x: 16 12 18 4 3 10 5 12

y: 87 88 89 68 78 80 75 83

Q. x : 25 28 35 32 31 36 29 38. 34 32

y : 43 46 49 41 36 32 31 30 33 39

$\Rightarrow$ $y = 59.248 - 0.664x$.

$x = 40.892 - 0.234y$

Q. Obtain the equations of two lines of regression for the following data

x: 65 66 67 67 68 69 70 72

y: 67 68 65 68 72 72 69 71

Also obtain the estimate of x for $Y = 70$

$\Rightarrow$ $Y = 0.665X + 23.78$

$X = 0.54Y + 30.74$

$Y = 70,\ X = 68.54$

Q. In a partially destroyed laboratory, record of an analysis of correlation data, the following results only for Legible.

$Var(x) = 9$, Regression equations: $8X - 10Y + 66 = 0$ and $40X - 18Y = 214$. What are: (i) the means of x and Y (ii) the correlation coefficient between X and Y and (iii) standard deviation of Y?

$\Rightarrow$ ① $\bar{X} = 13,\ \bar{Y} = 17$

② $b_{yx} = 4/5,\ b_{xy} = 9/20$

$r = 0.6$

③ $b_{yx} = r\dfrac{\sigma_y}{\sigma_x} \Rightarrow \sigma_y = 4$

Joint distribution (discrete case).

Let X and Y be two discrete random variables on the same sample space S. Suppose that X can assume the values $x_1, x_2, \ldots x_n$ and Y can assume the values $y_1, y_2 \cdots y_m$. The probability distribution or joint probability mass function of X and Y is defined by

$f(x_i, y_j) = P(X = x_i, Y = y_j);$ for $i = 1, 2, \cdots n$

$j = 1, 2, \cdots m$

Properties :- ① $f(x_i, y_j) \geq 0$

② $\sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i, y_j) = 1$

The probability that $X = x_i$ is given by

$P(x = x_i) = f_X(x_i) = \sum_{j=1}^{m} f(x_i, y_j)$; for $i = 1, 2, \cdots n$

and is called the marginal probability function of X or marginal distribution of X. Similarly, the probability that $Y = y_j$ is given by $P(Y = y_j) = f_Y(y_j) = \sum_{i=1}^{n} f(x_i, y_j)$, for $j = 1, 2, \ldots m$ and is called this marginal probability function of Y or marginal distribution of Y.

The joint probability distribution of X and Y can be represented in the form of a rectangular table as follows.

$X \backslash Y$	y_1	y_2	$\ldots$	y_m	$f_X \downarrow$
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\ldots$	$f(x_1, y_m)$	$f_X(x_1)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\ldots$	$f(x_2, y_m)$	$f_X(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	$f(x_n, y_1)$	$f(x_n, y_2)$	$\ldots$	$f(x_n, y_m)$	$f_X(x_m)$
$f_Y \rightarrow$	$f_Y(y_1)$	$f_Y(y_2)$	$\ldots$	$f_Y(y_m)$	1

The joint distribution function of X and Y is defined by

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f(u,v)$$

⇒ Two random variables X and Y are independent if $F(x,y) = F_x(u) \cdot F_y(y)$

⇒ The conditional probability density of X given Y is

$$f(x/y) = P[(X = x) | (Y = y)] = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x,y)}{f_y(y)}$$

and its conditional probability density of Y given X is $f(y/x) = P(Y = y | X = x)$

$$= \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x,y)}{f_X(x)}$$

Note:- If X and Y are independent, then

$$f(x/y) = \frac{f(x,y)}{f_y(y)} = \frac{f_x(x) f_y(y)}{f_y(y)} = f_x(x)$$

Similarly,

$$f(y/x) = \frac{f(x,y)}{f_x(x)} = \frac{f_x(x) f_y(y)}{f_x(x)} = f_y(y)$$

Q The joint pmf of X and Y are given by $f(x,y) = c(x + 2y)$, $x = 0, 1, 2$ and $y = 0, 1, 2, 3$.

a) find c b) find $P(X < 1)$, $P(Y \leq 2)$, $P(X \leq 1, Y < 2)$, $P(X \leq 1 / Y \leq 2)$, $P(X + Y \leq 3)$ and $P(X^2 + Y^2 \leq 4)$

(c) find the marginal pmf and conditional probablity distribution.

→ The joint pmf can be represented as.

x \ y	0	1	2	3	$f_x \downarrow$	
0	0	2c	4c	6c	12c	$f_x(0)$
1	c	3c	5c	7c	16c	$f_x(1)$
2	2c	4c	6c	8c	20c	$f_x(2)$
	3c	9c	15c	21c		
$f_y \rightarrow$	$f_y(0)$	$f_y(1)$	$f_y(2)$	$f_y(3)$		

(a) since $\sum_i \sum_j f(x_i, y_j) = 1$

$$\sum_{x=0}^{2} \sum_{j=0}^{3} c(x+2y) = 1$$

$\Rightarrow 48c = 1$

$\Rightarrow c = 1/48$

(b) $P(X \leq 1) = P(X = 0, 1)$

$= P(X=0) + P(X=1)$

$= \sum_{j=0}^{3} f(0, j) + \sum_{j=0}^{} f(1, j)$

$= (0 + 2c + 4c + 6c) + (c + 3c + 5c + 7c)$

$= 12c + 16c = 18c$

$= 18/48 = 7/12$

$P(Y \leq 2) = P(Y = 0, 1, 2)$

$= P(Y=0) + P(Y=1) + P(Y=2)$

$= \sum_{i=0}^{2} f(i, 0) + \sum_{i=0}^{2} f(i, 1) + \sum_{i=0}^{2} f(i, 2)$

$= (0 + c + 2c) + (2c + 3c + 4c) + (4c + 5c + 6c)$

$= 3c + 9c + 15c$

$= 27c$

$= 27/48 = 9/16$

0 1

0 1 2

$P(X \leq 1, Y \leq 2) = \sum_{i=0}^{1} \sum_{j=0}^{2} f(x_i, y_j)$

$= f(0,0) + f(0,1) + f(0,2) + f(1,0) + f(1,1) + f(1,2)$

$= 0 + 2c + 4c + c + 3c + 5c = 15c$

$= 15/48 = 5/16$

$$P(X \leq 1 \mid Y \leq 2) = \frac{P(X \leq 1, Y \leq 2)}{P(Y \leq 2)}$$

$$= \frac{5/16}{9/16} = 5/9$$

0 1 2

0 1 2 3

p.no.26

$P(X+Y \leq 3) = f(0,0) + f(0,1) + f(0,2) + f(0,3)$
$+ f(1,0) + f(1,1) + f(1,2) +$
$f(2,0) + f(2,1)$

$= 0 + 2c + 4c + 6c + c + 3c + 5c +$
$2c + 4c$

$= 27c$

0 1 2

0 1 2 3

$= 27/48 = 9/16$

$P(x^2+y^2 \le 4) = f(0,0)+f(0,1)+f(0,2)+$
$f(1,0)+f(1,1)+f(1,2)$
$= 0+2c+4c+c+3c+2c$
$= 12c$
$= \frac{12}{48}$
$= 1/4.$

c) The marginal pmf fx and fy are given by

$x =$	0	1	2	(Row totals)
$f_x =$	12/48	16/48	20/48	

$Y =$	0	1	2	3	(column totals)
$f_y =$	3/48	9/48	15/48	2/48.	

Conditional distribution.

$f(x/y)$ is calculated as each element of a column is divided by the corresponding column sum.

$$P(x=i \mid Y=j) = \frac{P(x=i, y=j)}{P(Y=j)} = \frac{f(x_i, y_j)}{\sum_i f(x_i, y_j)}$$

$f(x/y)$:-

X \ Y	0	1	2	3
0	0/3/48 = 0	2/48/9/48 = 2/9	4/48 / 15/48 = 4/15	6/48 / 21/48 = 6/21
1	1/48/3/8 = 1/3	3/48/9/8 = 3/9	5/48 / 15/48 = 5/15	7/48 / 21/48 = 7/21
2	2/48/3/48 = 2/3	4/48 / 9/48 = 4/9	6/48 / 15/48 = 6/15	8/48 / 21/48 = 8/21

Similarly, $f(y/x)$ is calculated as each element of a row is divided by the corresponding row sum, since

$$P(Y=j/X=i) = \frac{P(X=i, Y=j)}{P(X=i)} = \frac{f(x_i, y_j)}{\sum_j f(x_i, y_j)}$$

x \ y	0	1	3
0	0/12/48 = 0	4/48/12/48 = 4/12	6/48/12/48 = 6/12
1	2/48/16/48 = 1/16	5/48/16/48 = 5/16	7/48/16/48 = 7/16
2	2/48/20/48 = 2/20	6/48/20/48 = 6/20	8/48/20/48 = 8/20

$x = 1, y = 1$

Note:- $f_x(1) = 16/48$; $f_y(1) = 9/48$; $f(1,1) = 3/48$

$\neq f_X(1) \times f_Y(1)$

∴ X and Y are not independent

$f(Y/X)$:-

X \ Y	0	1	2	3
0	$\frac{0}{48}/\frac{12}{48} = 0$	$\frac{2}{48}/\frac{12}{48} = 2/12$	$\frac{4}{48}/\frac{12}{48} = 4/12$	$\frac{6}{48}/\frac{12}{48} = 6/12$
1	$\frac{1}{48}/\frac{16}{48} = 1/16$	$\frac{3}{48}/\frac{16}{48} = 3/16$	$\frac{5}{48}/\frac{16}{48} = 5/16$	$\frac{7}{48}/\frac{16}{48} = 7/16$
2	$\frac{2}{48}/\frac{20}{48} = 2/20$	$\frac{4}{48}/\frac{20}{48} = 4/20$	$\frac{6}{48}/\frac{20}{48} = 6/20$	$\frac{8}{48}/\frac{20}{48} = 8/20$

P.no-29 39ns. ~~Q The joint distrbn of X & Y is given by $f(x,y) = (x+y$~~

⇒ Note:-

Marginal distrbn of X is

$$f_X(x) = \sum_{y=0}^{2} f(x,y) \; ; \; x = 0, 1, 2.$$

Marginal distrbn of Y is

$$f_Y(y) = \sum_{x=0}^{2} f(x,y) \; ; \; y = 0, 1, 2.$$

X \ Y	0	1	2	$f_X(x)$
0	0	1/27	2/27	3/27
1	2/27	3/27	4/27	9/27
2	4/27	5/27	6/27	15/27
$f_Y(y)$	6/27	9/27	12/27	1

The marginal distrbn of X is :

x	0	1	2
$f_X(x)$	1/27	9/27	15/27

The marginal distbn of Y is :

Y	0	1	2
$f_Y(y)$	6/27	9/27	12/27

* No. X and Y are not independent random variables, bcoz

$f(x,y) \neq f_x(x) \cdot f_y(y)$, for some value of (x,y)

eg:- $f(0,0) = 0$, but $f_X(0) \times f_Y(0) = \frac{3}{27} \times \frac{6}{27} = \frac{2}{81}$

Hence $f(0,0) \neq f_X(0) \times f_Y(0)$

Also $f(1,1) = 1/27$; $f_X(1) \times f_Y(1) = \frac{9}{27} \cdot \frac{9}{27} = \frac{1}{9}$

Hence $f(1,1) \neq f_X(1) \times f_Y(1)$

∴ X & Y are not independent random variables.

Module - 4.

A finite subset of statistical individuals in a popl^n is called a sample and no. of individuals in a sample is called the sample size.

⇒ Types of Sampling:-

① Purposive Sampling:-

It is the one in which the sample units are selected with definite purpose in view.

eg:- If we want to give the picture that the standard of living has increased in the city of New Delhi, we may take individuals in the sample from rich and posh localities like Defence Colony, South Extension, Golf Links, Jor Bagh, Chanakyapuri, Greater Kailash etc and ignore localities where low income group and the middle class families live. This sampling suffers from the drawback of favouritism and nepotism and doesn't give a representative sample of the population.

② Random Sampling:-

In this case, the sample units are selected at random and the drawback of purposive sampling, viz, favouritism or subjective element, is completely overcome.

A random sample is one in which each unit of popl^n has an equal chance of being included in it.

Suppose we take a sample of size 'n' from a finite popl^n of size N. Then there are ^NC_n possible samples. A sampling technique in which each of the ^NC_n samples has an equal chance of being selected is known as random sampling and the sampling obtained by this technique is termed as a random sample.

The simplest method which is normally used is the lottery system, which is an eg of random sampling.

Suppose we want to select 'r' candidates out of 'n'. We assign numbers 1-n one number to each candidate and write these numbers (1-n) on 'n' slips ~~oo~~ which are made as homogeneous as possible in shape, size etc. These slips are then put in a bag and thoroughly shuffled and then 'r' slips are drawn one-by-one. The 'r' candidates corresponding to the numbers on the slips drawn, will constitute the random sample.

③ Simple Sampling:-
It is a random sampling in which each units of the popl^n has an equal chance, say p, of being included in the sample and that this probability is independent of the previous drawing. Thus a simple sample of size 'n' from the popl^n may be identified with series of 'n' independent trials with a constant probability 'p' of success for each trial.

④ Stratified Sampling:-
Here the entire heterogeneous popl^n is divided into a no. of homogeneous group, usually termed as a Strata, which differ from one another, but each of these groups is homogeneous within itself. Then units are sampled at random from each of these stratums the sample size in each stratum varies acco. to the relative importance of the stratum in the popl^n. The sample, which is the aggregate of the sampled units of each of the stratum, is termed as stratified sample and the technique of drawing this sample is known as stratified sampling.

⇒ Parameter and Statistic:-
- Any measurable func^n of the popl^n values is called popl^n parameter.
eg:- Mean, median, SD etc of a popl^n are it's parameter.
- Any measurable func^n of the Sample values is called a ~~Sample~~ Statistics.
eg:- The sample mean, sample SD, sample ~~mom~~ moments are Statistics.

⇒ Sampling Distribution:-

The no. of possible samples of size 'n'

that can be drawn from a finite popl^n of size N is ${}^NC_n$.
The total no of possible samples is:

$$^NC_n = \frac{N!}{n!(N-n)!} = t \text{ (say)}$$

For each of these samples we can compute a statistic, say 't'
ie; $t = t(x_1, x_2, x_3 \ldots x_n)$ which will obviously vary from sample to sample.

Sample No	Statistic		
	t	$\bar{x}$	s^2
1	t_1	$\bar{x}_1$	s_1^2
2	t_2	$\bar{x}_2$	s_2^2
⋮	⋮	⋮	⋮
n	t_n	$\bar{x}_n$	s_n^2

The aggregate of the various values of this statistic under consideration so obtained [one for each sample] may be grouped into a frequency distrb^n which is known as sampling distrb^n of this statistic.

Various Sampling Distrb^{ns} are:-
(i) Sampling distrb^n of Mean
(ii) Sampling distrb^n of population.
(iii) Students 't' distrb^n
(iv) F-distrb^n
(v) ch-square distrb^n.

⇒ Defn:-
Since statistics are random variables the distrb^n of the statistics are called sampling distrb^n.

⇒ Standard Error [SE]:-
The standard deviation of the sampling distrb^n of a statistic is known as its standard error, abbreviated as. SE.
SE plays a very important role in large sample theory and forms the basis of the testing of hypotheses. The concept of SE is extremely useful in the estimation of parameter.
If 'u' is any statistic, $\sqrt{V(u)}$ is its standard error. Then the test statistic,

$$t = \frac{\text{Sample statistic} - \text{popl}^n \text{ parameter}}{\text{SE of the statistic}}$$

ie; $$t = \frac{u - E(u)}{\sqrt{V(u)}} \sim N(0,1).$$

The std error (SE) of some of the well known statistic are given below, where 'n' is the sample size, σ^2 is the popl^n variance and P is the popl^n.

★ **Result 1 :-**

If a random sample of size 'n' is taken from a popl^n having the mean μ and variance σ^2, then

$$\text{SE of Mean} = \begin{cases} \sigma/\sqrt{n} \; ; \text{if popl}^n \text{ is infinite}; \; n \geq 30. \\ \sigma/\sqrt{n}\sqrt{\dfrac{N-n}{N-1}} \; ; \text{if popl}^n \text{ of size } N; \; n < 30. \\ s/\sqrt{n} \; ; \text{if } \sigma \text{ is unknown}. \end{cases}$$

★ **Result 2 : Central Limit Theorem:-**

If $\bar{x}$ is the mean of the sample size 'n' taken from a popl^n having mean μ, and variance σ^2, then $z = \dfrac{x-\mu}{SE}$ is a random variable where the distrb^n func^n approaches std normal distrb^n as $n \to \infty$.

ie; The sample distrb^n of means is normal with mean μ and SD $\dfrac{\sigma}{\sqrt{n}}$.

★ **Result 3 :-**

When the popl^n SD is not known and the sample is small, then the sampling distrb^n of the mean is given by this random variable,

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \; ; \text{where } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

follows a t-distrb^n with degrees of freedom $(n-1)$.

★ **Result 4 :-**

If a random sample of size 'n' is taken a popl^n distrb^n of proportion of a characteristic with probability of occurence P, then the standard error of the sampling distrb^n will be $\sqrt{PQ/n}$, where $q = 1-p$.

Also, if 'n' is large, then $z = \dfrac{\bar{p}-p}{SE}$ will be a random variable where distrb^n appro. std normal where $\bar{p}$ is the proportion of the sample.

★ **Result 5 :-**

If 2 independent random samples of size n_1 and n_2 are taken from two popl^n with SD σ_1 and σ_2, then SE of the difference of sample means

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

If the sampling are taken from the same popl^n with SD σ, then

$$SE = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

If σ is unknown, then

$$SE = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2$ and

$$s_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2$$

* Result 6:-

If 2 independent samples of size n_1 & n_2 are taken from 2 popl^n having proportion $\bar{P}_1$ and $\bar{P}_2$ respectively of a characteristic, then the SE of the sampling distrb^n of the difference of proportion is given by

$$SE = \sqrt{\frac{\bar{P}_1\bar{q}_1}{n_1} + \frac{\bar{P}_2\bar{q}_2}{n_2}}$$

If the samples are taken from the same popl^n, then

$$SE = \sqrt{P_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \text{ where}$$

$$P_0 = \frac{n_1\bar{P}_1 + n_2\bar{P}_2}{n_1+n_2} \text{ and } q_0 = 1 - P_0.$$

Note:- (1) SE of mean with σ unknown $= s/\sqrt{n}$ where $s^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2$

(2) SE of difference of means with σ_1 and σ_2 unknown

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

⇒ Estimation :-

Estimation is concerned with the estimation of popl^n parameter from the same statistic.

The sample statistics to estimate popl^n parameters are called estimators and values of these popl^n parameters are called estimates.

If the estimate of a popl^n parameter is specified as a range of values, the estimates is called interval estimates.

A good estimator [parameter] has the following ~~parameter~~ properties:

① Unbiased :-

If the mean value of the sampling distrib^{n} of the statistic is equal to the parameter estimated, then the estimator is unbiased.

② Efficient :-

If the estimator has a relatively small variance, then its efficient.
ie; The most efficient estimator has the smallest variance.

③ Sufficient :-

If the estimator use all the infm^{n} available from the sample, then it is sufficient.

④ Consistent :-

If the estimator approaches the parameter as the same size becomes larger, then it is consistent.

⇒ Confidence Level :-

The pb of an estimator to lie b/w the interval is called the confidence level or confidence coefficient.

95% confidence level of a parameter θ to lie in the interval (c_1, c_2) we mean, prob$(c_1 < \theta < c_2) = 0.95$. We usually denote the confidence level by $1-\alpha$ and α is called the significance level.
Confidence Level = 1 – Significance level.
ie; 95% confidence level means 5% significance level.
ie; an interval estimates, we determine 2 constants c_1 and c_2 such that
$P(c_1 < \theta < c_2) = 1-\alpha$; α – significance level.
The interval (c_1, c_2) is called confidence ~~band~~ interval and the limits c_1 & c_2 are called the confidence limits.

Suppose we want to test whether the given sample of size 'n' has been drawn from a popl^{n} with mean μ.

If σ is known and $n > 30$.
We can use the test statistic
$Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ follows $N(0,1)$, where σ is the popl^{n} SD and $\bar{x}$ is the sample mean.

If σ is unknown and $n < 30$, then we can use $z = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$, where s is the sample SD.

From the area of standard normal distrib^n, we can find $z_{\alpha/2}$ such that

$$P\left[|z| < z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[-z_{\alpha/2} < z < z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[-z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[-\sigma/\sqrt{n}\, z_{\alpha/2} < \bar{x}-\mu < \sigma/\sqrt{n}\, z_{\alpha/2}\right] = 1-\alpha.$$

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

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

$\bar{x} \pm \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}$ is the required confidence interval.

$z_{\alpha/2}$ is called the critical value or significance value.

Note : 1

If the sample size is large ($n \geq 30$), $z = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$ follows std normal distrib^n, where $SE = \dfrac{\sigma}{\sqrt{n}}$.

Note : 2

If 'n' is small and σ is unknown, then z follows t-distrib^n with $(n-1)$ degrees of freedom. Then z-follows t-distrib^n if n is small and σ is unknown, we use 's' instead of σ.

ie; If the popl^n SD σ is not known and the sample is small, we use the test statistic $z = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$; s = sample SD.

Confidence Interval is $\bar{x} \pm \dfrac{s}{\sqrt{n}} t_{\alpha/2}$

where $t_{\alpha/2}$ is that value from the table of t-distrib^n for which $P(-t_{\alpha/2} < z < t_{\alpha/2}) = 1-\alpha$ with $(n-1)$ degrees of freedom.

Note : 3

The following table gives the significant value or critical values of diff confidence levels.

Confidence level	90 %	95 %	99 %
$z_{\alpha/2}$	1.645	1.96	2.58

Note : 4.

In the case of proportions z-follows std normal distrib^n, confidence interval is

$$P \pm z_{\alpha/2} \sqrt{\frac{pq}{n}}$$

Q. A sample of 50 items taken from a popl^n $N(52.5, 16)$. find 95% of confidence interval of the popl^n mean.

→ Here $n = 50$, $\sigma = 16$, $\bar{x} = 52.5$
Since $n > 30$, the sample is large.
Also σ is given.
Hence $SE = \sigma/\sqrt{n} = 16/\sqrt{50} = 2.26$
For 95% confidence level, $z_{\alpha/2} = 1.96$
Hence confidence ~~lev~~ interval is
$\left[\bar{x} - z_{\alpha/2}\,\sigma/\sqrt{n},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

ie; $(52.5 - 1.96 \times 2.26,\ 52.5 + 1.96 \times 2.26)$
ie; $(48.07,\ 56.96)$

Q. A sample of 100 items gave a mean 7.4 kg and a SD 1.2 kg. find 95% confidence limits for the popl^n mean.

– Here $n = 100$, $\bar{x} = 7.4$, $s = 1.2$
Since $n > 30$ sample is large. Also σ is not known.
Then $S.E = \frac{s}{\sqrt{n}} = \frac{1.2}{\sqrt{100}} = 0.12$
$z_{\alpha/2} = 1.96$.

The confidence limits of μ are
$\bar{x} \pm \frac{s}{\sqrt{n}} z_{\alpha/2}$
ie; $7.4 \pm 1.96 \times 0.12$
7.4 ± 0.2352
$7.1648 \le \mu \le 0.2352$.

Q. If $n = 50$ and if $P(-5.88 < \bar{x} - \mu < 5.88) = 0.95$. find σ.

– Given $P(-5.88 < \bar{x} - \mu < 5.88) = 0.95$ — ①
But $P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ — ②

Compare ① & ②, we get
$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 5.88$ | $z_{\alpha/2} = 1.96$, $n = 50$
$1.96 \times \frac{\sigma}{\sqrt{50}} = 5.88$
$\Rightarrow \sigma = \frac{5.88 \times \sqrt{50}}{1.96} = 21.213$

Q. The mean operating life for a random sample of 10 light bulbs is 4000 hours with a SD 200 hours. Estimate a 95% confidence ~~level~~ interval for the popl^n mean.

– Here $n = 10$, $s = 200$, $\bar{x} = 4000$.

Since $n < 30$, sample is small.
Also σ is not known, Hence we use t-distrb^n with $(n-1)$ degrees of freedom.

Here $SE = \frac{s}{\sqrt{n}} = \frac{200}{\sqrt{10}} = 63.25$.

$t_{\alpha/2}$ at $(n-1)$ d.f $= t_{\alpha/2}$ at $(10-1) = 9$ df.
$= 2.262$ //

Hence the confidence limits of μ are $\bar{x} \pm t_{\alpha/2}\ SE$.

ie; $4000 \pm 2.262 \times 63.25$

3857 and 4143 //

Q. The mean operating life of a random sample of 15 bulbs taken from a pop^n with SD 500hrs is 8900 hours. Find
(a) 95% confidence limits
(b) 90% confidence limits for pop^n mean.

- Here $n = 15$; $\sigma = 500$; $\bar{x} = 8900$
Since $n < 30$, sample is small.
But σ is known. Hence we use normal distrb^n. The confidence limits are:

(a) $(\bar{x} \pm Z_{\alpha/2}\ \sigma/\sqrt{n})$
ie; $8900 \pm 1.96 \times \frac{500}{\sqrt{15}}$

ie; 8647 and 9153

(b) for 90%, $Z_{\alpha/2} = 1.645$
Confidence limits are
$8900 \pm (1.645) \times 500/\sqrt{15}$
ie; 8687 and 9113.

Q. In a random sample of 64 of 2400 intersections in a city, the mean no. of accidents per year was 3.2 with S.D 0.08. Find a 90% confidence interval for the mean no. of accidents per year.

- Here $n = 64$; $N = 2400$; $\bar{x} = 3.2$
$s = 0.08$ $\because s > 30$
we use normal distrb^n

$SE = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{0.08}{\sqrt{64}} \times \sqrt{\frac{2400-64}{2400-1}}$
$= 0.00987$ //

For 90%, $Z_{\alpha/2} = 1.645 = SE$

The confidence interval is
$(\bar{x} - Z_{\alpha/2}\ SE,\ \bar{x} + Z_{\alpha/2}\ SE)$
ie; $(3.2 - 1.645 \times 0.0099,\ 3.2 + 1.645 \times 0.0099)$

$= (3.184,\ 3.216)$ //

Q. In a sample of 400 people, 17 were male. Estimate the popl^n proportions at 95% confidence level.

– Here $n = 400$; $P = \frac{17}{400} = 0.43$

$q = 1 - p = 0.57$.

For 95%, $Z_{\alpha/2} = 1.96$

$SE = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.43 \times 0.57}{400}} = 0.0248$

Then confidence limits are;

$P \pm Z_{\alpha/2}\, SE = P \pm Z_{\alpha/2} \sqrt{\frac{pq}{n}}$

$\Rightarrow 0.43 \pm 1.96 \times 0.00248$

$\Rightarrow 0.3814$ & 0.4786

Q. Suppose the following 10 values represent random obsrv^n from a normal popl^n 2, 6, 7, 9, 5, 1, 0, 3, 5, 4. Construct a 99% confidence interval for the mean of this popl^n.

$\Rightarrow$ Here $\bar{x} = \frac{2+6+7+9+5+1+0+3+5+4}{10}$

$= 4.2$

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

$= \frac{1}{10-1}\Big[(2-4.2)^2 + (6-4.2)^2 + (7-4.2)^2 + (9-4.2)^2 + (5-4.2)^2 + (1-4.2)^2 + (3-4.2)^2 + (5-4.2)^2 + (4-4.2)^2\Big]$

$S^2 = 7.733$

$\therefore S = 2.78$

$1-\alpha = 0.9$
$\alpha = 0.01$
$\alpha/2 = 0.005$

Since n is small, σ is unknown, we use t-distrib^n.

$t_{\alpha/2}$ of $(n-1)$ df $= t_{\alpha/2}$ of $(10-1) = 9$ df.

$= 3.250$.

Here $SE = \frac{S}{\sqrt{n}} = \frac{2.78}{\sqrt{10}} = 0.8794$

Hence confidence limits are:

$\bar{x} \pm t_{\alpha/2}\, S/\sqrt{n}$

ie; $4.2 \pm (3.250)(0.8794)$

$4.2 \pm 2.888 = 1.342$ & 7.058

Q. In a random samples of 450 industrial accidents it was found that 230 were due to unsafe working conditions. Construct 95% confidence interval for corresponding true proportions.

– Here $P = \frac{230}{450} = 0.51$

$n = 450$; $q = \frac{220}{450} = 0.49$

Since n is large, we use normal distrib^n.

Now $Z_{\alpha/2} = 1.96$ and

$SE = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.51 \times 0.49}{450}} = 0.0236$

Thus confidence limits are:

$(0.51 - 1.96 \times 0.0236, 0.51 + 1.96 \times 0.0236)$

$= (0.4638, 0.5563)$

Q. Find the least sample size required if length of 95% confidence interval for the mean of a normal pop^n with SD 8 should be less than 10.

→ Let n be size of sample

95% confidence interval is

$(\bar{X} - Z_{\alpha/2} SE, \bar{X} + Z_{\alpha/2} SE)$

But $Z_{\alpha/2} = 1.96$ and

$SE = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{n}}$

Length of confidence interval =

$(\bar{X} + Z_{\alpha/2} SE) - (\bar{X} - Z_{\alpha/2} SE)$

$= 2 Z_{\alpha/2} SE$

$= 2(1.96) \frac{8}{\sqrt{n}} = \frac{31.36}{\sqrt{n}}$

Since width is to be less than 10, we have

$\frac{31.36}{\sqrt{n}} < 10 \Rightarrow \sqrt{n} > \frac{31.36}{10} = 3.136$

$\therefore n > (3.136)^2 = 9.834$

$\therefore$ least sample size $= 10$

Q. A random sample of 900 items with mean 3.5 is drawn from a pop^n with SD 2.61. Find 95% confidence interval for the mean.

→ $(3.33, 3.67)$

Q. A random sample of size 25 drawn from a ~~particular~~ pop^n having SD 10 is having mean 20. Find (i) 95% (ii) 99% confidence intervals of this pop^n mean.

→ $(16.08, 23.92)$

Q. Find the least sample size required if length of the 95% confidence interval for the mean of normal pop^n with SD 5 should be less than 6.

→ $n = 11$

Q. A random sample of 500 apples was taken from a large consignment and

60 were found to be bad. Obtain 99% confidence limits for the percentage of bad apples in the consignment.

→ $P = \frac{60}{500} = 0.12$; $q = 0.88$

$P \pm Z_{\alpha/2} \sqrt{\frac{pq}{n}}$

Module - 5
TESTING & HYPOTHESIS.

Testing of hypothesis deals with method for deciding either to accept or reject the hypothesis based on samples taken from the population.

Statistical hypothesis is a statement about the form of the distribution followed by one or more population or about the values of the parameters present in their distribution function.

eg: The weight of student in a university follows the normal distribution

Parametric hypothesis -

A hypothesis concerning the values of the parameters present in the distribution function of a population. The form of the distribution function assumed to be known ~~is called~~ as Parametric hypothesis

eg: the hypothesis $P = 1/2$ for a binomial distribution

Simple and Composite hypothesis

A statistical hypothesis which Completely Specified the distribution of a population is said to be simple hypothesis.
eg: Consider a population following normal distribution with parameter μ and σ
The hypothesis $H_0: \mu = 3, \sigma = 5$ is simple.
A statistical hypothesis which is not simple is said to be Composite.
eg: The hypothesis $H_0, \mu = 5$ is Composite.

Null hypothesis.

The hypothesis which we consider in a testing problem is called null hypothesis. usually null hypothesis are simple and it is ~~called~~ denoted by H_0. Null hypothesis is a hypothesis which we want to reject

Alternative hypothesis

Using a sample takes from the population we either reject or accept the hypothesis. Rejection of the null hypothesis resulted in the acceptance of another hypothesis. which is known as alternate hypothesis and is denoted by H_1
eg: let P denoted the probability of getting a head, when a Coin is tossed we can take $H_0: P = 1/2$. if this null hypothesis is rejected and then ~~if~~ it would mean the following three possible alternatives.
ie $P \neq 1/2, P < 1/2$ or $P > 1/2$
$H_0: P = 1/2$ may be tested against other alternatives such that
$H_1: P > 1/2$, $H: P < 1/2$ Or $H_1: P > 1/2$
ie ① $H_0: \mu = 1/2, H_1 = \mu \neq 1/2$
② $H_1: \mu = 1/2, H_1 = \mu < 1/2$
③ $H_1 = \mu = 1/2, H_1 = \mu > 1/2$
The hypothesis $\mu = 1/2$ are called Single hypothesis and if the type

$\mu \neq 1/2, \mu < 1/2, \mu > 2$ (ie non Specific) are called composite hypothesis
Note Null hypothesis is expressed

on a simple hypothesis and the alternate hypothesis is a composite hypothesis.

Statistical test (test statistics)

A statistical test is a procedure by which we make a decission either to accept or reject the hypothesis based on the samples taken from the population

Level of significance (α)

The probablity of rejecting a hypothesis H_0 when it is true is called the level of significance and it is denoted as α.

Type I and Type II error.

There are two types of errors

Type I error :- Rejecting H_0 whet it is actually true is called type I error.

Type - 2 error.

Accepting a null hypothesis H_0 when it is false is called type II error

These can be represent in tabular form as shown below.

state of nature / action	H_0 true.	H_0 false.
Reject H_0	Type I error	No error
Accept H_0	No error	Type II error

Critical region.

The basis of testing the hypothesis is the partition of a sample space into two exclusive region. namely the region of rejection and the region of acceptance. If the sample points falls in the region of rejection, H_0 is rejected. The regin of rejection is called critical region.

Critical region.

~~The basis of testing th~~ The value of test statistic which separates the critical region or (rejection region) and acceptance region is called critical value or significant value. The critical value is usually denote as z_α or t_α depending on the sampling distribution of the test statistic.

One-tailed and two tailed test

In any test, the critical region is represented by a portion of a area under the probablity curve of the sampling distribution of the statistic.

A test of any statistical hypothesis where the alternatives hypothesis is one tailed (right tailed or left tailed) is called a one-tailed test.

eg: A test for testing the mean of a population $H_0 = \mu = \mu_0$ against the alternative hypothesis.

$H_1 = \mu > \mu_0$ (right-tailed) or

$H_1 = \mu < \mu_0$ (left-tailed, is a single tailed test

In a particular problem, whether one tailed or two tailed test is to be applied depends entirely on the nature of the alternative hypothesis. If the alternative hypothesis is two-tailed, we apply two tailed test and if alternative hypothesis is one tailed, we apply one tailed test.

In the right-tailed test ($H_1: \mu > \mu_0$) the critical region lies entirely in the right tail of the sampling distribution of $\bar{x}$, while for the left-tailed test ($H_1: \mu < \mu_0$), the critical region is entirely in the left tail of the distribution.

A test of statistical hypothesis where the alternative hypothesis is two tailed such as $H_0: \mu = \mu_0$ against the alternative hypothesis

$H_1: \mu \neq \mu_0$ ($\mu > \mu_0$ and $\mu < \mu_0$) is known as two tailed test and in such a case the critical region is given by the portion of the area lying in both tails of the probablity curve of the test statistic.

We now summarize the steps involved in testing a statistical hypothesis

step1:- State the null hypothesis H_0 and the alternative hypothesis H_1.

ie; if we take $H_0 : \mu = \mu_0$ against alternatives:

(1) $H_1 : \mu = \mu_0$ [2-sided alternative]
(2) $H_1 : \mu < \mu_0$ } [one-sided alternatives]
(3) $H_1 : \mu > \mu_0$

Step 2:- Choose the level of significance α
ie; we specify 10%, 5% or 1% level of significance

Step 3:- Determine the test statistic.
Normally, we select normal distrb^n or t-distrb^n.

Step 4:- Determine the pb distrb^n of the test statistic.

Step 5:- Determine the best critical region

Step 6:- Calculate value of test statistic

Step 7:- If the calculated value of the test statistic falls in critical region, reject the null hypothesis H_0, otherwise accept it.
ie; If the calculated value exceeds the table value, reject H_0, otherwise accept it.

We are going to deal with the following tests:

(1) Testing the hypothesis concerning one mean.
(2) Testing the hypothesis concerning difference of 2 means.
(3) Testing the hypotheses concerning one proportion
(4) Testing the hypothesis concerning difference of 2 proportions.

In all the above tests the test statistic will be 'z' if the sample is large or the popl^n SD σ is known. In this case of 1 & 2, if the popl^n SD is not known, and also sample is small, we use t-statistic. Thus, we use z-test for all tests where popl^n SD is known or the sample is large ($n \geq 30$) and t-test for only those cases where σ is unknown and $n < 30$.

(1) Hypotheses Concerning a Mean:-
To test whether the difference b/w the sample mean $\bar{x}$ and popl^n mean μ is significant, we use the following

Sample mean $\to \bar{x}$
Popl^n mean $\to \mu$
popl^n SD $\to \sigma$

test statistic

(a) If sample size is large or this popl^n SD σ is known, then the test statistic is $z = \frac{\bar{X}-\mu}{SE}$ which is approximately std normal.

(b) If sample size is small and if popl^n std dev^n σ is unknown, then the test statistic is
$t = \frac{\bar{X}-\mu}{SE}$ which follows t-distrib^n, with $(n-1)$ degrees of freedom. Here

$$SE = \begin{cases} \sigma/\sqrt{n} \; ; \text{ if } \sigma \text{ is known, popl}^n \text{ infinite} \\ s/\sqrt{n} \; ; \text{ if } \sigma \text{ is unknown, } s \text{ is known and popl}^n \text{ infinite} \\ \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \; ; \text{ If popl}^n \text{ is finite and is of size } N \end{cases}$$

* Note :-
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Q. A sample of 900 members is found to have mean of 3.47 cm. Can it be reasonably regarded as a simple sample from a large popl^n with mean 3.23 cm and SD 2.31 cm ?

- Given $\mu = 3.23$; $\sigma = 2.31$
$\bar{x} = 3.47$ and $n = 900$.
Let $H_0 : \mu = 3.23$
$H_1 : \mu \neq 3.23$
As sample size $n = 900$ is large, the test statistic is :
$$z = \frac{\bar{x}-\mu}{SE} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{3.47-3.23}{2.31/\sqrt{900}}$$
$$= 3.1169.$$
H_1 is 2 sided and hence at 5%. level of significance, from normal table,
$Z_{\alpha/2} = 1.96$ ($\alpha = 0.05$)

Since $|z| > 1.96$
$\therefore H_0$ is rejected against H_1.

Q. A stenographer claims that she can takes dictations at the rate of 120 words per minute. Can we reject her claim on the basis of 100 trails in which she demonstrates a mean of 116 words with a SD of 15 words ? Use 5% level of significance.

- Given $n = 100$; $\mu = 120$
$\bar{x} = 116$ and $s = 15$
Let $H_0 : \mu = 120$
$H_1 : \mu \neq 120$

As the sample size 'n' is large, we consider the test statistic;

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

$$= \frac{116-120}{15/\sqrt{100}} = -2.67$$

At 5% level of significance,

$z_{\alpha/2} = 1.96$

Since $|z| > 1.96$, we reject H_0.

Q. The SD of the tubelight of a particular brand of ultraviolet tubes is 500 hours and the operating life of the tubes is normally distributed. The manufacturer claims that average tube life is atleast 9000 hours. Test the claims at 5% level of significance, given that a random sample of 15 tubes gave a mean life of 8,800 hours.

— Given $n = 15$; $\mu = 9000$; $\sigma = 500$; $\bar{x} = 8800$.

Since $n < 30$, the sample size is small, but popl^n SD, σ is given.

Let $H_0 : \mu = 9000$

$H_1 : \mu < 9000$.

$$z = \frac{\bar{x}-\mu}{SE} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{8800-9000}{500/\sqrt{15}}$$

$$= -1.549$$

At 5% level of significance,

H_1 being left tailed,

$z_\alpha = -1.645$

$|z| < |z_\alpha| = 1.645$

H_0 is accepted.

Q. An educator claims that the average IQ of American college students is almost 110 and that in a study made to test his claim, 150 American college students had an avg IQ of 111.2 with a SD 7.2. At 1% level of significance, test the claim of the educator.

— Given $n = 150$; $\mu = 110$; $\bar{x} = 111.2$; $s = 7.2$.

Since claim of the educator is almost 110. we take

$H_0 : \mu = 110$

$H_1 : \mu > 110$.

The sample size is large, hence take test statistic;

$$z = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{111.2-110}{7.2/\sqrt{150}} = -2.0412$$

As H_1 is one-tailed (right tailed), Z_α at 1% level of significance is 2.33.
Since $Z < Z_\alpha$, H_0 is accepted.

9. A company claims that the mean life of its bulbs produced is 1600 hours. A random sample of 100 bulbs gave a mean life is 1570 hours with a SD 120 hours. Test the claim
(a) at 5% level
(b) at 1% level

– Given $n = 100$; $\mu = 1600$; $\bar{x} = 1570$; $S = 120$.

Let $H_0 : \mu = 1600$
$H_1 : \mu \neq 1600$

As the sample size is large, we take the test statistic,

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{1570 - 1600}{120/\sqrt{100}} = -2.5$$

(a) At 5% level of significance; the critical value, $Z_{\alpha/2} = 1.96$. Since $|z| > 1.96$, we reject H_0.

(b) At 1% level, the critical value $Z_{\alpha/2} = 2.58$. Since $|z| < 2.58$, we Accept H_0.

9. 10 students are selected at random from a school and their heights are found to be in inches 50, 52, 52, 53, 55, 56, 57, 58, 58, 59. In the light of these data discuss the suggestion that the mean height of students of the school is 54 inches. Use 5% level of significance.

– Given $n = 10$; $\mu = 54$

Let $H_0 : \mu = 54$
$H_1 : \mu \neq 54$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
50	−5	25
52	−3	9
52	−3	9
53	−2	4
55	0	0
56	1	1
57	2	4
58	3	9
58	3	9
59	4	16
550		86

$n = 10$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{550}{10} = 55$$

$$S^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2 = \frac{1}{9} \times 86 = 9.56$$

$$S = \sqrt{9.56} = 3.09$$

Now, $$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{55 - 54}{3.09/\sqrt{10}} = 1.023$$

At 5% level of significance, the critical value,
t_α with $10-1=9$ d.f is 2.262
Since $|t| < 2.262$ (2-tailed)
H_0 is accepted.

(2) Hypothesis Concerning Difference of means of two Samples:-

To test whether the difference of means $\bar{x}_1 - \bar{x}_2$ of 2 samples $\bar{x}_1$ and $\bar{x}_2$ is significant, we use the following test statistic.

(a) If the sample size is large, or if the SD of the popl^n is known, we use the test statistic $z = \frac{\bar{x}_1 - \bar{x}_2}{SE}$ which follows the standard normal distrib^n.

(b) If sample size (s) is small and if the popl^n SD is not known, we use the test statistic $t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$ which follows t-distrib^n with n_1+n_2-2 degrees of freedom.

Here;

$$SE = \begin{cases} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ ; when samples are drawn from 2 dif popl}^{ns} \text{ with SD } \sigma_1 \text{ \& } \sigma_2 \\ \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \text{ ; when } \sigma_1 \text{ \& } \sigma_2 \text{ are not known} \\ \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \text{ ; when samples are taken from same popl}^n \text{ with SD } \sigma. \end{cases}$$

If σ is not known, we take σ as,

$$\sigma^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}.$$

Q. The means of 2 random samples of the 1000 and 2000 are 67.5 and 68 inches respectively. Can the samples be regarded to have been drawn from the same popl^n of SD 9.5 inches? Test at 5% level of significance.

- Let $H_0: \mu_1 = \mu_2$ (mean of popl^n)
$H_1: \mu_1 \neq \mu_2$

Given $n_1 = 1000$; $n_2 = 2000$;
$\bar{x}_1 = 67.5$; $\bar{x}_2 = 68$; $\sigma = 9.5$

The test statistic;

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{67.5-68}{9.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}} = -1.359$$

As it is 2-sided, At 5% level, critical value $Z_{\alpha/2} = 1.96$
Since $|z| < 1.96$, H_0 is accepted.
ie; Samples are from the same population.

Q. The mean produce of wheat of a sample of 100 fields is 200 lbs per acre, with a SD of 10 lbs. Another sample of 150 fields gives the mean of 220 lbs with a SD of 12 lbs. Can the two samples be considered to have been taken from the same popl^n whose SD is 11 lbs? Use 5% level of significance.

- Let $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$.

Given $n_1 = 100$; $n_2 = 150$; $\bar{x}_1 = 200$;
$\bar{x}_2 = 220$; $s_1 = 10$; $s_2 = 12$; $\sigma = 11$.

Since σ is given, we consider the test Statistic;

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{200 - 220}{11\sqrt{\frac{1}{100} + \frac{1}{150}}} = -14.08$$

As H_1 is two-sided, critical value at 5% level of significance is 1.96.
Since $|z| > 1.96$, H_0 is rejected.

Q. A random sample from 200 villages was taken from Kanpur district and the avg popl^n per village was found to be 420 with SD of 50. Another random sample of 200 villages from the same district gave an avg popl^n of 480 per village with a SD is 60. Is the difference b/w the avg of 2 samples statistically significant? Take 1% level of significance

- Let $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Given $n_1 = 200$; $n_2 = 200$; $\bar{x}_1 = 420$;
$\bar{x}_2 = 480$; $s_1 = 50$; $s_2 = 60$.

Since popl^n is same and σ is not given, we take

$$\sigma^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(199)(50)^2 + 199(60)^2}{398} = 3050$$

$\therefore \sigma = \sqrt{3050} = 55.23$

Since samples are large, we take test statistic;

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{420 - 480}{55.23\sqrt{\frac{1}{200} + \frac{1}{200}}} = -10.86$$

As H_1 is 2 sided, At 1% level of significance, the critical value $Z_{\alpha/2} = 2.58$

Since $|z| > 2.58$, H_0 is rejected.

Q The mean height of 50 male students who showed above avg participation in college athletics was 68.2 inches with a SD of 2.5 inches; while 50 male students who showed no interest in such participation had a mean height of 67.5 inches with a SD of 2.8 inches. Test the hypothesis that male students who participate in college atheletics are taller than other male students.

— Let $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$ [Athletic students are taller than other].

Given $n_1 = 50$; $n_2 = 50$; $\bar{x}_1 = 68.2$; $\bar{x}_2 = 67.5$; $s_1 = 2.5$; $s_2 = 2.8$

Since samples are larger and from d/f pop^n, we take the test statistic,

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{68.2 - 67.5}{\sqrt{\frac{(2.5)^2}{50} + \frac{(2.8)^2}{50}}} = 1.3186$$

As H_1 is right tailed, at 5% level of significance, the critical value. $Z_\alpha = 1.645$. Since $z < 1.645$;

H_0 is accepted.

Q The height of 6 randomly chosen sailors are inches 63, 65, 58, 69, 71, 72. The heights 10 randomly chosen soldiers are inches 61, 62, 65, 66, 69, 69, 70, 71, 72 & 73. Do these figures indicate that soldiers are on an avg shorter than sailors? Test at 5% level of significance.

— Let $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$ [Sailors taller than Soldiers].

Sample 1

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
63	-3.33	11.0889
65	-1.33	1.7689
58	-8.33	69.3889
69	2.67	7.1289
71	4.67	21.8089
72	5.67	32.1489
398		143.3334

$\bar{x}_1 = \frac{398}{6} = 66.33$

$s_1^2 = \frac{1}{n_1 - 1} \Sigma (x_i - \bar{x}_1)^2$

$= \frac{1}{5} \times 143.334$

$= 28.67$

$s_1 = \sqrt{28.67}$

$= 5.354.$

Sample 2.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
61	-6.8	46.24
62	-5.8	33.64
65	-2.8	7.84
66	-1.8	3.24
69	1.2	1.44
69	1.2	1.44
70	2.2	4.84
71	3.2	10.24
72	4.2	17.64
73	5.2	27.04
678		153.60

$\bar{x}_2 = \frac{678}{10} = 6.78$

$s_2^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$

$= \frac{1}{10-1} \times 153.60$

$= 17.07$

$\therefore s_2 = \sqrt{17.07}$

$= 4.13$

Since $n < 30$, we take test statistic,

$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ where

$\sigma^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$ ~~$5 \times (5.1$~~

$= \frac{5 (5.354)^2 + 9 (4.13)^2}{6 + 10 - 2} = 21.2$

$\therefore t = \frac{66.3 - 67.8}{\sqrt{21.2} \sqrt{\frac{1}{6} + \frac{1}{10}}} = -0.6183$

As H_1 is right tailed, the critical value of t at 5% level of significance with $n_1 + n_2 - 2 = 14$ d.f is 1.761

As $|t| < 1.761$

H_0 is accepted.

(3) Hypothesis Concerning one proportion:-

To test whether the difference b/w sample proportion $\bar{p}$ and the popl^n proportion 'p' is significant we use the following test statistic,

$z = \frac{\bar{p} - p}{SE}$ which follows std normal distrb^n.

Here $SE = \sqrt{\frac{pq}{n}}$; $q = 1 - p$.

? A coin is tossed 10,000 times and ~~head~~ turns up 5195 times. Is the coin unbiased ?.

- If the coin is unbiased, $p = 1/2 = 0.5$

$n = 10000$; $\bar{p} = \frac{5195}{10,000} = 0.5195$

Let $H_0 : P = 0.5$

$H_1 : P_0 \neq 0.5$

$SE = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.5 \times 0.5}{10000}} = 0.005$

The test statistic is

$$z = \frac{\bar{p} - p}{SE} = \frac{0.5195 - 0.5}{0.005} = 3.9$$

Since H_1 is 2 sided the critical value of z at 5% level of significance is 1.96.

Since $|z| > 1.96$, H_0 is rejected.

Q. A die is thrown 9000 times and of these 3220 yielded a 3 or 4. Can the die be regarded as unbiased?

- Here $n = 9000$

$$\bar{p} = \frac{3220}{9000} = 0.3578$$

p = pb(getting a 3 or 4)

$$= \frac{2}{6} = \frac{1}{3} = 0.333$$

$$\therefore q = 1 - p = \frac{2}{3} = 0.667$$

Let $H_0 : p = 1/3$

$H_1 : p \neq 1/3$

Test statistic is

$$z = \frac{\bar{p} - p}{\sqrt{pq/n}} = \frac{0.3578 - 0.3333}{\sqrt{\frac{(0.333)(0.667)}{9000}}}$$

$$= 4.93$$

Since $|z| > 3$, H_0 is rejected.

Q. In a big city, 325 men out of 600 men were found to be smokers. Does this inform^n support the conclusion that the majority of men in this city are smokers? Use 5% level of significance.

- Here $n = 600$; $\bar{p} = \frac{325}{600} = 0.5417$.

popl^n proportion, p is taken as $1/2$

Then $H_0 : p = 1/2$

$H_1 : p > 1/2$ (majority is smokers)

Test statistic is:

$$z = \frac{\bar{p} - p}{\sqrt{pq/n}} = \frac{0.5417 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}} = 2.04$$

Since H_1 is right tailed, the critical value z_α at 5% level of significance is 1.640.

$\therefore z > 1.645$. H_0 is rejected.

(4) Hypotheses Concerning the Difference of Sample Proportions:-

To test whether the difference of

2 sample ~~post~~ proportions is significant, we use test statistic,

$z = \frac{\bar{P}_1 - \bar{P}_2}{SE}$ which follows standard normal distbⁿ where $\bar{P}_1$ & $\bar{P}_2$ are the sample proportions.

$$SE = \begin{cases} \sqrt{\frac{\bar{P}_1 \bar{q}_1}{n_1} + \frac{\bar{P}_2 \bar{q}_2}{n_2}} \ ; \text{ when popl}^n \text{ are heterogeneous} \\ \sqrt{P_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \ ; \text{ when popl}^n \text{ are homogeneous} \ ; P_0 = \frac{n_1 \bar{P}_1 + n_2 \bar{P}_2}{n_1 + n_2} \end{cases}$$

Q. A machine puts out 16 imperfect articles in a sample of 500. After machine overhauled, it puts out 3 imperfect articles in a batch of 100. Has the machine improved? Test at 5% level of significance.

Given $n_1 = 500$; $\bar{P}_1 = \frac{16}{500} = 0.032$

$n_2 = 100$; $\bar{P}_2 = \frac{3}{100} = 0.03$

Let $H_0 : P_1 = P_2$

$H_1 : P_1 < P_2$

$P_0 = \frac{n_1 \bar{P}_1 + n_2 \bar{P}_2}{n_1 + n_2}$

$= \frac{500(0.032) + 100(0.03)}{500 + 100} = 0.0317$

$q = 1 - P_0 = 0.9683$.

The test statistic is:

$z = \frac{\bar{P}_1 - \bar{P}_2}{\sqrt{P_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.032 - 0.03}{\sqrt{0.0317 \times 0.9683 \left(\frac{1}{500} + \frac{1}{100}\right)}}$

$= 0.1042$

As H_1 is left tailed, the critical value of z at 5% level is -1.645.

Since $z > -1.645$

H_0 is accepted.

Q. In a random sample of 1000 person 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 b/w town A & town B, so far as the proportion of wheat consumers is concerned?

- let $H_0: P_1 = P_2$

$H_1: P_1 \neq P_2$

Given $n_1 = 1000$; $\bar{P_1} = \frac{400}{1000} = 0.4$

$n_2 = 800$; $\bar{P_2} = \frac{400}{800} = 0.5$

$$P_0 = \frac{n_1 \bar{P_1} + n_2 \bar{P_2}}{n_1 + n_2}$$

$$= \frac{1000(0.4) + 800(0.5)}{1000 + 800} = \frac{4}{9}$$

$q_0 = 1 - 4/9 = 5/9$

$$SE = \sqrt{P_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= \sqrt{\frac{4}{9} \times \frac{5}{9} \left(\frac{1}{1000} + \frac{1}{800}\right)} = 0.024$$

Test statistic Z

$$Z = \frac{\bar{P_1} - \bar{P_2}}{SE} = \frac{0.4 - 0.5}{0.024} = -4.17$$

At 5% level of significance, the critical value of $Z = 1.96$

Since $|Z| > 1.96$, H_0 is rejected.

Q The percentage of defective parts turned out by the same machine on 2 consecutive days 6 & 4. If 500 parts are turned out on each of the 2 days, would it be justified to claims that qty has improved. Use 1% level of significance

- $n_1 = n_2 = 500$; $\bar{P_1} = \frac{6}{100} = 0.06$

$\bar{P_2} = \frac{4}{100} = 0.04$

let $H_0: P_1 = P_2$.

$H_1: P_1 > P_2$

Now $P_0 = \frac{n_1 \bar{P_1} + n_2 \bar{P_2}}{n_1 + n_2} = \frac{500(0.06) + 500(0.04)}{1000}$

$= 0.05$

$q = 1 - P_0 = 0.95$

$SE = \sqrt{P_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{(0.05)(0.05)\left(\frac{1}{500} + \frac{1}{500}\right)}$

$= 0.0138$

Now $Z = \frac{\bar{P_1} - \bar{P_2}}{SE} = \frac{0.06 - 0.04}{0.0138} = 1.45$

At 1% level, the critical value $Z_\alpha = 2.33$. Since $Z < 2.33$

H_0 is accepted

$\therefore$ There is no improvement.

⇒ Mouse Events & Key Events

```
import java.applet.*;
import java.awt.*;
import java.awt.event.*;
public class events extends Applet
{           implements MouseListener
 String msg = "Initial Message";
 public void () {
     addMouseListener(this);
 }
 public void mouseClicked(MouseEvent me)
 { msg = "Mouse Clicked";
    repaint();
 } public void mousePressed(MouseEvent me)
 { msg = "Mouse Pressed";
    repaint();
 } public void mouseReleased(MouseEvent me)
 { msg = "Mouse Released";
   repaint();
 } public void mouseEntered(MouseEvent me)
 { msg = "Mouse Entered";
   repaint();
 } public void mouseExited(MouseEvent me)
 { msg = "Mouse Exited";
   repaint();
 } public void paint(Graphics g).
 } g.drawString(msg, 20, 20);
 } }        applet code.
```

⇒ Key Events:-

```
import java.awt.*;
import java.awt.event.*;
import java.applet.*;
public class key_events extends Applet
          Implements KeyListener
 {
  String msg = "——Type Here——";
 int x = 30, y = 50;
 public void init()   {
 addKeyListener(this);
 requestFocus();
 } public void keyTyped(KeyEvent ke)
 {
 msg += ke.getKeyChar();
 repaint();
 } public void keyReleased(KeyEvent ke)
 {
 showStatus("Key Up!!!");
 } public void keyPressed(KeyEvent ke)
 { showStatus("Key Down!!!");
 } public void paint(Graphics G)
 { G.drawString(msg, x, y);
   }
 }
```