NOTES DATE

A.M:

For population the sign of mean is μ & sample $\bar{X}$

For the frequency distribution $\bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i}$

Properties of A.M.

1) The sum of the durations of the variants value from the mean is zero. $(\Sigma(x-\bar{x}) = 0)$

NOTE :- Taking deviation means subtracting from that value.

2) Sum of the squares of the deviation of variants value from their mean is least $[(\Sigma x-\bar{x})^2 = 0]$

3) If we have the Arithmetic mean & no. of observation of two or more related groups the combined mean is

$$\bar{X}_{\Sigma} = \frac{N_1\bar{x}_1 + N_2\bar{X}_2 + N_3\bar{x}_3 + \dots m_n\bar{x}_n}{n_1 + n_2 + n_3 + m_n}$$

4) A.M. Changes on change of origin & scale.

Merits of A.M.

1) It is rigidly defined.
2) It is easy to understand & calculate
3) Least affected by fluctuation of sampling.
4) It is based on all observation.

Demerits of A.M.

1) It is very much affected by extreme values
2) It's value may not coinside with any of the given value.
3) It cannot be calculated for open ended frequency distribution.
4) Neither it can be located on frequency curve nor can be Obtained by inspection.

PAGE

NOTES DATE

* Measure of central tendency

↳ tendency of data to cluster around a central point is known as central tendency.

It may also be defined as a value of variable which is thoroughly representative of the series or the distribution as a whole. The measure of this character are known as averages.

If average is a single value which is considered as the most representative or typical value for a given set of ~~time~~ data.

Objective of the study of averages -

1) To get one single value that describes the characterstics of the entire data.
2) To facilitate comparison b/w 2 or more series.

Characterstics of a good average -

1) It should be rigidly defined, means it has one & only one interpretation. The estimation should not vary from individual to individual.
2) It's computation should be based on all observations.
3) It should be as little affected by fluctuation of sampling as possible.
4) It should be easily and rapidly competable.
5) It should be easy to understand/comprehensible.
6) It should be capable further algebric treatment.

There are two types of averages:

1) Mathematical avg's ⇒ Arithmetic mean, hormonic mean, geometric mean.
2) Positional avg's ⇒ Median, mode, partition value.

① Arithmetic mean ⇒ It is the most popular and widely used average. It's value is obtained by adding together all the observation & by dividing this total by the no. of observation.

PAGE

NOTES DATE

② Geometric mean– It is the n^{th} root of the product of observations.

$$G.M. = \sqrt[N]{X_1 \cdot X_2 \cdot X_3 \cdot \dots X_n} = (X_1 \cdot X_2 \cdot X_3 \cdot \dots X_n)^{\frac{1}{n}}$$

logarithm

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

Frequency distribution

$$G.M. = \sqrt[N]{X_1^{f_1} \cdot X_2^{f_2} \cdot X_3^{f_3} \dots X_n^{f_n}}$$

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

Merits of G.M.

1) It is rigidly defined.
2) It is based on all observation.
3) It is capable of further algebric treatment.
4) If any observation have 0 value than its observation value is 0.
5) If any observation is -ve value than its computation is also meaningless [can't be calculated].
6) It is specially used in following cases–
   a) When we have to give more weights to smaller values.
   b) When we are dealing with the averages of rates of changes as in index no.
   c) In case of frequency distribution which is markedly skewed to right G.M. is used to convert it in symmetrical distribution.
   d) When we have to deal with a quantity the changes of which are proportional to the quantity itself and have to determine average rate of change in % per thousand or per million etc.

** $A.M. \geq G.M. \geq H.M.$

PAGE

NOTES DATE

③ Hormonic mean ⇒ It is the reciprocal of the mean (A.M.) of the reciprocals of observation.

* $\frac{1}{H.M.} = \frac{1}{N}\left[\Sigma \frac{1}{X_i}\right]$ or $H.M. = \frac{1}{\frac{1}{N}\left[\Sigma \frac{1}{X_i}\right]} = \frac{N}{\Sigma \frac{1}{X_i}}$

In case of frequency distribution

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

Merits of H.M.

1) It is rigidly defined.
2) Calculation is based on all observations.
3) Not much affected by fluctuation of sampling.
4) It is capable of further Algebric treatment.

Demerits of H.M.

1) Neither easily calculated nor easily comprehensible.

uses

1) It is used when the data are given in terms of units purchased per rupee or km covered per hour.

④ Weighted mean ⇒ In a condition when all the observation are not to be given importance than each item is given a proper weight as per its importance and than the average which is calculated on the basis of these weight is called as weighted mean.

$$W.M. = \frac{\Sigma W_i X_i}{\Sigma W_i}$$

Median ⇒ It is the positional average.
When the data are arranged in ascending or descending order then the middle value is known as median.

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

PAGE

NOTES DATE

- If the data are even in no. than, mean of the $\frac{N^{th}}{2}$ or $\frac{N^{th}}{2}+1$ is the median.

Computation of the median from frequency

[1] Discrete Series ⇒

For computing median C.F. are calculated 1st and than with the help of value of the $\left(\frac{N+1}{2}\right)^{th}$, item is obtained. In case of discrete series median is generally a term of the series.

C.F. upto stage is the total of all the frequencies upto stage.

[2] Continuous Series ⇒ Here also the C.f. are calculated 1st and then the class interval in which median is located is formed and than median is calculated by following formula

$$\text{Median} = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

l_1 = lower limit of median class.
c = C.f. of the lower class than the median class.
f = frequency of median class.
i = Class interval.

Ques

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

$\frac{N}{2} \Rightarrow \frac{16}{2} = 8, \Rightarrow i = 10.$

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 20 + \frac{8-6}{6} \times 10 \Rightarrow 20 + \frac{10}{3} \Rightarrow \frac{70}{3}.$$

Merits of the median

1) It can be used for open ended series.
2) It is easily calculated.
3) It is not affected by extreme abrupt value.
4) It is rigidly defined.

PAGE

NOTES DATE

Demerits of median

1) It is not a satisfactory average when their is a great variation among the items of variation.
2) It is not capable of further algebric treatment.
3) It is more likely to be affected by fluctuation of sampling.
4) For any data median is only one and it divides the series into two equal parts.

Mode ⇒ This is that value of the variable which occurs most frequently and whose frequency is maximum.

In the continuous variable the mode is calculated by the following formula

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

l_1 = lower limit of modal class.
f_1 = frequency of modal class.
f_0 = frequency of the class previous to modal class.
f_2 = frequency of the class next to modal class.
i = class interval.

Merits

1) It can be obtained just by inspection.
2) It is easily comprehensible.
3) Not affected by extreme values.

Demerits

1) In many cases there is no single and well defined mode.
2) Its computation is not based on all observation.
3) It is not suitable for algebric treatment.

NOTE :-

- In case of symmetrical distribution.

  Mean = Mode = Median.

- In case of asymmetrical distribution.

  Mean − Mode = 3 [mean − Median]

PAGE

NOTES DATE

Partition of Quartiles ⇒ Quartiles are the value of the variant which divide the total no. of observation in no. of equal parts when arranged in ascending/descending order of magnitude. Ex - Median.

Quartiles ⇒ Divides the whole series into 4 equal parts
Total no. = 3.

Deciles ⇒ Divides the whole series into 10 equal parts.
Total no. = 9.

Percentile ⇒ Divides whole data into 100 equal parts.
Total no. = 99.

* Mean of the first 'n' natural number = $\frac{N+1}{2}$.

Measure of Dispersion

For Study a series, a study of the extent of scattering, 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 the 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 a typical value in the sense that it closely represents the individual value and it is reliable. On the other hand when dispersion is large the average may be quite unreliable.

There are two types of measure of dispersion:

① Absolute measure of dispersion ⇒ Absolute measure of dispersion are expressed in the same statistical unit in which the original data are given.

② Relative measure of dispersion ⇒ A measure of relative dispersion is the ratio of a measure of a abosolute measure of dispersion to an appropriate average. It is sometimes called the Co-efficient of dispersion.

PAGE

NOTES DATE

Dispersion measures the extent to which the items vary from some central value since measures of dispersion give an average of the differences of various items from an average they are also called as averages of the II^{nd} order.

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

$$R = X_H - X_L$$

Mean deviation ⇒ Mean of absolute deviations (the deviation without any +ve or −ve sign) of all the observation from the averages is known as mean deviation.
When deviation is taken from A.M. it is known as 'mean deviation about the mean'.

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

* M.D. When calculated about the median it is least or minimum.

Standard deviation ⇒ Concept given by Carl Pearson in 1823.
- It is most important and widely used in measure of dispersion.
- It is also known as root mean square deviation.

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

Population standard deviation (σ)

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

Sample S.D. ⇒ $\sqrt{\frac{\Sigma(X-\bar{X})^2}{N-1}} = \sqrt{\frac{\Sigma fx^2 - \frac{(\Sigma fx)^2}{N}}{N-1}}$

* S.D. didn't affected by origin but affected by the scale.

PAGE

Statistic ⇒ sample value of a character.

Parameter ⇒ The population value of a character.

NOTES DATE

Variants ⇒ It is the measurment of amt. of variation in data.

→ concept given by R.A. fisher in 1913.

- Simply variance is the square of S.D. So it is known as Mean square deviation.

Population (V) ⇒

$$\sigma^2 = \frac{\Sigma fx^2 - \frac{(\Sigma fx)^2}{N}}{N-1}$$

Sample (V)

$$V = \frac{\Sigma (x-\bar{x})^2}{N-1}$$

Quartile deviation ⇒ $Q_D = \frac{Q_3 - Q_1}{2}$ ∴∘ $QD = \frac{2}{3} S.D.$

Relative measure of dispersion / Co-efficient of variation (cv)

↳ one of the most important relative measure of dispersion. Which is used to compare the dispersion of two or more distribution having diff. units.

⇒ It is expressed in % tage.

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

- The value of C.V. is less it shows more consistency in the data.

Standard error ⇒ S.D. of the sampling distribution of a statistics is known as Standard error (S.E.) of that statistics.

$$S.E. = \frac{S.D.}{\sqrt{N}}$$

Sampling distribution ⇒ When all the possible samples of same size are drawn from a population than the distribution of all the sample means sampling distribution.

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

PAGE

NOTES DATE

Probable error ⇒ Quartile deviation of sampling distribution.

$P.E. = \frac{2}{3} S.E.$

CH-3 [objective & short ques mainly in exam].

Moments ⇒ 1st moment about the origin
↳ deviation from the 0.

About the Arithmetic mean
↳ deviation from A.M.

first moment = Mean of the deviations.
Second moment = Mean of square of deviation.
third moment = Mean of the cube of deviations.
fourth moment = Mean of the four power of deviation.

- first moment about the origin is A.M.
- r^{th} moment → r^{th} moment about any value is the mean of the r^{th} power of deviations of all varied value from that value

$$= \frac{\Sigma (x-A)^r}{N}$$

So, if A is the origin i.e. $A = 0$ $r = 1$

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

* 1st moment about the origin = A.M.
* 1st moment about the mean = 0. $\because \mu_1 = \frac{\Sigma (x-\bar{x})}{N} = 0$ $\because \Sigma (x-\bar{x}) = 0$
- The moments about the A.M. are k/a Central moments and denoted by μ.
- $\mu_2 = \sigma^2 = \frac{\Sigma (x-\bar{x})^2}{N}$ So, IInd Central moment is variance.

Skewness :- The term Skewness refers to lack of symmetry i.e. when a distribution is not symmetrical it is called as Skewed distribution.

PAGE

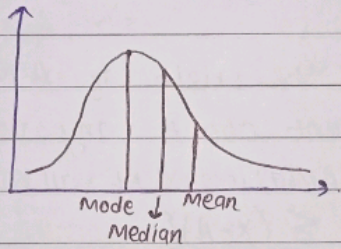
NOTES DATE

The concept of skewness can be understand by following discription:

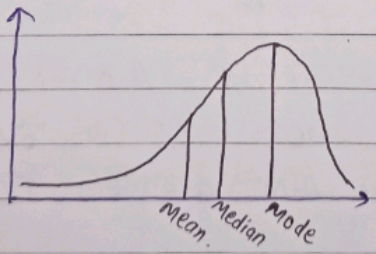
① Symmetrical Distribution ⇒ In symmetrical distribution the values of mean, mode and median are equal or co-incides so the spread of frequencies is the same on both sides of the centre point of the curve.

② Asymmetrical distribution ⇒ A distribution which is not symmetrical is k/a asymmetrical / skewed distribution. It is of two types:

(i) Positively skewed distribution ⇒ In this the value of the mean is maximum and that of mode is least. The median lies in btw the two. In this distribution the frequencies are spread out over a greater range of values on high value end of the curve i.e. right hand side than they are on the low value end.



(ii) Negatively skewed distribution ⇒ In this distribution the value of mode is maxim and that of Mean is least and median lies in btw two. There is a excess tail on left hand side.



In a symmetrical distribution all the odd moments would always be zero. But in asymmetrical they are not zero. So, to measure the skewness two co-efficient are used.

PAGE

NOTES DATE

(i) $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ NOTE for symmetrical distribution $\beta_1 = 0$ b/c $\mu_3 = 0$.

- The greater the value of β_1 the more skewed the distribution however β_1 cannot tell us about the direction of skewness. i.e. whether it is +ve or -ve.

Hence we calculate the γ_1 $[\gamma_1 = \sqrt{\beta_1}]$

So, now if μ_3 is +ve we have +ve skewness.
if μ_3 is -ve than there is -ve skewness.

Kurtosis ⇒ It refers to the degree of flatness or peakedness in the region about the mode of a frequency curve. The normal curve is called mesokurtic.

- If curve is more peaked than the normal curve it is k/a leptokurtic.
- In such a case items are more closely bunched around the mode. on the other hand, if a curve is more flat topped than the normal curve it is k/a platykurtic.

So, for a normal curve the kurtosis is measured by co-efficient β_2.

$$\beta_2 = \left(\frac{\mu_4}{\mu_2}\right)^2 \quad \because \text{for a normal curve } \beta_2 = 3$$

- If value of β_2 is greater than 3 the curve is leptokurtic.
- If value of β_2 is smaller than 3 the curve is platykurtic.
- An another coefficient γ_2 is also used for kurtosis:

$$\gamma_2 = \beta_2 - 3.$$

So for a normal curve value of $\gamma_2 = 0$
If γ_2 is +ve than curve is leptokurtic
If γ_2 is -ve than curve is platykurtic.

PAGE

NOTES

Probability

DATE

Girolama Cardano [1501-1576] → An italian Mathematician was the first men to write the book on 'Book of games of chance' which has published after his death in 1663.

Gallileo ⇒ He was also an italian mathematician was the first men to attempt quantitative measure of Probability.

Systematic & Scientific foundation of Mathematical theory of Probability was led by 2 french mathematician i.e B-Pascal & Pierre de fermat.

De-Moivre ⇒ He was a swiss Mathematician who postulated the doctorines of chance.

T-Bayes → Gave concept of inverse probability.

Pierre Simon de laplace - ⇒ He gave the 'Theory of Analytical Probabilities' [1812].

Probability ⇒ Probability of a given event is an expression of likelyhood or chance of occurance of an event

or

It is just the possibility of happening of an event.

Experiment → Any operation on certain objects which gives different results is k/a experiment.

Outcomes → The different possible results of an experiment are k/a outcomes.

Event → one outcome or grp of outcomes which does not give unique results even it though repeated under essential ideal conditions.

Exhaustive event → Total no. of all possible outcomes of an event.

PAGE

NOTES DATE

Equally likely cases → events are said to be equally likely when one does not occur more often then the others means the cases whose chance of happening are equal are known as Equally likely cases.

Simple event → When occurance of a single event is k/a simple event.

Compound event → When two or more events occur in Connection with each other the joint occurence is called Compound event.

Independent & dependent events → 2 or more events are said to be independent when the outcome of one does not affect the other and is not affected by others.

The dependents events are those in which occurence/non occurence of one event in any one trial affects the probability of other event in other trials.

Mutually exclusive events → 2 events are said to be mutually exclusive or incompatible when both cannot happen simultaneously in a single trial. Mutually exclusive events are always be Connected by the either or

PAGE

NOTES DATE

Principles of probability or laplace ⇒ The first principle of laplace. If an event can happen in 'A' ways & fail to happen in B ways where each of these ways is equally likely the probability (chance of its happening is $\frac{a}{a+b}$ & that of failing is $\frac{b}{a+b}$.

II nd law ⇒ 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'. When events are not mutually exclusive or in other words it is possible for both events to occur simultaneously the Addition rule must be modified in following manner-
in that case

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

$P(A \cap B)$ = polarity of occurence of A or B simultaneously.

Multiplication law ⇒ This theorem states that if two events A & B are independent the probability that they both will occur is equal to the product of their individual probability.

$$P(A \& B) = P(A) \times P(B)$$

* The range of probability is 0 to 1.

Conditional probability ⇒ For dependent events when we are computing, the probability of a particular event(A) or given information about the occurence of another event 'B' this probability is referred to as Conditional probability.

Probability of occuring of (A) ⇒ $P\left(\frac{B}{A}\right)$

$$P(A \cdot B) = P(A) \times P\left(\frac{B}{A}\right)$$

PAGE

NOTES DATE

Permutation ⇒ Each arrangement of given things that can be made taken either all at a time or sum of them at a time is k/a permutation.

Combination → Each of the group/selection which can be made by taking either all at a time or sum of a no. of things is called as combination.
In forming a combination we are concerned with a no. of things each group contain but in permutation we consider diff. arrangement of the things also.

→ No. of Permutations of 'N' dis-similar things when
i) All things are taken

$$\text{No. of permutation} = \underset{\text{factorial}}{\lfloor n} \qquad \left[\begin{matrix} \lfloor 0 = \infty \\ \lfloor 1 = 1 \end{matrix}\right]$$

Ex. $\lfloor 2 = 2\times1 = 2$
$\lfloor 3 = 3\times2\times1 = 6$
$\lfloor 4 = 4\times3\times2\times1 = 24$
$\lfloor 5 = 5\times4\times3\times2\times1 = 120$
$\lfloor 10 = 10\times9\times8\times7\times6\times5\times4\times3\times2\times1 =$

ii) When 'r' items/things taken at a time

$$\text{No. of permutation} = \frac{\lfloor n}{\lfloor n-r}$$

When the no. of permutation of N things taken all at a time When 'p' of the things are exactly alike of one kind 'q' of them exactly alike another kind are of them exactly alike of another kind & so on or the rest are different. then the permutation is $\frac{\lfloor n}{\lfloor p \lfloor q \lfloor r}$.

Probability of happening of an event, once, twice, thrice & so on in n trials when probability of its happening in one trial is known, If the probability of happening an event is 'p', not happening is '1-p = q', then the probability of happening exactly r times in n trials is

PAGE

NOTES DATE

$nEr\ p^r q^{n-r}$

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

The probability of happening the event at least one time = 1-P [All the event fail to happen]

* Probability distribution

↳ All possible outcomes of a random variable together with Corresponding Probabilities of their frequency distribution is k/a probability distribution. The distribution which are not obtained by actual experiment or observations but are based on Mathematically derived in certain assumptions are k/a theoritical distributions These provides models on basis of which the results of actual observation or experiment can be assesed.

There are 3 types of distributions

① Binomial ② Poisson ③ Normal/Gussian.

The first two distributions are for the discrete distb^n Whereas Normal distb^n for continuous variable

① Binomial ⇒ It is described by swiss mathematician James Bernoulli in 1713.

⇒ It is particular case of multinomial distb^n having application in research programmes associated with Probability & sampling.

Assumption ⇒

i) An experiment is performed under the same cond^n for fixed no. of trials i.e n.

ii) In each trial their are only 2 possible outcomes of Experim^{nt}

iii) The probability of success or desired event is 'p' & it is constant for trial to trial.

iv) The probability of failure is 1-P = q.

v) The trials are statistically independent means the result of the trial must be unaffected by the result of another trial.

PAGE

NOTES DATE

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

probable frequency of various outcomes is

$$(p+q)^n = p^n + {}^nC_1 p^1 q^{n-1} + {}^nC_2 p^2 q^{n-2} + {}^nC_3 \ldots\ldots$$

N set of N trials

$$N(p+q)^n = N\left(p^n + {}^nC_1 p^1 q^{n-1} + {}^nC_2 p^2 q^{n-2} + {}^nC_3 \ldots\ldots\right)$$

- Binomial distb^n is symmetrical if $p = 0.5$ and if p is more than 0.5 than it is skewed to right if it is less than 0.5 than it is skewed to left.

* Properties of Binomial distb^n

i) No. of trials are finite & fixed.
ii) Results of any trial can be defined into only on 2 categories i.e success / failure.
iii) Probability of success remain constant and same for whole trials.
iv) Trials are independent to each other.
v) The value of mean is NXP and the variance is the npq and S.D. is $\sqrt{npq}$.

(2) Poisson Distb^n ⇒ It was developed by a french mathematician 'simon Davis poisson' in 1837.
This distribution may be expected in cases where the chance of any individual event being a success is small. This distribution is used to describe the behaviour of rare events such as accidents on roads.

Properties

It is a discrete probability distribution and defined by Probability mass function.

$$pmf[x] = \frac{e^{-m} m^x}{L^x}$$

Pmf for variable x, X is the any variable having value 0 to ∞,

PAGE

NOTES DATE

e is the constant based on natural logarithm ($e = 2.7183$),
m = mean of the poisson distribution of mean of the no. of occurence of events. $m = nP$ (n: no. of trials; P: Probability of success)

All the poisson probability distribution are skewed to the right, now the mean is → $m = np$ and variance is M & S.D. $= \sqrt{M}$.

- Poisson distribution can be reasonable approximation of the binomial when 'n' is large and 'P' is small and np is ∞.

So, it is known that poisson is a good approximation of Binomial when $n \geq 20$ & $P \leq 0.5$.

③ Normal distribution ⇒ It was given by de moivre & later on develop by Laplace & Gauss. It is the theoritical distribution for continuous variable. It is limiting care of Binomial distribution, if no. of trials is very large but practically should be limited. Also normal distribution is approximated to poisson distribution when 'n' is large.

Properties

(i) Normal distribution can be defined by probability density function:

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

μ = Population mean, x = Value of continuous random variable
σ = Variance, e = mathematical logarithm ($e = 2.7183$), $\pi = 3.14$
$\sqrt{2\pi} = 2.5066$, σx = S.D.

** Normal distribution curve is bell shaped & symmetrical in appearance.

- In the normal distribution mean, median, mode are equal.
- The curve is asymptotic to the base on either side means it never touches the base line.
- Normal curve is unimodal usually but in some special cases it may be bimodal.

PAGE

NOTES DATE

- The area under the normal distribution curve is described as follows:

The area, $\mu \pm \sigma$ covers 68.2% area & $\mu \pm 2\sigma = 95.45\%$ & $\mu \pm 3\sigma = 99.73\%$

Constant of the normal distribution

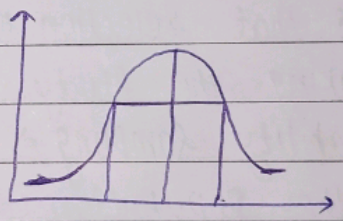
$\mu_1 = 0$

$\mu_1 = \frac{\Sigma[x - \bar{x}]}{N}$, $\mu_2 = \frac{\Sigma[x - \bar{x}]^2}{N}$ = variance = σ^2

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

The normal curve with the mean 0 & unit standard deviation is called as a Standard normal curve.

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



NOTES

Sampling

DATE

Sampling method

* Sample = A sample is a part of population which is selected for the purpose of investigation considering that it exhibits, the characterstics of population.

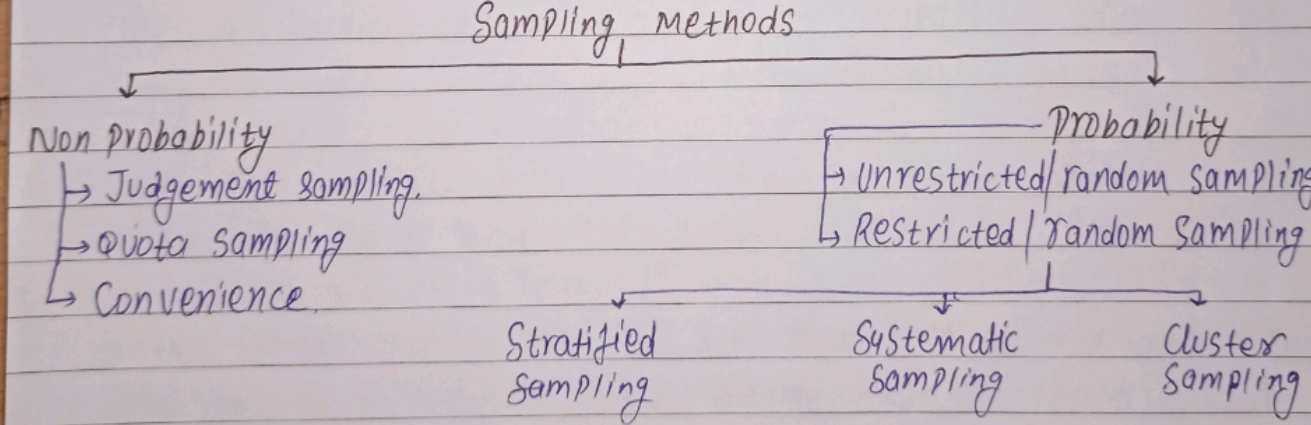
- Sampling is simply the process of learning about the population on the basis of population drawn from it.

* Population = Aggregate form of all the individuals under study.

- The various method of sampling can be grouped under two heads:
  1) Probability / Random Sampling 2) Non probability / Non random.

1) Probability sampling ⇒ Methods in which every item in the population has a known chance of being chosen for the sampling. This implies that selection of sample items is independent of the person making the study.

2) Non probability sampling ⇒ Methods which do not provide every item in the population of known chance of being included in the sample.



- Simple random sampling refers to that sampling technique in which each and every unit of population has an equal opportunity of being selected in the sample. In simple random sampling which item get selected in the sample is just a matter of chance. Personal bias of investigator does not influence the

PAGE

Selection. So to ensure randomness of selection one may adopt any of the following method :-

(1) Lottery method ⇒ It is very popular method of taking random sample, under this method all items of the population are numbered or named on seperate, slips of paper of identical size, shape and colour. These slips are than founded and mix them in a container or drum. A blind fold selection is than made of the no. of slips required to constitute the desired sample size. The selection of items thus depends entirely on Chance.

(2) Table of random No's ⇒ The lottery method become quite combustion as the size of population ↑ses. An alternative method of random selection is that of using of random no's Some of the common table of random no's are as follows.

1.) Tipett's [1927] ⇒ random no. table 41,600 random digits grouped into 10,400 sets.

2) fischer & yates [1938] ⇒ Table of random no's with 1500 random digits arranged into 15000 of 10 digit no.

Stratified sampling ⇒ one of the random method applies the following procedure. The population to be sampled is subdivided or stratified into groups which are mutually exclusive and include all items in the population. A simple random sample is then Choosen independently from each group. This sampling procedure differs from the simple random sampling in the later, the sample items are Choosen at random from the entire population. In stratified random sampling, the sampling is designed a designated no. of items is Choosen from each stratum.

Systematic sampling ⇒ It is formed by selecting one unit at random & then selecting additional units at evenly spaced internals

Until the sample has been formed. This method is properly used in these cases where a complete list of population from which sample is to be drawn is available.

Cluster or multistage Sampling ⇒ Under this method the random selection is made of primary intermediate & finally units from a given population or stratum. Their are several stages in which sampling process is carried out.
At first the first stage units are sampled by some suitable methods such as simple random sampling. Then a sample of second stage is selected in each of the selected first stage unit. Again by suitable method which may be same as or may diff. from the method employed from first stage unit. . further stages may be added as required.

Judgement Sampling ⇒ In this method of sampling the choice of sample items depends exclusively on the judgement of the investigator. In the sample which he thinks are most typical of the population with regard to the characterstics under investigation are selected. It is also k-a Purposive or deleberate sampling.

Quota Sampling ⇒ In a quota sample, quota's are setup according to some specified characteristics such as so many in each of several income groups. Each interviewer is then told to an interview a certain no. of person which constitute his quota. Within the quota the selection of sample items depends on personal judgement.
It is often used in public opinion studies

NOTES DATE

Convenience Sample ⇒ A convenience sample is obtained by selecting convenient population unite. The method of convenience sampling is also called the chunk. A chunk refers to that fraction of population being investigated which is selected neither by probability nor by judgement but by convenience.

- It is often used for making pilot studies questions may be tested and the preliminary information may be obtained before the final the sampling design is decided upon.

Correlation & Regression

NOTES DATE

Correlation ⇒ It is the degree of relationship or association between two variables.

It is measured by Co-efficient of Correlation which is denoted by 'r'. it was given by karl Pierrson.

- It is unitless and it's value ranges from −1 to +1.
- It may be +ve, −ve or zero.
- −1 means Perfect −ve correlation, +1 means Perfect +ve correlation. +ve correlation means that Changes in both variables will be in same direction whereas −ve correlation means the Changes will be in opposite direction.
- Zero correlation means ($r=0$) that both variables are independent to each other.

Types of Correlation −

On the basis of no. of traits

1) Simple correlation ⇒ Correlation Study btw two variables.
2) Partial Correlation ⇒ In this more than two traits are taken for analysis but at a time only two variables are taken under Consideration and rest of all others remain Constant.
3) Multiple correlation ⇒ In multiple correlation, association study is performed for more than two traits at a time.

On the basis of dirⁿ of Change

1) Linear correlation ⇒ In this Correlation a unit Change in one variable cause Constant Change in another variable and the graph is in straight line.
2) Curvilinear/Non-linear ⇒ In this correlation for the unit Change in one variable does not Cause a Constant Change in another variable and Shape of the Curve is Parabolic.

So,

$$r = \frac{Cov[xy]}{\sigma_x \cdot \sigma_y} \quad \text{or} \quad \frac{Cov[x \cdot y]}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}}$$

PAGE

NOTES DATE

$$Cov[x,y] = \frac{\Sigma(x-\bar{x})\cdot(y-\bar{y})}{N}$$ N = no. of pairs.

Properties:
i) Correlation coefficient is free from origin & scale.
ii) The correlation coefficient btw two variables is symmetrical.
iii) It is unitless.

Regression ⇒ It means going back.
Regression is the relationship btw the two or more variables which have cause and effect properties.
The effect variable termed as dependable variables whereas cause variable termed as independable variables.
- The concept of regression was given by Sir Francis Galton in 1885. He also kla as father of biostatistic.

b_{yx} = regression of Y on X , Y = dependable or X = Independable.
b_{xy} = regression of X on y. , X = dependable or Y = Independable.

$$b_{xy} = \frac{Cov[x,y]}{\sigma_y^2} \quad , \quad b_{yx} = \frac{Cov[x,y]}{\sigma_x^2}$$

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

- Regresion coefficient is defined as change in dependent variable for a unit change in Independent variable.

Unit of regression coefficient is unit of dependent variable.
Means, Unit of regression = unit of dependent variable.
* It's value ranges from $-\infty$ to $+\infty$.

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} \quad , \quad b_{yx} = r\frac{\sigma_y}{\sigma_x}.$$

PAGE

NOTES DATE

Correlation	Regression
1. Correlation tells about the Strength and direction of association.	1. Regression predict the change in Independent value due to unit change in Independable variable.
2. Correlation coefficient is symmetrical ($\gamma_{xy} = \gamma_{yx}$)	2. Regression coefficient are not symmetrical ($b_{xy} \neq b_{yx}$).
3. It is unitless.	3. Unit is unit of dependable variable
4. It's range is from -1 to $+1$.	4. It's ranges from $-\infty$ to $+\infty$.
5. There is no distinction of cause and effect variable.	5. There is a relationship of cause and effect variable.
6. It is not affected by change in origin and scale.	6. It is not affected by change in origin but affected by change in scale.
7. It provides a relative measurment of association.	7. It provides an absolute measure of regression.

Regression equation ⇒

Regression of 'y' on 'x'

$$y = a + b_{yx} \cdot x$$

Regression of 'x' on 'y'

$$x = a + b_{xy} \cdot y$$