Module-3

Statistical Inference-1

population or univarse

A large collection of Numerical data or attributes is called population.

Sample.

A small section Selected from the population or finite Subset of population is called sample

Sample size

The number of individuals or elements in the sample is called sample size

Sampling

The process of drawing sample from the population is called sampling.

Here two kinds

i) Sampling with replacement.

ii) Sampling without replacement.

Sampling with replacement

In Sampling if each ~~ea~~ element has more than one chance of getting selected, then it is sampling with replacement

(Here After the first draw the element is put back to the population before second draw.)

Mean $= \mu$

Variance $= \dfrac{\sigma^2}{n}$

Here μ : mean of population

σ : Variance of population

n : Sample Size

Sampling without replacement

In sampling if each element of population has only one chance of getting selected is called sampling without replacement.

(Here After the first draw the element is not put back to the population)

Here Mean = μ

variance = $\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

N: Size of population

n: Sample size

σ : Std deviation of population.

Testing of Hypothesis

Hypothesis

To reach decisions about the population on the basis of information revealed by sample, we make certain assumptions about population. which may or may not be true is called Hypothesis

or

Assumption about the population which may or maynot true is called Hypothesis.

Null Hypothesis

Hypothesis formulated for the purpose of its rejection under the assumption that it is true is called Null Hypothesis. "denoted by H_0"

Alternate Hypothesis

Any Hypothesis which is complimentary to the null Hypothesis is called Alternate hypothesis.
"Denoted by H_1"

Type-1 error

If a Hypothesis is rejected while it should have been accepted is called type-1 error.

Type-2 error

If a Hypothesis is accepted while it should have been rejected is called type-2 error.

Level of Signifiance

The probability level below which we reject the Hypothesis is called Significance level

Generally we have,

i) 5% level of Signifiance. (95% confidence level)

ii) 1% level of Significance. (99% confidence level)

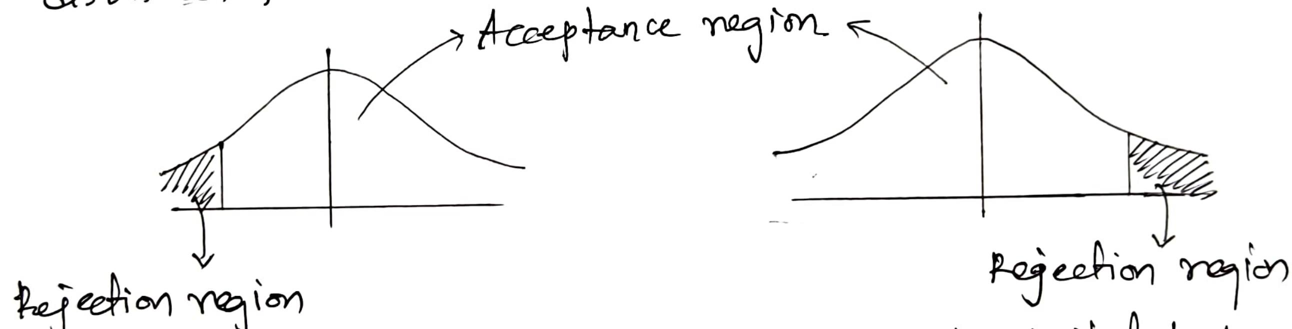
Confidence interval

The probability that the population parameter will fall between two set of values for certain proportion of time is called confidence interval.

one tailed & two tailed tests

one tailed test

Region of rejection lies entirely in one end of the distribution.

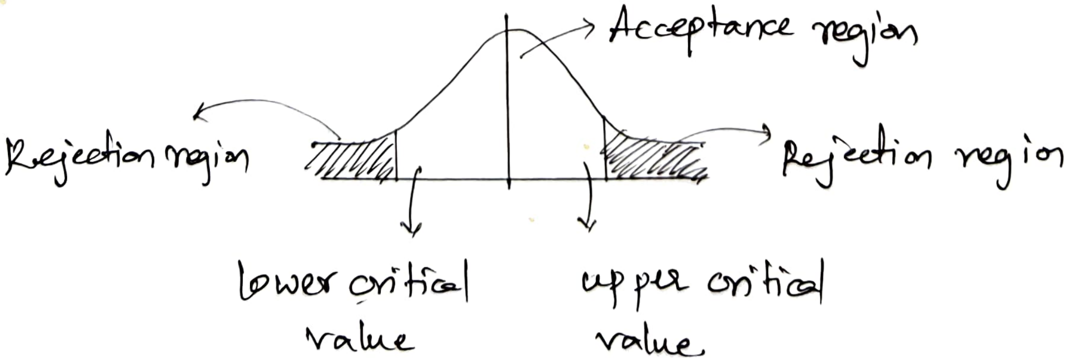


Left tailed test

Right tailed test

Two tailed test

Here critical region split into two equal parts placed at each tail of the distribution



Testing of Hypothesis

Procedures which enable to decide whether to accept or reject the hypothesis are called test of Hypothesis or test of Significance..

Note:

(i) If Sample size is less than 30 it is considered as Small sample

(ii) If Sample size is more than 30 it is considered as Large samples.

Test of Significance for Large Samples

Following are the important test of Significance

i) Test of Significance for single proportion.

ii) Test of Significance for difference of proportion.

iii) Test of Significance for single mean.

iv) Test of Significance for difference of mean..

Note:

1) At 5% level of Significance

(i) if $-1.96 < z < 1.96$ then Hypothesis accepted

ii) Else Hypothesis rejected

At 1% level of Significance

(i) If $-2.58 < z < 2.58$ then Hypothesis accepted

(ii) Else Hypothesis rejected

Test of Significance for Single proportion

This test is used to find the significant difference between the proportion of sample & population.

Here
$$\boxed{z = \frac{p - M}{\sigma} = \frac{p - np}{\sqrt{npq}}}$$

where p is sample ~~mean~~ proportion

Test of Significance for difference of proportion.

This test is used to ~~find~~ test the significance of the difference b/w the sample proportions, The test statistic under the null hypothesis H_0 that there is no significant difference b/w the sample proportions (i.e $P_1 = P_2$)

Here
$$\boxed{z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$$
where $p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$

$q = 1 - p$

Here P_1 is sample proportion of large sample with sample size n_1

P_2 is sample proportion of large sample with sample size n_2

'Also the above two samples are from different population.'

Test of Significance for Single mean

This test is used to find the significant difference between sample mean & population mean.

Here test statistic is $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$

where $\bar{x}$ - Sample mean

μ - population mean

σ - Standard deviation (population or Sample)

n - Sample size

Note: confidence limit of Mean of population

For population mean, the confidence limits are

(i) 95% confidence limit (0.05 Significance level)

$$\boxed{\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}}$$

(ii) 99% confidence limit (0.01 Significance level)

$$\boxed{\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}}$$

Test of Significance for difference of means

To test the equality of two ~~proportions~~ population means (ie to test $\mu_1 = \mu_2$)

or

To test the significance of difference between the two independent means (ie $\bar{x}_1 - \bar{x}_2$).

The test statistic under the null hypothesis Ho that there is no significant difference between sample means (ie $\bar{x}_1 = \bar{x}_2$)

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

where $\bar{x}_1$ & $\bar{x}_2$ are sample mean of different population

σ_1 & σ_2 are standard deviation of population respectively

n_1 & n_2 are sample size of respective population.

Note:

Nature of test	Level of Significance 1%	5%
Two tailed test	2.58	1.96
Right tailed test	2.33	1.645
Left tailed test	-2.33	-1.645

Test of Significance of single proportion.

This test is used to find the significant difference between, proportion of the sample & the population.

problems:

1) A coin was tossed 1000 times and head turned up 540 times. Can we infer that coin is unbiased at 1% level of significance

(unbiased means equal chance of getting & tail)

Sol: Given $n = 1000$ let ~~H₀~~: be coin is unbiased. is Hypothesis.

$\bar{x} = 540$

$\Rightarrow p = 1/2, \; q = 1/2$

w.k.t $\mu_x = np = 1000 \times \frac{1}{2} = 500$

$\sigma_x = \sqrt{npq} = \sqrt{1000 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{250} = 15.81$

$z = \dfrac{\bar{x} - \mu_x}{\sigma_x} = \dfrac{540 - 500}{15.81} = 2.53$ (Since $|z_{cal}| < 2.53$)

Since $z = 2.53 < 2.58$

$\Rightarrow$ At 1% level of Significance (99% confidence level) we can accept the Hypothesis that the coin is unbaiced.

Q. A die was thrown 9000 times & a throw 5 or 6 was obtained 3240 times on the assumption of random throwing, do the data indicate that the die is unbaised?

Sol: Given $P = 3240$, $n = 9000$

here $\boxed{p = \frac{2}{6} = \frac{1}{3}}$ (5 or 6 outcome is success)

$q = 1 - p = 1 - 1/3 \rightarrow \boxed{q = 2/3}$

let H_0: let the die is unbaised (ie $p = 1/3$)

H_1: Die is baised (i.e $H_1: p \neq 1/3$)

Now $z = \frac{P - np}{\sqrt{npq}} = \frac{3240 - 9000 \times 1/3}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}}$

$z = 5.3733$

$z_{cal} = 5.3733$

5% L.O.S

$|z_{cal}| > z_{tab}$

$5.377 > 1.96$

H_0 is rejected

1% L.O.S

$|z_{cal}| > z_{tab}$

$5.377 > 2.58$

H_0 is rejected

Q. A coin was tossed 400 times & the head turned up 216 times. Test the hypothesis at 5% Level of Significance that coin is unbaised.

Sol: Given $n = 400$, $P = 216$

clearly $p = 1/2$ (Getting head turn)

$q = 1/2$

let H_0 : coin is unbaised. (i.e $H_0 : p = 1/2$)

H_1 : coin is baised (ie $H_1 : p \neq 1/2$)

$$z = \frac{P - np}{\sqrt{npq}} = \frac{216 - 400 \times 1/2}{\sqrt{400 \times \frac{1}{2} \times \frac{1}{2}}}$$

$z = 1.6$

$z_{cal} = 1.6$

~~Since~~ $\cancel{|z_{cal}|} < \cancel{1.96}\ \cancel{z_{tab}}$ 1.96

Since $|z_{cal}| < |z_{tab}|$

$1.6 < 1.96$ at 5% L.O.S

$\Rightarrow$ H_0 is accepted ie coin is unbaised

Q: A die was thrown 1200 times and the number '6' was obtained 236 times, can the die is considered as fair at 0.01 level of significance.

Sol: Given $P = 236$ $n = 1200$

$$\boxed{p = 1/6}$$

$$q = 1 - p = 1 - 1/6 \Rightarrow \boxed{q = 5/6}$$

let H_0 : Die is fair (unbaiced)

H_1 : Die is not fair (baiced)

Now $$z = \frac{P - np}{\sqrt{npq}} = \frac{236 - 1200 \times \frac{1}{6}}{\sqrt{1200 \times \frac{1}{6} \times \frac{5}{6}}}$$

$$z = 2.7778$$

$$\boxed{z_{cal} = 2.778}$$

w.k.t At 1% level of significance $z_{tab} = 2.58$

Since $|z_{cal}| > |z_{tab}|$

$$2.778 > 2.58$$

$\Rightarrow$ H_0 is rejected

i.e coin is baiced one.

8. In a city A, 20% of a random sample of 900 boys has certain physical defect, in Another city B, 18.5% of random sample of 1600 boys had the same defect. Is the difference b/w the proportion Significant.?

Sol: Given $P_1 = 0.2$ (20%) $P_2 = 0.185$ (18.5%)

$n_1 = 900$ $n_2 = 1600$

let H_0: There is no significant difference b/w proportions

H_1: There is significant difference b/w proportions.

Now w.k.t
$$z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{—— (1)}$$

$$p = \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2} = \frac{900 \times 0.2 + 1600 \times 0.185}{900 + 1600} = 0.1904$$

$\boxed{p = 0.1904}$ $\Rightarrow$ $q = 1 - p$

$= 1 - 0.1904$

$\boxed{q = 0.8096}$

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

$z = 0.9146$ $\rightarrow$ $\boxed{z_{cal} = 0.9146}$

At 5% L.O.S

$z_{tab} = 1.96$

Since $|z_{cal}| < z_{tab}$

$0.9146 < 1.96$

$\Rightarrow$ H_0 is accepted

At 1% L.O.S

$z_{tab} = 2.58$

$|z_{cal}| < z_{tab}$

$0.9146 < 2.58$

$\Rightarrow$ H_0 is accepted

ie

Q. one type of air craft is found to develop engine trouble in 5 flights out of 100 and another type in 7 flights out of total 200 flights. Is there a significant difference in two types of air crafts so far as engine defects are concerned?

Sol. Given $P_1 = \frac{5}{100} = 0.5$ $\qquad P_2 = \frac{7}{200} = 0.035$

$n_1 = 100$ $\qquad n_2 = 200$

Let H_0: There is no significant difference between the air crafts so far as engine defects are concerned.

H_1: There is a significant difference b/w the air craft so far engine defects are concerned..

$$P = \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2} = \frac{0.5 \times 100 + 0.035 \times 200}{100 + 200}$$

$\boxed{P = 0.04}$ $\rightarrow$ $\boxed{q = 0.96}$

Now the test Statistic Ho is

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

$$z = \frac{0.05 - 0.035}{\sqrt{(0.04)(0.96)\left(\frac{1}{100}+\frac{1}{200}\right)}}$$

$$\boxed{z_{cal} = 0.625}$$

5% L.O.S

Since $z_{tab} = 1.96$

Since $|z_{cal}| < z_{tab}$

$0.625 < 1.96$

$\Rightarrow$ Ho accepted

ie No Significant difference.

1% L.O.S

$z_{tab} = 2.58$

Since $|z_{cal}| < z_{tab}$

$0.625 < 2.58$

$\Rightarrow$ Ho accepted

ie No Significant difference

Q) In a exit poll enquiry it was revealed that 600 voters in one locality and 400 voters from another locality favoured 55% & 48% respectively. a particular party to come to power. Test the Hypothesis that there is a difference in the locality in ~~rey~~ respect of the opinion.

Sol: Given. $P_1 = 0.55$ (55%) $\quad P_2 = 0.48$ (48%)

$n_1 = 600 \quad n_2 = 400$

let H_0: There is no significant difference in the locality in respect to the opinion

H_1: There is significant difference in the locality in respect to the opinion

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{600 \times 0.55 + 400 \times 0.48}{600 + 400}$$

$$\boxed{P = 0.522} \rightarrow q = 1 - p$$

$$= 1 - 0.522$$

$$\boxed{q = 0.478}$$

The test Statistic H_0 is

$$z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.55 - 0.48}{\sqrt{0.522 \times 0.478\left(\frac{1}{600} + \frac{1}{400}\right)}}$$

$$z = 2.174$$

$$\boxed{z_{cal} = 2.174}$$

At 5% L.O.S

w.k.t $z_{tab} = 1.96$

Since $|z_{cal}| > z_{tab}$

$2.174 > ~~2~~ 1.96$

$\Rightarrow$ H_0 rejected

At 1% L.O.S

$z_{tab} = 2.58$

Since $|z_{cal}| < z_{tab}$

$2.174 < 2.58$

$\Rightarrow$ H_0 is accepted

Q. A random Sample of 100 recorded deaths in past year Showed an average life Span of 71.8 years. Assuming population S.D of 8.9 years, does the data indicate that the average Span today is greater than 70 years, use 5% level of Significance.

Sol Given $n = 100$, $\bar{x} = 71.8$, $\sigma = 8.9$, test for $\mu > 70$

let H_0: The average life Span is 70 years

$H_0 : \mu = 70$

H_1: Average life Span is greater than 70 years

$H_1 : \mu > 70$ (right tailed test)

clearly $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{71.8 - 70}{\left(\frac{8.9}{\sqrt{100}}\right)}$

$$\boxed{z_{cal} = 2.022}$$

W.k.t At 5% L.O.S $z_{tab} = 1.645$

Since $z_{cal} > z_{tab}$

$2.022 > 1.645$

$\Rightarrow$ H_0 is rejected

$\Rightarrow$ The average life span is greater than 70 years

Q. A Sample of 200 tyres is taken from a lot. the mean life of tyres is found to be 40,000 kms with standard deviation of 3200 kms. Is it reasonable to assume the mean life of tyres in the lot of 41,000 kms ? also establish 95% confidence limits within which mean life of tyres in the lot is expected to lie.

Sol: Given $\bar{x} = 40,000$, $S = 3200$, $\mu = 41,000$ is test

$n = 200$

let H_0: population mean is 41000 $(\mu = 41,000)$

H_1: population mean $\neq$ 41,000 (two-tailed)

Test Statistic is

$$z = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{41000 - 40,000}{\left(\frac{3200}{\sqrt{200}}\right)} = -4.4194$$

$\therefore \; |z_{cal}| = 4.4194$

w.k.t At 5% L.O.S $\quad z_{tab} = 1.96$

Since $|z_{cal}| > z_{tab}$

$4.4194 > 1.96$

H_0 is rejected

i.e The mean life of the tyres in the lot is not equal to 41,000 km

Q. A Stenographer claims that she can type at rate of 120 wpm. can we reject her claim on the basis of 100 trails in which she demonstrates a mean of 116 wpm. with std deviation of 15 words? use 5% Level of significance

Sol. Given $\quad n = 100, \quad \bar{x} = 116, \quad \sigma = 15$

Test the hypothesis for $\mu = 120$.

let H_0: let $\mu = 120$ (stenographer claim)

H_1: let $\mu \neq 120$

Now $\quad z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{116 - 120}{15/\sqrt{100}} = -2.66$

$$\boxed{z_{cal} = -2.66}$$

At 5% L.O.S w.b.t $z_{tab} = 1.96$

Since $|z_{cal}| > z_{tab}$

$2.66 > 1.96$

$\Rightarrow H_0$ is rejected

ie Stenographer's claim that she type 120 wpm is rejected

Q. A Sample of 100 bulbs produced by company showed mean life of 1190 hours and standard deviation 90 hrs. also a sample of 75 bulbs produced by a company B showed a mean life of 1230 hours and standard deviation of 120 hrs. Is there a difference between the mean life of the bulbs produced by the two companies at (i) 5% Significance level

(ii) 1% Significance level

Sol: Given $n_1 = 100$, $\bar{x}_1 = 1190$, $\sigma_1 = 90$

$n_2 = 75$, $\bar{x}_2 = 1230$ $\sigma_2 = 120$

let H_0: There is no significant difference between the mean life time of bulbs produced by company

H_1: There is ~~no~~ significant difference b/w mean life time of bulbs produced by company.

Now the test Statistic for testing H_0 is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1190 - 1230}{\sqrt{\frac{90^2}{100} + \frac{120^2}{75}}} = 2.4209$$

i.e $\boxed{z_{cal} = 2.4209}$

At 5% L.O.S	1% L.O.S
w.k.t $z_{tab} = 1.96$	w.k.t $z_{tab} = 2.58$
Since $\lvert z_{cal} \rvert > 1.96$	Since $\lvert z_{cal} \rvert < 2.58$
$2.4209 > 1.96$	$2.4209 < 2.58$
$\Rightarrow H_0$ rejected	$\Rightarrow H_0$ is accepted
i.e There is significant difference	i.e There is no significant difference

Q. A random sample of 1000 workers in a company has mean wage of Rs 50 per day and standard deviation of Rs 15. Another sample of 1500 workers from another company has mean wage of Rs 45 per day and S.D of Rs 20. Does the mean rate of wages varies between the two companies? Find the 95% confidence limits for the difference of mean wages of the population of two companies.

Sol: Given $n_1 = 1000$ $\quad \bar{x}_1 = 50$ $\quad \sigma_1 = 15$

$n_2 = 1500$ $\quad \bar{x}_2 = 45$ $\quad \sigma_2 = 20$

H_0 : There is no significant difference b/w mean wages b/w two companies.

H_1 : There is significant difference b/w mean wages b/w two companies.

Test Statistic for testing H_0 is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{50-45}{\sqrt{\frac{15^2}{1000} + \frac{20^2}{1500}}} = 7.13$$

$$\boxed{z_{cal} = 7.13}$$

At 5% Level of Significance w.k.t $z_{tab} = 1.96$

Since $|z_{cal}| > z_{tab}$

$\Rightarrow$ $7.13 > 1.96$

$\rightarrow$ H_0 rejected (i.e There is significant difference)

At 1% level of Significance w.k.t $z_{tab} = 2.58$

Since $|z_{cal}| > z_{tab}$

$7.13 > 2.58$

$\rightarrow$ H_0 rejected (i.e There is significant difference)

Also The confidence limit is given by.

$$(\text{Statistic}) \pm \text{Critical vale} \times \left(\text{Standard error of Statistic}\right).$$

At 5% L.O.S 95% confidence limits

$$\therefore = (\bar{x}_1 - \bar{x}_2) \pm 1.96 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$= 5 \pm 1.96 (0.7012)$$

$$= 5 \pm 1.374.$$

ie 3.626 and 6.374.

Thus we say that 95% confidence that the difference of population mean of wages between two companies lies between. Rs 3.626 and 6.374

Q. Before an increase in excise duty on tea 800 persons out of sample of 1000 persons were found to be tea drinkers, After an increase in excise duty 800 people out of 1200 people found to be tea drinker. Using standard error of proportion state whether there is a significant decrease in the consumption of tea after the increase in excised duty?

Sol Here $n_1 = 1000$ $n_2 = 1200$

$$P_1 = \frac{800}{1000} \qquad P_2 = \frac{800}{1200}$$

$$P_1 = 0.80 \qquad P_2 = 0.67$$

let $H_0: P_1 = P_2$

$H_1: P_1 > P_2$ (Right tailed test)

Now
$$z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{0.800 \times 1000 + 0.67 \times 1200}{1000 + 1200}$$

$$\boxed{p = 0.73}$$

$$\Rightarrow q = 1 - p = 1 - 0.73 \Rightarrow \boxed{q = 0.27}$$

Now $$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.80 - 0.67}{\sqrt{0.73 \times 0.27 \left(\frac{1}{1000} + \frac{1}{1200}\right)}} = 6.5$$

$$\boxed{z_{cal} = 6.5}$$

Now At 5% L.O.S

$z_{tab} = \cancel{1.96}\ 1.645$

$|z_{cal}| > z_{tab}$

$6.5 > 1.645$

H_0 rejected

At 1% L.O.S

$z_{tab} = 2.33$

$|z_{cal}| > z_{tab}$

$6.5 > 2.33$

H_0 rejected

ie There is significant difference in the consumption of tea after increase in the exciseduty.