

UNIT-IV

TESTING OF HYPOTHESIS

- Test of significance: Null Hypothesis; Alternative Hypothesis-Type Error –Type II Error
- Large sample test for single proportion
- Difference of proportions
- Single mean
- Difference of means
- Test for single mean; Difference of means for small samples
- Test for ratio of variances for small samples

The logo for NRCM (National Research Center for Materials) features a stylized tree with a purple trunk and branches, and leaves in shades of yellow, orange, and red. Below the tree, the letters "NRCM" are written in a large, bold, purple font.

Population:

It is the aggregate or totality of statistical data forming a subject of investigation.

Sample:

A portion of the population which is examined with a view of determining the population characteristics.

There by studying characteristics of small group and extending them to the big group is called Sampling.

Sampling Distributions:

The collection of all samples of a given size defines a distribution called as Sampling Distribution.

Notation:

N

n

z

p

S

P

Types of Sampling:**1. Random Sampling:**

In this method of sampling, all members of the population have equal chance to be selected as a member of the sample. If N is the size of population, then every individual in that group have an equal probability of $\frac{1}{N}$ to be chosen as the member of the sample.

2. Stratified Sampling:

In this method the entire population is divided into small groups which are called Strata. These Strata are homogeneous in nature. The sample can be drawn by taking the members from each of these Strata.

3. Purposive Sampling:

In this type of sampling, the sample will be chosen according to the user's choice. The personal judgement plays a vital role in selecting the sample. No scientific method or mathematical calculations are included. This is purely biased sampling.

4. Sequential Sampling:

In this method, first sample is selected to estimate the characteristics of the population. If it is upto the mark, the process terminates, else the second sample will be chosen either with replacement or without replacement from the population. Hence the samples are selected in a sequential manner based on the results of the past.

5. Systematic Sampling:

In this method, the samples are chosen at regular intervals of time in a systematic manner. In this method the choice of the sample is based on the interval of time.

1. The number of samples that can be drawn from the population of size N with

sample size n are with replacement $= N^n$

2. With out replacement $= N_{c_n}$

3. Correction factor $= \frac{N-n}{N-1}$

4. Sample Mean ($\bar{x}$) $= \frac{1}{n} \sum_{i=1}^n x_i$

5. Variance (S^2) $= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \cong \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

6. standard Error $= \frac{\sigma}{\sqrt{n}}$

Problems:

1. Evaluate the correction factor and also, the number of samples that can be drawn from a population with replacement and without replacement, where $n = 100$ and $n = 5$

Sol: Correction factor $= \frac{N-n}{N-1}$

$$= \frac{100-5}{100-1} = \frac{95}{99} = 0.9596$$

No. of samples

with replacement $= N^n = 100^5 = 10000000000 = 10^{10}$

with out replacement $= N_{c_n} = 100_{c_5} = 75287520$

2. A population contains members 2,3,6,8,11. Evaluate the no. of samples of size 2 that can be drawn from this population, with replacement and hence evaluate the following (i) Mean of the population (ii) Standard deviation of the population (iii) Mean of the sampling distribution of means (iv) Standard deviation of the sampling distribution of means.

Sol: Given that $N = 5$ and $n = 2$

No. of samples with replacements $= 5^2 = 25$

$$(i) \text{ Mean } = \mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

$$(ii) \sigma^2 = \frac{1}{5} \sum_{i=1}^5 (x_i - \mu)^2$$

$$= \frac{1}{5} [(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2]$$

$$= \frac{1}{5} [(-4)^2 + (-3)^2 + 0 + (2)^2 + (5)^2]$$

$$= \frac{1}{5} [16 + 9 + 4 + 25] = \frac{54}{5}$$

$$\sigma^2 = 10.8$$

$$\sigma = \sqrt{10.8} = 3.2863$$

(iii) Sampling distribution is {(2,2), (2,3), (2,6), (2,8), (2,11), (3,2), (3,3), (3,6), (3,8), (3,11), (6,2), (6,3), (6,6), (6,8), (6,11), (8,2), (8,3), (8,6), (8,8), (8,11), (11,2), (11,3), (11,6), (11,8), (11,11)}

Sampling Distribution of mean is

$\{2, 2.5, 4.5, 6.5, 2.5, 3, 4.5, 5.5, 7, 4, 4.5, 6, 7, 8.5, 5, 5.5, 7, 8, 9.5, 6.5, 7, 8.5, 9.5, 11\}$

Mean of the sampling distribution of means

$$= \frac{1}{25} [2 + 2.5 + 4.5 + \dots + 11] = \frac{150}{25} = 6$$

(iv) variance of the sampling distribution means

$$= \frac{1}{25} [(2 - 6)^2 + (2.5 - 6)^2 + (4 - 6)^2 + \dots + (11 - 6)^2]$$

$$S^2 = 5.4$$

Standard deviation of the sampling distribution means $= \sqrt{5.4} = 2.32$

3. Consider a population consisting of the members 5, 10, 14, 18, 13, 24. Consider all possible samples of size 2 with out replacement and then evaluate the following (i)

Mean of the population (ii) Standard deviation of the population

(iii) Mean of the sampling distribution of means (iv) Standard deviation of the sampling distribution of means.

Sol: (i) Mean of the population $\mu = \frac{5 + 10 + 14 + 18 + 13 + 24}{6} = 14$

(ii) $\sigma^2 = \frac{1}{6} \sum_{i=1}^6 (x_i - \mu)^2$

$$= \frac{1}{6} [(5 - 14)^2 + (10 - 14)^2 + (14 - 14)^2 + (18 - 14)^2 + (13 - 14)^2 + (24 - 14)^2]$$

$$= \frac{1}{6} [(-9)^2 + (-4)^2 + 0 + (4)^2 + (1)^2 + (10)^2]$$

$$= \frac{1}{6} [81 + 16 + 16 + 1 + 100]$$

$$\sigma^2 = 35.6$$

$$\sigma = \sqrt{35.6} = 5.96$$

(iii) Sample distribution of samples of size 2

$= \{(5, 10), (5, 14), (5, 18), (5, 13), (5, 24), (10, 14), (10, 18), (10, 13), (10, 24), (14, 18), (14, 13), (14, 24), (18, 13), (18, 24), (13, 24)\}$

Sampling Distribution of mean is

$\{7.5, 9.5, 11.5, 9, 14.5, 12, 14, 11.5, 17, 16, 13.5, 19, 15.5, 21, 18.5\}$

Mean of the sampling distribution of means

$$= \frac{7.5 + 9.5 + 11.5 + \dots + 18.5}{15} = 14$$

(iv) variance of the sampling distribution means

$$= \frac{1}{15} [(7.5 - 14)^2 + (9.5 - 14)^2 + (11.5 - 14)^2 + \dots + (18.5 - 14)^2]$$

$$S^2 = 14.26$$

$$S = 3.77$$

4. Consider a population of size 4 with the members 1, 5, 6, 8 from which samples of size 2 are drawn. Evaluate the standard error.

Sol: $\mu = \frac{1 + 5 + 6 + 8}{4} = \frac{20}{4} = 5$

$$\begin{aligned}\sigma^2 &= \frac{1}{N} \sum_{i=1}^4 (x_i - \mu)^2 \\ &= \frac{1}{4} [(1-5)^2 + (5-5)^2 + (6-5)^2 + (8-5)^2] \\ &= \frac{1}{4} [16 + 1 + 9] = \frac{26}{4} = 6.5\end{aligned}$$

$$\therefore \text{Standard error} = \frac{\sigma}{\sqrt{n}} = \frac{2.54}{\sqrt{2}} = 1.79$$

5. What is the effect on the standard errors if a sample is drawn from a population where its size is increased from 400 to 900

$$\text{Sol: } S.E \propto \frac{1}{\sqrt{n}}$$

$$\begin{aligned}\frac{S.E_1}{S.E_2} &= \sqrt{\frac{n_2}{n_1}} \\ &= \sqrt{\frac{900}{400}} = \sqrt{\frac{9}{4}} = 1.5\end{aligned}$$

$$S.E_1 = (1.5) S.E_2$$

Testing Hypothesis:

Hypothesis: Any statement that can be made about the population parameter is called a hypothesis.

It is of two types:

(1) Null hypothesis:

It is denoted by H_0 and specifies the statement that there is no difference between the population and the sample parameters.

Ex: $H_0: \mu = 64$

(2) Alternative hypothesis:

It is denoted by H_1 and is defined in contradiction to the Null hypothesis.

Ex: $H_1: \mu \neq 64$ (2 tailed test)

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

$H_1: \mu < 64$ (Left tailed test)

The samples can be classified into two types

1. Small Samples ($n < 30$)

2. Large Samples ($n > 30$)

Testing Procedure:

1. Null Hypothesis

2. Alternative Hypothesis

3. Level of Significance:

It is denoted by α and can be given a value in contradiction to the confidence level. There are 3 different values for α . They are 1%, 5%, 10% which specify 99%, 95%, 90% confidence level.

4. Test Statistic:

It is denoted by z and is given with a small formula that varies from test to test

5. Conclusion:

In this step, a comparison between the calculated z and a table value z_α

If $|z| < z_\alpha$ then the conclusion H_0 is accepted (or)

$|z| > z_\alpha$ then the conclusion H_0 is rejected and H_1 is accepted

Type – I Error: Reject H_0 when it is true.

Type – II Error: Accept H_0 when it is false.

Acceptance region:

The region that provides acceptance to the Null hypothesis is called acceptance region

Critical region (or) Rejection region:

The region that gives rejection to H_0 is called the critical region.

z_α values

	Level of significance		
	1%	5%	10%
Two-Tailed Test	2.58	1.96	1.645
Right-Tailed Test	2.33	1.645	1.28
Left-Tailed Test	-2.33	-1.645	-1.28

Test of significance os single mean:

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

If σ is not known, $z = \frac{\bar{x} - \mu}{s / \sqrt{n}}$

Problems:

1. According to the norms established for a mechanical aptitude test for the persons with 18 years of age proved that, they have an average height of 73.2 with a S. D of 8.6. If 4 randomly selected persons were tested, there height was averaged as 76.7. Test the hypothesis $\mu = 73.2$ against the alternative hypothesis, $\mu > 73.2$ at 1% level of significance.

Sol: $\mu = 73.2$, $\sigma = 8.6$, $\bar{x} = 76.7$; $\alpha = 1\%$ and $n = 4$

Testing procedure:

Step1: Null hypothesis:

$$H_0: \mu = 73.2$$

Step2: Alternative hypothesis:

$$H_1: \mu > 73.2$$

Step3: Level of significance:

$$\alpha = 1\%$$

Step4: Test statistic:

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

$$= \frac{76.7 - 73.2}{8.6 / \sqrt{4}} = 0.81$$

Step5: Conclusion:

at 1% level of significance (L.O.S) for a right failed test,

$$z_{\alpha} = 2.33$$

$$|z| = 0.81$$

$$|z| < z_{\alpha} \Rightarrow H_0 \text{ is accepted}$$

$$\mu = 73.2$$

The true confidence limits for the population mean for the population proportion

$$\text{are } z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

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

$$1) 95\% \text{ confidence limits are } \bar{x} \pm (1.96) \frac{\sigma}{\sqrt{n}}$$

$$2) 98\% \text{ confidence limits are } \bar{x} \pm (2.33) \frac{\sigma}{\sqrt{n}}$$

$$3) 99\% \text{ confidence limits are } \bar{x} \pm (2.58) \frac{\sigma}{\sqrt{n}}$$

2. Consider the random samples of size 400 which is drawn from a population of standard deviation 10. It is observed that the sample mean is 40. Test the claim, whether this sample can be regarded as drawn from a population with a mean of 38. Also calculate the 95% confidence limits for the population mean.

Sol: $n = 400$; $\sigma = 10$; $\bar{x} = 40$ and $\mu = 38$

Testing procedure:

Step1: Null hypothesis:

$$H_0: \mu = 38$$

Step2: Alternative hypothesis:

$$H_1: \mu \neq 38 \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 5\%$$

Step4: Test statistic:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{40 - 38}{10 / \sqrt{400}} = 2 \times 2 = 4$$

Step5: Conclusion:

at 5% L.O.S for a right failed test,

$$z_{\alpha} = 1.96$$

$$|z| = 4$$

$$|z| > z_{\alpha} \Rightarrow H_0 \text{ is rejected}$$

$$\Rightarrow H_1 \text{ is accepted}$$

$$\Rightarrow \mu \neq 38$$

$\therefore$ The sample is not drawn from a population with mean 38

95% confidence limits for population mean $\bar{x} \pm (1.96) \frac{\sigma}{\sqrt{n}}$

$$\text{Lower limit is } \bar{x} - (1.96) \frac{\sigma}{\sqrt{n}}$$

$$= 40 - (1.96) \left(\frac{10}{\sqrt{400}} \right)$$

$$= 39.02$$

$$\text{Upper limit is } \bar{x} + (1.96) \frac{\sigma}{\sqrt{n}}$$

$$= 40 + (1.96) \left(\frac{10}{\sqrt{400}} \right)$$

$$= 40.98$$

The interval is (39.02, 40.98)

3. An ambulance service claims that on an average it takes less than 10 min to attend an emergency call. A sample of 36 calls has a mean of 11 mins with a variance of 16 mins. Test at 5% level of significance whether the claim is true.

Sol: $n = 36$; $s^2 = 16$ and $s = 4$; $\bar{x} = 11$ and $\mu = 10$

Testing procedure:

Step1: Null hypothesis:

$$H_0: \mu < 10$$

Step2: Alternative hypothesis:

$$H_1: \mu > 10 \text{ [right tailed test]}$$

Step3: Level of significance:

$$\alpha = 5\%$$

Step4: Test statistic:

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

$$= \frac{11 - 10}{4 / \sqrt{36}} = \frac{1}{4/6} = \frac{6}{4} = 1.52$$

Step5: Conclusion:

at 5% L.O.S for a right tailed test,

$$z_\alpha = 1.645$$

$$|z| = 1.5$$

$$|z| < z_\alpha \Rightarrow H_0 \text{ is accepted}$$

$\therefore$ The average time taken by ambulance to attend emergency call less than 10 min.

4. The mean life time of 100 electric bulbs produced by a company is 1560 hours with a S.D of 90 hours. Can this be regarded as drawn from the population of electric bulbs with mean life time of 1580 hours.

Sol: Given that $\mu = 1580$; $n = 100$; $\bar{x} = 1560$ & $S = 90$

Testing Procedure:

Step1: Null hypothesis:

$$H_0: \mu = 1580$$

Step2: Alternative hypothesis:

$$H_1: \mu \neq 1580 \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 5\%$$

Step4: Test statistic:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{1560 - 1580}{90 / \sqrt{100}} = -20 \times \frac{1}{9} = -2.2$$

Step5: Conclusion:

at 5% L.O.S for a two tailed test,

$$z_\alpha = 1.96$$

$$|z| = 2.2$$

$$|z| > z_\alpha \Rightarrow H_0 \text{ is rejected}$$

$$\Rightarrow H_1 \text{ is accepted}$$

$$\Rightarrow \mu \neq 1580$$

$\therefore$ The sample is not drawn from 1580

Test of significance for difference of means:

Consider two samples with the given parameter. The test statistic is given by

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

If σ is not known, i.e., replace σ_1 with s_1 and σ_2 with s_2 then $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

If both the samples are drawn from the same population, then $\sigma_1^2 = \sigma_2^2$ i.e., σ^2

$$\text{then } z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

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

1. The mean of two large samples of sizes 1000 and 2000 have the mean values 67.5 and 68 respectively. Can these samples be regarded as drawn from the same population with standard deviation 2.5

Sol: Given that $n_1 = 1000$; $n_2 = 2000$

$$\bar{x}_1 = 67.5; \bar{x}_2 = 68 \text{ and } \sigma = 2.5$$

Step1: Null hypothesis:

$$H_0: \mu_1 = \mu_2$$

Step2: Alternative hypothesis:

$$H_1: \mu_1 \neq \mu_2 \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 5\%$$

Step4: Test statistic:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{67.5 - 68}{2.5 \left[\sqrt{\frac{1}{1000} + \frac{1}{2000}} \right]} = -5.164$$

Step5: Conclusion:

at 5% L.O.S for a two tailed test,

$$z_\alpha = 1.96$$

$$|z| = 5.164$$

$$|z| > z_\alpha \Rightarrow H_0 \text{ is rejected}$$

$$\Rightarrow H_1 \text{ is accepted}$$

$\therefore$ The sample are not drawn from the same population.

2. In a survey of buying 400 women shoppers are chosen at random at super market located at a certain place in a city. Their average weekly food expenditure is 250 rupees with a standard deviation of 40 rupees. Another group of 600 women shopper were chosen at random in a super bazar located at other part of city. Their average weekly food expenditure is 220 rupees with a standard deviation of 55 rupees. Test at 1% level of significance, whether the two populations have the same average weekly food expenditure.

Sol: Given that $\bar{x}_1 = 250$; $n_1 = 400$ and $s_1 = 40$

$\bar{x}_2 = 220$; $n_2 = 600$ and $s_2 = 55$

Testing Procedure:

Step1: Null hypothesis:

$$H_0: \mu_1 = \mu_2$$

Step2: Alternative hypothesis:

$$H_1: \mu_1 \neq \mu_2 \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 1\%$$

Step4: Test statistic:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 \frac{1}{n_1} + s_2^2 \frac{1}{n_2}}} = \frac{250 - 220}{\sqrt{\frac{(40)^2}{400} + \frac{(55)^2}{600}}} = 9.97$$

Step5: Conclusion:

at 1% L.O.S for a two tailed test,

$$z_\alpha = 2.58$$

$$|z| = 9.97$$

$$|z| > z_\alpha \Rightarrow H_0 \text{ is rejected}$$

$\Rightarrow H_1$ is accepted

$\therefore$ The average weekly food expenditure of the two populations are not identical.

3. Two samples are drawn with the sizes 250 and 300 with means 63.2 and 75.4 respectively. The population of standard deviation is 23.7. Verify whether the samples are drawn from the same population at 10 % level of significance.

Sol: Given that $n_1 = 250$; $n_2 = 300$

$\bar{x}_1 = 63.2$; $\bar{x}_2 = 75.4$ and $\sigma = 23.7$

Testing Procedure:

Step1: Null hypothesis:

$$H_0: \mu_1 = \mu_2$$

Step2: Alternative hypothesis:

$$H_1: \mu_1 \neq \mu_2 \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 10\%$$

Step4: Test statistic:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{63.2 - 75.4}{23.7 \left[\sqrt{\frac{1}{250} + \frac{1}{300}} \right]} = -6.01$$

Step5: Conclusion:

at 10% L.O.S for a two tailed test,

$$z_\alpha = 1.6458$$

$$|z| = 6.01$$

$$|z| > z_\alpha \Rightarrow H_0 \text{ is rejected}$$

$\Rightarrow H_1$ is accepted

Test of significance for single proportion (Large Samples)

A single sample from a population drawn where p denotes the sample proportion

P denotes the the population proportion and $p + q = 1$

$$P + Q = 1$$

The test statistic is given by $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

The confidence limits for the population proportion are

$$1) \text{ at } 95\%, p \pm (1.96) \sqrt{\frac{PQ}{n}}$$

$$2) \text{ at } 98\%, p \pm (2.33) \sqrt{\frac{PQ}{n}}$$

$$3) \text{ at } 99\%, p \pm (2.58)\sqrt{\frac{PQ}{n}}$$

1. A manufacturer claims that 95% of his products are upto the specifications. A sample of two hundred items were picked up in which 18 were faulty. Test the claim at 5% level of significance.

$$\text{Sol: } P = 95\% = 0.95$$

$$Q = 1 - 0.95 = 0.05$$

$$n = 200$$

$$\text{Bad items} = 18$$

$$\text{Good ones} = 200 - 18 = 182$$

$$p = \frac{182}{200} = 0.91$$

$$q = 1 - 0.91 = 0.09$$

Testing Procedure:

Step1: Null hypothesis:

$$H_0: P = 0.95$$

Step2: Alternative hypothesis:

$$H_1: P \neq 0.95 \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 5\%$$

Step4: Test statistic:

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

$$= -2.59$$

Step5: Conclusion:

at 5% L.O.S for a two tailed test,

$$z_\alpha = 1.96$$

$$|z| = 2.59$$

$$|z| > z_\alpha \Rightarrow H_0 \text{ is rejected}$$

$$\Rightarrow H_1 \text{ is accepted}$$

$\therefore$ Manufacturer claim is not true.

2. In a sample of 1000 people from Karnataka state, 540 members are rice eaters. The remaining are considered as wheat eaters. Test the claim at 1% LOS that the rice eaters and wheat eaters are equally popular.

$$\text{Sol: } P = \frac{1}{2}; Q = \frac{1}{2}$$

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

$$q = 1 - 0.54 = 0.46$$

Testing Procedure:

Step1: Null hypothesis:

$$H_0: P = \frac{1}{2}$$

Step2: Alternative hypothesis:

$$H_1: P \neq \frac{1}{2} \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 1\%$$

Step4: Test statistic:

$$z = \frac{\frac{p - P}{\sqrt{\frac{PQ}{n}}}}{\frac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1000}}}} = 2.529$$

Step5: Conclusion:

at 5% L.O.S for a two tailed test,

$$z_\alpha = 2.58$$

$$|z| = 2.52$$

$$|z| < z_\alpha \Rightarrow H_0 \text{ is accepted}$$

$\therefore$ The rice eaters and wheat eaters are equally popular.

3. A random sample of 500 pineapple were stated from a large confinement and 65 were found to be bad. Find the two limits proportion of bad pineapples in confinement at 95% confidence.

Sol: $n = 500$

$$p = \frac{65}{500} = 0.13$$

$$q = 1 - 0.13 = 0.87$$

95% confidence limits are $p \pm 1.96\sqrt{\frac{pq}{n}}$

$$\text{Lower limit is } p - 1.96\sqrt{\frac{pq}{n}} = 0.13 - (1.96)\sqrt{\frac{(0.13)(0.87)}{500}} = 0.1005$$

$$\text{Upper limit is } p + 1.96\sqrt{\frac{pq}{n}} = 0.13 + (1.96)\sqrt{\frac{(0.13)(0.87)}{500}} = 0.1594$$

4. In a sample of 500 people from a village in Rajasthan, 280 are found to be wheat eaters and the rest are rice eaters. Can we assume that both rice and wheat eaters are equally popular in Rajasthan? Also calculate through the confidence limits for population proportion of wheat eaters.

Sol: $n = 500$

$$p = \frac{280}{500} = 0.56$$

$$q = 1 - 0.56 = 0.44$$

Testing Procedure:

Step1: Null hypothesis:

$$H_0: P = 0.5$$

Step2: Alternative hypothesis:

$$H_1: P \neq 0.5 \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 5\%$$

Step4: Test statistic:

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.56 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{500}}}$$

$$= 2.68$$

Step5: Conclusion:

at 5% L.O.S for a two tailed test,

$$z_{\alpha} = 1.96$$

$$|z| = 2.68$$

$|z| > z_{\alpha} \Rightarrow H_0$ is rejected

$\Rightarrow H_1$ is accepted

Confidence limits

$$\text{Lower limit is } p - 1.96\sqrt{\frac{PQ}{n}} = 0.56 - (1.96)\sqrt{\frac{(0.5)(0.5)}{500}} = 0.5162$$

$$\text{Upper limit is } p + 1.96\sqrt{\frac{PQ}{n}} = 0.56 + (1.96)\sqrt{\frac{(0.5)(0.5)}{500}} = 0.6038$$

Test of significance for difference of proportions:

Two samples are drawn from 2 populations, Test statistic is given by

$$z = \frac{p_1 - p_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}$$

If the population proportions P_1, Q_1 and P_2, Q_2 are not known, then replace them with sample values

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}$$

Method of pooling:

In this method, let $p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}$ then the Test statistic is given by

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

1. A random sample of 400 men and 600 women were asked, whether they would like to have flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that, the proportion of men and women in favour of the proposal are identical at 5% level of significance.

Sol: $n_1 = 400$; $n_2 = 600$

$$\begin{aligned}
 p_1 &= \frac{200}{400} = 0.5 ; p_2 = \frac{325}{600} = 0.5417 \\
 q_1 &= 1 - p_1 = 1 - 0.5 = 0.5 \\
 q_2 &= 1 - p_2 = 1 - 0.5417 = 0.4583 \\
 p &= \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400(0.5) + 600(0.5417)}{400 + 600} = 0.525 \\
 q &= 1 - p = 1 - 0.525 = 0.475 \\
 \text{Testing Hypothesis:} \\
 \text{Step1: Null hypothesis:} \\
 H_0: p_1 &= p_2 \\
 \text{Step2: Alternative hypothesis:} \\
 H_1: p_1 &\neq p_2 \text{ [two tailed test]} \\
 \text{Step3: Level of significance:} \\
 \alpha &= 5\% \\
 \text{Step4: Test statistic:} \\
 z &= \frac{p_1 - p_2}{\sqrt{(pq) \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}} \\
 &= \frac{0.5 - 0.5417}{\sqrt{(0.525)(0.475) \left[\frac{1}{400} + \frac{1}{600} \right]}} = -1.2936
 \end{aligned}$$

Step5: Conclusion:

at 5% L.O.S for a two tailed test,

$$z_\alpha = 1.96$$

$$|z| = 1.2936$$

$$|z| < z_\alpha \Rightarrow H_0 \text{ is accepted}$$

$\therefore$ The proportion of men and women in favour of the proporsal are identical.

2. On the basis of total score 200 candidates of civil service exam are divided into two groups. The upper 30% and the remaining 70% consider the 1st question of the examination among the 1st group. 40 had the correct answer. In the 2nd group 80 had the correct answer. On the basis of these results, can we conclude that the 1st question, answering specifies that, the abilities of the 2nd people into two groups are identical.

Sol: $n_1 = 60$; $n_2 = 140$

$$x_1 = 40; x_2 = 80$$

$$p_1 = \frac{x_1}{n_1} = \frac{40}{60} = 0.6667 ; p_2 = \frac{x_2}{n_2} = \frac{80}{140} = 0.5714$$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{60(0.6667) + 140(0.5714)}{60 + 140} = 0.6$$

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

Testing Hypothesis:

Step1: Null hypothesis:

$$H_0: p_1 = p_2$$

Step2: Alternative hypothesis:

$$H_1: p_1 \neq p_2 \text{ [two tailed test]}$$

Step3: Level of significance:

$$\alpha = 5\%$$

Step4: Test statistic:

$$z = \frac{p_1 - p_2}{\sqrt{(pq) \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$
$$= \frac{0.6667 - 0.5714}{\sqrt{(0.6)(0.4) \left[\frac{1}{60} + \frac{1}{140} \right]}} = 1.2607$$

Step5: Conclusion:

at 5% L. O. S for a two tailed test,

$$z_\alpha = 1.96$$

$$|z| = 1.2607$$

$$|z| < z_\alpha \Rightarrow H_0 \text{ is accepted.}$$

