26/11/2[illegible]
Tuesday

Agb

M T W T F S S
Page No.:
Date:
YOUVA

Measures of central tendency

- tendency of data to ~~closure~~ cluster around a central point is known as central tendency.
- It may also defined as a value of variable which is thoroughly representative of series or distribution as a whole.
- measure of this character are known as averages.
- an average is a single value which is considered as the most representative or typical value for a given set of data.

Objectives of the study of averages

- to get one single value that describes the characteristic of entire data.
- to facilitate comparison b/w 2 or more series.

Characteristic of a good average

- it should be rigidly defined means it has one & only one interpretation.
- estimation shouldn't vary from individual to individual.
- its computation should be based on all observations.
- It should be as little effected by fluctuations of sampling as possible.
- It should be easily & rapidly computable.
- It should be easy to understand & easily comprehensionable.
- It should be capable of further ~~algi~~ algebric treatment.

- there are 2 types of averages
① Mathematical averages ⇒ arithmetic mean, harmonic mean, geometric mean.

② positional averages ⇒ median, mode, ~~mean~~ partition value

Ⓐ Arithmetic mean (AM) ⇒ most popular & widely used average. Its value is obtained by adding together all the observations & by dividing this total by no. of observations

$$\cancel{\bar{X} = \frac{\sum x_i}{N}} \qquad \boxed{\bar{X} = \frac{\sum_{i=1}^{i=n} x_i}{N}}$$

for population denoted by greek mark μ
& sample $\bar{X}$

$$\boxed{\bar{X} = \frac{\sum f_i x_i}{\sum f_i}} \rightarrow \text{for frequency distribution}$$

$$X = \frac{\sum f_i m_i}{\sum f_i}$$

properties of AM

- sum of deviation of various value from mean is zero.

$$\boxed{\sum (x - X) = 0}$$

deviation जिस से लेना
substract करना

taking deviation means we have to substract the value & answer is zero ^

- sum of squares of deviation of variants value from their mean is the least.

$$\boxed{\sum (x - X)^2 = \text{least}}$$

- If we have the AM & no. of observation of two or more than two unrelated group. the combined mean is

$$\bar{X}_G = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 - - - - N_n\bar{X}_n}{N_1 + N_2 + N_3 + - - - + N_n}$$

★★ Arithmetic mean changes on change of origin & scale.

merits of AM
- It is rigidly defined.
- It is ~~to~~ easy to understand.
- least affected by fluctuations of sampling.
- It based on all observations.

★★★
demerits of AM
- It is very much affected by extreme value.
- It's value mayn't coincide with any of given values.
- It can't be calculated for open ended frequency distribution.
- neither it can be located on frequency curve nor can be obtained by inspection

27/11/24 Wednesday

(3) Geometric mean

It is n^{th} root of product of N observation.

$$G.m = \sqrt[N]{x_1 \cdot x_2 \cdot x_3 - - - x_n}$$
$$= (x_1 \cdot x_2 \cdot x_3 - - - x_n)^{1/n}$$

Logarithm

$$G.m = \text{antilog}\left[\frac{\Sigma \log x}{N}\right]$$

frequency distribution

$$G.m. = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_3^{f_3} - - - x_n^{f_n}}$$

$$G.M. = \text{antilog}\left[\frac{1}{N} \Sigma f_i \log x_i\right]$$

A.M ≥ G.M. ≥ H.M. ★★

$27 \times 3 = 8\sqrt{3}$

Single the average average the

x	f	log x
3	1	log 3
4	3	" 4
7	2	" 7

$\frac{3 \cdot 4}{6}$

$GM = \text{antilog}\,(6)^n$

$\frac{20}{21}$

	2	7, 10, 12
4	2	7, 5, 6
21	3	7, 5, 3
	7	
	5	

Merits

- It is rigidly defined.
- It is based on all observation
- It is capable of further algebric treatment
- If any obs^n has -ve its computation is meaningless.

~~Demerits~~

- ~~Its~~ If any obs^n has zero value then its value be zero.
- It is especially used in following cases:-

a) when we have to give more weights to smaller value. (एक ही संख्या 2 या चार संख्या Same no)

b) when we are dealing with averages of rates of changes in ratio as in index number.

In case of frequency dist^n when it is markedly skewed (तिरछा) to right GM is used to convert it in symmetrical dist^n.

d) when we have to deal the quantity the changes of which are proportional to quantity itself & have to determined average rate of change in % per thousand per million etc.

④ Harmonic mean ⇒ reciprocal of the arithmetic mean of the reciprocals of obs^n value.

$$\frac{1}{H.M.} = \frac{1}{N}\left[\sum \frac{1}{x_i}\right]$$

$$H.M. = \frac{1}{\frac{1}{N}\left[\sum \frac{1}{x_i}\right]} \Rightarrow \frac{N}{\sum \frac{1}{x_i}}$$

$$-\frac{1}{H.M.} = \frac{1}{\Sigma f}\left[\Sigma \frac{f_i}{x_i}\right]$$

merits

It is rigidly defined & calculⁿ based on all obsⁿ.

- not much affected by sampling.
- & it is capable of further algebric treatment

Demerits

- It is neither easily calculated nor comprehensible

uses

- It is used when data are given in terms of units purchased per rupee or km covered per hour

$$HM = \frac{1}{N}\left[\Sigma \frac{1}{x_i}\right]$$

$$= \frac{1}{2}\left[\frac{1}{a}+\frac{1}{b}\right] = \frac{2ab}{a+b}$$

(5) Weighted Mean

In a condⁿ when all items on obsevⁿ aren't to be given equal imp then each item is given a proper wt. as per its imp & then the average which is calculated on the basis of these wt. is cd. as weighted mean.

$$W.M. = \frac{\Sigma w_i x_i}{w_i}$$

$$\boxed{W.M. = \frac{\Sigma w_i x_i}{\Sigma w_i}}$$

Median

- positional average
- when data are arrⁿ in ascending or descending order then the middle value is

known as median.

- If data are odd in no. then observation at $\left(\frac{N+1}{2}\right)^{th}$ posn is median.

Even in no. data.

$$\text{median} = \left(\frac{N}{2}\right)^{th} \& \left(\frac{N^{th}}{2}+1\right)$$

28/11/24 Thursday

median is the value in X column.

X	f	CF
5	2	2
12	3	5
17	1	6
19	5	11

median

$\frac{N+1}{2} = 6th$

- so the computation of median series.

1) Discrete series

for computing median CF are calculated first & then with their help value of the $\left(\frac{n+1}{2}\right)^{th}$ item is obtained In case of discrete series median is generally a term of series.

- CF upto ~~n stage~~ any stage is the total of all frequencies upto that stage.

Continous series

here also CF is calculated.

If first & then class interval in which median is located & then median is calculated by following formula.

$$\boxed{\text{median} = l_1 + \frac{\frac{N}{2} - c}{f} \times i}$$

l_1 = lower limit of median class

N = total No. of observation

C = cumulative freq of lower class than median class
f = freq. of median class
i = class interval

Ex

X	f	c.f.
0-10	3	3
10-20	3	6
20-30	6	12
30-40	4	16

$\frac{N}{2} = \frac{16}{2} = 8$

$m = 20 + \frac{8-6}{6} \times 10$

$= 20 + \frac{20}{6} = \frac{120+20}{6} = \frac{140}{6} = 23.3$

Merits of median
- median can be used for open ended series.
- easily calculated, understable.
- not affected by extreme absolute value.

Demerits
- not a satisfactory average when there is great variation among item of population.
- It is not capable for further algebric treatment.
- more likely to be affected by fluctuation of sampling.
- for any data median is only one & its divides the whole series into equal two parts.

MODE
- this is that value of variable which occurs most frequently or whose frequency is maximum
- in continous variable mode is calculated by following formula. here —

M	T	W	T	F	S	S

Page No.:
Date:
YOUVA

$$\boxed{\text{mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i}$$

l_1 = lower limit of modal clas ⇒ 20.
f_1 = highest frequency, freq. of modal class ⇒ 6
f_0 = frequency of class previous to model class ⇒ 3
f_2 = frequency of class next to model class ⇒ 4
i = class interval = 10

X	f
0-10	3
10-20	3
20-30	6
30-40	4

$$\text{mode} = 20 + \frac{6-3}{12-3-4} \times 10$$
$$= 20 + \frac{30}{5} = 26$$

Merits → easiest to find
- it can be obtained by just by inspection or observation.
- easily comprehensible & not affected by extreme values.

Demerits
- in many cases there is no signal & well defined mode.
- it computation isn't based on all observations.
- it isn't suitable for algebric treatment

# Symmetrical distⁿ
mean = mode = median

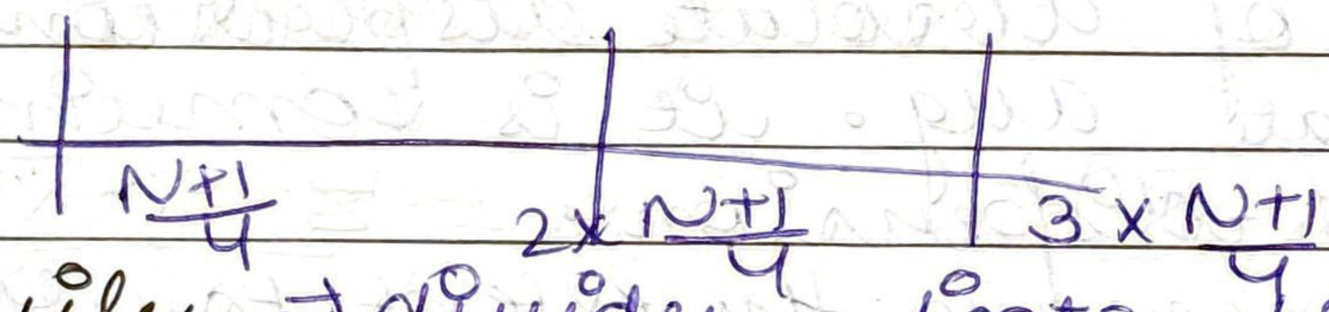
Assymmetrical distn

mean - mode = 3(mean - median)

Partition of quantiles / Partition values

- quantiles are value of variant which divide the total no. of observn into a no. of equal parts when arrn in ascending or descending order of magnitude.

Quartiles ⇒ divides the series into 4 equal parts (note this not cut mein notion)



deciles ⇒ divides into 10 parts & 5th one is equal to median.

percentile ⇒ divides into 100 equal parts 99 equal parts.

mean of first 'n' natural no. ⇒ $\boxed{\frac{N+1}{2}}$

3/12/24 Tuesday

Measure of dispersion

- for study a series, a study of extent of scattering of the observation or dispersion is also essential along with the study of the central tendency In order to throw more light on the nature of series.
- measure of variation or dispersion point out as to how far an average is representative of the mass.
- when dispersion is small the average is typical value in the sense that it closely represents the individual value &

it is reliable. On the other hand when dispersion is large the average may be quite unreliable.

- there are 2 types of measure (MOD) of dispersion

① Absolute MOD ⇒ are expressed in the same statistical unit in which the original data are given.

② Relative measure of dispersion ⇒ a measure of dispersion is ratio of a measure of absolute dispersion to a appropriate avg. It is sometimes cd. coefficient of dispersion.

- dispersion measures the extent to which the items vary from some central value. since measures of dispersion give an avg. of the differences of various items from an average they are also cd. as averages of second order.

Ⓐ Range ⇒ is the simplest measure of dispersion. It is the difference b/w highest & lowest observation of a series.

$$R = X_H - X_L$$

Ⓑ Mean deviation ⇒ mean of absolute deviation [the deviation without any '+' or '−' sign of all the observations from the average is known as Mean deviation.

$$d = X - Avg.$$

$$M.D = \frac{\Sigma d}{N}$$

when deviation is taken from arithmatic mean then it is known as mean deviation about the mean (taking deviation from mean)

$$M.D = \frac{\Sigma |X - \bar{X}|}{N}$$

$$M.D. = \frac{\Sigma f |X - \bar{X}|}{N}$$

~~Med~~ Mean deviation calculated abt the median it is least or minimum.

(3) Standard deviation → It was introduced by Karl Pearson - 1823. It is the most imp & widely used measure of deviation.
- standard deviation is also known as root mean square deviation.

$$S.D. = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}}$$

for population standard deviation (σ)

$$\sigma = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}}$$

or

$$\sqrt{\frac{\Sigma f (X - \bar{X})^2}{\Sigma f}}$$

Sample standard deviation ⇒

$$\sqrt{\frac{\Sigma (X - \bar{X})^2}{N-1}} = \sqrt{\frac{\Sigma f (X - \bar{X})^2}{\Sigma f - 1}}$$

$$S.D. = \sqrt{\frac{\Sigma f X^2 - \frac{(\Sigma f X)^2}{N}}{N-1}}$$

★★★ standard deviation divide effect by origin but effected by scale.

Variance (σ^2) / (V) → population / sample

- it is measurement of amt of variation in the data.
- the concept of variance was given by R.A. fisher 1913.
- simply it square of S.D. & that it is also known as mean square deviation.

$$\boxed{\sigma^2 = \frac{\Sigma (x-\bar{x})^2}{N}}$$

★ $$\boxed{V = \frac{\Sigma (x-\bar{x})^2}{N-1}}$$ mathematical use have good results

4/12/24 Wednesday

Quartile deviation

$$\boxed{Q.D. = \frac{Q_3 - Q_1}{2}}$$

Q_3 = third quartile
Q_1 = first quartile

$$\boxed{Q.D. = \frac{2}{3} SD}$$

Relative measure of dispersion

① Coefficient of variation (C.V.) ⇒ one of most imp relative measure of dispersion which is used to compare dispersion of 2 or more series of different units.

- it is expressed in %.

formula of $\boxed{C.V. = \frac{S.D.}{A.M.} \times 100}$

Population → Parameter. Mean = μ S.D. = σ

Variation = σ^2

★★Parameter → popn value of a character is known as Parameter & these are denoted by greek letters

- the value of C.V. is less it shows more consistency of data.
(gathering around central value).

C.V. ↓ consistency ↑ C.V. ↑ consistency ↓

★★★ Imp

Standard error (S.E.)

- standard deviation of the sampling distn of a statistics is known as standard error of that statistics.

★ Statistic ⇒ sample of value of a character.

- when all the possible samples of a same size are drawn from a popn then distn of all the samples means is known as sampling distn.

$$\boxed{S.E. = \frac{SD}{\sqrt{N}}}$$ (standard deviation)

- the term S.E. of any estimate is used for a measure of the average magnitude of the difference b/w the sample estimate & population parameter taken over all the possible sample of same size from the popn.

Probable error (PE)

- quantile deviation of sampling distn.

$$\boxed{PE = \frac{2}{3} SD}$$

—— x —— x —— x —— x ——

4/12/24 Wednesday

Ch 3

Moment & Kurtosis

about the origin means deviation from the zero.
about the arithmatic mean " " " arithmatic mean.

first moment mean ~~sum~~ mean of the deviation.
~~fir~~ second " " mean " " Square of deviation
third " " mean " " cube of deviation
four " " mean " " of 4 power " "

First moment about origin is known as arithmetic mean.
first moment about mean is always zero.

$$\frac{\Sigma|x-\bar{x}|}{N} = 0$$

Second moment about mean is known as variant (mean sq. deviation)

$$V = \frac{\Sigma(x-\bar{x})^2}{N}$$

5/12/24 Thursday r^{th} moment ⇒ Let any value is the mean of the rth power of deviations of all varied values from that value.

r^{th} moment about any value $= \frac{\Sigma(x-A)^r}{N}$

$A \rightarrow$ origin $\rightarrow$ zero

first moment at origin

$r = 1 \Rightarrow \frac{\Sigma(x-0)^1}{N}$

$= \frac{\Sigma(x)^1}{N} = \frac{\Sigma x}{N} \rightarrow$ Arithmatic mean

first moment about mean $= \frac{\Sigma(X-\bar{X})}{N}$

$$\Sigma(x-\bar{x}) = 0$$

$$\frac{0}{N} = 0$$

Moments abt A.M are known as central moments & they are denoted by μ_r.

value of first central moment is always zero

second moment abt mean

$$= \frac{\Sigma(x-\bar{x})^2}{N}$$

$$\cancel{\mu_2 \quad \Sigma(x-\bar{x}) = 0}$$

$$\cancel{\frac{0}{N} = 0}$$

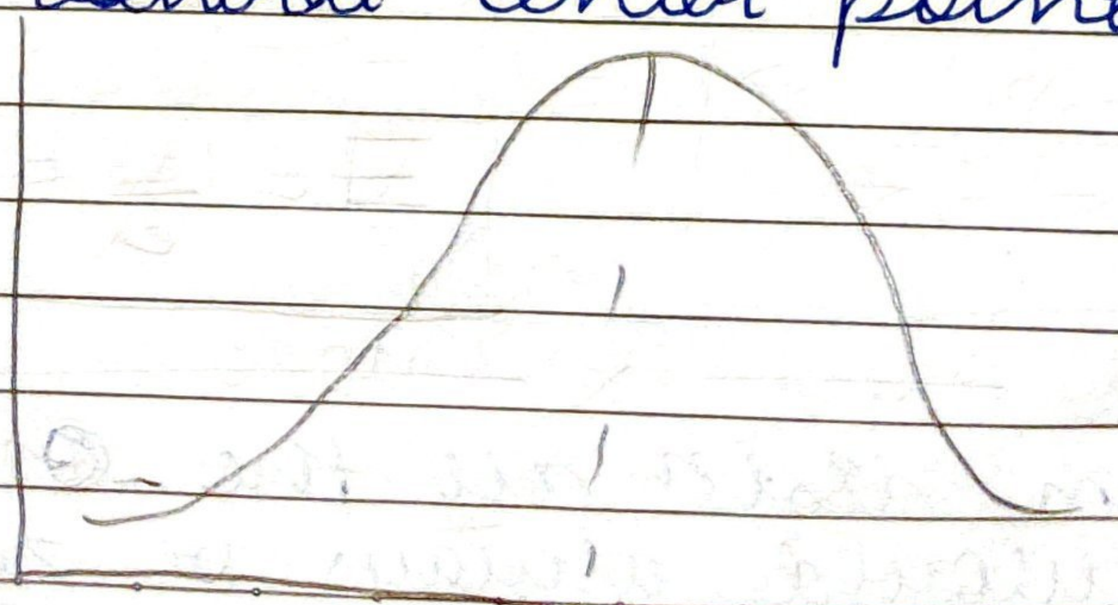
second moment is known as variance

$$\boxed{\mu_2 = \sigma^2}$$

~~Sequence~~ skewness → term 'skewness' refers to lack of symmetry i.e. when a distribution is not symmetrical it is known as skewed distn.

- concept of skewness can be understand by following description.

① Symmetrical distn → values of mean, mode & median are equal or coincide. so the spread of frequencies is the same on both sides of ~~central~~ center point of the curve.

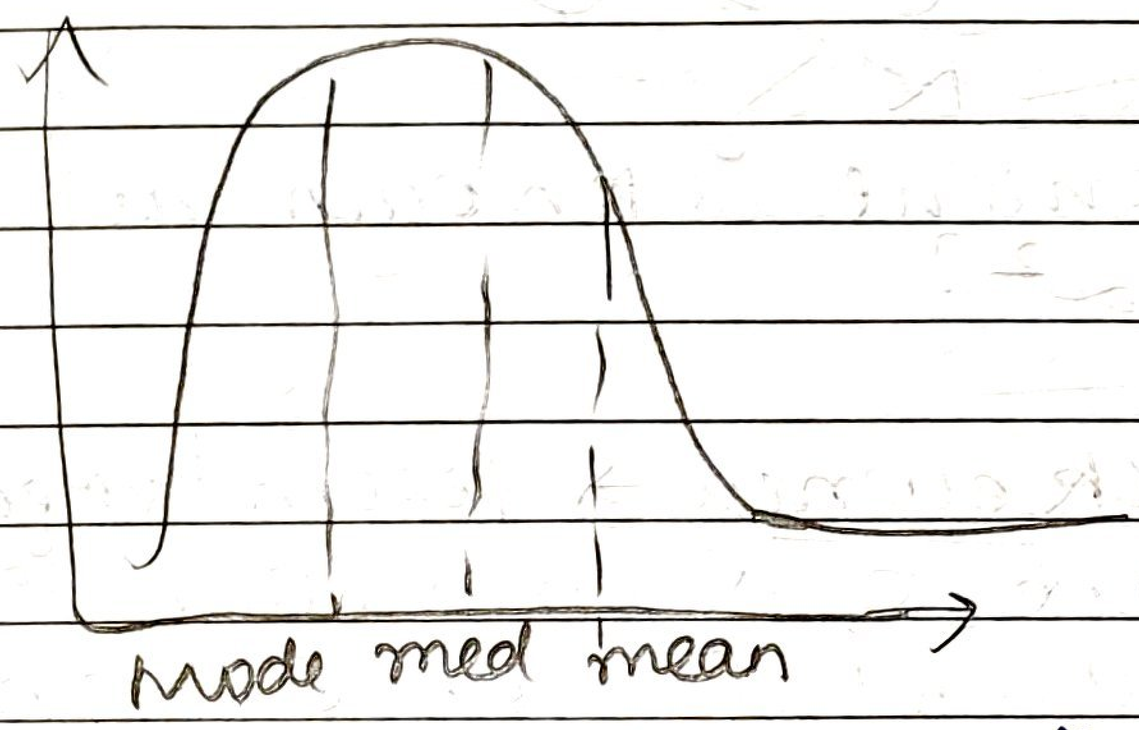


mode = mean = median

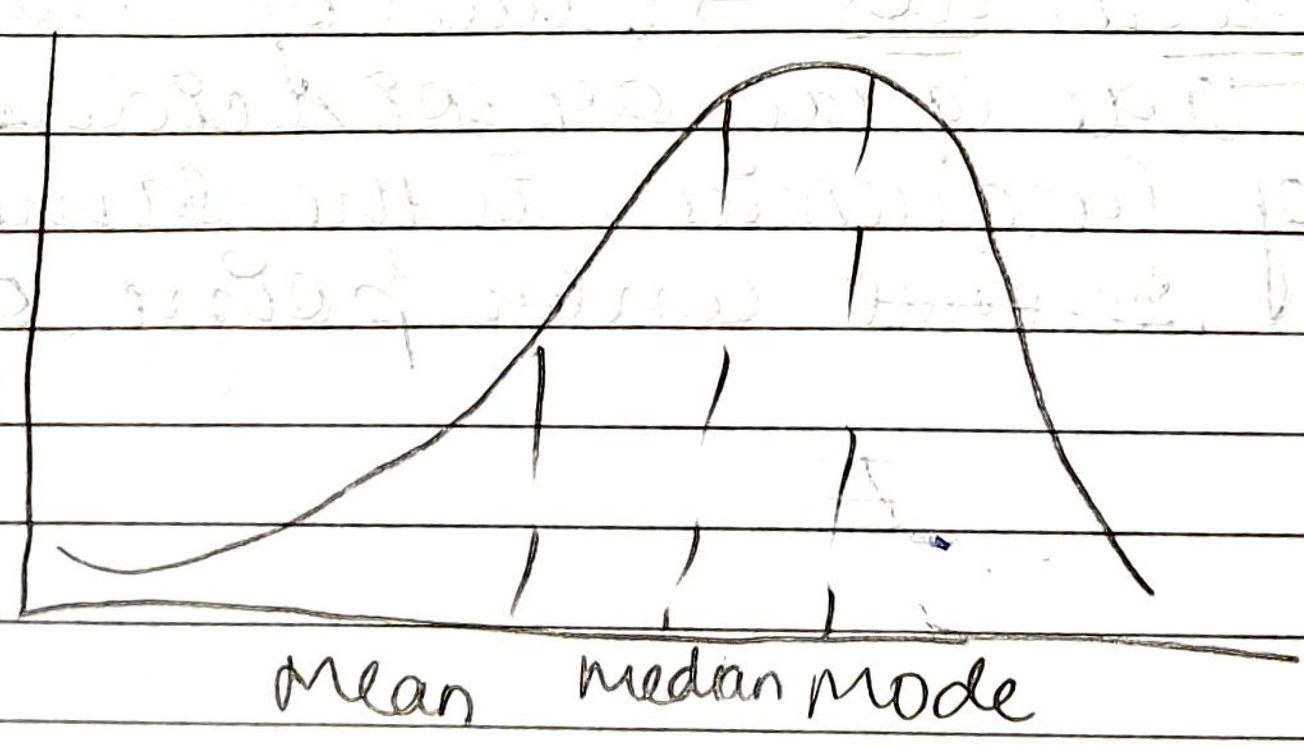
② Asymmetrical distⁿ

- a distⁿ which is not symmetrical is known as asymmetrical or skewed distⁿ
- It is of 2 types:-

① +vely skewed distⁿ ⇒ the value of the mean is maximum & that of mode is least. median lies in b/w the two. In +vely skewed distⁿ the frequencies are spread out over a greater range of value on the high value & of the curve. i.e, right hand side then they are on the low value end.



② -vely skewed distⁿ ⇒ ~~In -vely~~ value of mode is max. & that of mean is least & median lies b/w two. there is an excess tail on left hand side



- In a symm distⁿ all the odd moments i.e. μ_1, μ_3 would always be zero. but in asymmetrical distⁿ they ~~all~~ all are zero.

To measure the skewness two coefficient are used.

1) $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

for symmetrical distr^n $\beta_1 = ?$

$\beta_1 =$ zero $\quad \mu_3 = 0$

greater value of β_1 the more skewed the distribution, however β_1 cannot tell us abt the dir^n of skewness that is whether it is +ve or -ve.

hence we calculate γ_1 (Carl's Pearson's constant)

$$\boxed{\gamma_1 = \sqrt{\beta_1}}$$

now if μ_3 is +ve we has +ve skewness if μ_3 is -ve there is -ve skewness.

Kurtosis

It refers to the degree of flatness or peakedness in the region abt the mode of a frequency curve.

the normal curve is mesokurtic.

If a curve is more peaked then the normal curve then it is cd. leptokurtic.

In such a case items are more closely bunched around the mode.

- On other hand if a curve is more flat topped then normal curve it is cd as platykurtic.
- kurtosis is measured by coefficient β_2

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2}}$$

- for a normal curve the value of β_2 is three. when the value of β_2 is greater than 3 curve is leptokurtic when

When ~~of~~ less than 3
value of β_2 is ~~&~~. It is <u>platykurtic</u> &
~~also~~ another coefficient is also used
as γ_2 is also used for kurtosis.

$$\boxed{\gamma_2 = \beta_2 - 3}$$

for normal curve value of γ_2 is zero
~~of~~ $\{\beta_2 = 3 \quad \gamma_2 = 0\}$

<u>Normal curve</u> :- $\beta_1 = 0$ $\quad \beta_2 = 3$
$\mu_1 = 0$ $\quad \gamma_2 = 0$
$\mu_3 = 0$
$\gamma_1 = 0$

— X — X — X —

10/12/24
Tuesday

Probability

Girolamo Cardano (1507 – 1576) ~~first~~ an Italian mathematician was the first man to give concept on the subject entitled 'Books on game of chance' which was published after his death in 1663.

Galileo ⇒ Italian mathematican was first man to attempt quantative measure of probability.

* Systemic & scientific foundⁿ of mathematical theory of probability was led by two french mathematican B. Pascal & Picrie de fermat.

De Moivree ⇒ swiss mathematician who postulated the 'the doctarin of chance'.

Thomas Bayes ⇒ gave concept of inverse probability.

Pierre Simon de Laplace ⇒ theory of Analytical Probabilities → in 1812.

Probability
- the Probability of a given event is an expression of likelihood or chance of occurrance of an event.
- It is just possibility of happening of an event.

Experiment → any operation on certain objects which gives different results. is known as

experiment.

Outcomes ⇒ the different possible result of an experiment are known as outcomes.

Event ⇒ one outcome or group of outcomes which doesn't give unique results even it through repeated under essential ideal conditions.

Exhaustive events ⇒ total no. of all possible outcomes of ~~any~~ an event.

~~f~~ equally ~~likely~~ likely cases ⇒ events are said to be equally likely when one doesn't occur more than the others means the cases who chance of happening are equal are known as equally likely cases.

Simple event ⇒ when occurance of a single event.

Compound event ⇒ when two or more events occurs in connection with each other the joint occurance is k/a compound event.

Independent & dependent events ⇒ two or more events are said to be independent when the outcome of one doesn't affect the other & isn't affected by others.

- The dependent events are those in which occurrance or non-occurrance of one event in any one trial affects the probability of other event in other trials.

Mutually exclusive events ⇒ two events are said to be mutually exclusive or incompatible when both can't happen simultaneously in a single trial.

M T W T F S S
Page No.:
Date: Words

- they are always being connected by "either or"

Principles / theorems of Probability / Laplace

A) First principal of Laplace → If an event can happen in 'a' ways & fail to happen in 'b' ways where each of these ways is equally likelily the probability or chance of its happening is $\boxed{\frac{a}{a+b}}$ & that of failing is $\boxed{\frac{b}{a+b}}$

2) Addition law → It states that if two events A & B are mutually exclusive the probability of either 'A' or 'B' is the sum of the individual probability of A & B.

$p(A)$ = probability of occurrance of event A

$$p(A \text{ or } B) = p(A) + p(B)$$

when events aren't mutually exclusive or in other words it is possible for both events to occur ~~simult~~ simultaneously the add^n rule must be modified in following manner -

In that case :-

$$P(A \text{ or } B) = p(A) + P(B) - p(A \cap B)$$

$p(A \cap B)$ = probability of occurrance of A & B simultaneously.

3) Multiplication law → this ~~theorem~~ theorem states that if two events A & B are independent the probability that they both will occur is

equal to the product of ~~the~~ their individual probability.

$$\boxed{P(A \text{ and } B) = P(A) \times P(B)}$$

11/12/24 Wednesday

the range of probability is zero to one.
zero → impossible event
1 → sure events.

Conditional probability

for dependent events when ~~we~~ we are computing the probability of a particular event 'A' given information abt the occurrance of another event B this probability is referred to as conditional probability.

- for ex there are 2 events.
A, B
'B' is dependent on 'A'
Probability ⇒ $P\left(\frac{B}{A}\right)$

$$P(AB) \Rightarrow P(A) \times P\left(\frac{B}{A}\right)$$

ex club → $\frac{13}{52} = \frac{1}{4}$ $[P(A)]$ $= \frac{1}{4} \times \frac{4}{51}$ //

Queen ⇒ $\frac{4}{51}\left(P\left(\frac{B}{A}\right)\right)$

Permutation

- each arr^n of given things that can be made taken either all at a time or some of them at a time is known as permutation

Combination

- each of group or selection which can be made by taking at a time either or or sum of a number of things is cd. as combination.
- in forming a combination we are concerned with the no. of things each group contain but in per mutation we consider different arrn of things also.
- no. of per mutations of n dissimilar things when →

1) all the things are taken → factorial

No. of permutation = $\lfloor n$

$\lfloor 0 = 1$

$\lfloor 1 = 1$

$\lfloor 2 = 2 \times 1 = 2$

$\lfloor 3 = 3 \times 2 \times 1 = 6$

$\lfloor 4 = 4 \times 3 \times 2 \times 1 = 24$

$\lfloor 10 = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$

$\lfloor n = n \times (n-1) \times (n-2) \times (n-3)$

↳ n dissimilar thing

2) when r items things are taken at a time.

$$\boxed{\frac{\lfloor n}{\lfloor n-r}}$$

eg 5612

$$\frac{\lfloor n}{\lfloor n-r} = \frac{\lfloor 4}{\lfloor 4-2} = \frac{\lfloor 4}{\lfloor 2} = \frac{24}{2} = 12$$

- no. of per mutation on n things taken on a time when p of things are exactly alike of one kind q of them exactly
" " another kind ~~are~~ r of them "
" " third " & so on & rest are different then the permutation is –

128 18 8 6 1024 22 2 2 2 2 2 2 16 8 12

$$\boxed{\frac{\lfloor n}{\lfloor p \lfloor q \lfloor r}}$$

12/12/24 Thursday

Combination

$$^{n}C_{r} = \frac{\lfloor n}{\lfloor r \lfloor n-r}$$

* probability of happening ~~one~~ an event once, twice, thrice & so on ... in n trials when probability of its happening in one trial is known.

If probability of happening an event is p & not-happening $1-p = q$ then probability of happening exactly r times in n times

$$\boxed{^{n}C_{r}\, p^{r} \cdot q^{n-r}}$$

$$\boxed{(p+q)^{n} = p^{n} + {}^{n}C_{1}\, p q^{n-1} + {}^{n}C_{2}\, p^{2} q^{n-2}}$$

eg $N = 10$ Head $= 7$

$p = \frac{1}{2}$ $q = \frac{1}{2}$

$$^{10}C_{7} \cdot \left(\frac{1}{2}\right)^{7} \left(\frac{1}{2}\right)^{3}$$

$$^{n}C_{r} = \frac{\lfloor n}{\lfloor r \lfloor n-r}$$

$$^{10}C_{7} = \frac{\lfloor 10}{\lfloor 7 \lfloor 3}$$

the probability of ~~that~~ happening the event atleast one time → $1 - p$ (All the events fail to happen)

17/12/24
Tuesday

Probability distribution (P.D.)

- all possible outcome of a random variable together with corresponding probabilities of their frequency distribution is known as P.D.
- the P.D. which aren't obtained by actual observation on experiment but are based on mathematically derived on certain assumption are known as theoretical distribution.
- these provide models on basis of which the results of actual observation or experiment can be assessed.
- there are 3 main distribution:-

A) binomial B) Poisson C) Normal/gussian

- the first 2 distributions are for discrete variable whereas last one is for continous).

A) Binomial ⇒ It was described by Swiss mathematician James Bernoulli 1713. It is a particular case of multinomial distn having appn in research programm associated with probability & sampling.

Assumptions ⇒ 1) an experiment is performed under the same conditions for fixed number of trials i.e. 'N'.

2) in each trial there are only two possible outcome of the experiment i.e. yes or no, success or failure, true or false.

3) the probability of sucess or desired event is 'p' & is constant from trial to trial.

probability of failure is 1-p i.e. q

4) the trials ~~Statistically~~ statistically independent means a result of trial must be unaffected by result of another ~~trat~~ trial.

$r \Rightarrow$ no. of desired events.

$$\boxed{P(r) = {}^nC_r \, p^n \, q^{n-r}}$$

- probable frequency of various outcomes in 'n' trials will be

$$\boxed{(p+q)^n = p^n + nC_1 p^1 q^{n-1} + nC_2 p^2 q^{n-2} + \cdots}$$

* N set of ~~n~~ trials

$$\boxed{N(p+q)^n = N\left(p^n + nC_1 p^1 q^{n-1} + nC_2 p^2 q^{n-2} + \cdots\right)}$$

- ~~Ss~~
- binomial dist^n is symmetrical if $p = 0.5$ & if p is more than 0.5 then it is skewed to right. & if it is less than 0.5 then it is)) to left.

Properties

1. no. of trials are finite & fixed.
2. results of any trial can be defined into only on 2 categories i.e. success or failure.
3. probability of success will ~~be same~~ remain constant for each trials & all ~~trials~~ same for all trials.
4. trials are independent of each other.
5. Value of mean $= np$

   Variance $= npq$

   S.D. $= \sqrt{npq}$

B) Poisson

developed by french mathematician Simon Damis Poission in 1837. this distⁿ may be expected in cases where chance of any individual event being a success is small. this distⁿ is used to described the behaviour of rare events. such as no. of accidents on road.

properties

- discrete probability distⁿ & it is defined by probability mass function:-

$$\boxed{pmf(x) = \frac{e^{-m} m^x}{\lfloor x}}$$

here pmf (x) is probability mass function of x. ~~has~~ x is any variable having value from 0 to ∞

e is constant base logarithm e = 2.7183

m is mean of poission distrⁿ or average no. of occurance of an event.

$$\boxed{m = np}$$

n → no. of trials

p → probability of success.

- all the poission probability distⁿ are skewed to right.

now mean is the constant.

mean = m = np

variance = m

S.D. = $\sqrt{m}$

** poission distribution can be reasonable approximable of binomial when n is large & p is small & np is finite i.e. mean is finite so it known that poission is good approximation to binomial when

$n > 20$ & $p < 0.5$ then poission distribution ≡ Binomial))

18/12/24 Wednesday

Normal distribution

It was given by de moivre & later on developed by Laplace & Gauss that's why it is also known as Gaussian dist^n.

- theoretical dist^n for continous dist^n.
- It is a limiting case of binomial dist^n if no. of trials is very large but practically it should be limited.
- also ~~normal dist~~ normal Dist is approx to B.D. ~~when~~ when m is large it is equal to poission Dist^n.

Properties

- probability density function ⇒

$$\cancel{P(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}}}$$

$$\boxed{P(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}}}$$

μ = population mean

x = Value of continous random variable ~~for~~ ~~standard deviati~~

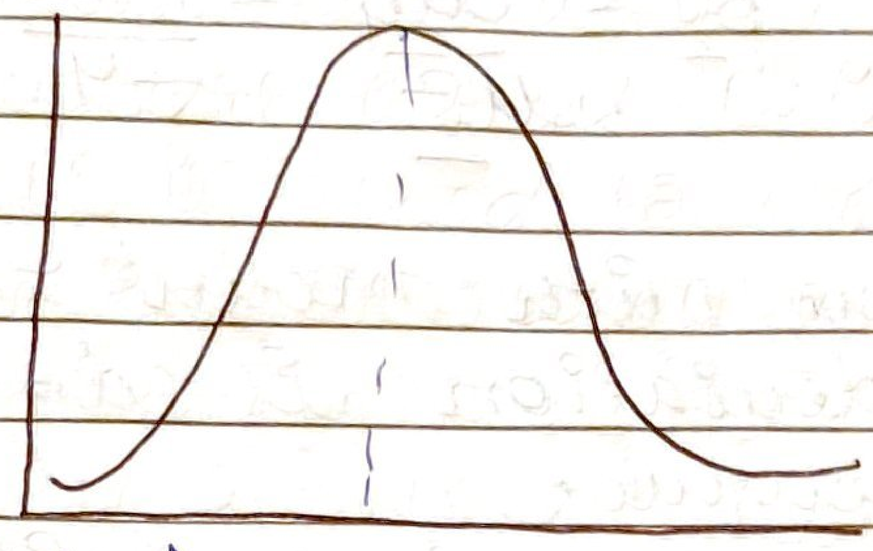
σ_x = stand deviation

σ^2 = Variable

e = logarithm constant = 2.7183

$\sqrt{2\pi}$ = ~~3.14~~ 2.5066

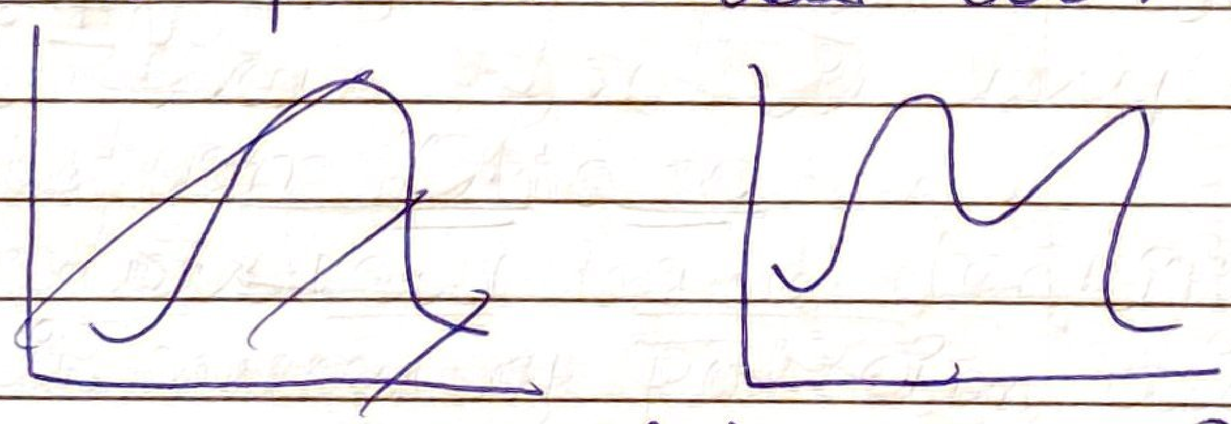
2) Normal distⁿ curve is bell shaped & symmetrical in its appearance.



3) In N.D. mean, mode, median are equal.

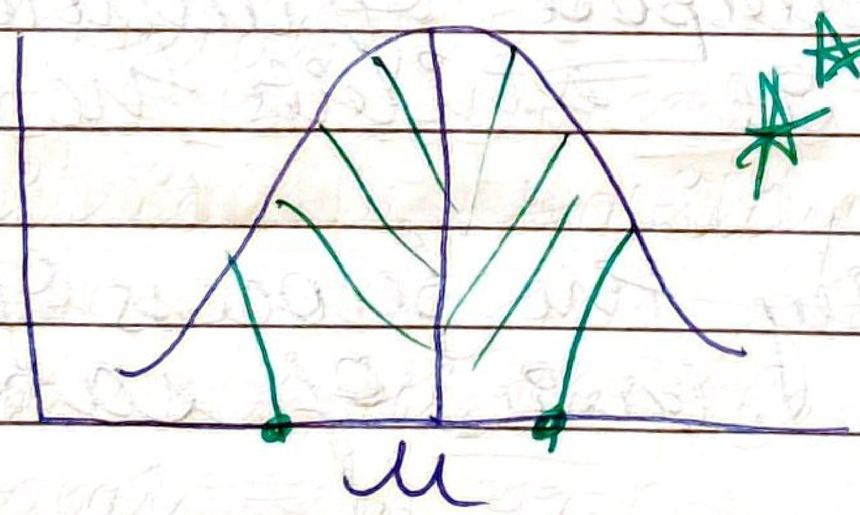
4) curve is asymptotic to base on either side means it never touches the base line.

5) normal curve is unimodel usually but in some special cases it may be bimodal.



6) area under N.D curve is described as follows :-

the area $\mu \pm \sigma$ covers 68.27%



$\mu \pm 2\sigma = 95.45\%$

$\mu \pm 3\sigma = 99.73\%$

Constant of N.D.

1) μ_1 & μ_3 are zero.

$$\mu_1 = \frac{\Sigma(x-\bar{x})}{N}$$

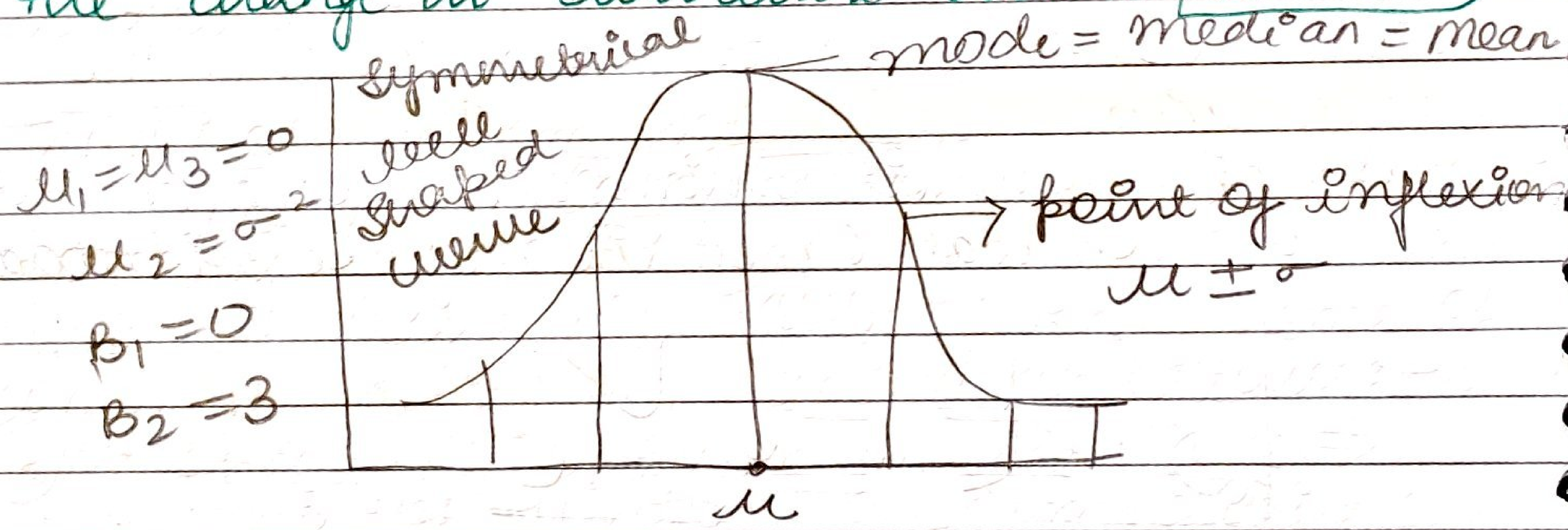
$$\mu_2 = \frac{\Sigma(x-\bar{x})^2}{N} = \text{variance}$$

$\mu_3 = 0$ $\qquad$ $\mu_4 = 3\sigma^4$

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$ $\qquad$ $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3$

★★★ a normal curve with mean zero & unit standard deviation is cd. standard normal ~~deviat~~ curve.

- the point of inflexion that is the point where the change in curvature occurs $\boxed{\bar{x} = \pm\sigma}$.



- ~~a sample~~

Sampling Method

- a sample is a part of popⁿ which is selected for a purpose of investigation considering that it exhibits the characteristics of the popⁿ.
- It is ~~the~~ simply the process of learning abt popⁿ on a basis of sample drawn from it.
- population is aggregate form of all individuals under study.

Various methods of sampling can be grouped under two heads.

① Probability / random Sampling ⇒ are those

In which every item in the popⁿ has a non-chance or probability of being chosen for a sample. This implies that selection of sample items is independent of the person making the study.

2) Non-Probability / non-random → are those which don't provide any every item in the popⁿ with the non-chance of being included in the sample.

Sampling Method

- Non-Probability
  1. Judgement sampling
  2. Quota sampling
  3. Convience ,,
- Probability
  1. Unrestricted random sampling
  2. Restricted random sampling
     - Stratified sampling
     - Systematic sampling
     - Cluster samp

19/12/24
Thursday

(SRS)
Simple Random sampling refers to that sampling technique in which each & every unit of population has an equal opportunity of being selected in the sample.

- in SRS which item get selected in the sample is just a matter of chance personal bias of investigators doesn't influence the selection.
- to ensure randomness of selection one may adopt any of following method.

1) Lottery method → very popular method of taking random sample. Under this method all items of the population are numbered or named separate slips of paper of identical shape, size & colour. These slips are then folded & mixed up in a container or drum. A blindfold selection of is made of no. of slips required to constitute the desired sample size. Selection of items thus depends entirely on chance!

(TORN)

② Table of Random numbers ⇒ lottery method becomes quite complusion as size of popⁿ ↑ses. An alternative method of R.S is that of using of Table of RN. Some of common table of RN are as follows:-

(A) Tippetts (1927) Random No. Tables ⇒ consisting of 41,600 random digits grouped into 10,400 sets of four digit distinct number.

(B) Fischer & Yates (1938) ⇒ with 15000 Random digit arranged into 1500 sets of 10 digit no.

Stratified Sampling (SRS)

- It is one of random method applies the following procedure.
- The popⁿ to be sampled is sub-divided or stratified into groups which are mutually exclusive & include all items in the popⁿ. A simple random sample is then chosen independently from each grp.

This sampling procedure differs from SRS in that in the later the sample items are chosen at random from entire popn. In SRS the sampling is designed so that a dessicanated no. of items is choosen from each stratum.

Systematic Sampling (SS)

It is formed by selecting one unit random & then selecting additional unit at evenly spaced intervals until the sample has been formed.

- this method is properly used in those cases where a complete list of popn from which sample is to be drawn is available.

Cluster Sampling [CS]

- under this method the random selection is made up of primary intermediate & final units from a given data or sample.
- there are several stages in which sampling is carried out.
- at first the first stage unit are sampled by suitable method such as simple random sampling.
- then a sample of second stage unit is selected from each of the selected first class.
- again by some suitable method which may be same as or different method from first stage unit.
- further stages may be added as required

Judgement Sampling

- In this method the choice of sample items depends exclusively on the judgement of investigator. In the sample which ~~the things~~ he thinks are most typical of popn with regard to characteristic under investigation are selected. It is also known as purposive or deliberate sampling.

Quota Sampling

- In this quota's are setup according to some specified characteristic such as so many in each of several income groups.
- each interviewer is then told to an interview a certain no. of person which consist his quota.
- within the quota the selection of sample items depends on personal judgement.
- It is often used in public opinion studies.

Convience Sampling (CS)

- a convience sample is obtained by selecting convinient popn units. The method of CS is also cal. the chunk.
- a chunk refers to that fraction of popn being investigated which is selected neither by probability nor by judgement but by convienence.
- It is often used for making pilot (or study) studies.
- Questions may be tested & preliminary information may be obtained by chunk before the final sampling design is decided upon

Tuesday

Correlation & Regression

- It is degree of relationship or assn b/w 2 variables.
- It is measured by coefficient of correlation which is denoted by 'r' given by Karl pearson
- It is unitless & its value ranges from -1 to 0 to +1.
- It may be +ve, -ve or zero.
- -1 means perfect -ve correlation
- +1 " " " +ve "
- +ve correlation means that changes in both variables will be in same dirn whereas -ve correlation means changes in opp dirn.
- Zero correlation means that both variables are independent to each other.

Types of correlation

A) On the basis of no. of traits

i) Simple Correlation ⇒ correlation study b/w 2 variables.

ii) partial " ⇒ In this more than 2 traits are taken for analysis but at a time only 2 variables are taken under consideration & rest of all others remain constant.

iii) Multiple correlation ⇒ assn study is performed for more than 2 traits at a time.

B) On the basis of dirn of change

1) Linear correlation ⇒ In this a unit change in one variable cause constant change " another " & graph is straight line.

ii) Curvilinear / Non-linear ⇒ for unit change in one variable doesn't ~~change~~ cause a constant change in another variable. graph is Curved Shaped. (Parabola)

$$\boxed{r_{xy} = \frac{Cov(X \cdot Y)}{\sigma_x \cdot \sigma_y}}$$

$$= \frac{Cov(X \cdot Y)}{\sigma^2_x \cdot \sigma^2_y}$$

$$\boxed{Cov(X \cdot Y) = \frac{\Sigma(X-\bar{Y})(Y-\bar{Y})}{N}}$$

N = no. of pairs

Properties

① the correlation coefficient is free from origin & scale.

② the correlation b/w 2 variables is symmetrical & it is unitless value ⇒ −1 to +1.

Regression

- It means going back on more
- relationship b/w 2 variables which have cause & effect property.
- so effect variable is termed as dependent variable. whereas cause variable are known as Independent variable.
- concept was given by Francis Galton 1877 whose studied the ass^n b/w father & son. "father of Regression"

dependent

- it is denoted by "b"

b_{yx} = regression of Y on X
Y - dependent X - independent
b_{xy} = regression of X on Y
X - dependent Y - independent

$$b_{xy} = \frac{Cov(X \cdot Y)}{\sigma^2 y}$$

(जो नीचे है वो नीचे होगा)

$$b_{yx} = \frac{Cov(X \cdot Y)}{\sigma^2 x}$$

★★★ $$r = \sqrt{b_{xy} \cdot b_{yx}}$$

★ correlation coefficient is geometric mean of regression coefficient.

★ regression coefficient is defined as change in dependent variable for a unit change in independent ".

- ~~It's unit is of~~ unit of regression coefficient is unit of dependent variable.
- values ranges from $-\infty$ to $+\infty$.

$$b_{xy} = r \frac{\sigma x}{\sigma y}$$

$$b_{yx} = r \frac{\sigma y}{\sigma x}$$

Correlation	Regression
- it tells abt strength & dirn	- It predict the change in dependent value due to unit change in independent value.

Correlation	Regression
correlation coefficient are symmetrical $r_{xy} = r_{yx}$	these are not symmetrical. $b_{xy} \neq b_{yx}$
- It is unitless.	unit of regression coefficient is unit of dependent variable
- Range ⇒ -1 to $+1$	Range ⇒ $-\infty$ to $+\infty$
- in this there is no distinction of cause & effect variable	there is relationship of cause & effect variable.
- It isn't affected by change of origin & scale	- not effected by change in origin but effected by change of scale.
- Correlation provide a relative measure of assn	- absolute measure of regrn.

Q. $b_{xy} = 3$ value of $b_{yx} = ?$

(a) 4 (b) 5 (c) -5 (d) 0.1

$r = \sqrt{b_{xy} \cdot b_{yx}}$

$4 = \sqrt{4 \times 3} = \sqrt{12} = 3.$

$5 = \sqrt{5 \times 3} = \sqrt{15} = 4.$

$-5 = \sqrt{-5 \times 3} = \sqrt{-15} = -4$

$0.1 = \sqrt{0.1 \times 3} = \sqrt{0.3} =$

if value of one regression coefficient is more than 1 then value of 2nd b will be less than 1.

Regression equation

regression of Y on X

$Y = a + bx$ (b_{yx})

regression of X on Y

$X = a + by$ (b_{xy})

a = intercept

b = regression coefficient

2/1/25 Thursday

HYPOTHESIS

Null hypothesis (H_0)
↳ hypothesis of no difference.

when we rejecting the hypothesis when it is true
↳ Type I error (RTI) (Reject the true type-I error)
Accepting the hypothesis when it is false
↳ Type II error

level of significance
- Maximum probability of type-I error.
ex 5%, 1%, 0.1%

Degree of freedom
N − (K) → no. of constraint
no. of values which can be assigned at your own will.

Z-Test ⇒ To compare the means when sample size is large or population SD is known.
σ → ~~large~~ known
N → large - more than 30.

$$Z = \frac{\bar{X} - \mu}{S.E.}$$

$$\text{Standard error (S.E)} = \frac{\sigma}{\sqrt{N}}$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

Ex $N = 50$ $\bar{X} = 32$ $\mu = 30$ $\sigma = 4$

$$\frac{32-30}{\frac{4}{\sqrt{50}}} \quad \frac{2}{4/\sqrt{50}} = \frac{2}{4/7} = \frac{2 \times 7}{4\ 2} = 4$$

7/1/25 Tuesday

~~t-test~~ the test in which there is no assumption abt the pop^n parameters from which samples are drawn are known as non-paramatric test. these are very much useful in behavioural science.

M T W T F S S

Page No.:

Date:

YOUVA

(χ^2)

The Kia square test is one of the most commonly used non-parametric ~~use~~ test.

- It is used in ** enumeration statistics. **
- It was first used by Karl Pearson in 1900.
- The quantity of Kia square ~~destroys~~ describes the magnitude of discrepency (difference) b/w theory & observation.
- It is computed by following formula:-

$$\boxed{\chi^2 = \sum \frac{(O-E)^2}{E}}$$

here O = observed frequency

E = expected "

- range of χ^2 value is zero to ∞.
- it always positive.

** It is statistic not a parameter.

- If the value of χ^2 is zero it means that the observed & expected frequency completely coincidence.
- The greater the value of χ^2 there is more discrepency b/w observed & expected frequency.

(observed nation)

$$\boxed{\text{Degree of freedom} = N-1 \quad (\cancel{n}\ (r-1))}$$

N ⇒ no. of categories/class (not no. of samples)

no of classes (new-1)

- The calculated value is compared with tabulated value for given degree of freedom at a certain ~~level~~ certified level of significance.
- If the calculated value is equal to or greater than tabulated value then we reject the null hypothesis & concludes that difference b/w observed & expected frequency is significant & it is

not just due to sampling of chance.
- If calculated value is less than then null hypothesis is accepted & the difference b/w expected & observed frequencies is nonsignificant & difference is just only due to chance.

~~Here~~. To test the independence of attributes the data are cross classified in a specific manner. the table giving simultaneous class^n of the body of the data in 2 different ways is known as ~~Quantigenous table~~. contigency table

row

Gender	Graduate	Post Graduate	
Male	50 (a)	20 (b)	$R_1 = 70$
female	55 (b)	17 (d)	$R_2 = 72$
	$a+c = 105$	37	
column	C_1	C_2	$a+b+c+d = N = 150$

$$\boxed{\chi^2 = \frac{(ad - bc)^2 N}{R_1 R_2 C_1 C_2}}$$

(degree of freedom)

$$\boxed{D.F = (r-1)(c-1)}$$

$\binom{M}{N}$ r = no. of row, c = No. of column

- If frequency of any cell is below 5 then some correction is needed in the Yate's correction

ad aur bc mein jo bada hai usme subtract karte hai

$$\boxed{\frac{\left(ad - bc - \frac{N}{2}\right)^2 N}{R_1 R_2 C_1 C_2}}$$

Uses of Kia square

- Kia square test is used to test the "goodness of fit" it enables us to ascertain how appropriately the theoritical distrⁿ fit with the empirical distⁿ. so, it is the "test of concordance".

- it is a test to test the Independence of attributes.
- with the help of Kia square we can find out whether two or more attributes are associated or not.
- It is also used to test the homogeneity.
- It helps to determine whether two or more Independent sample are drawn from the same popⁿ or different popⁿ.

8/1/25 Wednesday

ANOVA (Analysis of Variance)

- Observations noted from an experiment of any character show variation. this variation is due to an ~~ano~~ no. of factors known as sources of variations.
- the portion of " caused by different sources are known as components of variations.
- the ~~ano~~ ANOVA is a simple arithmatical process of sorting out the components of variations in a given data. It was developed by R. A. fischer in 1920. It is a tool by which total variation may be split up into several physical assignable components.

Role / functions of ANOVA

- It ~~states~~ sorts & estimates the variance components
- It provides the test of significance.

A) Treatment ⇒ any factor in the experiment that is controlled at different levels or values & supposed to be have an effect on the experimental units.
In other words any controlled factor or manipulations deliberately applied to the experimental units observed in exp.

B) Sum of Squares ⇒ It means the sum of squares of deviations of the variats from their mean

C) Mean Sum of Square ⇒ Synonyms of variance
It is obtained by dividing the sum of square by corresponding degree of freedom

d) error / Residual component ⇒ component of variation which remains unexplained by different sources of variation is considered due to error.

Types

- 2 types of ANOVA

① One Way
↳ in this data are classified acc to only one criterion.
- There are 2 ~~way~~ components of variation one is due to variation b/w the grps or treatment.
- & other is due to)) among the variants variates within the grp.

② Two Way

second component is known as error variance
- if μ is the overall mean & α is the component due to grp & ε (~~atto~~ alsilog) is component due to error then any observed value will be the sum of three parts.

(A) an overall mean.
B) deviation due to the grp
C) ~~trans or~~ chance error

statistical model ★★ $\boxed{X_{ij} = \mu + \alpha_i + \varepsilon_{ij}}$

X_{ij} = j^{th} variate of the ith grp
μ = overall mean
α_i = effect of ith grp
ε_{ij} = error component / chance.

F-test → is named in the honour of great statistician R.A. fischer by the George A. Snedecor
9/1/25 Thursday - object of F-test is to find out whether the two independent estimates of population variants significantly or whether the two samples may be regarded as drawn from normal pop^n having the same variants.
- for carrying out the test of significance we calculate the ratio F & formula is.

(S1 upper लोटा नीचे) $\boxed{F = \frac{S_1^2}{S_2^2}} = \boxed{\frac{\text{larger variance}}{\text{smaller variance}}}$

S_1 is always the larger ~~variance~~ estimates of variants.
- here there are 2 degree of freedom.
1) $\mu_1 = n_1 - 1$ 2) $\mu_2 = n_2 - 1$

the calculated value of F is compared with tabulated value for both degree of freedom at 5% or 1% level of significance.

- If the calculated value is higher than the tabulated value then we reject null hypothesis difference is significant & lesser then null hypothesis & difference isn't significant.

1 → null hypothesis
2 → alternate))
3 → level of significance
4 → assumption
(5) → computation → i) correction factor

T_1	T_2	T_3	T_4
15	1	7	10
21	2	8	2
32	32	11	3
45	4	12	4
50	5	14	5

$$\boxed{\text{correction factor (C.F.)} = \frac{\left(\sum x_{ij}\right)^2}{N}}$$

$\sum x_{ij}$ ⇒ sum of all observations
N → no. of samples / categories

$T_1 \neq T_2 \neq T_3 \neq T_4$

ii) Total sum of square (TSS)

$$\boxed{\sum x^2_{ij} - C.F.}$$

iii) Between sum of square (BSS)

$$\cancel{\frac{\sum T_i^2}{n_i}} \quad \boxed{\sum \frac{T_i^2}{n_i} - C.F.}$$

mean sum of squares equal to variance

(4) within Sum of square (WSS)

$$\boxed{WSS = TSS - BSS}$$

ANOVA Table

Source of Variation	Degree of freedom	Sum of sq.	mean sum of square	F-Value
(1) Between treatment of grp -	T - 1	BSS	$MSB = \frac{BSS}{T-1}$	$\boxed{F = \frac{MSB}{MSW}}$
(2) within grp / treatment error component	N - T	WSS	$MSW = \frac{WSS}{N-T}$	
(3) Total	N - 1	TSS		

15/1/26 Wednesday

Two Way ANOVA

- in 2 way ANOVA the data are classified on the basis of 2 criteria -

In this there are 2 null hypothesis
no difference b/w 1 criteria
" " " 2 "
alternate hypothesis → difference b/w 1 criteria
" " " 2 "
It is maximum probability of type-1 error →

	G_1 x2	G_2 x2	G_3 x2	sum of each grp
T_1	7 49	8 64	6	21
T_2	10 100	11 121	3	24
T_3	4 16	7	9	20
T_4	8 64	6	9	23
sum of each	29	32	27	88

Correction factor $= \frac{(\Sigma X)^2}{N} = \frac{(88)^2}{12}$

Total sum of square (TSS

$$TSS = \Sigma X^2 - C.F.$$

Between sum of square / Treatment

$$BSST = \Sigma X^2 - C.F.$$

$$BSST = \left[\frac{(\Sigma T_1)^2}{N_1} + \frac{(\Sigma T_2)^2}{N_2} + \frac{(\Sigma T_3)^2}{N_3} + \frac{(\Sigma T_4)^2}{N_4}\right] - C.F.$$

$$BSSG = \left\{\frac{(\Sigma G_1)^2}{K_1} + \frac{(\Sigma G_2)^2}{K_2} + \frac{(\Sigma G_3)^2}{K_3}\right\} - C.F.$$

Within sum of square

$$WSS = TSS - BSST - BSSG$$

$$WSS = TSS - (BSST + BSSG)$$

ANOVA Table

Source of Variation	Degree of freedom	Sum of sq.	mean sum of sq.	F-value
(1) B/w treatment	T-1	BSST	$\frac{BSST}{T-1}$	$\frac{MSBT}{MSW}$
(2) B/w grp (block) ★★★★	G-1	BSSG	$\frac{BSSG}{G-1}$	$\frac{MSBG}{MSW}$
(3) Within sum of square/component	(T-1)(G-1)	WSS	$\frac{WSS}{(T-1)(G-1)}$	
(4) Total	N-1	TSS		

Design of Experiment

- concept ~~wd~~ of design of experiment given by R.A. fischer.
- experimental design concerned the arranging of treatments or variables in a such a manner that the inferences & conclusions regarding the effects of these treatments can be easily done & their reliability is measured.

** Statistical inference ⇒ Systematic drawing of a conclusion or conclusions from given data.

16/11/25 Thursday - Principles of design of Exp
- there are 3 " " " " " .

① replication ⇒
② Randomization
③ local control

- the repitation of treatments under investigation is known as replication.
- it is essential that treatment should atleast be duplicated or preferably replicated so that an estimate of error may be obtained & the significance of the observed treatment difference is be trusted.
- repⁿ provides means for estimating error for different test of significance.
- so there 2 different advantages of replication :-

1) It ↑ses the precision by ↓ing the error.
2) It provide the estimate of error.

- the no. of replication should be such as to provide atleast 12 degree of freedom for error.

Design of Experiment

- concept ~~wa~~ of design of experiment given by R.A. fischer.
- experimental design concerned the arranging of treatments or variables in a such a manner that the inferences & conclusions regarding the effects of these treatments can be easily done & their reliability is measured.

** Statistical inference ⇒ Systematic drawing of a conclusion or conclusions from given data.

16/11/25 Thursday - Principles of design of Exp
- there are 3 " " " " " .

① replication ⇒
② Randomization
③ local control

- the repitation of treatments under investigation is known as replication.
- it is essential that treatment should atleast be duplicated or preferably replicated so that an estimate of error may be obtained & the significance of the observed treatment difference is be trusted.
- rep^n provides means for estimating error for different test of significance.
- so there 2 ~~different~~ advantages of replication :-

1) It ↑ses the precision by ↓ing the error
2) It provide the estimate of error.

- the no. of replication should be such as to provide atleast 12 degree of freedom for error.

② Randomization
- the allocation of ~~interval~~ treatments to the different experimental units by a random process.
- It minimizes the chances of biasness.
- It facilitates equal environmental conditions to different treatments.
- It should be accompanied by sufficient no. of replⁿ.
- by replⁿ & randomization valid estimate of error is obtained.

③ local control ⇒ the principal of making use of greater homogenity in a grp of experimental units for reducing experimental error.
- this is done by dividing the experimental material into homogenous grp of units.
- local control is meaningful when the treatments are replicated several times.

Advantages
① when experimental material is heterogenous w.r.t certain characters by local control. we are in posⁿ to study all the treatments for their relative merits under different sets of homogenous conditions.
② the chance error is also diminished & precision of estimates of treatment effects is ↑ed.
- by replⁿ & local control error is diminished.

I
Replication

II
Randomization

III
Local control

Valid estimate the error

diminish the error

Accuracy

the accuracy of a measurement signifies the closeness ~~which~~ with which a measurement approaches the true value / closness to true value.
- the lesser the deviation from true value the greater is the accuracy of any measurement.

Precision

- denotes closness with which measurement approaches the average value of measurement
- the lesser the variability from their average value the greater the precision of these measurements & their estimates.
- accuracy implies the lack of bias. while precision involves the control of variability.
- by bias we mean the systematic error.
- by variability means the random error.
- random error may be defined as one whose expectations is zero that is such that in a series of observations difference b/w the true value

Two way Anova ← (RBD → Randomized block design)

(population parameter) & the sample estimates ↓ses on the average as the no. of observation ↑ses.
- a systematic error or bias on other hand that will never be diminished nearly by ↑ing the no. of observations bcz that is due to such a source of error which can't be controlled by ↑ing the no. of observations or repeating the experiments unless the source of error is removed.

28/1/25 Tuesday

CRD

Completely randomized design
It is equivalent ~~two~~ to One way Anova.
- this is the simplest type of design in which whole experimental material is divided into no. of experimental units depending upon the no. of treatment & no. of replications you each.
After that treatments are allotted by units entirely by chance (Randomly)
* random allocation of treatment *

Str of ANOVa

no. of treatment = n / t
" " replication = r
total no. of experimental units = n × r / t × r = N
If treatment have vary no. of replication r_1, r_2, r_3 --- so on then total no. of experimental units = $r_1 + r_2 + r_3$ ---
there are two sources of variation -
① within treatment → component provides a basis for
② B/W "

estimation of error.
degree of freedom total = $n-1$
b/w treatment = $t-1 / n-1$
within treatment → $N-T / N-n$ N = total no. of unit
n = no. of treatment.
after this write the whole procedure of ANOVA
(put table & procedure likna hai)

** #Merits
- In this design any no. of treatments & replicates may be used.
- the no. of replicates may be vary from treatment to treatment.
- the stastical analysis of data is very easy.
- the method of analysis remain simple when result from some units are rejected or missing.
- the relative loss of information due to missing data is smaller than that with any other design.
- this design is specially useful in small experiment where the experimental material is scared / limited & homogenous.
- this design provides max. no. of degree of freedom for estimation of error as compared to other designs for a given no. of treatment & a given no. of experimental units.

demerits
- It is not much accurate as no restriction on the randomization of treatment.
** in CRD principle of local control is not followed. **

Uses → It should be used when experimental material is limited & homogenous.

when it is expected that some of units will destroyed or will fail to respond.

RBD (randomized block design).

- In this design whole exp. material is divided into homogenous grp each of which constitutes a single replication. each of these groups further divided into a no. of experimental units which are equal in all aspects.
- treatments are applied to these units by any random process.
- In this design the no. of blocks must be equal to no. of replication fixed for each treatment whereas no. of plots/individual in each block should be equal to no. of treatment.
- exp. error within each block are kept to be as small as practically possible & variation from block to block as great as possible.
- randomization of treatment in each block afresh.

Merits

when exp. material is heterogenous we go for RBD & any no. of treatment & any no. of replication.

It is ~~repli~~ equivalent to two way Anova

Degree of freedom in RBD.

(7) Grouping

(2) no. of Individuals in a block / group = no. of treatments.

no. of Individual for a treatment = no. of block / group.

in treatment = 4
group / block = 3

	T_1	T_2	T_3	T_4
B_1	X_{11}	X_{12}	X_{13}	X_{14}
B_2	X_{21}	X_{22}	X_{23}	X_{24}
B_3	X_{31}			
	$\cancel{X_{41}}$			

Two way ANOVA

$\boxed{(B-1)\ (T-1)}$

$B-1 = 2$

$T-1 = 3$

$(B-1)(T-1) = 6$

$N-1 = 11$

29/1/25
Wednesday

Computer

Characteristics of computer

① Automatic ⇒ Given a job, computer can work on it automatically without human intervention

② Diligence

③ Versatility

④ Power of Remembering

⑤ No I.Q.

⑥ No feelings

Evolution of computer

- Blaise Pascal invented the first mechanical
- Charles Babbage' is considered to be father of modern computer.

Computer Generations

- Generation in computer talk is a step.

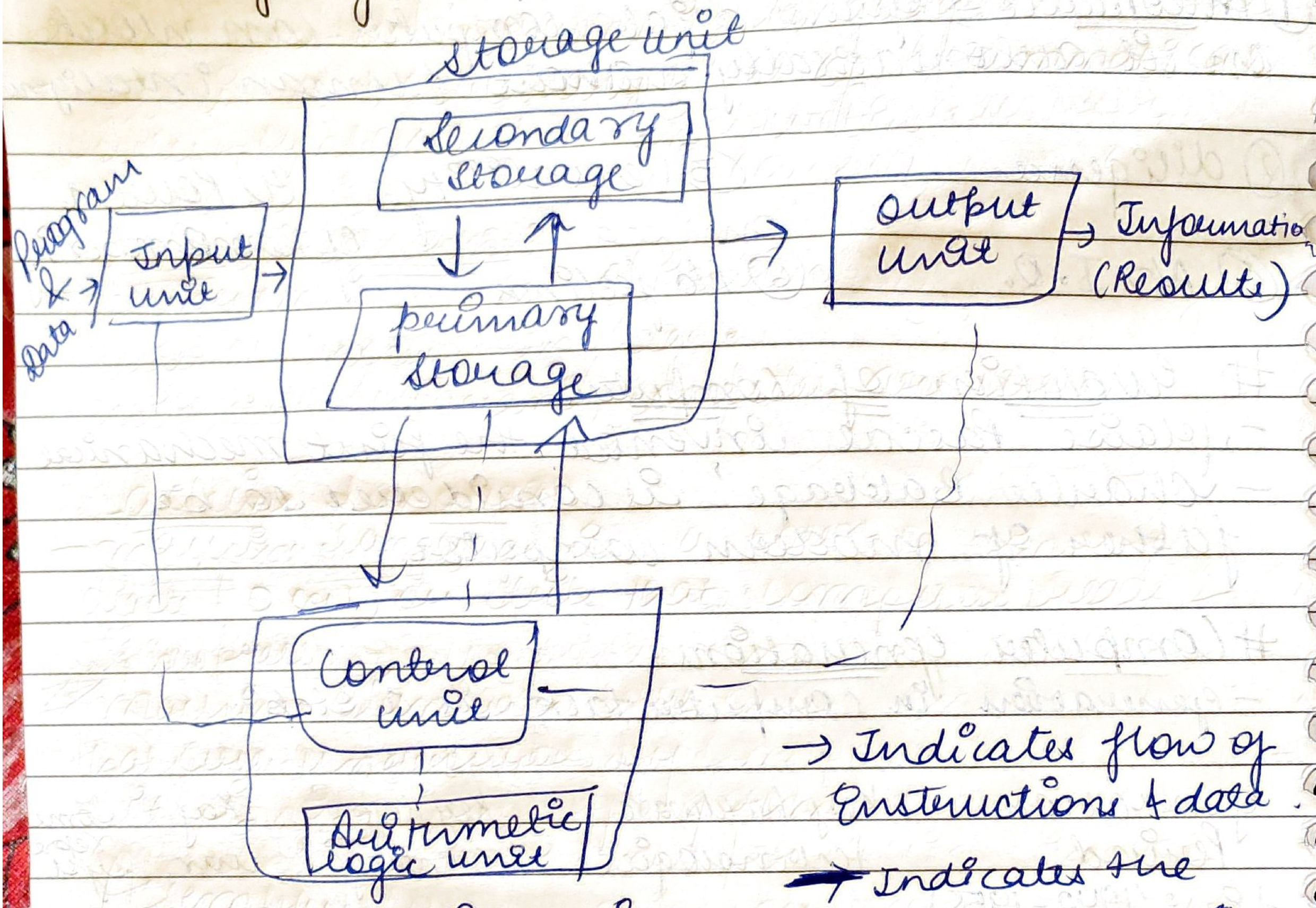
Generation Period	Key hardware technologies	Key soft technologies	Key User	Some rep syst
First (1942-1955)				
Third (1964-1975)	ICs with SSI & MSI ~~techniques~~ ~~technologies~~ large magnetic			IBM PC
Fourth (1975-1989)				
☆ Fifth (1989-present	ICS with ULSI technology	JAVA WWW		PARAM 10000

ISAR के Supra computer → Ashoka.

Garbage-In-Garbage-Out (GIGO)
Graphic User Interface (GUI)

basic organization of computer System/

Anatomy of computer



Two types of storage.

Primary storage
- used to hold running program instructions

Secondary Storage

ALU → Arithmatic logic unit
CU → Control Unit

ALU + CU → Central processing unit

Memory capacity

- 1 byte = 8 bit
- 1 Kilo bytes (KB) = 1024 (2^{10}) bytes
- 1 Megabytes = 1,048,576 (2^{20}) bytes
  1 Gigabytes = 1,073,741824 (2^{30}) bytes
  1 TB = 1024 (GB).

RAM → Random access Memory
ROM → Read only Memory

- Manufacture programmed ROM
- User Programmed ROM or Programmable ROM
- ** Erasable PROM [EPROM]
- Ultra violet

Class^n of commonly used Secondary storage devic

Secondary storage devices

- sequential Access device
  - Magnetic Type
- Direct Access device
  - Magnetic disks
    - floppy disk
    - Hard disks
      - zip disk
      - Disk pack
      - Winchester disk
  - Optical disks
    - CD ROM
    - WORM (CD-R
  - Memory storage devices

30/1/25
Thursday

M T W T F S S
Page No.:
Date:
YOUVA

Computer languages
- a grp of instructions that are used to create computer programs are known as computer languages or programming languages.
- It is a vocabulary & set of grammatical rules (syntax) for instructing.

Computer programs
a program is a set of instruction following rules of chosen language.
- without programs, computer are useless.

Machine language
- machine language is a com^n of binary digits (0 or 1) or bits that computer reads & interprets.
☆☆ machine language are only languages that are understood by computer.
- It is in binary form.

Assembly language
- a program written in assembly language consists of a series of instructions mnemonics that correspond to a stream of executable.

Assembly language programs are translated into machine language by a program cd. an assembler.
- these are second generation languages.

High-level language
- This allow us to write computer code using instructions.

Page No.:
Date:

Compiler & Interpreter
Reads the whole source code & translates it into a complete machine code program

Interpreter is a program that executes instructions written in a high-level language. An interpreter reads source code one instruction or line at a time.

FORTRAN → formula translation
ALGOL → Algorithmic language
C

COBOL → common business oriented language
SQL → structured query language.
BASIC → Beginner's all purpose symbolic instruction code.
Pascal
Logo
Hypertalk.

HTML → HyperText Markup lang.
XML → Extensible " "
WWW → World Wide Web.
URL → Uniform resource locator.

Computer → ~~comm~~
CPU → central processing unit
PDF → Portable Document Format.
~~Lor~~ DVD → Digital Versatile Disk.
HTTPS → HyperText Transfer Protocol secure.
UPS → Uninterruptible Power Supply
VIRUS → Vital Inf.
Yahoo → Yet another hierarchical officious Oracle
USB → Universal Serial Bus.
OMR →

** Internet ⇒ Is a network of computers that links many different types of computers all over the world.

E-mail → electronic mail → allow user to send a mail (message) to another Internet user in any part of world in a near-real time manner.

File Transfer Protocol (FTP) → allows user to move a file from one computer to another on the Internet.

~~Telenet → allow user to~~ Telenet.

Electronic Mail

WWW (world wide web) → HTML → special language ☆ ☆
→ HTTP

WWW browsers

IP address → Numerical Internet Protocol address is an addressing system recognised by Internet. IP address take 4 no. separated by 3 dots this no. is alloted by local networking agencies as per guidelines of Internet national Information Center.

TCP/IP → Transmission Control Protocol is a basic communication language or protocol of Internet.

It is a 2 layer programm:-

1) The higher layer transmission control protocol manages the assembly of message or file into smaller packets that are transmitted over the Internet & received by a TCP layer that reassembles the packets into original message. ~~the lo~~

2) lower layer IP handles the address part of each packet so that it gets to the right destination.