1

SAMPLING DISTRIBUTION

Population:- A group of individuals under study is called population (or) Population is the collection of objects.

A Population may be finite or infinite. If the population contains a finite no. of units then it is called finite population.

Ex:- No. of students in a college.

If the population contains an infinite no. of units called infinite population.

Ex:- Stars in the sky

$x_1, x_2, \ldots \ldots x_N$ are population units.

Sample:- A finite subset of statistical individuals in a population is called a sample (or) subset of a population

Ex:- Blood ~~sam~~testing

Sample size:- The no. of objects in the sample is called sample size. and it is denoted by n.

NOTE:- Population size is denoted by N.

TYPES OF SAMPLING:-

There are four types of sampling

① Purposive sampling

② Random sampling

③ Stratified sampling

④ Systematic sampling

① Purposive Sampling:- If the sample elements are selected with a definite purpose in mind, then the sample selected is called purposive sampling.

② Random sampling (or Probability sampling):- It is the process of drawing a sample from a population in such a way that each member of the population has an equal chance of being included in the sample. The sample obtained by the process of random sampling is called a random sample.

Ex:- Selecting randomly 20 words from a dictionary is a random sample.

If each element of a population may be selected more than once then it is called sampling with replacement whereas if the element cannot be selected more than once, it is called sampling without replacement.

NOTE:- If N is the size of a population and n is the sample size, then

(i) The number of samples with replacement $= N^n$

(ii) The number of samples without " $= {}^N C_n$

③ Stratified sampling:- This method is useful when the population is heterogeneous. In this type of sampling, the population is first sub-divided into several parts (or small groups) called strata according to some relevant characteristics so that each stratum is more or less homogeneous. Each stratum is called a sub-population. Then a small sample (called sub-sample) is selected from each stratum at random. All the sub-samples are combined together to form the stratified sample which represents the population properly.

The Process of obtaining and obtaining a stratified sample with a view to estimating the characteristic of the population is known as stratified sampling.

2

④ Systematic Sampling:- If the Population size is finite, all the units of the population are arranged in some order, Then from the first k items, one unit is selected at random. This unit and every kth unit of the serially listed population combined together constitute a systematic sample. This type of sampling is known as systematic sampling.

CLASSIFICATION OF SAMPLES:-

Samples are classified in two ways.

1) Large sample: If the size of the sample $(n) \geq 30$, the sample is said to be large sample.

2) Small sample:- If the size of the sample $(n) < 30$, the sample is said to be small sample (or) exact sample.

PARAMETERS AND STATISTICS:-

Statistical measures computed from the sample observations is known as statistic. Let $x_1, x_2 \ldots x_n$ are n sample observations.

mean $\bar{x}$ & Standard deviation (s) are known as Statistic.

In other words, the mean, median, mode, S.D. variance measures of the population are called parameters and the measures obtained from the sample of the population are called statistics.

The parameter refers to population while statistic refer to sample.

<u>Sample mean</u>:- If $x_1, x_2, \ldots x_n$ represents a random sample of size n, then the sample mean is denoted by the statistic

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

<u>Sample variance</u>:- If $x_1, x_2 \ldots x_n$ represent a random of sample size n, then the sample variance is denoted by the statistic.

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

CENTRAL LIMIT THEOREM :-

If $x_i, (i=1,2,\ldots n)$ be independent random variables such that $E(x_i) = \mu_i$ and $V(x_i) = \sigma_i^2$, then under certain very general conditions, the random variable $S_n = x_1 + x_2 + \ldots + x_n$, is asymptotically normal with mean μ and standard deviation σ where $\mu = \sum_{i=1}^{n} \mu_i$ and $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$

SAMPLING DISTRIBUTION OF MEAN (σ KNOWN) :-

The probability distribution of $\bar{x}$ is called the sampling distribution of means. The sampling distribution of a statistic depends on the size of the population, the size of the samples, and the method of choosing the samples.

<u>Infinite Population</u>:- Suppose the samples are drawn from an infinite population (or) sampling is done with replacement, then

the mean of the sampling distribution of means,

$$\mu_{\bar{x}} = \frac{\mu + \mu + \mu + \ldots + \mu}{n} = \mu$$

and variance $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2 + \sigma^2 + \ldots + \sigma^2}{n} = \dfrac{\sigma^2}{n}$.

8

$\therefore$ S.D of mean, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

The total no. of sample with replacement $= N^n$

The sampling distribution of $\bar{x}$ will be approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ provided that the sample size is large.

Standard sample mean, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

Finite Population:- Consider a finite population of size N with mean μ and standard deviation σ. Draw all possible samples of size n without replacement, from this population.

The mean of the sampling distribution of mean (for $N>n$) is $\mu_{\bar{x}} = \mu$.

The variance is $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \left(\frac{N-n}{N-1}\right)$

and S.D. is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Here, the factor $\left(\frac{N-n}{N-1}\right)$, often called the finite population corrector factor.

The number of samples without replacement $= {}^{N}C_{n}$

① Find the value of the finite population correction factor for $n=10$ and $N=1000$.

Sol: Given N = The size of the finite population = 1000

n = The size of the sample = 10

$$\therefore \text{Correction factor} = \frac{N-n}{N-1} = \frac{1000-10}{1000-1}$$

$$= \frac{990}{999}$$

$$= 0.991.$$

② How many different samples of size two can be chosen, from a finite population of size 25.

Sol:. We can take $^{N}C_n$ samples of size 'n' from the population of size N.

Here $N=25$, $n=2$

$\therefore$ we can take $^{25}C_2 = 300$ samples of size 2 from finite population of size 25.

③ A population consists of five numbers 2, 3, 6, 8 and 11. Consider all possible samples of size two which can be drawn with replacement from this population. Find

a) The mean of the Population.

b) The standard deviation of the population.

c) The mean of the sampling distribution of means and

d) The standard deviation of the sampling distribution of means (i.e., standard error of means).

Sol: a) Mean of the population is given by

$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

b) Variance of the population (σ^2) is given by

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

$$= \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$$

$$= 10.8$$

$$\therefore \sigma = \sqrt{10.8}$$

i.e., $\sigma = 3.29$

(c) Sampling with replacement (Infinite Population):

The total no. of samples with replacement is

$N^n = 5^2 = 25$ samples of size 2.

Here N = Population size and n = Sample size listing all possibles samples of size 2 from population 2, 3, 6, 8, 11 with replacement we get 25 samples.

$$\begin{Bmatrix} (2,2) & (2,3) & (2,6) & (2,8) & (2,11) \\ (3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\ (6,2) & (6,3) & (6,6) & (6,8) & (6,11) \\ (8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\ (11,2) & (11,3) & (11,6) & (11,8) & (11,11) \end{Bmatrix}$$

Now compute the arithmetic mean for each of these 25 samples. The set of 25 means $\bar{x}$ of these 25 samples, gives rise to the distribution of means of the samples known as sampling distribution of means.

The sample mean are

$$\begin{Bmatrix} 2 & 2.5 & 4 & 5 & 6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{Bmatrix} \text{———} (I)$$

and the mean of sampling distribution of means is the mean of these 25 means.

$$\mu_{\bar{x}} = \frac{\text{sum of all sample means in (I)}}{25} = \frac{150}{25} = 6$$

Illustrating that $\mu_{\bar{x}} = \mu$.

d) The variance $\sigma_{\bar{x}}^2$ of the sampling distribution of means is obtained by subtracting the mean 6 from each number in (I) and squaring the result, adding all 25 members thus obtained, and dividing by 25.

$$\sigma_{\bar{x}}^2 = \frac{(2-6)^2 + \cdots\cdots + (11-6)^2}{25} = \frac{135}{25} = 5.40$$

and thus $\sigma_{\bar{x}} = \sqrt{5.40} = 2.32$

Clearly, for finite population involving sampling with replacement (or infinite population)

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \text{ or } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{3.29}{\sqrt{2}} = 2.32.$$

(2) above problem as without replacement

a & b same

(c) Sampling without replacement (finite population):

The total no. of samples without replacement is $N_{C_n} = {}^5C_2 = 10$ samples of size 2.

The 10 samples are

$$\begin{Bmatrix} (2,3) & (2,6) & (2,8) & (2,11) \\ (3,6) & (3,8) & (3,11) & \\ (6,8) & (6,11) & & \\ (8,11) & & & \end{Bmatrix}$$

The selection (2,3) is considered same as (3,2)

The corresponding sample means are.

$$\begin{Bmatrix} 2.5 & 4 & 5 & 6.5 \\ 4.5 & 5.5 & 7 & \\ 7 & 8.5 & & \\ 9.5 & & & \end{Bmatrix}$$

The mean of the sampling distribution of means is

$$\mu_{\bar{x}} = \frac{(2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5)}{10} = 6$$

Illustrating that $\mu_{\bar{x}} = \mu$

d) The variance of sampling distribution of means

$$\sigma_{\bar{x}}^2 = \frac{(2.5-6)^2+(4-6)^2+\dots+(9.5-6)^2}{10}$$

$$= 4.05$$

$$\sigma_{\bar{x}} = 2.01$$

Showing that $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

$$= \frac{10.8}{2}\left(\frac{5-2}{5-1}\right) = 4.05$$

for sampling without replacement.

② Find the mean and standard deviation of sampling distribution of variances for the population 2,3,4,5 by drawing samples of size two (a) with replacement (b) without replacement.

③ A population consists of six numbers 4,8,12,16,20,24. Consider all samples of size two which can be drawn without replacement from this population. Find

a) The Population mean

b) The Population standard deviation

c) The mean of the sampling distribution of means

d) The S.D. of the sampling distribution of means.

④ Samples of size 2 are taken from the population 1,2,3,4,5,6 (i) with replacement & (ii) without replacement.

a, b, c, d. (above)

⑤ If the population is 3, 6, 9, 15, 27

a) list all possible samples of size 3 that can be taken without replacement from the finite population

b) Calculate the mean of each of the sampling distribution of means.

c) Find the S.D. of sampling distribution of means

Sol:. Mean of the population $\mu = \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12$

standard deviation of the population.

$$\sigma = \sqrt{\frac{(3-12)^2+(6-12)^2+\cdots\cdots+(27-12)^2}{5}}$$

$$= \sqrt{\frac{360}{5}} = 8.4853$$

a) Sampling without replacement (finite population):

The total number of samples without replacement is

$$N_{C_n} = {}^5C_3 = 10$$

The 10 samples are

(3, 6, 9), (3, 6, 15), (3, 9, 15), (3, 6, 27), (3, 9, 27), (3, 5, 27), (6, 9, 15), (6, 9, 27), (6, 15, 27), (9, 15, 27).

b) Mean of the sampling distribution of mean is

$$\mu_{\bar{x}} = \frac{6+8+9+10+12+13+14+15+16+17}{10} = \frac{120}{10} = 12$$

c) $\sigma_{\bar{x}}^2 = \frac{1}{9}\left[(6-12)^2+(8-12)^2+\cdots+(17-12)^2\right]$

$$= \frac{120}{9}$$

$$= 13.3$$

$\therefore \sigma_{\bar{x}} = \sqrt{13.3} = 3.651.$

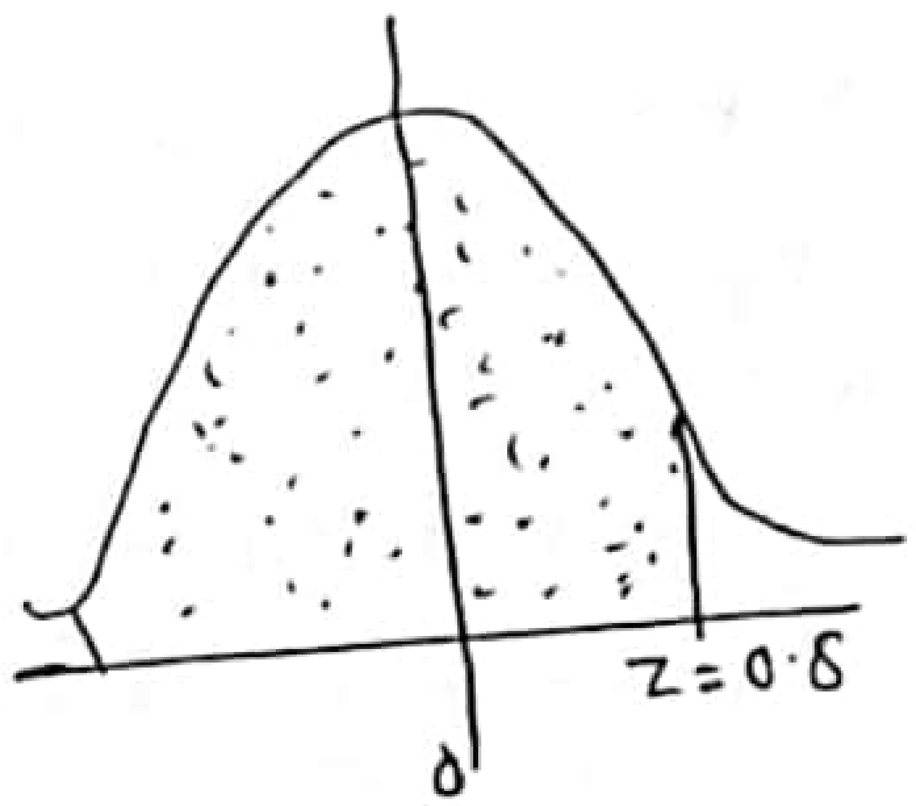
⑥ The mean height of students in a college is 155cms and standard deviation is 15. What is the probability that the mean height of 36 students is less than 157 cms.

Sol:- μ = mean of the population
= mean height of students of a college = 155cm

σ = S.D. of population = 15cms

n = Sample size = 36

$\bar{x}$ = Mean of sample = 157cms



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

$= \dfrac{157-155}{15/\sqrt{36}} = 0.8$

$\therefore P(\bar{x} \le 157) = P(z<0.8) = 0.5 + P(0 \le z < 0.8)$

$= 0.5 + 0.2881 = 0.7881$

Thus the probability that the mean height of 36 students is less than 157 = 0.7881.

⑦ A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$ and the variance $\sigma^2 = 256$. What is the probability that $\bar{x}$ will be between 75 and 78.

Sol:- n = size of the sample = 100

μ = mean of the population = 76

σ^2 = variance of the population = 256

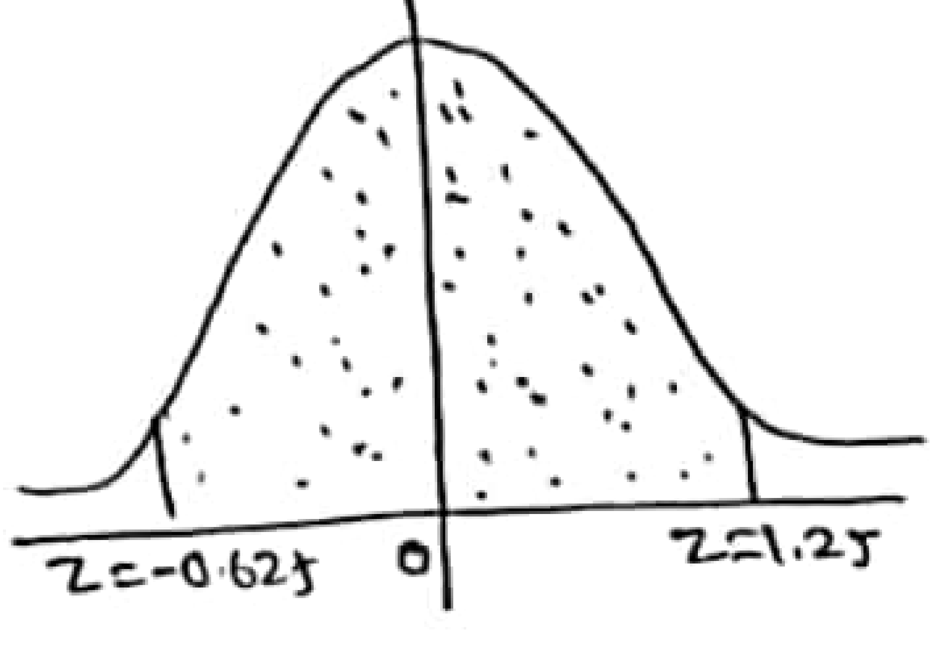
$\therefore \sigma = 16$

Since n is large, the sample mean $\bar{x} \sim N(\mu, \sigma^2/n)$

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

When $\bar{x}_1 = 75$

$Z_1 = \dfrac{75-76}{16/\sqrt{100}} = -0.625$



and $\bar{x}_2 = 78$, $z_2 = \frac{\bar{x}_2-\mu}{\sigma/\sqrt{n}} = \frac{78-76}{16/\sqrt{100}} = 1.25$

$\therefore P(75 \le \bar{x} \le 78) = P(z_1 \le z \le z_2)$

$= P(-0.625 \le z \le 1.25)$

$= P(-0.625 \le z \le 0) + P(0 \le z \le 1.25)$

$= 0.2334 + 0.3944$

$= 0.628$

⑧ A normal population has a mean of 0.1 and S.D. of 2.1. Find the probability that mean of a sample of size 900 will be negative.

Sol:- Given $\mu = 0.1$, $\sigma = 2.1$ and $n = 900$

The standard normal variate

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-0.1}{2.1/\sqrt{900}}$$

$\therefore \bar{x} = 0.1 + 0.007z$, where $z \sim N(0,1)$

$\therefore$ The required probability, that the sample mean is negative is given by

$P(\bar{x} < 0) = P(0.1 + 0.07z < 0)$

$= P(0.07z < -0.1)$

$= P\left(z < \frac{-0.1}{0.07}\right)$

$= P(z < -1.43)$

$= 0.50 - P(0 < z < 1.43)$

$= 0.50 - 0.4236$

$= 0.0764.$

⑨ A random sample of size 64 is taken from a normal population with $\mu = 51.4$ and $\sigma = 68$. What is the probability that the mean of the sample will (a) exceed 52.9

(b) fail between 50.5 and 52.3

(c) be less than 50.6.

SAMPLING DISTRIBUTION OF DIFFERNCES AND SUMS:-

Let μ_{S_1} and σ_{S_1} be the mean and S.D. of sampling distribution of statistic S_1 obtained by computing S_1 for all possible samples of size n_1 draw from a population A. Also let μ_{S_2} and σ_{S_2} be the mean and standard deviation of sampling distribution of statistic S_2 obtained by computing S_2 for all possible samples of size n_2 drawn from another different population B.

Now compute the statistic S_1-S_2, the difference of the statistic from all the possible combinations of these samples from the two populations A and B.

Then The mean $\mu_{S_1-S_2}$ and the S.D. $\sigma_{S_1-S_2}$ of the sampling distribution of differences are given by

$$\mu_{S_1-S_2} = \mu_{S_1} - \mu_{S_2} \text{ and } \sigma_{S_1-S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

assuming that the samples are independent.

Sampling distribution of sum of statistics has mean $\mu_{S_1+S_2}$ and S.D. $\sigma_{S_1+S_2}$ given by

$$\mu_{S_1+S_2} = \mu_{S_1} + \mu_{S_2} \text{ and } \sigma_{S_1+S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

① Let $U_1=(3,7,8)$, $U_2=(2,4)$ Find

a) μ_{U_1} (b) μ_{U_2} (c) Mean of the sampling distribution of the difference of means $\mu_{U_1-U_2}$

d) σ_{U_1} (e) σ_{U_2} (f) the standard deviation of the sampling distribution of the difference of means $\sigma_{U_1-U_2}$

Sol: Given $U_1=\{3,7,8\}$ and $U_2=\{2,4\}$

$U_1-U_2=\{1,5,6,4,3,-1\}$

Now

(a) $\mu_{U_1} = \frac{3+7+8}{3} = 6$

Scanned with CamScanner

(b) $\mu_{u_2} = \frac{2+4}{2} = 3$

(c) $\mu_{u_1-u_2} = \frac{1+5+6+4+3-1}{6} = 3$

(d) $\sigma_{u_1} = \sqrt{\frac{(6-3)^2+(6-7)^2+(6-8)^2}{3}} = \sqrt{\frac{14}{3}}$

(e) $\sigma_{u_2} = \sqrt{\frac{(2-3)^2+(3-4)^2}{2}} = 1$

(f) $\sigma_{u_1-u_2} = \sqrt{\frac{(1-3)^2+(5-3)^2+(6-3)^2+(4-3)^2+(3-3)^2+(-1-3)^2}{6}}$

$= \sqrt{\frac{34}{6}} = \sqrt{\frac{17}{3}}.$

SAMPLING DISTRIBUTION OF THE MEAN (σ UNKNOWN):-

For large sample of size ($n \geq 30$), even if standard deviation σ of Population is not known, it does not make any difference. Hence we can substitute the sample S.D. 's' in the place of σ and the sample S.D. 's', is calculated using the sample mean $\bar{x}$ by the formula

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

For small sample of size ($n<30$), when σ is unknown, it can be substituted by s. Provided we make the assumption that the sample is taken from normal population. we will discuss the t-distribution, F-distribution and χ^2-distribution in next chapter.

ESTIMATION

(1)

Estimate:- To find an unknown population parameter is a estimate statement.

Estimator:- The method of determining unknown population parameter is called estimator. For instance, sample mean is an estimator of population mean because sample mean is a method of determining the population mean.

A parameter can have one or two or many estimators. Basically, there are two kinds of estimates to determine the statistic of the population parameter namely.

(a) Point estimation

(b) Interval estimation

① Point estimation:- If a single value is calculated as an estimate from an unknown population parameter. The procedure to find the parameter is called point estimation.

Properties of good estimator:- An estimator is said to be a good estimator if it is,

(i) unbiased (ii) consistent (iii) Efficient and sufficient.

(1) Unbiased estimator:- A statistic $\hat{\theta}$ is said to be an unbiased estimator if and only if the mean of the sampling distribution of estimator is equal to the parameter θ.

i.e, E[Statistic] = Parameter

$$E[\hat{\theta}] = \theta$$

(2) Consistent:- Any statistic $\bar{\theta}$ or $\hat{\theta}$ is said to be consistent if and only if, $E(\hat{\theta}) = \theta$ and $V(\hat{\theta}) \to 0$ as $n \to \infty$

(3) Efficient:- Suppose $\hat{\theta}_1, \hat{\theta}_2$ be two unbiased estimator of popl^n

Scanned with CamScanner

Parameter θ and $V(\hat{\theta}_1)$ & $V(\hat{\theta}_2)$ be the variance of statistic. if $V(\hat{\theta}_1) < V(\hat{\theta}_2)$ then $\hat{\theta}_1$ is said to be efficient (or) more efficient unbiased estimator of θ.

Interval estimation:- In general, point estimator does not coincide with a true value of the parameter, so it is preferred to obtain arrange of values in an interval in which the parameter value lies.

Suppose α is the probability that the interval (a,b) does not include that the parameter θ.

Then $1-\alpha = P[a<\theta<b] = P[\theta, a, b]$

The interval (a,b) is called confidence interval $(1-\alpha)$ is called confidence coefficient and is generally given as $(1-\alpha)100\%$. a & b are also known as confidence limits of parameter θ.

MAXIMUM ERROR OF ESTIMATE E FOR LARGE SAMPLES:-

The sample mean estimate very rarely equals to the mean of population μ, a point estimate is generally accompanied with a statement of error which gives difference between estimate and the quantity to be estimated; the estimator.

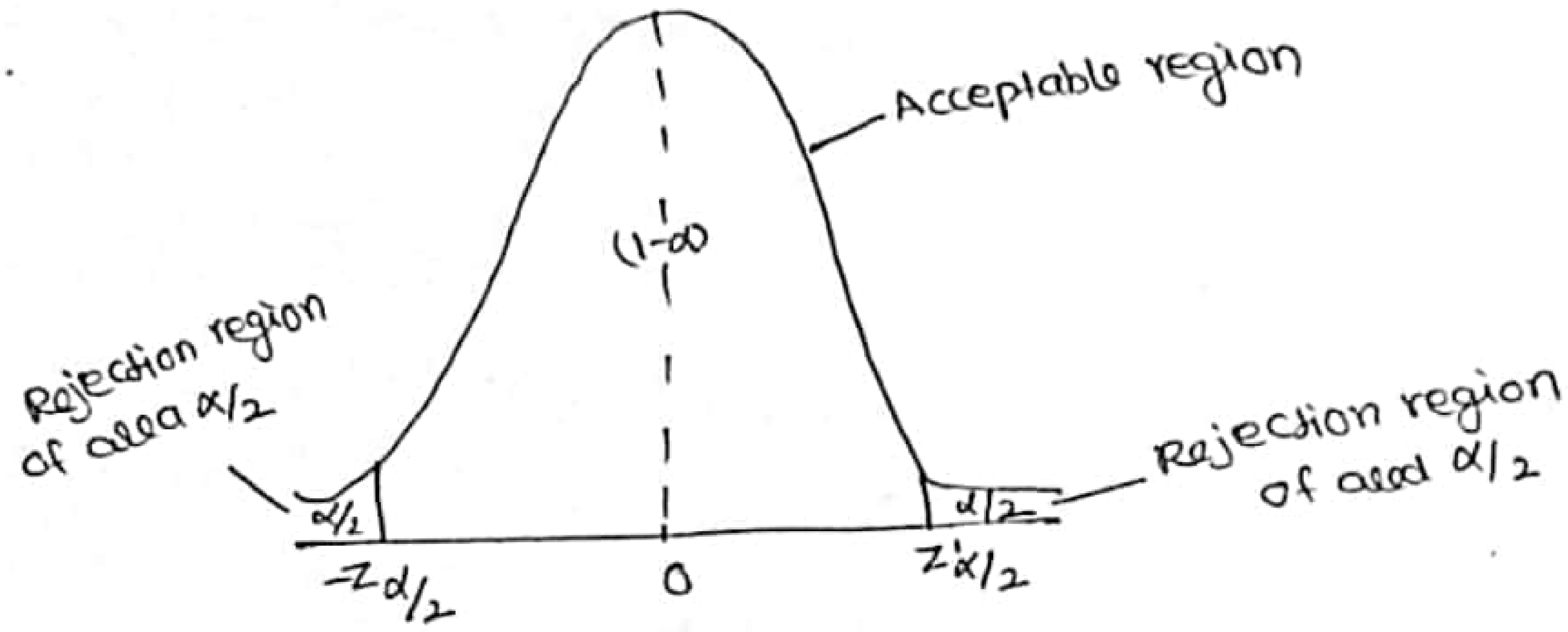
Thus error is $|\bar{x}-\mu|$

For large n, the random variable $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is normal variate approximately

Then $P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$

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

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



$$P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1-\alpha$$

Multiplying each term in the inequality by $\sigma/\sqrt{n}$ and then subtracting $\bar{x}$ from each term and multiplying -1.

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

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

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

Confidence interval for μ, σ known:-

If $\bar{x}$ is the mean of a random sample of size n from the population with known variance σ^2, $(1-\alpha)100\%$ confidence interval for μ is given by

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

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

So, the maximum error of estimate E with $(1-\alpha)$ probability is given by

$$\boxed{E = z_{\alpha/2}(\sigma/\sqrt{n})}$$

Sample size:- When α, E, σ are known, the sample size n is given by

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2 \quad (or) \quad E_{max} = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

when σ is unknown ; σ is replace by S, S is the standard deviation of sample to determine E.

Thus the maximum Error estimate

$$E = t_{\alpha/2}\left(S/\sqrt{n}\right) \text{ with } (1-\alpha) \text{ probability}$$

Confidence interval μ, σ unknown:-

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $(1-\alpha)100\%$ confidence interval for μ is

$$\bar{x} - t_{\alpha/2}\left(S/\sqrt{n}\right) < \mu < \bar{x} + t_{\alpha/2}\left(S/\sqrt{n}\right)$$

where $t_{\alpha/2}$ is the t-value with $v = n-1$ d.f., leaving an area of $\alpha/2$ to the right.

$\therefore$ The maximum error of estimate for small samples is given by

$$E = t_{\alpha/2}\, S/\sqrt{n-1}$$

where n = sample size, S = S.D. of means.

BAYESIAN ESTIMATION:-

Bayesian estimation is used to obtain mean and variance of posterior distribution of a population.

If the prior distribution on parameters mean μ_0 and variance σ_0^2 of a population are known then find the posterior distribution parameter of a given population. This is called Bayesian estimation.

$$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2+\sigma^2} \quad \text{and } \sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}}$$

n = Sample size

$\bar{x}$ = sample mean

s = standard deviation of sample use $s = \sigma$.

σ_0^2 = prior variance

σ^2 = sample variance

μ_0 = prior mean.

Here μ_1 and σ_1 are known as the mean and S.D. of the posterior distribution. In the computation of μ_1 and σ_1, σ^2 is assumed to be known, when σ^2 is unknown, which is generally the case, is replaced by sample variance s^2 provided $n \geq 30$ (large sample)

Bayesian interval for μ:

$(1-\alpha)100\%$ Bayesian interval for μ is given by

$$\mu_1 - Z_{\alpha/2}\cdot\sigma_1 < \mu < \mu_1 + Z_{\alpha/2}\cdot\sigma_1$$

Sufficient:- $T = t(x_1, x_2 \dots x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2 \dots x_n$ of size n from the population with density $f(x, \theta)$. such that the conditional distribution of $x_1, x_2 \dots x_n$ given T is independent of θ, then T is a sufficient estimator of θ.

① In a study of an automobile insurence a random sample of 80 body repair costs had a mean of Rs. 472.36 and the S.D. of Rs 62.35. If $\bar{x}$ is used as a point estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exceed Rs. 10?

Sol. Size of random sample $n = 80$.

The mean of random sample $\bar{x}$ = Rs 472.36

Standard deviation, σ = Rs 62.35

Maximum error of estimate, E_{max} = Rs 10

we have $E_{max} = Z_{\alpha/2} \cdot \sigma/\sqrt{n}$

$$\Rightarrow Z_{\alpha/2} = \frac{E_{max} \cdot \sqrt{n}}{\sigma} = \frac{10\sqrt{80}}{62.35}$$

$$\Rightarrow Z_{\alpha_2} = 0.9236$$

$$\Rightarrow 1-\alpha/2 = Z_{\alpha/2} = 0.9236 \Rightarrow \alpha = 0.1528$$

$\therefore$ confidence $= (1-\alpha)100\% = 84.72\%$.

② If we can assert with 95% that the maximum error is 0.05 and P=0.2 find the size of the sample.

Sol: Given P=0.2, E=0.05

we know that maximum error $E = Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$

$$\Rightarrow 0.05 = 1.96\sqrt{\frac{0.2\times0.8}{n}}$$

$$\Rightarrow n = \frac{0.2\times0.8\times(1.96)^2}{(0.05)^2}$$

$$n = 246$$

③ Assuming that $\sigma = 20.0$, 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 3.0 points?

Sol: Given maximum error E = 3.0 and $\sigma = 20.0$

we have $Z_{\alpha/2} = 1.96$

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

$$n = \left(\frac{1.96\times20}{3}\right)^2 = 170.74$$

$$n \cong 171.$$

(4) What is the maximum error one can expect to make with probability 0.90, when using the mean of a random sample of size $n=64$ to estimate the mean of population with $\sigma^2 = 2.56$

10

sol: Here $n=64$

The probability $=0.90$

$\sigma^2 = 2.56 \Rightarrow \sigma = \sqrt{2.56} = 1.6$

Confidence limit $=90\%$.

$\therefore Z_{\alpha/2} = 1.645$

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

$= 1.645 \times \frac{1.6}{\sqrt{64}} = 0.329.$

(5) A random sample of size 100 has a standard deviation of 5. what can you say about the maximum error with 95% confidence.

sol: Given $s=5$, $n=100$

$Z_{\alpha/2}$ for 95% confidence $=1.96$

W.K.T. Maximum error $E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$\therefore E = (1.96) \frac{5}{\sqrt{100}} = 0.98.$

(6) A random sample of 400 items is found to have mean 82 and S.D. of 18. Find the maximum error of estimation at 95% confidence interval. Find the confidence limits for the mean if $\bar{x}=82$.

sol: Give S.D. $\sigma = 18$

$n = 400$

$Z_{\alpha/2}$ for 95% confidence $= 1.96$

Sample mean $= \bar{x} = 82$

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

$= \frac{1.96 \times 18}{20} = 1.764$

The limits for the confidence are

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

$\therefore$ Confidence limits are 80.236 and 83.764.

(7) A sample of size 300 was taken whose variance is 225 and mean 54. Construct 95% confidence interval for the mean.

Sol: Since the sample size 300 is large (>30), normal distribution is used as the sampling distribution.

Here $n = 300$, $\bar{x}$ = Sample mean = 54, $\sigma = \sqrt{225} = 15$

$$\therefore \text{S.E. of } \bar{x} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{300}} = 0.866$$

95% confidence limits for the population mean are.

$$\bar{x} \pm 1.96 (\text{S.E. of } \bar{x}) = 54 \pm 1.96(0.866)$$

$$= 55.697 \ \& \ 52.3$$

$\therefore$ The required confidence interval is (55.697, 52.3)

(8) A population random variable has mean 100 and S.D. 16. What are the mean and S.D. of the sample mean for the random sample of size 4 drawn with replacement.

Sol. Given $\mu = 100$, $\sigma = 16$, $n = 4$

Since the sampling is done with replacements, the population may be considered as infinite.

we have to find $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$.

$$\therefore \mu_{\bar{x}} = \mu = 100 \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{4}} = 8$$

(9) Find 95% confidence limits for the mean of a normality distri-buted population from which the following sample was taken.

15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

(10) A random sample of size 81 was taken whose variance is 20.25 and mean is 32, construct 98% confidence interval.

(7) (12)

TEST OF HYPOTHESIS

Test of hypothesis:- When parametric values are unknown, we estimate them through sample values. But the problem arises when the sample provides a value, which is either exactly equal to the parametric value, not too far, In that situation one has to develop some procedure which enables one to decide whether to accept a value or not on the basis of sample values, such a procedure known as testing of hypothesis.

In many circumstances, we are to make decisions about population on the basis of only sample information. For example, on the basis of sample data.

(i) a drug chemist is to decide whether a new drug is really effective in curing a disease.

(ii) a quality control manager is to determine whether a process is working properly.

TEST OF STATISTICAL HYPOTHESIS:-

A statistical test of hypothesis is a rule or procedure which makes one to decide about the acceptence or rejection of the hypothesis it can be denoted by H.

Hypothesis are two types

(i) Null Hypothesis

(ii) Alternative Hypothesis.

1) Null Hypothesis:- For applying the tests of significance, we irst setup a hypothesis which is a statement about the pulation parameter. Such a hypothesis is useually hypothesy

no difference is called a null hypothesis.

i.e., There is no significance difference between two parameters and it is denoted by H_0.

<u>Alternative Hypothesis</u>:- Any hypothesis which is complementary the null hypothesis is called an alternative hypothesis and it is denoted by H_1.

i.e., There is a significance difference b/w two parameters.

<u>Critical region</u>:- Let $x_1, x_2 \ldots x_n$ be the sample observations and S be the sample space. we devide the whole sample space 'S' into two disjoint parts w and S-w. A region in the sample space S. which amounts to rejection to H_0 is called critical region (or) region of rejection.

w is the critical region and S-w is the acceptance region. If the sample points falls in the subset w. H_0 is rejected otherwise H_0 is accepted

<u>ERRORS IN SAMPLING</u>

There are two types of errors in sampling

① Type I error

② Type II error.

Type I error (or) first kind of error :-

Reject H_0 . when H_0 is true. it is also known as rejection error. Probability of type I error it is denoted by α.

i.e., $\alpha = P[\text{Reject } H_0 / H_0 \text{ is true}]$.

Type II error (OR) Second kind of error :-

Accept H_0 when H_0 is false. It is also known as acceptance error. Probability of type-II error is denoted by β.

i.e, $\beta = P[\text{accept } H_0 / H_0 \text{ is false}]$

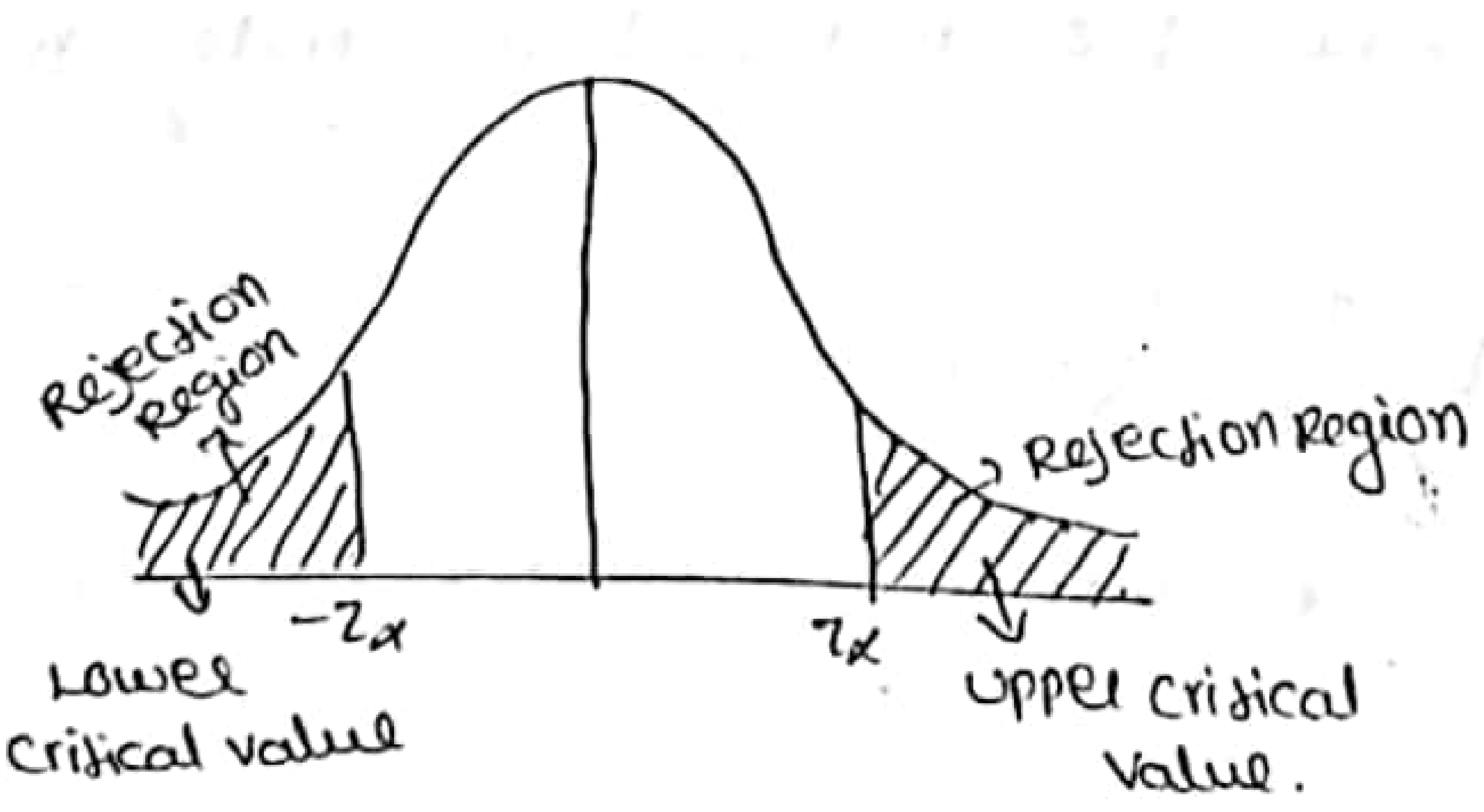
Type II error more serious than type I error.

Level of significance :- The probability of type I error is known as level of significance. The l.o.s. usually employed in testing of hypothesis are 5% and 1%. , α is always fixed in advance before collecting the sample information.

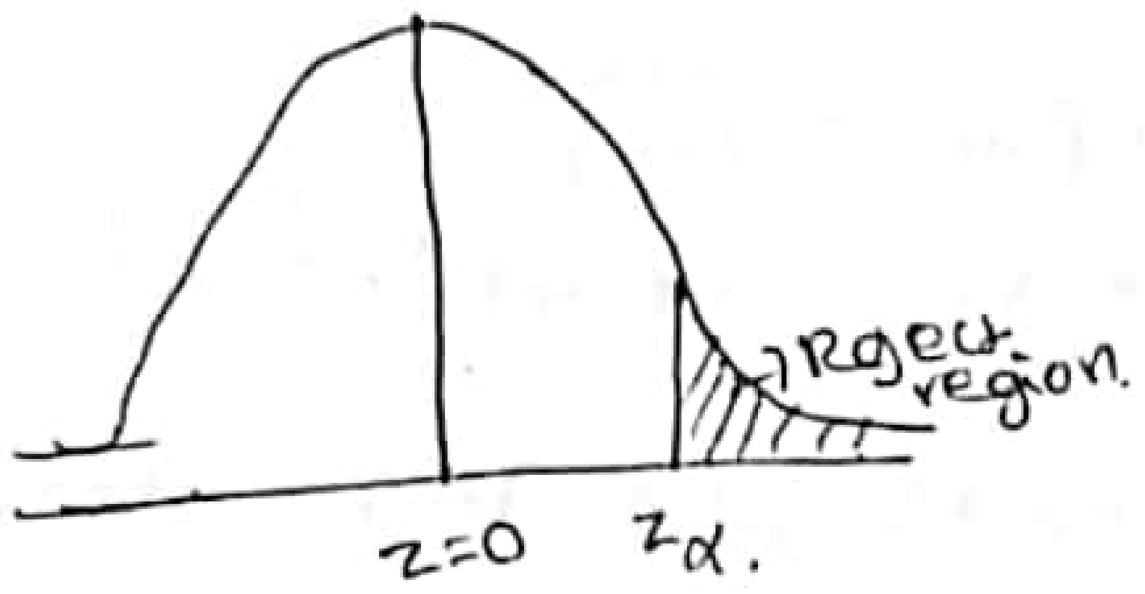
TWO-TAILED AND ONE-TAILED TESTS :-

Hypothesis is such that it leads two alternatives to the H_0 , it is said to be a two-tailed test.

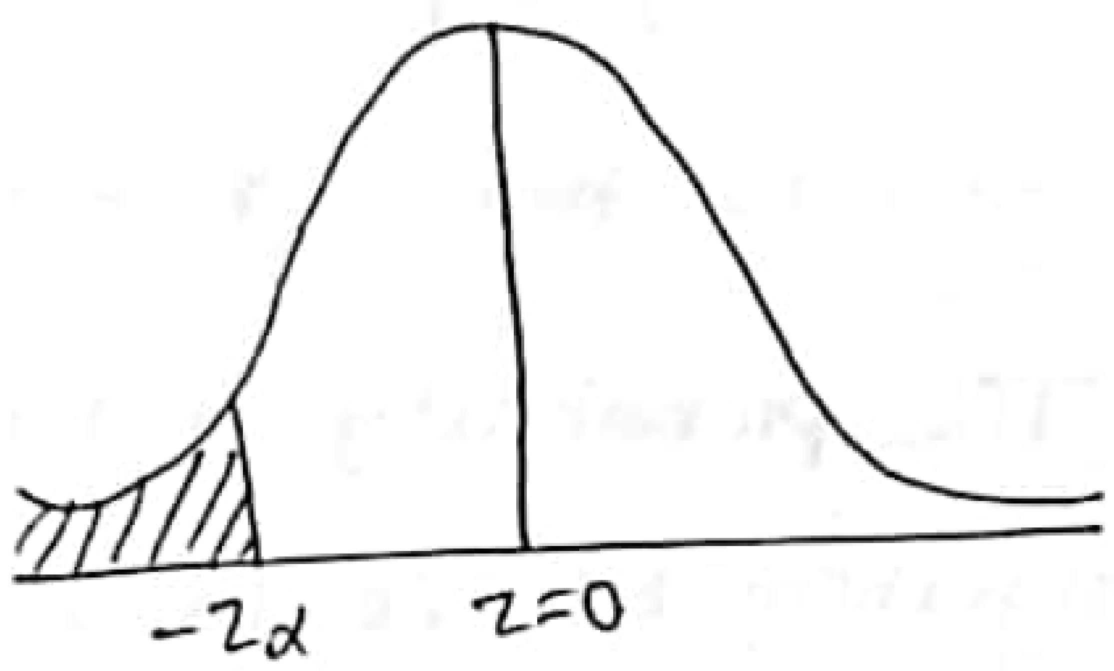
i.e., $H_1: \mu \neq \mu_0$ $[\mu > \mu_0 \text{ or } \mu < \mu_0]$



Right tailed test:- If alternative hypothesis $H_1: \mu > \mu_0$ to the H_0 it is said to be a right tailed test. In this situation the half of the area of critical region lies half on the right.



Left tailed test:- If $H_1: \mu < \mu_0$ to the H_0 it is known as left tailed test.



PROCEDURE FOR TESTING OF HYPOTHESIS:

① setup the Null hypothesis H_0: There is no significant difference b/w statistic and parameter

② setup alternative hypothesis H_1.

③ choose the appropriate level of significance α.

(1% or 5% or any percentage)

④ Calculate the value of z, test statistic under the null hypothesis.

$$|z| = \frac{t - E(t)}{S.E(t)} \sim N(0,1)$$

where t = statistic.

⑤ Conclusion:- Compare calculated value of z with the tabulated value at α l.o.s.

If cal.value $z <$ table value of z, then we accept the null hypothesis at α l.o.s. and otherwise reject H_0.

Critical value of z — l.o.s.

z_α	1%	5%	10%
Two tailed test	2.58	1.96	1.645
Right tailed test	2.33	1.645	1.28
Left tailed test	-2.33	-1.645	-1.28

TEST OF SIGNIFICANCE FOR LARGE SAMPLES:-

Standard normal variate $z = \frac{x-\mu}{\sigma}$

$$\mu = np$$
$$\sigma = \sqrt{npq}$$

① A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance.

Sol: Here $n=960$, $x=183$

P = probability of getting head $= 1/2$

$\therefore q = 1-p = 1/2$

1) Null Hypothesis H_0: The coin is unbiased

2) Alternative Hypothesis H_1: The coin is biased

3) Level of significance; $\alpha = 0.05$

4. The test Statistic is:

Standard normal variate $z = \frac{x-\mu}{\sigma}$

$p = 1/2$, $q = 1/2$

$n = 960$

$\mu = np = 960(1/2) = 480$

$\sigma = \sqrt{960(1/2)(1/2)}$

$= 15.49$

$z = \frac{x-\mu}{\sigma}$

$= \frac{183-480}{15.49}$

$= \frac{-297}{15.49} = -19.17$

$\therefore |z| = 19.17$

At $Z_{cal} \not< Z_{tab}$

$\therefore$ Ho is rejected at 5% l.o.s.

$\therefore$ The coin is biased.

(2) A die is tossed 960 times and it falls with 5 upwards 184 times. Is the die unbiased at a level of significance of 0.01?

As:- accept

(3) A coin was tossed 400 times and returned heads 216 times. Test the hypothesis that the coin is unbiased. Use a 0.05 L.O.S.

As: accept.

LARGE SAMPLE TESTS

Under large sample tests, we will see four important tests to test the significance.

1) Testing of significance for single Proportion
2) " " for difference of Proportion
3) " " for single mean
4) " " for difference of means.

TEST OF SIGNIFICANCE OF A SINGLE MEAN :-

Let a random sample of size $n (n \geq 30)$ has the sample mean $\bar{x}$, and μ be the population mean. Also the pop mean μ has a specified value μ_0.

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

we use $s = \sigma$

NOTE:-

① 95% confidence limits $\bar{x} \pm 1.96\, \sigma/\sqrt{n}$

② 99% confidence limits $\bar{x} \pm 2.58\, \sigma/\sqrt{n}$

③ 98% confidence limits $\bar{x} \pm 2.33\, \sigma/\sqrt{n}$

① An oceanographer wants to check whether the depth of the ocean in a certain region is 57.4 fathoms, as had previously been re--corded. What can be conclude at the 0.05 level of significance, if readings taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a standard deviation of 5.2 fathoms.

sol: Given $n = 40$, $\bar{x} = 59.1$ and $\sigma = 5.2$

1. Null Hypothesis H_0: $\mu = 57.4$

2. Alternative hypothesis H_1: $\mu \neq 57.4$

3) Level of significance $= 0.05$

4) The test statistic is $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$$= \frac{59.1-57.4}{5.2/\sqrt{40}} = 2.067$$

$$\cancel{= \frac{1.2}{0.7875} = 1.5238}$$

Tabulated value of z at 5% level of significance is 1.96

Hence calculated $z >$ tabulated z.

$\therefore$ The null hypothesis H_0 is rejected.

② A sample of 900 members has a mean of 3.4 cms and S.D 2.61 cms. Is the sample from a large population of mean 3.25 cms and S.D 2.61 cms. If the population is normal and its mean is unknown find the 95% fiducial limits of true mean.

Sol Given $n = 900$ $\mu = 3.25$

$\bar{x} = 3.4$ cm $\sigma = 2.61$

and $S = 2.61$

1) Null Hypothesis H_0: Assume that the sample has been drawn from the population with mean $\mu = 3.25$.

2) Alternative Hypothesis H_1: $\mu \neq 3.25$

3) The test statistic is, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

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

$$= 1.724$$

i.e., $z = 1.724 < 1.96$

$\therefore$ we accept the null hypothesis H_0.

i.e., The sample has been drawn from the population with mean $\mu = 3.25$.

95% confidence limits are given by

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96 \times \frac{2.61}{\sqrt{900}}$$

$$= 34 \pm 0.1705$$

i.e., 3.57 and 3.2295

③ A sample of 400 items is taken from a population whose S.D. is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also calculate 95% confidence interval for the population.

Sol: Given $n=400$, $\bar{x}=40$, $\mu=38$ and $\sigma=10$

1. Null hypothesis $H_0: \mu = 38$
2. Alternative hypothesis $H_1: \mu \neq 38$
3. Level of significance, $\alpha = 0.05$
4. The test statistic is $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$$= \frac{40-38}{10/\sqrt{400}} = 4$$

i.e, $z = 4 > 1.96$

$\therefore$ we reject the null hypothesis H_0.

i.e., the sample is not from the population whose mean is 38.

95% confidence interval is $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

i.e., $\left(40 - \frac{1.96(10)}{\sqrt{400}}, 40 + \frac{1.96(10)}{\sqrt{400}}\right)$

$= \left(40-\frac{1.96\times10}{20}, 40+\frac{1.96\times10}{20}\right)$

$= (39.02, 40.98)$

4) In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis $\mu = 32.6$ minutes in favour of alternative null hypothesis $\mu > 32.6$ at $\alpha = 0.025$ l.o.s.

Sol: Given $n = 60$, $\bar{x} = 33.8$, $\mu = 32.6$ and $\sigma = 6.1$

1. Null hypothesis $H_0: \mu = 32.6$.
2. Alternative hypothesis $H_1: \mu > 32.6$
3. Level of significance $\alpha = 0.025$
4. The test statistic is $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$= \frac{33.8-32.6}{6.1/\sqrt{60}} = \frac{1.2}{0.7875}$

$= 1.5238$

Tabulated value of Z at 0.025 l.o.s is 2.58

Hence calculated Z < tabulated Z

$\therefore$ The null hypothesis H_0 is ~~rejected~~ accepted.

5) A sample of 64 students have a mean weight of 70 kgs. Can this be regarded as a sample from a population with mean weight 65 kgs and standard deviation 25 kgs.

6) The mean life time of a sample of 100 light tubes produced by a company is found to be 1560 hrs with a population S.D. of 90 hrs. Test the hypothesis for $\alpha = 0.05$ that the mean life time of the tubes produced by the company is 1580 hrs.

TEST FOR EQUALITY OF TWO MEANS

(TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS OF TWO LARGE SAMPLES)

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means of two independent large random samples sizes n_1 and n_2 drawn from two populations having mean μ_1 and μ_2 and standard deviations σ_1 and σ_2. To test whether the population means are equal.

The test statistic is

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

where $\delta = \mu_1 - \mu_2$ = given constants

$H_0: \mu_1 = \mu_2$, then the test statistic

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

use $\sigma_1^2 = \sigma_2^2 = \sigma^2$

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

σ is not known we can use an estime of

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

① The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches.

Sol: Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 1000$, $n_2 = 2000$

and $\bar{x}_1 = 67.5$ inches , $\bar{x}_2 = 68$ inches.

Population S.D. $\sigma = 2.5$ inches.

1. Null Hypothesis H_0: The sample have been draw from the same population of S.D. 2.5 inches.

   i.e., $\mu_1 = \mu_2$ and $\sigma = 2.5$ inches.

2. Alternative hypothesis H_1: $\mu_1 \neq \mu_2$

3. The test statistic is $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}}$

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

$$\Rightarrow Z = \frac{-0.5}{0.0968} = -5.16$$

$\therefore |Z| = 5.16 > 1.96$

i.e., the calculated value of $Z >$ the table value of Z.

Hence the null hypothesis H_0 is rejected at 5% l.o.s. and we conclude that the samples are not drawn from the same population of S.D. 2.5 inches.

② Samples of students were drawn from two universities and from their weights in kilograms, mean and standard deviations are calculated and show below. Make a large sample test to test the significance of the difference between the means.

	Mean	S.D.	Size of the sample
University A	55	10	400
University B	57	15	100.

Sol'. Given $\bar{x}_1 = 55$, $\bar{x}_2 = 57$, $n_1 = 400$, $n_2 = 100$.

$S_1 = 10$ and $S_2 = 15$

1. Null Hypothesise H_0: $\bar{x}_1 = \bar{x}_2$ i.e., there is no difference

2. Alternative Hypothesis H_1: $\bar{x}_1 \neq \bar{x}_2$

3. Level of significance, $\alpha = 0.05$

4. Critical region : Accept H_0 if $-1.96 < Z < 1.96$

5. The test statistic is

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

$$= \frac{55-57}{\sqrt{\frac{100}{400} + \frac{225}{100}}} = \frac{-2}{\sqrt{\frac{1}{4} + \frac{9}{4}}}$$

$$= -1.26.$$

$\therefore |Z| = 1.26 < 1.96$

Hence, we accept the Null Hypothesis H_0 at 5% l.o.s. and conclude that there is no significant difference between the means.

(3) A researcher want to know the intelligence of students in a school. He selected two groups of students. In the first group there 150 students having mean IQ of 75 with a S.D. of 15 in the second group there are 250 students having mean IQ of 70 with S.D. is 20.

(4) A simple sample of the height of 6400 Englishmen has a mean of 67.85 inches and a S.D. of 2.56 inches while a simple sample of heights of 1600 Austrians has a mean of 68.55 inches and S.D. of 2.52 inches. Do the data indicate the Austrains are on the average ~~of~~ taller than the Englishmen? (use α is 0.01)

TEST OF SIGNIFICANCE FOR SINGLE PROPORTION:-

Suppose a large random sample of size n ~~has a sample proportion p of members~~ is taken from a normal population. To test the significance difference between the sample proportion p and the Population P, we use the test Statistic

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$ where n is the sample size.

NOTE ①:- Limits for population proportion P are given by $p \pm 3\sqrt{\frac{pq}{n}}$ where $q = 1-p$

② Confidence interval for proportion P of large sample at α l.o.s. is

$$p - Z_{\alpha/2}\sqrt{\frac{PQ}{n}} < P < p + Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$ where $Q = 1-P$ and

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

$Z_{\alpha/2} = 2.33$ (for 98%)

$Z_{\alpha/2} = 2.58$ (for 99%)

① A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% l.o.s.

sol. Given sample size $n = 200$

Number of pieces confirming to specification $= 200-18 = 182$

$\therefore p =$ proportion of pieces confirming to specification.

$$= \frac{182}{200} = 0.91$$

P = population proportion = $\frac{95}{100} = 0.95$.

1. Null Hypothesis H_0: The proportion of pieces confirming to specifications

i.e., P = 95%.

2. Alternative Hypothesis H_1: $P < 0.95$ (left - tail test)

3. The test statistic is

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91-0.95}{\sqrt{\frac{0.95 \times 0.05}{200}}}$$

$$Z = \frac{-0.04}{0.0154} = -2.59$$

Since alternative hypothesis is left tailed, the tabulated value of z at 5% l.o.s. is 1.645.

Since calculated value of $|z| = 2.6$ is greater than 1.645, we reject the Null Hypothesis H_0 at 5% l.o.s. and Conclude that the manufacturer's claim is rejected.

② 20 people were attacked by a disease and only 18 survived. will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of the hypothesis that is more at 5% l.o.s?

Sol:

n = Sample size = 20

X = number of survived sample = 18

p = proportion of survived people = $\frac{X}{n} = \frac{18}{20} = 0.9$

P = 0.85

$\therefore Q = 1 - P = 1 - 0.85$

$= 0.15$

1. Null Hypothesis H_0: P = 0.85

2. Alternative " H_1: P > 0.85 (Right tailed)

3. Level of significance, $\alpha = 0.5$

4. The test statistic is.

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

$$= \frac{0.9-0.85}{\sqrt{\frac{0.85\times0.15}{20}}} = \frac{0.5}{0.8} = 0.625$$

$\therefore$ Calculated $z = 0.625$

Tabulated z at 5% l.o.S. $z_\alpha = 1.645$

Since calculated $z <$ tabulated z,

we accept the N.H. H_0.

i.e., The proportion of the suvived people is 0.85.

③ A random sample of 500 pineapples was taken from a lage consignment and 65 were found to be bad. Find the percenta-age of bad pineapples in the consignment.

Sol. $n = 500$

p = Proprotion of bad pineapples in the sample

$= \frac{65}{500} = 0.13$

$q = 1-p = 0.87$

we know that the limits for population proportion P are given by

$$p \pm 3\sqrt{\frac{pq}{n}} = 0.13 \pm 3\sqrt{\frac{0.13\times0.87}{500}} = 0.13 \pm 0.045$$

$$= (0.085, 0.175)$$

$\therefore$ The percentage of bad pineapples in the consignment lies between 8.5 and 17.5.

(4) In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. construct a 99%, con-fidence interval for the corresponding true percentage.

(20)

Sol: we have $x=24$

$n=160$

$P=\frac{24}{160}=0.15, \quad Q=1-P=0.85$

Now $\sqrt{\frac{PQ}{n}}=\sqrt{\frac{0.15\times0.85}{160}}=0.028$

Confidence interval at 99%. l.o.s. is

$\left(P-3\sqrt{\frac{PQ}{n}},\ P+3\sqrt{\frac{PQ}{n}}\right)$

i.e.. $(0.15-3\times0.028,\ 0.15+3\times0.028)$

$(0.065, 0.234)$

(5) In a sample of 500 from a village in Andra pradesh, 280 are found to be rice eaters and the rest wheat eaters. Can we assume that the both articles are equally popular.

(6) In a hospital 480 feamales and 520 male babies were born in a week. Do these figure confirm the hypothesis that males and females are born in equal number?

TEST FOR EQUALITY OF TWO PROPORTIONS (OR TEST OF SIGNIFICANCE OF DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS - LARGE SAMPLES)

Let p_1 and p_2 be the proportions in two large random samples of sizes n_1 and n_2 drawn from two populations having P_1 and P_2.

$\therefore$ Standard Error of Difference $=S.E.(P_1-P_2)=\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}$

a) When the population proportions P_1 and P_2 are known.

Hence the test statistic is

$$Z = \frac{P_1 - P_2}{S.E.(P_1 - P_2)} = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

When the population proportions P_1 and P_2 are known but sample proportions p_1 and p_2 are known.

Hence the test statistic is

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

where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ so that $q = 1 - p$.

① Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same, at 5% level?

Sol: Given sample size $n_1 = 400$, $n_2 = 600$

Proportion of mean $p_1 = \frac{200}{400} = 0.5$

Proportion of women $p_2 = \frac{325}{600} = 0.541$

1) Null hypothesis H_0: Assume that there is no significant difference between the option of men and women as far as proposal of flyover is concerned.

i.e., $H_0: P_1 = P_2 = P$

2) Alternative hypothesis $H_1: P_1 \neq P_2$ (two tailed)

21

3. The test statistic is

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

where $p = \dfrac{n_1p_1 + n_2p_2}{n_1+n_2}$

$$= \frac{400 \times \frac{200}{400} + 600 \times \frac{325}{600}}{400+600} = \frac{525}{1000} = 0.525$$

and $q = 1-p = 1-0.525$

$= 0.475$

$$\therefore\ z = \frac{0.5 - 0.541}{\sqrt{0.525 \times 0.425\left(\frac{1}{400}+\frac{1}{600}\right)}}$$

$$= \frac{-0.041}{0.032} = -1.28$$

$|z| = 1.28$

since $|z| < 1.96$, we accept the null hypothesis H_0 at 5% l.o.s.

i.e., there is no difference of opinion between men and women as far as proposal of flyover is concerned.

② In two large populations, there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations.

Sol: Given $n_1 = 1200$, $n_2 = 900$

P_1 = Proportion of fair haired people in the first population.

$$P_1 = \frac{30}{100} = 0.3.$$

P_2 = Proportion of fair haired people in the second population

$$P_2 = \frac{25}{100} = 0.25.$$

1. Null Hypothesis H_0: Assume that the sample proportions are equal i.e., the difference in population proportions is likely to be hidden in sampling.

i.e., $H_0: P_1 = P_2$

2. Alternative Hypothesis: $H_1: P_1 \neq P_2$

3. The test statistic is $$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

where $Q_1 = 1 - P_1 = 1 - 0.3 = 0.7$

$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$

$$\therefore Z = \frac{0.3 - 0.25}{\sqrt{\frac{0.3 \times 0.7}{1200} + \frac{0.25 \times 0.75}{900}}}$$

$= 2.55$

i.e., $Z = 2.55$

Since $Z > 1.96$, therefore we reject the Null Hypothesis H_0 at 5% l.o.s. i.e., the sample proportions are not equal. Thus we conclude that the difference in population proportions is unlikely that the real difference will be hidden

③ In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the Proportion of wheat is concerned.

④ A cigarette manufacturing firm claims that its brand A line of cigarettes outsells its brand B by 8%. If it is found that 42 out of a sample of 200 smokers prefer brand A and 18 out of another sample of 100 smokers prefer brand B. test whether the 8% difference is valid claim.

SMALL SAMPLES

t-test:- If a sample with less than 30 items then we use t-test for testing of hypothesis. In this we estimate the population variance from the sample values.

i.e., If $x_1, x_2 \ldots x_n$ are the values in the sample with sample mean $\bar{x}$ and standard deviation s drawn from a population with mean μ and S.D. σ.

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

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

is a random variable having the t-distribution with $n-1$ d.f. and with p.d.f.

$$f(t) = y_0\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$ where $v = n-1$ and y_0 is a constant.

APPLICATIONS OF THE t-DISTRIBUTION:-

The t-distribution has a wide number of applications in Statistics, some of them are given below:

1. To test the significance of the sample mean, when population variance is not given.
2. To test the significance of the mean of the sample i.e., to test if the sample mean differs significantly from the population mean.
3. To test the significance of the difference between two sample means or to compare two samples
4. To test the significance of an observed sample correlation coefficient and sample regression coefficient.

① A random sample of size 25 from a normal population has the mean $\bar{x}=47.5$ and the standard deviation $S=8.4$. Does this information tend to support or refuse the claim that the mean of the population is $\mu=42.5$?

Sol: Given n = The size of the sample = 25

$\bar{x}$ = The mean of the sample = 47.5

μ = The population mean = 42.5

S = S.D of Sample = 8.4

we have t-distribution

$$t=\frac{\bar{x}-\mu}{S/\sqrt{n}}=\frac{47.5-42.5}{8.4/\sqrt{25}}$$

$$=\frac{5\sqrt{25}}{8.4}=2.98$$

This value of t has 24 d.f. ($\because \nu = n-1 = 25-1=24$)

From the table of t-distribution for $\nu=24$ with $\alpha=0.005$ is 2.797. we conclude that the information given in the data of this example tend to refuse the claim that the mean of the population is $\mu=42.5$ (i.e, μ cannot be 42.5)

F-test:- F-test can be used for test of equality of population variances.

Suppose we want to test whether two independent samples x_i $(i=1,2,\dots,n_1)$, y_j $(1,2,\dots,n_2)$ has been drawn from the normal population with same variances σ^2. we consider $H_0: \sigma_x^2=\sigma_y^2=\sigma^2$.

i.e., Population variances are equal.

To test above H_0, the test statistic is given by

$$F=\frac{S_1^2}{S_2^2}\sim F((n_1-1),(n_2-1))\text{ d.f.}$$

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

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

Conclusion:- Compare calculated F with $F_{(n_1-1),(n_2-1)}$ d.f.

at certain l.o.s. (5% or 1%)

if $F_{cal} < F_{tab}$, $F_{cal} > F_{tab}$

accept H_0 Reject H_0.

χ^2-distribution:- Let s^2 be the sample variance of size n, taken from a normal population having the variance σ^2.

Then $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^2}$ $\left[\because S^2 = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}\right]$

is the value of a random variable having the χ^2-distribution with $\nu = n-1$ d.f.

APPLICATIONS OF χ^2 DISTRIBUTION:-

(1) To test the goodness of fit.

(2) To test the independence of attributes

(3) To test the homogeneity of independent estimation of the population variance.

(4) To test the homogeneity of independent estimation of the population correlation coefficient.

(1) For an F-distribution, find

(a) $F_{0.05}$ with $\nu_1 = 7$ and $\nu_2 = 15$

(b) $F_{0.01}$ with $\nu_1 = 24$ and $\nu_2 = 19$

c) $F_{0.95}$ with $\nu_1=19$ and $\nu_2=24$

d) $F_{0.99}$ with $\nu_1=28$ and $\nu_2=12$

Sol:

a) From table $F_{0.05}$ with $\nu_1=7$ and $\nu_2=15$ is 2.71

b) $F_{0.01}$ with $\nu_1=24, \nu_2=19$ is 2.92

c) $F_{0.95}(19,24) = \dfrac{1}{F_{0.05}(24,19)} = \dfrac{1}{2.11} = 0.473$

d) $F_{0.99}(28,12) = \dfrac{1}{F_{0.01}(12,28)} = \dfrac{1}{2.90} = 0.34482$

TEST OF SIGNIFICANCE FOR SMALL SAMPLES :-

There are three important tests for, that is

(i) Students 't' Test.

(ii) F-test

(iii) χ^2-test.

① Students t-test: single mean

The Students t is defined by the test

Statistic $\quad t = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n-1}}$

$\bar{x}$ = mean of a sample

n = size of the sample

σ = S.D. of the sample

μ = mean of the population supposed to be normal

where $\sigma^2 = \sum_{i=1}^{n} \dfrac{(x_i-\bar{x})^2}{n}$.

① The average breaking strength of the steel rods is specified to be 18.5 thousand pounds. To test this sample of 14 rods were tested. The mean and standard deviations obtained were 17.85 and 1.955 respectively. Is the result of experi- -ment significant?

Sol: Given sample size $n = 14$

Sample mean $\bar{x} = 17.85$

S.D. (S) $= 1.955$

population mean, $\mu = 18.5$

Degrees of freedom $= n-1 = 13$

1. Null Hypothesis H_0: The resut of the expeeiment is not significant.

2. Alternative Hypothesis H_1: $\mu \neq 18.5$

3. Level of significance $\alpha = 0.05$

4. The test statistic is

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

$$= \frac{17.85-18.5}{1.955/\sqrt{13}} = \frac{0.65}{0.542} = -1.199$$

$\therefore |t| = 1.199$

i.e., calculated $t = 1.199$

Tabulated t at 5% l.o.s. for 13 d.f. for two tailed test $= 2.16$

Since calculated value $t <$ table value t, we accept the Null Hypothesis H_0 at 5% level and conclude that the result of the expeeiment is not significant.

② A random sample of six steel beams has a mean compressive strength of 58,392 P.S.i (Pounds per square inch) with a S.D. of 648 p.s.i. use this information and l.o.s $\alpha = 0.05$ to test whether the true avg compressive strength of the steel from which the sample came is 58,000 P.S.i. Assume normality.

Sol. we have

n = sample size (number of steel beams) $= 6 < 30$

$\therefore$ The sample is small.

$\bar{x}$ = Sample mean (average compressive strength) = 58392 psi.

S = S.D. of six Beams = 648 psi

Degrees of freedom $= n-1 = 6-1 = 5$

In this problem σ is known and $n < 30$. Hence we use t-distribution.

1. Null Hypothesis: $H_0: \mu = 58000$

2. Alternative Hypothesis $H_1: \mu \neq 58000$

3. l.o.s. $\alpha = 0.05$

4. Critical region: since A.H. is of the type '$\neq$' the test is two-tailed and the critical region is $-3.365 < t < 3.365$

5. The test statistic is $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$

$$= \frac{58392-58000}{648/\sqrt{5}}$$

$$= 1.353$$

Since $t = 1.353 < 3.365 = t_{\alpha/2}$, we accept the null hypothesis H_0.

Hence the average compressive strength of the steel beam is not equal to 58000 psi.

③ A sample of 100 iron bars is said to be drawn from a large number of bars whose lengths are normally distributed with mean 4 feet and S.D. 6ft. If the sample mean is 4.2 feet. Can the sample be regarded as a truly random sample?

④ A sample of 155 members has a mean 67 and S.D. 5.2. In this sample has been taken from a large population of mean 70.

Problems related to student's t-Test (when S.D. of the sample is not given directly)

① A random sample of 10 boys had the following I.Q's: 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100 (a) Do these data support the assumption of a population mean I.Q. of 100? (b) Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.

Sol: (a) Here S.D. and mean of the sample is not given directly. we have to determine these S.D. and mean as follows.

Mean $\bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2$.

x	$x-\bar{x}$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
972		1833.60

We know that $s^2 = \frac{1}{n-1}\sum_{1}(x_i-\bar{x})^2$

$$= \frac{1833.60}{9}$$

$\therefore$ Standard deviation $s = \sqrt{203.73}$

$$= 14.27$$

1. Null Hypothesis H_0: The data support the assumption of a population mean I.Q. of 100 in the population

2. Alternative Hypothesis: $H_1: \mu \neq 100$ (two-tailed test)

3. L.O.S. $\alpha = 0.05$

4. The test statistic is

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{97.2-100}{14.72/\sqrt{10}} = -0.62$$

$\therefore |t| = 0.62$ i.e., calculated value of $t = 0.62$

Tab. value of t for (10-1) d.f. i.e., 9 d.f. at 5% l.o.s is 2.26.

Since cal. value t > tab value t. we accept the null hypothesis H_0, i.e., the data support the assumption of mean I.Q. of 100 in the population.

(b) The 95% confidence limits are given by $\bar{x} \pm t_{0.05} s/\sqrt{n}$

$$= 97.2 \pm 2.26 \times 4.512 = 107.4 \text{ and } 87$$

$\therefore$ The 95% confidence limits within which the mean I.Q. values of sample of 10 boys will be (87, 107.40)

(2) The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the avg height is greater than 64 inches? Test at 5% l.o.s. assuming that for 9 d.f. ($t = 1.833$ at $\alpha = 0.05$).

we first compute the sample means and standard deviations.

$\bar{x}$ = mean of first sample

$= \frac{1}{7}(28+30+32+33+33+29+34)$

$= \frac{1}{7}(219) = \cancel{29.85}\ 31.286$

$\bar{y}$ = Mean of second sample

$= \frac{1}{6}(29+30+30+24+27+29)$

$= \frac{1}{6}(169) = 28.16$

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
28	-3.286	10.8	29	0.84	0.7056
30	-1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	-4.16	17.3056
33	1.714	2.94	27	-1.16	1.3456
29	-2.286	5.226	29	0.84	0.7056
34	2.714	7.366			
219	31.4358	169			26.8336

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

$= \frac{1}{11}[31.4358 + 26.8336]$

$= \frac{1}{11}[58.2694]$

$= 5.23.$

$\therefore s = \sqrt{5.23}$

$= 2.3$

STUDENT'S t-TEST FOR DIFFERENCE OF MEANS :-

To test the significance difference between the sample means $\bar{x}$ and $\bar{y}$ of two independent samples of sizes n_1 and n_2 with the same variance, we use statistic.

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim n_1+n_2-2 \text{ d.f.}$$

where $\bar{x} = \frac{\sum x_i}{n}$, $\bar{y} = \frac{\sum y_i}{n}$

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

or

$$S^2 = \frac{1}{n_1+n_2-2}\left[(n_1-1)s_1^2+(n_2-1)s_2^2\right]$$

But for large samples, the following estimate is used.

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

where s_1, s_2 are two sample standard deviations

(1) Two horses A and B were tested according to the time (in seconds) to run a particular track with the following result.

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29.	

Test whether the two horses have the same running capacity.

Sol:- Given $n_1 = 7$, $n_2 = 6$

(27)

1. Null Hypothesis $H_0: \mu_1 = \mu_2$

2. Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

3. Level of significance, $\alpha = 0.05$

4. The test statistic is

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

$$= \frac{31.286 - 28.16}{(2.3)\sqrt{\frac{1}{7} + \frac{1}{6}}} = 2.443.$$

Tabulated value of t for $7+6-2 = 11$ d.f at 5% l.o.s. is 2.2.

since cal. value $t >$ Tab. value t, we reject the null hypothesis H_0. and conclude that both horses A and B do not have the same running capacity.

(2) Find the maximum difference that we can expect with probability 0.95 between the means of samples of sizes 10 and 12 from a normal population if their S.D. are found to be 2 and 3 respectively.

Sol: we have $n_1 = 10$, $n_2 = 12$, $S_1 = 2$ & $S_2 = 3$

$$\therefore S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = \frac{1}{10+12-2}\left[10(2)^2 + 12(3)^2\right]$$

$$= 7.4$$

$$\therefore S = \sqrt{7.4} = 2.72$$

1. Null Hypothesis $H_0: \mu_1 = \mu_2$

2. Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

3. Level of significance $\alpha = 0.05$

4. The test statistic is

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

$$\Rightarrow |\bar{x}-\bar{y}| = t_{\alpha}.S.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= (2.086)(2.72)\sqrt{\frac{1}{10}+\frac{1}{12}}$$

$$= 2.43$$

[$\because$ Tabled value of t for $10+12-2=20$ d.f at 5% l.o.s is 2.086 (two-tailed)]

Hence the maximum difference between the means is 2.43.

3) To compare two kinds of bumper guards, 6 of each kind were mounted on a car and then the car was run into a concrete wall. The following are the costs of repair.

Guard 1	107	148	123	165	102	119
Guard 2	134	115	112	151	133	129

Use the 0.01 l.o.s. to test whether the difference between two sample means is significant.

4) The IQs (Intelligence quotients) of 16 students from one area of a city showed a mean of 107 with a S.D. of 10, while the IQs of 14 students from another area of the city showed a mean of 112 with a S.D. of 8. Is there a significant difference between the IQs of the two groups at a 0.05 l.o.s?

(2) Memory capacity of 10 students were tested before and after training state whether the training was effective or not from the following scores.

Before training	12	14	11	8	7	10	3	0	5	6
After training	15	16	10	7	5	12	10	2	3	8

(3) Scores obtained in a shooting competition by 10 soldiers before and after intensive training are given below:

Before	62	24	57	55	63	54	56	68	33	43
After	70	38	58	58	56	67	68	75	42	38

Test whether the intensive training is useful at 5% l.o.s.

SNEDECOR'S F-TEST OF SIGNIFICANCE

Test for Equality of Two population variances:-

Let two independent random samples of sizes n_1 and n_2 be drawn two normal populations.

To test the hypothesis that the two population variance σ_1^2 and σ_2^2 are equal.

The estimates of σ_1^2 and σ_2^2 are given by

$$S_1^2 = \frac{n_1 s_1^2}{n_1-1} = \frac{\sum(x_i-\bar{x})^2}{n_1-1} \text{ and } S_2^2 = \frac{n_1 s_2^2}{n_2-1} = \frac{\sum(y_i-\bar{y})^2}{n_2-1}$$

where s_1^2 and s_2^2 are the variances of the two samples.

Assuming that Ho is true.

the test statistic $F = \frac{S_1^2}{S_2^2}$ or $\frac{S_2^2}{S_1^2}$ according as $S_1^2 > S_2^2$ or $S_2^2 > S_1^2$ follows F-distribution with (n_1-1, n_2-1) d.f.

① Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins, show the sample standard deviations of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test hypothesis that the true variances are equal.

Sol: Let the Null Hypothesis be

$H_0: \sigma_1^2 = \sigma_2^2$ i.e., the variances of sources are equal.

The Alternative Hypothesis is $H_1: \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 11$, $n_2 = 9$, $s_1 = 0.8$, $s_2 = 0.5$

Here samples S.D's are given

we have to determine the population variances S_1^2 and S_2^2

by using the relations.

$$n_1 s_1^2 = (n_1 - 1) S_1^2 \quad \text{and} \quad n_2 s_2^2 = (n_2 - 1) S_2^2$$

$$\Rightarrow S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11 \times (0.8)^2}{10} = 0.704$$

$$\Rightarrow S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{9 \times (0.5)^2}{8} = 0.281$$

The test statistic is $F = \frac{S_1^2}{S_2^2} = \frac{0.704}{0.281}$

$= 2.5 \quad (\because S_1^2 > S_2^2)$

$\therefore$ calculated $F = 2.5$

Tab. value of F for (10, 8) d.f at 5% l.o.s. is 3.35

Since calculated F < tabulated F, we accept the null hypothesis H_0 at 5% l.o.s. with (10, 8) d.f and conclude that the variances of the two populations is the same and therefore, the two samples have the same variance.

PAIRED - SAMPLE t-TEST

If $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the pairs of sales data before and after the sales promotion in a business concern. we apply paired t-test to examine the significance of the difference of the two situation.

Let $d_i = x_i - y_i$ or $y_i - x_i$ for $i = 1, 2, 3 \ldots n$

The test statistic for n paired observations (which are dependent) by taking the differences $d_1, d_2, \ldots d_n$ of the paired data.

$$t = \frac{\bar{d} - \mu}{S/\sqrt{n}} = \frac{\bar{d}}{S/\sqrt{n}} \quad (\because \mu = 0)$$

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

or $S^2 = \frac{\sum d^2 - n(\bar{d})^2}{n-1}$ or $\frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right]$

are the mean and variance of the differences $d_1, d_2, \ldots d_n$ respectively and μ is the mean of the population of differences.

The above statistic follows student's t-distribution with $(n-1)$ d.f.

① The Blood Pressure of 5 women before and after intake of a certain drug are given below:

Before	110	120	123	132	125
After	120	118	125	136	121

Test whether there is significant change in Blood pressure at 1% l.o.s.

Sol: Let the Null hypothesis be $H_0: \mu_1 = \mu_2$
i.e., there is no significant difference in blood pressure.

before and after intake of drug.

The the Alternative Hypothesis $H_1: \mu_1 < \mu_2$

Assuming that H_0 is true, the test statistic is

$$t = \frac{\bar{d}}{S/\sqrt{n}} \quad \text{where } \bar{d} = \frac{\Sigma d}{n}, \; d = y - x$$

$$\text{and } S^2 = \frac{\Sigma (d-\bar{d})^2}{n-1} = \frac{\Sigma d^2 - (\bar{d})^2 \times n}{n-1}$$

calculations for $\bar{d}$ and S

Womens	B.P. before intake of drug (x)	B.P. after intake of drug (y)	$d = y - x$	d^2
1	110	120	10	100
2	120	118	-2	4
3	123	125	2	4
4	132	136	4	16
5	125	121	-4	16
			$\Sigma d = 10$	$\Sigma d^2 = 140$

$$\therefore \bar{d} = \frac{\Sigma d}{n} = \frac{10}{5} = 2 \text{ and } S^2 = \frac{140 - (2)^2 \times 5}{4} = 30$$

$$\therefore S = \sqrt{30}$$

$$\therefore t = \frac{\bar{d}}{S/\sqrt{n}} = \frac{2}{\sqrt{30}/\sqrt{5}} = 0.82$$

Degree of freedom $= n - 1 = 5 - 1 = 4$

Thus $t = 0.82 < 4.6$ at 1% l.o.f with 4 d.f

Since the calculated value of $t <$ the table value with 4 d.f. at 1% level, we accept H_0 at 1% l.o.s. and conclude that there is no significant change in Blood pressure after intake of a certain drug.

① Four methods are under development for making discs of a super conducting material. Fifty discs are made by each method and they are checked for super conductivity when cooled with liquid

	1st method	2nd method	3rd method	4th method
Superconductors	31	42	22	25
Failures	19	8	28	25

Test the significant difference between the proportions of super conductor at 5% level.

② From the following data, find whether there is ~~any~~ any significant liking in the habit of taking soft drinks among the categories of employees.

	Employees		
Soft Drinks	clerks	Teachers	officers
pepsi	10	25	65
Thumsup	15	30	65
Fanta	50	60	30

CHI-SQUARE TEST FOR POPULATION VARIANCE :-

Suppose that a random sample x_i $(i=1,2,\ldots n)$ is drawn from a normal population with mean μ and variance σ^2. To test the hypothesis that the population variance σ^2 has a specified value σ_0

The test statistic is $\chi^2 = \sum \frac{(x_i-\bar{x})^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2}$

where s^2 = sample variance $= \frac{\sum(x_i-\bar{x})^2}{n}$ and $ns^2 = (n-1)S^2$

Assuming that H_0 is true, the test statistic χ^2 follows χ^2-dist. with $(n-1)$ d.f.

Conclusion: If cal. value of χ^2 > tab value, we reject H_0 otherwise accept H_0.

① A firm manufacturing rivets wants to limit variations in their length as much as possible. The lengths (in cms) of 10 rivets manufactured by a new process are

2.15	1.99	2.05	2.12	2.17
2.01	1.98	2.03	2.25	1.93

Examine whether the new process can be considered superior to the old if the old population has S.D. 0.145 cm?

Sol: we have

$n = 10$, $\bar{x} = \frac{\sum x_i}{n} = \frac{20.68}{10} = 2.068$

$S^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{0.09096}{10} = 0.0091$

and $\sigma_0 = 0.145$

1. Null Hypothesis $H_0 : \sigma^2 = \sigma_0^2$
2. Alternative , $H_1 : \sigma^2 > \sigma_0^2$
3. l.o.s. $\alpha = 0.05$
4. Assuming that H_0 is true, the test statistic is

$$\chi^2 = \frac{ns^2}{\sigma_0^2} = \frac{0.09096}{(0.145)^2} = 4.236 \quad (\sigma_0 = 0.145)$$

$\therefore$ calculated $\chi^2 = 4.3$

d.f. $= n-1 = 10-1 = 9$

Tab. value χ^2 at 9 d.f at 5% l.o.s is 16.919

Since cal. value $\chi^2 <$ tab χ^2, we accept the null Hypothesis H_0

i.e., The new process cannot be considered superior to the old process.

② A random sample of size 20 from a normal population gives a mean of 42 and a variance of 25. Test the hypothesis that the population standard deviation is 8 at 5% level of significance.

$\therefore$ calculated $F = 4.075$

Tabulated value of F for $(n_2-1, n_1-1) = (5, 4)$ d.f at 5%. 1.0.5 is 6.26. since calculated F < tabulated F, we accept the null Hypoesis H_0 i.e., the variances are equal.

i.e., $\sigma_1^2 = \sigma_2^2$

③ Two random samples gave the following results

Sample	Size	Sample mean	Sum of squares of Deviations from the mean
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population.

④ In one sample of 10 observations, the sum of the squares of the deviations of the sample values from sample mean was 120 and in the other sample of 12 observations, it was 314. Test whether. Test whether the difference is significant at 5% level?

CHI-SQUARE (χ^2) TEST :-

Def:- If a set of events $A_1, A_2 \ldots A_n$ are observed to occur with frequencies $O_1, O_2 \ldots O_n$ respectively and according to probability rules $A_1, A_2 \ldots A_n$ are expected to occur with frequencies $E_1, E_2, \ldots E_n$ respectively with $O_1, O_2 \ldots O_n$ are called observed frequencies and $E_1, E_2 \ldots E_n$ are called expected frequencies.

If O_i $(i=1,2..n)$ is a set of observed (experimental) frequencies and E_i $(i=1,2 \ldots n)$ is the corresponding set of expected (theoretical) frequencies, then χ^2 is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \text{ with } (n-1) \text{ d.f.}$$

② The nicotine contents in milligrams in two samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	-
Sample B	27	30	28	31	22	36

Can it be said that the two samples have come from the same normal population.

Sol: Given $n_1 = 5$, $n_2 = 6$

Calculation for mean's and S.D's of the samples.

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
24	0.6	0.36	27	-2	4
27	2.4	5.76	30	1	1
26	1.4	1.96	28	-1	1
21	3.6	12.96	31	2	4
25	0.4	0.16	22	-7	49
			36	7	49
123		21.2	174		108

$$\therefore \bar{x} = \frac{\Sigma x}{n_1} = \frac{123}{5} = 24.6. \quad \bar{y} = \frac{\Sigma y}{n_2} = \frac{174}{6} = 29$$

$$\Sigma(x_i-\bar{x})^2 = 21.2 \qquad \Sigma(y_i-\bar{y})^2 = 108$$

$$\therefore S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} = \frac{21.2}{4} = 5.3 \text{ and}$$

$$S_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1} = \frac{108}{5} = 21.6$$

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

The test statistic is $F = \frac{S_2^2}{S_1^2} = \frac{21.6}{5.3}$

$= 4.075$

Sol: Expected frequency of accidents each week $= \frac{100}{10} = 10$.

Null Hypothesis H_0: The accident conditions were the same during the 10 week period.

Alternative Hypothesis: The accidents conditions are different during the 10 week period.

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)$	$\frac{(O_i-E_i)^2}{E_i}$
12	10	2	0.4
8	10	-2	0.4
20	10	10	10.0
2	10	-8	6.4
14	10	4	1.6
10	10	0	0
15	10	-5	2.5
6	10	-4	1.6
9	10	-1	0.1
4	10	-6	3.6
100	10		
100	100		26.6

Now $\chi^2 = \sum \frac{(O_i-E_i)^2}{E_i} = 26.6$.

i.e., calculated $\chi^2 = 26.6$

Here $n = 10$ observations are given

$\therefore$ D.f. $= n-1 = 10-1 = 9$

Tab. $\chi^2 = 16.9$

Since cal value $\chi^2 >$ Tab. χ^2. therefore, the Null Hypothesis is rejected and conclude that the accident conditions were not the same during the 10 week period.

(31)

χ^2 is used to test whether differences between observed and expected frequencies are significant.

Chi-square Distribution is an important continuous prob. distribution and it is used both large and small tests. In chi-square tests, χ^2-distribution is mainly used.

(i) To test the goodness of fit

(ii) to test the independence of Attributes.

(iii) To test the population has a specified value of the variance σ^2.

① χ^2 TEST AS A GOODNESS OF FIT :-

Let $O_1, O_2 \ldots O_n$ be a set of observed frequencies and $E_1, E_2 \ldots E_n$ the corresponding set of expected frequencies.

Then the test statistic χ^2 is given by.

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i-E_i)^2}{E_i}\right] = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \cdots + \frac{(O_n-E_n)^2}{E_n}$$

Assuming Ho is true, the test statistic χ^2 follows chi-square distribution with $(n-1)$ d.f.

Conclusion: If the calculated value of χ^2 > tabulated value of χ^2 at α level, the Null Hypothesis Ho is rejected, otherwise Ho is accepted.

① The number of automobile accidents per week in a certain community are as follows: 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

① on the basis of information given below about the treatement of 200 patients suffering from a disease, State whethe the new treatement is comparatively superior to the conventi-
-onal treatement.

	Favourable	Not favourable	Total
New	60	30	90
conventional	40	70	110

Sol:. Null Hypothesis H_0: No difference between new and conventional treatement (or) New and conventional trea-
-tement are independent.

The number of degrees of freedom is $(2-1)(2-1)=1$

Expected frequencies are given in the table.

$$\text{Expected frequency} = \frac{\text{Row total} \times \text{column total}}{\text{Grand total}}$$

$\frac{90\times100}{200}=45$	$\frac{90\times100}{200}=45$	90
$\frac{100\times110}{200}=55$	$\frac{100\times110}{200}=55$	110
100	100	200

Caluclation of χ^2

observed frequen (O_i)	Expected frequen (E_i)	$(O_i-E_i)^2$	$(O_i-E_i)^2/E_i$
60	45	225	5
30	45	225	5
40	55	225	4.09
70	55	225	4.09
200	200		18.18

(32)

② A die is thrown 264 times with the following results. Show that the die is biased. (Given $\chi^2_{0.05} = 11.07$ for 5 d.f.)

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

③ A pair of dice are thrown 360 times and the frequency of each sum is indicated below:

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the chi-square test at 0.05 l.o.s.

② CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES:-

The test statistic

$$\chi^2 = \sum\left[\frac{(O_i - E_i)^2}{E_i}\right] \sim (r-1)(c-1) \text{ d.f.}$$

where expected frequency E_i of any cell

$$= \frac{\text{Row total} \times \text{column total}}{\text{Grand total.}}$$

(or)

The value of χ^2 is given by $\chi^2 = \dfrac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$

where ~~a+~~ $N = a+b+c+d$ with d.f. $= (2-1)(2-1) = 1$

$\frac{60\times40}{150}=16$	$\frac{30\times40}{150}=8$	$\frac{60\times40}{150}=16$	40
$\frac{60\times50}{150}=20$	$\frac{30\times50}{150}=10$	$\frac{60\times50}{150}=20$	50
$\frac{60\times60}{150}=24$	$\frac{30\times60}{150}=12$	$\frac{60\times60}{150}=24$	60
60	30	60	150

Calculation of χ^2

Observed frequency (O_i)	Expected Frequency (E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
15	16	1	0.0625
5	8	9	1.125
20	16	16	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.042
15	12	9	0.~~7~~5
20	24	16	0.666
Total			3.6458

$$\therefore \chi^2 = \sum \frac{(O_i-E_i)^2}{E_i} = 3.6458$$

i.e., calculated $\chi^2 = 3.6458$

Table χ^2 for $(3-1)(3-1) = 4$ d.f at 5% l.o.s. is 9.488

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., The hair colour and eye colour are independent

i.e., The hair colour and eye colour are not associated.

Scanned with CamScanner

$$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 18.18$$

Tab value χ^2 for 1.d.f at 5%.1.0.5 is 3.841.

Since cal. χ^2 > tab χ^2 we reject the null Hypothesis Ho.

i.e., new and conventional treatement are not independent. The new treatement is comparatively superior to conventional treatement.

② The following table gives the classification of 100 workers according to sex and nature of work. Test whether the nature of work is independent of the sex of the worker.

	Stable	unstable	Total
Males	40	20	60
Females	10	30	40
Total	50	50	100.

③ Given the following contingency table for hair colour and eye colour. Find the value of χ^2. Is there good association between the two?

		Haircolour			
		Fair	Brown	Black	Total
Eyecolour	Blue	15	5	20	40
	Grey	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

Sol: Null Hypothesis Ho: The two attributes, hair and eye colour are independent

Table of expected frequencies,

$\frac{60\times40}{150}=16$	$\frac{30\times40}{150}=8$	$\frac{60\times40}{150}=16$	40
$\frac{60\times50}{150}=20$	$\frac{30\times50}{150}=10$	$\frac{60\times50}{150}=20$	50
$\frac{60\times60}{150}=24$	$\frac{30\times60}{150}=12$	$\frac{60\times60}{150}=24$	60
60	30	60	150

Calculation of χ^2

Observed frequency (O_i)	Expected Frequency (E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
15	16	1	0.0625
5	8	9	1.125
20	16	16	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.042
15	12	9	0.75
20	24	16	0.666
Total			3.6458

$$\therefore \chi^2 = \sum \frac{(O_i-E_i)^2}{E_i} = 3.6458$$

i.e., calculated $\chi^2 = 3.6458$

Table χ^2 for $(3-1)(3-1) = 4$ d.f at 5% l.o.s. is 9.488

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., The hair colour and eye colour are independent

i.e., The hair colour and eye colour are not associated