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 parameter using sample statistics.

Estimator :- The procedure or rule to determine unknown population parameter using sample statistics is called an Estimator

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

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

$\therefore$ $\bar{x}$, 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 Estimators :-

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

Ex :- 1) $E(\bar{x}) = \mu$

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

Biased estimator :-

A sample statistic $\hat{\theta}$ is said to be ~~unbiased~~ 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 $x_1, x_2, x_3, \ldots x_n$ 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 $E[x_i] = \mu$, $\forall i$

$$E[\bar{x}] = \frac{1}{n} E[x_1 + x_2 + x_3 + \ldots x_n]$$

$$= \frac{1}{n} [E(x_1) + E(x_2) + \ldots + E(x_n)]$$

$$= \frac{1}{n} [\mu + \mu + \ldots + \mu]$$

$$= \frac{1}{\cancel{n}} [\cancel{n} \mu]$$

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

Theorem:

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

Proof

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

consider

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

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

$$\Rightarrow \sum_{i=1}^{n} (x_i - \mu)^2 - 2(\bar{x} - \mu)\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \mu\right) + n(\bar{x} - \mu)^2.$$

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

$$\Rightarrow \sum^1 (x_i - \mu)^2 - 2n(\bar{x} - \mu)^2 + n(\bar{x} - \mu)^2 \rightarrow (1).$$

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

$$E(s^2) = \frac{1}{n-1} E\left[\sum^1 (x_i - \bar{x})^2\right]$$

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

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} \sigma^2 - n \cdot \sigma_{\bar{x}}^2\right]$$

$$\Rightarrow \frac{1}{n-1}\left[n\sigma^2 - \cancel{n} \cdot \frac{\sigma^2}{\cancel{n}}\right] \qquad \left[\because \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\right]$$

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

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

Note:-

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

Scanned by CamScanner

3) The population parameter θ can consists many estimators.

Most efficient parameter estimator :-

If $\hat{\theta}_1$, $\hat{\theta}_2$ are two unbiased estimators of some population parameter 'θ'. If $V(\hat{\theta}_1) < V(\hat{\theta}_2)$ 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 $n \to \infty$ then $\hat{\theta}_n$ is consistent.

ii) Unbiasedness :-

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

iii) Efficiency :-

The 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 a 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. ie, $(163, 167)$.

The interval of population parameter θ is of the form $\hat{\theta}_L$ is $\hat{\theta}_L < \theta < \hat{\theta}_u$ 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 'ϵ'.

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

$$\epsilon = |\bar{x}-\mu| = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

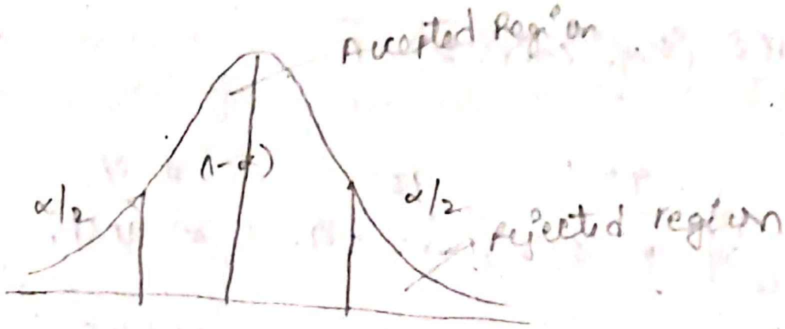
$$\bar{x}-\mu = \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

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

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \longrightarrow (*)$$

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

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



The interval * 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) :-

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

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

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

Maximum Error & confidence Interval of μ (small sample):-

(σ & σ is unknown.)

Maximum error is

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

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

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

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

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

Sample size (n) :-

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

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

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

Table Values ($Z_{\alpha/2}$ (or) Z_α)

$(1-\alpha)$	99%	95%	90%	
level of significance $(\alpha) \rightarrow$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 2\%$
Two-tailed Test	$Z_{\alpha/2} = 2.58$	1.96	1.645	2.33
Right tailed Test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	1.28	
Left tailed Test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	-1.28	

$Z_{\alpha/2}$ is z-value leaving Area $\alpha/2$ to the right of Normal curve.

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

Given $n = 100$

Mean of sample $(\bar{x}) = 212.3$

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

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

Max. Error $(e) = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

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

$= 0.098$

Confidence Interval for μ is

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

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

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

$$212.202 < \mu < 212.398$$

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 $(n) = 100$

$S.D = \sigma = 5$

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

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

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

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 ~~the~~ μ.

Given sample size $= 100$.

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

$\sigma = 5.1$

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

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

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

$= 0.9996$

Confidence Interval for μ is

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

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

$21.6 - \overset{0.996}{<} \mu < 21.6 - 0.9996$

$\Rightarrow$

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 $(E) = 3$.

$\sigma = 20$

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

$Z_{\alpha/2} = 1.96$

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

$n = \left(\frac{Z_{\alpha/2}\,\sigma}{E}\right)^2$

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

5) In a Study of Automobile Insurance a random sample of 80 body repair cost had mean of ₹ 472.36 and S.D of ₹ 62.35. 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 $(E) = 10/-$

Sample size $(n) = 80$

Sample Mean $(\bar{x}) = 472.36$

$\sigma = 62.35$

$Z_{\alpha/2} = ?$

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

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

$Z_{\alpha/2} = \frac{10 \times \sqrt{80}}{62.35}$

$= 1.43$

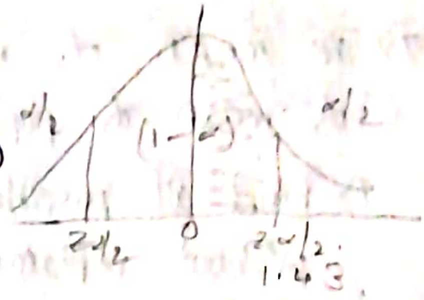
from Tables,

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

confidence $= (1-\alpha) = 2 \times (0.4236)$

$= 0.8472.$

$\Rightarrow 84.72\%.$



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 doesnot differ from the true mean by more than 30 secs.

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

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

$z_{\alpha/2} = \frac{0.5(\sqrt{40})}{2.06} = 1.535.$

from tables.

$P(0 < z < 1.535) = \frac{1-\alpha}{2} = 0.4382$

confidence $= (1-\alpha) = 2 \times (0.4382)$

$= 0.8764.$

$= 87.64\%.$

Scanned by CamScanner

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 (i) and assigning each sample a probability of 1/10. Verify that mean of these $\bar{x}$ 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 $= {}^{N}C_{n} = {}^{5}C_{3}$

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

Samples = {(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 $(\mu) = \frac{3+6+9+15+27}{5} = \frac{60}{5}$,

$\mu = 12$.

Assigning prob. 1/10 to each sample mean.

$\bar{x}$	6	8	9	10	12	13	14	15	16	17.
$p(\bar{x})$	$\frac{1}{10}$	$\frac{1}{10}$	.	.	.	.	.	.	.	.

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

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

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

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

$\Rightarrow 12$

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

i) show that x/n 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 15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

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

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

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 ~~so~~not unbiased estimator of p.

9) Confidence limits

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

Sample ~~mean~~ variance

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

Sample = {15, 17, 10, 18, 16, 9, 7, 11, 13, 14}

$$\text{Sample Mean } (\bar{x}) = \frac{15+17+10+18+16+9+7+11+13+14}{10}$$

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

$$\Rightarrow 13$$

$$S^2 = \frac{1}{10-1}\{(15-13)^2+(17-13)^2+\cdots\cdots+(14-13)^2\}$$

$$= \frac{1}{9}[4+16+9+25+9+16+36+4+0+1].$$

$$= \frac{120}{9} \Rightarrow \frac{40}{3}.$$

$$\boxed{S = 3.65}$$

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

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 desease.

Test of Hypothesis :-

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

2m
***
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 H1. defined as

i) $H_1 : \mu \neq \mu_0$ ($\mu > \mu_0$ (or) $\mu < \mu_0$) (Two-tailed).

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 Ho. 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 Ho when it is True).

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

P(Reject Ho when it is True) = α - Error.

It is also called procedure's risk.

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

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

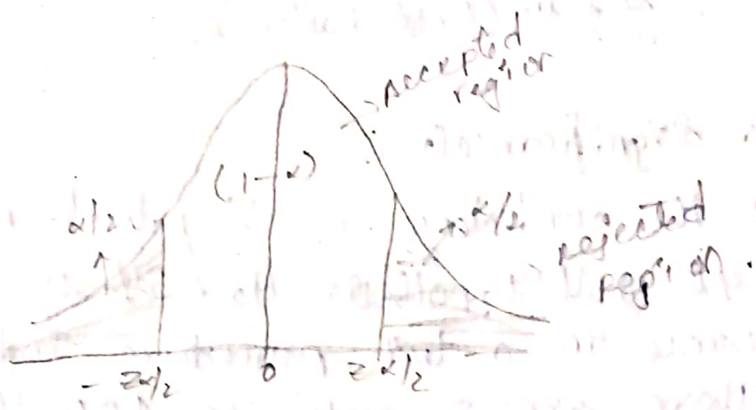
It is also called consumer's risk.

Scanned by CamScanner

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

ii) 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.

iii) 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 Ho is true (or) Ho is accepted.

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

critical values :-

	(α) level of significance			
	1%	5%	10%	2%
Two-tailed	$\lvert z_{\alpha/2}\rvert = 2.58$	1.96	1.645	2.33
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 mean :-

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 :- $H_0: \mu \neq \mu_0$

2) Alternative Hypothesis :-

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

ii) $H_1 : \mu > \mu_0$

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

3) Level of significance :-

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

4) Test statistic :-

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

5) Conclusion :-

if $|z| < z_{\alpha/2}$ Ho True

if $|z| > z_{\alpha/2}$ Ho false

1) 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.

Sample size (n) = 400
Sample mean ($\bar{x}$) = 40
std population (σ) = 10
Population mean (μ) = 38

i) Null Hypothesis :- $\mu = 38$

ii) Alternative Hypothesis :- $H_0 : \mu \neq 38$

iii) Level of significance :-

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

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

iv) Test statistic :-

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \Rightarrow \frac{40-38}{10/\sqrt{400}} \Rightarrow \frac{2 \times 20}{10} = 4.$$

v) Conclusion :-

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

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

$\therefore H_0$ is false

$\therefore H_1$ is True

$\mu \neq 88$ //

confidence limits with 95% for μ is

$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ (or) $\bar{x} \pm \epsilon$.

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

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

confidence limits.

$(\bar{x} - \epsilon ., \bar{x} + \epsilon)$

$(40 - 0.98, 40 + 0.98)$

$(39.02 + 40.98)$

2) A Sample of 900 items has mean 3.4 and S.D 2.61. Ic this sample has been taken from a large population of mean 3.25 and S.D 2.61. If the population is normal and 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 and 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.

2) Given

Sample size(n) = 900

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

population s.D(σ) = 2.61

population mean(μ) = 3.25.

i) Null Hypothesis :-

$H_0 : \mu = 3.25$

ii) Alternative Hypothesis.

$H_1 : \mu \neq 3.25$

Two tailed.

iii) Level of significance :-

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

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

iv) Test statistic :-

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{3.4-3.25}{2.61/\sqrt{900}} = 1.724.$$

v) Conclusion :-

$z = 1.724$, $z_{\alpha/2} = 1.96$

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

H_0 is True

$\therefore \mu = 3.25$.

Confidence limits :-

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

(or)

$$\bar{x} \pm e$$

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

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

$$e = 0.17$$

confidence interval is = $(\bar{x}-e, \bar{x}+e)$

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

$= (3.23, 3.57)$.

Scanned by CamScanner

3) Given,

Sample size $(n) = 36$ calls.

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

Population S.D $(\sigma^2) = 16$ mins $\Rightarrow \sigma = 4$ mins.

Population mean $(\mu) = 10$ mins.

i) Null Hypothesis :-

$H_0 : \mu = 10$ mins.

ii) Alternative Hypothesis :-

$H_1 : \mu < 10$ mins.

(left tailed)

iii) Level of significance :-

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

Table value $z_\alpha = -1.645$.

iv) Test Statistic :-

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \Rightarrow \frac{11 - 10}{4/\sqrt{36}} = 1.5.$$

v) Conclusion :-

$z = 1.5$, $z_\alpha = -1.645$

$|z| = 1.5$, $|z_\alpha| = 1.645$.

$|z| < |z_\alpha|$

$\therefore H_0$ is True

$\mu = 10$ mins

4) Given

Sample size $(n) = 64$

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

population S.D $(\sigma) = 35$ kgs

population mean $(\mu) = 56$ kgs.

i) Null Hypothesis :-

$H_0 : \mu = 56$.

Scanned by CamScanner

ii) Alternative Hypothesis :-

$H_1 : \mu \neq 56$

Two tailed.

iii) Level of significance.

$\alpha = 5\%$.

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

iv) Test statistic :-

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \Rightarrow \frac{70-56}{25/\sqrt{64}}.$$

$z = 4.48$

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

$\mu \neq 56$ kg's.

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

i) Null Hypothe

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 :-

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

ii) Alternative Hypothesis :-

i) $H_1 : \mu_1 \neq \mu_2$ ($\mu_1 > \mu_2$ (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)

Table value = $Z_{\alpha/2}$ (or) Z_α

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 s $n_1 = 1000$ and 2000 members are 67.5 ~~and~~ inches 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 sizes $n_1 = 1000$ $n_2 = 2000$

population S.D = 2.5 inches.

i) Null type :-

$H_0 : \mu_1 = \mu_2$

ii) Alternative Hypothesis :-

$H_1 : \mu_1 \neq \mu_2$ (Two-tailed)

iii) Level of significance (α) = 5%

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

Scanned by CamScanner

iv) Test Statistic :-

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

$$= \frac{67.5 - 68}{(2.5)\sqrt{\frac{1}{1000} + \frac{1}{2000}}} \Rightarrow -5.16$$

v) Conclusions :-

$Z = -5.16$ $\quad Z_{\alpha/2} = 1.96$

$|z| = 5.16$ $\quad Z_{\alpha/2} = 1.96$

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

$\therefore$ H_0 is false

$\therefore$ H_1 is True

$\mu_1 \neq \mu_2$

$\therefore$ 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 250 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 ~~ma~~ 50 male students who have not participated in sports is 67.2 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) 1^{st} group => Sample size (n_1) = 150
Sample mean ($\bar{x}$) = 75
population S.D (σ_1) = 15

2^{nd} group => Sample size (n_2) = 250
Sample mean ($\bar{x}$) = 70
population ~~size~~ (σ_2) = 20.

i) Null Hypothesis :-
$H_1 : \mu_1 = \mu_2$.

ii) Alternative Hypothesis.
$H_1 : \mu_1 \neq \mu_2$.

iii) Level of significance
$\alpha = 5\%$.
Table value $z_{\alpha/2} = 1.96$

iv) Test ~~of~~ statistic :-

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

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 ($\bar{y}$) = 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)

$\alpha = 5\% = 0.05$

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

iv) Test significance :-

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \Rightarrow \frac{72 - 70}{\sqrt{\frac{(8)^2}{32} + \frac{6^2}{36}}} \Rightarrow 1.154.$$

v) Conclusion :-

$Z = 1.154$ $\quad Z_{\alpha/2} = 1.645$

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

$\therefore H_0$ is True (Accepted).

$\mu_1 = \mu_2$.

$\therefore$ 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 not 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)

$\alpha = 5\% = 0.05.$

Table value $Z_{\alpha/2} = 1.645$

iv) Test statistic :-

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \Rightarrow \frac{68.2 - 67.2}{\sqrt{\frac{(2.5)^2}{50} + \frac{(0.8)^2}{50}}} \Rightarrow 1.88,$$

v) Conclusion.

$z = 1.88 \quad Z_{\alpha/2} = 1.645$

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

$\therefore$ 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 'p' which is taken from a population proportion 'P'.

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

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

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) $H_1 : P \neq P_0$ ($P > P_0$ (or) $P < P_0$).

ii) $H_1 : P > P_0$

iii) $H_1 : P < P_0$

iii) Level of Significance :- (α)

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

iv) Test statistic.

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} \qquad Q = 1-P.$$

v) Conclusion :-

if $|Z| < Z_{\alpha/2}$ Ho True

if $|Z| > Z_{\alpha/2}$ Ho False.

Confidence Limit for P :-

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

(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 ~~one~~ 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%.

$P = 0.5$.

$Q = 1-P = 1-0.5 = 0.5$

i) Null Hypothesis :-

Ho : $P = 50\% = 0.5$.

Scanned by CamScanner

2) Alternative Hypothesis.

$H_1 : P \neq 0.5$.

3) Level of significance (α)

$\alpha = 1\%$.

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

4) Test statistic :-

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

5) Conclusion :-

$Z = 2.532$, $Z_{\alpha/2} = 2.58$

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

$\therefore H_0$

2) 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 piece of equipment revealed that 18 were faulty. Test its 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 true population.

5) The experience had shown that 20% of a manufactured product is of top quality. In one days 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 :- $P = 50\%$.
ii) Alternative Hypothesis :- H_1 : $P > 50\%$.
iii) Level of Significance :- (α)
$\alpha = 5\%$.
Table value $= Z_\alpha = 1.645$.

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

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 $(P) = 95\% = 0.95$
$Q = 1 - P$.
$\Rightarrow 1 - 0.95$
$\Rightarrow 0.05$.

Scanned by CamScanner

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

ii) Alternative Hypo :-

$H_1 : p \neq 95\%$.

iii) Level of significance (α)

$\alpha = 5\%$.

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

iv) Test statistic :-

$$z = \frac{P - p}{\sqrt{\frac{PQ}{n}}} \Rightarrow \frac{0.91 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{200}}} \Rightarrow -2.59$$

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) $= 900$

No. of chapathi eaters (x) $= 90$.

sample proportion of chapathi eaters $= P = \frac{x}{n} = \frac{90}{900}$

$= 0.1$

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

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

Confidence Interval is

$$P - Z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}} < P < P + Z_{\alpha/2} \sqrt{\frac{PQ}{n}}.$$

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

$\Rightarrow 0.07 < P < 0.13$.

$\therefore (0.07, 0.13)$

5) Sample size (n) = 400

no. of top quality (x) = 50.

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

population proportion of top quality (P) = 20%

$Q = 1 - 0.2 => 0.8$ = 0.2.

i) Null Hypothesis :-

$P = 20\%$

ii) Alternative Hypothesis

$H_1 : P \neq 20\%$

iii) Level of significance :- (α)

$\alpha = 5\%$

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

iv) Test statistic :-

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

v) Conclusion :-

$z = -3.75$, $Z_{\alpha/2} = 1.96$

$|z| = 3.75$, $Z_{\alpha/2} = 1.96$

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

H_0 is false

H_1 is True.

$P \neq 20\%$.

note :-

Test statistic

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$ where $\sqrt{\frac{PQ}{n}}$ is standard Error.

Max Error = $Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$

sample proportion

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

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

Scanned by CamScanner

Test of significance of difference b/w two proportions:

Let p_1, p_2 be two large sample proportions of size's n_1, n_2 drawn from population proportions P_1, P_2. To test whether the two samples have been drawn from sample 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:-

i) $H_1 : P_1 \neq P_2$

ii) $H_1 : P_1 > P_2$

iii) $H_1 : P_1 < P_2$.

iii) Level of significance

iv) Test statistic.

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

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

where $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}.$

$$Q = 1 - P.$$

case (ii) if $P_1 - P_2 \neq 0$

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

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

$$Q = 1 - P.$$

v) Conclusion :-

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

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

Standard Error $(P_1 - P_2)$ is

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

Test — UNIT-IV — Test Statistic

Large sample tests ($n \geq 30$)

1) Test of significance of single mean

$\bar{x}, n$

Table value $= Z_{\alpha/2}$ (or) Z_α

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

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

2) Test of sign. difference b/w Two means

$\bar{x}$	$\bar{y}$
n_1	n_2
σ_1	σ_2

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

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

3) Test of sign. of single proportion

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

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

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

$Q = 1-P$

$H_0 : P = P_0$
$H_1 : P \neq P_0$

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

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

x_1	x_2
n_1	n_2

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

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

$H_0 : P_1 = P_2$
$H_1 : P_1 \neq P_2$.

where p = combined proportion.

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

$q = 1-p$.

Small Sample Test ($n < 30$)

1) t-test

Test of significance of single mean.

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

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

case i) if 's' is given

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

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

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

for Two tailed:-

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

for one - tailed:-

Table value $= t_\alpha$.

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

$\bar{x}$ $\bar{y}$

n_1 n_2

s_1^2 s_2^2

for t

degrees of freedom:-

$\nu = n_1 + n_2 - 2$.

Two tailed test:-

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

one tailed test:-

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.

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

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

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

3) paired t-test:-

x_i	Before	
y_i	After.	

$d_i = x_i - y_i$ (or) $d_i = y_i - x_i$

d.f $(\nu) = (n-1)$

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

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

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

F- Tes

F- test :-

variances

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

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

$$F = \frac{S_1^2}{S_2^2}, \text{ if } S_1^2 > S_2^2.$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1}$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{\sum (y_i - \bar{y})^2}{n_2 - 1}$$

Table value :-

$F_\alpha (\nu_1, \nu_2)$

$\nu_1 = (n_1 - 1)$

$\nu_2 = (n_2 - 1)$

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 :-

$d.f (\nu) = n - 1$

χ^2_α

2) Independence of attributes.

H_0 : there is no association b/w attributes.

H_1 : There is a association b/w attributes.

	C	D
A	a	b
B	c	d

O_i

$\Rightarrow$

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

$N = a+b+c+d$

E_i

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

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

Table value $= \chi^2_\alpha$