Unit-3

Testing of Hypothesis

Population

A collection of individuals is Called a Population (or) Universe

Sample

A finite subset of a Population is Called a sample and the Process of Selecting each Sample is Called Sampling.

Sample Size

The number of Individuals in a Sample is Called Sample Size

Parameters and Statistics

Statistical measures like mean μ, Variance σ^2 Calculated on the basis of ~~Pop~~ Population Values of a are Called Parameters.

Corresponding measures mean $\bar{x}$, Variance s^2 computed on the sample observation is called Statistics.

A sample statistic is denoted by t.

Sampling distribution

The Probability distribution of a statistic 't' is called the sampling distribution of t.

Null Hypothesis (H_0)

A hypothesis which assumes that there is no significant difference between the sample statistic and the corresponding parameters, such a hypothesis of no difference is called null hypothesis and is denoted by H_0.

A hypothesis which is different from null hypothesis is called alternative hypothesis (H_1)

A Procedure to decide whether to accept (or) reject null hypothesis (alternative hypothesis) is called a test of hypothesis

Standard Error

The standard deviation of the Sampling distribution of a Statistic is Called Standard error of Statistic.

Critical region and Acceptance region

When the sample Statistic lies in a Certain region or interval is Called Critical region (or) region of rejection

The region Complementary to Crical region is called region of acceptance.

(or)

Critical region

A region in the sample

Space S which is the rejection of H_0 is known as Critical region (or) region of rejection. The region Complementary to the Critical region is called acceptance region.

Type 1 & Type 2 error

Reject H_0 when it is true. → Type 1

Accept H_0 when it is false → Type 2

Critical Value [Significant Value]

The Value of Statistic Z for which the Critical region and acceptance region are Separated is called Critical Value and is denoted by Z_α where α is the level of Significance.

One Tail & two tail test

Set $H_0 : \mu = \mu_0$

Then H_1 is any one of the following

$H_1 : \mu \neq \mu_0$ (ie $\mu > \mu_0$, $\mu < \mu_0$ Two tail)

$H_1 : \mu > \mu_0$ (right tail)

$H_1 : \mu < \mu_0$ (Left tail)

Nature	0.01 1%	0.02 2%	0.05 5%	0.10 10%
2 Tail	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 2.33$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Right tail	$Z_\alpha = 2.33$	$Z_\alpha = 2.005$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left tail	$Z_\alpha = -2.33$	$Z_\alpha = -2.005$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

Procedure for testing of hypothesis

1) set H_0
2) set H_1 and check whether it is 2 tail (or) 1 tail
3) Z_α is noted
4) find test statistic value $|Z|$ [use formula]
5) If $|Z| < |Z_\alpha|$ accept H_0.
   $|Z| > |Z_\alpha|$ reject H_0.

Large Sample

when size of a sample is greater than 30 then it is called Large sample otherwise it is Small sample.

Large Sample Test

Test 1

Test of Significance between ~~sam~~ Sample Proportion and Population Proportion.

test Statistic $Z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$

Where p - Sample Proportion

$P \rightarrow$ Population Proportion

$n \rightarrow$ Size of Sample

$Q = 1 - P$.

95% Confidence limits are

$$\frac{|p - P|}{\sqrt{\dfrac{pq}{n}}} \leq 1.96$$

i.e $\left(p - 1.96\sqrt{\dfrac{pq}{n}} \,,\; p + 1.96\sqrt{\dfrac{pq}{n}} \right)$.

Problems

1) 20% of manufactured Product is of top quality. In one day Production of 400 articles only 50 are of top Quantity. Verify the hypothesis and also find 95% confidence limit.

Soln

$$H_0 : p = P$$ [no difference between Sample Proportion and Population Proportion]

$$H_1 : p \neq P$$

$$P = 20\% = \frac{20}{100} = \frac{1}{5}$$

$$Q = 1 - P = 4/5$$

$$p = \frac{50}{400} = \frac{1}{8}, \quad n = 400$$

Let level of significance is 5%.

$$Z_\alpha = 1.96$$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \cancel{\frac{1/5 - 1/8}{\sqrt{\frac{1/8(7/8)}{400}}}}$$

$= -3.75$

$|z| > z_{\alpha}$

Reject Ho.

Sample Proportion and Population Proportion are not same.

ie 20% of the product manufactured is not of top quality.

95% Confidence limit

$$p - 1.96\sqrt{\frac{pq}{n}} \leq P \leq p + 1.96\sqrt{\frac{pq}{n}}$$

$$0.093 \leq P \leq 0.0157$$

2) A coin is tossed 256 times and 132 heads are obtained. would you conclude that the coin is biased one.

Sol

H_0 : Coin is unbiased ($P = \frac{1}{2}$)

H_1 : $P \neq \frac{1}{2}$ (biased)

Given $P = \frac{1}{2}$, $Q = 1 - P = \frac{1}{2}$

$$p = \frac{132}{256} = 0.5156 \quad (n = 256)$$

Let level of significance be 5%

$$Z_\alpha = 1.96.$$

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = 0.4992$$

$$|Z| < Z_\alpha.$$

Accept H_0.

Coin is unbiased one.

3) The fatality rate of typhoid Patients is believed to be 17.26%. In a certain year 640 Patients

suffering from typhoid in a hospital and only 63 patients dead. Can you consider the hospital sufficient.

<u>Sol</u>

$$\text{Let } P = 17.26\% = 0.1726$$

$$Q = 1 - P = 0.8274$$

$$p = \frac{63}{640} = 0.0984$$

$$n = 640.$$

$H_0 : p = P$ [Hospital is not sufficient]

$H_1 : p < P$. [~~Typhoid~~ typhoid patient is less. Hospital is sufficient]

(Left tail)

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = -4.96$$

(Sample < Pop Left tail)

Let the level of significance be 1%.

$$Z_\alpha = -2.33$$

$$|Z_\alpha| = 2.33$$

$$|Z| > |Z_\alpha| \quad \text{Reject } H_0.$$

~~The~~ Hospital is Sufficient

4) A salesman in a departmental stores claims that atmost 60% of the shoppers entering the store leaves without making the purchase. A random sample of 50 shoppers showed that 35 of them left without making a purchase. Are these sample results ~~Const~~ Consistent with the claim of the significance. Use a level of significance of 5%.

Sol

$$P = 60\% = 0.6$$

$$Q = 1 - 0.6 = 0.4$$

$p = \frac{35}{\cancel{50}\,50} = 0.7$ $\quad (n = 50)$

~~Let~~ $H_0 : p = P$

$H_1 : p > P$ (Right tail) [sample > pop] (Right)

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = 1.443$$

Let the level of ~~to~~ significane be 5%

$Z_\alpha = 1.645$

$|Z| < Z_\alpha$

Accept H_0.

60% of the shoppers entering the store leaves without making the purchase.

Test 2

①

Test Of Significance Of the difference between two sample proportions

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

$P \rightarrow$ Pop. Proportion $\quad Q = 1 - P$.

If P is not known

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

②

Problems ①

1) In a city, 20% of random sample of 900 boys had a slight physical defect. In another city 18.5% of a random sample of 1600 school boys had the defect. Is the differences between the proportions significant

Sol

$p_1 = 0.2 \qquad p_2 = 0.185$

$n_1 = 900 \qquad n_2 = 1600$

① T.O.P Pbm.

$z_\alpha = 1.96$ (2)

$H_0 : p_1 = p_2$

$H_1 : p_1 \neq p_2$

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

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.1904$$

$Q = 0.8096$

$$Z = \frac{0.2 - 0.185}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 0.92$$

$|z| < z_\alpha$.

Accept Z_0.

There is no significant diff bt P_1 & P_2 at 5% level

(3) (b) T.P Pbm

2) 15.5% of a random sample of 1600 UG were smokers whereas 20% of random sample of 900 PG were smokers in a state. Can we conclude that less no. of UG are smokers than PG

Sol

$p_1 = 0.1555$ $\quad p_2 = 0.2$

$n_1 = 1600$ $\quad n_2 = 900$

3

$H_0 : P_1 = P_2$

$H_1 : P_1 < P_2$ (one tail) (Left tail)

Let LOS be 5%

$Z_\alpha = -1.645$ $\quad P = 0.1712$

$Q = 0.8288$

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$= -2.87$

$|z| > |z_\alpha|$

Reject H_0 4

pbm

① 3 T.P

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

Sol

$$p_1 = \frac{800}{1000} = \frac{4}{5}$$

$$p_2 = \frac{800}{1200} = \frac{2}{3}$$

$H_0 : p_1 = p_2$

$H_1 : p_1 > P_2$ (Right tail)

Let LOS be 1%.

$Z_\alpha = 2.33$

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 6.82$$

$|z| > Z_\alpha$.

Reject H_0

4

Large sample (when n > 30)

(Normal test)
Test of significance for single mean

Type 3

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

where $\bar{x} \rightarrow$ sample mean

$\mu \rightarrow$ pop. mean

$\sigma \rightarrow$ S.D. of Popu

If Population S.D is not known

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

where $s \rightarrow$ Sampe S.D.

Note

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ are called

95% confidence limit

$\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$ 99% Confidence limit.

These fixed values are

	1%	5%
Two tail	2.58	1.96
Right tail	2.33	1.645
Left tail	-2.33	-1.645

A sample of 900 members (7) has a mean of 3.4 cm and S.D 2.61 cm. Is the sample from a large population of mean 3.25 cm and S.D 2.61 cm. Find the 95% confidence limits

Large Sample Test 3

Sol

Test 3 (1):

$n = 900$

$\bar{x} = 3.64$

$s = 2.61$

$\mu = 3.25$

$\sigma = 2.61$

H_0: Assume that the sample has been drawn from the pop with mean $\mu = 3.25$ $\bar{x} = \mu$

H_1: $\mu \neq 3.25$

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

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

$$= 1.724 < 1.96$$

~~tabt 1.96~~

we accept H_0.

The sample has been drawn from the pop with mean $\mu = 3.25$

95% confidence limits are

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

$$= 3.4 \pm 1.96 \times \frac{2.61}{\sqrt{900}}$$

$$= 3.4 \pm 0.1705$$

$$= 3.5 \text{ and } 3.2295$$

2) The mean breaking strength of cables is 1800. with a S.D 100. By a new technique in manufacturing process, it is claimed that breaking strength of cables have increased. In order to test this claim a 50 cables is test and found the mean breaking strength is 1850. Can we support the claim at 1% level of significance.

Sol

$n = 50.$

$\bar{x} = 1850$

$\mu = 1800$

$\sigma = 100$

test

(10)

$H_0 : \mu = 1800$

$H_1 : \mu > 1800$ (one tail)

~~Large sample test 3~~

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

$$= \frac{1850 - 1800}{100/\sqrt{50}} = \frac{50 \times \sqrt{50}}{100} \quad \frac{\cancel{50}}{\cancel{100}_2}$$

cal $z = 3.53$

tab $z = 2.33$

Reject H

ie, There is increase in breaking strength.

Test of Significance of difference of Mean (11)

(3)

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

D.M Last B

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

$\sigma_1 = \sigma_2 = \sigma$.

1) If σ_1, σ_2 is not known σ_1, σ_2 can be approximated by s_1, s_2

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

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

DMO

(1)

1) The mean of 2 large samples of 1000 and 2000 numbers are 67.5 inches and 68.0 inches resply. Can the samples the regarded as drawn from the same pop of S.D 2.5 inches

$n_1 = 1000$ $\qquad$ $n_2 = 2000$

$\bar{x}_1 = 67.5$ $\qquad$ $\bar{x}_2 = 68$

$\sigma = 2.5$

(12)

H_0 : The samples have been drawn from the same pop of SD 2.5 inches

ie, $\mu_1 = \mu_2$ and $\sigma = 2.5$

H_1 : $\mu_1 \neq \mu_2$

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

$$= \frac{67.5 - 68}{\sqrt{2.5^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}} = 5.16$$

> 1.96

Reject H_0.

1) In a Survey of buying habits, 400 women shoppers are chosen at random in super market A. Their average weekly food expenditure is Rs 250 with a S.D of Rs 50. For 400 women shoppers in super market B the average weekly food expenditure is Rs 220 with a S.D of Rs 55. Test at 1% level of significance whether the average weekly food expenditure of the two shoppers are equal.

Test 4

(14)

$n_1 = 400$ $\quad \bar{x}_1 = 250$ $\quad s_1 = 40$

$n_2 = 400$ $\quad \bar{x}_2 = 220$ $\quad s_2 = 55$

$H_0: \bar{x}_1 = \bar{x}_2$

$H_1: \bar{x}_1 \neq \bar{x}_2$

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

$$= \frac{250 - 220}{\sqrt{\frac{(40)^2}{400} + \frac{(55)^2}{400}}}$$

Cal $z = 8.82$

tab $z = 2.58$ (1% level)

Cal $z >$ tab z

H_0 Reject

Difference of Means (Problems in Normal test)

1) In a random sample of size 500, the mean is 20. In another indepL sample of size 400 the mean is 15. could the samples have been drawn from the same population with SD 4

Sol Given $n_1 = 500$, $n_2 = 400$

$\bar{x}_1 = 20$, $\bar{x}_2 = 15$, $\sigma = 4$.

$H_0 : \bar{x}_1 = \bar{x}_2$ the samples have been drawn from the same pop.

$H_1 : \bar{x}_1 \neq \bar{x}_2$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{20 - 15}{4\sqrt{\frac{1}{500} + \frac{1}{400}}}$$

$= 18.6$

Test ②

let LOS be 5%. (6)

tab z = ~~2.58~~ 1.96.

Cal z > tab z

The Ho Rejected.

The Samples could not have been drawn from the same Population.

Test 4 (2)

(b) A sample of heights of 6400 English men has a mean of 170 cm and S.D of 6.4 cm while a simple sample of heights of 1600 American has a mean of 172 cm and a S.D of 6.3 cm. Do the data indicate that Americans are on average taller than Englishmen

Sol

$n_1 = 6400$ $n_2 = 1600$ Test 4

$\bar{x}_1 = 170$ $\bar{x}_2 = 172$ (8)

$n_1 = 6.4$ $n_2 = 6.3$

$H_0 : \mu_1 = \mu_2$ (or) $\bar{x}_1 = \bar{x}_2$

$H_1 : \mu_1 < \mu_2$. Left (tail) (vi)

tab $Z_\alpha = 2.33$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = 11.32$$

$$|Z| > |Z_\alpha^{tab}|.$$

Reject H_0.

The Americans are on the average taller than Englishmen

6

Test 4 (3)

3) The average marks scored by 32 boys is 72 with a S.D of 8 while that for 36 girls is 70, with a S.D of 6. Test at 1% level of significance whether the boys perform better than girls

(18)

Sol $H_0 : \bar{x}_1 = \bar{x}_2$

$H_1 : \bar{x}_1 > \bar{x}_2$ (R Tail Test)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = 1.15$$

tab $z = 2.33$

Test 4

(4)

cal z < tab z.

H_0 Accept.

Case II

we cannot conclude that boys perform better than girls