Unit-III

Testing of Hypothesis (theory part also be asked)

Population : An aggrigate of object under study. It is also called universe.

Sample : A finite subset of population. It is small part of universe.

Sample Size : The no. of individuals in the sample.

Sampling : The process of selecting the sample from the universe.

Parameter : The statestical consepts of the population. Such as mean, variance etc.

Statastics : The statastical concept from the sample from the members of the sample.

Population mean and variance are given by μ and σ^2

Sample mean and variance are given by $\bar{x}$ and s^2

Test of Significance : It unables us to decide on the basis of the results of the sample whether

(a) Deviation b/n observed sample statistic and hypothetical parameter value or

(b) Deviation b/n two sample statistic

is significant or might be attributed to pluctuations of in samplings.

For applying the test of Significance

(i) null hypothesis H_0 and alternate hypothesis H_1 are set up.

null hypothesis H_0 is a definite statement and alternate hypothesis H_1 is complementary of null hypothesis.

Example : Null hypothesis that population has specified meaning.

null → $H_0 : \mu = \mu_0$ (mean = specified mean μ_0)

alternative → H_1 : $\mu \neq \mu_0$ (Two tailed test)

$\mu > \mu_0$ (right tailed test)

$\mu < \mu_0$ (left tailed test)

Critical region : A region corresponding to statistics in the sample space which amounts to the rejection of null hypothesis is called critical region or region of rejection. The region of sample space which amounts to the acceptance of a null hypothesis is called region of acceptance.

Level of Significance : The probability of the value of a variate falling in the critical region is known as level of significance

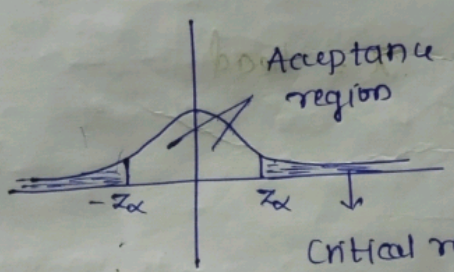
Errors

type 1 : When null hypothesis is true it is rejected.

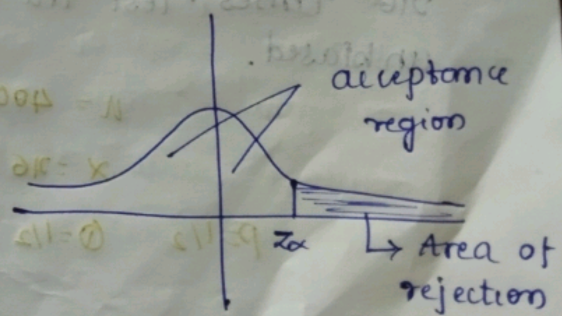
type 2 : When null hypothesis is false it is accepted.

Critical values or significant values (Z_α) : The values of the test statastics which seperates the critical region and acceptance region.

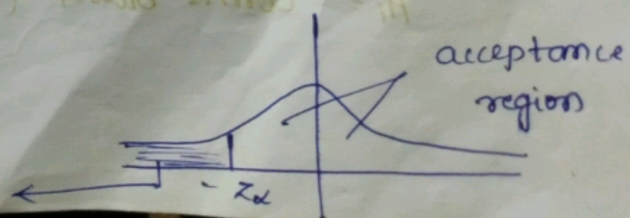
Two tailed test



Right tailed test



Left tailed test



Level of Significance

	1% (0.01)	5% (0.05)	10% (0.1)
Two tailed test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.966$	$\lvert Z_\alpha \rvert = 0.645$
Right tailed test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left tailed test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

Test of Significance for large Samples: If the sample size $n > 30$ the sample is taken as large sample.

* Testing of Significance for single propotion

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

where $p = \frac{X}{n}$ is the observed propotion of success. i.e X the no. of successes in n independent trials with constant probability P of success of each trial. and Q is $1 - P$. Z is called test statistics which is used to test significant difference of sample in population popotion.

① A coin was tossed 400 times and head turned up 216 times. Test the hypothesis that the coin is unbiased.

$n = 400$ $\qquad$ x = head

$x = 216$

$p = 1/2$ $\quad$ $Q = 1/2$

H_0 = coin is unbiased ($\mu = \bar{x}$)

H_1 = Coin is biased ($\mu \neq \bar{x}$) (two tailed test)

$$p = \frac{x}{n} = \frac{216}{400} = 0.54$$

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

$z = 1.6 < 1.966$ at 5% level of significance.

$\therefore$ Ho is accepted.

(3) 20 people where attacked by a desease and only 18 servive. will you reject the hypothesis that the servival rate if attacked by this deasese is 85% in favour of the hypothesis that it is ~~known~~ more. test at 5% level of significance (using large sample test).

$n = 20 \qquad X = 18$ (no. of successes)

$p = \frac{X}{n} = \frac{18}{20}$

$P = 0.85 \qquad Q = 1 - 0.85 = 0.15$

Hypothesis.

H_0: Servival rate after attack of desease $(P = 0.85)$

H_1: $\cancel{P}$ $(P > 0.85) \rightarrow$ right tailed test.

$$Z = \frac{p - P}{\sqrt{PQ/n}} = \frac{0.9 - 0.85}{\sqrt{0.85 \times 0.15/20}} = 0.626 < 1.645 \ (5\%)$$

$\therefore H_0$ is accepted.

<u>Note</u>: Probable limits for observed proportion of sucess. is given by $p \pm Z_\alpha \sqrt{\frac{pq}{n}}$ and $q = 1 - p$.

$$p - Z_\alpha \sqrt{\frac{pq}{n}} < \mu < p + Z_\alpha \sqrt{\frac{pq}{n}}$$

4. A random sample of 500 apples was taken from a large consignment and 60 were found bad. Obtain the 99% confidence limits for the percentage of bad apples of consignment.

$n = 500$

$X = 60$

99% of confidence limits means 1% of significance so $Z_\alpha = 2.58$

$p = \frac{60}{500} = 0.12$

$q = 0.88$

$Z_\alpha = 2.58$

Probable limit for observed proportion of success. $\Rightarrow p \pm \sqrt{\frac{pq}{n}}\, z_\alpha$

$$\Rightarrow 0.12 \pm \sqrt{\frac{0.12 \times 0.88}{500}} \times 2.58$$

$$\Rightarrow 0.1574 \text{ and } 0.0825$$

$$\Rightarrow 8.25 < P < 15.74$$

$$\Rightarrow 8\% < P < 16\%$$

Hence 99% confidence limit for percentage of bad apple in the consignment is $[8\%, 16\%]$.

5. 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 are equal popular in this State at 1% level of significance.

$n = 1000$ $\quad P = 0.5$ (probability of success i.e probability of eating rice 1/2 we have only 2 option)

$x = 540$ $\quad Q = 0.5$

$p = 0.54$

$H_0 : P = 0.5$

$H_1 : P \neq 0.5$ (two tailed test)

$Z = 2.52 < 2.58$ at 1% level of significance

$\therefore$ Ho is accepted.

We can assume that both rice and wheat eaters are equal popular in this state.

Test of difference b/n proportions

Consider two samples X_1 and X_2 of sizes of n_1 and n_2 respectively. taken from two different populations.
Two test the significance of the difference b/n the sample proportion P_1 and P_2 the test statistic under the null hypothesis that there is no significant difference b/n two sample proportions is

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

where $\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$ $\quad \hat{Q} = \frac{n_1 q_1 + n_2 q_2}{n_1 +}$

$\hat{Q} = 1 - \hat{P}$

6. Before an increase in tax on Tea 800 people out of 1000 people were found to be tea drinkers. After an increase in the tax 800 people were known to be tea drinkers in a sample of 1200. Do you think there has been a significant decrease in consumption of tea after the increase in tax.

$n_1 = 1000 \qquad n_2 = 1200$

$x_1 = 800 \qquad x_2 = 800$

Ho (Null hypothesis) : there is no significance diff b/n tea drinkers after and before tax.

H_1 : $t_1 > t_2$ (right tailed test)

t_1 i.e tea drinkers before tax are greater than after tax.

$$P_1 = \frac{x_1}{n_1} = \frac{800}{1000} = 0.8$$

$$P_2 = \frac{x_2}{n_2} = \frac{800}{1200} = 0.66$$

$\hat{P} = 0.72$

$\hat{Q} = 0.28$

$Z = 7.28 > 1.645$ at 5% level of significance

Ho is rejected

$H_1 > P_2$

Test of Significance for single mean

Here the difference between sample mean and population mean is significant or not is tested.
Under the null hypothesis there is no difference b/n sample mean and population mean.

$$Z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$$

$\bar{x}$ - Sample mean
μ = population mean
σ - standard deviation of population
n - Sample size

Note:

1. If σ is not known then $Z = \frac{\bar{x} - \mu}{(S/\sqrt{n})}$

   where, S - standard deviation of sample.

2. The limit of the population mean μ are given by

$$\left(\bar{x} - z_{\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)\right) < \mu < \left(\bar{x} + z_{\alpha}\left(\frac{\sigma}{\sqrt{n}}\right)\right)$$

1) A normal population has a mean of 6.8 and standard deviation of 1.5. A sample of 400 members gave a mean of 6.75 is the difference significant?

$\mu = 6.8$ $\quad n = 400$
$\sigma = 1.5$
$\bar{x} = 6.75$

H_0 - there is no difference b/n sample and population mean i.e $\bar{x} \not= \cancel{\mu\ 6.8}$ $\mu = 6.80$

H_1 - there is diff b/n $\bar{x}$ and μ i.e $\overset{\mu \neq 6.80}{\cancel{\bar{x} \neq \mu}}$ (two tailed test)

two tailed test so mod.

$$Z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} = \frac{6.75 - 6.8}{(1.5/20)} = |-0.6661| = 0.67 < 2.58$$

So Z is accepted at 1% and 5% level of Signifance.

2) The mean weight obtained from a random sample of size 100 is 64 grms. The standard deviation of the weight distribution of the population is 3 grms. Test the statement that the mean weigut of popution is 64 grms at 5% level of significance. Also set up 99% confidence limit of mean weigut of the population.

$\mu = 67$ $\quad n = 100$

$\bar{x} = 64$ $\quad \sigma = 3$

Ho = there is no significance diff b/n μ and $\bar{x}$ i.e $\cancel{\mu \neq \bar{x}}$ $\cancel{\bar{x} \neq 67}$ $\mu = 67$

Ho = therr is a significance diff i.e $\cancel{\mu \neq \bar{x}}$ $\cancel{\bar{x} \neq 67}$ $\mu = 67$ (two tailed test)

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{64 - 67}{3/\sqrt{100}} = |-10| = 10 > 2.58 \text{ at } 1\%$$

level of significance

so it is rejected.

limit of mean at 99% confidence i.e 1% significant.

$$\bar{x} - z_\alpha(\sigma/\sqrt{n}) < \mu < \bar{x} + z_\alpha(\sigma/\sqrt{n})$$

$$64 - 2.58\left(\frac{3}{10}\right) < \mu < 64 + 2.58\left(\frac{3}{10}\right)$$

$\cancel{61 < \mu < 67}$

$63.226 < \mu < 64.77$

$[63\cancel{.}, 65\cancel{.}]$

3) The average marks in mathematics of a sample of 100 students was 51 with standard deviation 6 marks. Could this have been from a random ~~sample from~~ population with average marks 50.

$\mu = 50$ $\quad s = 6$

$\bar{x} = 51$ $\quad n = 100$

$\cancel{\sigma}$ $\quad$ Ho : $\mu = 50$

H_1 : $\mu \neq 50$ (two tailed test)

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{51 - 50}{6/\sqrt{100}} = \frac{1}{6/10} = 1.66 < 2.58 \text{ and } < 1.96$$

So accepted at both significant levels.

4) A sample of 900 members has a mean 3.4cm and standard deviation 2.61cm. Is the sample from the large population of mean 3.25cm and standard deviation 2.61cm. If the population is normal and its mean is not known find the 95% and 99% confidence limits of the true mean.

$n = 900$ $\bar{x} = 3.4$. $\sigma = 2.61$ $s = 2.61$ $\mu = 3.25$

H0 : there is no significance diff b/n mean ($\mu = 3.25$)

H1 : there is a significant diff ($\mu \neq 3.25$) (two tailed)

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.25}{2.61/\sqrt{900}}$$

$$= 1.724 < 2.58 \text{ and } < 1.966$$

So accepted at both signifacnt levels 1% and 5%.

limit of true mean at 1% signifancant level

$$\bar{x} - z_\alpha\left(\sigma/\sqrt{n}\right) < \mu < \bar{x} + z_\alpha\left(\sigma/\sqrt{n}\right)$$

$$3.4 - 2.58\left(\frac{2.61}{30}\right) < \mu < 3.4 + 2.58\left(\frac{2.61}{30}\right)$$

$$3.17 < \mu < 3.62$$

$[3.17, 3.62]$

limit of true mean at 5% significant level

$$3.4 - 1.966\left(\frac{2.61}{30}\right) < \mu < 3.4 + 1.966\left(\frac{2.61}{30}\right)$$

$$3.22 < \mu < 3.57$$

$[3.22, 3.57]$

5) The guranted average life of a certain type of electric bulbs is 1000 hrs and standard deviation 125 hrs. It is decided to sample the output so as to ensure that 90% of the bulbs do not fall short of the guranted average by more than 2.5%. What must be the minimum size of the sample.

$\mu = 1000$

$\sigma = 125$

$\bar{x} > 1000 - 2.5\%$ (should not fall short of guranted avg).

$\bar{x} > 975$

Since we donot want the sample mean to be less than the guarented mean more than 2.5% :

we have $\bar{x} > 975$

Let n be the size of the sample and $z = \dfrac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

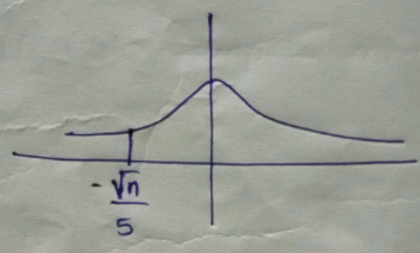
$$z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{975-1000}{\frac{125}{\sqrt{n}}} = -\frac{\sqrt{n}}{5}$$

$P\left(z > -\frac{\sqrt{n}}{5}\right) = 0.9$ (Hence we have considered $\bar{x} > 975$ then $P(z > \sqrt{n}/5)$)

$$P\left(z > -\frac{\sqrt{n}}{5}\right) = 0.9$$

$$0.5 + A\left(\frac{\sqrt{n}}{5}\right) = 0.9$$

$$A\left(\frac{\sqrt{n}}{5}\right) = 0.4$$



given $\boxed{A(1.28) = 0.4}$

$$\frac{\sqrt{n}}{5} = 1.28$$

$$n = 40.96$$

$$n = 41$$

Test of Significance for difference of means.

Let $\bar{x}_1$ be the mean of a sample of size n_1 and $\bar{x}_2$ be the mean of sample of size n_2 with the population mean μ_1 and variance σ_1^2 and population mean μ_2 and variance σ_2^2 respectively then

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

Under the null hypothesis that the samples are drawn from the same population i.e $\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$

then

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Note: (1) If σ_1 and σ_2 are not known and σ_1 is not equal to σ_2 $(\sigma_1 \neq \sigma_2)$, $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

(2) If $\sigma_1 = \sigma_2 = \sigma$ and σ is not known then

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

1) The means of two single large samples of 1000 and 2000 members are 67.5 inches and 68 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches. Test at 5% level of significance.

$n_1 = 1000$ $\quad n_2 = 2000$

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

$\sigma = 2.5$

H_0 = null hypothesis : samples are from same population
$(\sigma_1 = \sigma_2)$

H_1 = alternate hypothesis : samples are from diff population
$(\sigma_1 \neq \sigma_2)$ (two tailed test)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \cancel{\frac{1000 - 2000}{2.5\sqrt{\frac{1}{67.5} + \frac{1}{68}}}} =$$

$$= \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$z = |-5.16| = 5.16 > 1.96$$

Hence ~~Ho~~ is rejected at 5% significance level

2) ~~Intdigent~~ test were given to two groups of boys and girls

	Mean	S.D	Size
Girls	75	8	60
Boys	73	10	100

~~Ex~~ Examinad. if there is a diff b/n the mean score which is significant.

$\bar{x}_1 = 75$ $\bar{x}_2 = 73$

$S_1 = 8$ $S_2 = 10$

$n_1 = 60$ $n_2 = 100$

null hypothesis : H_0 = there is no diff b/n ~~samples~~ mean score.

alternate : there is a diff b/n mean score

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{75 - 73}{\sqrt{\frac{8^2}{60} + \frac{10^2}{100}}} = 1.39 < 1.96 @ 5\%$$

$\therefore$ z is accepted

3) The mean height of 50 male students who showed above average participation in college athletics was 68.2 inches with standard deviation of 2.5 inches. While 50 male students who showed no interest in such participation has mean height of 67.5 inches with standard deviation of 2.8 inches. ⓐ Test the hypothesis that male students who participate in college athletics are taller than other male students.

(i) $n_1 = 50$ $\quad \bar{x}_1 = 68.2$ $\quad s_1 = 2.5$

$n_2 = 50$ $\quad \bar{x}_2 = 67.5$ $\quad s_2 = 2.8$

H_0: There is no diff in the height of male students $(\bar{x}_1 = \bar{x}_2)$

H_1: The height is greater (male who participated in athlets) than other male students who showed no interest. $(\bar{x}_1 > \bar{x}_2)$ (right tailed test)

ⓑ By how much should the sample size of each of 2 groups be increased in order that the observed difference of 0.7 inches in the mean height be significant at 5% level of significance.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 1.318 = 1.32 < 1.64$$ at 5% level of significance (right tailed test)

H_0 is accepted.

$z > 1.645$

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{z_1^2}{n_1} + \frac{z_2^2}{n_2}}} > 1.645$$ (insted of n_1 and n_2 take as n we want one sample size n)

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{z_1^2/n + z_2^2/n}} > 1.645$$

The diff b/n the mean height of 2 groups, each of size n will be significant if z is greater than or equal to 1.645

$$\frac{0.7}{\sqrt{\frac{(2.5)^2+(2.8)^2}{n}}} \geq 1.645 \qquad \frac{0.7}{\sqrt{\frac{1}{\sqrt{n}}}\sqrt{(2.5)^2+(2.8)^2}} \geq 1.645$$

$$\frac{1}{\sqrt{n}} \geq \frac{0.7}{\frac{\sqrt{14.09}}{1.645}} \qquad \frac{1}{\sqrt{n}} \leq \frac{0.7}{\frac{\sqrt{14.09}}{1.645}}$$

$$\frac{1}{\sqrt{n}} \geq \frac{0.7}{6.174} \qquad \frac{1}{\sqrt{\eta}} \leq 0.43$$

0.012

77.8

$\simeq 78$

$$\frac{1}{n} =$$

$$n \geq 77.79 \qquad n = \left(\frac{0.7}{\frac{\sqrt{14.09}}{1.645}}\right)^2$$

$$n \simeq 78 \qquad \frac{1}{n} = 0.094$$

The increase in sample size so that it would have increased is, Hence the sample size of each of 2 the groups should be increased by 78-50 i.e 28. in order that the diff b/n the mean heights of 2 groups is significant.

10M

Test of Significance for the diff of Standard deviation.

If s_1 and s_2 are standard deviation of 2 independent samples then under the null hypothesis that the sample standard deviation don't differ significantly.

$$z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

σ_1 and σ_2 are S.D of population.

Note: If population standard deviation is not known then

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

s_1 and s_2 are ~~variance of~~ S.D of sample.

① 10M

① Random samples drawn from 2 countries gave the following data relating to the height of adults.

10M

	Country A	Country B
Mean height (inches)	67.42	67.25
standard deviation	2.58	2.50
No. of members	1000	1200

(i) Is the diff b/n means is significant ?

(ii) Is the diff b/n S.D is significant ?

(i) $\bar{x}_1 = 67.42$ $\bar{x}_2 = 67.25$

$s_1 = 2.58$ $s_2 = 2.50$

$n_1 = 1000$ $n_2 = 1200$

Ho : there is no diff b/n means

H1 : there is a diff b/n means (two tailed test)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{67.42 - 67.25}{\sqrt{\frac{(2.58)^2}{1000} + \frac{(2.50)^2}{1200}}}$$

$z = 1.56 < 1.966$ at 5% level of significance

Ho is accepted.

(ii) Ho : there is no diff b/n S.D's

H1 : there is a diff b/n S.D's (two tailed test)

$$z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} = 1.03 < 1.966$$ at 5% level of significance

Ho is accepted.