$p_2 = \text{Sample prop of consume alcohol in B} \quad p_2 = \frac{x_2}{n_2} = \frac{100}{200} = 0.5 \quad \& \quad q_2 = 0.5$$

Null Hypothesis:

$$H_0 : P_1 = P_2 \quad i.e., \text{ there is no difference between A \& B}$$

Alternative hypothesis:

$$H_1 : P_1 \neq P_2 \quad (\text{Two tailed})$$

Here Population Proportion P is not known.

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

$$P = \frac{100 * 0.6 + 200 * 0.5}{100 + 200} = 0.533$$

$$\therefore P = 0.533 \quad and \quad Q = 0.467$$

The test statistic is given by $z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$

$$z = \frac{0.6 - 0.5}{\sqrt{[0.533 * 0.467] \left(\frac{1}{100} + \frac{1}{200} \right)}} = \frac{0.1}{0.061}$$

$$z = 1.63 \text{ app}$$

At 5% level of significance, the table value for $z_{\alpha} = 1.96$.

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|1.63| < 1.96 \Rightarrow \text{Accept } H_0$$

Conclusion: We accept the null hypothesis. That is the two villages don't differ in respect of consume alcohol.

Example: 11

In a referendum submitted by the students to the body at a university, 850 men and 560 women voted. 500 men and 320 women noted favorably. Does this indicate a significant difference of opinion between men and women on this matter at 1% level of significance?

Solution:

Given that $n_1 = 850$, $x_1 = 500$ and $n_2 = 560$, $x_2 = 320$

$$p_1 = \text{Prop of men who voted favourably} \quad p_1 = \frac{x_1}{n_1} = \frac{500}{850} = 0.59 \quad \& \quad q_1 = 0.41$$

$$p_2 = \text{Prop of women who voted favourably} \quad p_2 = \frac{x_2}{n_2} = \frac{320}{560} = 0.57 \quad \& \quad q_2 = 0.43$$

Null Hypothesis:

$H_0 : P_1 = P_2$ i.e., there is no significant opinion between men & women

Alternative hypothesis:

$H_1 : P_1 \neq P_2$ (Two tailed)

Here Population Proportion P is not known.

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

$$P = \frac{850 * 0.59 + 560 * 0.57}{850 + 560} = 0.58$$

$$\therefore P = 0.58 \text{ and } Q = 0.42$$

The test statistic is given by $z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$

$$z = \frac{0.59 - 0.57}{\sqrt{[0.58 * 0.42] \left(\frac{1}{850} + \frac{1}{560} \right)}} = \frac{0.02}{0.0268}$$

$$z = 0.74 \text{ app}$$

At 1% level of significance, the table value for $z_{\alpha} = 2.58$.

$$|\text{calculated value}| \leq \text{tabulated value} \Rightarrow \text{Accept } H_0$$

$$|0.74| < 2.58 \Rightarrow \text{Accept } H_0$$

Conclusion: We accept the null hypothesis. That is we conclude that there is no significant difference between opinion of men and women.

Example: 12

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

Solution:

Given that

$$n_1 = 956 \text{ and } n_1 + n_2 = 1406$$

$$n_2 = 1406 - n_1 = 1406 - 956$$

$$n_2 = 450$$

$$p_1 = \text{Sample proportion of male births in town A} = 0.525$$

$$\text{Let } p_2 = \text{Sample proportion of male births in town B} = ?$$

Also the combined proportion of males in two town is $P = 0.496$

$$i.e., \quad P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \Rightarrow 0.496 = \frac{956 * 0.525 + 450 * p_2}{1406}$$

$$i.e., \quad 0.496 * 1406 = 501.9 + 450 * p_2$$

$$i.e., \quad 450 p_2 = 697.376 - 501.9$$

$$i.e., \quad 450 p_2 = 195.476$$

$$p_2 = \frac{195.476}{450}$$

$$p_2 = 0.434$$

Hence $p_1 = 0.525$, $p_2 = 0.434$, $P = 0.496$, $Q = 0.504$, $n_1 = 956$, $n_2 = 450$

Null Hypothesis:

$H_0 : P_1 = P_2$ i. e., there is no difference between male births in two towns

Alternative hypothesis:

$H_1 : P_1 \neq P_2$ (Two tailed)

The test statistic is given by $z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$

$$z = \frac{0.525 - 0.434}{\sqrt{[0.496 * 0.504] \left(\frac{1}{956} + \frac{1}{450} \right)}} = \frac{0.091}{0.0286}$$

$$z = 3.18 \text{ app}$$

At 5% level of significance, the table value for $z_\alpha = 1.96$.

$|calculated\ value| \leq tabulated\ value \Rightarrow Accept\ H_0$

But $|3.18| > 1.96 \Rightarrow Reject\ H_0$

Conclusion: We reject the null hypothesis. That is we conclude that there is significant difference between male births of two towns..

Example: 13

A cigarette manufacturing firm claims that its brand A cigarette out sells its brand B by 8%. If it is found that 42 out of a sample of 200 smokers prefer brand A and 18 out of another random sample of 100 smokers prefer brand B, test whether the 8% difference is a valid claim.

TESTING OF HYPOTHESIS ABOUT POPULATION MEAN

The test statistic for the population mean is given by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \text{ app } N(0,1)$$

Example: 14

The mean lifetime of a sample of 100 light tubes produced by a company is found to be 1580 hours with standard deviation of 90 hours. Test the hypothesis that the mean lifetime of the tubes produced by the company is 1600 hours.

Solution:

Given that $n = 100$, $\bar{x} = 1580$, $s = 90$ and $\mu = 1600$

Null Hypothesis: $H_0: \mu = 1600$ i.e., there is no difference between sample mean and hypothetical population mean.

Alternative Hypothesis: $H_1: \mu \neq 1600$ (Two tailed alternative)

The test statistic is given by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$
$$z = \frac{1580 - 1600}{\frac{90}{\sqrt{100}}} = -2.22$$

$$\therefore z = -2.22 \quad [\text{Calculated value}]$$

At 5% significance level the tabulated value for z_α is 1.96.

$|\text{Calculated value}| \leq \text{Tabulated value}$ then Accept H_0

But $|2.22| > 1.96$ So we reject H_0

Conclusion:

$|z| > z_\alpha$, we reject Null Hypothesis and hence the mean life time of tubes produced by the company may not be 1600 hours.

Example: 15

The mean breaking strength of a cables supplied by a manufacturer is 1800 with the S.D of 100. By a new technique in the manufacturing process, it is claimed that the breaking strength of the cable has increased. To test this claim a sample of 50 cables is tested and is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance?

Solution:

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

Null Hypothesis : $H_0: \mu = 1800$ i.e., the mean breaking strength of the cables is 1800

Alternative Hypothesis: $H_1: \mu > 1800$ (Right Tailed)

The test statistic is given by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$
$$z = \frac{1850 - 1800}{\frac{100}{\sqrt{50}}} = 3.54$$
$$\therefore z = 3.54 \quad [\text{Calculated value}]$$

At 1% significance level the tabulated value for z_{α} is 2.33.

$|\text{Calculated value}| \leq \text{Tabulated value} \quad \text{then Accept } H_0$

But $|3.54| > 2.33$ So we reject H_0

Conclusion:

$|z| > z_{\alpha}$, we reject Null Hypothesis. That is the mean breaking strength of the cable is increased.

Example: 16

A sample of 100 students is taken from a large population. The mean height of the students in this sample is 160cm. Can it be reasonably regarded that this sample is from a population of mean 165 cm and S.D 10 cm? Also find the 95% fiducial limits for the mean.

Solution:

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

Null Hypothesis: $H_0: \mu = 165$ i.e., there is no difference between sample mean and population mean.

Alternative Hypothesis: $H_1: \mu \neq 165$ (Two tailed alternative)

The test statistic is given by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$
$$z = \frac{160 - 165}{\frac{10}{\sqrt{100}}} = -5$$
$$\therefore z = -5 \quad [\text{Calculated value}]$$

At 5% significance level the tabulated value for z_{α} is 1.96.

$|Calculated\ value| \leq Tabulated\ value$ then Accept H_0

But $|5| > 1.96$ So we reject H_0

Conclusion:

$|z| > z_{\alpha}$, we reject Null Hypothesis. That is there is a significant difference between the sample mean and population means.

95% fiducial limits (Confidence Interval)

The confidence interval for the population mean is given by

$$\begin{aligned}\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\ 160 \pm (1.96) \frac{10}{\sqrt{100}} &= 160 \pm 1.96 \\ 158.04 < \mu < 161.96\end{aligned}$$

Example: 17

A sample of 900 members has a mean 3.4 cm and S.D 2.61 cm. Is the sample from a large population of mean 3.25 cm and S.D 2.61 cm. If the population is normal and the mean is unknown, find the 95% confidence limits for the mean.

Solution:

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

Null Hypothesis : $H_0: \mu = 3.25$ i.e., there is no difference between sample mean and population mean.

Alternative Hypothesis : $H_1: \mu \neq 3.25$ (Two tailed alternative)

The test statistic is given by

$$\begin{aligned}z &= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}} = 1.724 \\ \therefore z &= 1.724 \quad [Calculated\ value]\end{aligned}$$

At 5% significance level the tabulated value for z_{α} is 1.96.

$|Calculated\ value| \leq Tabulated\ value$ then Accept H_0

But $|1.724| < 1.96$ So we Accept H_0

Conclusion:

$|z| < z_{\alpha}$, we accept Null Hypothesis. That is there is no significant difference between the sample mean and population means.

95% fiducial limits (Confidence Interval)

The confidence interval for the population mean is given by

$$\begin{aligned}\bar{x} \pm z_{\alpha} \frac{\sigma}{\sqrt{n}} \\ 3.4 \pm (1.96) \frac{2.61}{\sqrt{900}} &= 3.4 \pm 0.17052 \\ 3.23 < \mu < 3.57\end{aligned}$$

TESTING OF HYPOTHESIS ABOUT DIFFERENCE BETWEEN TWO MEANS

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad app \quad N(0,1)$$

Example: 18

The buyer of electric bulbs bought 100 bulbs each of two famous brands. Upon testing these he found that brand A had a mean life of 1500 hours with a standard deviation of 50 hours whereas brand B had a mean life of 1530 hours with a standard deviation of 60 hours. Can it be concluded at 5% level of significance, that the two brands differ significantly in quality?

Solution:

Given that $\bar{x}_1 = 1500$, $\bar{x}_2 = 1530$, $s_1 = 50$, $s_2 = 60$, $n_1 = 100$ and $n_2 = 100$

Null Hypothesis

$H_0 : \mu_1 = \mu_2$ i.e., the two brands of bulbs do not differ significantly in quality.

Null Hypothesis:

$$H_0 : \mu_1 \neq \mu_2 \text{ (Two tailed alternative)}$$

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1500 - 1530}{\sqrt{\frac{50^2}{100} + \frac{60^2}{100}}} = -\frac{30}{7.81} = -3.84$$

At 5% significance level the tabulated value for z_α is 1.96.

$|Calculated\ value| \leq Tabulated\ value$ then Accept H_0

But $|3.84| > 1.96$ So we Reject H_0

Conclusion:

$|z| > z_\alpha$, we reject the Null Hypothesis. That is the two brands of bulbs differ significantly in quality.

Example: 19

Intelligence test given to two groups of boys and girls gave the following information

	Mean Score	S.D	Number
Girls	75	10	50
Boys	70	12	100

Is the difference in the mean scores of boys and girls statistically significant?

Solution:

Given that $\bar{x}_1 = 75$, $\bar{x}_2 = 70$, $s_1^2 = 100$, $s_2^2 = 144$, $n_1 = 50$ and $n_2 = 100$

Null Hypothesis

$H_0 : \mu_1 = \mu_2$ i.e., the difference in mean score of boys and girls are not significant.

Null Hypothesis:

$$H_0 : \mu_1 \neq \mu_2 \text{ (Two tailed alternative)}$$

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{75 - 70}{\sqrt{\frac{100}{50} + \frac{144}{100}}} = \frac{5}{\sqrt{3.44}} = 2.6958$$

At 5% significance level the tabulated value for z_α is 1.96.

|Calculated value| ≤ Tabulated value then Accept H_0

But |2.69| > 1.96 So we Reject H_0

Conclusion:

$|z| > z_\alpha$, we reject the Null Hypothesis. That is the difference in mean score of boys and girls are statistically significant.

Example: 20

A simple sample of heights of 6400 Englishmen has a mean of 170 cm and S.D of 6.4 cm, while a simple sample of heights of 1600 Americans has mean of 172 cm and S.D of 6.3 cm. Do the data indicate that Americans are on the average taller than Englishmen's?

Solution:

Given that $\bar{x}_1 = 170$, $\bar{x}_2 = 172$, $s_1 = 6.4$, $s_2 = 6.3$, $n_1 = 6400$ and $n_2 = 1600$

Null Hypothesis

$H_0 : \mu_1 = \mu_2$ i.e., there is no significant difference between heights of Americans and Englishmen.

Alternative Hypothesis:

$H_0 : \mu_1 < \mu_2$ (Left Tailed) Americans are taller than Englishmen.

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{170 - 172}{\sqrt{\frac{6.4^2}{6400} + \frac{6.3^2}{1600}}} = \frac{-2}{\sqrt{\quad}} = -11.321$$

At 5% significance level the tabulated value for z_α is 1.96.

|Calculated value| ≤ Tabulated value then Accept H_0

But |11.32| > 1.96 So we Reject H_0

Conclusion:

$|z| > z_{\alpha}$, we reject the Null Hypothesis. That is Americans are on the average, taller than Englishmen.

Example: 21

In a certain factory there are two independent processes manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 Ozs, with a s.d of 12 Ozs, while the corresponding figures in a sample of 400 items from the other process are 124 Ozs and 14 Ozs. Is the difference between the two sample means significant? Also find 99% confidence limits.

Solution:

Given that $\bar{x}_1 = 120$, $\bar{x}_2 = 124$, $s_1 = 12$, $s_2 = 14$, $n_1 = 250$ and $n_2 = 400$

Null Hypothesis

$H_0 : \mu_1 = \mu_2$ i.e., the sample means do not differ significantly.

Alternative Hypothesis:

$H_0 : \mu_1 \neq \mu_2$ (Two tailed alternative)

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{120 - 124}{\sqrt{\frac{144}{250} + \frac{196}{400}}} = \frac{-4}{1.0324} = -3.87$$

At 1% significance level the tabulated value for z_{α} is 2.58.

$|Calculated\ value| \leq Tabulated\ value$ then Accept H_0

But $|3.87| > 2.58$ So we Reject H_0

Conclusion:

$|z| > z_{\alpha}$, we reject the Null Hypothesis. That is there is a significant difference between the sample means.

Example: 22

Random samples drawn from two places gave the following data relating to the heights of male adults:

	Place A	Place B
Mean Height (in Inches)	68.5	65.5
S.D (in Inches)	2.5	3
No. Of Adult males in sample	1200	1500

Test at 5% level of significance that the mean height is the same for adults in the two places.

Example: 23

In comparing the average protein content μ_1 and μ_2 of two brands of dog food, a consumer testing service finds that fifty 5-pund packages of brand A dog food had an average protein content $\bar{x}_1 = 11$ ounces per package and a s.d of $s_1 = 1$ ounce, while sixty-5 pound packages of brand B dog food has an average protein content $\bar{x}_2 = 9$ ounces per package and a s.d of $s_2 = 0.5$ ounce. A difference of 0.5 ounces is considered to be not sufficiently important to report as a consumer issue. Therefore, a decision was made to test the hypothesis.

$H_0 : \mu_1 - \mu_2 = 0.5$ versus $H_1 : \mu_1 - \mu_2 > 0.5$. use the observed data to test these hypothesis at 0.01 level of significance.

Solution:

Given that $\bar{x}_1 = 11$, $\bar{x}_2 = 9$, $s_1 = 1$, $s_2 = 0.5$, $n_1 = 50$ and $n_2 = 60$

Null Hypothesis

$H_0 : \mu_1 - \mu_2 = 0.5$ i.e., the difference is sufficiently important.

Alternative Hypothesis:

$H_0 : \mu_1 - \mu_2 > 0.5$ (Two tailed alternative)

The test statistic is given by

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(11 - 9) - (0.5)}{\sqrt{\frac{1^2}{50} + \frac{0.5^2}{60}}} = \frac{1.5}{0.1554} = 9.65$$

At 1% significance level the tabulated value for z_α is 2.33.

$|Calculated\ value| \leq Tabulated\ value$ then Accept H_0

But $|9.65| > 2.33$ So we Reject H_0

Conclusion:

$|z| > z_{\alpha}$, we reject the Null Hypothesis.

SMALL SAMPLES

TEST OF HYPOTHESIS ABOUT THE POPULATION MEAN

The test statistic is defined as

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \text{ with } n - 1 \text{ degrees of freedom}$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \text{ with } n - 1 \text{ degrees of freedom}$$

Example: 26

Ten oil tins are taken at random from an automatic filling machine. The mean weight of the tins 15.8 kg and standard deviation of 0.5 kg. Does the sample mean differ significantly from the intended weight of 16 kg?

Solution:

Given that $n = 10$, $\bar{x} = 15.8$, $s = 0.50$, $\mu = 16$

Null Hypothesis : $H_0 : \mu = 16$

Alternative Hypothesis : $H_1 : \mu \neq 16$ (two tailed test)

The test statistic is given by

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \text{ with } n - 1 \text{ degrees of freedom}$$

$$t = \frac{15.8 - 16}{\frac{0.5}{\sqrt{10-1}}} = \frac{-0.2}{\frac{0.5}{\sqrt{9}}} = \frac{-0.2}{0.1667} = -1.2$$

$$|t| = 1.2$$

The critical value for t for a two tailed test at 5% level of significance with $10 - 1 = 9$ degrees of freedom is 2.26.

Calculated value = 1.2 and Tabulated value = 2.26

$|Calculated\ value| \leq Tabulated\ value$ then accept H_0

$|1.2| < 2.26$ accept H_0

Conclusion:

$|t_{\alpha}| < t$, we accept H_0 . That is sample mean is not differ from the intended weight.

Example: 27

A random sample of 10 boys had the following I.Q's: 70, 120, 110, 101, 88, 83, 95, 98, 107, and 100. Do these data support the assumption of a population mean I.Q of 100? Find the reasonable range in which most of the mean I.Q values of samples of 10 boys lie.

Solution:

x_i	70	120	110	101	88	83	95	98	107	100	$\sum x = 972$
$x_i - \bar{x}$	-27.2	22.8	12.8	3.8	-9.2	-14.2	-2.2	0.8	9.8	2.8	
$(x_i - \bar{x})^2$	739.84	519.84	163.84	14.44	84.64	201.64	4.84	0.64	96.04	7.84	$\sum (x_i - \bar{x})^2 = 1833.6$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{972}{10} = 97.2$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{1833.6}{10 - 1} = 203.7$$

$$\therefore S = \sqrt{203.7} = 14.27$$

Hence $n = 10$, $\bar{x} = 97.2$, $S = 14.27$, $\mu = 100$

Null Hypothesis : $H_0 : \mu = 100$

Alternative Hypothesis : $H_1 : \mu \neq 100$ (two tailed test)

The test statistic is given by

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \text{ with } n - 1 \text{ degrees of freedom}$$

$$t = \frac{97.2 - 100}{\frac{14.27}{\sqrt{10}}} = \frac{-2.8}{4.5126} = \frac{-0.2}{0.1667} = -0.62$$

$$|t| = 0.62$$

The critical value for t for a two tailed test at 5% level of significance with $10 - 1 = 9$ degrees of freedom is 2.26.

Calculated value = 0.62 and Tabulated value = 2.26

$|Calculated\ value| \leq Tabulated\ value$ then accept H_0

$|0.62| < 2.26$ accept H_0

Conclusion:

$|t_{\alpha}| < t$, we accept H_0 . That is the data are consistent with the assumption of mean I.Q of 100 in the population.

Confidence Interval :

The confidence interval for the small sample mean is given by

$$\bar{x} \pm t_{\frac{\alpha}{2}} \left(\frac{S}{\sqrt{n}} \right) \text{ with } n - 1 \text{ D.O.F}$$

$$97.2 \pm 2.26 \left(\frac{14.27}{\sqrt{10}} \right) = 97.2 \pm [2.26 * 4.5125]$$

$$= 97.2 \pm 10.19$$

$$\therefore 87 < \mu < 107.4$$

Example: 28

The height of 10 males of a given locality is found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches?

Solution:

x_i	70	67	62	68	61	68	70	64	64	66	$\sum x = 660$
$x_i - \bar{x}$	4	1	-4	2	-5	2	4	-2	-2	0	
$(x_i - \bar{x})^2$	16	1	16	4	25	4	16	4	4	0	$\sum (x_i - \bar{x})^2 = 90$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{660}{10} = 66$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{90}{10 - 1} = \frac{90}{9} = 10$$

$$\therefore S = \sqrt{10} = 3.162$$

Hence $n = 10$, $\bar{x} = 66$, $S = 3.162$, $\mu = 64$

Null Hypothesis : $H_0 : \mu = 64$ i.e., the average height is equal to 64 inches.

Alternative Hypothesis : $H_1 : \mu > 64$ (Right tailed test)

The test statistic is given by

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \text{ with } n - 1 \text{ degrees of freedom}$$

$$t = \frac{66 - 64}{\frac{3.162}{\sqrt{10}}} = \frac{2}{1} = 2$$

$$|t| = 2$$

The critical value for t for a right tailed test at 5% level of significance with $10 - 1 = 9$ degrees of freedom is 1.833.

Calculated value = 2 and Tabulated value = 1.833

$|Calculated \text{ value}| \leq Tabulated \text{ value}$ then accept H_0

But $|2| > 1.833$ Reject H_0

Conclusion:

$|t_{\alpha}| > t$, we reject H_0 . That is the average height is greater than 64 inches.

TESTING OF HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO MEANS

(USING T TEST)

The test statistic for the difference between two means for small samples is given by

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Where $\bar{x} = \frac{\sum x}{n_1}$ and $\bar{y} = \frac{\sum y}{n_2}$

Example: 29

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

	Type I	Type II
sample size	$n_1 = 8$	$n_2 = 7$

sample mean	$\bar{x}_1 = 1234 \text{ hrs}$	$\bar{x}_2 = 1036 \text{ hrs}$
sample S.D	$s_1 = 36 \text{ hrs}$	$s_2 = 40 \text{ hrs}$

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life?

Solution:

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

Null Hypothesis : $H_0 : \mu_1 = \mu_2$ i.e., the two types I & type II of electric bulbs are same.

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

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \quad \text{with } (n_1 + n_2 - 2) \text{ degrees of freedom}$$

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

$$|t| =$$

The critical value for t for a right tailed test at 5% level of significance with $8 + 7 - 2 = 13$ degrees of freedom is 1.77.

Calculated value = 9.39 and Tabulated value = 1.77

$|Calculated \text{ value}| > Tabulated \text{ value}$ then Reject H_0

But $|9.39| > 1.77$ Reject H_0

Conclusion:

$|t_{\alpha}| > t$, we reject H_0 . That is the type I is superior than type II.

Example: 30

In a test given to two groups of students the marks obtained were as follows.

First Group	18	20	36	50	49	36	34	49	41
Second Group	29	28	26	35	30	44	46		

Examine the significant difference between the means of marks secured by students of the above two groups.

Solution:

x_i	18	20	36	50	49	36	34	49	41	$\sum x = 333$
$x_i - \bar{x}$	-19	-17	-1	13	12	-1	-3	12	4	
$(x_i - \bar{x})^2$	361	289	1	169	144	1	9	144	16	$\sum (x_i - \bar{x})^2 = 1134$
y_i	29	28	26	35	30	44	46	238		$\sum y = 238$
$y_i - \bar{y}$	-5	-6	-8	1	-4	10	12			
$(y_i - \bar{y})^2$	25	36	64	1	16	100	144	386		$\sum (y_i - \bar{y})^2 = 386$

$$\bar{x} = \frac{\sum x_i}{n_1} = \frac{333}{9} = 37$$

$$and \quad \bar{y} = \frac{\sum y_i}{n_2} = \frac{238}{7} = 34$$

$$s_1^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1} = \frac{1134}{9 - 1} = 141.75 \quad and \quad s_2^2 = \frac{\sum (y_i - \bar{y})^2}{n_2 - 1} = \frac{386}{7 - 1} = 64.33$$

$$\therefore s_1 = \sqrt{141.75} = \quad and \quad s_2 = \sqrt{64.33} =$$

Hence $n_1 = 9, n_2 = 7, \bar{x}_1 = 37, \bar{x}_2 = 34, s_1 = , s_2 =$

Null Hypothesis : $H_0 : \mu_1 = \mu_2$

i.e., there is no significant difference between the mean marks secured by the two groups.

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

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \quad with (n_1 + n_2 - 2) \text{ degrees of freedom}$$

$$t = \frac{37 - 34}{\sqrt{\left(\frac{141.75}{9} + \frac{64.33}{7}\right)}} =$$

$$|t| = 0.57$$

The critical value for t for a two tailed test at 5% level of significance with $9 + 7 - 2 = 14$ degrees of freedom is 1.76.

Calculated value = 0.57 and Tabulated value = 1.76

$|Calculated \text{ value} | < Tabulated \text{ value}$ then Accept H_0

But $|0.57| < 1.76$ Accept H_0

Conclusion:

$|t_\alpha| < t$, we Accept H_0 . That is there is no significant difference between the mean marks secured by the two groups.

Example: 31

Two independent samples of 8 and 7 items respectively had the following values.

Sample I	9	11	13	11	15	9	12	14
Sample II	10	12	10	14	9	8	10	

Is the difference between the means of samples significant?

Solution:

x_i	9	11	13	11	15	9	12	14	$\sum x = 94$
$x_i - \bar{x}$	-2.75	-0.75	1.25	-0.75	3.25	-2.75	0.25	2.25	
$(x_i - \bar{x})^2$	7.563	0.563	1.563	0.563	10.56	7.563	0.063	5.063	$\sum (x_i - \bar{x})^2 = 33.5$
y_i	10	12	10	14	9	8	10		$\sum y = 73$
$y_i - \bar{y}$	-0.43	1.57	-0.43	3.57	-1.43	-2.43	-0.43		
$(y_i - \bar{y})^2$	0.185	2.465	0.185	12.74	2.045	5.905	0.185		$\sum (y_i - \bar{y})^2 = 23.71$

$$\bar{x} = \frac{\sum x_i}{n_1} = \frac{94}{8} = 11.75$$

and

$$\bar{y} = \frac{\sum y_i}{n_2} = \frac{73}{7} = 10.43$$

$$s_1^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1} = \frac{33.5}{8 - 1} = 4.79$$

and

$$s_2^2 = \frac{\sum (y_i - \bar{y})^2}{n_2 - 1} = \frac{23.71}{7 - 1} = 3.95$$

$$\therefore s_1 = \sqrt{4.79} =$$

and

$$s_2 = \sqrt{3.95} =$$

Hence $n_1 = 8$, $n_2 = 7$, $\bar{x}_1 = 11.75$, $\bar{x}_2 = 10.43$ $s_1 =$, $s_2 =$

Null Hypothesis : $H_0 : \mu_1 = \mu_2$

i.e., there is no significant difference between the mean of two samples.

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

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \quad \text{with } (n_1 + n_2 - 2) \text{ degrees of freedom}$$

$$t = \frac{11.75 - 10.43}{\sqrt{\left(\frac{8}{8} + \frac{7}{7}\right)}} =$$

$$|t| = 1.215$$

The critical value for t for a two tailed test at 5% level of significance with $8 + 7 - 2 = 13$ degrees of freedom is 1.771.

Calculated value = 1.215 and Tabulated value = 1.771

$|Calculated\ value| < Tabulated\ value$ then Accept H_0

But $|1.215| < 1.771$ Accept H_0

Conclusion:

$|t_{\alpha}| < t$, we Accept H_0 . That is there is no significant difference between the mean of two samples.

DESIGN OF EXPERIMENTS

Completely Randomized Design (CRD) Or One Way Classification

Example: 35

A random sample is selected from each of three makes of ropes and their breaking strength (in pounds) are measured with the following results:

I	II	III
70	100	60
72	110	65
75	108	57
80	112	84
83	113	87
	120	73
	107	

Test whether the breaking strength of the ropes differs significantly.

Solution:

I	II	III
-10	20	-20
-8	30	-15
-5	28	-23
0	32	4
3	33	7

$$\begin{array}{ccc} & 40 & -7 \\ & 27 & \\ \hline T_1 = -20 & T_2 = 210 & T_3 = -54 \end{array}$$

Let us take the null hypothesis that the breaking strength of the ropes does not differ significantly.

$$i.e., H_0 : \mu_1 = \mu_2 = \mu_3$$

For simplifying calculations we take 80 as common origin and the new values are given below:

$$\begin{aligned} \text{Here } G &= \sum \sum y_{ij} = \text{Grand total} \\ &= -20 + 210 - 54 = 136 \end{aligned}$$

$$\therefore \text{Correction factor} = \frac{G^2}{N} = \frac{(136)^2}{18} = 1027.56$$

$$\text{Total sum of squares } SST = 7992 - 1027.56 = 6964.44$$

Between ropes sum of squares

$$R_2 = SSW = SST - R_1 = 6964.44 - 5838.44$$

$$R_2 = 1126$$

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of squares	Mean square	F-ratio
Between ropes	2	5838.44	2919.22	$F = \frac{2921.22}{75.07}$ $= 38.89$
Error	15	1126	75.07	
Total	17	6964.44		

The table value for $F_{(2,15)}$ at 5% level of significance is 3.68.

Conclusion:

Since the calculated value of F is greater than the table value, the null hypothesis is rejected and hence we conclude that the breaking strengths of the ropes differ significantly.

Example: 36

The following are the number of mistakes made in 5 successive days by 4 technicians working for a photographic laboratory. Test whether the difference among the four sample means can be attributed to chance. [Test at a level of significance $\alpha = 0.01$].

I	II	III	IV
6	14	10	9
14	9	12	12
10	12	7	8
8	10	15	10
11	14	11	11

Solution:

	I	II	III	IV
	6	14	10	9
	14	9	12	12
	10	12	7	8
	8	10	15	10
	11	14	11	11
<i>Total T_i</i>	$T_1 = 49$	$T_2 = 59$	$T_3 = 55$	$T_4 = 50$
T_i^2	$T_1^2 = 2401$	$T_2^2 = 3481$	$T_3^2 = 3025$	$T_4^2 = 2500$

Null Hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

i.e., the difference among the four sample means can be attributed to chance.

Alternative Hypothesis:

H_1 : There is a significant difference among the four sample means.

Level of significance: $\alpha = 0.01$

$$\begin{aligned} \text{Grand total } G &= \sum \sum y_{ij} = 49 + 59 + 55 + 50 \\ &= 213 \end{aligned}$$

$$\therefore \text{Correction factor} = \frac{G^2}{N} = \frac{(213)^2}{20} = 2268.45$$

$$\text{Total sum of squares } SST = 2383 - 2268.45 = 114.55$$

Between ropes sum of squares

$$R_2 = SSW = SST - R_1 = 114.55 - 12.95 = 101.6$$

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of squares	Mean sum of squares	F-ratio
Between technicians	$k - 1 = 4 - 1 = 3$	12.95	$\frac{12.95}{3} = 4.32$	$F = \frac{6.35}{4.32}$
Error	$N - k = 20 - 4 = 16$	101.6	$\frac{101.6}{16} = 6.35$	$= 1.47$
Total	$N - 1 = 20 - 1 = 19$	114.55		

The table value for $F_{(16,3)}$ at 1% level of significance is 8.7.

Conclusion:

Since the calculated value of F is less than the table value, the null hypothesis is accepted. That is the difference between the sample means is due to chance.

Example: 37

As part of the investigation of the collapse of the roof of a building, a testing laboratory is given all the available bolts that connected all the steel structure at three different positions on the roof. The forces required to shear each of these bolts (coded values) are as follows:

Position 1	90	82	79	98	83	91	
Position 2	105	89	93	104	89	95	86
Position 3	83	89	80	94			

Solution:

For simplifying calculations we take 90 as origin. The new values are

								T_i	T_i^2
Position 1	0	-8	-11	8	-7	1		-17	289
Position 2	15	-1	3	14	-1	5	-4	31	961
Position 3	-7	-1	-10	4				-14	196

Null Hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3$

i.e., the difference among the sample means at the three positions is not significant.

Alternative Hypothesis: The differences between the sample means are significant.

Level of significance: $\alpha = 0.05$

$$\text{Grand total } G = \sum \sum y_{ij} = -17 + 31 - 14 = 0$$

$$\therefore \text{Correction factor} = \frac{G^2}{N} = \frac{(0)^2}{20} = 0$$

$$\text{Total sum of squares } SST = 938 - 0 = 938$$

$$\text{Sum of the squares between positions} = SSB = R_1$$

$$= \left[\frac{289}{6} + \frac{961}{7} + \frac{196}{4} \right] - 0$$

$$R_1 = 234.452$$

$$\text{Error sum of squares} = SSW = R_2 = SST - R_1 = 703.548$$

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of squares	Mean sum of squares	F-ratio
Between Positions	$3 - 1 = 2$	234.452	117.226	$F = \frac{117.226}{50.253}$
Error	$17 - 3 = 14$	703.548	50.253	$= 2.333$
Total	16	938		

The table value for $F_{(2,14)}$ at 5% level of significance is 3.74.

Conclusion:

Since the calculated value of F is less than the table value, the null hypothesis is accepted. That is the difference among the sample means is not significant.

Example: 38

A completely randomized design experiment with 10 plots and 3 treatments gave the following results:

Plot No	1	2	3	4	5	6	7	8	9	10
Treatment	A	B	C	A	C	C	A	B	A	B
Yield	5	4	3	7	5	1	3	4	1	7

Analyse the results for treatment effects.

OR

A completely randomized design experiment with ten plots and three treatments gave the results given below. Analyze the results for the effects of treatments.

Treatment	Replications			
A	5	7	1	3
B	4	4	7	
C	3	1	5	

Solution:

Treatment	Replications				T_i	T_i^2
A	5	7	1	3	16	256
B	4	4	7		15	225
C	3	1	5		9	81

Null Hypothesis: H_0 : There is no difference in the effects of treatments.

Alternative Hypothesis: H_1 : There is significant difference in the effects of treatments.

Level of significance: $\alpha=0.05$

$$\text{Grand total } G = 16 + 15 + 9 = 40$$

$$\therefore \text{Correction factor} = \frac{G^2}{N} = \frac{(40)^2}{10} = 160$$

$$\begin{aligned} \text{Between treatments sum of squares } R_1 = SSB &= \sum \frac{T_i^2}{r_i} - C.F \\ &= \left[\frac{256}{4} + \frac{225}{3} + \frac{81}{3} \right] - 160 \\ R_1 &= 6 \end{aligned}$$

$$\text{Error sum of squares } R_2 = SSW = SST - R_1 = 40 - 6 = 34$$

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of squares	Mean sum of squares	F-ratio
Between Positions	$3 - 1 = 2$	6	3	$F = \frac{4.857}{3}$
Error	$10 - 3 = 7$	34	4.857	$= 1.619$
Total	9	40		

The table value for $F_{(7,2)}$ at 5% level of significance is 19.4.

Conclusion:

$F < F_{0.05}$. We accept H_0 and conclude that there is no difference in the effects of treatments.

RANDOMIZED BLOCK DIAGRAM OR TWO WAY CLASSIFICATION (RBD)

Example: 39

The following data represents the number of units of production per day turned out by different workers using 4 different types of machines.

		Machine Type			
		A	B	C	D
Workers	1	44	38	47	36
	2	46	40	52	43
	3	34	36	44	32
	4	43	38	46	33
	5	38	42	49	39

1. Test whether the five men differ with respect to mean productivity and
2. Test whether the mean productivity is the same for the four different machine types.

Solution:

Let us take the null hypothesis that

1. The 5 workers do not differ with respect to mean productivity

$$i.e., H_{01} : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

2. The mean productivity is the same for the four different machines.

$$i.e., H_{02} : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

To simplify calculation let us subtract 40 from each value, the new values are

		Machine Type				
		A	B	C	D	Total
Workers	1	4	-2	7	-4	5
	2	6	0	12	3	21
	3	-6	-4	4	-8	-14
	4	3	-2	6	-7	0
	5	-2	2	9	-1	8
Total		5	-6	38	-17	20

$$\text{Correction factor} = C.F = \frac{G^2}{N} = \frac{(20)^2}{20} = 20$$

$$SST = \text{Total sum of squares} = \sum_i \sum_j y_{ij}^2 - C.F$$

$$= [(4)^2 + (-2)^2 + (7)^2 + (-4)^2 + (6)^2 + (0)^2 + (12)^2 + (3)^2 + (-6)^2 + (-4)^2 + (4)^2 + (-8)^2 + (3)^2 + (-2)^2 + (6)^2 + (-7)^2 + (-2)^2 + (2)^2 + (9)^2 + (-1)^2] - 20$$

$$= 574$$

Between workers sum of squares

$$R_1 = \frac{(5)^2}{4} + \frac{(21)^2}{4} + \frac{(-14)^2}{4} + \frac{(0)^2}{4} + \frac{(8)^2}{4} - 20 = 181.5 - 20 = 161.5$$

Between machines sum of squares

$$R_2 = \frac{(5)^2}{5} + \frac{(-6)^2}{5} + \frac{(38)^2}{5} + \frac{(-17)^2}{5} - 20 = 358.8 - 20 = 338.8$$

Error sum of squares

$$R_3 = SST - R_1 - R_2 = 574 - 161.5 - 338.8 = 73.7$$

ANOVA table for two-way classification				
Source of variation	Degrees of freedom	Sum of squares (SS)	Mean sum of squares (MS)	Variance Ration (F-Ratio)

Workers	$5 - 1 = 4$	161.5	$\frac{161.5}{4} = 40.375$	$F_1 = \frac{40.375}{6.14} = 6.576$
Machines	$4 - 1 = 3$	338.8	$\frac{338.8}{3} = 112.93$	$F_2 = \frac{112.93}{6.14} = 18.393$
Error	$4 * 3 = 12$	73.7	$\frac{73.7}{12} = 6.14$	
Total	19	574		

$$F_{0.05}(4,12) = 3.26 \text{ and } F_{0.05}(3,12) = 3.49$$

Conclusion:

1. $F_1 > F_{0.05}(4,12)$. Hence H_{01} is accepted. That is the 5 workers differ respect to mean productivity.
2. $F_2 > F_{0.05}(3,12)$. Hence H_{02} is rejected. That is the mean productivity is not the same for the four machines.

Example: 40

A company appoints 4 salesmen's A,B , C and D and observes their sales in 3 seasons: summer, winter and monsoon. The figures (in lakhs of Rs.) are given in the following table:

Salesmen				
Season	A	B	C	D
Summer	45	40	38	37
Winter	43	41	45	38
Monsoon	39	39	41	41

Carry out an analysis of variance.

Solution:

Let us take the null hypothesis that there is no significant difference between the sales in the three seasons and also between the sales of the 4 salesmen.

To simplify calculation let us subtract 40 from each value, the new values are

Salesmen					
Season	A	B	C	D	Total
Summer	5	0	-2	-3	0
Winter	3	1	5	-2	7
Monsoon	-1	-1	1	1	0
Total	7	0	4	-4	7

$$\text{Correction factor} = C.F = \frac{G^2}{N} = \frac{(7)^2}{12} = 4.083$$

$$SST = \text{Total sum of squares} = \sum_i \sum_j y_{ij}^2 - C.F$$

$$= [(5)^2 + (0)^2 + (-2)^2 + (-3)^2 + (3)^2 + (1)^2 + (5)^2 + (-2)^2 + (-1)^2 + (-1)^2 + (1)^2 + (1)^2] - 4.083$$

$$= 76.917$$

Between seasons sum of squares

$$R_1 = \frac{(0)^2}{4} + \frac{(7)^2}{4} + \frac{(0)^2}{4} - 4.083 = 8.167$$

Between salesmen sum of squares

$$R_2 = \frac{(7)^2}{3} + \frac{(0)^2}{3} + \frac{(4)^2}{3} + \frac{(-4)^2}{3} - 4.083 = 22.917$$

Error sum of squares

$$R_3 = SST - R_1 - R_2 = 76.917 - 22.917 - 8.167 = 45.833$$

ANOVA table for two-way classification				
Source of variation	Degrees of freedom	Sum of squares (SS)	Mean sum of squares (MS)	Variance Ration (F-Ratio)
Seasons	3 - 1 = 2	8.167	4.0835	$F_1 = \frac{7.639}{4.0835} = 1.87$

Salesmen	$4 - 1 = 3$	22.917	7.639	$F_2 = \frac{7.639}{7.639} = 1$
Error	$3 * 2 = 6$	45.833	7.639	
Total	11	76.917		

$$F_{0.05}(6,2) = 19.3 \text{ and } F_{0.05}(3,6) = 4.76$$

Conclusion:

1. $F_1 < F_{0.05}(6,2)$. Hence we accept the null hypothesis. That is there is no difference between the sales in the seasons.
2. $F_2 < F_{0.05}(3,6)$. Hence we accept the null hypothesis. That is there is no difference between in the sales of the 4 salesmen.

Example: 41

Four different, though supposedly equivalent forms of a standardized reading achievement test were given to each of 5 students, and the following are the scores which they obtained.

		Machine Type			
		A	B	C	D
Workers	1	44	38	47	36
	2	46	40	52	43
	3	34	36	44	32
	4	43	38	46	33
	5	38	42	49	39

1. Test whether the five men differ with respect to mean productivity and
2. Test whether the mean productivity is the same for the four different machine types.

Solution: Let us take the null hypothesis that

1. The 5 workers do not differ with respect to mean productivity

$$i.e., H_{01} : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

2. The mean productivity is the same for the four different machines.

$$i.e., H_{02} : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

To simplify calculation let us subtract 40 from each value, the new values are

		Machine Type				
		A	B	C	D	Total
Workers	1	4	-2	7	-4	5
	2	6	0	12	3	21
	3	-6	-4	4	-8	-14
	4	3	-2	6	-7	0
	5	-2	2	9	-1	8
Total		5	-6	38	-17	20

$$\text{Correction factor} = C.F = \frac{G^2}{N} = \frac{(20)^2}{20} = 20$$

$$SST = \text{Total sum of squares} = \sum_i \sum_j y_{ij}^2 - C.F$$

$$= [(4)^2 + (-2)^2 + (7)^2 + (-4)^2 + (6)^2 + (0)^2 + (12)^2 + (3)^2 + (-6)^2 + (-4)^2 + (4)^2 + (-8)^2 + (3)^2 + (-2)^2 + (6)^2 + (-7)^2 + (-2)^2 + (2)^2 + (9)^2 + (-1)^2] - 20$$

$$= 574$$

Between workers sum of squares

$$R_1 = \frac{(5)^2}{4} + \frac{(21)^2}{4} + \frac{(-14)^2}{4} + \frac{(0)^2}{4} + \frac{(8)^2}{4} - 20 = 181.5 - 20 = 161.5$$

Between machines sum of squares

$$R_2 = \frac{(5)^2}{5} + \frac{(-6)^2}{5} + \frac{(38)^2}{5} + \frac{(-17)^2}{5} - 20 = 358.8 - 20 = 338.8$$

Error sum of squares

$$R_3 = SST - R_1 - R_2 = 574 - 161.5 - 338.8 = 73.7$$

ANOVA table for two-way classification				
Source of variation	Degrees of freedom	Sum of squares (SS)	Mean sum of squares (MS)	Variance Ration (F-Ratio)
Workers	$5 - 1 = 4$	161.5	$\frac{161.5}{4} = 40.375$	$F_1 = \frac{40.375}{6.14} = 6.576$
Machines	$4 - 1 = 3$	338.8	$\frac{338.8}{3} = 112.93$	$F_2 = \frac{112.93}{6.14} = 18.393$
Error	$4 * 3 = 12$	73.7	$\frac{73.7}{12} = 6.14$	
Total	19	574		

$$F_{0.05}(4,12) = 3.26 \text{ and } F_{0.05}(3,12) = 3.49$$

Conclusion:

- $F_1 > F_{0.05}(4,12)$. Hence H_{01} is accepted. That is the 5 workers differ respect to mean productivity.
- $F_2 > F_{0.05}(3,12)$. Hence H_{02} is rejected. That is the mean productivity is not the same for the four machines.

LATIN SQUARE DESIGN (LSD) OR THREE WAY CLASSIFICATION

Example: 45

Set up the analysis of variance for the following results of a Latin Square Design. Use 0.01 level of significance.

A	C	B	D
12	19	10	8
C	B	D	A
18	12	6	7
B	D	A	C
22	10	5	21
D	A	C	B
12	7	27	17

Solution:

Null hypothesis: There is no significant difference between rows, columns and between the treatments.

Columns (j) / Rows (i)	1	2	3	4	Total
1	12	19	10	8	49
2	18	12	6	7	43
3	22	10	5	21	58
4	12	7	27	17	63
Total	64	48	48	53	213

Treatment total $A = 31$, $B = 61$, $C = 85$, $D = 36$

Here $G = 213$ and $N = 16$

$$\text{Correction factor} = \frac{G^2}{N} = \frac{(213)^2}{16} = 2835.56$$

$$\text{Total sum of squares} = \sum_i \sum_j y_{ij}^2 - C.F$$

$$SST = 3483 - 2835.56 = 647.44$$

Between row sum of squares

$$\begin{aligned} R_1 &= \frac{1}{4} (49^2 + 43^2 + 58^2 + 63^2) - 2835.56 \\ &= 2895.25 - 2835.56 = 60.19 \end{aligned}$$

Between column sum of squares

$$\begin{aligned} R_2 &= \frac{1}{4} (64^2 + 48^2 + 48^2 + 53^2) - 2835.56 \\ &= 2878.25 - 2835.56 = 42.69 \end{aligned}$$

Between treatment sum of squares

$$\begin{aligned} R_4 &= \frac{1}{4} (31^2 + 61^2 + 85^2 + 36^2) - 2835.56 \\ &= 3300.75 - 2835.56 = 60.19 \end{aligned}$$

Error sum of squares

$$\begin{aligned} R_4 &= SST - R_1 - R_2 - R_3 \\ R_4 &= 647.744 - 60.19 - 42.69 - 465.19 \\ R_4 &= 79.37 \end{aligned}$$

ANOVA TABLE				
Source of variation	D.o.f	Sum of squares (SS)	Mean Sum of squares (MS)	Variance Ratio (F-ratio)
Between rows	3	60.19	20.06	$F_1 = \frac{20.06}{13.23} = 1.52$
Between columns	3	42.69	14.23	$F_2 = \frac{14.23}{13.23} = 1.08$
between treatments	3	465.19	155.06	$F_1 = \frac{155.06}{13.23} = 11.72$
Error	6	79.37	13.23	
Total	15	647.74		

Here the tabulated value $F_{0.01}(3,6) = 9.78$

Conclusion:

Since $F_1 < F_{0.01}(3,6)$ and $F_2 < F_{0.01}(3,6)$, we accept the null hypothesis and hence we may conclude that there is no significant difference between the rows and columns.

The calculated value of $F_3 > F_{0.01}(3,6)$, and so we conclude that the treatments are significantly different.

Example: 46

Analyze the variance in the Latin Square of yields (in kgs) of paddy where P,Q,R,S denote the different methods of cultivation.

S122	P121	R123	Q122
Q124	R123	P122	S125
P120	Q119	S120	R121
R122	S123	Q121	P122

Examine whether the different methods of cultivation have given significantly different yields.

Solution:

Null hypothesis: There is no significant difference between the different methods of cultivation.

To simplify calculations , we subtract 120 from the given values.

Columns (j) / Rows (i)	1	2	3	4	Total
1	S2	P1	R3	Q2	8
2	Q4	R3	P2	S5	14
3	P0	Q-1	S0	R1	0
4	R2	S3	Q1	P2	8
Total	8	6	6	10	30

Treatment total $P = 31$, $Q = 61$, $R = 85$, $S = 36$

Here $G = 30$ and $N = 16$

$$\text{Correction factor} = \frac{G^2}{N} = \frac{(30)^2}{16} = 56.25$$

$$\text{Total sum of squares} = \sum_i \sum_j y_{ij}^2 - C.F$$

$$SST = 92 - 56.25 = 35.75$$

Between row sum of squares

$$R_1 = \frac{1}{4} (8^2 + 14^2 + 0^2 + 8^2) - 26.25 = 24.75$$

Between column sum of squares

$$R_2 = \frac{1}{4} (8^2 + 6^2 + 6^2 + 10^2) - 56.25 = 2.75$$

Between treatment sum of squares

$$R_4 = \frac{1}{4} (5^2 + 6^2 + 9^2 + 10^2) - 56.256 = 4.25$$

Error sum of squares

$$R_4 = SST - R_1 - R_2 - R_3$$

$$R_4 = 35.75 - 24.75 - 2.75 - 4.2$$

$$R_4 = 4$$

ANOVA TABLE				
Source of variation	D.o.f	Sum of squares (SS)	Mean Sum of squares (MS)	Variance Ration (F-ratio)
Between rows	3	24.75	8.25	$F = \frac{1.417}{0.667} = 2.124$
Between columns	3	2.75	0.917	
between treatments	3	4.25	1.417	
Error	6	4	0.667	
Total	15	35.75		

Here the tabulated value $F_{0.05}(3,6) = 4.76$

Conclusion:

Since $F < F_{0.05}(3,6)$, we accept the null hypothesis and hence we may conclude that there is no significant difference between the different methods of cultivation.

Example: 47

The figures in the following 5*5 Latin square are the numbers of minutes, engines E_1, E_2, E_3, E_4 and E_5 tuned up by mechanics M_1, M_2, M_3, M_4 and M_5 , ran with a gallon of fuel A, B, C, D and E.

	E_1	E_2	E_3	E_4	E_5
M_1	A 31	B 24	C 20	D 20	E 18
M_2	B 21	C 27	D 23	E 25	A 31
M_3	C 21	D 27	E 25	A 29	B 21
M_4	D 21	E 25	A 33	B 25	C 22
M_5	E 21	A 37	B 24	C 24	D 20

Use the level of significance $\alpha = 0.01$ to test

1. The null hypothesis H_0 that there is no difference in the performance of the five engines.
2. H_0 that the persons who tuned up these engines have no effect on their performance.
3. H_0 that the engines perform equally well with each of the fuels.

Solution:

Null hypothesis: there is no significant difference in the performance due to the engines, the persons who tuned them up, or the fuels.

To simplify calculations, we subtract 25 from the given values.

	E_1	E_2	E_3	E_4	E_5	Total
M_1	A6	B-1	C-5	D-5	E-7	-12
M_2	B-4	C2	D-2	E0	A6	2
M_3	C-4	D2	E0	A4	B-4	-2
M_4	D-4	E0	A8	B0	C-3	1
M_5	E-4	A12	B-1	C-1	D-5	1
Total	-10	15	0	-2	-13	-10

Treatment total $A = 36$, $B = -10$, $C = -11$, $D = -14$, $E = -11$

Here $G = -10$ and $N = 25$

$$\text{Correction factor} = \frac{G^2}{N} = \frac{(-10)^2}{25} = 4$$

$$\text{Total sum of squares} = SST = 524 - 4 = 520$$

Between mechanics sum of squares

$$R_1 = 26.8$$

Between engine sum of squares

$$R_2 = 95.6$$

Between fuel sum of squares

$$R_4 = 362.8$$

Error sum of squares

$$R_4 = SST - R_1 - R_2 - R_3$$

$$R_4 = 520 - 26.8 - 95.6 - 362.8 = 34.8$$

$$R_4 = 34.8$$

ANOVA TABLE				
Source of variation	D.o.f	Sum of squares (SS)	Mean Sum of squares (MS)	Variance Ration (F-ratio)
Mechanics	4	26.8	6.7	$F_1 = 2.31$
Engines	4	95.6	23.9	$F_2 = 8.24$
Fuels	4	362.8	90.7	$F_1 = 31.28$
Error	$(5-1)(5-1)=12$	34.8	2.9	
Total	$n^2 - 1 = 24$			

Here the tabulated value $F_{0.01}(4,12) = 5.41$

Conclusion:

1. $F_1 < F_{0.01}(4,12)$, we accept the null hypothesis and we may conclude that the person who tuned up the engines have no effect on their performances.
2. $F_2 > F_{0.01}(4,12)$, we reject the null hypothesis and we may conclude that there is difference in the performance of the five engines.
3. $F_3 > F_{0.01}(4,12)$, we reject the null hypothesis and we may conclude that the engines do not perform equally well with each of the fuels.