Unit – IV

Test of significance: Large sample test for single proportion, difference of proportions, single mean, difference of means, and difference of standard deviations.

Small samples: Test for single mean, difference of means and correlation coefficients – test for ratio of variances – Chi–Square test for goodness of fit and independence of attributes. (12 hours)

Parameters and Statistics

The statistical constants such as mean (μ), variance (σ2) etc of the population is called as parameters and the statistical measures computed from sample such as mean ( ), variance (s2), etc are called as statistic.

Sampling Distributions

Consider all possible samples of size 'n' drawn from a given population at random. We calculate mean values of these samples. If we group these different means according to their frequencies, the frequency distribution so formed is called sampling distribution.

Null hypothesis

It is a definite statement about the population parameter which is usually a hypothesis of no difference and is denoted by H0.

In case of signal statistic, H0 will be that the sample statistic does not differ significantly from the hypothetical parameter value.

In case of two statistics, H0 will be that the sample statistics do not differ significantly.

Alternative hypothesis

Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by H1.

Types of sampling.

(i) Purposive sampling

(ii) Random sampling

(iii) Stratified sampling

(iv) Systematic sampling

One tailed and two tailed test.

When only one tail of the sampling distribution of the normal curve is used, the test is called one-tailed test. One tailed test will be either left tailed or right tailed test. When two tails of the sampling distribution of the normal curve are used, the relevant test is called two-tailed test.

Critical value Level of significance (α)

Level of significance			
Critical value			
	1%	5%	10%
1. Two tailed test	2.58	1.96	1.645
2.Right tailed test	2.33	1.645	1.28
3.Left tailed test	- 2.33	- 1.645	- 1.28

Standard Error

The standard deviation of the sampling distribution of a statistic is known as its standard error, abbreviated as S.E

Test of significance

The procedure which enables us to decide whether to accept or reject the hypothesis is called the test of significance.

Sampling Theory:

1.If $n < 30$ is called small sample.

2.If $n > 30$ is called large sample.

Statistics and parameters:

$\mu, \sigma =$ mean and standard deviation of population.

$x, s =$ mean and standard deviation of sample.

$\sigma^2 =$ Variance of population.

Level of significance		
Critical value		
	5%	1%
1. Two tailed test	1.96	2.58
2.Right tailed test	1.645	2.33
3.Left tailed test	1.645	2.33

If the size of the sample $n > 30$. Then it is called large sample.

Five types of large samples:

i) Test of significance of single proportion.

ii) Test of significance of different proportion.

iii) Test of significance of single mean.

iv) Test of significance of difference of mean.

v) Test of significance of difference standard deviation

Type 1 - Test of significance of single proportion.

PROCEDURE:

Step 1: write P, p, n,

$p = \frac{X}{n}$ = sample proportion

P = Population proportion

n = sample size

2. Assume null hypothesis, H_0

   Alternative hypotheis, H_1

3. Calculate Z value

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

$$Q = 1 - P$$

4. Table value at 5% (or) 1% level of significance.
5. Calculte value less than table value, null hypothesis H_0 is accepted.
6. Calculate value greater than table value, null hypothesis H_0 is rejected.

PROBLEMS:

1. The manufactures claimed that at least 95% of equipment"s which he supplied to a factory conformed to the specifications. An examination of sample of 200 pieces of equipment"s revealed at 18 were faulty, test his claim at 5% level of significance.

Solution:

Step 1:

Here, n = 200

X = no of pieces containing to speciation"s in the sample

X = 200 – 18 = 182

P = 95% = 0.95

Q = 1 – P = 0.05

p = sample proportion conformed to specification

$$p = \frac{x}{n} = \frac{200-18}{200}$$

p = 0.91

Step 2:

Assume null hypothesis H_0 :

P = 0.95

Alternative hypothesis, H_1 :

P $\neq$ 0.95 [Two tailed test]

Step 3: Test statistic

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

$$z = \frac{|0.91-0.95|}{\sqrt{(0.95\times 0.05)/200}}$$

$$z = |-2.5955|$$

$$z_c = 2.5955$$

Step 4:

$\therefore$Table value for 5% level of significance is $z_t = 1.96$

$\therefore$ Calculate value is greater than table value i.e. 2.59 > 1.96

$\therefore H_0$ is rejected.

Conclusion:

Hence the manufacture claim is rejected.

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

Solution:

Step 1:

Given n =1000

X = 540

P = 50% = 0.5

Q = 1 – P = 1 – ½ = ½ = 0.5

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

Step 2:

Assume null hypothesis, H_0

P = 0.5

Alternate hypothesis, H_1:

P $\neq$ 0.5 [Two tailed test]

Step 3: Test statistic

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

$$z = \frac{0.54-0.5}{\sqrt{(0.5\times0.5)/1000}}$$

$z_c = 2.5298$

Step 4:

$\therefore$ Table value for 1% level of significance, $z_t = 2.58$.

$\therefore$ Calculate value is lesser than table value i.e 2.52 < 2.58.

$\therefore H_0$ is accepted.

Conclusion:

We may conclude that both rice and wheat eaters are equally popular in this state.

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

Solution:
Step 1:

Given n= 600

X = 325

P = 50% = 0.5

Q = 1 – P = 0.5

$$p = \frac{x}{n} = \frac{325}{600} = 0.5416$$

Step 2:

Assume null hypothesis, H_0

P = 0.5

Assume Alternate hypothesis, H_1

$P \neq 0.5$ [two tailed test]

Step 3: Test statistic:

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

$$z = \frac{0.5416 - 0.5}{\sqrt{(0.5 \times 0.5)/600}}$$

$$z_c = 2.0379$$

Step 4:

Table value for 5% level of significance, $z_t = 1.96$

$\therefore$ Calculate value isgreater than table value i.e) 2.04 > 1.96

$\therefore$ H_0 is rejected.

Conclusion:

$\therefore$ Majority of men are smokers in the city.

4. In a factory sample 400 parts are manufactured, the no. of deflective parts was found to be 30, the company however claims that only 5% of that product is deflective. [May 2016, Dec 10]

Solution:
Step 1:

Given n = 400

X = 30

$$p = \frac{x}{n} = \frac{30}{400} = 0.075$$

P = population proportion = 5% = 0.05

Q = 1 – P = 0.95

Step 2:

Assume null hypothesis H_0 :

P = 0.05

Assume alternate hypothesis H_1:

P G 0.05 [two tailed test]

Step 3 Test statistic:

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

$$z = \frac{0.075-0.05}{\sqrt{(0.05\times0.95)/400}}$$

$$z_c = 2.941$$

Step 4:

$\therefore$Table value for 5% level of significance is $z_t = 1.96$

$\therefore$Calculate value is greater than the tabulated value. i.e) 2.2941 > 1.96

$\therefore$ H_0 is rejected.

Conclusion: $\therefore$The company claims that 5% of sample is not acceptable.

5. A die was thrown 9000 times and a throw is observed 3220 times. Consider 3 or 4 will be accepted. Show that the die cannot be regarded as an unbiased. [MAY 2017]

Solution:
Step 1:

Given n = 9000

$$p = \frac{x}{n} = \frac{3220}{9000} = 0.3758$$

P = Population proportion of success

= (1/6) + (1/6) = 2/6 = 1/3 = 0.3333

Q = 1 – P = 0.6667.

Step 2:

Assume null hypothesis H_0 :

P = 0.3333

Assume alternate hypothesis H_1:

P G 0.3333[two tailed test]

Step 3: Test statistic:

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

$$z = \frac{0.3578 - 0.333}{\sqrt{(0.333 \times 0.6667)/9000}}$$

$$z_c = 4.9306$$

Step 4:

The table value for 5% level of significance is $z_t = 1.96$

$\therefore$Calculate value is greater than the tabulated value. i.e. 4.9306 > 1.96

$\therefore$ H_0 is rejected.

Conclusion:

Therefore the dice is biased.

6. A manufacture claims that only 4% of his products supplied by him are defective. A random sample of 600 products contained 36 defectives. Test the claim at 5% level of significance.

Solution:

Step 1:

Given n = 600

X = number of defectives = 36

Sample proportion = $p = \frac{x}{n} = \frac{36}{600} = 0.06$

P = 4% = 0.04

Q = 1 – P = 1 – 0.04 = 0.96

Step 2:

Assume null hypothesis, H_0 : P = 0.04 is the proportion of defective is 4%

Alternate hypothesis, H_1 : P > 0.04 [one tailed test]

Step 3: Test statistic

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

$$z = \frac{0.06 - 0.04}{\sqrt{\frac{0.04 \times 0.96}{600}}} = \frac{0.02}{0.008} = 2.5$$

$$z_c = 2.5$$

Step 4:

$\therefore$ Table value for 5% level of significance, z_t = 1.645.

$\therefore$ Calculate value is lesser than table value i.e 2.5 > 1.645.

$\therefore H_0$ is Rejected.

Conclusion:

We may conclude that the manufacture claims that only 4% defective are not accepcted.

<u>NOTE:</u>

1. 98% of confident limit,

$$P = p \pm 2.33\sqrt{pq/n}$$

2. 99% of confident limit,

$$P = p \pm 2.58\sqrt{pq/n}$$

3. 95% of confident limit,

$$P = p \pm 1.96\sqrt{pq/n}$$

4. Limits for population $P = p \pm 3\sqrt{pq/n}$

1. A random sample of 500 apples was taken from a large container and 60 were found to be bad. Obtain 98% of confident limit for the number of bad apples in the container.

Solution:

Step 1:

Given n = 500, X = 60

p = 60/500 = 0.12

q = 1 – p = 0.88

Step 2:

w.k.t.

98% of confident limit for population proportion is given by,

$$P = p \pm 2.33\sqrt{pq/n}$$

$$P = 0.12 \pm 2.33\sqrt{0.12 \times 0.88/500}$$

$$P = 0.12 \pm 0.388$$

P = 0.1538 ; P = 0.0862

Conclusion:

∴ 98% of confident limit for the no. of bad appler in the container are 0.1538,0.0862.

2. A random sample of 500 pineapples was taken from a large container and 65 were found to be bad. Find the percentage of bad pineapples in the container.

Solution:
Step 1:

Given n = 500

X = 65

p = 65/500 = 0.13

q = 1 – p = 0.87

Step 2:

Limits for population proportion is given by

$$P = p \pm 3\sqrt{pq/n}$$

$$P = 0.13 \pm 3\sqrt{0.13 \times 0.87/500}$$

$$P = 0.12 \pm 0.0451$$

P = 0.1751 ; P = 0.0849

Conclusion:

∴ The percentage of bad pineapples in the container are 0.1751 , 0.0849.

Type 2 - Test of significance of different proportion:

1. If sample proportions $p_1,\ p_2$ are given then $z = \dfrac{|p_1 - p_2|}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

To find P:

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

To find Q:

Q = 1 – P

2. If the sample proportions are not given

   i.e Population proportion are given, then $z = \dfrac{|p_1 - p_2|}{\sqrt{\left(\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}\right)}}$

3. If difference is given in proportion, then $z = \dfrac{|(p_1 - p_2) - (P_1 - P_2)|}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

4. Test the signi difference b/w both sample and population proportion $z = \dfrac{|p_1 - P|}{\sqrt{\dfrac{n_2 PQ}{n_1(n_1 + n_2)}}}$

1. A random sample of 400 men and 600 women were asked whether they would like to have a school near this residency. 200 men and 325 women were in favor of their proposal. Test the hypothesis that proportion of men and women in favor of the proposal are same at 5% level. [MAY 2017]

SOLUTION:

Step 1:

$n_1 = 400$	$n_2 = 600$
$X_1 = 200$	$X_2 = 325$
$p_1 = 200/400$	$p_2 = 325/600$
$p_1 = 0.5$	$p_2 = 0.5417$
$q_1 = 1 - P_1 = 0.5$	$q_2 = 1 - P_2 = 0.4583$

Step 2:

Null hypothesis, $H_0: \; p_1 = p_2$

Alternate hypothesis, $H_1 : \; p_1 \neq p_2$ (Two tailed test)

Step 3: Test Statistic

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

Where

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

$$P = \frac{(400 \times 0.5) + (600 \times 0.541)}{400 + 600}$$

$P = 0.5250$

Q = 1 – P = 0.4749

Substutite in z

$$z = \frac{|0.5 - 0.54171|}{\sqrt{(0.5250 \times 0.4749)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$z_c = 1.2937$

Step 4:

At 5% level of significance, z_t = 1.96

$\therefore$calculated value < Tabulated value

i.e H_0 is accepted.

Conclusion:

$\therefore$ The proportion of men and women in favour of proposals are same.

2. In a two large population there are 30% and 25% respectively of fair haired people. In this difference likely to be hidden in a sample of 1200 and 900 respectively from the two population.

SOLUTION:

Step 1:

n_1 = 1200	n_2 = 900
P_1 = 30% = 0.3	P_2 = 25% = 0.25
Q_1 =1 – P_1 =0.7	Q_2 =1 – P_2 = 0.75

Step 2:

Assume Null hypothesis, $H_0: p_1 = p_2$

Alternate hypothesis, $H_1 : p_1 \neq p_2$ (Two tailed test)

Step 3: Test Statistic

$$z = \frac{|p_1 - p_2|}{\sqrt{\left(\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}\right)}}$$

$$z = \frac{|0.3 - 0.25|}{\sqrt{\frac{0.3\times 0.7}{1200} + \frac{0.25\times 0.75}{900}}}$$

$$z_c = 2.5537$$

Step 4:

At 5% level of significance, $z_t = 1.96$.

The calculated value > tabulated value

i.e H_0 is rejected.

Conclusion:

Therefore the difference is unlikely to hidden in a samples.

3. The manufacturing form claim that its brand A product out sells the brand B product by 8%, it found that 42 out of sample of 200 persons prefer brand A and 18 out of another sample of 100 persons prefer brand B. Test whether 8% difference is valid claim.

SOLUTION:

$n_1 = 200$ $\quad$ $n_2 = 100$

$X_1 = 42$ $\quad$ $X_2 = 18$

$p_1 = 42/200$ $\quad$ $p_2 = 18/100$

$p_1 = 0.21$ $\quad$ $p_2 = 0.18$

$q_1 = 1 - P_1 = 0.79$ $\quad$ $q_2 = 1 - P_2 = 0.82$

The difference between population proportion is $P_1 - P_2 = 8\ \% = 0.08$

Step 2:

Assume Null hypothesis, $H_0 : P_1 - P_2 = 0.08$

Assume alternate hypothesis, $H_1: \ P_1 - P_2 \neq 0.08$

Step 3: Test Statistic

$$z = \frac{\left|(p_1 - p_2) - (P_1 - P_2)\right|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P = \frac{(0.21 \times 200) + (0.18 \times 100)}{200 + 100} = 0.2$$

Q = 1 – P = 0.8

$$z = \frac{\left|(0.21 - 0.18) - 0.08\right|}{\sqrt{(0.2 \times 0.8)\left(\frac{1}{200} + \frac{1}{100}\right)}}$$

$$z = \frac{0.05}{0.0489}$$

z_c = 1.0206.

Step 4:

At 5% level of significance, z_t = 1.96.

∴ The tabulated value is greater than th calculated value

i.e. H_0 is accepted.

Conclusion:

∴ The 8% difference in brand A and brand B is a valid claim.

4. In a random sample of 400 students of university teaching department, it was found that 300 students failed in the examination. In another random sample of 500 students of the affiliated college, the no. of failures in the same examination was found to be 300. Find out whether the proportion of failures in the university teaching department is significantly greater than the proportion of failures in the university teaching department and affiliated colleges taken together.

SOLUTION:

Step 1:

$n_1 = 400$ $\qquad$ $n_2 = 500$

$X_1 = 300$ $\qquad$ $X_2 = 300$

p_1 = 300/400 $\qquad$ p_2 = 300/500

p_1 =0.75 $\qquad$ p_2 =0.6

q_1 =1 – P_1 = 0.25 $\qquad$ q_2 =1 – P_2 = 0.4

p_1 = population proportion of university teaching department.

P_1 = Population proportion of both university teaching department and affiliated colleges

Step 2:

Assume Null hypothesis, $H_0: P_1 = p$

Alternate hypothesis, $H_1 : P_1 \neq p$ (Two tailed test)

Step 3: Test Statistic

$$z = \frac{|p_1 - P_1|}{\sqrt{\dfrac{n_2 PQ}{n_1(n_1+n_2)}}} \qquad P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$P = \frac{(400\times 0.75)+(500\times 0.6)}{400+500}$$

= 0.6667

Q = 1 – P = 0.3333

$$z = \frac{|0.75 - 0.6667|}{\sqrt{\dfrac{500\times 0.6667\times 0.333}{400(900)}}}$$

z_c = 4.7416

Step 4:

At 5% level of significance, z_t = 1.96

∴ Calculated value is greater than the tabulated value.

i.e. H_0 is rejected.

Conclusion: The proportion of failures in the university teaching department is greater than proportion of failures in the university teaching department and affiliated colleges taken together

5. A sample survey results show that out of 800 literate people 480 are employed whereas out of 600 illiterate people only 350 are employed. Can the difference between two proportions of employed persons be ascribed due to sampling fluctuations?

Solution:

$n_1 = 800$ $\quad n_2 = 600$

$X_1 = 480$ $\quad X_2 = 350$

$p_1 = 480/800$ $\quad p_2 = 350/600$

$p_1 = 0.6$ $\quad p_2 = 0.58$

Step 2:

Null hypothesis, H_0: $p_1 = p_2$

Alternate hypothesis, $H_1 : p_1 \neq p_2$ (Two tailed test)

Step 3: Test Statistic

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

Where

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

$$P = \frac{(800)\times(0.6)+(600)\times(0.58)}{800+600}$$

$P = 0.5914$

Q = 1 – P = 0.4086

$$Z = \frac{|0.6 - 0.58|}{\sqrt{(0.5914*0.4086)\left(\frac{1}{800} + \frac{1}{600}\right)}}$$

$z_c = 0.7533$

Step 4:

At 5% level of significance, $z_t = 1.96$

∴calculated value < Tabulated value

i.eH_0 is accepted.

Conclusion:

∴ The difference between two proportions of employed persons are same.

6. In a large city A, 20% of random sample of 900 school boys had a slight physical defect. In another large city B, 18.5% of random sample of 1600 school boys had the same defect. Is there any difference between the proportions of significance?

SOLUTION:

Step 1:

$n_1 = 900$ $n_2 = 1600$

$p_1 = 20\% = 0.2$ $p_2 = 18.5\% = 0.185$

$q_1 = 1 - P_1 = 0.8$ $q_2 = 1 - P_2 = 0.815$

Step 2:

Null hypothesis, H_0: $p_1 = p_2$

Alternate hypothesis, H_1 : $p_1 \neq p_2$ [two tailed test]

Step 3: Test Statistic

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

Where

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

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

$P = 0.1904$

Q = 1 – P = 0.8096

Substutite in z

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

$z_c = 0.9169$

Step 4:

At 5% level of significance, the tabulated value for z_t = 1.96.

Here, tabulated value is greater than calculated value.

$\therefore H_0$ is accepted.

Conclusion: Thus there is no significant difference between population proportion and sample proportion.

7. A machine produced 20 defective units in a sample of 400. After overhauling the machine it produced 10 defective units in a batch of 300. Has the machine improved in production due to Overhauling? Test it a 5% L.O.S.

Step 1:

$n_1 = 400$	$n_2 = 300$
$X_1 = 20$	$X_2 = 10$
$p_1 = 20/400$	$p_2 = 10/300$
$p_1 = 0.05$	$p_2 = 0.033$

Step 2:

Null hypothesis, H_0: $p_1 = p_2$

Alternate hypothesis, H_1 : $p_1 \neq p_2$ (Two tailed test)

Step 3: Test Statistic

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

Where

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

$$P = \frac{(400)\times(0.05) + (300)\times(0.033)}{400 + 300}$$

= 0.0427

Q = 1 – P = 0.9573

$$Z = \frac{|0.05 - 0.033|}{\sqrt{(0.0427 * 0.9573)\left(\frac{1}{400} + \frac{1}{300}\right)}}$$

z_c = 1.103

Step 4:

At 5% level of significance, z_t = 1.96

$\therefore$calculated value <Tabulated value

i.e H_0 is accepted.

Conclusion:

∴ The proportion of men and women in favour of proposals are same.

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 batch of 100.Has the machine improved?

Solution:

Step 1:

$n_1 = 500$	$n_2 = 100$
$X_1 = 16$	$X_2 = 3$
$p_1 = 16/500$	$p_2 = 3/100$
$p_1 = 0.032$	$p_2 = 0.03$

Step 2:

Null hypothesis, H_0: Assuming that there is no improvement in the machine.

H_0: $p_1 = p_2$

Alternate hypothesis, H_1 : $p_1 \neq p_2$

Step 3: Test Statistic

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

Where

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

$$P = \frac{(500 \times 0.032) + (100 \times 0.03)}{500 + 600}$$

$= 0.0316$

Q = 1 – p = 0.9684

$$Z = \frac{|0.032 - 0.03|}{\sqrt{(0.0316 * 0.9684)\left(\frac{1}{500} + \frac{1}{100}\right)}}$$

$z_c = 1.04$

Step 4:

At 5% level of significance, tabulated $z_t = 1.64$

∴calculated value $<$ Tabulated value

i.e H_0 is accepted.

Conclusion:

∴ There is no improvement in the machine.

TYPE 3 - TEST OF SIGNIFICANCE OF SINGLE MEAN:

To test whether the difference between population mean and sample mean is significant (or) not,

i) If σ is given, then $z = \dfrac{|\bar{x}-\mu|}{\sigma/\sqrt{n}}$

ii) If σ is not given, then $z = \dfrac{|\bar{x}-\mu|}{s/\sqrt{n}}$

Where $\dot{x}$ = sample mean

μ = population mean

σ = population standard deviation

s = sample standard deviation

n = sample size

CONFIDENT LIMIT (OR) FIDUCIAL LIMIT:

1. 95% of confident limit, $\mu = \bar{x} \pm 1.96\dfrac{\sigma}{\sqrt{n}}$
2. 98% of confident limit, $\mu = \bar{x} \pm 2.33\dfrac{\sigma}{\sqrt{n}}$
3. 99% of confident imit, $\mu = \bar{x} \pm 2.58\dfrac{\sigma}{\sqrt{n}}$

1. A sample of 900 members has a mean of 3.4cm and S.D. is 2.61cm. If the sample from a large population of mean 3.25cm and S.D is 2.61cm. If the population is normal and its mean is unknown, find the 95% of fiducial limit of the true mean.

SOLUTION:

Step 1:

n = 900

$\bar{x} = 3.4cm$

s = 2.61cm

$\mu = 3.25cm$

$\sigma = 2.61cm$

Step 2:

Assume null hypothesis, $H_0 : \mu = 3.25$

Assume alternate hypothesis, $H_1 : \mu \neq 3.25$

Step 3: Test Statistic

Here, σ is given then $z = \frac{|\bar{x} - \mu|}{\sigma / \sqrt{n}}$

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

$$z_c = 1.7241$$

Step 4:

At 5% level of significance, the tabulated value of z_t is 1.96.

The calculated value <tabulated value .

i.e, H_0 is accepted.

Conclusion:

There is no significant difference between population mean and sample mean

At 95% of confident limit are

$$\mu = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

$$= 3.4 \pm 1.96 \left(2.61 / \sqrt{900} \right)$$

$$= 3.4 \pm 0.1705$$

$$\mu = 3.5705; \ \mu = 3.2295$$

2. **A normal population has a mean of 6.48 and a S.D of 1.5 In a sample of 400 members, mean is 6.75. Is the difference significant?**

Solution:

Step 1:

GIVEN:

n = 400

$\mu = 6.48$

$\sigma = 1.5$

$\bar{x} = 6.75$

Step 2:

Assume null hypothesis, $H_0 : \mu = 6.48$

Assume alternate hypothesis, $H_1 : \mu \neq 6.48$

Step 3: Test Statistic

Here, σ is given then $z = \dfrac{|\overline{x} - \mu|}{\sigma / \sqrt{n}}$

$$z = \frac{|6.75 - 6.48|}{1.5 / \sqrt{400}}$$

$z_c = 3.6.$

Step 4:

At 5% level of significance, the tabulated value of z_t is 1.96.

$\therefore$ Tabulated value < calculated value.

i.e. H_0 is rejected.

Conclusion:

$\therefore$ There is significant difference between population mean and sample mean.

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

Age:	**16 - 20**	**21 – 25**	**26 – 30**	**31 – 35**	**36 – 40**
No. of persons:	**12**	**22**	**20**	**30**	**16**

Calculate the arithmetic mean and S.D. of this distribution and use these values to test his claim at 5% level of significance.

SOLUTION:

Step 1:

n = 100

μ = 30.5

A= middle value of x_i = 28

h = common difference = 5

Age	No. of persons (f)	Mid pt x_i	$d = \frac{x_i - A}{h}$	fd	fd^2
16 – 20	12	18	-2	-24	48
21 – 25	22	23	-1	-22	22
26 – 30	20	**28 = A**	0	0	0
31 – 35	30	33	1	30	30
36 – 40	16	38	2	32	64
	100			16	164

Arithmetic mean, $\bar{x} = A + \frac{h\sum fd}{\sum f}$

$$= 28 + \frac{5 \times 16}{100}$$

$$= 28.8$$

Standard deviation, $s = h\sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$

$$= \quad s = 5\sqrt{\frac{164}{100} - \left(\frac{16}{100}\right)^2}$$

s = 6.453

Step 2:

Assume null hypothesis, H_0:

$\mu = 30.5$

Assume alternate hypothesis, H_1 :

μ G 30.5

Step 3: Test Statistic

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

$$z = \frac{|28.8-30.5|}{\frac{6.3529}{\sqrt{100}}}$$

$$z_c = 2.6344$$

Step 4:

At 5% level of significance the tabulated value of $z_t = 1.96$

$\therefore$ Calculatedvalue >Tabulated value.

i.e. $H_0$is rejected.

Conclusion:

His claim is rejected.

TYPE – 4

TEST OF SIGNIFICANCE OF DIFFERENCE OF MEAN

To test whether the significant difference between $\bar{x}_1$ and $\bar{x}_2$

1. If the population S.D is given (i.e) σ is given then,

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

2. If $\sigma_1 = \sigma_2 = \sigma$

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

3. If the population standard deviation is not given (i.e) σ is not given

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

PROBLEMS

1. **The mean of two large sample of size 2000 and 1000 of the mean are 68.0 and 67.5 respectively can the sample mean regarded as drawn from the same population standard deviation of 2.25. [Nov 2014]**

Solution:

Step 1:

$$\bar{x}_1 = 68$$

$$\bar{x}_2 = 67.5$$

$$n_1 = 2000$$

$$n_2 = 1000$$

$$\sigma = 2.25$$

Step 2:

Null hypothesis $H_0\ \mu_1 = \mu_2$

Alternate hypothesis $H_1\ \mu_1 \neq \mu_2$ (two tailed test)

Step 3: Test statistics

Here $\sigma_1 = \sigma_2 = \sigma$

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

$$z = \frac{68 - 67.5}{2.25\sqrt{\frac{1}{2000} + \frac{1}{1000}}}$$

$$z = \frac{0.5}{2.25(0.03873)}$$

$$z_c = 5.7377$$

Step 4:

The tabulated value for 5% level of significance of two tailed test is

$$z_t = 1.96$$

$$z_c > z_t$$

$$\therefore H_0 \text{ is rejected}$$

Conclusion:

$\therefore$ The sample mean cannot be regarded as drawn from the same population standard deviation of 2.25

2. Given that the information relating to two places A and B test whether there is any significant difference between two mean wages

	Place A	Place B
Mean wages	**47**	**49**
S.D	**28**	**40**
No. of workers	**1000**	**1500**

Solution:

Step 1:

$n_1 = 1000 \qquad n_2 = 1500$

$\bar{x}_1 = 47 \qquad \bar{x}_2 = 49$

$s_1 = 28 \qquad s_2 = 40$

Step 2:

Null hypothesis $H_0 \quad \mu_1 = \mu_2$

Alternate hypothesis $H_1 \; \mu_1 \neq \mu_2$ (two tailed test)

Step 3: Test Statistics

Here σ is not given

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

$$z = \frac{47 - 49}{\sqrt{\frac{28^2}{1000} + \frac{40^2}{1500}}}$$

$$z = \frac{2}{1.3604}$$

$$z_c = 1.4701$$

Step 4: The tabulated value of 5% level of significance for two tailed test z_t = 1.96

$$z_c > z_t$$

$\therefore H_0$ is rejected

Conclusion: $\therefore$ There is no significance difference between two mean wages

3. Two random sample of size 1600 and 2000 of farms gave a yield of 2000kg and 2050kg respectively. The variance of the farms in country may be taken as 100kg. Examine whether the two samples differ significantly in the yield.

Solution:

Step 1:

$n_1 = 1600 \qquad n_2 = 2000$

$\overline{x}_1 = 2000 \qquad \overline{x}_2 = 2050$

$\sigma^2 = 100 \qquad \sigma = 10$

Step 2:

Null hypothesis H_0

$\mu_1 = \mu_2$

Alternate hypothesis H_1

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

Step 3: Test statistics

Here σ is given

$\sigma_1 = \sigma_2 = \sigma$

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

$$z = \frac{2000 - 2050}{10\sqrt{\frac{1}{1600} + \frac{1}{2000}}}$$

$$z = \frac{50}{10 \times 0.03354}$$

$$z = \frac{50}{0.3354}$$

$$z_c = 149.076$$

Step 4:

The tabulates value for two tailed test of 5% level of significance $z_t = 1.96$

$z_c > z_t$

$\therefore H_0$ is rejected

Conclusion:

$\therefore$ There is significance difference between samples in the yield

4. The following data are got from an investigation.

	No. of cases	Mean wages	S.D of the wages
Sample I	**400**	**Rs.47.4**	**Rs.3.1**
Sample I	**900**	**Rs.50.3**	**Rs.3.3**

Find out whether the two mean wages differ significantly.

Solution:

Step 1:

$n_1 = 400 \qquad n_2 = 900$

$\overline{x}_1 = 47.4 \qquad \overline{x}_2 = 50.3$

$s_1 = 3.1 \qquad s_2 = 3.3$

Step 2:

Null hypothesis $H_0 \quad : \overline{x_1} = \overline{x_2}$

Alternate hypothesis $H_1 : \overline{x_1} \neq \overline{x_2}$ (two tailed test)

Step 3: Test Statistics

Here σ is not given

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

$$z = \frac{47.4 - 50.3}{\sqrt{\frac{3.1^2}{400} + \frac{3.3^2}{900}}}$$

$$= \frac{-2.9}{\sqrt{0.0240 + 0.0121}} = \frac{-2.9}{0.19} = -15.263$$

$\therefore \quad |Z| = 15.263$

Step 4:

The tabulated value of 5% level of significance for two tailed test z_t = 1.96

$z_c > z_t$

$\therefore H_0$ is rejected

Conclusion:

$\therefore$ There is a significant difference between two mean wages.

Type 5: Test of Significance of SD

1. If the sample S.D (or) population S.D is given then

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma^2{}_1}{2n_1} + \frac{\sigma_2^{\ 2}}{2n_2}}}$$

2. If σ is not given

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

Problem:

1. The S.D of the random sample of 900 members is 4.6 & that of another independent sample of 1600 members is 4.8 .Examine together the S.D is significantly different. [Nov 13, 16, 17]

Step 1: Given:

$n_1 = 900$ $n_2 = 1600$ $s_1 = 4.6$ $s_2 = 4.8$

Step 2:

Assume null Hypothesis H_o

$\sigma_1 = \sigma_2$

Alternate Hypothesis H_1

$\sigma_1 \neq \sigma_2$(Two tailed test)

Step 3: Test statistics

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

$$Z = \frac{-0.2}{\sqrt{\frac{4.6}{1800} + \frac{4.8}{3200}}}$$

$Z_c = 1.4526$

Step 4:

The tabulated value of z at 5% level of Significance is 1.96.

$Z_c < Zt$

$H_o$is accepted

Conclusion: There is no significant difference between S.D

2. **A random sample drawn from the 2 countries A&B .If the following data relating to the height of the male adults**

	A	**B**
Maximum height	**67.42**	**67.25**
No.of males	**1000**	**1200**
S.D	**2.58**	**2.5**

(i) **If the difference of the significance of two mean**

(ii) **If the difference of the significance of two S.D**

Solution:

Step1:

Given:

$n_1 = 1000$ $\quad$ $n_2 = 1200$

$\overline{x}_1 = 67.42$ $\quad$ $\overline{x}_2 = 67.25$

$s_1 = 2.58$ $\quad$ $s_2 = 2.5$

Step 2:

(i) **Difference between S.D**

Assume null Hypothesis H_o

$$\sigma_1 = \sigma_2$$

Alternate Hypothesis H_1

$$\sigma_1 \neq \sigma_2 \text{(two tailed test)}$$

Step 3: Test statistics

Here σ is not given, then

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

$$= \frac{0.03}{\sqrt{\frac{2.58}{2000} + \frac{2.5}{2400}}}$$

$Z_c = 1.038$

Step 4:

The Tabulated value of z at 5% level of significance is 1.96

$Z_c < Zt$

H_0 is accepted.

Conclusion:

There is no significance difference between the two S.D.

(ii) Difference between mean

Step 1: Same

Step 2:

Assume null Hypothesis H_o

$\mu_1 = \mu_2$

Alternate Hypothesis H_1

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

Step 3: Teststatistics

$$Z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s^2{}_1}{n_1} + \frac{s^2{}_2}{n_2}}}$$

$$Z = \frac{67.42 - 67.25}{\sqrt{\frac{(2.58)^2}{1000} + \frac{(2.5)^2}{1200}}}$$

$Z_c = 1.567$

Step 4:

The tablulated value of Z at 5% level of significance is 1.96

$Z_c < Z_t$

$H_o$is accepted

Conclusion:

Ther is no significance difference between two mean

3. The standard deviation of a random sample of 1000 is found to be 2.6 and the standard deviation 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 standard deviation. [MAY 2017]

Solution:

Step 1: <u>Given:</u>

$n_1 = 1000$, $s_1 = 2.6$

$n_2 = 500$ $s_2 = 2.7$

Step 2:

Assume null Hypothesis H_o

$\sigma_1 = \sigma_2$

Alternate Hypothesis H_1

$\sigma_1 = \sigma_2$(Two tailed test)

Step 3: Test statistics

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

$$z = \frac{|2.6 - 2.7|}{\sqrt{\frac{6.76}{2000} + \frac{7.29}{1000}}}$$

$$z = \frac{|-0.1|}{\sqrt{3380+1822.5}} = \frac{0.1}{\sqrt{5202.5}}$$

$$z_c = 1.386$$

Step 4:

The tabulated value of z at 5% level of Significance is 1.96.

$Z_c < Zt$

H_o is accepted

Conclusion: There is no significant difference between S.D

Small Samples

STUDENT‟S „t‟ TEST FOR SINGLE MEAN

Suppose we want to test

(a) If a random sample x_i of size n has been drawn from a normal population with a specified mean μ_0 .

(b)If the sample mean differs significantly from the hypothetical value μ_0 of the population mean. In this case the statistic is given by

1. If the S.D of the samples „s‟ or S.D of the population „σ‟ is given directly then

$$t = \frac{|\bar{x} - \mu|}{\sigma / \sqrt{n-1}} \quad or \quad t = \frac{|\bar{x} - \mu|}{s / \sqrt{n-1}}$$

2. If the SD of the sample is not given directly then

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

Where $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$ *where* $\bar{x} = \frac{\sum x}{n}$

where $\bar{x}$ = Sample mean, μ = Population Mean,

s = Sample standard deviation, n = sample size

Problems

1. A machinist is making engine parts with axle diameters of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a S.D of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specification. [May 2015]

Solution:

Step 1:

Here the sample size n = 10 < 30.

Hence the sample is small sample.

Given $\bar{x}$ = 0.742 inches μ = 0.700 inches

S.D of s = 0.040 inches n = 10

Step 2:

Null Hypothesis $H_0 = \mu$ = 0.700.The product is conforming to specification

Alternative Hypothesis $H_1 : \mu \neq 0.70$

Step 3: The test statistic

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

$$t = \frac{0.742 - 0.700}{0.040 / \sqrt{10-1}}$$

t = 3.15

Step 4:

Degrees of Freedom (d.f) = n – 1 = 10 – 1 = 9

The table value of t at 5% level with 9 degrees of freedom, $t_{0.05} = 2.26$

Since calculated value of t > tabulated value of t

H_0 is rejected.

Conclusion:

Therefore, the product is not meeting the specification.

2. **The mean weekly sales of soap bars in departmental stores was 146.3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a S.D. of 17.2. Was the advertising campaign successful?**

Solution:

Step 1:

Here the sample size n = 22 < 30.

Hence the sample is small sample.

Here n = 22 $\bar{x}$ = 153.7

μ = 146.3 s = 17.2

Step 2:

Null Hypothesis $H_0 = \mu = 146.3$ The advertising campaign was not successful

Alternative Hypothesis $H_1: \mu \neq 146.3$ (Two tailed test)

Step 3: The test statistic

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

$$t = \frac{153.7-146.3}{17.2 \Big/ \sqrt{22-1}}$$

t = 1.97

Step 4:

Degrees of Freedom (d.f) = n – 1 = 22 – 1 = 21

The table value of t at 5% level of significance is 2.08

Since calculated value of t < tabulated value of t

H_0 is accepted

Conclusion:

We accept the assumption that the advertising campaign was not successful.

i.e. Advertising campaign was successful.

3. The average breaking strength of steel rods is specified to be 18.5 thousand pounds. To test this a sample of 14 rods was tested. The mean and standard deviations obtained were 17.85 and 1.955 respectively. Is the result of the experiment significant?

Solution:

Step 1:

Here the sample size n = 14 < 30.

Hence the sample is small sample.

Given Also sample mean $\bar{x}$ = 17.85 and the population mean μ = 18.5 are given.

S.D(s) = 1.955

Step 2:

Null Hypothesis H_0 : μ = 18.5 The result of the experiment is not significant

Alternative Hypothesis H_1: $\mu \neq 18.5$

Step 3: The test statistic

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

$$t = \frac{17.85-18.5}{1.955 \Big/ \sqrt{14-1}} \quad = \quad -1.199$$

|t| = 1.199

Step 4:

Degrees of Freedom = n – 1 = 13

The table value of t at 5% level with 13 degrees of freedom, $t_{0.05} = 2.16$

Since calculated value of t < tabulated value of t

we accept H_0 at 5% level.

Conclusion:

The result of the experiment is not significant.

4. A random sample of size 16 valves from a normal population showed a mean of 53 and a sum of squares of deviation from the mean equals to 150. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% confidence limits of the mean of the population.

Solution:

Step 1:

Given, Sample size, n = 16

Sample Mean, $\bar{x} = 53$

$$\sum(x-\bar{x})^2 = 150$$

Therefore, $s = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n}} = \sqrt{\dfrac{150}{16}} = \sqrt{9.375}$

s = 3.0618

Step 2:

Null Hypothesis $H_0 = \mu = 56$ The sample is taken from the population having 56 as mean.

Alternative Hypothesis $H_1 : \mu \neq 56$

Step 3: The test statistic

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

$$t = \frac{53-56}{3.0618/\sqrt{15}} = -3.7947$$

$|t| = 3.7947$

Step 4:

Degrees of freedom d.f = n – 1 = 15

Since calculated t value > tabulated t value

H_0 is rejected.

Conclusion:

The sample cannot be regarded as taken from the population

The 95% confidence limit of the mean of the population is given by

$$\mu = \dot{x} \pm t_{0.05}\frac{s}{\sqrt{n}} = 53 \pm 2.13 \times 0.79 = 53 \pm 1.6827 = 54.68 \text{ and } 51.31$$

Hence 95% confidence limit is [54.68, 51.31].

STUDENT"S t - TEST (When S.D. of the sample is not given directly)

5. A random sample of 10 boys had the following I.Q."s of 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. To these data support the assumption of a population mean of I.Q of 100? Find a reasonable range in which most of the mean I.Q values of samples of 10 boys lie. [May 16]

Solution:

Step1:

Given n = 10

$\mu = 100$

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

Here S.D. and mean of sample is not given directly. We have to determine these S.D. and mean as follows:

x	$(x-\bar{x})$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.48

98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
972		**1833.60**

We know that Standard Deviation $s = \sqrt{\frac{\sum(x-\overline{x})^2}{n}} = \sqrt{\frac{1833.60}{10}} = \sqrt{183.6}$

Therefore, $s = 13.541$

Step 2:

Null Hypothesis H_0 = μ = 100 The data support the assumption of a population mean I.Q. of 100 in the population.

Alternative Hypothesis $H_1 : \mu \neq 100$

Step 3: The test statistic

$$t = \left| \frac{\overline{x} - \mu}{s / \sqrt{n-1}} \right|$$

$$t = \frac{97.2 - 100}{13.541 / \sqrt{9}} = -0.6203$$

$|t| = 0.6203$

Step 4:

The table value of t at 5% level of significance for 9d.f. for two tailed test is 2.26

Since calculated t value < tabulated t value

we accept the null hypothesis H_0.

Conclusion:

Hence the data support the assumption of mean I.Q. of 100 in the population.

The 95% confidence limit of the mean of the population is given by

$$\mu = \dot{x} \pm t_{0.05} \frac{s}{\sqrt{n}} = 97.2 \pm 2.26 \times 4.514 = 97.2 \pm 10.20$$

$$\mu = 107.40 \; and \; 87.00$$

Hence 95% confidence limit is [87.00, 107.40].

6. Certain pesticide is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to weigh (in kg) as follows: 50, 49, 52, 44, 45, 48, 46, 45, 49, 45. Test if the average packing can be taken to be 50 kg. ($t_{0.05}$ for 9d.f. = 2.262) [Nov 2016, Dec 2010]

Solution:

Step 1:

Given n = 10

$\mu = 50$

Mean $\bar{x} = \frac{\sum x}{n} = \frac{473}{10} = 47.3$

Calculation for sample mean and S.D

x	$(x-\bar{x})$	$(x-\bar{x})^2$
50	2.7	7.29
49	1.7	2.89
52	4.7	22.09
44	-3.3	10.89
45	-2.3	5.29
48	0.7	0.49
46	-1.3	1.69
45	-2.3	5.29
49	1.7	2.89
45	-2.3	5.29
473		**64.1**

We know that $s = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} = \sqrt{\frac{64.1}{10}} = \sqrt{6.41}$

Therefore, $s = 2.478$

Step 2:

Null Hypothesis $H_0 : \mu = 50$ The average packing is 50 kg. i.e. $\mu = 50$ Kg

Alternative Hypothesis H_1: $\mu \neq 50$

Step 3: The test statistic

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

$$t = \frac{47.3 - 50}{2.478/\sqrt{9}} = -3.27$$

$|t| = 3.27$

Step 4:

The table value of t at 5% level of significance for 9 d.f. for two tailed test is 2.262

Since calculated t value > tabulated t value

we reject the null hypothesis H_0.

Conclusion:

The average packing is not 50 Kg.

9. Test whether the sample having the values 63, 63, 64, 55, 69, 70, 70 and 71 has been chosen from a population with mean of 65 at 5% level of significance.

Solution:

Step 1:

n = 8 $\mu = 65$

Step 2:

Null Hypothesis H_0 = There is no significance difference between sample mean and population mean. i.e. $\mu = 65$

Alternative Hypothesis H1: $\mu \neq 65$

Step 3:

$$\bar{x} = \frac{\sum x}{n} = \frac{525}{8} = 67.63$$

Calculation for sample means and S.D.s

x	$(x-\bar{x})^2$
63	21.437
63	21.437
64	13.1769
55	159.52
69	1.877
70	5.617
70	5.617
71	11.36
	$\sum(x-\bar{x})^2 = 240.04$

Therefore,

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{240.04}{8}} = \sqrt{30}$$

$$s = 5.4772$$

Therefore, $s = 5.4772$

Now,

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

$$t = \frac{67.63-65}{5.4772/\sqrt{7}} = \frac{2.63}{2.07}$$

$$t = 1.2704$$

i.e. Calculated $t = 1.2704$

Step 4:

Degrees of freedom = n – 1 = 7

Tabulated t for 7 d.f. at 5% level = 2.365.

Since calculated t < tabulated t

we accept the null hypothesis H_0.

Conclusion:

There is no significance difference between sample mean and population mean.

STUDENT"S „t" TEST FOR DIFFERENCE OF MEANS

To test the significant difference between two means $\overline{x_1}$ and $\overline{x_2}$ of samples of sizes n_1 and n_2 , use the statistic

$$t = \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where, $$s = \sqrt{\frac{\sum\left(x_1 - \overline{x}_1\right)^2 + \left(x_2 - \overline{x}_2\right)^2}{n_1 + n_2 - 2}}$$

or If S_1 and S_2 are given directly then

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

s_1, s_2 sample standard deviation.

Degrees of Freedom d.f. = $n_1 + n_2 - 2$

PROBLEMS:

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

	Type I	Type II
Sample Number	8	7
Sample Means	1234 hours	1036 hours
Sample S.D.	36 hours	40 hours

Is the difference in the means sufficient to warrant that type I is superior to type II regarding the length of life. [Dec 11]

Solution:

Step 1:

Given

$n_1 = 8$	$n_2 = 7$
$\overline{x}_1$ = 1234 hours	$\overline{x}_2$ = 1036 hours
s_1 = 36 hours	s_2 = 40 hours

Step 2:

Null Hypothesis H_0 : $\mu_1 = \mu_2$ The two types I and II of electric bulbs are identical.

Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

Step 3: Test statistics

Since two sample means $\overline{x}_1$ *and* $\overline{x}_2$ are given and also sample standard deviation s_1 and s_2 are given directly we use the statistic

$$t = \frac{|\overline{x}_1 - \overline{x}_2|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where $$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \qquad s = \sqrt{\frac{8(36)^2 + 7(40)^2}{8 + 7 - 2}} = \sqrt{1659.08}$$

s = 40.73

$$t = \frac{1234 - 1036}{40.73\sqrt{\frac{1}{8} + \frac{1}{7}}} = 9.39$$

Step 4:

Degrees of Freedom (d.f.) = $n_1 + n_2 - 2$ = 8 + 7 – 2 = 13 at 5% level.

Tabulated value of t for 13 d.f. at 5% level is 2.16

Since calculated t value > tabulated t value ,

We reject the null hypothesis H0.

Conclusion: The two types I and II of electric bulbs are not identical.

2. To verify whether a course in accounting improved performance, a similar test was given to 12 participants both before and after the course. The marks are

Before : 44 40 61 52 32 44 70 41 67 72 53 72

After : 53 38 69 57 46 39 73 48 73 74 60 78Was the course useful?

Solution:

Step 1: Given n_1 = 12 n_2 = 12

Step 2:

Null Hypothesis H0: There is no significant difference in the performance between before and afterthe course. i.e. $\mu_1 = \mu_2$

Alternatice Hypothesis H1 : $\mu_1 \neq \mu_2$

Calculation of sample means and S.D.‟s

$$\bar{x} = \frac{\sum x}{n} = \frac{648}{12} = 54 \qquad \bar{y} = \frac{\sum y}{n} = \frac{708}{12} = 59$$

x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
44	-10	100	53	-6	36
40	-14	196	38	-21	441
61	7	49	69	10	100
52	-2	4	57	-2	4
32	-22	484	46	-13	169
44	-10	100	39	-20	400
70	16	256	73	14	196
41	-13	169	48	-11	121
67	13	169	73	14	196
72	18	324	74	15	225
53	-1	1	60	1	1
72	18	324	78	19	361
		2176			**2250**

$$\sum (x-\bar{x})^2 = 2176 \qquad \sum (y-\bar{y})^2 = 2250$$

Therefore,

$$s = \sqrt{\frac{\sum\left(x_1 - \bar{x}_1\right)^2 + \left(x_2 - \bar{x}_2\right)^2}{n_1 + n_2 - 2}}$$

$$s = \sqrt{\frac{2176 + 2250}{12 + 12 - 2}} = \sqrt{\frac{4426}{22}} = \sqrt{201.18}$$

Therefore, s = 14.18

Now,

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{54 - 59}{14.18\sqrt{\frac{1}{12} + \frac{1}{12}}} = -0.8637$$

i.e. Calculated |t| = 0.8637

Step 4:

Degrees of freedom = $n_1 + n_2 - 2$ = 12 + 12 – 2 = 22

Therefore, tabulated t for 22 d.f. at 5% level is 2.07.

Since calculated t < tabulated t

we have to accept the null hypothesis.

Conclusion:

The course is useful.

3. Below are given the gain in weights (in lbs) of pigs fed of two diets A and B.

Diet A	**25**	**32**	**30**	**34**	**24**	**14**	**32**	**24**	**30**	**31**	**35**	**25**	**-**	**-**	**-**
Diet B	**44**	**34**	**22**	**10**	**47**	**31**	**40**	**30**	**32**	**35**	**18**	**21**	**35**	**29**	**22**

Test if the two diets differ significantly as regards their effect on increase in weight.

Solution:

Step 1: $n_1 = 12$ $n_2 = 15$

Step 2: Null Hypothesis H_0 = There is no significant difference between the mean increase in weight dure to diets A and B. i.e. $\mu_1 = \mu_2$

Alternative Hypothesis H1: $\mu_1 \neq \mu_2$

Step 3: $\bar{x} = \frac{\sum x}{n} = \frac{336}{12} = 28$ $\bar{y} = \frac{\sum y}{n} = \frac{450}{15} = 30$

Calculation for sample means and S.D.s

x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
25	-3	9	44	14	196
32	4	16	34	4	16
30	2	4	22	-8	64
34	6	36	10	-20	400
21	-4	16	47	17	289
14	-14	196	31	1	1
32	4	16	40	10	100
24	-4	16	30	0	0
30	2	4	32	2	4
31	3	9	35	5	25
35	7	49	18	-12	144
25	-3	9	21	-9	81
			35	5	25
			29	-1	1
			22	-8	64
336		**380**	**450**		**1410**

$\sum(x-\bar{x})^2 = 380$ $\sum(y-\bar{y})^2 = 1410$

Therefore,

$$s = \sqrt{\frac{\sum (x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

$$s = \sqrt{\frac{380 + 1410}{12 + 15 - 2}} = \sqrt{\frac{1790}{25}} = \sqrt{71.6}$$

Therefore, s = 8.46

Now,

$$t = \frac{|\bar{x} - \bar{y}|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{28 - 30}{8.46\sqrt{\frac{1}{12} + \frac{1}{15}}} = -0.609$$

i.e. Calculated |t| = 0.8637

Step 4:

Degrees of freedom = $n_1 + n_2 - 2 = 12 + 15 - 2 = 25$

Tabulated t for 25 d.f. at 5% level = 2.06.

Since calculated t < tabulated t

we accept the null hypothesis H_0.

Conclusion:

There is no significant difference between the mean increase in weight

4. The mean of two random sample sizes 9 and 7 are 196.42 and 198.82. The sum of the squares of the deviations from the mean are 26.94 and 18.73 respectively. can the sample be considered drawn from normal populations of equal means.

Solution:

Step 1:

Given

n_1= 9	**n_2=7**
$\bar{x}_1$ **=196.42**	$\bar{x}_2$ **= 198.82**
$\sum (x_1 - \bar{x}_1)^2 = 26.94$	$\sum (x_2 - \bar{x}_2)^2 = 18.72$

Step 2:

Null Hypothesis H_0: $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$

Step 3:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{where} \quad s = \sqrt{\frac{\sum (x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

$$t = \frac{196.42 - 198.82}{1.8059\sqrt{\left(\frac{1}{9} + \frac{1}{7}\right)}}$$

t = 2.6371

Step 4:

Degree of freedom = $n_1+n_2 - 2$= 9 + 7 – 2 = 14

At 5% L.O.S for t test the table value is 2.15

Therefore t calculated value > t tabulated value

Therefore H_0 is rejected

Conclusion:

The sample can not be considered as drawn from the same normal population

5. The height of six randomly chosen sailors are (in inches): 72, 71, 69, 68, 65, 63. Those of ten randomly chosen army men are 73, 72, 71, 70, 69, 69, 66, 65, 62, 61. Analyse the concept that the army men are shorter, on the average, than sailors.

Solution:

Step 1: $n_1 = 6$ $n_2 = 10$

Step 2:

Null Hypothesis H_0 = The concept that the army men are shorter, on the average, than sailors**.**

. i.e. $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$

Step 3:

$$\bar{x} = \frac{\sum x}{n} = \frac{408}{6} = 68 \qquad \bar{y} = \frac{\sum y}{n} = \frac{678}{10} = 67.8$$

Calculation for sample means and S.D.s

x	$(x-\bar{x})^2$	y	$(y-\bar{y})^2$
72	16	73	27.04
71	9	72	17.64
69	1	71	10.24
68	0	70	4.84
65	9	69	1.44
63	25	69	1.44
		66	3.24
		65	7.84
		62	33.64
		61	46.24
	$\sum(x-\bar{x})^2=60$		$\sum(y-\bar{y})^2=153.6$

Therefore,

$$s=\sqrt{\frac{\sum(x-\bar{x})^2+(y-\bar{y})^2}{n_1+n_2-2}}$$

$$s=\sqrt{\frac{60+153.6}{6+10-2}}=\sqrt{\frac{213.6}{14}}=\sqrt{15.26}$$

$$s=3.906$$

Therefore, $s=3.906$

Now,

$$t=\frac{|\bar{x}-\bar{y}|}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$t = \frac{68 - 67.8}{3.906\sqrt{\frac{1}{6} + \frac{1}{10}}} = \frac{0.2}{3.906\sqrt{0.2667}}$$

$$t = \frac{0.2}{3.906(.5164)} = 0.9915$$

i.e. Calculated $t = 0.9915$

Step 4:

Degrees of freedom = $n_1 + n_2 - 2 = 6+10-2 = 14$

Tabulated t for 14 d.f. at 5% level = 2.145.

Since calculated t < tabulated t

we accept the null hypothesis H_0.

Conclusion:

The concept that the army men are shorter, on the average, than sailors.

F – TEST FOR VARIANCE

To test whether if there is any significant difference between two estimates of population variance (or) To test if the two samples have come from the same population, we use F-Test.

In this case we set up null hypothesis $H_0 = \sigma_1 = \sigma_2$. i.e. Population variances are same.

Under H_0, the test statistic is

1. If S_1^2 is greater, then

$$F = \frac{s_1^2}{s_2^2}$$

2. If S_2^2 is greater, then

$$F = \frac{s_2^2}{s_1^2}$$

Where $s_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1}$, $s_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1}$

The degrees of freedom are $\gamma_1 = n_1 - 1$, $\gamma_2 = n_2 - 1 = (n_1 - 1, n_2 - 1)$

Note:We will take greater of the variances S_1^2 or S_2^2 in the numerator and adjust for the degrees of freedom accordingly .i.e. $F = \frac{Greater\ \ Variance}{Smaller\ \ Variance}$

If sample variance or sample SD are given directly then

$$F = \frac{n_1 s_1^2 / n_1 - 1}{n_2 s_2^2 / n_2 - 1}$$

1. In one sample of 10 observations from a normal population, the sum of the squares of the deviations of the sample values from the sample mean is 102.4 and in another sample of 12 observations from another normal population, the sum of the squares of the deviations of the sample values from the sample mean is 120.5. Examine whether the two normal populations have the same variances. [Nov 2016]

Solution:

Step 1:

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

$$\sum\left(x_1 - \bar{x}_1\right)^2 = 102.4 \qquad \sum\left(x_2 - \bar{x}_2\right)^2 = 120.5$$

Therefore,

$$s_1^2 = \frac{\sum\left(x_1 - \bar{x}_1\right)^2}{n_1 - 1}, \qquad s_2^2 = \frac{\sum\left(x_2 - \bar{x}_2\right)^2}{n_2 - 1}$$

$$s_1^2 = \frac{102.4}{9} \qquad s_2^2 = \frac{120.5}{11}$$

$$s_1^2 = 11.37 \qquad s_2^2 = 10.95$$

Step 2:

Null Hypothesis H_0: The two normal populations have the same variances.

$$\sigma_1^2 = \sigma_2^2$$

Alternative Hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$

Step 3: The test statistic

$$F = \frac{s_1^2}{s_2^2} \quad \because\ s_1^2 > s_2^2$$

$$F = \frac{11.37}{10.95} = 1.038$$

i.e. Calculated F = 1.038

Step 4:

Tabulated value of F for (9,11) d.f. at 5% level of significance is 2.91 (approx).

Since calculated value of F < tabulated F

we accept the null hypothesis H_0.

Conclusion:

The two normal populations have the same variances.

2. Two random samples gave the following results:

Sample	Size	Sample Mean	Sum of Squares of Deviation from the Mean
1	10	15	90
2	12	14	108

Test whether the samples come from the same normal population. [Nov 11]

Solution:

Step 1:

Given

$n_1 = 10$	$\bar{x}_1 = 15$	$\sum (x_1 - \bar{x}_1)^2 = 90$
$n_2 = 12$	$\bar{x}_2 = 14$	$\sum (x_2 - \bar{x}_2)^2 = 108$

Step 2:

Null Hypothesis H_0: The two samples have been drawn from the same normal population.

i.e. $H_0 = \mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$

Here we have to use two tests (i) To test equality of variances by F-test (ii) To test equality of means by t-test.

(i) F – test (Equality of Variances)

Step 3:

$$s_1^2 = \frac{\sum (x_1 - \bar{x}_1)^2}{n_1 - 1}, \qquad s_2^2 = \frac{\sum (x_2 - \bar{x}_2)^2}{n_2 - 1}$$

$$s_1^2 = \frac{90}{9} = 10 \qquad s_2^2 = \frac{108}{11} = 9.82$$

Now

$$F = \frac{s_1^2}{s_2^2} \qquad \because s_1^2 > s_2^2$$

$$F = \frac{10}{9.82} = 1.018$$

i.e. Calculated F = 1.018

Step 4:

Tabulated F at 5% level for (9,11) d.f. is 2.90. i.e. $F_{0.05}(9,11) = 2.90$

Since Calculated F < Tabulated F

we accept the null hypothesis H_0. i.e. μ1 = μ2

Conclusion:

The samples come from the same normal population.

(ii) **t-test** (To Test Equality of Means)

Step 2: Null Hypothesis H0: μ1=μ2

Given,

n_1= 10	$\bar{x}_1 = 15$	$\sum(x_1 - \bar{x}_1)^2 = 90$
n_2 = 12	$\bar{x}_2 = 14$	$\sum(x_2 - \bar{x}_2)^2 = 108$

$$s = \sqrt{\frac{\sum(x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

$$s = \sqrt{\frac{90+108}{10+12-2}} = \sqrt{\frac{198}{20}} = \sqrt{9.9}$$

Therefore, s = 3.15

Now,

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

$$t = \frac{15-14}{3.15\sqrt{\frac{1}{10} + \frac{1}{12}}} = 0.74$$

Therefore, Calculated value of t = 0.74

Step 4:

Tabulated value of t for 20 d.f. ($n_1 + n_2$ -2) at 5% level of significance is 2.086.

Since calculated value of t < tabulated value of t.

We accept the null hypothesis H0. i.e. $\mu_1 = \mu_2$.

Conclusion: Hence from (i) and (ii) , the given samples have been drawn from the same normal population.

3. It is known that the mean diameters of rivets produced by two firms A and B are practically the same but the standard deviations may differ. For 22 rivets produced by the firm A, the standard deviation is 2.9 mm, while for 16 rivets manufactured by firm B, the standard deviation is 3.8 mm. Compute the statistic you would use to test whether the products of firm A have the same variability as those of firm B and test it"s significance

Solution:

Given $n_1 = 22$ $n_2 = 16$

$s_1 = 2.9$ $s_2 = 3.8$

Here the S.D."s of the samples s1 and s2 are given. Therefore, the population variance S_1^2 and S_2^2 are obtained by using the relations,

Step 2:

Null Hypothesis H0: The products of both the firms A and B have the same variability.

i.e. $\sigma_1^2 = \sigma_2^2$

Alternative hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$

Step3: the test statistic

$$F = \frac{n_2 s_2^2 / n_2 - 1}{n_1 s_1^2 / n_1 - 1}$$

$$F = \frac{16 \times (3.8)^2 / 15}{22 \times (2.9)^2 / 21}$$

$$F = 1.748$$

i.e. Calculated F = 1.75

Step 4:

Tabulated value of F with (15,21) d.f. at 5% of significance is 2.20.

Since calculated F < tabulated F

we accept the null hypothesis H_0.

Conclusion: The products of both firms A and B have the same variability.

4. Two independent samples of sizes 9 and 7 from a normal population had the following values of the variables.

Sample I	**18**	**13**	**12**	**15**	**12**	**14**	**16**	**14**	**15**
Sample II	**16**	**19**	**13**	**16**	**18**	**13**	**15**		

Do the estimates of the population variance differ significantly at 5% level? [MAY 2017]

Solution:

Step 1: Given $n_1 = 9$ $n_2 = 7$

$$\bar{x} = \frac{\sum x}{n} = \frac{129}{9} = 14.33 \qquad \bar{y} = \frac{\sum y}{n} = \frac{110}{7} = 15.71$$

x	$(x-\bar{x})^2$	y	$(y-\bar{y})^2$
18	13.469	16	0.0841
13	1.769	19	10.824
12	5.429	13	7.344
15	0.449	16	0.0841
12	5.429	18	5.2441
14	0.109	13	7.3441
16	2.789	15	0.5041
14	0.109		
15	0.449		
	$\sum(x-\bar{x})^2 = 30.001$		$\sum(y-\bar{y})^2 = 31.429$

Therefore,

$$s_1^2 = \frac{\sum(x-\bar{x})^2}{n_1 - 1}, \qquad s_2^2 = \frac{\sum(x_2-\bar{x}_2)^2}{n_2 - 1}$$

$$s_1^{\,2} = \frac{30.001}{8} = 3.75 \qquad\qquad s_2^{\,2} = \frac{31.429}{6} = 5.238$$

$$s_1^{\,2} = 3.75 \qquad\qquad s_2^{\,2} = 5.238$$

Step 2:

Null Hypothesis H_0: The two normal populations have the same variances. $\sigma_1^2 = \sigma_2^2$

Alternative Hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$

Step 3: The test statistic

$$F = \frac{S_2^{\,2}}{S_1^{\,2}} \qquad \because S_2^{\,2} > S_1^{\,2}$$

$$F = \frac{5.238}{3.75} = 1.397$$

i.e. Calculated $F = 1.397$

Step 4:

Tabulated value of F for (6,8) d.f. at 5% level of significance is 3.44 (approx).

Since calculated value of F < tabulated F

we accept the null hypothesis H_0.

Conclusion:

The two normal populations have the same variances.

5. Test whether the two sets of data are drawn from same population

I set	**17**	**27**	**18**	**25**	**27**	**29**	**27**	**23**	**17**
II set	**16**	**16**	**20**	**16**	**20**	**17**	**15**	**21**	**-**

Solution:

T -test

Step 1: $n_1 = 9$ $n_2 = 8$

Step 2:

Null Hypothesis $H_0 = \mu_1 = \mu_2$

Alternative Hypothesis H1: $\mu_1 \neq \mu_2$

Step 3:

$$\bar{x} = \frac{\sum x}{n} = \frac{210}{9} = 23.3$$

$$\bar{y} = \frac{\sum y}{n} = \frac{141}{8} = 17.625$$

Calculation for sample means and S.D.s

x	$(x-\bar{x})^2$	y	$(y-\bar{y})^2$
17	39.69	16	2.64
27	13.69	16	2.64
18	28.09	20	5.64
25	2.89	16	2.64
27	13.6	20	5.64
29	32.49	17	0.39
27	13.67	15	6.891
23	0.09	21	11.3906
17	39.69	-	-
	$\sum(x-\bar{x})^2 = 183.97$		$\sum(y-\bar{y})^2 = 37.8716$

Therefore,

$$s = \sqrt{\frac{\sum(x-\bar{x})^2 + (y-\bar{y})^2}{n_1 + n_2 - 2}}$$

$$s = \sqrt{\frac{183.97 + 37.8716}{9+8-2}} = \sqrt{\frac{221.84}{15}} = \sqrt{14.79} \qquad s = 3.8457$$

Now,

$$t = \frac{|\bar{x} - \bar{y}|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{|23.3 - 17.625|}{3.8457\sqrt{\frac{1}{9} + \frac{1}{8}}} = \frac{5.675}{3.8457\sqrt{0.2361}}$$

$$t = \frac{5.675}{3.8457 \times 0.4859} = 3.0369$$

i.e. Calculated $t = 3.0369$

Step 4:

Degrees of freedom = $n_1 + n_2 - 2 = 15$

Tabulated t for 15 d.f. at 5% level = 2.131.

Since calculated t > tabulated t

we Rejectt the null hypothesis H_0.

ii) F –test

$$F = \frac{S_1^2}{S_2^2} \qquad \because S_1^2 > S_2^2$$

$$s_1^2 = \frac{\sum(x - \bar{x})^2}{n_1 - 1}, \qquad s_2^2 = \frac{\sum(y - \bar{y})^2}{n_2 - 1}$$

$$s_1^2 = \frac{183.97}{8} = 22.996 \qquad s_2^2 = \frac{17.625}{7} = 2.5179$$

Step 2:

Null Hypothesis H_0: $\sigma_1^2 = \sigma_2^2$

Alternative Hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$

Step 3: The test statistic

$$F = \frac{22.996}{2.5179} = 9.133$$

i.e. Calculated $F = 9.133$

Step 4:

Tabulated value of F for (8,7) d.f. at 5% level of significance is 3.73 (approx).

Since calculated value of F > tabulated F

we Reject the null hypothesis H_0.

Conclusion:

The two sets of data are not drawn from same population

Chi square test for goodness of fit and independence of attributes

1. The following table gives number of aircrafts accidents that occurred during various days. Use T^2 test and find whether the accidents are uniformly distributed over the week

Days :	Sun	Mon	Tues	Wed	Thurs	Fri	Sat
No of Accidents:	14	16	8	12	11	9	14

[May 15, 16]

Solution:

Step 1:

Given

n = 7

Step 2:

Null hypothesis H_0 : There are no significance difference between O_i and E_i

Step 3:

$$E = \frac{\sum O}{n} = \frac{84}{7} = 12$$

$$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
14	12	4	0.333
16	12	16	1.333
8	12	16	1.333
12	12	0	0
11	12	1	0.0835
9	12	9	0.75
14	12	4	0.333
		$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$ = 4.165	

Step 4:

DOF = n – 1 = 7 – 1 = 6

At 5% $\chi^2_{tab} = 12.592$

$\chi^2_{cal} < \chi^2_{tab}$

Hence H_0 is accepted.

Conclusion:

Hence the accidents are uniformly distributed over the week

2. The following table gives the number of aircraft accidents that occur during the various days of the week. Test whether the accidents are uniformly distributed over the week.

Days	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Total**
No. of accidents	**14**	**18**	**12**	**11**	**15**	**14**	**84**

Solution:

Step 1:

Given

n = 6

Step 2:

Null hypothesis H_0 : There are no significance difference between O_i and E_i

Step 3:

$$E = \frac{\sum o}{n} = 84/6=14$$

$$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$$

O	E	$O-E$	$\frac{(O_i - E_i)^2}{E_i}$
14	14	0	0
18	14	4	1.143
12	14	-2	0.286
11	14	-3	0.643
15	14	1	0.071
14	14	0	0

$$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right] = 2.143$$

Step 4:

DOF = n – 1 = 6 – 1 = 5

At 5% $T^2{}_{tab} = 11.07$

$T^2{}_{ca;} < T^2{}_{tab}$. H_0 is accepted.

Conclusion:

Hence the accidents are uniformly distributed over the week

3. The theory predicts the proportion of beams in the 4 groups A,B,C,D be 9 : 3 : 3 : 1. In an experiment with 1600 beams the numbers in 4 groups where 882, 313, 287, 118. Does the experimental result support the theory. [May 14]

Solution:

Step 1:

Given n = 4

Step 2:

Null hypothesis H_0 : There is no significance difference between O and E

Step 3:

If we divide 1600 in the ratio 9:3:3:1 We get expected frequency

$$E = \frac{total\ no\ of\ \ product}{sum\ of\ \ ratio} \times Individual\ ratio$$

$E_i(A) = \frac{1600}{16} X9 = 900$

$E_i(B) = \frac{1600}{16} X3 = 300$

$E_i(C) = \frac{1600}{16} X3 = 300$

$E_i(D) = \frac{1600}{16} X1 = 100$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
882	900	324	0.360
313	300	169	0.563
287	300	169	0.563
118	100	324	3.24
$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right] = 4.726$			

Step 4:

D.O.F = 4 – 1 = 3

χ^2_{tab} at 5% level of significance = 7.815

$\chi^2_{cal} < \chi^2_{tab}$

H_0 is accepted.

Conclusion:

Thus, the experimental result supports the theory.

4. The survey of 320 families with 5 childrens each gave the following distribution:

No of boys	**5**	**4**	**3**	**2**	**1**	**0**
No of girls	**0**	**1**	**2**	**3**	**4**	**5**
No of families	**14**	**56**	**110**	**88**	**40**	**12**

Use T^2 test whether the male and female birth are equally popular.

Solution:

Step 1:

Given N = 320 n = 5

Observed frequency	14	56	110	88	40	12

Step 2:

Null hypothesis H_0 : The female and male birth are equally popular

Step 3:

We use binomial distribution to find the expected frequency.

Expected frequency are given by $f(x) = N \times nc_x p_x q_{n-x}$

Where N = 320; n = 5

P = probability of getting a girl = $\frac{1}{2}$

q = $1 - \frac{1}{p} = 1 - \frac{1}{2} = \frac{1}{2}$

x = 0,1,2,3,4,5

$f(x) = N \times nc_x p_x q_{n-x}$

f(0) = $320 \times 5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5$ = 10

f(1) = $320 \times 5C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4$ = 50

f(2) = $320 \times 5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3$ = 100

f(3) = $320 \times 5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2$ = 100

f(4) = $320 \times 5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1$ = 50

$$f(5) = 320 \times 5C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = 10$$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
14	10	16	1.6
56	50	36	0.72
110	100	100	1
88	100	144	1.44
40	50	100	2
12	10	4	0.4
			$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right] = 7.16$

$$\chi^2_{cal} = \sum\left[\frac{(O-E)^2}{E}\right] = 7.16$$

Step 4:

D.O.F = 6 – 1 = 5

χ^2_{tab} at 5% = 11.07

$\chi^2_{cal} < \chi^2_{tab}$

H_0 is accepted.

Conclusion: Therefore, the male and female birth are equally popular.

5. Examine the goodness of fit for the Poisson distribution to the following data. [Nov 10 May 15, Nov 16]

x	0	1	2	3	4
f	**109**	**65**	**22**	**3**	**1**

Solution:

Step 1:

Given n = 4

Step 2:

Null hypothesis H_0 : there is no significance difference between O and E

Step 3:

To find expected frequency :

$$f(x) = N\frac{e^{-\lambda}\lambda^x}{x!}$$

Here , mean = variance = λ

To find λ:

x	f	fx
0	109	0
1	65	65
2	22	44
3	3	9
4	1	4
	Σf = 200	Σfx = 122

Mean = $\frac{\Sigma f}{\Sigma f}$ λ = 0.61

Here, N = Σf = 200

$$f(x) = N \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$f(0) = 200 \frac{e^{-0.61}(0.61)^{0}}{0!} = 108.6701$$

$$f(1) = 200 \frac{e^{-0.61}(0.61)^{1}}{1!} = 66.2888$$

$$f(2) = 200 \frac{e^{-0.61}(0.61)^{2}}{2!} = 20.2181$$

$$f(3) = 200 \frac{e^{-0.61}(0.61)^{3}}{3!} = 4.111$$

$$f(4) = 200 \frac{e^{-0.61}(0.61)^{4}}{4!} = 0.6269$$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
109	108.6701	0.1088	1.0015 x 10^{-3}
65	66.2888	1.661	0.0251
22	20.2181	3.1751	0.157
3	4.111	1.2343	0.3002
1	0.6269	0.1392	0.2221
		$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$	= 0.7054

$$\chi^2_{cal} = \sum\left[\frac{(O-E)^2}{E}\right] = 0.7054$$

Therefore, $\chi^2_{cal} = 0.7054$

Step 4:

D.O.F = n – 2 = 5 – 2 = 3

χ^2_{tab} at 5% = 7.815

$\chi^2_{cal} < \chi^2_{tab}$

Hence H_0 is accepted.

Conclusion:

There is no significance difference between O and E

6. Fit a binomial distribution for the following data and also test the goodness of fit.

X	**0**	**1**	**2**	**3**	**4**
F	**5**	**29**	**36**	**25**	**5**

Solution:

Step 1:

Given:

O	5	29	36	25	5

Step 2:

Null hypothesis H_0 : There is no significance difference between O and E

Step 3: To find expected frequency

f(x) = N . $nc_x p^x q^{n-x}$

Here, n = 4

To find p : In binomial distribution ;

X	f	fx
0	5	0
1	29	29
2	36	72
3	25	75
4	5	20
	100	196

$$\text{Mean} = \frac{\sum fx}{\sum f} = n\,p$$

$$\text{Mean} = \frac{196}{100} = np$$

$n\,p = 1.96 \quad \Rightarrow 4\,p = 1.96$

$p = 0.49$

$q = 1 - p = 0.15$

$f(x) = N\ .\ nc_x\, p^x q^{n-x}$

$f(0) = 100\ .\ 4c_0\,(0.49)^0(0.51)^4 = 6.7652$

$f(1) = 100\ .\ 4c_1\,(0.49)^1(0.51)^3 = 25.9995$

$f(2) = 100.\ 4c_2\,(0.49)^2(0.51)^2 = 37.470$

$f(3) = 100.\ 4c_3\,(0.49)^3(0.51)^1 = 24.0004$

$f(4) = 100\ .\ 4c_4\,(0.49)^4(0.51)^0 = 5.7648$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	6.7052	3.1159	0.4606
29	25.9995	9.0030	0.3463
36	37.47	2.1609	0.0577
25	24	1	0.0417
5	5.7648	0.5849	0.1014
			$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right] = 1.0077$

Step 4:

D.O.F = n - 1 = 5 – 1 = 4

χ^2_{tab} at 5% = 9.488

$\chi^2_{cal} < \chi^2_{tab}$

Hence, H_0 is accepted.

Conclusion: Therefore , there is no significant difference between O and E.

7. From the following information state whether the condition of the home is associated with the condition of the child .

Condition of Home

Condition Of Child		**Clean**	**Dirty**
	Clean	**69**	**51**
	Fairly clean	**81**	**20**
	Dirty	**35**	**44**

Solution:

Step 1: Null hypothesis H_0: The condition of the home is associated with the condition of the child .

Step 2: Given O: 69 51 81 20 35 44

Condition of Home

Condition Of Child		Clean	Dirty	Total
	Clean	69	51	120
	Fairly clean	81	20	101
	Dirty	35	44	79
	Total	185	115	N = 300

$$E(69) = \frac{120 \times 185}{300} = 74$$

$$E(51) = \frac{120 \times 115}{300} = 46$$

$$E(81) = \frac{101 \times 185}{300} = 62.283$$

$$E(20) = \frac{101 \times 115}{300} = 38.717$$

$$E(35) = \frac{79 \times 185}{300} = 48.717$$

$$E(44) = \frac{79 \times 115}{300} = 30.283$$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
69	74	25	0.3378
51	46	25	0.3435
81	62.283	350.3261	5.6247
20	38.717	350.3261	9.0484
35	48.717	188.1561	3.8622
44	30.283	188.1561	6.2133
			$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right] = 25.6299$

Step 3:

D.O.F = (m – 1) (n – 1) = (3 – 1) (2 – 1) = 2

χ^2_{tab} at 5% = 5.991

$\chi^2_{cal} > \chi^2_{tab}$

H_0 isrejected.

Conclusion:

The condition of the home is associated with the condition of the child .

8. In an experiment of immunization of cattle from TB , The following results were

	Affected	**Non-affected**
Inocalated	**12**	**28**
Non – Inocalated	**13**	**7**

Examine the effect of vaccine controlling the incidents of TB

Solution:

Step 1:

Null hypothesis H_0: The effect of vaccine controlling the incidents of TB

Step 2:

O	12	28	13	7

To find Expected frequency:

	Affected	Non-affected	Total
Inoculated	12	28	40
Non – Inoculated	13	7	20
Total	25	35	60

E(12) = $\frac{40 \times 25}{60}$ = 16.6667   E(13) = $\frac{25 \times 20}{60}$ = 8.3333

E(28) = $\frac{35 \times 40}{60}$ = 23.3333   E(7) = $\frac{35 \times 20}{60}$ = 11.6667

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
12	16.6667	21.7781	1.3067
28	23.3333	21.7781	0.9333
13	8.3333	21.7781	2.6134
7	11.6667	21.7781	1.8667
$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$ = 6.7201			

D.O.F = (m – 1) (n – 1) = 1

χ^2_{tab} at 5% $= 3.841$

$\chi^2_{cal} > \chi^2_{tab}$

H_0 is rejected.

Conclusion: The effect of vaccine not controlling the incidents of TB

9. The following data are for a sample of 300 car owners. Who where classified with respect to age and the number of accidents they had during the past 2 years . Test whether there is any relationship between these 2 variables. [May 2012]

Number of Accidents

		1	**1 or 2**	**More than 2**
Age	**Less than 30**	**8**	**23**	**14**
	Between 30 and 50	**21**	**42**	**12**
	More than 50	**71**	**90**	**19**

Solution:

Step 1:

Null hypothesis H_0 There is relationship between these 2 variables

O	8	23	14	21	42	12	71	90	19

Step 2:To find expected frequency :

	1	1 or 2	More than 2	
Less than 30	8	23	14	45
Between 30 and 50	21	42	12	75
More than 50	71	90	19	180
	100	155	45	N = 300

E(8) $= \frac{45 \text{ K } 100}{300} = 15$

E(23) $= \frac{45 \text{ K } 155}{300} = 23.25$

E(14) $= \frac{45 \text{ K } 45}{300} = 6.75$

E(21) $= \frac{75 \text{ K } 100}{300} = 25$

E(42) $= \frac{75 \text{ K } 155}{300} = 38.75$

$E(12) = \frac{75 \text{ K } 45}{300} = 11.25$

$E(71) = \frac{180 \text{ K } 100}{300} = 60$

$E(90) = \frac{180 \text{ K } 155}{300} = 93$

$E(19) = \frac{180 \text{ K } 45}{300} = 27$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
8	15	49	3.2667
23	23.25	0.0625	0.0027
14	6.75	52.5625	7.787
21	25	16	0.64
42	38.75	10.5625	0.2756
12	11.25	0.5625	0.05
71	60	121	2.0167
90	93	9	0.0968
19	27	64	2.3704
$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right] = 16.5059$			

Step 4:

D.O.F = (m – 1) (n – 1)

= (3 – 1) (3 – 1) = 4

χ^2_{tab} at 5% = 9.488

$\chi^2_{cal} > \chi^2_{tab}$

Therefore, H_0 is rejected.

Conclusion:

There is no relationship between these 2 variable

10. The data given below are taking during an epidemic of cholera. Test the effectiveness of inoculation in preventing the attack of cholera

	Attacked	**Non-Attacked**	**Total**
Inocalated	**31**	**469**	**500**
Non – Inocalated	**185**	**1315**	**1500**
Total	**216**	**1784**	**2000**

Solution:

Step 1:

O	12	28	13	7

Null hypothesis H_0 : The effect of vaccine controlling the incidents of TB

Step 2:

To find Expected frequency:

	Attacked	**Non-Attacked**	**Total**
Inocalated	**31**	**469**	**500**
Non – Inocalated	**185**	**1315**	**1500**
Total	**216**	**1784**	**2000**

E(31) = $\frac{500 \text{ K } 216}{2000}$ = 54

E(469) = $\frac{500 \text{ K } 1784}{2000}$ = 446

E(185) = $\frac{1500 \text{ K } 216}{2000}$ = 162

E(1315) = $\frac{1500 \text{ K } 1784}{2000}$ = 1338

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
31	54	529	9.7963
469	446	529	1.1861
185	162	529	3.2654
1315	1338	529	0.3954
$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right] =$ 14.643			

D.O.F = (m – 1) (n – 1) = 1

χ^2_{tab} at 5% = 3.841

$\chi^2_{cal} > \chi^2_{tab}$

H_0 is rejected.

Conclusion: **T**he effectiveness of inoculation is not preventing the attack of cholera

11. In a certain sample of 2000 families 1400 families are consumers of tea. Out of 1800 Hindu families 1236 families consumes tea. Use χ^2 test and state whether there is any significant difference between consumption of tea among Hindu and Non – Hindu families

Solution:

The given datas are in table form

i.e)

	Hindu	Non – Hindu	Total
Consuming Tea	1236	164	1400
Non – consuming tea	564	36	600
Total	1800	200	2000

Step 1:

Null hypothesis H_0: there is significant difference between consumption of tea among Hindu and Non – Hindu families

Step 2:

To find Expected frequency:

$$E(1236) = \frac{1400 \times 1800}{2000} = 1260$$

$$E(164) = \frac{1400 \times 200}{2000} = 140$$

$$E(564) = \frac{600 \times 1800}{2000} = 540$$

$$E(36) = \frac{600 \times 200}{2000} = 60$$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1236	1260	576	0.4571
164	140	576	4.1143
564	540	576	1.0666
36	60	576	9.6
$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$ = 15.238			

D.O.F = (m – 1) (n – 1) = 1

χ^2_{tab} at 5% = 3.841

$\chi^2_{cal} > \chi^2_{tab}$

H_0 isrejected.

Conclusion: there is no significant difference between consumption of tea among Hindu and Non – Hindu families

12. On the basis of information noted below, find out whether the new treatment is comparatively superior to the conventional one. [NOV 2017]

	Favourable	Non - Favourable	Total
Conventional	**40**	**70**	**110**
New	**60**	**30**	**90**
Total	**100**	**100**	**N = 200**

Solution:

Step 1:

Null hypothesis H_0: there is significant difference between consumption of tea among Hindu and Non – Hindu families

Step 2:

To find Expected frequency:

$$E(40) = \frac{110 \times 100}{200} = 55$$

$$E(70) = \frac{110 \times 100}{200} = 55$$

$$E(60) = \frac{90 \times 100}{200} = 45$$

$$E(30) = \frac{90 \times 100}{200} = 45$$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
40	55	225	4.090
70	55	225	4.090
60	45	225	5
30	45	225	5
$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$ = 18.18			

D.O.F = (m – 1) (n – 1) = 1

χ^2_{tab} at 5% = 3.841