

UNIT-IV

Estimation & Tests of Hypotheses.

Short Answer Questions

1) Define Null Hypothesis and Alternative Hypothesis

Sol:- Null Hypothesis :- A definite statement about the population parameter for applying the test of significance is called null hypothesis, which is usually a hypothesis of no difference and is denoted by H_0

Alternative Hypothesis :- Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis and is denoted by H_1

2) Define Type I error and Type II error

Type I error :- if we reject the null hypothesis when it is true, and it is also known as producer risk

$$P[\text{reject } H_0 \text{ when it is true}] = \alpha$$

Type II error :- if we accept the null hypothesis when it is wrong, and it is also known as consumer risk.

$$P[\text{accept } H_0 \text{ when it is wrong}] = \beta$$

- ③ if a sample number is '500' and The S.D is '15'
 find maximum error with 95% confidence

Sol:-

$$\text{Maximum error } E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$z_{\alpha/2} = 1.96 \text{ at 95%}$$

$$\sigma = 15$$

$$n = 500$$

$$E = 1.96 \frac{15}{\sqrt{500}}$$

$$E = 1.31$$

- ④ The mean & S.D. of population are 11795 & 14504
 respectively if $n=50$, find 95% confidence
 Interval for the mean

Sol:-

$$\text{Interval} = \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$$\bar{x} = 11795 = 14504$$

$$= \left(11795 - (1.96) \frac{14504}{\sqrt{50}}, 11795 + (1.96) \frac{14504}{\sqrt{50}} \right)$$

$$= (7899.42, 15690.57)$$

- ⑤ if we can assert with 95% that The maximum error
 is 0.05 and $\rho=0.2$ find sample size

Sol:- Max error $E = z_{\alpha/2} \sqrt{\frac{pq}{n}}$ $p=0.2$ $q=1-p$
 $= 1-0.2$
 $= 0.8$

$$0.05 = 1.96 \sqrt{\frac{(0.2)(0.8)}{n}}$$

$$n = \frac{(0.2)(0.8) \times (1.96)^2}{(0.05)^2} = 246$$

⑥ A Random Sample of size '100' has a S.D of '5'. What can u say about the max error with 95% Confidence

Sol: Given That $n=100$ $\sigma=5$

$$Z_{\alpha/2} = 1.96 \text{ (95% Confidence)}$$

$$\text{Max error } E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 1.96 \frac{5}{\sqrt{100}} = 0.98$$

① a) Construct 95% Confidence Interval for True proportion of Computer literacy if 47 out of 150 persons from rural areas are computer literate

$$\text{proportion } p = \frac{47}{150} = 0.313 \quad Q = 1 - p \\ = 1 - 0.313 \\ = 0.687$$

Confidence Interval

$$(p - Z_{\alpha/2} \sqrt{\frac{pq}{n}}, p + Z_{\alpha/2} \sqrt{\frac{pq}{n}})$$

$$(0.313 - 1.96 \sqrt{\frac{0.313 \cdot 0.687}{150}}, 0.313 + 1.96 \sqrt{\frac{0.313 \cdot 0.687}{150}})$$

$$(0.313 - 0.0742, 0.313 + 0.0742)$$

$$(0.2388, 0.3872)$$

b) A sample of size 9 was taken from a population gave $s^2 = 10.9$, $\bar{x} = 15.8$ obtain 99% Confidence Interval for μ

$$(\bar{x} \pm z_{\alpha/2} s/\sqrt{n})$$

$$(15.8 - 2.58(3.3/\sqrt{9}), 15.8 + 2.58(3.3/\sqrt{9}))$$

$$(12.96, 18.6)$$

Critical Value (Significant values): The value of test statistic separates a rejection region and the acceptance region is called the Critical Value.

Critical value of Z	Level of Significance (α)		
	1%.	5%.	10%.
Two-tailed	2.58	1.96	1.64
Right-tailed	2.33	1.64	1.28
Left-tailed	-2.33	-1.64	-1.28

Long Answer Questions

2)

The mean height of 90 students in a class is 180cm test 90% level whether the sample has been drawn from a population mean is 170 cm and standard deviation 35.

Sol:- Given data. Sample mean $\bar{x} = 180$

Sample size $n = 90$

Population mean $\mu = 170$

Population S.D $\sigma = 35$

Null Hypothesis: H_0 :- The sample has been drawn from a population ($\bar{x} = \mu$)

Alternative Hypothesis H_1 :- The sample has not drawn from a population ($\bar{x} \neq \mu$) (Two tailed)

Test Statistic:

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

$$= \frac{180 - 170}{35 / \sqrt{90}} = \frac{10 \times 9.48}{35} = 2.71$$

$$Z_{cal} =$$

level of significance: $\alpha = 10\%$

$$Z_{tab} = 1.64$$

Conclusion: $|Z_{cal}| > Z_{tab}$ ($|2.71| > 1.64$)

So Alternative hypothesis accepted

The sample is not drawn from same population

③ The means of two large samples of size 1000 & 2000 member are 67.5 Inches & 68.0 Inches
 Can the sample be regarded as drawn from the same population of S.D 2.5 Inches.

Sol:- Given data

first Sample mean $\bar{x}_1 = 67.5$

Second Sample mean $\bar{x}_2 = 68$

pop S.D $= \sigma_1 = \sigma_2 = \sigma = 2.5$

$n_1 = 1000$ $n_2 = 2000$

Null Hypothesis H_0 : Two samples drawn from the same population i.e. ($\mu_1 = \mu_2$)

Alternative Hypothesis H_1 : Two samples not drawn from the same population i.e. ($\mu_1 \neq \mu_2$)

Test Statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{67.5 - 68}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}} = -5.16$$

level of Significance

$\alpha = 5\%$

$$|Z| > Z_{tab} \quad (| -5.16 | > 1.96)$$

Conclusion: Alternative hypothesis accepted

The Samples are drawn from the same population

4) A manufacturer claimed that at least 98% of the steel pipes which she supplied to a factory conformed to specifications. ~~test his claim at a significance level 0.05~~. An examination

$$\text{population proportion } P = \frac{98}{100} = 0.98$$

of 500 pieces of pipe revealed that 30 were defective. Test this claim at a significance level of 0.05

Sol:-

$$\text{population proportion } P = \frac{98}{100} = 0.98$$

500 pipes 30 are defective so non defective pipes = $500 - 30 = 470$

$$\text{Sample proportion of non defective pipe } p = \frac{x_1}{n_1} = \frac{470}{500} = 0.94$$

$$n_1 = 500$$

Null Hypothesis H_0 : There is no diff b/w

Sample proportion with population proportion
i.e. ($p=P$)

Alternative Hypothesis H_1 : There is diff b/w

Sample proportion with population proportion
i.e. ($p \neq P$)

Test Statistic:- $Z = \frac{P - P}{\sqrt{\frac{PQ}{n}}}$

$$\begin{aligned} P &= 0.98 & Q &= 1 - P \\ &&&= 1 - 0.98 \\ &&&= 0.02 \end{aligned}$$

$$Z_{\text{cal}} = \frac{0.94 - 0.98}{\sqrt{\frac{(0.98)(0.02)}{500}}} = -2.02$$

Level of Significance: $\alpha = 0.05$
 $= 5\%$

$$Z_{\text{tab}} = 1.96$$

Conclusion:- $|Z_{\text{cal}}| > Z_{\text{tab}}$ ($-2.02 > 1.96$)

Alternative Hypothesis accepted
There is diff b/w Sample proportion with pop proportion.

(5) In a certain city 125 men in a sample of 500 were found to be smokers. In another city the number of smokers was 375 in a random sample of 1000. Does this indicate that there is a greater population smokers in the second city than the first city?

Sol:-

$$\text{Smokers proportion in first city } p_1 = \frac{x_1}{n_1} = \frac{125}{500} = 0.25$$

$$\text{Smokers proportion in second city } p_2 = \frac{x_2}{n_2} = \frac{375}{1000} = 0.375$$

Null Hypothesis H_0 : There is no diff b/w Two cities smokers proportion ($p_1 = p_2$)

Alternative Hypothesis H_1 : There is a greater population smokers in the second city than the first city

$$(p_2 > p_1) \quad (\text{Right tailed})$$

Test Statistic

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

$$p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{125 + 375}{500 + 1000} = \frac{500}{1500} = 0.33$$

$$\begin{aligned}q &= 1 - p \\&= 1 - 0.33 \\&= 0.67\end{aligned}$$

$$z = \frac{(0.25) - (0.375)}{\sqrt{(0.33)(0.67)\left(\frac{1}{500} + \frac{1}{1000}\right)}} = 1.506$$

Level of significance: $\alpha = 5\%$ (Right-tailed)

$$z_{tab} = 1.64$$

Conclusion: $|z_{cal}| < z_{tab}$ ($|1.506| < 1.64$)

Null Hypothesis accepted

There is a greater population smokers in the
Second City than the first City

⑥ The owner of a machine shop must decide which of two snack vending machines to install in this shop. If each is tested '250' times the first machine fails to work 13 times and second machine fails to work 7 times. Test at 0.05 level of significance whether the difference b/w the corresponding sample proportions is significant.

Sol:-

first sample size $n_1 = 250$

Second sample size $n_2 = 250$

first sample proportion to fails machine $p_1 = \frac{x_1}{n_1} = \frac{13}{250} = 0.052$

Second sample proportion

$$p_2 = \frac{x_2}{n_2} = \frac{7}{250} = 0.028$$

Null Hypothesis ~~H₀~~ :- $p_1 = p_2$

Alternative Hypothesis: H_1 :- $p_1 \neq p_2$

Test Statistic $Z = \frac{p_1 - p_2}{\sqrt{pq(\frac{1}{n_1} + \frac{1}{n_2})}}$

$$p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{13 + 7}{250 + 250} = \frac{20}{500} = 0.04$$

$$q = 1 - 0.04 = 0.96$$

$$Z_0 = \frac{0.052 - 0.028}{\sqrt{(0.04)(0.96) \left(\frac{1}{250} + \frac{1}{250} \right)}} = 1.38$$

$$Z_{01} = 1.38$$

Level of Significance: - $\alpha = 5\%$

$$Z_{tab} = 1.96$$

Conclusion: $|Z_{01}| < Z_{tab}$ ($1.38 < 1.96$)

Null Hypothesis accepted

There is no diff b/w Sample proportions

UNIT - IV

parameters:- The population measurements of observations are called parameters. It is denoted by θ .

Ex:- μ, σ^2, N

Statistics:- The statistical measurements of observations in a sample are called statistics. It is denoted by $\hat{\theta}$.

Ex:- \bar{x}, s^2, n

Estimate:- It is a statement made to find unknown population parameters using sample statistics.

Estimator:- The procedure or rule to determine unknown population parameters using sample statistics, is called an estimator.

Ex:- i) $E(\bar{x}) = \mu$

ii) $E(s^2) = \sigma^2$

i) \bar{x}, s^2 are estimators.

Types of estimations:-

1) point estimation 2) interval estimation

1) point estimation:-

If an estimate of the population parameter is given by a single value then that estimate is called point estimation of parameter.

Ex:- The mean height of a student in a

college is 165 cm

Unbiased Estimator :-

A Sample statistic $\hat{\theta}$ is said to be unbiased estimator of population parameter θ . If $E[\hat{\theta}] = \theta$.

$$\text{Ex:- 1) } E(\bar{x}) = \mu \\ 2) E(s^2) = \sigma^2$$

Biased estimator :-

A Sample statistic $\hat{\theta}$ is said to be biased estimator of population parameter θ if $E[\hat{\theta}] \neq \theta$.

Theorem :-

Show that sample mean \bar{x} is unbiased estimator of population mean μ .

Proof :-

Let $x_1, x_2, x_3, \dots, x_n$ be a random sample from a population with mean μ .

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$E[\bar{x}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right]$$

We know that $E[x_i] = \mu$, $\forall i$

$$\begin{aligned} E[\bar{x}] &= \frac{1}{n} E[x_1 + x_2 + x_3 + \dots + x_n] \\ &= \frac{1}{n} [E(x_1) + E(x_2) + \dots + E(x_n)] \\ &= \frac{1}{n} [\mu + \mu + \dots + \mu] \\ &= \frac{1}{n} [n\mu] \end{aligned}$$

$$\therefore E[\bar{x}] = \mu.$$

Theorem:

Show that sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ is unbiased estimator of population mean σ^2 i.e., $E(s^2) = \sigma^2$.

Proof

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

Consider

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n [(x_i - \mu) - (\bar{x} - \mu)]^2 \\ &= \sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu) \sum_{i=1}^n (x_i - \mu) + \sum_{i=1}^n (\bar{x} - \mu)^2 \\ &\stackrel{1}{=} \sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu) \left(\sum_{i=1}^n x_i - \sum_{i=1}^n \mu \right) + n(\bar{x} - \mu)^2. \\ \stackrel{1}{=} \sum_{i=1}^n (x_i - \mu)^2 &- 2(\bar{x} - \mu)(n\bar{x} - n\mu) + n(\bar{x} - \mu)^2 \\ &\stackrel{2}{=} \sum_{i=1}^n (x_i - \mu)^2 - 2n(\bar{x} - \mu)^2 + n(\bar{x} - \mu)^2 \rightarrow 0. \end{aligned}$$

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

$$E(S^2) = \frac{1}{n-1} E[\sum_{i=1}^n (x_i - \bar{x})^2]$$

$$E(S^2) = \frac{1}{n-1} E[\sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^n E[(x_i - \mu)^2] - n E[\bar{x} - \mu]^2 \right].$$

$$\stackrel{3}{=} \frac{1}{n-1} \left[\sum_{i=1}^n \sigma^2 - n \cdot \sigma_x^2 \right]$$

$$\stackrel{4}{=} \frac{1}{n-1} \left[n\sigma^2 - n \cdot \sigma_x^2 \right] \quad \left[\because \sigma_x^2 = \frac{\sigma^2}{n} \right]$$

$$= \frac{1}{n-1} (n-1) \cdot \sigma^2$$

$$\therefore E(S^2) = \underline{\underline{\sigma^2}}$$

Note:-

If $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ then S^2 is biased estimator of σ^2 i.e., $E(S^2) \neq \sigma^2$.

3) The population parameter θ can consist many estimators.

Most efficient parameter estimator :-

If $\hat{\theta}_1, \hat{\theta}_2$ are two unbiased estimators of some population parameter ' θ '. If $V(\hat{\theta}_1) \leq V(\hat{\theta}_2)$ then $\hat{\theta}_1$ is most efficient estimator of ' θ ' than $\hat{\theta}_2$.
(or)

All possible unbiased estimators of parameter θ , the one with smallest variance is called most efficient parameter estimator of θ .

Good Estimator:-

A good estimator is the one which is as close to the true value of parameter as possible.
(or)

It satisfies following properties :-

i) Consistency :-

Estimator $\hat{\theta}_n$ converges to θ , if $n \rightarrow \infty$ then $\hat{\theta}_n$ is consistent.

ii) Unbiasedness :-

The estimator $\hat{\theta}$ is called unbiased if $E[\hat{\theta}] = \theta$.

iii) Efficiency :-

For all possible unbiased estimators of θ , the one with smallest variance is called most efficient estimation.

iv) Sufficiency :-

An estimator is said to be sufficient for a parameter if it contains all the information in the sample regarding parameter.

Interval estimation:-

If an estimate of population parameter is given by 2 different values, in which the parameter lie, then the estimate is called interval estimation.

Ex:- The mean height of college student is 165 ± 2 cm, i.e., $(163, 167)$.

The population parameter θ is of the form $\hat{\theta}_L$ if $\hat{\theta}_L < \theta < \hat{\theta}_U$ and $P(\hat{\theta}_L < \theta < \hat{\theta}_U) = (1-\alpha)$.

Then $(1-\alpha)$ is called confidence coefficient, and $(1-\alpha)100\%$ is called confidence percentage.

The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$ is called confidence interval with $(1-\alpha)100\%$ confidence.

Maximum Error & Confidence Interval of μ (Large Sample)

The difference

The modulus of difference b/w sample mean and population mean is called maximum error. It is denoted by ' E '.

$$\text{i.e., } |\bar{x} - \mu| = \text{max. error} = E$$

$$E = |\bar{x} - \mu| = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

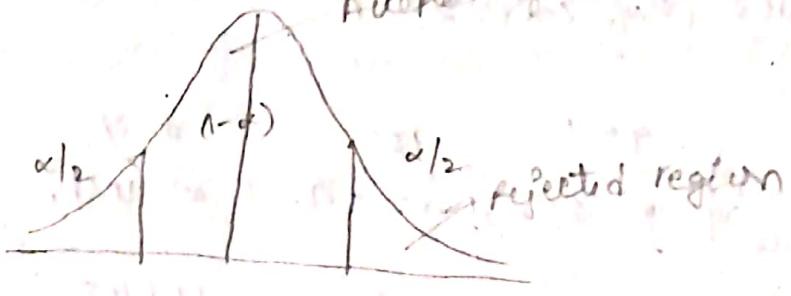
$$\bar{x} - \mu = \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\mu = \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \textcircled{*}$$

$$\bar{x} - E \leq \mu \leq \bar{x} + E$$

$$P\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$



The interval $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ is called confidence interval for population mean μ with $(1-\alpha)100\%$ confidence.

The confidence $(1-\alpha)$ is given by 99%, (or) 95%, (or) 90%.

Sample size (n):

$$e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{e}$$

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{e} \right)^2$$

Maximum error & confidence Interval of μ (small sample):

Maximum error is e (as σ is unknown.)

$$e = |\bar{x} - \mu| = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\bar{x} - \mu = \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\mu = \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \rightarrow \textcircled{*}$$

$$\bar{x} - e < \mu < \bar{x} + e$$

Sample size (n):

$$e = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\sqrt{n} = \frac{t_{\alpha/2} \cdot s}{e}$$

$$n = \left(\frac{t_{\alpha/2} \cdot s}{e} \right)^2$$

Table Values ($z_{\alpha/2}$ or z_α)

$(1-\alpha)$	99%	95%	90%
level of α significance	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
two-tailed test	$z_{\alpha/2} = 2.58$	1.96	1.645
right tailed test	$z_\alpha = 2.33$	1.645	1.28
left tailed test	$z_\alpha = -2.83$	-1.645	-1.28

$z_{\alpha/2}$ is z-value leaving area $\alpha/2$ to the right of normal curve.

- D) Find 95% confidence interval for mean of Normal distribution with variance 0.25 using a sample of $n=100$ values with mean 212.3.

Given $n=100$

mean of sample (\bar{x}) = 212.3

$$S.D = \sigma = \sqrt{0.25} = 0.5$$

$$(1-\alpha) = 95\%, z_{\alpha/2} = 1.96$$

$$\text{Margin of Error } (E) = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$= (1.96) \cdot \frac{0.5}{\sqrt{100}}$$

$$= 0.098$$

Confidence Interval for N 95% :

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$212.3 - 0.098 < \mu < 212.3 + 0.098$$

$$212.202 < \mu < 212.398$$

2) A Random sample of size 100 has a std. deviation of 5, what will you say about Max. error with 95% confidence.

Given Sample size (n) = 100

$$S.D = \sigma = 5$$

$$\text{Now } (1-\alpha) = 95\%, z_{\alpha/2} = 1.96$$

$$\text{Max. Error } (\epsilon) = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$= (1.96) \frac{5}{\sqrt{100}} \Rightarrow 1.96 \left(\frac{1}{2}\right) = 0.98$$

3) A Random sample of size 100 is taken from a population with $\sigma = 5.1$, Given that sample mean $\bar{x} = 21.6$. Construct 95% confidence limits for population mean μ .

Given sample size = 100

$$\text{Mean } (\bar{x}) = 21.6$$

$$\sigma = 5.1$$

$$(1-\alpha) = 95\%, z_{\alpha/2} = 1.96$$

$$\text{Max. Error } (\epsilon) = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$= (1.96) \frac{5.1}{\sqrt{100}}$$

$$= 0.9996$$

Confidence Interval for μ is

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - \epsilon < \mu < \bar{x} + \epsilon$$

$$21.6 - 0.9996 < \mu < 21.6 + 0.9996$$

\Rightarrow

4) Assume that $\sigma = 20$, How large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3 points.

Given Max. error (ϵ) = 3.

$$\sigma = 20$$

$$(1 - \alpha) = 0.95 = 95\%$$

$$z_{\alpha/2} = 1.96$$

$$\epsilon = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$n = \left(\frac{z_{\alpha/2} \sigma}{\epsilon} \right)^2$$

$$= \left[\frac{(1.96)20}{3} \right]^2 = (13.06)^2 = 26.12$$

5) In a study of Automobile Insurance a random sample of 80 body repair cost had mean of ₹ 472.36 and S.D of ₹ 62.85. If \bar{x} is used as a point estimate to the true average repair cost with what confidence, we can assert that the max. error does not exceed ₹ 10/-

Max. error (ϵ) = 10/-

Sample size (n) = 80

Sample Mean (\bar{x}) = 472.36

$$\sigma = 62.85$$

$$z_{\alpha/2} = ?$$

$$\epsilon = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$10 = z_{\alpha/2} \cdot \frac{62.85}{\sqrt{80}}$$

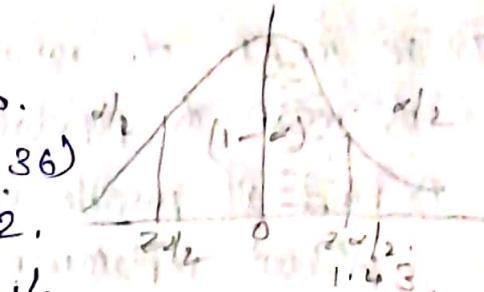
$$z_{\alpha/2} = \frac{10 \times \sqrt{80}}{62.85}$$

$$= 1.43$$

from Tables,

$$P(0 < Z < 1.43) = \frac{1-\alpha}{2} = 0.4236.$$

$$\begin{aligned}\text{Confidence} &= (1-\alpha) = 2 \times 0.4236 \\ &= 0.8472 \\ &\Rightarrow 84.72\%.\end{aligned}$$



- 6) The efficiency expert of computer company tested 40 engineers to estimate the average time it takes to assemble a certain computer component, getting mean of 12.73 mins and S.D 2.06 mins
- i) If \bar{x} is 12.73 mins is used as point estimate of Actual average time required to perform the task, find max. error with 99%. confidence.
 - ii) Construct 98% confidence interval for true average time it takes to do the job.
 - iii) With what confidence can be assert that the sample mean does not differ from the true mean by more than 30 secs.

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

$$0.5 = z_{\alpha/2} \cdot \frac{2.06}{\sqrt{40}}$$

$$z_{\alpha/2} = \frac{0.5(\sqrt{40})}{2.06} = 1.585.$$

From tables.

$$P(0 < Z < 1.585) = \frac{1-\alpha}{2} = 0.4382$$

$$\begin{aligned}\text{Confidence} &= (1-\alpha) = 2 \times 0.4382 \\ &= 0.8764 \\ &\Rightarrow 87.64\%.\end{aligned}$$

7) The mean of random sample is unbiased estimate of mean of population 3, 6, 9, 15, 27.

i) Find all possible samples of size '3' that can be taken without replacement from finite population.

ii) Calculate the mean of each of samples in ① and assigning each sample a probability of $\frac{1}{10}$. Verify that mean of these \bar{x} 's is equal to 12. and verify $E(\bar{x}) = \mu$.

i) Population = {3, 6, 9, 15, 27}

Population size (N) = 5

Sample size (n) = 3.

No. of samples without replacement = $N^C_n = 5^C_3$

$$= \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10.$$

Sample S = { (3, 6, 9), (3, 6, 15), (3, 6, 27), (3, 9, 15),
(3, 9, 27), (3, 15, 27), (6, 9, 15), (6, 9, 27), (9, 15, 27)
(6, 15, 27) }

Sample Mean = { 6, 8, 12, 9, 18, 15, 10, 14, 17, 16 },

Population Mean (μ) = $\frac{3+6+9+15+27}{5} = \frac{60}{5}$,

$$\mu = 12.$$

Assigning prob. $\frac{1}{10}$ to each sample mean.

$$\bar{x} \quad 6 \quad 8 \quad 9 \quad 10 \quad 12 \quad 13 \quad 14 \quad 15 \quad 16 \quad 17.$$

$$P(\bar{x}) \quad \frac{1}{10} \quad \frac{1}{10} \quad \dots$$

$$E(\bar{x}) = \sum \bar{x}_i P(\bar{x}_i)$$

$$= 6\left(\frac{1}{10}\right) + 8\left(\frac{1}{10}\right) + \dots + 17\left(\frac{1}{10}\right)$$

$$\Rightarrow \frac{1}{10}[6 + 8 + \dots + 17]$$

$$\Rightarrow \frac{120}{10}$$

$$\Rightarrow 12$$

8) Suppose that we observe a R.V having binomial distribution and getting x success in n trials.

i) Show that \bar{x} is an unbiased estimate of binomial parameter ' p '.

ii) Show that $\frac{x+1}{n+2}$ is not unbiased estimate of binomial parameter ' p '.

9) Find 95% confidence limit for the mean of normal distributed pop. from which the sample was taken {5, 17, 10, 18, 16, 9, 7, 11, 13, 14}.

8) i) If X is R.V having Binomial dist. mean(μ) = $E(X)$ = np

$$E\left[\frac{\bar{x}}{n}\right] = \frac{E(x)}{n} = \frac{np}{n} = p,$$

$$\text{ii) } E\left[\frac{x+1}{n+2}\right] = \frac{1}{n+2} E[x+1] \Rightarrow \frac{1}{n+2} [E(x)+1].$$

$$\Rightarrow \frac{np+1}{n+2} \neq p.$$

$\therefore \frac{x+1}{n+2}$ is not unbiased estimator of p .

9) Confidence limits

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}. \quad [\sigma \text{ is unknown}]$$

Sample Mean Variance

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

$$\text{Sample} = \{5, 17, 10, 18, 16, 9, 7, 11, 13, 14\}$$

$$\text{Sample Mean}(\bar{x}) = \frac{5+17+10+18+16+9+7+11+13+14}{10}$$

$$= \frac{130}{10}$$

$$\Rightarrow 13$$

$$\begin{aligned}
 S^2 &= \frac{1}{10-1} \left\{ (15-18)^2 + (17-13)^2 + \dots + (14-18)^2 \right\} \\
 &= \frac{1}{9} [4+16+9+25+9+16+8+6+4+0+1] \\
 &= \frac{120}{9} \Rightarrow \frac{40}{3}.
 \end{aligned}$$

$$\boxed{S = 3.65}$$

$$(1-\alpha)95\%, \quad Z_{\alpha/2} = 1.96, \quad n = 10.$$

confidence interval is $13 \pm (1.96) \frac{(3.65)}{\sqrt{10}}$

$$\Rightarrow 13 \pm 2.262$$

$$\Rightarrow (13 - 2.262, 13 + 2.262)$$

$$\Rightarrow (10.74, 15.26).$$

Hypothesis:-

The value of parameter whether to accept or reject a statement about parameter then that statement is called hypothesis.

- Ex:- i) The majority of men in the city are smokers
 ii) A drunk chemist is to decide whether a new drug is really effective in curing a disease.

Test of Hypothesis:-

The procedure which enables us to decide on the basis of sample result whether a hypothesis is true or not, is called test of hypothesis.

* * Null Hypothesis:-

A hypothesis of no difference is called null hypothesis.

It is denoted by H_0 and defined as

$$H_0 : \mu = \mu_0$$

Alternative Hypothesis :-

The opposite statement of null hypothesis is called alternative hypothesis. It is denoted by H_1 .
Defined as

- i) $H_1 : \mu \neq \mu_0 (\mu > \mu_0 \text{ or } \mu < \mu_0)$ (Two-tail d).
- ii) $H_1 : \mu > \mu_0$ (Right tailed)
- iii) $H_1 : \mu < \mu_0$ (Left tailed).

Level of Significance (α) :-

It is the confidence with which we reject or accept null hypothesis H_0 . 5% of level of significance in a test procedure indicates that there are 5 cases in 100 that reject null hypothesis.



Types of Errors :-

- 1) Type-I Error :- (Reject H_0 when it is True).

If the null hypothesis H_0 is true but it is rejected by test procedure. Then the error made is called Type-I error (or) α -error.

$$P(\text{Reject } H_0 \text{ when it is true}) = \alpha - \text{Error}$$

It is also called procedure's risk.

- 2) Type-II Error :- (Accept H_0 when it is false).

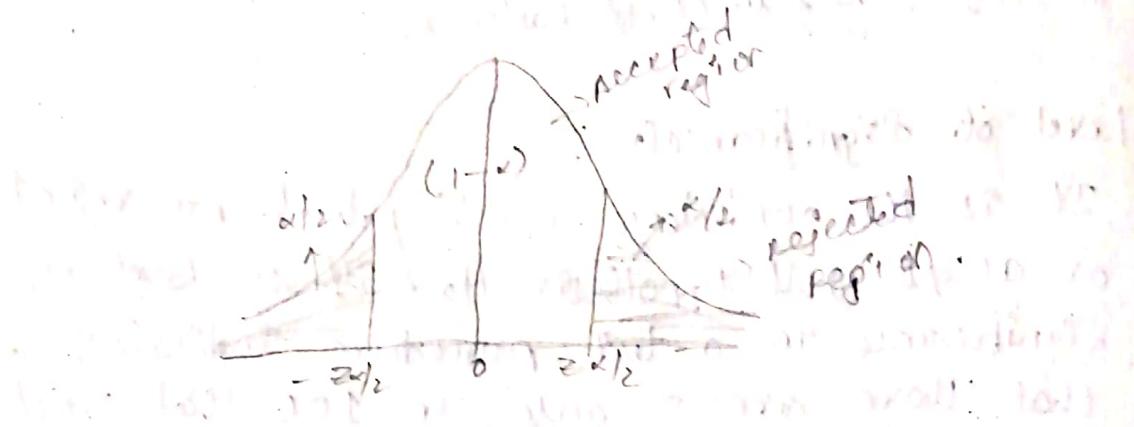
If the null hypothesis H_0 is false but it is accepted by test procedure, Then the error made is called Type-II error (or) β -error.

It is also called consumer risk.

Critical Region :-

A region corresponding to statistic t^* in the sample space S which leads to the rejection of H_0 is called critical region (or) Rejection region.

The region which leads to acceptance of H_0 is called acceptance region.



Critical values :-

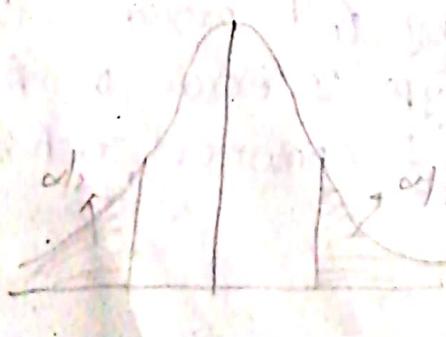
The value of test statistic which separates the critical region and acceptance region is called critical value (or) significant value.

* Two tailed test & one-tailed test

i) Two tailed test :-

If the alternative hypothesis is of the form

$H_1 : \mu \neq \mu_0 (\mu > \mu_0 \text{ or } \mu < \mu_0)$ in a test of hypothesis then the test process is called two-tailed test.



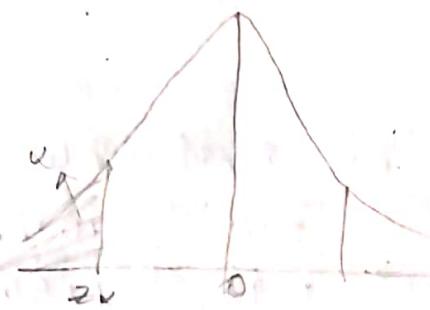
i) Right tailed test:-

If the alternative hypothesis of the form $H_1: \mu > \mu_0$ in a test of hypothesis then the test is called Right Tailed test.



ii) Left Tailed Test:-

If the alternative hypothesis of the form $H_1: \mu < \mu_0$ in a test of Hypothesis then it is called Left tailed test.



Process of testing of Hypothesis:-

Step 1) Null Hypothesis:- $H_0: \mu = \mu_0$

A hypothesis with no difference is called null hypothesis.

Step 2) Alternative hypothesis:-

It is defined as one of the following

i) $H_1: \mu \neq \mu_0$ ($\mu > \mu_0$ or $\mu < \mu_0$)

ii) $H_1: \mu > \mu_0$

iii) $H_1: \mu < \mu_0$.

Step 3) Level of Significance :-

usually it is 1%. (or) 5%. (or) 10%.

Table value = $z_{\alpha/2}$ (or) z_{α}

Step 4) Test statistic :-

$$z = \frac{t - E(t)}{S.E.(t)}$$

where t = statistic

Step 5) Conclusion:-

If $|z| < z_{\alpha/2}$ then H_0 is true (or) H_0 is accepted.

If $|z| > z_{\alpha/2}$ then H_0 is false (or) H_0 is rejected.
so, the H_1 is true.

Critical values :-

(a) Level of significance				
	1%	5%	10%	2%
Two-tailed	$ z_{\alpha/2} = 1.96$	1.96	1.645	1.33
Right-tailed	$z_{\alpha} = 2.33$	1.645	1.28	
Left-tailed	$z_{\alpha} = -2.33$	-1.645	-1.28	

Large sample :-

Test of significance of sample means :-

A Random sample of size n has the sample mean \bar{x} , which is taken from the population with mean μ and S.D σ and the population mean μ has specified value μ_0 .

i) Null Hypothesis :-

$$H_0: \mu = \mu_0$$

g) Alternative Hypothesis :-

i) $H_1 : \mu \neq \mu_0$

ii) $H_1 : \mu > \mu_0$

iii) $H_1 : \mu < \mu_0$

5) Level of Significance :-

Table Value = $Z_{\alpha/2}$.

6) Test statistic :-

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

7) Conclusion :-

If $|Z| < Z_{\alpha/2}$ H_0 True

If $|Z| > Z_{\alpha/2}$ H_0 false

i) Sample of 400 items are taken from a population whose std. deviation 10. The mean of sample is 40. Test whether the sample has come from a population with mean 38. Also calculate 95% confidence limits.

ii) Null Hypotheses :- $\mu = 38$

iii) Alternative Hypotheses :- $H_0 : \mu \neq 38$

iv) Level of Significance :-

$$(1 - \alpha) = 95\% \Rightarrow \alpha = 5\%$$

Table value = $Z_{\alpha/2} = 1.96$ (Two-tailed)

v) Test statistic :-

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

vi) Conclusion :-

$$z = 4, Z_{\alpha/2} = 1.96$$

$$|z| > z_{\alpha/2}$$

$\therefore H_0$ is false

$\therefore H_1$ is true

$$\mu \neq 88.4$$

confidence limits with 95% for n is

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (\text{or}) \quad \bar{x} \pm E$$

$$E = (1.96) \frac{10}{\sqrt{400}} \Rightarrow (1.96) \frac{10}{20}$$

$$\Rightarrow \frac{1.96}{2} \Rightarrow 0.98$$

confidence limits.

$$(\bar{x} - E, \bar{x} + E)$$

$$(40 - 0.98, 40 + 0.98)$$

$$(39.02, 40.98)$$

Q) A sample of 900 items has mean 3.4 and S.D 2.61.

If this sample has been taken from a large population of mean 3.25 and S.D 2.61. If the population is normal find 95% confidence limits for mean.

3) An Ambulance service claims that it takes on the average less than 10 mins to reach its destination in emergency calls. A sample of 36 calls has mean of 11 mins. and variance 16 mins. Test the claim at 5% level of significance

4) A sample of 64 students have mean weight 70 kgs. can this be regarded as a sample from population with mean weight 56 kgs and std-deviation 25 kgs.

Q) Given

$$\text{Sample size}(n) = 900$$

$$\text{Sample mean}(\bar{x}) = 3.4$$

$$\text{Population S.D}(\sigma) = 2.61$$

$$\text{Population mean}(\mu) = 3.25$$

v) Null Hypothesis :-

$$H_0: \mu = 3.25$$

ii) Alternative Hypothesis.

$$H_1: \mu \neq 3.25$$

Two tailed

iii) Level of significance :-

$$(1-\alpha) = 95\% \Rightarrow \alpha = 5\%$$

$$\text{Table value } Z_{\alpha/2} = 1.96$$

iv) Test statistic :-

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.25}{2.61/\sqrt{900}} = 1.924$$

v) Conclusion :-

$$Z = 1.924, Z_{\alpha/2} = 1.96$$

$$|Z| < Z_{\alpha/2}$$

H_0 is True

$$\therefore \mu = 3.25$$

Confidence limits :-

$$\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

(or)

$$\bar{x} \pm E$$

$$\text{Max. error} (E) = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$= (1.96) \frac{2.61}{\sqrt{900}}$$

$$E = 0.17$$

$$\text{confidence interval is } (\bar{x} - E, \bar{x} + E)$$

$$= (3.4 - 0.17, 3.4 + 0.17)$$

$$= (3.23, 3.57)$$

3) Given,

Sample size (n) = 86 balls.

Sample's mean (\bar{x}) = 11 mins.

Population s.d (σ) = 16 mins $\Rightarrow \sigma = 4$ mins.

Population mean (μ) = 10 mins.

i) Null Hypothesis :-

$$H_0 : \mu = 10 \text{ mins.}$$

ii) Alternative Hypothesis :-

$$H_1 : \mu < 10 \text{ mins.}$$

(let t-tailed)

iii) Level of Significance (α) :-

$$(1 - \alpha) = 95\% \Leftrightarrow \alpha = 5\%.$$

Table value $z_{\alpha} = -1.645$.

iv) Test statistic :-

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \Rightarrow \frac{11 - 10}{4 / \sqrt{86}} = 1.5.$$

v) Conclusion :-

$$Z = 1.5, z_{\alpha} = -1.645$$

$$|Z| = 1.5, |z_{\alpha}| = 1.645.$$

$$|Z| < |z_{\alpha}|$$

$\therefore H_0$ is True

$$\mu = 10 \text{ mins}$$

4) Given

Sample size (n) = 64

Sample mean (\bar{x}) = 70 kgs

Population s.d (σ) = 25 kgs

Population mean (μ) = 56 kgs.

i) Null Hypothesis :-

$$H_0 : \mu = 56.$$

i) Alternative Hypothesis :-

$$H_1: \mu \neq 56$$

Two tailed.

ii) Level of significance .

$$\alpha = 5\%$$

$$\text{Table value } Z_{\alpha/2} = 1.96.$$

iv) Test statistic :-

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow \frac{70 - 56}{55/\sqrt{64}}$$

$$Z \approx 4.48$$

v) Conclusion :-

$$Z = 4.48 \quad Z_{\alpha/2} = 1.96$$

$$|Z| > Z_{\alpha/2}$$

$\therefore H_0$ is false

$\therefore H_1$ is true

$$\mu \neq 56 \text{ kgs.}$$

Test of Equality of Two means (or) Test of
Significance difference b/w two means.

i) Null hypothesis

Let \bar{x} and \bar{y} be sample means of two independent large samples sizes n_1 and n_2 drawn from two populations having means μ_1 and μ_2 with the std-deviations σ_1 & σ_2 . To test whether the two population means are equal,

i) Null Hypothesis :-

$$H_0: \mu_1 = \mu_2 \text{ (or) } \mu_1 - \mu_2 = 0.$$

ii) Alternative Hypothesis :-

i) $H_1: \mu_1 \neq \mu_2 \text{ (or) } (\mu_1 > \mu_2 \text{ or } \mu_1 < \mu_2)$ (Two tailed)

ii) $H_1: \mu_1 > \mu_2$ (right tailed)

iii) $H_1: \mu_1 < \mu_2$ (left tailed)

iii) Level of significance :- (α)

$$\text{Table value} = z_{\alpha/2} \text{ (or)} z_{\alpha}$$

iv) Test statistic :-

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

If $\sigma_1 = \sigma_2 = \sigma$ (say)

$$Z = \frac{\bar{x} - \bar{y}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

v) Conclusion :-

If $|Z| < z_{\alpha/2}$, $\therefore H_0$ True

If $|Z| > z_{\alpha/2}$, $\therefore H_0$ False
 $\therefore H_1$ is True

1) The means of two large samples of size n_1 and n_2

$n_1 = 1000$ and 2000 members are 67.5 and 68 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches

Given,

Sample mean $\bar{x} = 67.5$ inches $\bar{y} = 68$ inches

sample size $n_1 = 1000$

$n_2 = 2000$

population S.D = 2.5 inches

i) Null type :-

$$H_0: \mu_1 = \mu_2$$

ii) Alternative Hypothesis :-

$$H_1: \mu_1 \neq \mu_2 \text{ (two-tailed)}$$

iii) Level of Significance (α) = 5%

$$\text{Table value} c = z_{\alpha/2} = 1.96$$

(v) Test statistic:-

$$\begin{aligned} z &= \frac{\bar{x} - \bar{y}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= \frac{67.5 - 68}{(2.5) \sqrt{\frac{1}{1000} + \frac{1}{2000}}} \Rightarrow -5.16 \end{aligned}$$

∴ Conclusion:-

$$z = -5.16 \quad |z|_2 = 1.96$$

$$|z| = 5.16 \quad |z|_2 = 1.96$$

$$|z| > |z|_2$$

∴ H_0 is false

∴ H_1 is True

$$\mu_1 \neq \mu_2$$

∴ Two samples has drawn from 2 different populations

2) A Researcher wants to know the intelligence of students in a school is selected 2 groups of students, in the first group there are 150 students having mean IQ of 75 with S.D 15, in the second group there are 850 students having mean IQ of 70 with S.D 20. Test whether the 2 groups of students have taken from same school.

3) The average marks scored by 32 boys is 72 with S.D 8, while that for 36 girls is 70 with S.D 6, Does this indicate that boys perform better than girls at 5% level of significance?

4) The mean height of 50 male students who participated in sports is 68.2 inches with S.D 2.5. The mean height of another 50 male students who have not participated in sports is 67.8 inches with S.D 2.8. Test the hypothesis the height of students who participated in sports is more than the students who have not participated in sports.

2) 1st group \Rightarrow Sample size (n_1) = 150
Sample mean (\bar{x}) = 75
Population S.D (σ_1) = 15

2nd group \Rightarrow Sample size (n_2) = 250
Sample mean (\bar{x}) = 70
Population S.D (σ_2) = 20.

i) Null Hypothesis:-

$$H_0: \mu_1 = \mu_2.$$

ii) Alternative Hypothesis.

$$H_1: \mu_1 \neq \mu_2.$$

iii) Level of Significance

$$\alpha = 5\%.$$

$$\text{Table value } z_{\alpha/2} = 1.96$$

iv) Test statistic:-

$$z = \frac{\bar{x} - \bar{A}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \Rightarrow \frac{75 - 70}{\sqrt{\frac{(15)^2}{150} + \frac{(20)^2}{250}}} \Rightarrow 2.83$$

v) Conclusion:-

$$z = 2.83, z_{\alpha/2} = 1.96$$

$$|z| > z_{\alpha/2}$$

$\therefore H_0$ is False (Rejected)

$\therefore H_1$ is True

$$\mu_1 \neq \mu_2$$

3) Boys:-

Sample size (n_1) = 32

Sample mean (\bar{x}) = 72

Population S.D (σ_1) = 8

Girls

Sample size (n_2) = 36

Sample mean (\bar{x}) = 70

Population S.D (σ_2) = 6.

i) Null Hypothesis :-

$$H_0 : \mu_1 = \mu_2$$

ii) Alternative Hypothesis :-

$$H_1 : \mu_1 > \mu_2$$

iii) Level of Significance (α)

$$\alpha = 5\% = 0.05$$

$$\text{Table value } z_{\alpha/2} = 1.645.$$

iv) Test Statistic :-

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \Rightarrow \frac{72 - 70}{\sqrt{\frac{(8)^2}{32} + \frac{6^2}{36}}} \Rightarrow 1.154$$

v) Conclusion :-

$$z = 1.154 \quad z_{\alpha/2} = 1.645$$

$$1.154 < 1.645$$

$\therefore H_0$ is TRUE (Accepted).

$$\mu_1 = \mu_2$$

- Both boys & girls perform equally

4) male in sports

Sample size (n_1) = 50

Sample mean (\bar{x}) = 68.2

Population S.D (σ_1) = 2.5

Male Prof in sports

Sample size (n_2) = 50

Sample mean (\bar{x}) = 67.8

Population S.D (σ_2) = 2.8

i) Null Hypothesis :-

$$H_0 : \mu_1 = \mu_2$$

ii) Alternative Hypothesis :-

$$H_1 : \mu_1 > \mu_2$$

iii) level of significance (α)

$$\alpha = 5\% = 0.05$$

$$\text{Table value } Z_{0.05} = 1.645$$

iv) Test statistic:-

$$Z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \Rightarrow \frac{68.2 - 67.2}{\sqrt{\frac{(2.5)^2}{50} + \frac{(0.8)^2}{50}}} \Rightarrow 1.88$$

v) Conclusion:-

$$Z = 1.88 \quad Z_{0.05} = 1.645$$

$$|Z| > Z_{0.05}$$

$\therefore H_0$ is False (Rejected)

$\therefore H_1$ is True.

$$\mu_1 > \mu_2$$

Test of significance of single proportion

Suppose a large random sample of size 'n' has a sample proportion ' \hat{p} ' which is taken from a population proportion ' p '.

proportion = $\frac{\text{No. of Success}}{\text{Total observations}} = \frac{x}{n}$

Sample proportion = $\hat{p} = \frac{x}{n}$

Population proportion = p .

To test the population proportion ' p ' which has specified value p_0 .

i) Null Hypothesis :-

$$H_0 : p = p_0$$

ii) Alternative Hypothesis :-

$$i) H_1 : p \neq p_0 \quad (p > p_0 \text{ or } p < p_0)$$

$$ii) H_1 : p > p_0$$

$$iii) H_1 : p < p_0$$

iii) Level of Significance :- (α)

$$\text{Table value} = z_{\alpha/2} \text{ or } Z_{\alpha}$$

iv) Test statistic :-

$$Z = \frac{P - P_0}{\sqrt{\frac{P_0 Q}{n}}} \quad Q = 1 - P$$

v) Conclusion :-

If $|Z| < z_{\alpha/2}$, H_0 True

If $|Z| > z_{\alpha/2}$, H_0 False.

Confidence Limit for P :-

$$1) P - z_{\alpha/2} \sqrt{\frac{PQ}{n}} < P < P + z_{\alpha/2} \sqrt{\frac{PQ}{n}}$$

(or)

$$P - 3\sqrt{\frac{PQ}{n}} < P < P + 3\sqrt{\frac{PQ}{n}}$$

If, P is unknown,

$$P - 3\sqrt{\frac{PQ}{n}} < P < P + 3\sqrt{\frac{PQ}{n}}$$

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

Sample size (n) = 1000

No. of rice eaters (x) = 540

Sample proportion (p) = $\frac{x}{n} = \frac{540}{1000} = 0.54$.

Population proportion (P) = 50.1.

$$P = 0.5$$

$$\alpha = 1 - P = 1 - 0.5 = 0.5$$

i) Null Hypothesis :-

$$H_0: P = 50.1 = 0.5$$

2) Alternative Hypothesis.

$$H_1: p \neq 0.5.$$

3) Level of significance (α)

$$\alpha = 1\%.$$

$$\text{Table value} = z_{\alpha/2} = 2.58.$$

a) Test statistic:-

$$z = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} \Rightarrow \frac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1000}}} = 2.532$$

5) Conclusion:-

$$z = 2.532, z_{\alpha/2} = 2.58$$

$$|z| > z_{\alpha/2}$$

$\therefore H_0$

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

3) A manufacturer claimed that 95% of equipment which he supplied to a factory, conformed to specifications. And examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance.

4) Among 900 people in a state 90 are found to be chapathi eaters. Construct 99% confidence interval for the population.

5) The experience had shown that 80% of a manufactured product is of top quality. In one day's production of 400 articles, only 50 are top quality. Test the hypothesis at 0.05 level.

2) Sample size of men (n) = 600

No. of men smokers (x) = 325.

sample proportion of smokers (p) = $\frac{x}{n} = \frac{325}{600}$

$$p = 0.5417$$

population proportion (P) = 50%.

i) Null Hypothesis :- $H_0 : p = 50\%$.

ii) Alternative Hypothesis :- $H_1 : p \neq 50\%$.

iii) Level of Significance, α :- (α)

$$\alpha = 5\%$$

$$\text{Table value} = Z_\alpha = 1.645.$$

iv) Test statistic.

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \Rightarrow \frac{0.5417 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}} = 2.04.$$

v) Conclusion :-

$$Z = 2.04, Z_\alpha = 1.645.$$

$$|Z| > Z_\alpha.$$

$\therefore H_0$ is false

$\therefore H_1$ is true.

$$p > 50\%$$

majority of men are smokers.

3) Sample size (n) = 200

No. of faulty = 18

No. of good condition items (x) = $200 - 18$

$$= 182$$

Sample proportion of good (p) = $\frac{x}{n} = \frac{182}{200} = 0.91$

Population proportion (P) = 95% = 0.95

$$Q = 1 - P$$

$$\Rightarrow 1 - 0.95$$

$$\Rightarrow 0.05.$$

i) Null Hypothesis :- $P = 95\%$.

ii) Alternative Hypothesis :-

$$H_1 : P \neq 95\%.$$

iii) Level of significance (α)

$$\alpha = 5\%.$$

$$\text{Table value } z_{\alpha/2} = z_{0.025} = 1.96.$$

iv) Test statistic :-

$$z = \frac{P - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} \Rightarrow \frac{0.91 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{900}}} = -2.59$$

v) Conclusion :-

$$z = -2.59, z_{\alpha/2} = 1.96$$

$$|z| = 2.59, z_{\alpha/2} = 1.96.$$

$$|z| > z_{\alpha/2}$$

$\therefore H_0$ is False

H_1 is True

$$P \neq 95\%.$$

4) Sample size (n) = 900

$$\text{No. of chapatti eaters (x)} = 90.$$

$$\text{Sample proportion of chapatti eaters} = P = \frac{x}{n} = \frac{90}{900} = 0.1$$

$$q = 1 - P = 1 - 0.1 = 0.9$$

$$(1 - \alpha) \approx 99\%, z_{\alpha/2} = 2.58.$$

Confidence Interval is

$$P - z_{\alpha/2} \cdot \sqrt{\frac{Pq}{n}} < P < P + z_{\alpha/2} \cdot \sqrt{\frac{Pq}{n}}$$

$$0.1 - 2.58 \cdot \sqrt{\frac{0.1 \times 0.9}{900}} < P < 0.1 + 2.58 \cdot \sqrt{\frac{0.1 \times 0.9}{900}}$$

$$\Rightarrow 0.02 < P < 0.13.$$

$$\therefore (0.02, 0.13)$$

3) Sample size (n) = 400

No. of top quality (x) = 50.

Sample proportion of top quality (P) = $\frac{x}{n} \Rightarrow \frac{50}{400} = 0.125$

Population proportion of top quality (p) = 20%.

$$Q = 1 - 0.2 = 0.8$$

$$= 0.82.$$

i) Null Hypothesis :-

$$P = 20\%.$$

ii) Alternative Hypothesis

$$H_1: P \neq 20\%.$$

iii) Level of Significance :- (α)

$$\alpha = 5\%.$$

$$\text{Table Value} = z_{\alpha/2} = 1.96.$$

iv) Test statistic :-

$$Z = \frac{P - p}{\sqrt{\frac{pq}{n}}} \Rightarrow \frac{0.125 - 0.2}{\sqrt{\frac{(0.2)(0.8)}{400}}} = -3.75.$$

v) Conclusion :-

$$|Z| = 3.75, z_{\alpha/2} = 1.96$$

$$|Z| > z_{\alpha/2}$$

H_0 is false
 H_1 is True.

$$P \neq 20\%.$$

Note:-

Test statistic

$$Z = \frac{P - p}{\sqrt{\frac{pq}{n}}} \quad \text{where } \sqrt{\frac{pq}{n}} \text{ is standard error.}$$

$$\text{Max. Error} = z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \text{where } \frac{\sigma}{\sqrt{n}} \text{ is S.E.}$$

$$\text{Max. Error} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

Test of significance of difference b/w two proportions

Let p_1, p_2 be two sample proportions of size n_1, n_2 drawn from population proportions P_1, P_2 .

To test whether the two samples have been drawn from same population.

$$P_1 = \frac{x_1}{n_1}$$

$$P_2 = \frac{x_2}{n_2}$$

i) Null Hypothesis:

$$H_0 : P_1 = P_2$$

ii) Alternative Hypothesis:-

$$\text{i)} H_1 : P_1 \neq P_2$$

$$\text{ii)} H_1 : P_1 > P_2$$

$$\text{iii)} H_1 : P_1 < P_2$$

iii) Level of significance

iv) Test statistic:

case(i) if $P_1 = P_2 = 0$

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

$$\text{where } p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$q = 1 - p$$

Case(ii) if $P_1 \neq P_2 \neq 0$

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

$$p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$q = 1 - p$$

v) Conclusion :-

If $|z| < Z_{\alpha/2}$, H_0 True

$|z| > Z_{\alpha/2}$, H_1 True

Standard Error $(p_1 - p_2)$ is

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

Large Sample Test ($n \geq 30$)

- 1) Test of significance of single mean

 \bar{x}, n Table value = $Z_{\alpha/2}$ or Z_{α}

- 2) Test of sign. difference

b/w two means

$$\begin{array}{ll} \bar{x} & \bar{y} \\ n_1 & n_2 \\ \sigma_1 & \sigma_2 \end{array}$$

- 3) Test of sign. of single proportion

 $\bar{x}, n \rightarrow$ not given.

$$p = \frac{\bar{x}}{n}$$

- 4) Test of sign. difference b/w two proportions.

 $\bar{x}, n \rightarrow$ not given.

$$\begin{array}{ll} x_1 & x_2 \\ n_1 & n_2 \end{array}$$

$$p_1 = \frac{x_1}{n_1}, \quad p_2 = \frac{x_2}{n_2}$$

Small Sample Test ($n \leq 30$)

- 1) t-test

Test of significance of single mean.

 $\bar{x}, n, v = (n-1)$

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

Test Statistic

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

$H_0: \mu = \mu_0$
 $H_1: \mu \neq \mu_0$

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

$H_0: \mu_1 = \mu_2$
 $H_1: \mu_1 \neq \mu_2$

$$z = \frac{p - p_0}{\sqrt{\frac{pq}{n}}}$$

$H_0: p = p_0$
 $H_1: p \neq p_0$

$$q = 1 - p$$

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

$H_0: p_1 = p_2$
 $H_1: p_1 \neq p_2$

where $p = \text{combined proportion}$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$q = 1 - p$$

case i.) if 's' is given

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

case ii.) if s is calculated (S)
 i.e., S is not given directly

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

for two tailed

$$\text{Table value} = t_{\alpha/2}$$

for one tailed

$$\text{Table value} = t_\alpha$$

8) Test of sign difference b/w
two means.

$$\begin{array}{cc} \bar{x}_1 & \bar{y} \\ n_1 & n_2 \\ s_1^2 & s_2^2 \end{array}$$

for t
degrees of freedom:

$$v = n_1 + n_2 - 2$$

Two tailed test:-

$$\text{Table value} = t_{\alpha/2}$$

One tailed test:-

$$\text{Table value} = t_\alpha$$

Case i) If s_1, s_2 are given
directly.

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

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

Case ii) If s_1, s_2 are not given
directly.

$$s^2 = \frac{1}{n_1 + n_2 - 2} \left[\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2 \right]$$

$$\bar{x} = \frac{1}{n_1} \sum x_i \quad \bar{y} = \frac{1}{n_2} \sum y_j$$

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

8) Paired t-test:-

x_i	before
y_i	After

$$d_i = x_i - y_i \quad (\text{or}) \quad d_i = y_i - x_i$$

$$\text{d.f. (v)} = (n - 1)$$

F-Test

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

$$\bar{d} = \frac{1}{n} \sum d_i$$

$$s^2 = \frac{1}{n-1} \sum (d_i - \bar{d})^2$$

F-test:-

Variance

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2$$

$F = \frac{\text{Greater variance}}{\text{smaller variance}}$.

$$F = \frac{s_1^2}{s_2^2}, \text{ if } s_1^2 > s_2^2.$$

Table value:-

$$F \propto (N_1, N_2)$$

$$N_1 = (n_1 - 1)$$

$$N_2 = (n_2 - 1)$$

$$s_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \sum \frac{(x_i - \bar{x})^2}{n_1 - 1}$$

$$s_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \sum \frac{(y_j - \bar{y})^2}{n_2 - 1}$$

χ^2 - test

1) Goodness of fit :-

$$H_0 : O_i = E_i$$

$$H_1 : O_i \neq E_i$$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Table value :-

$$d.f (v) = n - 1$$

$$\chi^2_{\alpha}$$

2) Independence of attributes.

H_0 : there is no association b/w attributes.

H_1 : There is a association b/w attributes.

		C	D	
		$\frac{(a+b)(a+d)}{N}$	$\frac{(b+c)(b+d)}{N}$	$a+b$
A	B	$\frac{(a+c)(c+d)}{N}$	$\frac{(b+d)(a+d)}{N}$	$c+d$
		$a+c$	$b+d$	$a+b+c+d$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$d.f (v) = (\text{No. of rows} - 1) \times (\text{No. of columns} - 1)$$

$$\text{Table value} = \chi^2_{\alpha}$$