STATISTICS

DEFINITION OF STATISTICS

SINGULAR SENSE : **scientific method that is employed for collecting, analysing and presenting data to draw statistical inferences**

PLURAL SENSE : **data qualitative as well as quantitative, that are collected, usually with a view of having statistical analysis.**

Language	Word
LATIN	STATUS
ITALIAN	STATISTA
GERMAN	STATISTIK
FRENCH	STATISTIQUE

LIMITATIONS OF STATISTICS

I. Statistics deals with the aggregates and not individual data.

II. Statistics is concerned with quantitative data. However, qualitative data also can be converted to quantitative data by providing a numerical description to the corresponding qualitative data.

Quantitative information shown as number

DATA

PRIMARY

SECONDARY

The data which are collected for the first time by an investigator or agency

collected data used by a different person or agency.

Variable is a measurable data

VARIABLE

DISCRETE VARIABLE

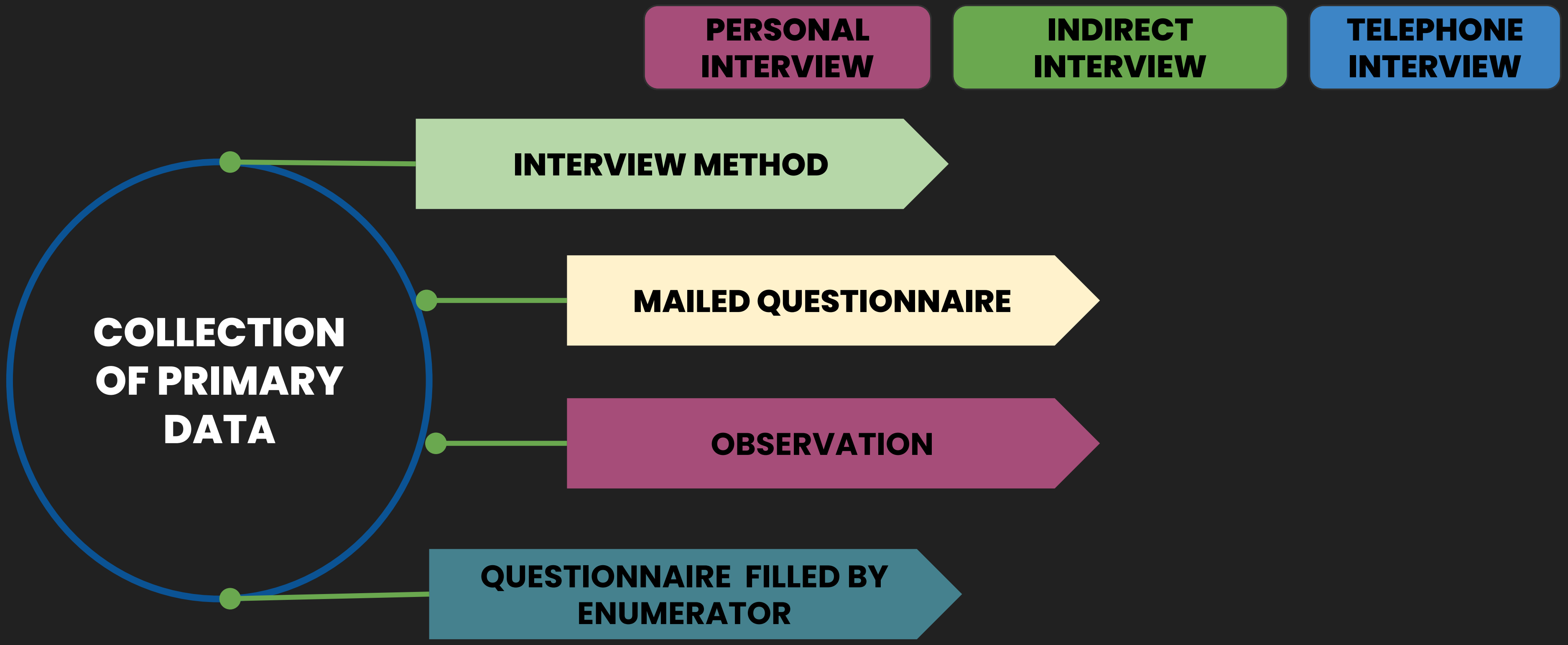
- When a variable assumes a finite or a countably infinite number of isolated values, it is known as a discrete variable.
- EXAMPLE : Number of petals in a flower, the number of road accidents in locality

CONTINUOUS VARIABLE

- When a variable assumes any value from a given interval.
- EXAMPLE : height, weight

ATTRIBUTE

- A qualitative characteristic is known as an attribute.
- The gender of a baby, the nationality of a person, the colour of a flower etc. are examples of attributes.


PERSONAL INTERVIEW
INDIRECT INTERVIEW
TELEPHONE INTERVIEW
INTERVIEW METHOD
COLLECTION OF PRIMARY DATA
MAILED QUESTIONNAIRE
OBSERVATION
QUESTIONNAIRE FILLED BY ENUMERATOR

INTERVIEW METHOD

PERSONAL INTERVIEW METHOD

The investigator meets the respondents directly and collects the required information .

Highly accurate

EXAMPLE : *natural calamity like a super cyclone or an earthquake or an epidemic like plague,*

INDIRECT INTERVIEW METHOD

- When reaching respondent is difficult, data is collected by contacting associated persons .
- Highly accurate , low coverage
- **EXAMPLE :** *rail accident*

TELEPHONE INTERVIEW METHOD

Data is collected over phone

Quick and non -expensive method

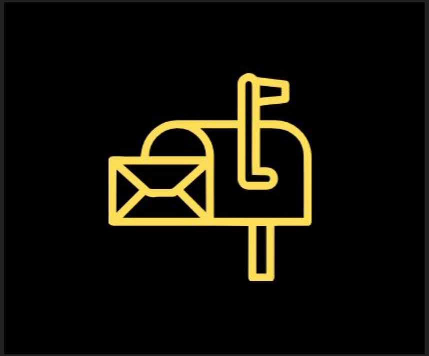
Non-responses is maximum

Low accuracy

High coverage

MAILED QUESTIONNAIRE METHOD

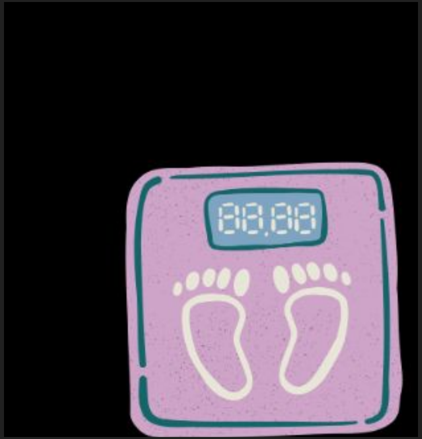
- In this method well-drafted and soundly-sequenced questionnaire covering all the important aspects of the data requirement is sent to the respondents for filling .



- Coverage is wide but amount of non responses will be maximum

OBSERVATION METHOD

- In this method data is collected by direct observation or using instrument .
- **EXAMPLE :** *data on the height and weight of a group of students.*
- more accurate
- time consuming,
- laborious
- covers only a small area.



QUESTIONNAIRE FILLED AND SENT BY ENUMERATORS

- Enumerator means a Person who directly interacts with respondent and fills the questionnaire.
- It is generally used in case of surveys and census.

SOURCES OF SECONDARY DATA

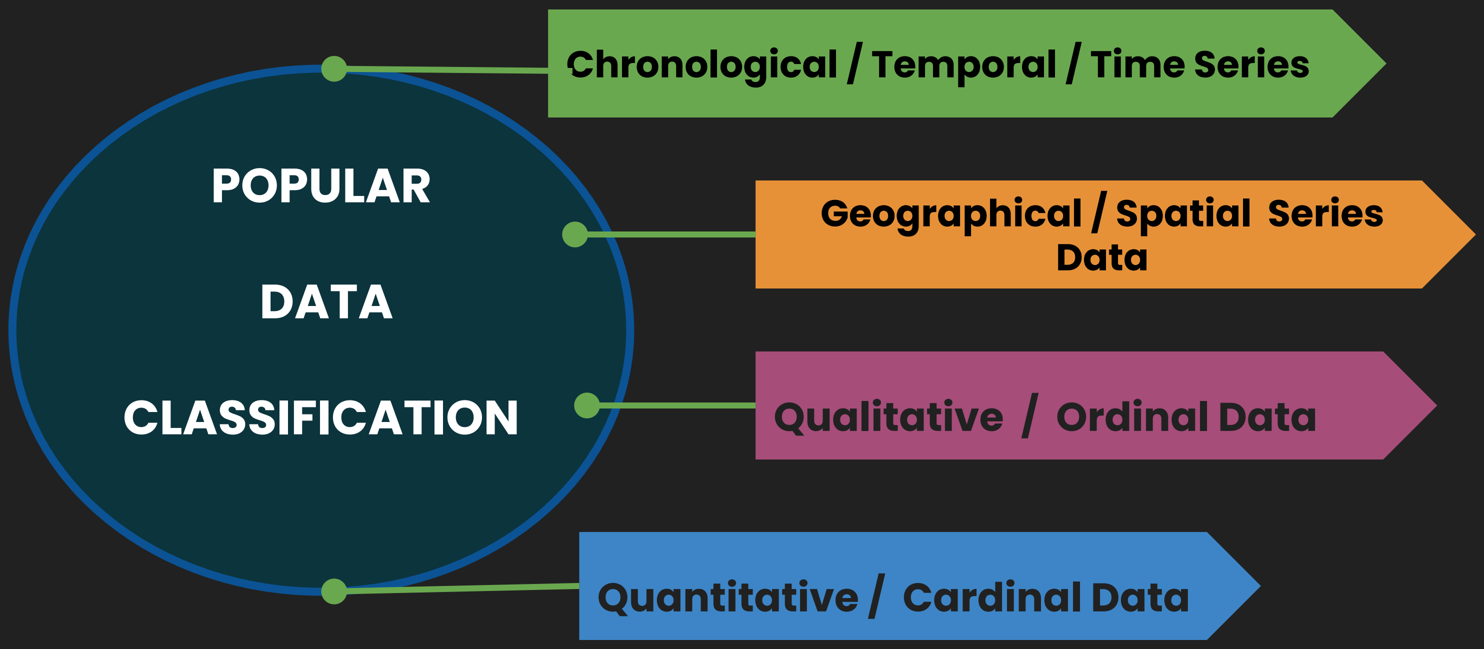
- International sources : WHO, ILO, IMF, World Bank etc.
- Government sources : Statistical Abstract by CSO,
- Private and quasi-government sources : ISI, ICAR, NCERT etc.
- Unpublished sources of various research institutes, researchers etc.

SCRUTINY OF DATA

- Checking accuracy and consistency of data
- No hard and fast rules can be recommended for the scrutiny of data. One must apply his intelligence, patience and experience while scrutinising the given information.

INTERNAL CONSISTENCY

When two or more series of related data are given , we should check consistency among them


POPULAR
DATA
CLASSIFICATION
Chronological / Temporal / Time Series
Geographical / Spatial Series Data
Qualitative / Ordinal Data
Quantitative / Cardinal Data

DATA CLASSIFICATION

Chronological / Temporal / Time Series

data are classified in respect of successive time points

Population of India (in crores)

Year	Population (Crores)
1951	35.7
1961	43.8
1971	54.6
1981	68.4
1991	81.8
2001	102.7
2011	121.0

Geographical / Spatial Series Data

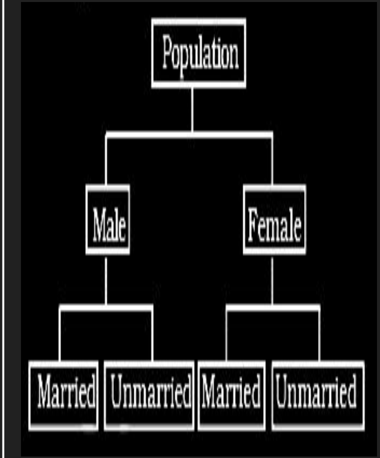
Data arranged region wise

Yield of Wheat for Different Countries (2013)

Country	Yield of wheat (kg/hectare)
Canada	3594
China	5055
France	7254
Germany	7998
India	3154
Pakistan	2787

QUALITATIVE / ORDINAL DATA

Data classified in respect of an attribute

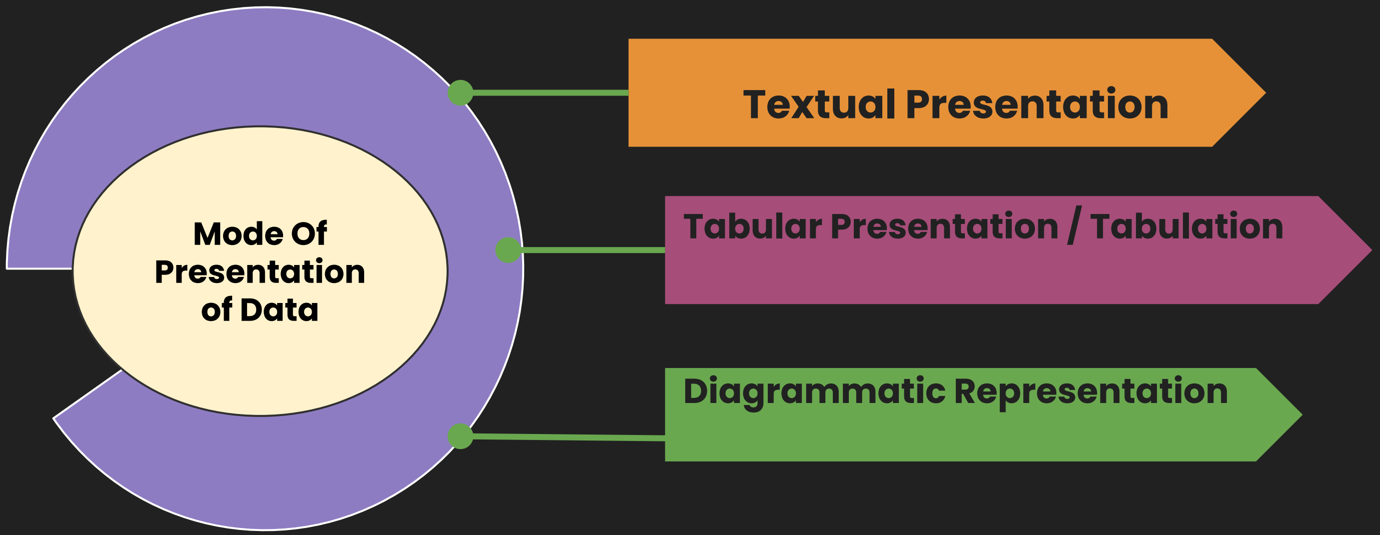


Quantitative / Cardinal Data

data are classified in respect of a variable

Frequency Distribution of Marks in Mathematics of 100 Students

Marks	Frequency
0–10	1
10–20	8
20–30	6
30–40	7
40–50	21
50–60	23
60–70	19
70–80	6
80–90	5
90–100	4
Total	*100*


Mode Of Presentation of Data
Textual Presentation
Tabular Presentation / Tabulation
Diagrammatic Representation

TEXTUAL PRESENTATION

- This method comprises presenting data with the help of a paragraph or a number of paragraphs.

- EXAMPLE
- 'In 2009, out of a total of five thousand workers of Roy Enamel Factory, four thousand and two hundred were members of a Trade Union. The number of female workers was twenty per cent of the total workers out of which thirty per cent were members of the Trade Union.

TEXTUAL PRESENTATION

MERITS

- Even a layman can present data by this method
- The observations with exact magnitude can be presented with the help of textual presentation.

DEMERITS

- It is dull, monotonous and comparison between different observations is not possible in this method.

TABULAR PRESENTATION / TABULATION

Tabulation may be defined as systematic presentation of data with the help of a statistical table .

MERITS

- It facilitates comparison between rows and columns.
- Complicated data can also be represented using tabulation.
- It is a must for diagrammatic representation.
- Without tabulation, statistical analysis of data is not possible.

TABULAR PRESENTATION / TABULATION

BOX HEAD : entire upper part of the table which includes columns and sub-column numbers, unit(s) of measurement along with caption.

CAPTION : the upper part of the table, describing the columns and sub-columns,

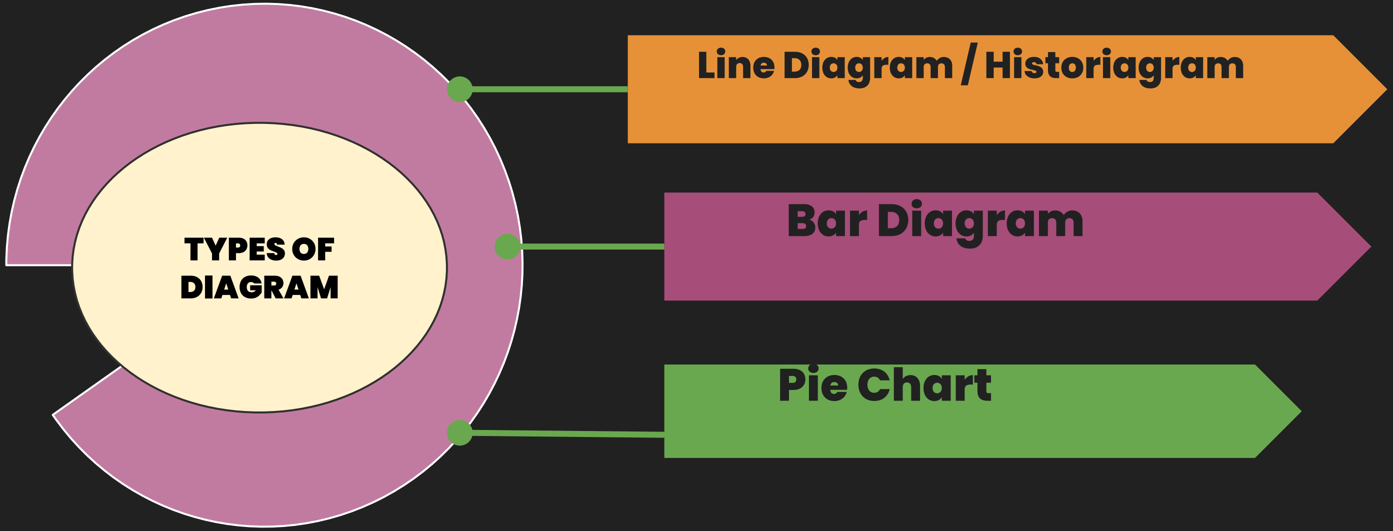
STUB : left part of the table providing the description of the rows.

BODY : main part of the table that contains the numerical figures.

FOOTNOTE : source of the data at the bottom of table

DIAGRAMMATIC REPRESENTATION OF DATA

- An attractive representation of statistical data
- can be used for both the educated section and uneducated section of the society.
- Any hidden trend present in the given data can be noticed only in this mode of representation.
- Compared to tabulation, this is less accurate. So, if there is a priority for accuracy, we have to recommend tabulation.


TYPES OF DIAGRAM
Line Diagram / Historiagram
Bar Diagram
Pie Chart

Line Diagram or Historiagram

- Generally used for time series .

wide fluctuation	LOG CHART OR RATIO CHART
two or more series of same unit	MULTIPLE LINE CHART
two or more series of distinct unit	MULTIPLE AXIS CHART

BAR DIAGRAM

Bars i.e. rectangles of equal width and usually of varying lengths drawn either horizontally or vertically.

- **Horizontal Bar diagram :** Qualitative data or Data varying over space (Geography)
- **Vertical Bar diagram :** Quantitative data or Time series data
- **Multiple or grouped bar diagram :** to compare related series
- **Component Bar diagram :** representing data divided into a number of components.
- **Percentage Bar diagram :** For relative comparison to whole

Pie chart

It is used for circular presentation of relative data

$$\text{Segment angle} = \frac{(\text{segment value} \times 360^0)}{(\text{total value})}$$

FREQUENCY And DISTRIBUTION

FREQUENCY : Number of times a particular observation is repeated.

FREQUENCY DISTRIBUTION TABLE : It is a table which contains observation or class intervals in one column and corresponding frequency in the other

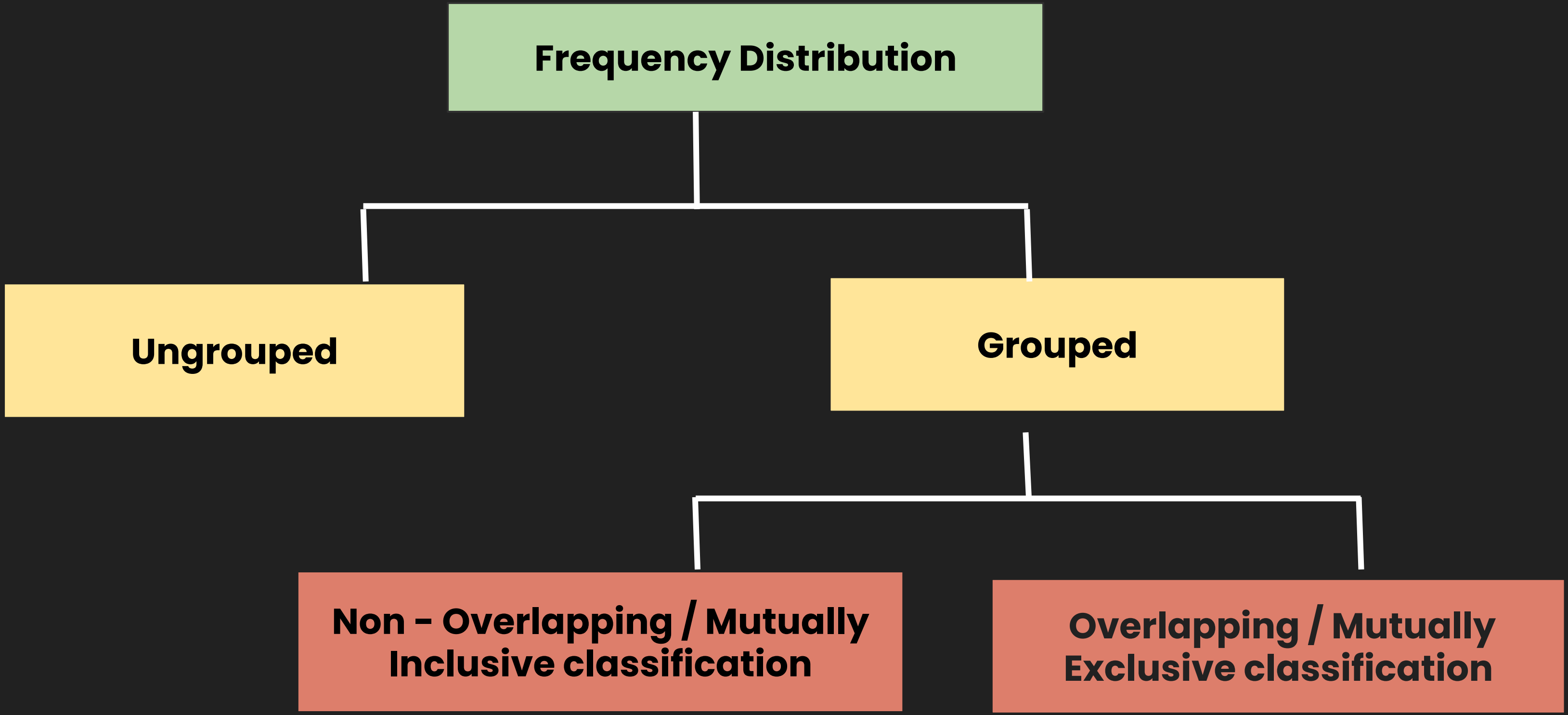
TYPES OF FREQUENCY DISTRIBUTION

Ungrouped / Simple Frequency Distribution

When there are limited number of distinct observations , frequency can be assigned to each one of them

Grouped Frequency distribution

When there are large number of observations , grouping is done among them , each group is called class interval and frequency is assigned to group and not individual value .


Frequency Distribution
Ungrouped
Grouped
Non - Overlapping / Mutually Inclusive classification
Overlapping / Mutually Exclusive classification

Class Limit

- For a class interval, the class limits may be defined as the minimum value and the maximum value the class interval
- Minimum value = lower class limit (LCL)
- Maximum value = upper class limit (UCL).

CLASS INTERVAL	FREQUENCY	LCL	UCL
44-48	3	44	48
49 - 53	4	49	53
54 -58	5	54	58

Non - Overlapping / Mutually Inclusive classification

CLASS INTERVAL	LCL	UCL
44-48	44	48
49 - 53	49	53
54 -58	54	58

Includes UCL

Overlapping / Mutually Exclusive classification

CLASS INTERVAL	LCL	UCL
40 - 50	40	50
50 - 60	50	60
60 - 70	60	70

Excludes UCL

CLASS BOUNDARY

OVERLAPPING / MUTUALLY EXCLUSIVE

CLASS INTERVAL	LCL	UCL	LCB	UCB
40 - 50	40	50	40	50
50 -60	50	60	50	60
60-70	60	70	60	70

Class limit = Class boundary

CLASS BOUNDARY

OVERLAPPING / MUTUALLY EXCLUSIVE

CLASS INTERVAL	LCL	UCL	LCB	UCB
40 - 50	40	50	40	50
50 -60	50	60	50	60
60-70	60	70	60	70

Class limit = Class boundary

CLASS BOUNDARY

NON-OVERLAPPING / MUTUALLY INCLUSIVE

CLASS INTERVAL	LCL	UCL	LCB	UCB
44-48	44	48	43.5	48.5
49 - 53	49	53	48.5	53.5
54 -58	54	58	53.5	58.5

LCB = LCL - 0.5
UCB = UCL + 0.5

CLASS LENGTH

Class length = UCB - LCB

CLASS INTERVAL	LCL	UCL	LCB	UCB	CLASS LENGTH
44-48	44	48	43.5	48.5	5
49 - 53	49	53	48.5	53.5	5
54 -58	54	58	53.5	58.5	5

Mid-Point or Mid-Value or Class Mark

$$\text{mid-point} = \frac{LCL + UCL}{2}$$

$$= \frac{LCB + UCB}{2}$$

CLASS INTERVAL	LCL	UCL	LCB	UCB	MID POINT
44-48	44	48	43.5	48.5	46
49 - 53	49	53	48.5	53.5	51
54 -58	54	58	53.5	58.5	56

CUMULATIVE FREQUENCY

- These are of two types -

  Less than type cumulative frequency

  More than type cumulative frequency

- At any class interval

  Less than type CF + More than type CF = Total frequency

Class Interval	Frequency	CB	Less than type CF	More than type CF
44 - 48	3	43.5	**0**	**36**
49- 53	4	48.5	**3**	**33**
54- 58	5	53.5	**7**	**29**
59 -63	7	58.5	**12**	**24**
64 - 68	9	63.5	**19**	**17**
69 -73	8	68.5	**28**	**8**
		73.5	**36**	**0**

36

Frequency Density

$$\text{Frequency Density} = \frac{\text{Class Frequency}}{\text{Class Length of Class}}$$

Relative Frequency

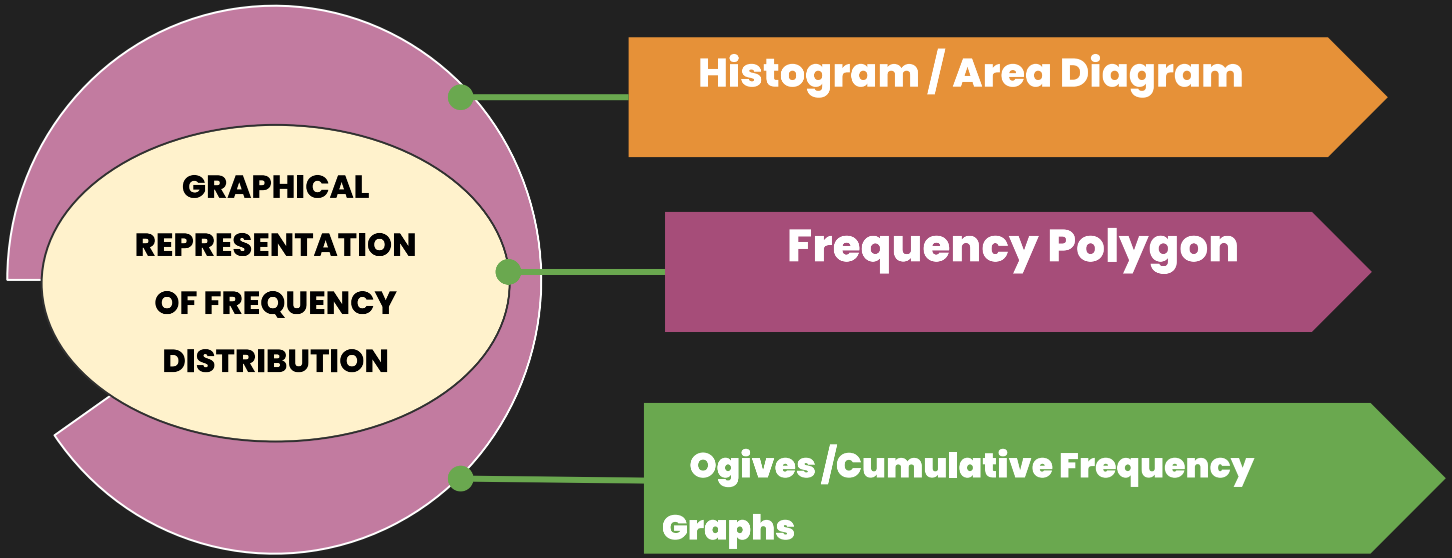
$$\text{Relative frequency} = \frac{\text{Class Frequency}}{\text{Total Frequency}}$$

Relative frequencies add up to unity

Percentage Frequency

$$\text{Percentage Frequency} = \frac{\text{Class Frequency}}{\text{Total Frequency}} \times 100$$

percentage frequencies add up to one hundred.


GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTION
Histogram / Area Diagram
Frequency Polygon
Ogives /Cumulative Frequency Graphs

GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTION

Histogram / Area Diagram

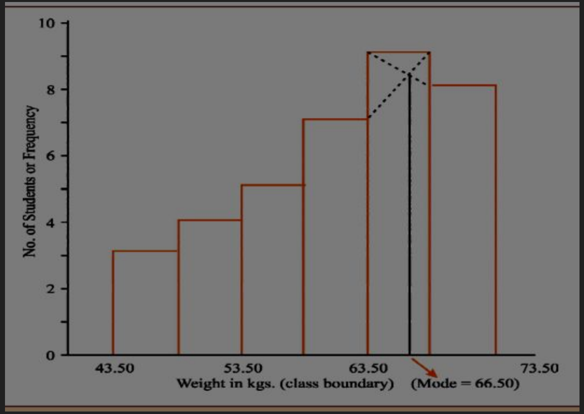
- This is a very convenient way to represent a frequency distribution.
- Comparison between frequency of two different classes are possible
- It is used to calculate **MODE.**

Frequency Polygon

- Usually frequency polygon is meant for simple / Ungrouped frequency distribution.
- However, we also apply it for grouped frequency distribution provided the width of the class intervals remains the same.

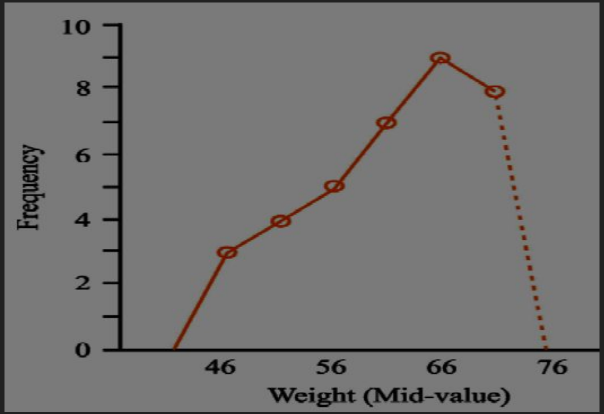
Ogives /Cumulative Frequency Graphs

- Made by both type of cumulative frequency and called as less than ogive , more than ogive
- It is used to calculate **quartiles , median**

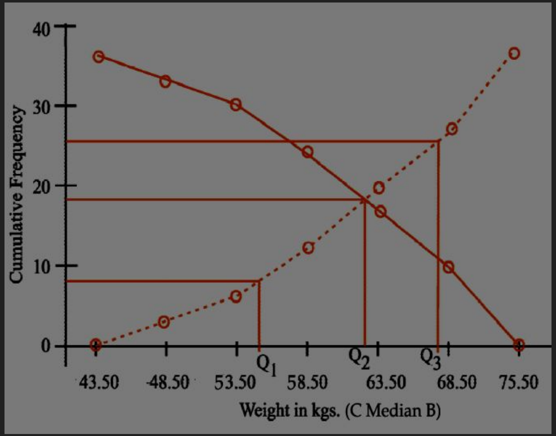


Histogram

Mode



Frequency Polygon



Ogives

Quartiles , Medians

FREQUENCY CURVE

- It is a limiting form of a histogram or frequency polygon.
- The frequency curve for a distribution can be obtained by drawing a smooth and free hand curve through the mid-points of the upper sides of the rectangles forming the histogram.

FREQUENCY CURVE

4 TYPES

BELL SHAPED CURVE

- commonly used distributions
- The distribution of height, weight, mark, profit etc. usually belong to this category.

J - SHAPED CURVE

U - SHAPED CURVE

MIXED CURVE

Que. The frequency of class 20 - 30 in the following data is

Class	0-10	10-20	20-30	30-40	40-50
Cumulative Frequency	5	13	28	34	38

(a) 5

(b) 28

(c) 15

(d) 13

C

Que. The data given below refers to the marks gained by a group of students:

Marks	Below 10	Below 20	Below 30	Below 40	Below 50
No.of Students	15	38	65	84	100

The no. of students getting marks more than 30 would be?

(a) 50

(b) 53

(c) 35

(d) 62

C

Que. Cost of Sugar in a month under the heads raw materials, labour, direct production and others were 12, 20, 35 & 23 units respectively. The difference between their central angles for the largest & smallest components of the cost of Sugar is

(a) 92

(b) 72

(c) 48

(d) 56

a

Unit 1 Exercise
Set A

Que 1. Which of the following statements is false?

(a) Statistics is derived from the Latin word 'Status'

(b) Statistics is derived from the Italian word 'Statista'

(c) Statistics is derived from the French word 'Statistik'

(d) None of these.

C

Que. 2 Statistics is defined in terms of numerical data in the

(a) Singular sense

(b) Plural sense

(c) Either (a) or (b)

(d) Both (a) and (b).

b

Que 3. Statistics is applied in

(a) Economics

(b) Business management

(c) Commerce and industry

(d) All these.

d

Que 4. Statistics is concerned with

(a) Qualitative information

(b) Quantitative information

(c) (a) or (b)

(d) Both (a) and (b).

d

Que 5. An attribute is

(a) A qualitative characteristic

(b) A quantitative characteristic

(c) A measurable characteristic

(d) All these.

a

Que 6. Annual income of a person is

(a) An attribute

(b) A discrete variable

(c) A continuous variable

(d) (b) or (c).

b

Que 7. Marks of a student is an example of

(a) An attribute

(b) A discrete variable

(c) A continuous variable

(d) None of these.

b

Que. 8 Nationality of a student is

(a) An attribute

(b) A continuous variable

(c) A discrete variable

(d) (a) or (c).

a

Que 9 Drinking habit of a person is

(a) An attribute

(b) A variable

(c) A discrete variable

(d) A continuous variable.

a

Que 10. Age of a person is

(a) An attribute

(b) A discrete variable

(c) A continuous variable

(d) A variable.

C

Que 11. Data collected on religion from the census reports are

(a) Primary data

(b) Secondary data

(c) Sample data

(d) (a) or (b).

b

Que.12 The data collected on the height of a group of students after recording their heights with a measuring tape are

(a) Primary data

(b) Secondary data

(c) Discrete data

(d) Continuous data.

a

Que 13. The primary data are collected by

(a) Interview method

(b) Observation method

(c) Questionnaire method

(d) All these.

d

Que 14. The quickest method to collect primary data is

(a) Personal interview

(b) Indirect interview

(c) Telephone interview

(d) By observation.

C

Que 15. The best method to collect data, in case of a natural calamity, is

(a) Personal interview

(b) Indirect interview

(c) Questionnaire method

(d) Direct observation method.

a

Que 16. In case of a rail accident, the appropriate method of data collection is by

(a) Personal interview

(b) Direct interview

(c) Indirect interview

(d) All these.

C

Que 17. Which method of data collection covers the widest area?

(a) Telephone interview method

(b) Mailed questionnaire method

(c) Direct interview method

(d) All these.

b

Que 18. The amount of non-responses is maximum in

(a) Mailed questionnaire method

(b) Interview method

(c) Observation method

(d) All these.

a

Que 19. Some important sources of secondary data are

(a) International and Government sources

(b) International and primary sources

(c) Private and primary sources

(d) Government sources.

a

Que 20. Internal consistency of the collected data can be checked when

(a) Internal data are given

(b) External data are given

(c) Two or more series are given

(d) A number of related series are given.

d

Que 21. The accuracy and consistency of data can be verified by

(a) Internal checking

(b) External checking

(c) Scrutiny

(d) Both (a) and (b).

c

Que22. The mode of presentation of data are

(a) Textual, tabulation and diagrammatic

(b) Tabular, internal and external

(c) Textual, tabular and internal

(d) Tabular, textual and external.

a

Que23. The best method of presentation of data is

(a) Textual

(b) Tabular

(c) Diagrammatic

(d) (b) and (c).

b

Que24. The most attractive method of data presentation is

(a) Tabular

(b) Textual

(c) Diagrammatic

(d) (a) or (b).

C

Que 25. For tabulation, 'caption' is

(a) The upper part of the table

(b) The lower part of the table

(c) The main part of the table

(d) The upper part of a table that describes the column and sub-column.

d

Que 26. ‘Stub’ of a table is the

(a) Left part of the table describing the columns

(b) Right part of the table describing the columns

(c) Right part of the table describing the rows

(d) Left part of the table describing the rows.

d

Que 27. The entire upper part of a table is known as

(a) Caption

(b) Stub

(c) Box head

(d) Body.

C

Que28. The unit of measurement in tabulation is shown in

(a) Box head

(b) Body

(c) Caption

(d) Stub.

a

Que 29. In tabulation source of the data, if any, is shown in the

(a) Footnote

(b) Body

(c) Stub

(d) Caption.

a

Que 30. Which of the following statements is untrue for tabulation?

(a) Statistical analysis of data requires tabulation

(b) It facilitates comparison between rows and not columns

(c) Complicated data can be presented

(d) Diagrammatic representation of data requires tabulation.

b

Que 31. Hidden trend, if any, in the data can be noticed in

(a) Textual presentation

(b) Tabulation

(c) Diagrammatic representation

(d) All these.

C

Que. 32 Diagrammatic representation of data is done by

(a) Diagrams

(b) Charts

(c) Pictures

(d) All these.

d

Que33. The most accurate mode of data presentation is

(a) Diagrammatic method

(b) Tabulation

(c) Textual presentation

(d) None of these.

b

Que 34. The chart that uses logarithm of the variable is known as

(a) Line chart

(b) Ratio chart

(c) Multiple line chart

(d) Component line chart.

b

Que 35. Multiple line chart is applied for

(a) Showing multiple charts

(b) Two or more related time series when the variables are expressed in the same unit

(c) Two or more related time series when the variables are expressed in different unit

(d) Multiple variations in the time series.

b

Que 36. Multiple axis line chart is considered when

(a) There is more than one time series

(b) The units of the variables are different

(c) (a) or (b)

(d) (a) and (b).

d

Que 37. Horizontal bar diagram is used for

(a) Qualitative data

(b) Data varying over time

(c) Data varying over space

(d) (a) or (c).

d

Que 38. Vertical bar diagram is applicable when

(a) The data are qualitative

(b) The data are quantitative

(c) When the data vary over time

(d) (b) or (c).

d

Que 39. Divided bar chart is considered for

(a) Comparing different components of a variable

(b) The relation of different components to the table

(c) (a) or (b)

(d) (a) and (b).

d

Que 40. In order to compare two or more related series, we consider

(a) Multiple bar chart

(b) Grouped bar chart

(c) (a) or (b)

(d) (a) and (b).

C

Que 41. Pie-diagram is used for

(a) Comparing different components and their relation to the total

(b) Representing qualitative data in a circle

(c) Representing quantitative data in circle

(d) (b) or (c).

a

Que42. A frequency distribution

(a) Arranges observations in an increasing order

(b) Arranges observation in terms of a number of groups

(c) Relates to a measurable characteristic

(d) All these.

d

Que 43. The frequency distribution of a continuous variable is known as

(a) Grouped frequency distribution

(b) Simple frequency distribution

(c) (a) or (b)

(d) (a) and (b).

a

Que 44. The distribution of shares is an example of the frequency distribution of

(a) A discrete variable

(b) A continuous variable

(c) An attribute

(d) (a) or (c).

a

Que 45. The distribution of profits of a blue-chip company relates to

(a) Discrete variable

(b) Continuous variable

(c) Attributes

(d) (a) or (b).

b

Que 46. Mutually exclusive classification

(a) Excludes both the class limits

(b) Excludes the upper class limit but includes the lower class limit

(c) Includes the upper class limit but excludes the upper class limit

(d) Either (b) or (c).

b

Que 47. **Mutually inclusive classification is usually meant for**

(a) A discrete variable

(b) A continuous variable

(c) An attribute

(d) All these.

a

Que 48. **Mutually exclusive classification is usually meant for**

(a) A discrete variable

(b) A continuous variable

(c) An attribute

(d) Any of these.

b

Que 49. The LCB is

(a) An upper limit to LCL

(b) A lower limit to LCL

(c) (a) and (b)

(d) (a) or (b).

b

Que 50. The UCB is

(a) An upper limit to UCL

(b) A lower limit to LCL

(c) Both (a) and (b)

(d) (a) or (b).

a

Que 51. length of a class is

(a) The difference between the UCB and LCB of that class

(b) The difference between the UCL and LCL of that class

(c) (a) or (b)

(d) Both (a) and (b).

a

Que 52. For a particular class boundary, the less than cumulative frequency and more than cumulative frequency add up to

(a) Total frequency

(b) Fifty per cent of the total frequency

(c) (a) or (b)

(d) None of these.

a

Que 53. Frequency density corresponding to a class interval is the ratio of

(a) Class frequency to the total frequency

(b) Class frequency to the class length

(c) Class length to the class frequency

(d) Class frequency to the cumulative frequency.

b

Que 54. Relative frequency for a particular class

(a) Lies between 0 and 1

(b) Lies between 0 and 1, both inclusive

(c) Lies between –1 and 0

(d) Lies between –1 to 1.

a

Que 55. Mode of a distribution can be obtained from

(a) Histogram

(b) Less than type ogives

(c) More than type ogives

(d) Frequency polygon.

a

Que 56. Median of a distribution can be obtained from

(a) Frequency polygon

(b) Histogram

(c) Less than type ogives

(d) None of these.

C

Que 57. A comparison among the class frequencies is possible only in

(a) Frequency polygon

(b) Histogram

(c) Ogives

(d) (a) or (b).

b

Que 58. Frequency curve is a limiting form of

(a) Frequency polygon

(b) Histogram

(c) (a) or (b)

(d) (a) and (b).

C

Que 59. Most of the commonly used frequency curves are

(a) Mixed

(b) Inverted J-shaped

(c) U-shaped

(d) Bell-shaped.

d

Que 60. The distribution of profits of a company follows

(a) J-shaped frequency curve

(b) U-shaped frequency curve

(c) Bell-shaped frequency curve

(d) Any of these.

C

POPULATION / UNIVERSE

- All items ,elements , or observations of interest having similar properties are known as population .
- It may be defined as the aggregate of all the units under consideration .
- Example : Population of students enrolled for CA Course
- The number of units belonging to a population is known as population size(N).

POPULATION

- If a population comprises only a finite number of units, then it is known as a **finite population.**
- **EXAMPLE :** Population of students enrolled for CA Course

- If the population contains an infinite or uncountable number of units, then it is known as an **infinite population.**
- **EXAMPLE :** population of stars, the population of mosquitoes

- A population consisting of real objects is known as an **existent population.**

- A population that exists just hypothetically like the population of heads when a coin is tossed infinitely is known as a hypothetical or an **imaginary population**

Census

- Study of every elements of population is called census .

SAMPLE

- A sample may be defined as a part of a population so selected with a view to representing the population in all its characteristics .

- If a sample contains n units, then n is known as sample size.

- The units forming the sample are known as “Sampling Units”.

- A detailed and complete list of all the sampling units is known as a “Sampling Frame”.

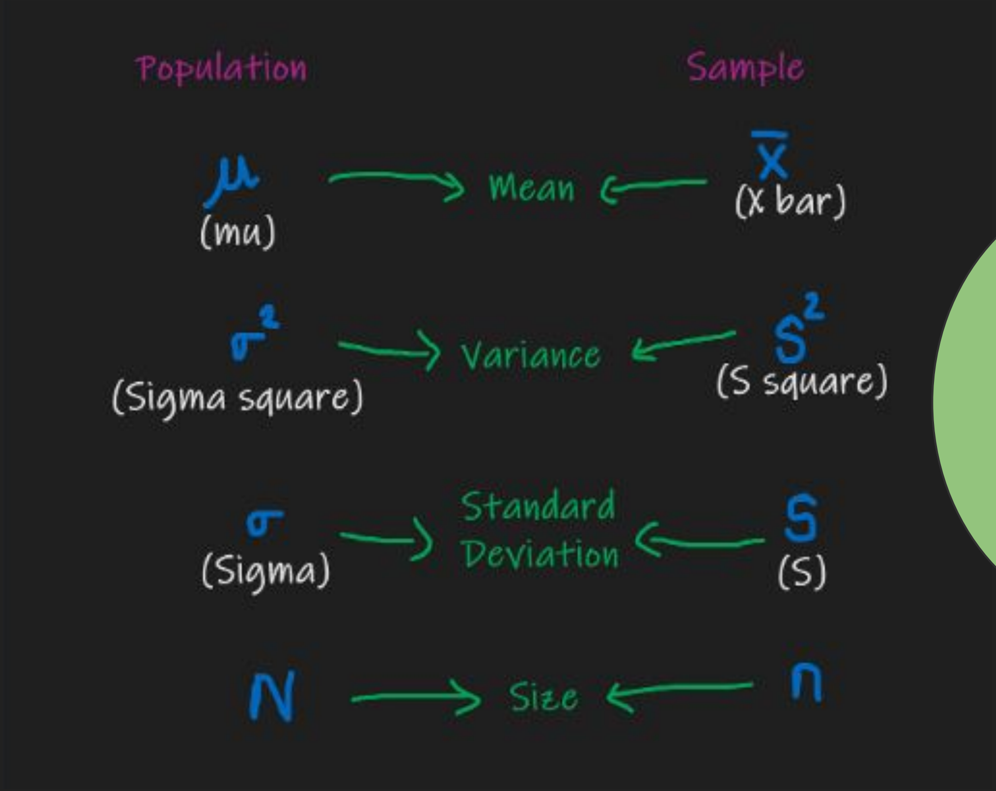
There are different statistical measures in statistics such as mean , median , mode , standard deviation , variance , proportion etc . These can be computed for both population and sample .

PARAMETER

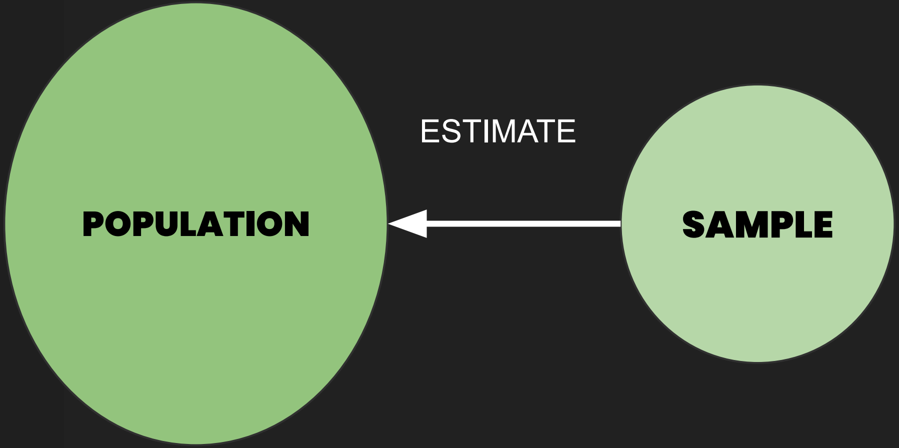
- It is the statistical measures computed from population.
- A parameter may be defined as a characteristic of a population based on all the units of the population

STATISTIC

- It is the statistical measures computed from Sample .
- A statistic may be defined as a statistical measure of sample observation and as such it is a function of sample observations

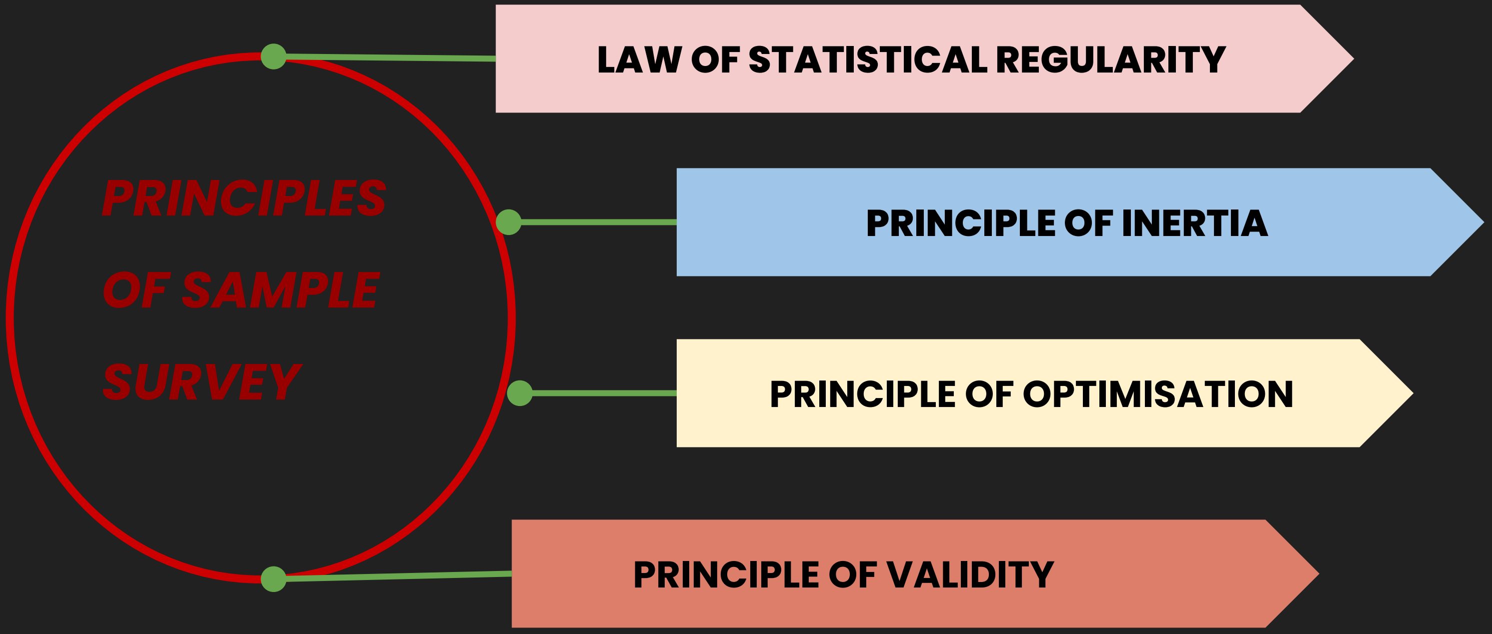


$P \longrightarrow$ *Proportion* $\longleftarrow \hat{P}$



- *A statistic is used to estimate a particular population parameter*

- *Sample Survey is the study of the unknown population on the basis of a proper representative sample drawn from it .*



PRINCIPLES OF SAMPLE SURVEY

LAW OF STATISTICAL REGULARITY :

If a sample of fairly large size is drawn from the population under discussion at random, then on an average the sample would possess the characteristics of that population.

PRINCIPLE OF INERTIA

As sample size increases , the results are likely to be more reliable , accurate and precise , provided other factors are kept constant

PRINCIPLE OF OPTIMISATION :

The principle of optimization ensures that an optimum level of efficiency at a minimum cost or the maximum efficiency at a given level of cost can be achieved with the selection of an appropriate sampling design.

PRINCIPLE OF VALIDITY :

The principle of validity states that a sampling design is valid only if it is possible to obtain valid estimates and valid tests about population parameters.

Only a probability sampling ensures this validity.

Deviation between the value of population parameter as obtained from a sample and its observed value.

TYPES OF ERROR

SAMPLING ERROR

Since only a part of population is investigated in sampling , every sampling design is subjected to this type of errors

Factors : defective sampling design, faulty demarcation,wrong choice of statistics,Variability in the population

NON SAMPLING ERROR

Errors due to recording observations, biases on the part of the enumerators, wrong and faulty interpretation of data is termed as non-sampling errors.

This type of errors happen both in sampling and complete enumeration

Factors : Ignorance , Communication gap ,Non response bias , Incomplete coverage

TYPES OF SAMPLING

- PROBABILITY SAMPLING
- MIXED SAMPLING
- NON – PROBABILITY SAMPLING

PROBABILITY SAMPLING

- In the Probability sampling there is always a fixed, pre assigned probability for each member of the population to be a part of the sample taken from that population

- Some important probability sampling are :

  simple random sampling ,

  stratified sampling,

  Multi Stage sampling, Multi Phase Sampling, Cluster Sampling and so on.

SIMPLE RANDOM SAMPLING

- When the units are selected independent of each other in such a way that each unit belonging to the population has an equal chance of being a part of the sample, the sampling is known as Simple random sampling or just random sampling.

SIMPLE RANDOM SAMPLING IS EFFECTIVE IF

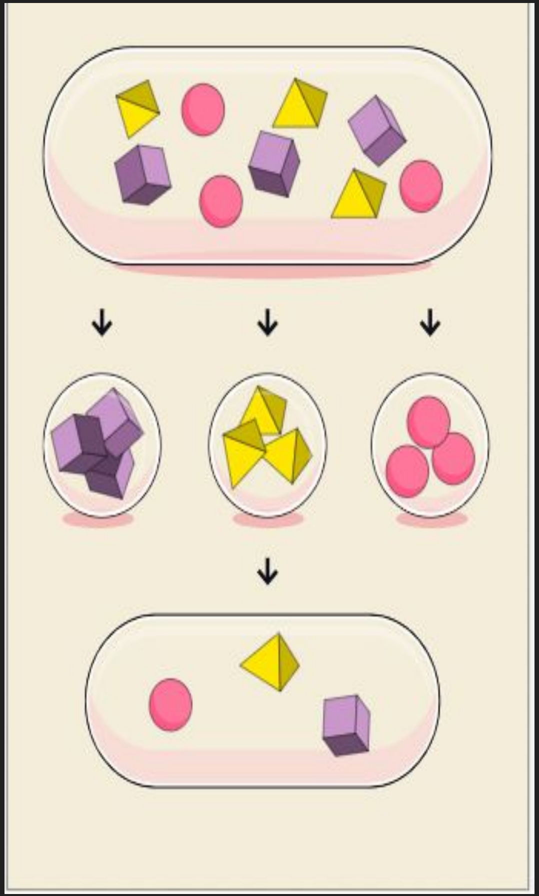
- the population is not very large
- the sample size is not very small
- the population under consideration is not heterogeneous

STRATIFIED SAMPLING

- In this method , the universe or the entire population is divided into a number of groups or strata and then certain number of items are taken from each group at random .

- Its basic purpose is to ensure that all the characteristics of a heterogeneous population are adequately represented in the sample .

- It helps in reduction of variability and thereby an increase in precision.



- There are two types of allocation of sample size.

- “Proportional allocation” or “Bowely’s allocation
- When there is not much variation between the strata variances
- sample sizes for different strata are taken as proportional to the population sizes.

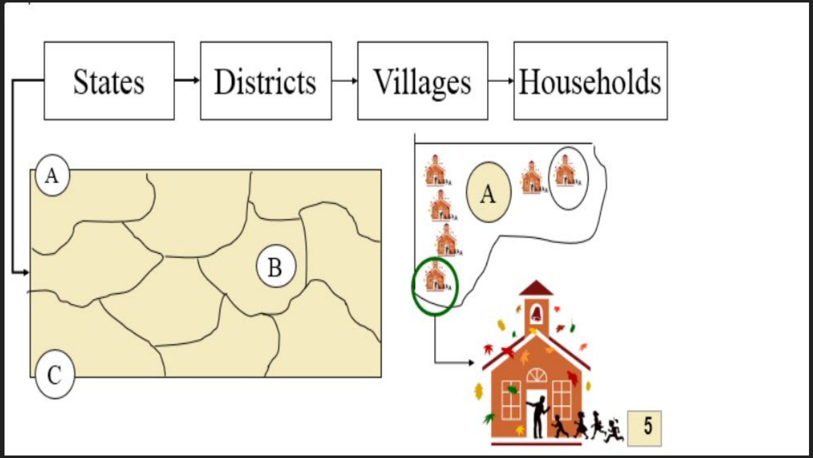
- “Neyman’s allocation”
- When the strata-variances differ significantly among themselves
- sample size vary jointly with population size and population standard deviation

STRATIFIED SAMPLING

- The purpose of stratified sampling are
- (i) to make representation of all the sub populations
- (ii) to provide an estimate of parameter not only for all the strata but also and overall estimate
- (iii) reduction of variability and thereby an increase in precision.
- Stratified sampling not advisable if
- (i) population is not large
- (ii) some prior information is not available
- (iii) there is not much heterogeneity among the units of population

MULTISTAGE SAMPLING

- **In this type of complicated sampling, the population is supposed to compose of first stage sampling units, each of which in its turn is supposed to compose of second stage sampling units, each of which again in its turn is supposed to compose of third stage sampling units and so on till we reach the ultimate sampling unit.**



Suppose we want to take a sample of 5000 households from India

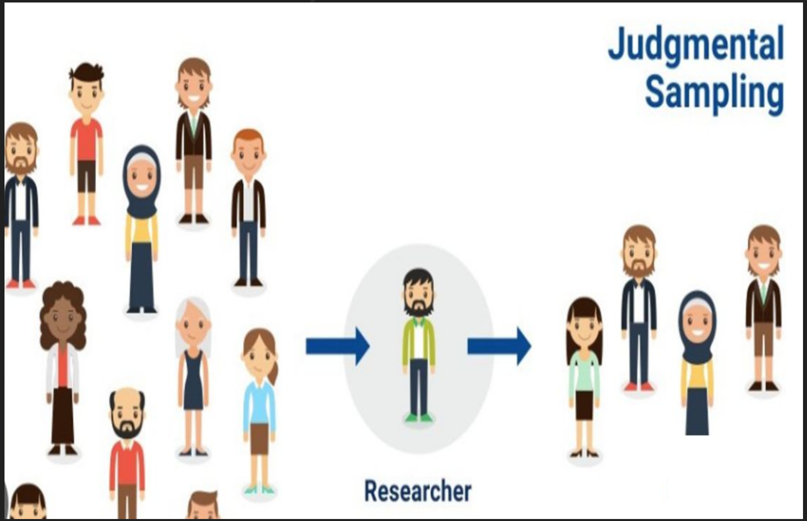
MULTISTAGE SAMPLING

- The coverage in case of multistage sampling is quite large.
- It also saves computational labour and is cost-effective.
- It adds flexibility into the sampling process which is lacking in other sampling schemes.
- However, compared to stratified sampling, multistage sampling is likely to be less accurate.

NON - PROBABILITY SAMPLING

- In non- probability sampling , no probability attached to the member of the population and as such it is based entirely on the judgement of the sampler.
- Non-probability sampling is also known as Purposive or Judgemental Sampling

PURPOSIVE OR JUDGEMENTAL SAMPLING



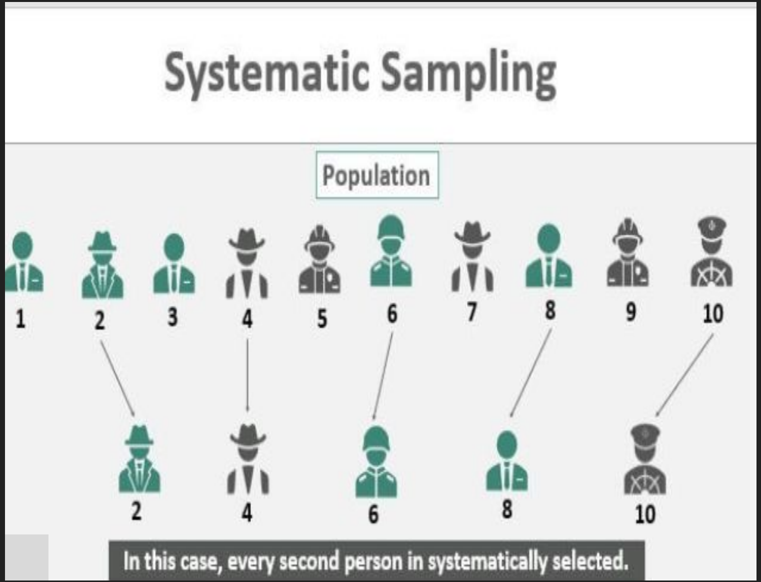
- This type of sampling is dependent solely on the discretion of the sampler and he applies his own judgement based on his belief, prejudice, whims and interest to select the sample.
- Since this type of sampling is non-probabilistic, it is purely subjective and, as such, varies from person to person.
- No statistical hypothesis can be tested on the basis of a purposive sampling

MIXED SAMPLING

- Mixed sampling is based partly on some probabilistic law and partly on some pre decided rule.
- Systematic sampling belongs to this category.

SYSTEMATIC SAMPLING

- It refers to a sampling scheme where the units constituting the sample are selected at regular interval after selecting the very first unit at random i.e., with equal probability.
- Systematic sampling is partly probability sampling in the sense that the first unit of the systematic sample is selected probabilistically and partly non- probability sampling in the sense that the remaining units of the sample are selected according to a fixed rule which is non-probabilistic in nature.



SYSTEMATIC SAMPLING

- If the population size N is a multiple of the sample size n i.e. **N = nk**, for a positive integer k which must be less than n, then the systematic sampling comprises selecting one of the first k units at random, usually by using random sampling number and thereby selecting every k^{th} unit till the complete, adequate and updated sampling frame comprising all the members of the population is exhausted. This type of systematic sampling is known as "*linear systematic sampling* ". K is known as "*sample interval*".

SYSTEMATIC SAMPLING

- **However, if N is not a multiple of n, then we may write $N = nk + p$, $p < k$ and as before, we select the first unit from 1 to k by using random sampling number and thereafter selecting every kth unit in a cyclic order till we get the sample of the required size n. This type of systematic sampling is known as "*circular systematic sampling*."**

SAMPLING FLUCTUATION

It is the variation in the value of a statistic computed from different samples .

- If we compute the value of a statistic, say mean, it is quite natural that the value of the sample mean may vary from sample to sample as the sampling units of one sample may be different from that of another sample.

SAMPLING DISTRIBUTION

It is the probability distribution of a given statistic

- The mean of the statistic, as obtained from its sampling distribution, is known as "Expectation"
- standard deviation of the statistic is known as the "Standard Error (SE)" .
- SE can be regarded as a measure of precision achieved by sampling.
- SE is inversely proportional to the square root of sample size.

SAMPLING DISTRIBUTION AND STANDARD ERROR OF STATISTIC

- Starting with a population of N units, we can draw many a sample of a fixed size n.
- In case of sampling with replacement, the total number of samples that can be drawn is N^n
- When it comes to sampling without replacement , the total number of samples that can be drawn is ${}^{N}C_{n}$

EXERCISE- Set (A)

Answer the following questions. Each question carries one mark.

Que. 1 Sampling can be described as a statistical procedure

(a) To infer about the unknown universe from a knowledge of any sample

(b) To infer about the known universe from a knowledge of a sample drawn from it

(c) To infer about the unknown universe from a knowledge of a random sample drawn from it

(d) Both (a) and (b).

C

Answer the following questions. Each question carries one mark.

Que. 2 The Law of Statistical Regularity says that

(a) Sample drawn from the population under discussion possesses the characteristics of the population

(b) A large sample drawn at random from the population would possess the characteristics of the population

(c) A large sample drawn at random from the population would possess the characteristics of the population on an average

(d) An optimum level of efficiency can be attained at a minimum cost.

C

Answer the following questions. Each question carries one mark.

Que. 3 **A sample survey is prone to**

(a) Sampling errors

(b) Non-sampling errors

(c) Either (a) or (b)

(d) Both (a) and (b)

d

Answer the following questions. Each question carries one mark.

Que. 4 The population of roses in Salt Lake City is an example of

(a) A finite population

(b) An infinite population

(c) A hypothetical population

(d) An imaginary population.

b

Answer the following questions. Each question carries one mark.

Que. 5 Statistical decision about an unknown universe is taken on the basis of

(a) Sample observations

(b) A sampling frame

(c) Sample survey

(d) Complete enumeration

a

Answer the following questions. Each question carries one mark.

Que. 6 Random sampling implies

(a) Haphazard sampling

(b) Probability sampling

(c) Systematic sampling

(d) Sampling with the same probability for each unit.

d

Answer the following questions. Each question carries one mark.

Que. 7 A parameter is a characteristic of

(a) Population

(b) Sample

(c) Both (a) and (b)

(d) (a) or (b)

a

Answer the following questions. Each question carries one mark.

Que. 8 A statistic is

(a) A function of sample observations

(b) A function of population units

(c) A characteristic of a population

(d) A part of a population.

a

Answer the following questions. Each question carries one mark.

Que. 9 Sampling Fluctuations may be described as

(a) The variation in the values of a statistic

(b) The variation in the values of a sample

(c) The differences in the values of a parameter

(d) The variation in the values of observations.

a

Answer the following questions. Each question carries one mark.

Que. 10 The sampling distribution is

(a) The distribution of sample observations

(b) The distribution of random samples

(c) The distribution of a parameter

(d) The probability distribution of a statistic.

d

Answer the following questions. Each question carries one mark.

Que. 11 Standard error can be described as

(a) The error committed in sampling

(b) The error committed in sample survey

(c) The error committed in estimating a parameter

(d) Standard deviation of a statistic.

d

Answer the following questions. Each question carries one mark.

Que. 12 A measure of precision obtained by sampling is given by

(a) Standard error

(b) Sampling fluctuation

(c) Sampling distribution

(d) Expectation.

a

Answer the following questions. Each question carries one mark.

Que. 13 As the sample size increases, standard error

(a) Increases

(b) Decreases

(c) Remains constant

(d) Decreases proportionally.

b

Answer the following questions. Each question carries one mark.

Que. 14 If from a population with 25 members, a random sample without replacement of 2 members is taken, the number of all such samples is

(a) 300

(b) 625

(c) 50

(d) 600

a

Answer the following questions. Each question carries one mark.

Que. 15 A population comprises 5 members. The number of all possible samples of size 2 that can be drawn from it with replacement is

(a) 100

(b) 15

(c) 125

(d) 25

d

Answer the following questions. Each question carries one mark.

Que. 16 Simple random sampling is very effective if

(a) The population is not very large

(b) The population is not much heterogeneous

(c) The population is partitioned into several sections.

(d) Both (a) and (b)

d

Answer the following questions. Each question carries one mark.

Que. 17 Simple random sampling is

(a) A probabilistic sampling

(b) A non- probabilistic sampling

(c) A mixed sampling

(d) Both (b) and (c).

a

Answer the following questions. Each question carries one mark.

Que. 18 According to Neyman's allocation, in stratified sampling

(a) Sample size is proportional to the population size

(b) Sample size is proportional to the sample SD

(c) Sample size is proportional to the sample variance

(d) Population size is proportional to the sample variance.

a

Answer the following questions. Each question carries one mark.

Que. 19 Which sampling provides separate estimates for population means for different segments and also an over all estimate?

(a) Multistage sampling

(b) Stratified sampling

(c) Simple random sampling

(d) Systematic sampling

b

Answer the following questions. Each question carries one mark.

Que. 20 Which sampling adds flexibility to the sampling process?

(a) Simple random sampling

(b) Multistage sampling

(c) Stratified sampling

(d) Systematic sampling

b

Answer the following questions. Each question carries one mark.

Que. 21 Which sampling is affected most if the sampling frame contains an undetected periodicity?

(a) Simple random sampling

(b) Stratified sampling

(c) Multistage sampling

(d) Systematic sampling

d

Answer the following questions. Each question carries one mark.

Que. 22 Which sampling is subjected to the discretion of the sampler?

(a) Systematic sampling

(b) Simple random sampling

(c) Purposive sampling

(d) Quota sampling.

C

Answer the following questions. Each question carries one mark.

Que. 23 If a random sample of size 2 with replacement is taken from the population containing the units 3,6 and 1, then the samples would be

(a) (3, 6),(3, 1),(6, 1)

(b) (3, 3),(6, 6),(1, 1)

(c) (3, 3),(3, 6),(3, 1),(6, 6),(6, 3),(6, 1),(1, 1),(1, 3),(1, 6)

(d) (1, 1),(1, 3),(1, 6),(6, 1),(6, 2),(6, 3),(6, 6),(1, 6),(1, 1)

C

Answer the following questions. Each question carries one mark.

Que. 24 If a random sample of size two is taken without replacement from a population containing the units a,b,c and d then the possible samples are

(a) (a, b),(a, c),(a, d)

(b) (a, b),(b, c), (c, d)

(c) (a, b), (b, a), (a, c),(c,a), (a, d), (d, a)

(d) (a, b), (a, c), (a, d), (b, c), (b, d), (c,d)

d

12 MARKS

CHAPTER 14

MEASURES OF CENTRAL TENDENCY AND DISPERSION

BY : SHIVANI SHARMA

Different Measures of Central Tendency

- **Mean**
  - Arithmetic Mean
  - Geometric Mean
  - Harmonic Mean
- **Partition Values**
  - Median
  - Quartiles
  - Deciles
  - Percentiles
- **Mode**

MEAN

Arithmetic Mean

Geometric Mean

Harmonic Mean

ARITHMETIC MEAN

Discrete Observation

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

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

Simple Frequency Distribution

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \dots + f_nx_n}{f_1 + f_2 + f_3 + \dots + f_n}$$

$$= \frac{\sum f_i x_i}{\sum f_i}$$

$$\bar{X} = \frac{\sum f_i x_i}{N}$$

where ,

$N = \Sigma f_i$.

Grouped Frequency Distribution

$$\bar{X} = \frac{\sum f_i x_i}{N}$$

where ,

x_i= mid point of class interval

$N = \Sigma f_i$.

PROPERTIES OF ARITHMETIC MEAN

If all the observations are constants, say k, then the AM is also constant , k.

the algebraic sum of deviations of a set of observations from their AM is zero

i.e. for unclassified data , $\sum (x_i - \bar{x}) = 0$ [redacted]

- **AM is affected both due to change of origin and scale.**

  If y = a + b x then $\bar{y} = a + b\bar{x}$.

PROPERTIES OF ARITHMETIC MEAN

If there are two groups containing n_1 and n_2 observations and $\overline{x}_1$ and $\overline{x}_2$ as the respective arithmetic means, then the combined AM is given by

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

BEST measure of Central Tendency

Amenable to Mathematical Property

ARITHMETIC MEAN

Based on ALL OBSERVATION

Cannot be used in case of open end classifications

Affected by Sampling Fluctuations

GEOMETRIC MEAN

Average Rates , percentages

- For a given set of n positive observations, the geometric mean is defined as the n-th root of the product of the observations.

Discrete Observation

$$G = (x_1 \times x_2 \times x_3 \ldots\ldots\ldots\ldots \times x_n)^{1/n}$$

Frequency Distribution

$$G = (x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \ldots\ldots\ldots\ldots\ldots\ldots \times x_n^{f_n})^{1/N}$$

PROPERTIES OF GEOMETRIC MEAN

if all the observations assumed by a variable are constants, say $K > 0$, then the GM of the observations is also K.

GM of the product of two variables is the product of their GM's i.e. if $z = xy$, then

GM of z = (GM of x) × (GM of y) [redacted]

[redacted] GM of the ratio of two variables is the ratio of the GM's of the two variables i.e. if $z = x/y$ then

$$\text{GM of } z = \frac{\text{GM of } x}{\text{GM of } y}$$ [redacted]

PROPERTIES OF GEOMETRIC MEAN

Logarithm of G for a set of observations is the AM of the logarithm of the observations; i.e.

$$\log G = \frac{1}{n}\sum \log x$$ [redacted]

HARMONIC MEAN

Average Rates ,SPEED

- For a given set of non-zero observations, harmonic mean is defined as the reciprocal of the AM of the reciprocals of the observation.

Discrete Observation

$$H = \frac{n}{\Sigma(1/x_i)}$$

Frequency Distribution

$$H = \frac{N}{\Sigma\left[\frac{f_i}{x_i}\right]}$$

PROPERTIES OF HARMONIC MEAN

- If all the values assumed by a variable are constant , say k , then the Harmonic Mean is also k .

HM of 1,1/2, 1/3,.................1/n is given by

$$\frac{2}{(n+1)}$$

- To calculate Average speed , use Harmonic Mean .

  The harmonic mean of two numbers x and y is given by $\frac{2xy}{x + y}$

PROPERTIES OF HARMONIC MEAN

- If there are two groups containing n_1 and n_2 observations with respective harmonic means as H_1 and H_2 , then combined Harmonic Mean is given by

$$\frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$$

RELATION BETWEEN AM ,GM ,HM

When all the observations are distinct → $AM > GM > HM$

When all the observations are same → $AM = GM = HM$

When nothing is mentioned → $AM \geq GM \geq HM$

RELATION BETWEEN AM ,GM ,HM

$$GM^2 = AM \times HM$$

This result holds for only two positive observations

Values dividing a given set of observations into a number of equal parts.

Partition Values

- Median
- Quartiles
- Deciles
- Percentiles

MEDIAN – PARTITION VALUE

FOR DISCRETE OBSERVATION

$$\text{Median} = \begin{cases} \left(\dfrac{n+1}{2}\right)\text{th observation, if } n \text{ is odd} \\[2ex] \dfrac{\left(\dfrac{n}{2}\right)\text{th observation} + \left(\dfrac{n}{2}+1\right)\text{th observation}}{2}, \text{ if } n \text{ is even} \end{cases}$$

MEDIAN – PARTITION VALUE

FOR SIMPLE FREQUENCY DISTRIBUTION

- Arrange the series into ascending or descending order.
- Calculate cumulative frequency .
- Calculate $\frac{N+1}{2}$
- Check cumulative frequency which is greater than $\frac{N+1}{2}$
- The value of x corresponding to this cumulative frequency would be the median .

FOR GROUPED FREQUENCY DISTRIBUTION

Make sure series is exclusive

We proceed stepwise as follows:

Step 1. For the given frequency distribution and obtain $N = \Sigma f_i$.

Step 2. Find $\frac{N}{2}$.

Step 3. Look at the cumulative frequency just greater than $\frac{N}{2}$ and find the corresponding class, known as *median class* (as the middle-most observation lies in this class).

Step 4. Compute the median using the formula:

$$\text{Median, } M_e = l + \left\{ h \times \frac{\left(\frac{N}{2} - cf\right)}{f} \right\}, \text{ where}$$

l = lower limit of median class;

h = width of median class;

f = frequency of median class;

cf = cumulative frequency of the class preceding the median class;

$N = \Sigma f_i$.

PROPERTIES OF MEDIAN

- For a set of observations, the sum of absolute deviations is minimum when the deviations are taken from the median.

  $\Sigma\,|x_i - A|$ is minimum if we choose A as the median.

- If x and y are two variables, to be related by $y = a + bx$ for any two constants a and b, then the median of y is given by

  $y_{me} = a + b\,x_{me}$

Also called as POSITIONAL AVERAGE

Can be used in case of open end classifications

MEDIAN

NOT Based on ALL OBSERVATION

NOT Affected by Sampling Fluctuations

PARTITION VALUES

Name of PV	No . of equal parts	No. of PVs	Symbol
Median	2	1	Me
Quartile	4	3	Q_1 , Q_2 , Q_3
Decile	10	9	D_1 , D_2 , ...D_9
Percentile	100	99	P_1 ,P_2 , P_3.......P_{99}

PARTITION VALUE

DISCRETE OBSERVATIONS

$$(n + 1)\, p^{th} \text{ term}$$

Where

n denotes the total number of observations

- $p = 1/4, 2/4, 3/4$ for Q_1, Q_2 and Q_3 respectively.
- $p = 1/10, 2/10, \ldots\ldots\ldots\ldots 9/10$. For $D_1, D_2, \ldots\ldots, D_9$ respectively.
- $p = 1/100, 2/100, \ldots\ldots, 99/100$ for $P_1, P_2, P_3 \ldots P_{99}$ respectively.

MODE

- **For a given set of observations, mode may be defined as the value that occurs the maximum number of times.**
- **If two or more observations are having maximum frequency then there are multiple modes . { MULTIMODAL DISTRIBUTION }.**
- **If all observations are having same frequency then distribution has no mode .**
- *MODE IS NOT RIGIDLY DEFINED .*

MODE

Grouped FREQUENCY DISTRIBUTION

STEPS

Find the class interval with the highest frequency .

This class interval is called MODAL CLASS

Make sure that series is exclusive

MODE

Grouped FREQUENCY DISTRIBUTION

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

where l = lower limit of the modal class,

h = size of the class interval [redacted],

f_1 = frequency of the modal class,

f_0 = frequency of the class preceding the modal class,

f_2 = frequency of the class succeeding the modal class.

PROPERTIES OF MODE

- **Mode is affected due to change in scale and due to change in origin .**
- if $y = a + bx$, then

$$y_{mo} = a + bx_{mo}$$

NOT Based on ALL OBSERVATION

MODE

NOT RIGIDLY DEFINED

NOT Amenable to Mathematical Property

RELATIONSHIP BETWEEN MEAN , MODE AND MEDIAN

FOR SYMMETRIC DATA

Mean = Median = Mode

In case of MODERATELY SKEWED DISTRIBUTION (EMPIRICAL RELATIONSHIP)

Mean – Mode = 3(Mean – Median)

Or

Mode = 3 Median – 2 Mean

WEIGHTED AVERAGE

When the observations under consideration have a hierarchical order of importance, we take recourse to computing weighted average, which could be either weighted AM or weighted GM or weighted HM.

1 2 3 n

Que. The mean weight of 15 students is 110 kg. The mean weight of 5 of them is 100 kg. and of another five students is 125 kg. then the mean weight of the remaining students is :

(a) 120

(b) 105

(c) 115

(d) None of these

b

Que. The median of following numbers, which are given is ascending order is 25. Find the Value of X.

11 ,13, 15 ,19 ,(x + 2) ,(x + 4) ,30 ,35 ,39 ,46

(a) 22

(b) 20

(c) 15

(d) 30

a

Que. The third decile for the numbers 15, 10, 20, 25, 18, 11, 9, 12, is:

(a) 13

(b) 10.70

(c) 11

(d) 11.50

b

Que. If the first quartile is 142 and semi-inter quartile range is 18, then the value of median is:

(a) 151

(b) 160

(c) 178

(d) None of these

b

Que. The average salary of a group of unskilled workers is Rs 10,000 and that of a group of skilled workers is Rs 15,000 . If the combined salary is Rs 12,000 , then what is the percentage of skilled workers ?

(a) 40%

(b) 50%

(c) 60%

(d) None of these

Ans a

Que. For open - end classification , which of the following is the best measure of central tendency ?

(a) AM

(b) GM

(c) Median

(d) Mode

Ans c

Que. If there are two groups with 75 and 65 as harmonic means containing 15 and 13 observations , then combined HM is given by

(a)70

(b) 72.25

(c) 78

(d) 76

Ans a

Que. Given the weights for the numbers 1,2,3 ....n are respectively 1^2 , 2^2 , 3^2.................. n^2 then weighted HM is

(a)(2n +1)/4

(b) (2n+1)/6

(c) (2n+1)/3

(d) (2n+1)/2

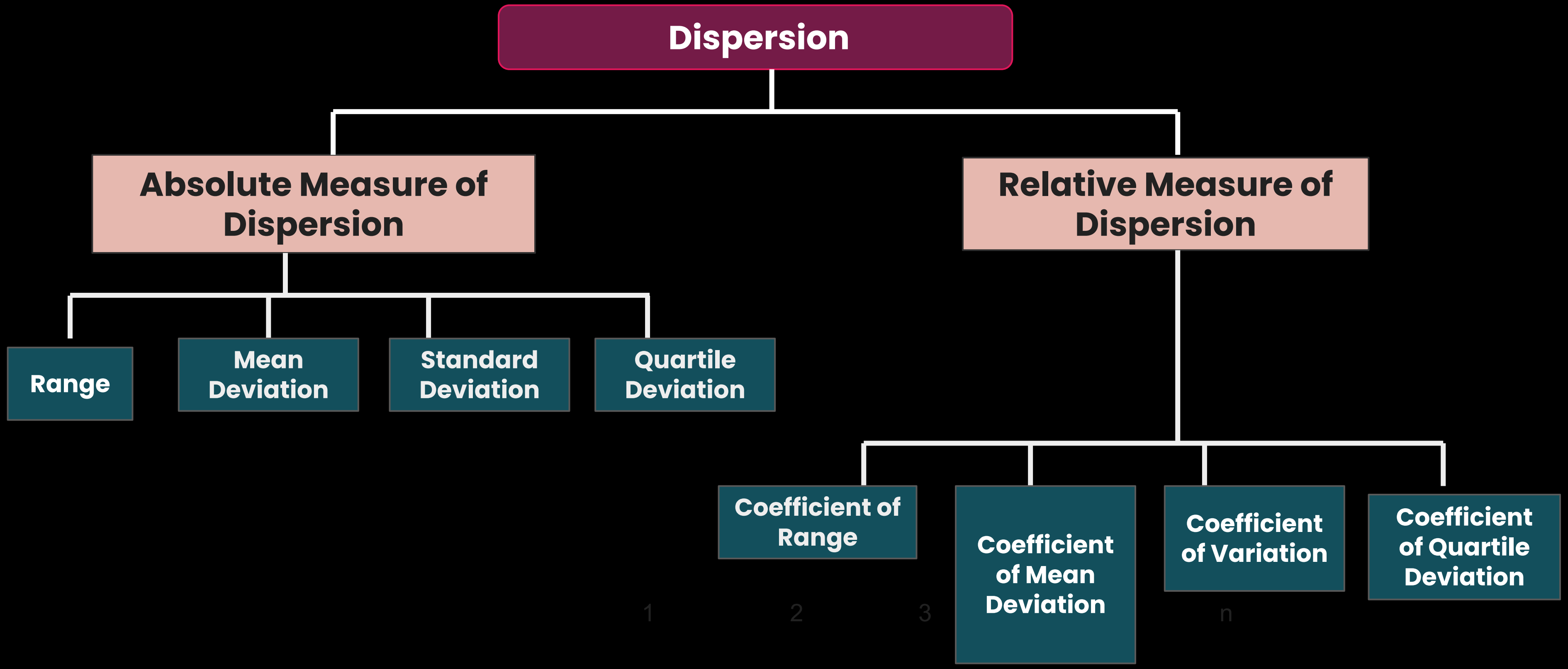
Ans c

USE MY CODE : SS12

DISPERSION

- Dispersion for a given set of observations may be defined as the amount of deviation of the observations, usually, from an appropriate measure of central tendency.

Absolute Measures Of Dispersion	Relative Measures Of Dispersion
I. Absolute measures are with units	I. Relative measures of dispersion are unit free.
2. These are not useful for comparison of two variables with different units	2. These are useful for comparison of two variables with different units
Example :	**Example** :


Dispersion
Absolute Measure of Dispersion
Relative Measure of Dispersion
Range
Mean Deviation
Standard Deviation
Quartile Deviation
Coefficient of Range
Coefficient of Mean Deviation
Coefficient of Variation
Coefficient of Quartile Deviation

RANGE

Discrete Observation

Range = L – S

Where ,

L : largest observations

S : smallest observations

COEFFICIENT OF RANGE

$$\text{Coefficient of range} = \frac{L-S}{L+S} \times 100$$

RANGE

Grouped Frequency distribution

Range = Uppermost Class Boundary – Lowermost Class Boundary

COEFFICIENT OF RANGE

$$\frac{\text{Uppermost class boundary} - \text{Lowermost class boundary}}{\text{Uppermost class boundary} + \text{Lowermost class boundary}} \times 100$$

Properties RANGE

If all the observations are constant i.e. equal, then the range is zero.

Range remains unaffected due to a change of origin but affected in the same ratio due to a change in scale .

If for any two constants a and b, two variables x and y are related by

$y = a + bx$, Then the range of y is given by

$$R_y = |b| \times R_x$$

MEAN DEVIATION

- For a given set of observation, mean deviation is defined as the arithmetic mean of the absolute deviations of the observations from an appropriate measure of central tendency.

- This appropriate measure of central tendency could be , Mean , Median , or Mode .

MEAN DEVIATION

Discrete Observation

$$MD_A = \frac{1}{n}\Sigma|x_i - A|$$

Frequency Distribution

$$MD_A = \frac{1}{N}\Sigma f\,|x - A|$$

COEFFICIENT OF Mean deviation

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation about A}}{\text{A}} \times 100$$

PROPERTIES MEAN DEVIATION

If all the observations are constant i.e. equal, then the mean deviation is zero

- Mean Deviation takes its minimum value when deviations are taken from Median .

- Mean Deviation remains unaffected due to a change of origin but affected in the same ratio due to a change in scale

STANDARD DEVIATION

- Standard deviation for a given set of observations is defined as the root mean square deviation when the deviations are taken from the AM of the observations.
- It is denoted by σ or SD
- The square of standard deviation is known as variance

STANDARD DEVIATION

DISCRETE OBSERVATION

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

Or

$$s = \sqrt{\frac{\sum x_i^{\,2}}{n} - \bar{x}^2}$$

FREQUENCY DISTRIBUTION

$$s = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{N}}$$

Or

$$\sqrt{\frac{\sum f_i x_i^{\,2}}{N} - \bar{x}^2}$$

Coefficient of Variation

$$\text{Coefficient of Variation (CV)} = \frac{\text{SD}}{\text{AM}} \times 100$$

PROPERTIES OF STANDARD DEVIATION

- If all the observations are constant i.e. equal, then the SD is zero.

- SD remains unaffected due to a change of origin but is affected in the same ratio due to a change of scale i.e.,

  if there are two variables x and y related as $y = a + bx$ for any two constants a and b,

  then SD of y is given by

  - $S_y = |b| S_x$

PROPERTIES OF STANDARD DEVIATION

SD for any two numbers

$$S = \frac{\text{Range}}{2}$$

SD for first n natural numbers

$$\sqrt{\frac{n^2-1}{12}}.$$

PROPERTIES OF STANDARD DEVIATION

- ***If there are two groups containing n_1 and n_2 observations, $\overline{x}_1$ and $\overline{x}_2$ as respective AM's, s_1 and s_2 as respective SD's , then the combined SD is given by***

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

where, $d_1 = \overline{x}_1 - \overline{x}$

$d_2 = \overline{x}_2 - \overline{x}$

and $\overline{x} = \dfrac{n_1 \overline{x}_1 + n_2 \overline{x}_2}{n_1 + n_2}$ = combined AM

QUARTILE DEVIATION

- Another measure of dispersion is provided by quartile deviation or semi-inter - quartile range which is given by

$$Q_d = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

PROPERTIES OF QUARTILE DEVIATION

- Quartile deviation provides the best measure of dispersion for open-end classification.

- It is also less affected due to extreme observations or sampling fluctuations.

- Quartile deviation remains unaffected due to a change of origin but is affected in the same ratio due to change in scale.

RELATIONSHIP BETWEEN SD, MD AND QD

$4\ SD = 5\ MD = 6\ QD$

Or

$SD : MD : QD = 15 : 12 : 10$

Que. Mean Deviation of data 3, 10 , 10 , 4, 7 ,18 , 5 from mode is

(a) 4.39

(b) 4.70

(c) 4.14

(d) 5.24

Ans c

Que. SD of first five consecutive natural numbers is

(a) $\sqrt{10}$

(b) $\sqrt{8}$

(c) $\sqrt{3}$

(d) $\sqrt{2}$

Ans d

USE MY CODE : SS12

Que. SD from numbers 1, 4 , 5 , 7, 8 is 2.45 . If 10 is added to each then SD will be

(a) 12.45

(b)24.5

(c)12

(d) will not change

Ans d

Que. If the mean and SD of X are a and b respectively , then SD of $(x-a)/b$ is

(a) a/b

(b) -1

(c) 1

(d) ab

Ans c

Que. If x and y are related by $y = 2x+5$ and the SD and AM of x are known to be 5 and 10 respectively , then the coefficient of variation of y is

(a)25

(b)30

(c)40

(d) 20

Ans c

Que. The sum of the squares of deviation from the mean of 10 observation is 250 . Mean of the data is 10 . Find the coefficient of variation

(a)10%

(b)25%

(c)50%

(d) 0%

Ans c

EXERCISE- Set (A)

Write down the correct answers. Each question carries 1 mark.

Que. 1 Measures of central tendency for a given set of observations measures

(a) The scatterness of the observations

(b) The central location of the observations

(c) Both (a) and (b)

(d) None of these.

b

Write down the correct answers. Each question carries 1 mark.

Que. 2 While computing the AM from a grouped frequency distribution, we assume that

(a) The classes are of equal length

(b) The classes have equal frequency

(c) All the values of a class are equal to the mid-value of that class

(d) None of these.

C

Write down the correct answers. Each question carries 1 mark.

Que. 3 Which of the following statements is wrong?

(a) Mean is rigidly defined

(b) Mean is not affected due to sampling fluctuations

(c) Mean has some mathematical properties

(d) All these

b

Write down the correct answers. Each question carries 1 mark.

Que. 4 Which of the following statements is true?

(a) Usually mean is the best measure of central tendency

(b) Usually median is the best measure of central tendency

(c) Usually mode is the best measure of central tendency

(d) Normally GM is the best measure of central tendency

a

Write down the correct answers. Each question carries 1 mark.

Que. 5 For open-end classification, which of the following is the best measure of central tendency?

(a) AM

(b) GM

(c) Median

(d) Mode

C

Write down the correct answers. Each question carries 1 mark.

Que. 6 The presence of extreme observations does not affect

(a) AM

(b) Median

(c) Mode

(d) Any of these.

b

Write down the correct answers. Each question carries 1 mark.

Que. 7 In case of an even number of observations which of the following is median?

(a) Any of the two middle-most value

(b) The simple average of these two middle values

(c) The weighted average of these two middle values

(d) Any of these

b

Write down the correct answers. Each question carries 1 mark.

Que. 8 The most commonly used measure of central tendency is

(a) AM

(b) Median

(c) Mode

(d) Both GM and HM.

a

Write down the correct answers. Each question carries 1 mark.

Que. 9 **Which one of the following is not uniquely defined?**

(a) Mean

(b) Median

(c) Mode

(d) All of these measures

C

Write down the correct answers. Each question carries 1 mark.

Que. 10 Which of the following measure of the central tendency is difficult to compute?

(a) Mean

(b) Median

(c) Mode

(d) GM

d

Write down the correct answers. Each question carries 1 mark.

Que. 11 Which measure(s) of central tendency is(are) considered for finding the average rates?

(a) AM

(b) GM

(c) HM

(d) Both (b) and (c)

d

Write down the correct answers. Each question carries 1 mark.

Que. 12 For a moderately skewed distribution, which of he following relationship holds?

(a) Mean – Mode = 3 (Mean – Median)

(b) Median – Mode = 3 (Mean – Median)

(c) Mean – Median = 3 (Mean – Mode)

(d) Mean – Median = 3 (Median – Mode)

a

Write down the correct answers. Each question carries 1 mark.

Que. 13 **Weighted averages are considered when**

(a) The data are not classified

(b) The data are put in the form of grouped frequency distribution

(c) All the observations are not of equal importance

(d) Both (a) and (c).

C

Write down the correct answers. Each question carries 1 mark.

Que. 14 Which of the following results hold for a set of distinct positive observations?

(a) AM $\geq$ GM $\geq$ HM

(b) HM $\geq$ GM $\geq$ AM

(c) AM > GM > HM

(d) GM > AM > HM

C

Write down the correct answers. Each question carries 1 mark.

Que. 15 When a firm registers both profits and losses, which of the following measure of central tendency cannot be considered?

(a) AM

(b) GM

(c) Median

(d) Mode

b

Write down the correct answers. Each question carries 1 mark.

Que. 16 Quartiles are the values dividing a given set of observations into

(a) Two equal parts

(b) Four equal parts

(c) Five equal parts

(d) None of these

b

Write down the correct answers. Each question carries 1 mark.

Que. 17 Quartiles can be determined graphically using

(a) Histogram

(b) Frequency Polygon

(c) Ogive

(d) Pie chart.

C

Write down the correct answers. Each question carries 1 mark.

Que. 18 **Which of the following measure(s) possesses (possess) mathematical properties?**

(a) AM

(b) GM

(c) HM

(d) All of these

d

Write down the correct answers. Each question carries 1 mark.

Que. 19 Which of the following measure(s) satisfies (satisfy) a linear relationship between two variables?

(a) Mean

(b) Median

(c) Mode

(d) All of these

d

Write down the correct answers. Each question carries one mark.

Que. 2 **Dispersion measures**

(a) The scatterness of a set of observations

(b) The concentration of a set of observations

(c) Both (a) and (b)

(d) Neither (a) and (b).

a

Write down the correct answers. Each question carries one mark.

Que. 3 When it comes to comparing two or more distributions we consider

(a) Absolute measures of dispersion

(b) Relative measures of dispersion

(c) Both (a) and (b)

(d) Either (a) or (b).

b

Write down the correct answers. Each question carries one mark.

Que. 4 Which one is easiest to compute?

(a) Relative measures of dispersion

(b) Absolute measures of dispersion

(c) Both (a) and (b)

(d) Range

d

Write down the correct answers. Each question carries one mark.

Que. 5 Which one is an absolute measure of dispersion?

(a) Range

(b) Mean Deviation

(c) Standard Deviation

(d) All these measures

d

Write down the correct answers. Each question carries one mark.

Que. 6 **Which measure of dispersion is most useful ?**

(a) Standard deviation

(b) Quartile deviation

(c) Mean deviation

(d) Range

a

Write down the correct answers. Each question carries one mark.

Que. 7 Which measures of dispersions is not affected by the presence of extreme observations?

(a) Range

(b) Mean deviation

(c) Standard deviation

(d) Quartile deviation

d

Write down the correct answers. Each question carries one mark.

Que. 8 Which measure of dispersion is based on the absolute deviations only?

(a) Standard deviation

(b) Mean deviation

(c) Quartile deviation

(d) Range

b

Write down the correct answers. Each question carries one mark.

Que. 9 **Which measure is based on only the central fifty percent of the observations?**

(a) Standard deviation

(b) Quartile deviation

(c) Mean deviation

(d) All these measures

b

Write down the correct answers. Each question carries one mark.

Que. 10 Which measure of dispersion is based on all the observations?

(a) Mean deviation

(b) Standard deviation

(c) Quartile deviation

(d) (a) and (b) but not (c)

d

Write down the correct answers. Each question carries one mark.

Que. 11 The appropriate measure of dispersion for open-end classification is

(a) Standard deviation

(b) Mean deviation

(c) Quartile deviation

(d) All these measures.

C

Write down the correct answers. Each question carries one mark.

Que. 12 The most commonly used measure of dispersion is

(a) Range

(b) Standard deviation

(c) Coefficient of variation

(d) Quartile deviation.

b

Write down the correct answers. Each question carries one mark.

Que. 13 Which measure of dispersion has some desirable mathematical properties?

(a) Standard deviation

(b) Mean deviation

(c) Quartile deviation

(d) All these measures

a

Write down the correct answers. Each question carries one mark.

Que. 14 If the profits of a company remains the same for the last ten months, then the standard deviation of profits for these ten months would be ?

(a) Positive

(b) Negative

(c) Zero

(d) (a) or (c)

C

Write down the correct answers. Each question carries one mark.

Que. 15 Which measure of dispersion is considered for finding a pooled measure of dispersion after combining several groups?

(a) Mean deviation

(b) Standard deviation

(c) Quartile deviation

(d) Any of these

b

Write down the correct answers. Each question carries one mark.

Que. 16 **A shift of origin has no impact on**

(a) Range

(b) Mean deviation

(c) Standard deviation

(d) All these and quartile deviation.

d

Write down the correct answers. Each question carries one mark.

Que. 17 The range of 15, 12, 10, 9, 17, 20 is

(a) 5

(b) 12

(c) 13

(d) 11.

d

Write down the correct answers. Each question carries one mark.

Que. 18 The standard deviation of 10, 16, 10, 16, 10, 10, 16, 16 is

(a) 4

(b) 6

(c) 3

(d) 0.

C

Write down the correct answers. Each question carries one mark.

Que. 19 For any two numbers SD is always

(a) Twice the range

(b) Half of the range

(c) Square of the range

(d) None of these.

b

Write down the correct answers. Each question carries one mark.

Que. 20 **If all the observations are increased by 10, then**

(a) SD would be increased by 10

(b) Mean deviation would be increased by 10

(c) Quartile deviation would be increased by 10

(d) All these three remain unchanged.

d

Write down the correct answers. Each question carries one mark.

Que. 21 If all the observations are multiplied by 2, then

(a) New SD would be also multiplied by 2

(b) New SD would be half of the previous SD

(c) New SD would be increased by 2

(d) New SD would be decreased by 2.

a

HISTORY

The first application of probability was made by a group of mathematicians in Europe about three hundreds years back to enhance their chances of winning in different games of gambling.

DIVISIONS OF PROBABILITY

SUBJECTIVE PROBABILITY

- Personal judgement
- Experience
- Influenced by the personal belief, attitude and bias of the person applying it

OBJECTIVE PROBABILITY

- Based on Rules and Maths

RANDOM EXPERIMENT

- If all possible outcomes of an experiment are known but the exact output cannot be predicted in advance then the experiment is known as Random Experiment

EXAMPLE :

Rolling a dice (or any number of dice),

Drawing items from a box containing both defective and non—defective items,

Drawing cards from a pack of well shuffled fifty two cards etc. are all random experiments.

SAMPLE SPACE

- The set of all possible outcomes of an experiment is called the sample space.

- **Example:**
  - List the sample space in tossing a fair coin.

    $S = \{H, T\}$

COINS

- Total number of elements in sample space while tossing a coin is given by 2^n

1. If a coin is tossed once $2^1 = 2$

   {H, T}

1. If two coins are tossed once or one coin tossed twice $2^2 = 4$

   {HH, HT, TH, TT}

1. If three coins are tossed once or one coin is tossed thrice

   $2^3 = 8$

   {HHH, HHT, HTH, THH, TTT, TTH, THT, HTT}

DICE

- Total number of elements in sample space while tossing a dice is given by 6^n

1. If a dice is rolled once $6^1 = 6$

   {1, 2, 3, 4, 5, 6}

DICE

2. If two die is rolled once or one dice is rolled twice

$6^2 = 36$

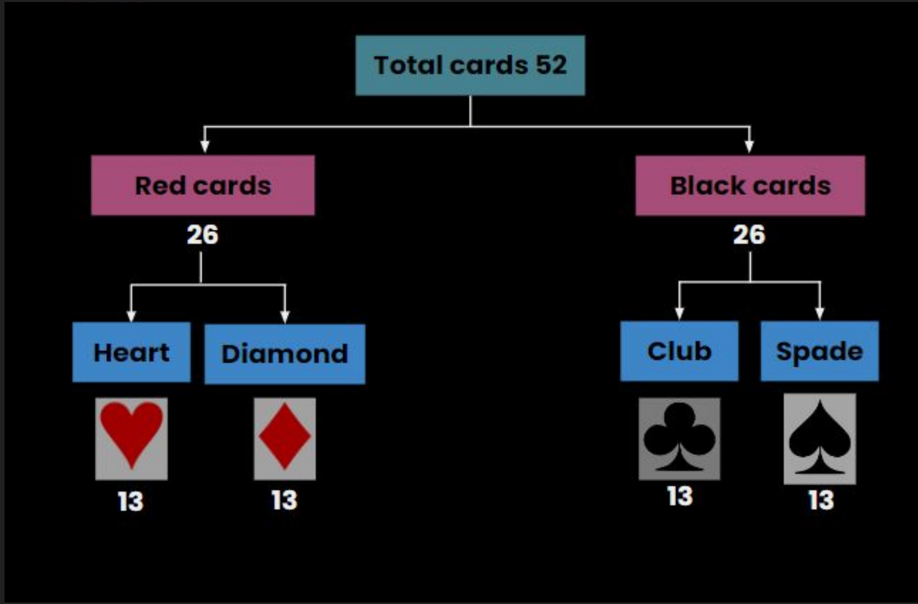
	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

DICE

3. If three dice are rolled once or one dice is rolled thrice

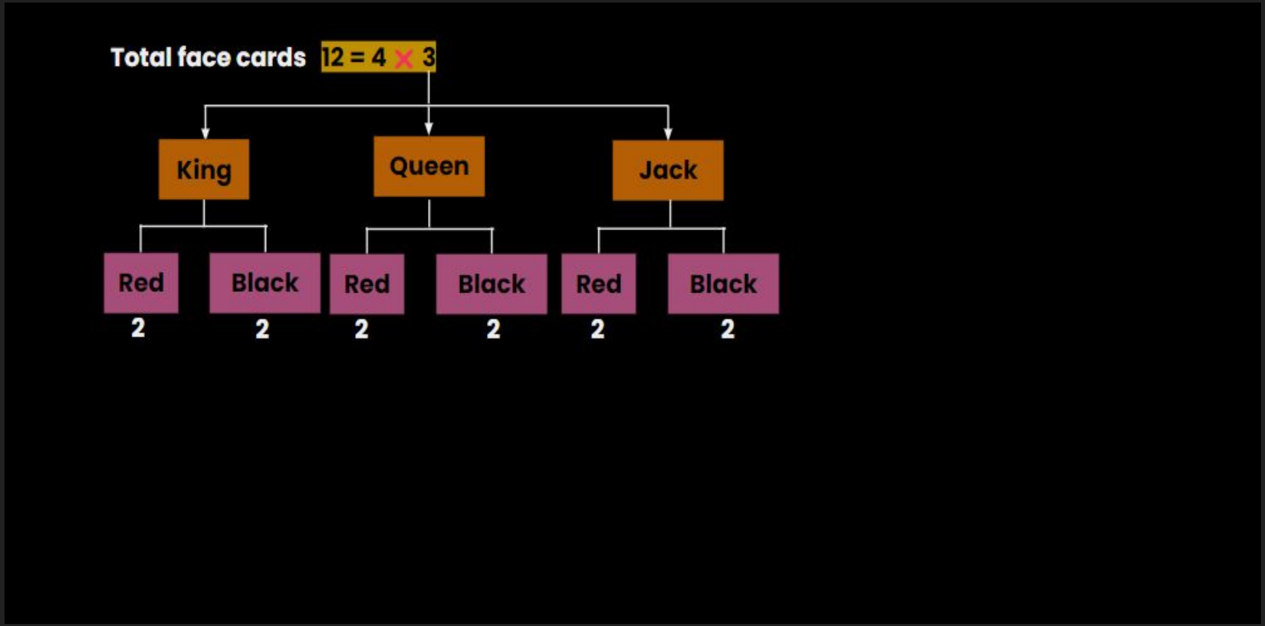
$6^3 = 216$

CARDS



Total cards 52
Red cards
26
Black cards
26
Heart
Diamond
Club
Spade
13
13
13
13

CARDS


Total face cards 12 = 4 × 3
King
Queen
Jack
Red
2
Black
2
Red
2
Black
2
Red
2
Black
2

EVENT

Any subset of a sample space is called event.

- The results or outcomes of a random experiment are known as events

EXAMPLE

The Event of getting a Prime number in a single throw of a die

$E = \{ 2, 3, 5 \}$

SIMPLE EVENT / Elementary Event

- Event at its simplest form
- Number of elements = 1

EXAMPLE : Tossing a coin once provides us two simple events namely Head and Tail

COMPOSITE / COMPOUND EVENT

- Event that can be subdivided into further events
- Number of elements > 1

EXAMPLE : Getting a head when coin is tossed twice

CLASSICAL DEFINITION OF PROBABILITY

- Also called Prior definition of Probability
- This formula is Event based
- It is given by Bernoulli and Laplace

$$P(A) = \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes}}$$

SURE EVENT

- If probability of occurrence of an event is 1
- Example

  Getting a number less than 7 on a throw of a single dice

IMPOSSIBLE EVENT

- If probability of occurrence of an event is 0
- Example

  Getting a number 7 on a throw of a single dice

RESULTS

(a) The probability of an event lies between 0 and 1, both inclusive.

i.e. $0 \leq P(A) \leq 1$

COMPLIMENTARY PROBABILITY

Probability of non - occurrence of an event A is denoted by $P(A')$ or $P(\overline{A})$ is called as complimentary event of A

$$P(A) + P(A') = 1$$

ODDS IN FAVOUR

Odds in favour of an event A

$$= \frac{\text{no of favorable events to A}}{\text{no of unfavorable events to A}}$$

ODDS AGAINST AN EVENT

Odds against an event A

$$= \frac{\text{no of unfavourable events to A}}{\text{no of favourable events to A}}$$

PROBABILITY OF AN EVENT

$$P(A) = \frac{\text{no of favourable events to A}}{\text{no of favourable + no of unfavourable}}$$

LIMITATIONS OF CLASSICAL PROBABILITY

i. It is applicable only when the total no. of events is finite.

ii. It can be used only when the events are equally likely or equi-probable.

iii. This definition has only a limited field of application like coin tossing, dice throwing, drawing cards etc. where the possible events are known well in advance

EQUALLY LIKELY EVENTS / MUTUALLY SYMMETRIC EVENTS / EQUI - PROBABLE EVENTS

- Events are equally likely if they have same probability of occurrence .

Example

- Getting head and getting tail on the toss of fair coins .

MUTUALLY EXCLUSIVE / DISJOINT / INCOMPATIBLE EVENTS

- Two events A and B are said to be mutually exclusive or disjoint if their simultaneous occurrence is impossible .
- If A and B are mutually exclusive then $A \cap B = \Phi$.
- EXAMPLE   Random Experiment : Throwing a dice

  A : getting odd number

  B : getting even number

EXHAUSTIVE EVENTS

- The events A1, A2, A3, ………… are known to form an exhaustive set if one of these events must necessarily occur.
- Example, the two events Head and Tail, when a coin is tossed once, are exhaustive as no other event except these two can occur.
- Events whose union is equal to sample space .

ADDITION THEOREM OF PROBABILITY

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If A and B are mutually exclusive

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- If A ,B and C are mutually exclusive

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

RESULT

- Two events A and B are exhaustive if

$$P(A \cup B) = 1$$

- Three events A, B and C are exhaustive if

$$P(A \cup B \cup C) = 1$$

- If A, B and C are mutually exclusive and exhaustive events,

then $P(A) + P(B) + P(C) = 1$

RESULT

- Three events A, B and C are equally likely if

$$P(A) = P(B) = P(C)$$

- Probability that only event A occurs

$$P(A-B) = P(A \cap B') = P(A) - P(A \cap B)$$

- Probability that only event B occurs

$$P(B-A) = P(B \cap A') = P(B) - P(A \cap B)$$

CONDITIONAL PROBABILITY

Event for which we are finding Conditional Probability

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Event which is occured

CONDITIONAL PROBABILITY

$$\boxed{P(B/A) = \frac{P(B \cap A)}{P(A)}}$$

Provided $P(A) \neq 0$

- We use the notation P(B/A), to be read as 'probability of the event B given that the event A has already occurred

CONDITIONAL PROBABILITY

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

if $P(B) \neq 0$.

- We use the notation P(A/B), to be read as 'probability of the event A given that the event B has already occurred

COMPOUND PROBABILITY / JOINT PROBABILITY

- The probability of occurrence of two events A and B simultaneously is known as the Compound Probability or Joint Probability of the events A and B and is denoted by $P(A \cap B)$

- In a similar manner, the probability of simultaneous occurrence of K events $A_1, A_2, \ldots A_k$, is denoted by $P(A_1 \cap A_2 \cap \ldots \cap A_k)$.

COMPOUND PROBABILITY / JOINT PROBABILITY

WITHOUT REPLACEMENT

(DEPENDENT EVENT)

$$P(A \cap B) = P(A) \, . \, P(B \mid A)$$

WITH REPLACEMENT

(INDEPENDENT EVENT)

$$P(A \cap B) = P(A) \, . \, P(B)$$

It is used when we have to find simultaneous occurrence of two or more events

RESULT

- It may be further noted that if two events A and B are independent, then the following pairs of events are also independent:

  (i) A and B′

  (ii) A′ and B

  (iii) A′ and B′

RANDOM VARIABLE / PROBABILITY DISTRIBUTION

If a coin is tossed three times

S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

X denotes the number of heads , then X is a random variable variable.

X = { 0 , 1, 2 ,3 }

X	0	1	2	3
P	1/8	3/8	3/8	1/8

RANDOM VARIABLE / PROBABILITY DISTRIBUTION

X	0	1	2	3
P	1/8	3/8	3/8	1/8

Expected Value

$$\mu = E(x) = \sum p_i x_i$$

$$E(X) = 0 \times 1/8 + 1 \times 3/8 + 2 \times 3/8 + 3 \times 1/8$$
$$= 12/8$$
$$= 1.5$$

RANDOM VARIABLE / PROBABILITY DISTRIBUTION

X	0	1	2	3
P	1/8	3/8	3/8	1/8

Variance of x, to be denoted by , σ^2 is given by

$$V(x) = \sigma^2 = E(x-\mu)^2$$

$$= E(x^2) - \mu^2$$

$E(X^2) = \frac{1}{8} \times 0^2 + \frac{3}{8} \times 1^2 + \frac{3}{8} \times 2^2 + \frac{1}{8} \times 3^2 = 3$

$E(x) = 1.5$

$v(x) = 0.75$

$SD = \sqrt{0.75}$

EXERCISE- Set (A)

Write down the correct answer. Each question carries 1 mark.

Que. 1 Initially, probability was a branch of

(a) Physics

(b) Statistics

(c) Mathematics

(d) Economics

Write down the correct answer. Each question carries 1 mark.

Que. 2 **Two broad divisions of probability are**

(a) Subjective probability and objective probability

(b) Deductive probability and non–deductive probability

(c) Statistical probability and Mathematical probability

(d) None of these

Write down the correct answer. Each question carries 1 mark.

Que. 3 **Subjective probability may be used in**

(a) Mathematics

(b) Statistics

(c) Management

(d) Accountancy

Write down the correct answer. Each question carries 1 mark.

Que. 5 An event that can be split into further events is known as

(a) Complex event

(b) Mixed event

(c) Simple event

(d) Composite event

Write down the correct answer. Each question carries 1 mark.

Que. 6 Which of the following pairs of events are mutually exclusive?

(a) A : The student reads in a school. B : He studies Philosophy.

(b) A : Raju was born in India. B : He is a fine Engineer.

(c) A : Ruma is 16 years old. B : She is a good singer.

(d) A : Peter is under 15 years of age. B : Peter is a voter of Kolkata.

Write down the correct answer. Each question carries 1 mark.

Que. 8 If $P(A \cap B) = 0$, then the two events A and B are

(a) Mutually exclusive

(b) Exhaustive

(c) Equally likely

(d) Independent.

Write down the correct answer. Each question carries 1 mark.

Que. 9 If for two events A and B, $P(A \cup B) = 1$, then A and B are

(a) Mutually exclusive events

(b) Equally likely events

(c) Exhaustive events

(d) Dependent events.

Write down the correct answer. Each question carries 1 mark.

Que. 10 If an unbiased coin is tossed once, then the two events Head and Tail are

(a) Mutually exclusive

(b) Exhaustive

(c) Equally likely

(d) All these (a), (b) and (c).

Write down the correct answer. Each question carries 1 mark.

Que. 11 If P(A) = P(B), then the two events A and B are

(a) Independent

(b) Dependent

(c) Equally likely

(d) Both (a) and (c).

Write down the correct answer. Each question carries 1 mark.

Que. 12 If for two events A and B, $P(A \cap B) \neq P(A) \times P(B)$, then the two events A and B are

(a) Independent

(b) Dependent

(c) Not equally likely

(d) Not exhaustive.

Write down the correct answer. Each question carries 1 mark.

Que. 14 If two events A and B are independent, then

(a) A and the complement of B are independent

(b) B and the complement of A are independent

(c) Complements of A and B are independent

(d) All of these (a), (b) and (c).

Write down the correct answer. Each question carries 1 mark.

Que. 15 If two events A and B are independent, then

(a) They can be mutually exclusive

(b) They can not be mutually exclusive

(c) They can not be exhaustive

(d) Both (b) and (c).

Write down the correct answer. Each question carries 1 mark.

Que. 16 If two events A and B are mutually exclusive, then

(a) They are always independent

(b) They may be independent

(c) They can not be independent

(d) They can not be equally likely.

Write down the correct answer. Each question carries 1 mark.

Que. 18 The probability of an event can assume any value between

(a) – 1 and 1

(b) 0 and 1, including 0 and 1

(c) – 1 and 0

(d) none of these.

Write down the correct answer. Each question carries 1 mark.

Que. 19 If $P(A) = 0$, then the event A

(a) will never happen

(b) will always happen

(c) may happen

(d) may not happen.

Write down the correct answer. Each question carries 1 mark.

Que. 20 If $P(A) = 1$, then the event A is known as

(a) symmetric event

(b) dependent event

(c) improbable event

(d) sure event.

Write down the correct answer. Each question carries 1 mark.

Que. 21 If p : q are the odds in favour of an event, then the probability of that event is

(a) p/q

(b) $p/(p + q)$

(c) $q/(p + q)$

(d) none of these

Write down the correct answer. Each question carries 1 mark.

Que. 22 If $P(A) = 5/9$, then the odds against the event A is

(a) 5 : 9

(b) 5 : 4

(c) 4 : 5

(d) 5 : 14

Write down the correct answer. Each question carries 1 mark.

Que. 23 **If A, B and C are mutually exclusive and exhaustive events, then P(A) + P(B) + P(C) equals to**

(a) 1/3

(b) 1

(c) 0

(d) any value between 0 and 1.

Write down the correct answer. Each question carries 1 mark.

Que. 25 P(B/A) is defined only when

(a) A is a sure event

(b) B is a sure event

(c) A is not an impossible event

(d) B is an impossible event.

Write down the correct answer. Each question carries 1 mark.

Que. 26 P(A/B') is defined only when

(a) B is not a sure event

(b) B is a sure event

(c) B is an impossible event

(d) B is not an impossible event.

Write down the correct answer. Each question carries 1 mark.

Que. 27 For two events A and B, $P(A \cup B) = P(A) + P(B)$ only when

(a) A and B are equally likely events

(b) A and B are exhaustive events

(c) A and B are mutually independent

(d) A and B are mutually exclusive.

Write down the correct answer. Each question carries 1 mark.

Que. 28 Addition Theorem of Probability states that for any two events A and B,

(a) $P(A \cup B) = P(A) + P(B)$

(b) $P(A \cup B) = P(A) + P(B) + P(A \cap B)$

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(d) $P(A \cup B) = P(A) \times P(B)$

Write down the correct answer. Each question carries 1 mark.

Que. 30 For any two events A and B,

(a) $P(A-B) = P(A) - P(B)$

(b) $P(A-B) = P(A) - P(A \cap B)$

(c) $P(A-B) = P(B) - P(A \cap B)$

(d) $P(B-A) = P(B) + P(A \cap B)$.

Write down the correct answer. Each question carries 1 mark.

Que. 31 The limitations of the classical definition of probability

(a) it is applicable when the total number of elementary events is finite

(b) it is applicable if the elementary events are equally likely

(c) it is applicable if the elementary events are mutually independent

(d) (a) and (b).

Write down the correct answer. Each question carries 1 mark.

Que. 32 According to the statistical definition of probability, the probability of an event A is the

(a) limiting value of the ratio of the no. of times the event A occurs to the number of times the experiment is repeated

(b) the ratio of the frequency of the occurrences of A to the total frequency

(c) the ratio of the frequency of the occurrences of A to the non-occurrence of A

(d) the ratio of the favourable elementary events to A to the total number of elementary events.

Write down the correct answer. Each question carries 1 mark.

Que. 33 The Theorem of Compound Probability states that for any two events A and B.

(a) $P(A \cap B) = P(A)\ P(B/A)$

(b) $P(A \cup B) = P(A) \times P(B/A)$

(c) $P(A \cap B) = P(A) \times P(B)$

(d) $P(A \cup B) = P(B) + P(B) - P(A \cap B)$.

Write down the correct answer. Each question carries 1 mark.

Que. 34 If A and B are mutually exclusive events, then

(a) $P(A) = P(A-B)$.

(b) $P(B) = P(A-B)$.

(c) $P(A) = P(A \cap B)$.

(d) $P(B) = P(A \cap B)$.

Write down the correct answer. Each question carries 1 mark.

Que. 35 If $P(A-B) = P(B-A)$, then the two events A and B satisfy the condition

(a) $P(A) = P(B)$.

(b) $P(A) + P(B) = 1$

(c) $P(A \cap B) = 0$

(d) $P(A \cup B) = 1$

Write down the correct answer. Each question carries 1 mark.

Que. 36 The number of conditions to be satisfied by three events A, B and C for complete independence is

(a) 2

(b) 3

(c) 4

(d) any number.

Write down the correct answer. Each question carries 1 mark.

Que. 37 If two events A and B are independent, then $P(A\ \ B)$

(a) equals to $P(A) + P(B)$

(b) equals to $P(A) \times P(B)$

(c) equals to $P(A) \times P(B/A)$

(d) equals to $P(B) \times P(A/B)$.

Write down the correct answer. Each question carries 1 mark.

Que. 42 If a random variable x assumes the values x_1, x_2, x_3, x_4 with corresponding probabilities p_1, p_2, p_3, p_4 then the expected value of x is

(a) $p_1 + p_2 + p_3 + p_4$

(b) $x_1p_1 + x_2p_3 + x_3p_2 + x_4p_4$

(c) $p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4$

(d) none of these.

Write down the correct answer. Each question carries 1 mark.

Que. 43 f(x), the probability mass function of a random variable x satisfies

(a) $f(x) > 0$

(b) $\sum_{x} f(x) = 1$

(c) both (a) and (b)

(d) $\hat{f}(x) \geq 0$ and $\sum_{x} f(x) = 1$

Write down the correct answer. Each question carries 1 mark.

Que. 44 Variance of a random variable x is given by

(a) $E(X - \mu)^2$

(b) $E[X - E(X)]^2$

(c) $E(X^2 - \mu)$

(d) a or b

Write down the correct answer. Each question carries 1 mark.

Que. 47 **If an unbiased coin is tossed twice, the probability of obtaining at least one tail is**

(a) 0.25

(b) 0.50

(c) 0.75

(d) 1.00

Write down the correct answer. Each question carries 1 mark.

Que. 49 A bag contains 15 one rupee coins, 25 two rupee coins and 10 five rupee coins. If a coin is selected at random from the bag, then the probability of not selecting a one rupee coin is

(a) 0.30

(b) 0.70

(c) 0.25

(d) 0.20

Write down the correct answer. Each question carries 1 mark.

Que. 50 A, B, C are three mutually independent with probabilities 0.3, 0.2 and 0.4 respectively. What is $P(A \cap B \cap C)$?

(a) 0.400

(b) 0.240

(c) 0.024

(d) 0.500

Write down the correct answer. Each question carries 1 mark.

Que. 51 If two letters are taken at random from the word HOME, what is the Probability that none of the letters would be vowels?

(a) 1/6

(b) 1/2

(c) 1/3

(d) 1/4

Write down the correct answer. Each question carries 1 mark.

Que. 52 If a card is drawn at random from a pack of 52 cards, what is the chance of getting a Spade or an ace?

(a) 4/13

(b) 5/13

(c) 0.25

(d) 0.20

Que. If x be the sum of two numbers obtained when two die are thrown simultaneously then $P(x \geq 7)$ is

(a) 5/12

(b) 7/12

(c) 11/15

(d) 3/8

Ans c

Que. If x be the sum of two numbers obtained when two die are thrown simultaneously then $P(x \geq 7)$ is

(a) 5/12

(b) 7/12

(c) 11/15

(d) 3/8

b

Que. Three coins are tossed together, the probability of getting exactly two head is: [2015-DEC]

(a) 5/8

(b) 3/8

(c) 1/8

(d) None

Ans : b

Que. The probability that a leap year has 53 Wednesday is

[2018-NOV]

(a) 2/7

(b) 3/5

(c) 2/3

(d) 1/7

Ans : a

Que. Ram is known to hit a target in 2 out of 3 shots where as Shyam is known to hit the same target in 5 out of 11 shots. What is the probability that the target would be hit if they both try? [2018-NOV]

(a) 9/11

(b) 3/11

(c) 10/33

(d) 6/11

Ans : a

Que. If in a class, 60% of the student study. Mathematics and science and 90% of the student study science, then the probability of a student studying mathematics given that he/she is already studying science is: [2021-JULY]

(a) 1/4

(b) 2/3

(c) 1

(d) 1/2

Ans : b

Que. The value of K for the probability density function of a variate X is equal to:

[2021-JULY]

X	0	1	2	3	4	5	6
P(x)	5k	3k	4k	6k	7k	9k	11k

(a) 3/9

(b) 1/40

(c) 1/49

(d) 1/45

Ans : d

Que. Assume that the probability for rain on a day is 0.4. An umbrella salesman can earn ₹ 400 per day in case of rain on that day and will lose ₹ 100 per day if there is no rain. The expected earnings in (in ₹) per day of the salesman is

[2021-DEC]

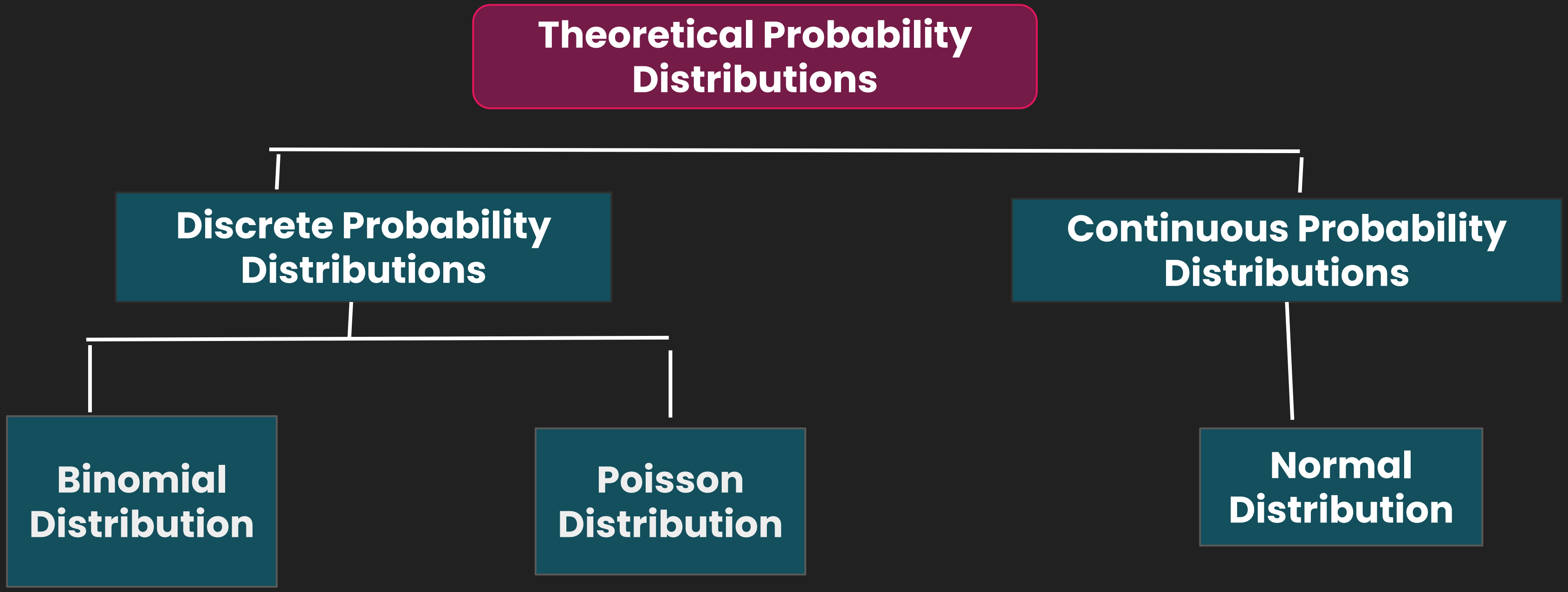
(a) 400

(b) 200

(c) 100

(d) 0

Ans : c


Theoretical Probability Distributions
Discrete Probability Distributions
Continuous Probability Distributions
Binomial Distribution
Poisson Distribution
Normal Distribution

NOTE

RANDOM VARIABLE	PROBABILITY FUNCTION
Discrete	Probability mass function
Continuous	Probability Density function

BINOMIAL DISTRIBUTION

- It is derived from a particular type of random experiment known as Bernoulli process named after the famous mathematician

CHARACTERISTICS OF BERNOULLI TRIALS

i. Each trial is associated with two mutually exclusive and exhaustive outcomes (one is 'success' and other is 'failure')

ii. The trials are independent.

iii. The probability of a success (p) and failure,($q = 1-p$), remain unchanged throughout the process.

iv. The number of trials is a finite positive integer.

BINOMIAL DISTRIBUTION

(bi - parametric discrete probability distribution)

- A **discrete random variable X** is defined to follow binomial distribution with parameters n and p,

$$X \sim B(n, p),$$

Probability Mass Function

$$f(x) = p(X = x) = {}^{n}c_{x}\, p^{x}\, q^{n-x} \quad \text{for } x = 0, 1, 2, \dots, n$$

BINOMIAL DISTRIBUTION

(bi - parametric discrete probability distribution)

Mean

$$\mu = n\,p$$

Variance

- The variance of the binomial distribution is given by

$$\sigma^2 = n\,p\,q$$

Variance of a binomial variable is always less than its mean.

Variance of X attains its maximum value at $p = q = 0.5$ and this maximum value is $n/4$.

BINOMIAL DISTRIBUTION

MODE

$(n+1)p$

INTEGER

- $\mu_0 = (n+1)p$
- $\mu_0 = (n+1)p - 1$

Bi - Modal

NON - INTEGER

μ_0 = the largest integer contained in $(n+1)p$

Uni- Modal

ADDITIVE PROPERTY

If X and Y are two independent variables such that

$X \sim B(n_1, p)$ and $Y \sim B(n_2, p)$

Then $(X+Y) \sim B(n_1 + n_2, p)$

POISSON DISTRIBUTION

(UNI- parametric
discrete probability distribution)

- Poisson distribution is applied when the total number of events is pretty large but the probability of occurrence is very small.

- A discrete random variable X that follows Poisson Distribution denoted as

$$X \sim P(m)$$

POISSON DISTRIBUTION

- A discrete random variable X that follows Poisson Distribution denoted as

$$X \sim P(m)$$

Probability Mass Function

$$f(x) = P(X = x) = \frac{e^{-m} . m^{x}}{x!} \text{ for } x = 0, 1, 2, \dots \infty$$

where ,

$$e = 2.71828$$

$$m = n\,p$$

POISSON DISTRIBUTION

Mean

- The mean of Poisson distribution is given by

$$\mu = m$$

Variance

- The variance of Poisson distribution is given by

$$\sigma^2 = m$$

Standard Deviation

$$\sqrt{m}$$

POISSON DISTRIBUTION

MODE

m

INTEGER

- $\mu_0 = m$
- $\mu_0 = m - 1$

Bi - Modal

NON - INTEGER

μ_0 = the largest integer contained in m

Uni- Modal

ADDITIVE PROPERTY

- If X and Y are two independent variables such that

$X \sim P(m_1)$ and $Y \sim P(m_2)$

$$X + Y \sim P(m_1 + m_2)$$

Poisson Model

- **Let us think of a random experiment under the following conditions:**

I. **The probability of finding success in a very small time interval (t, t + dt) is kt, where k (>0) is a constant.**

II. **The probability of having more than one success in this time interval is very low.**

III. **The probability of having success in this time interval is independent of t as well as earlier successes.**

NORMAL DISTRIBUTION

(BI - parametric
CONTINUOUS probability distribution)

- A continuous random variable x is defined to follow normal distribution with parameters μ and σ^2, to be denoted by

$$X \sim N(\mu, \sigma^2)$$

NORMAL DISTRIBUTION

(BI - parametric
CONTINUOUS probability distribution)

Probability Density Function

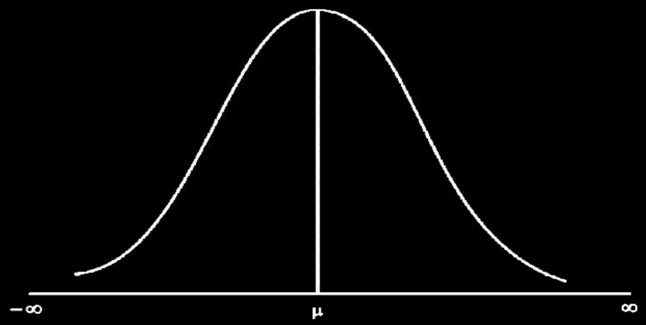
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-(\bar{x}-u)^2/2\sigma^2}$$

for $-\infty < x < \infty$.

- e = 2.71828
- X = random variable
- **μ = mean of normal random variable x**
- **σ = standard deviation of the given normal distribution**

NORMAL CURVE

- The normal curve is bell shaped .
- The line drawn through x = μ has divided the normal curve into two parts which are equal in all respect.
- Normal distribution is symmetrical about x = μ. As such, its skewness is zero
- The two tails of the normal curve extend indefinitely on both sides of the curve and both the left and right tails never touch the horizontal axis.
- The total area of the normal curve or for that any probability curve is taken to be unity i.e. one.



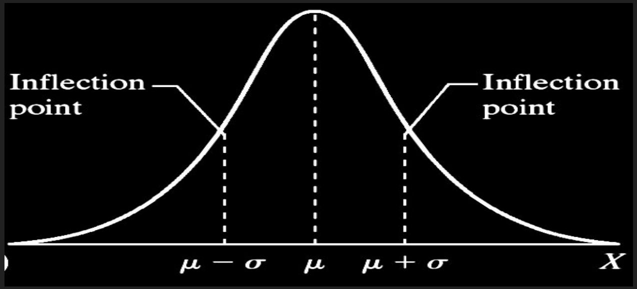
Normal curve /probability curve,

The area under this curve gives us the probability .

The area between −∞ to μ = the area between to μ to ∞ = 0.5

NORMAL DISTRIBUTION

MEAN = MEDIAN = MODE = μ (Symmetric distribution)

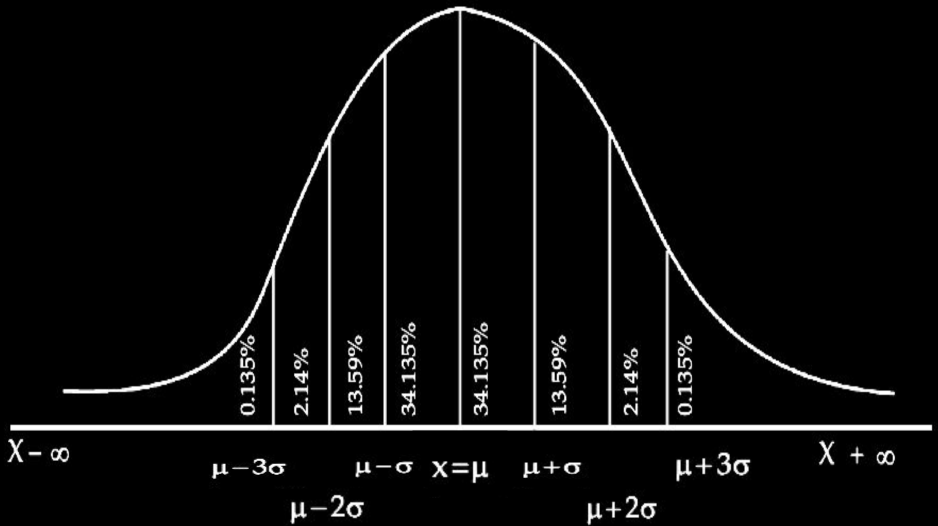


VARIANCE	σ^2 (given in question)
Standard deviation	σ
Mean deviation	$0.8\ \sigma$
Quartile Deviation	$0.675\ \sigma$
Quartiles	$Q_1 = \mu - 0.675\sigma$ $Q_3 = \mu + 0.675\sigma$

Two points of inflexion

- $\mu - \sigma$ and $\mu + \sigma$

NORMAL CURVE



$$P(\mu - \sigma < x < \mu + \sigma) = 0.6828$$

$$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9546$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$$

NORMAL CURVE

- If x and y are independent normal variables with means and standard deviations as μ_1 and μ_2 and σ_1 and σ_2, respectively, then z = x + y also follows normal distribution with

  $SD = \sqrt{\sigma_1^2 + \sigma_2^2}$ respectively.

- mean $(\mu_1 + \mu_2)$ and

STANDARD NORMAL DISTRIBUTION

- If we take $\mu = 0$ and $\sigma = 1$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{for } -\infty < z < \infty$$

- The random variable z is known as standard normal variate (or variable) or standard normal deviate.

- It is given by $z = \dfrac{x - \mu}{\sigma}$

IMPORTANT RESULTS of STANDARD NORMAL DISTRIBUTION

- Mean = Median = Mode= 0
- The standard normal distribution is symmetrical about z = 0
- Variance = 1
- Standard deviation = 1
- Point of Inflexion = -1 and 1
- Mean deviation = 0.8
- Quartile deviation = 0.675

Cumulative Distribution Function

$$P(z \leq k) = \phi(k)$$

$$P(x < a) = P\left[\frac{x-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right]$$

$$= P(z < k), (k = a - \mu/\sigma)$$

$$= \phi(k) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16.27)$$

Also $P(x \leq a) = P(x < a)$ as x is continuous.

$$\phi(-k) = 1 - \phi(k)$$

$$P(x > b) = 1 - P(x \leq b)$$

$$= 1 - \phi(b - \mu/\sigma) \ldots\ldots\ldots\ldots (16.28)$$

$$P(a < x < b) = \phi(b - \mu/\sigma) - \phi(a - \mu/\sigma)$$

- $\phi(k)$ gives the area from $-\infty$ to the point K
- Z table gives us the probability of values z = 0 to any value of z

Que. If the points of inflexion of a normal curve are 40 and 60 respectively, then its mean deviation is: [2019-JUNE]

(a) 8

(b) 45

(c) 50

(d) 60

Ans : a

Que. If the points of inflexion of a normal curve are 40 and 60 respectively, then its mean deviation is: [2019-JUNE]

(a) 8

(b) 45

(c) 50

(d) 60

Ans : a

Write down the correct answers. Each question carries 1 mark.

Que. 2 Probability distribution may be

(a) discrete.

(b) continuous.

(c) infinite.

(d) (a) or (b).

Write down the correct answers. Each question carries 1 mark.

Que. 3 An important discrete probability distribution is

(a) Poisson distribution.

(b) Normal distribution.

(c) Cauchy distribution.

(d) Log normal distribution.

Write down the correct answers. Each question carries 1 mark.

Que. 4 An important continuous probability distribution

(a) Binomial distribution.

(b) Poisson distribution.

(c) Geometric distribution.

(d) Normal distribution.

Write down the correct answers. Each question carries 1 mark.

Que. 8 **The important characteristic(s) of Bernoulli trials**

(a) each trial is associated with just two possible outcomes.

(b) trials are independent.

(c) trials are infinite.

(d) both (a) and (b).

Write down the correct answers. Each question carries 1 mark.

Que. 9 **The probability mass function of binomial distribution is given by**

(a) $f(x) = p^x q^{n-x}$.

(b) $f(x) = {}^n c_x p^x q^{\,n-x}$.

(c) $f(x) = {}^n c_x q^x q^{\,n-x}$..

(d) $f(x) = {}^n c_x p^{n-x} q^x$.

Write down the correct answers. Each question carries 1 mark.

Que. 10 If x is a binomial variable with parameters n and p, then x can assume

(a) any value between 0 and n.

(b) any value between 0 and n, both inclusive.

(c) any whole number between 0 and n, both inclusive.

(d) any number between 0 and infinity.

Write down the correct answers. Each question carries 1 mark.

Que. 11 A binomial distribution is

(a) never symmetrical.

(b) never positively skewed.

(c) never negatively skewed.

(d) symmetrical when $p = 0.5$.

Write down the correct answers. Each question carries 1 mark.

Que. 12 The mean of a binomial distribution with parameter n and p is

(a) $n(1-p)$.

(b) $np(1-p)$.

(c) np.

(d) $\sqrt{np(1-p)}$.

Write down the correct answers. Each question carries 1 mark.

Que. 13 The Variance of a binomial distribution with parameter n and p is

(a) $np^2(1-p)$.

(b) $\sqrt{np(1-p)}$.

(c) $nq(1-q)$.

(d) $np^2pp^2(1-p)p^2$

Write down the correct answers. Each question carries 1 mark.

Que. 14 An example of a bi-parametric discrete probability distribution is

(a) binomial distribution.

(b) poisson distribution.

(c) normal distribution.

(d) both (a) and (b).

Write down the correct answers. Each question carries 1 mark.

Que. 15 For a binomial distribution, mean and mode

(a) are never equal.

(b) are always equal.

(c) are equal when $q = 0.50$.

(d) do not always exist.

Write down the correct answers. Each question carries 1 mark.

Que. 16 The mean of binomial distribution is

(a) always more than its variance.

(b) always equal to its variance.

(c) always less than its variance.

(d) always equal to its standard deviation.

Write down the correct answers. Each question carries 1 mark.

Que. 18 The maximum value of the variance of a binomial distribution with parameters n and p is

(a) n/2.

(b) n/4.

(c) np (1 – p).

(d) 2n.

Write down the correct answers. Each question carries 1 mark.

Que. 19 The method usually applied for fitting a binomial distribution is known as

(a) method of least square.

(b) method of moments.

(c) method of probability distribution.

(d) method of deviations.

Write down the correct answers. Each question carries 1 mark.

Que. 20 Which one is not a condition of Poisson model?

(a) the probability of having success in a small time interval is constant.

(b) the probability of having success more than one in a small time interval is very small.

(c) the probability of having success in a small interval is independent of time and also of earlier success.

(d) the probability of having success in a small time interval $(t, t + dt)$ is kt for a positive constant k.

Write down the correct answers. Each question carries 1 mark.

Que. 21 Which one is uniparametric distribution?

(a) Binomial.

(b) Poisson.

(c) Normal.

(d) Hyper geometric.

Write down the correct answers. Each question carries 1 mark.

Que. 22 For a Poisson distribution,

(a) mean and standard deviation are equal.

(b) mean and variance are equal.

(c) standard deviation and variance are equal.

(d) both (a) and (b).

Write down the correct answers. Each question carries 1 mark.

Que. 24 Poisson distribution is

(a) always symmetric.

(b) always positively skewed.

(c) always negatively skewed.

(d) symmetric only when m = 2.

Write down the correct answers. Each question carries 1 mark.

Que. 25 A binomial distribution with parameters n and p can be approximated by a Poisson distribution with parameter m = np is

(a) $n \to \infty$

(b) $p \to 0$.

(c) $n \to \infty$ and $p \to 0$.

(d) $n \to \infty$ and $p \to 0$ so that np remains finite..

Write down the correct answers. Each question carries 1 mark.

Que. 26 For Poisson fitting to an observed frequency distribution,

(a) we equate the Poisson parameter to the mean of the frequency distribution.

(b) we equate the Poisson parameter to the median of the distribution.

(c) we equate the Poisson parameter to the mode of the distribution.

(d) none of these.

Write down the correct answers. Each question carries 1 mark.

Que. 27 **The most important continuous probability distribution is known as**

(a) Binomial distribution.

(b) Normal distribution.

(c) Chi-square distribution.

(d) Sampling distribution.

Write down the correct answers. Each question carries 1 mark.

Que. 28 **The probability density function of a normal variable x is given by**

(a) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for $-\infty < x < \infty$.

(b) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{-(x-\mu)^2}{2\sigma^2}}$ for $0 < x < \infty$.

(c) $f(x) = \frac{1}{\sqrt{2\pi\sigma}} . e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for $-\infty < x < \infty$.

(d) none of these.

Write down the correct answers. Each question carries 1 mark.

Que. 29 The total area of the normal curve is

(a) one.

(b) 50 per cent.

(c) 0.50.

(d) any value between 0 and 1.

Write down the correct answers. Each question carries 1 mark.

Que. 30 The normal curve is

(a) Bell-shaped.

(b) U- shaped.

(c) J-shaped.

(d) Inverted J-shaped.

Write down the correct answers. Each question carries 1 mark.

Que. 31 The normal curve is

(a) positively skewed.

(b) negatively skewed.

(c) symmetrical.

(d) all these.

Write down the correct answers. Each question carries 1 mark.

Que. 32 Area of the normal curve

(a) between $-\propto$ to μ is 0.50.

(b) between $\propto$ to μ is 0.50.

(c) between $-\propto$ to $\propto$ is 0.50.

(d) both (a) and (b).

Write down the correct answers. Each question carries 1 mark.

Que. 34 The mean and mode of a normal distribution

(a) may be equal.

(b) may be different.

(c) are always equal.

(d) (a) or (b).

Write down the correct answers. Each question carries 1 mark.

Que. 35 The mean deviation about median of a standard normal variate is

(a) 0.675 .

(b) 0.675.

(c) 0.80 .

(d) 0.80.

Write down the correct answers. Each question carries 1 mark.

Que. 36 The quartile deviation of a normal distribution with mean 10 and SD 4 is

(a) 0.675.

(b) 67.50.

(c) 2.70.

(d) 3.20.

Write down the correct answers. Each question carries 1 mark.

Que. 37 **For a standard normal distribution, the points of inflexion are given by**

(a) $\mu - \sigma$ and $\mu + \sigma$.

(b) $-\sigma$ and σ.

(c) -1 and 1.

(d) 0 and 1.

Write down the correct answers. Each question carries 1 mark.

Que. 38 The symbol (a) indicates the area of the standard normal curve between

(a) 0 to a.

(b) a to ∞.

(c) – ∝ to a.

(d) – ∝ to ∝.

Write down the correct answers. Each question carries 1 mark.

Que. 39 **The interval $(\mu - 3\sigma, \mu + 3\sigma)$ covers**

(a) 95% area of a normal distribution.

(b) 96% area of a normal distribution.

(c) 99% area of a normal distribution.

(d) all but 0.27% area of a normal distribution.

Write down the correct answers. Each question carries 1 mark.

Que. 40 Number of misprints per page of a thick book follows

(a) Normal distribution.

(b) Poisson distribution.

(c) Binomial distribution.

(d) Standard normal distribution.

Write down the correct answers. Each question carries 1 mark.

Que. 41 The results of ODI matches between India and Pakistan follows

(a) Binomial distribution.

(b) Poisson distribution.

(c) Normal distribution.

(d) (b) or (c).

Write down the correct answers. Each question carries 1 mark.

Que. 42 The wage of workers of a factory follow

(a) Binomial distribution.

(b) Poisson distribution.

(c) Normal distribution.

(d) Chi-square distribution.

Que. Find mode when n = 15 and p = 1/4 in binomial distribution?

[2019-NOV]

(a) 4

(b) 4 and 3

(c) 4.2

(d) 3.75

Ans : b

Que. For a certain type of mobiles , the length of time between charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours . A person owns one of these mobiles and wants to know the probability that the length of time will be between 50 and 70 hours is $\phi(1.33) = 0.9082$, $\phi(0) = 0.5$ [2021- July]

(a) - 0.4082

(b) 0.5

(c) 0 .4082

(d) - 0.5

Ans : c

[2019-NOV]

Que. What is the SD and mean x if

$$f(x) = \frac{\sqrt{2}}{\sqrt{\pi}} e^{-2(x-3)^2}, -\infty < x < \infty.$$

(a) 3, 1/2

(b) 3, 1/4

(c) 2, 1/2

(d) 2, √2

a

Que. If Standard Deviation is 1.732 then what is the value of poisson distribution. The P[-2.48 < x < 3.54] is [2022-JUNE]

(a) $13\ e^{-3}$

(b) $9\ e^{-3}$

(c) $4\ e^{-2}$

(d) e^{-6}

a

Que. The average number of advertisements per page appearing in a newspaper is 3. What is the probability that in a particular page zero number of advertisements are there? [2021-DEC]

(a) e^{-3}

(b) e^{0}

(c) e^{+3}

(d) e^{-1}

a

5 MARKS

CHAPTER 17

CORRELATION AND REGRESSION

BY: SHIVANI SHARMA

BIVARIATE DATA

Example Prepare a Bivariate Frequency table for the following data relating to the marks in Statistics (x) and Mathematics (y):

(15, 13),	(1, 3),	(2, 6),	(8, 3),	(15, 10),	(3, 9),	(13, 19),
(10, 11),	(6, 4),	(18, 14),	(10, 19),	(12, 8),	(11, 14),	(13, 16),
(17, 15),	(18, 18),	(11, 7),	(10, 14),	(14, 16),	(16, 15),	(7, 11),
(5, 1),	(11, 15),	(9, 4),	(10, 15),	(13, 12)	(14, 17),	(10, 11),
(6, 9),	(13, 17),	(16, 15),	(6, 4),	(4, 8),	(8, 11),	(9, 12),
(14, 11),	(16, 15),	(9, 10),	(4, 6),	(5, 7),	(3, 11),	(4, 16),
(5, 8),	(6, 9),	(7, 12),	(15, 6),	(18, 11),	(18, 19),	(17, 16)
(10, 14)						

- When data are collected on two variables simultaneously, they are known as bivariate data and the corresponding frequency distribution, derived from it, is known as Bivariate Frequency Distribution.

Bivariate Frequency Distribution of Marks in Statistics and Mathematics.

		MARKS IN MATHS					
X \ Y		0-4	4-8	8-12	12-16	16-20	Total
MARKS IN STATS	0–4	I (1)	I (1)	II (2)			4
	4–8	I (1)	IIII (4)	~~IIII~~ (5)	I (1)	I (1)	12
	8–12	I (1)	II (2)	IIII (4)	~~IIII~~ I (6)	I (1)	14
	12–16		I (1)	III (3)	II (2)	~~IIII~~ (5)	11
	16–20			I (1)	~~IIII~~ (5)	III (3)	9
	Total	3	8	15	14	10	50

- No . of cells = m x n

  where ,

  m = no. of class interval of x

  n = no. of class interval of y

MARGINAL DISTRIBUTION

Marginal Distribution of Marks in Statistics

Marks	No. of Students
0-4	4
4-8	12
8-12	14
12-16	11
16-20	9
Total	50

Marginal Distribution of Marks in Mathematics

- No . of Marginal Distributions in Bivariate data = 2

CONDITIONAL DISTRIBUTION

Conditional Distribution of Marks in Statistics for Students having Mathematics Marks between 8 to 12

Marks	No. of Students
0-4	2
4-8	5
8-12	4
12-16	3
16-20	1
Total	15

- No . of Conditional Distributions = m +n

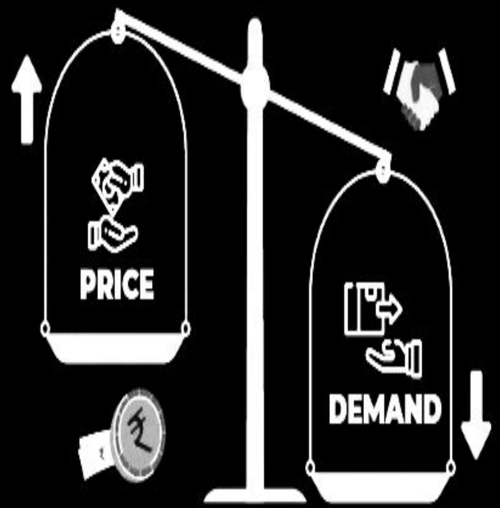
  where ,

  m = no. of class interval of x

  n = no. of class interval of y

Correlation

- In a bivariate data , if change in one variable causes change in another variable either directly or inversely, then the two variables are known to be associated or correlated.



TYPES OF CORRELATION

POSITIVE CORRELATION

- If two variables move in the same direction i.e. an increase (or decrease) on the part of one variable introduces an increase (or decrease) on the part of the other variable, then the two variables are known to be positively correlated.

- As for example, yield and rainfall, are positively correlated.

NEGATIVE CORRELATION

- if the two variables move in the opposite directions i.e. an increase (or a decrease) on the part of one variable results a decrease (or an increase) on the part of the other variable, then the two variables are known to have a negative correlation.

- As for example , the price and demand of an item, is negative correlation.

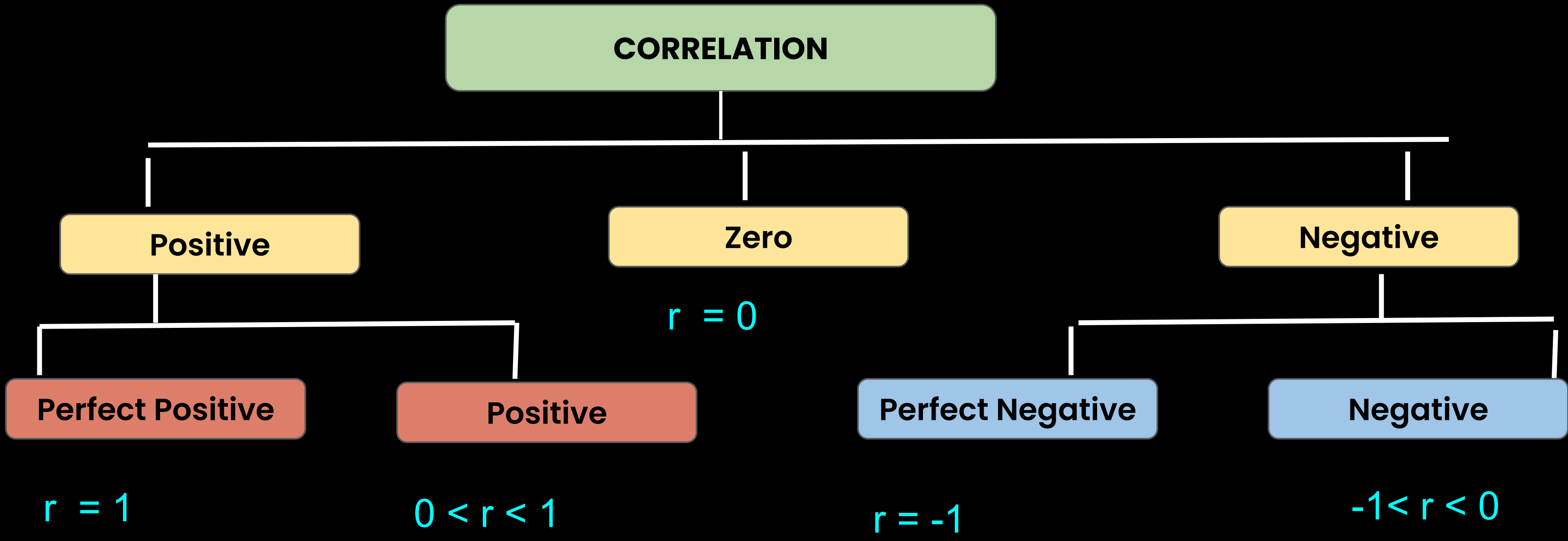
- As an example, there could be a positive correlation between production of rice and that of iron in India for the last twenty years due to the effect of a third variable time on both these variables. It is necessary to eliminate the influence of the third variable before computing correlation between the two original variables.

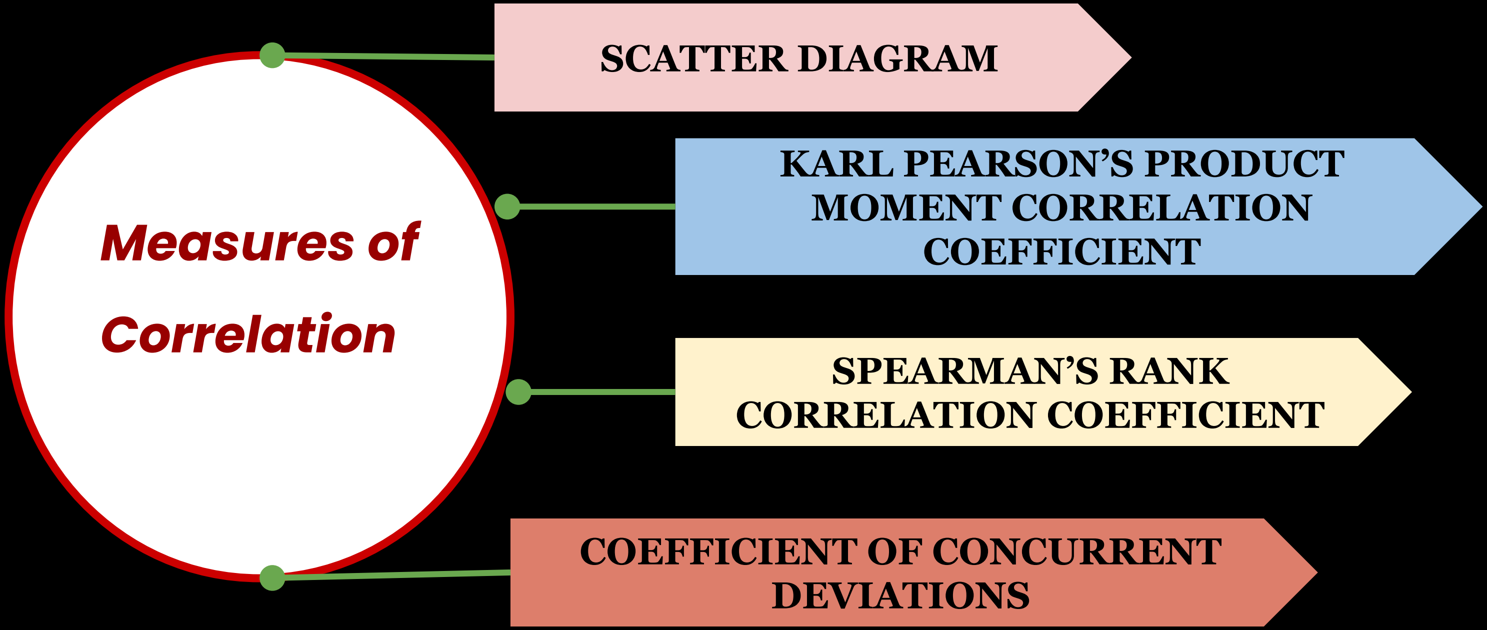
SPURIOUS CORRELATION

There are some cases when we may find a correlation between two variables although the two variables are not causally related. This is due to the existence of a third variable which is related to both the variables under consideration. Such a correlation is known as spurious correlation or nonsense correlation.

Correlation

- Correlation is expressed using r
- The value of correlation ranges from - 1 to 1 , both inclusive $-1 \leq r \leq 1$.




Measures of Correlation
SCATTER DIAGRAM
KARL PEARSON’S PRODUCT MOMENT CORRELATION COEFFICIENT
SPEARMAN’S RANK CORRELATION COEFFICIENT
COEFFICIENT OF CONCURRENT DEVIATIONS

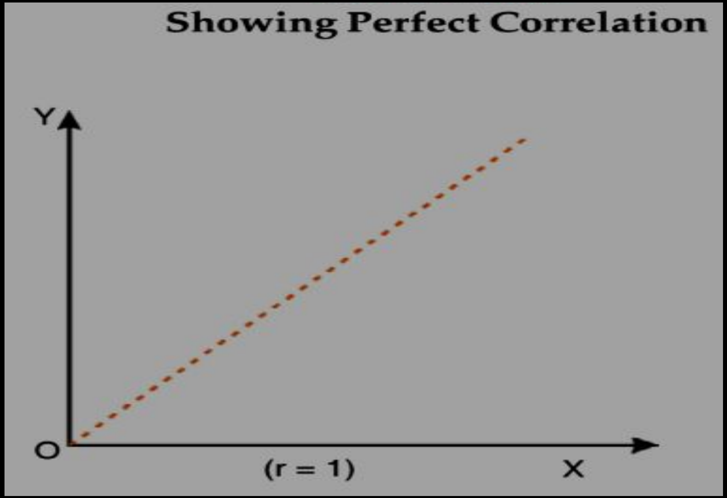
SCATTER DIAGRAM

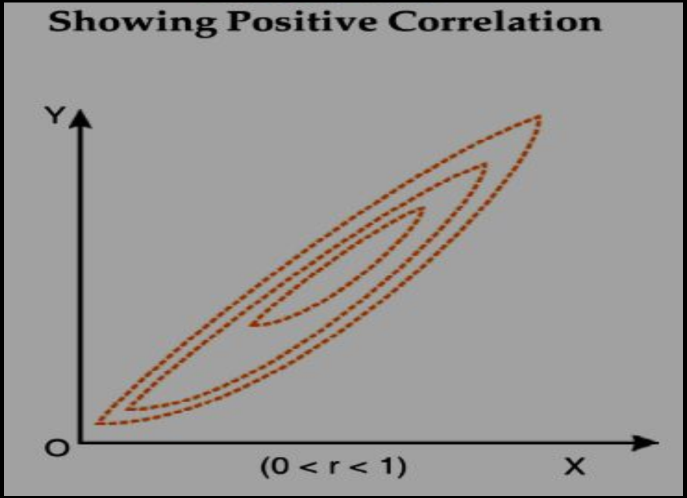
- This is a simple diagrammatic method to establish correlation between a pair of variables.

- scatter diagram can be applied for any type of correlation –linear as well as non-linear i.e. curvilinear.

- Scatter diagram can distinguish between different types of correlation although it fails to measure the extent of relationship between the variables.

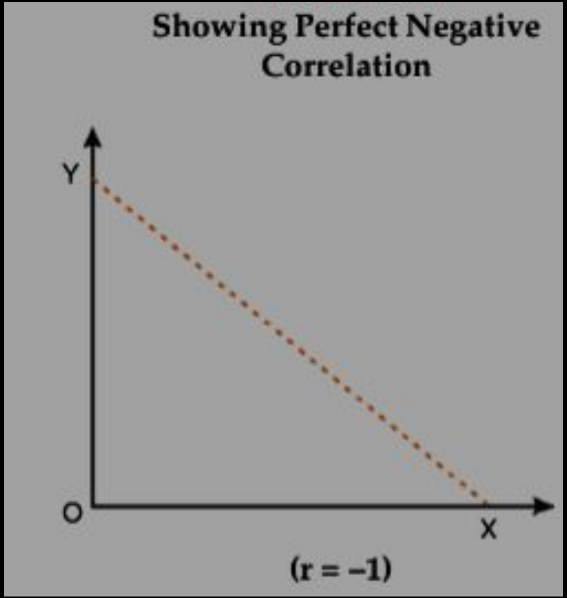
SCATTER DIAGRAM

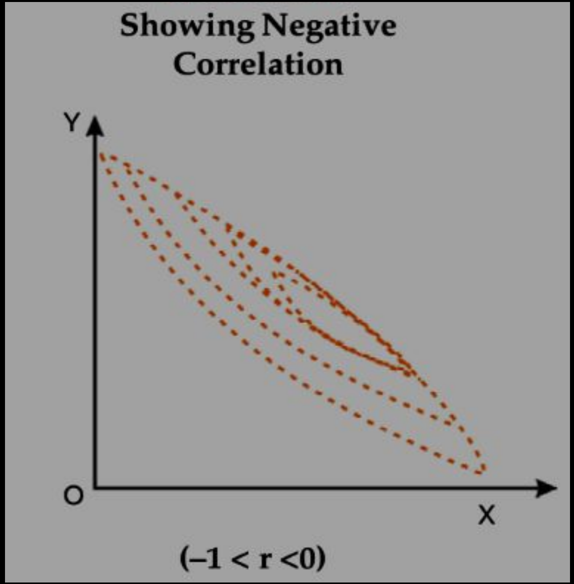




- The plotted points lie from lower left corner to upper right corner

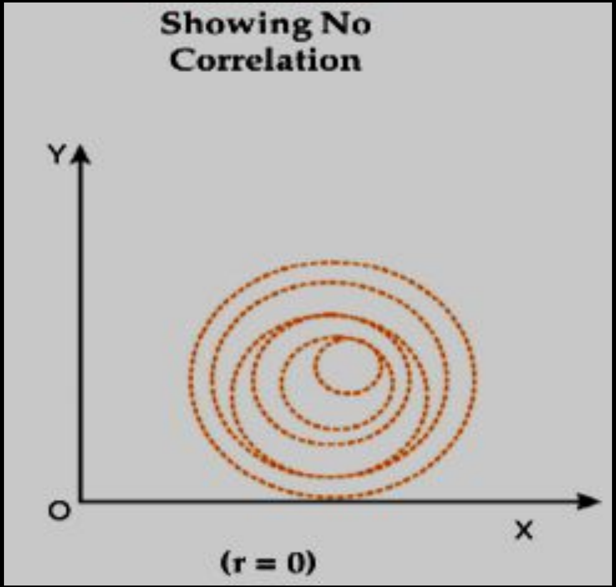
SCATTER DIAGRAM

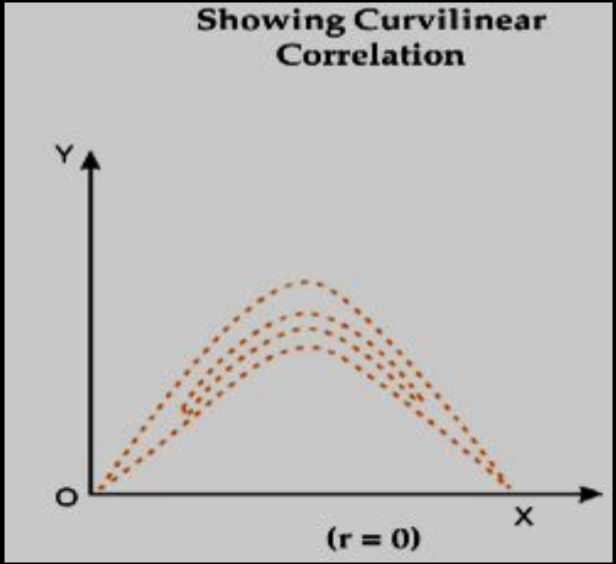




- **The plotted points concentrate from upper left to lower right**

SCATTER DIAGRAM





- **The plotted points would be equally distributed without depicting any particular pattern.**

KARL PEARSON’S PRODUCT MOMENT CORRELATION COEFFICIENT

- This is by for the best method for finding correlation between two variables provided the relationship between the two variables is linear

KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

- $$r = r_{xy} = \frac{\text{Cov}(x, y)}{S_x \times S_y} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

where

$$\text{cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x}\,\bar{y} \ldots$$

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

KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

- and $S_y = \sqrt{\dfrac{\Sigma\left(y_i - \bar{y}\right)^2}{n}} = \sqrt{\dfrac{\Sigma y_i^2}{n} - \bar{y}^2}$ ..

A single formula for computing correlation coefficient is given by

$$r = \frac{n\Sigma x_i y_i - \Sigma x_i \times \Sigma y_i}{\sqrt{n\Sigma x_i^2 - \left(\Sigma x_i\right)^2}\sqrt{n\Sigma y_i^2 - \left(\sum y_i\right)^2}}$$

PROPERTIES OF CORRELATION COEFFICIENT

(i) The Coefficient of Correlation is a unit-free measure.

(ii) The coefficient of correlation always lies between –1 and 1, including both the limiting values $-1 \le r \le 1$.

PROPERTIES OF CORRELATION COEFFICIENT

(iii) If two variables are related by a linear equation , then correlation coefficient will always be perfect +1 or -1 depends on the sign of slope of equation .

PROPERTIES OF CORRELATION COEFFICIENT

- Change of Origin : NO Impact
- Change of Scale : No Impact of value but affected by sign

Original		Change
x	$\longrightarrow$	u
y	$\longrightarrow$	v
r_{xy}	$\longrightarrow$	r_{uv}

- If sign of both change of scale are same

$$r_{uv} = r_{xy}$$

- If sign of both change of scale are different

$$r_{uv} = -r_{xy}$$

SPEARMAN'S RANK CORRELATION COEFFICIENT

- When we need finding correlation between two qualitative characteristics, say, beauty and intelligence, we take recourse to using rank correlation coefficient.

- Rank correlation can also be applied to find the level of agreement (or disagreement) between two judges so far as assessing a qualitative characteristic is concerned..

SPEARMAN'S RANK CORRELATION COEFFICIENT

- Spearman's rank correlation coefficient is given by

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

where r_R denotes rank correlation coefficient and it lies between –1 and 1 inclusive of these two values.

$d_i = x_i - y_i$ represents the difference in ranks for the i-th individual and n denotes the number of individuals.

COEFFICIENT OF CONCURRENT DEVIATIONS

- A very simple and casual method of finding correlation when we are not serious about the magnitude of the two variables .

$$r_c = \pm\sqrt{\pm\frac{(2c - m)}{m}}$$

where c is the number of concurrent deviations (same direction)

m is number of pairs compared , m = n-1

COEFFICIENT OF CONCURRENT DEVIATIONS

- **If $(2c-m) > 0$, then we take the positive sign both inside and outside the radical sign and if $(2c-m) < 0$, we are to consider the negative sign both inside and outside the radical sign.**

- **Like Pearson's correlation coefficient and Spearman's rank correlation coefficient, the coefficient of concurrent deviations also lies between –1 and 1, both inclusive.**

REGRESSION ANALYSIS

- In regression analysis, we are concerned with the estimation of one variable for a given value of another variable on the basis of an average mathematical relationship between the two variables.

Estimation of Y when X is given → **Y on X**

Y : Dependent

X : Independent

$$y = a + bx$$

Estimation of X when Y is given → **X on Y**

X : Dependent

Y : Independent

$$x = a + by$$

REGRESSION

Estimation of Y when X is given

Regression line of Y on X

$$Y - \overline{Y} = b_{yx}(X - \overline{X})$$

Estimation of X when Y is given

Regression line of X on Y

$$X - \overline{X} = b_{xy}(Y - \overline{Y})$$

METHOD OF LEAST SQUARES

REGRESSION COEFFICIENT

Regression Coefficient of Y on X

$$b_{yx} = \frac{Cov(x,y)}{\text{Var of } x}$$

$$b_{yx} = r \cdot \frac{SD_y}{SD_x}$$

REGRESSION COEFFICIENT

Regression Coefficient of X on Y

$$b_{xy} = \frac{Cov(x,y)}{\text{Var of } y}$$

$$b_{xy} = r \cdot \frac{SD_x}{SD_y}$$

Example If the relationship between two variables x and u is $u + 3x = 10$ and between two other variables y and v is

$2y + 5v = 25$, and the regression coefficient of y on x is known as 0.80, what would be the regression coefficient of v on u?

The regression coefficients remain unchanged due to a shift of origin but change due to a shift of scale.

PROPERTIES REGRESSION LINES / COEFFICIENTS

(ii) The two lines of regression intersect at the point $(\bar{x}, \bar{y})$ mean where x and y are the variables under consideration.

According to this property, the point of intersection of the regression line of y on x and the regression line of x on y is $(\bar{x}, \bar{y})$ i.e. the solution of the simultaneous equations in x and y.

PROPERTIES REGRESSION LINES / COEFFICIENTS

(iii) The coefficient of correlation between two variables x and y is the simple geometric mean of the two regression coefficients. The sign of the correlation coefficient would be the common sign of the two regression coefficients.

$$r = \pm \sqrt{b_{yx} \times b_{xy}}$$

If both the regression coefficients are negative, r would be negative and if both are positive, r would assume a positive value.

NOTE

- Product of the regression coefficient must be numerically less than unity .
- This can be applied, unlike correlation for any type of relationship linear as well as curvilinear.
- The two lines of regression coincide i.e. become identical when $r = -1$ or 1 or in other words, there is a perfect negative or positive correlation between the two variables under discussion.
- If $r = 0$ Regression lines are perpendicular to each other

Coefficient of Determination / Explained Variance / Accounted Variance

- Correlation coefficient measuring a linear relationship between the two variables indicates the amount of variation of one variable accounted for by the other variable.

$$r^2$$

Coefficient of Non- Determination / Unexplained Variance / Unaccounted Variance

- **The 'coefficient of non-determination' is given by $(1-r^2)$ and can be interpreted as the ratio of unexplained variance to the total variance.**

$$\boxed{\text{Coefficient of non-determination} = (1-r^2)}$$

Que. If the relationship between two variables u and v are given by $2u + v + 7 = 0$ and if the AM of u is 10 then the AM of v is

(a) 17

(b) -17

(c) -27

(d) None of these

Ans c

If cov(x, y) = 15, what restrictions should be put for the standard deviations of x and y?

(a) No restriction.

(b) The product of the standard deviations should be more than 15.

(c) The product of the standard deviations should be less than 15.

(d) The sum of the standard deviations should be less than 15.

Ans b

Que If the covariance between two variables is 20 and the variance of one of the variables is 16, what would be the variance of the other variable?

(a) $S^2_y \geq 25$

(b) More than 10

(c) Less than 10

(d) More than 1.25

Ans a

Que If $u + 5x = 6$ and $3y - 7v = 20$ and the correlation coefficient between x and y is 0.58 then what would be the correlation coefficient between u and v?

(a) 0.58

(b) −0.58

(c) −0.84

(d) 0.84

Ans b

Que From the following data

x:	2	3	5	4	7
y:	4	6	7	8	10

The coefficient of correlation was found to be 0.93. What is the correlation between u and v as given below?

u:	–3	–2	0	–1	2
v:	–4	–2	–1	0	2

(a) −0.93

(b) 0.93

(c) 0.57

(d) −0.57

Ans b

Que If the relationship between two variables x and y is given by $2x + 3y + 4 = 0$, then the value of the correlation coefficient between x and y is

(a) 0

(b) 1

(c) −1

(d) negative.

Ans C

For 10 pairs of observations, No. of concurrent deviations was found to be 4. What is the value of the coefficient of concurrent deviation?

(a) $\sqrt{0.2}$

(b) $\sqrt{-0.2}$

(c) 1/3

(d) −1/3

Ans d

The two lines of regression become identical when

(a) $r = 1$

(b) $r = -1$

(c) $r = 0$

(d) (a) or (b).

Ans d

What are the limits of the correlation coefficient?

(a) No limit

(b) −1 and 1, excluding the limits

(c) 0 and 1, including the limits

(d) −1 and 1, including the limits

Ans d

What are the limits of the two regression coefficients?

(a) No limit

(b) Must be positive

(c) One positive and the other negative

(d) Product of the regression coefficient must be numerically less than unity.

Ans d

If the regression line of y on x and of x on y are given by $2x + 3y = -1$ and $5x + 6y = -1$ then the arithmetic means of x and y are given by

(a) $(1, -1)$

(b) $(-1, 1)$

(c) $(-1, -1)$

d) $(2, 3)$

Ans a

Given the regression equations as $3x + y = 13$ and $2x + 5y = 20$, which one is the regression equation of y on x?

(a) 1st equation

(b) 2nd equation

(c) both (a) and (b)

(d) none of these.

Ans b

Given the following equations: $2x - 3y = 10$ and $3x + 4y = 15$, which one is the regression equation of x on y ?

(a) 1st equation

(b) 2nd equation

(c) both the equations

(d) none of these

Ans d

USE MY CODE : SS12

If $u = 2x + 5$ and $v = -3y - 6$ and regression coefficient of y on x is 2.4, what is the regression coefficient of v on u?

(a) 3.6

(b) −3.6

(c) 2.4

(d) −2.4

Ans b

If $4y - 5x = 15$ is the regression line of y on x and the coefficient of correlation between x and y is 0.75, what is the value of the regression coefficient of x on y?

(a) 0.45

(b) 0.9375

(c) 0.6

(d) none of these

Ans a

If the regression line of y on x and that of x on y are given by $y = -2x + 3$ and $8x = -y + 3$ respectively, what is the coefficient of correlation between x and y?

(a) 0.5

(b) $-1/\sqrt{2}$

(c) -0.5

(d) none of these

Ans c

INDEX NUMBERS

INDEX NUMBERS

- *An index number is a ratio of two or more time periods , one of which is the base time period. The value at the base time period serves as the standard point of comparison.*

- **Example:**
- **WPI**
- **CPI**
- **NIFTY**

IMPORTANT ISSUES IN INDEX CREATION

- ***Selection of data:*** It is important to understand the purpose for which the index is used. If it is used for purposes of knowing the cost of living, there is no need of including the prices of capital goods which do not directly influence the living.
- Index numbers are often constructed from the sample. It is necessary to ensure that it is representative. Random sampling, and if need be, a stratified random sampling can ensure this.
- It is also necessary to ensure comparability of data.

IMPORTANT ISSUES IN INDEX CREATION

- ***Base Period:*** **It is a point of reference in comparing various data.**
- **The period should be normal i.e., one of the relative stability, not affected by extraordinary events like war, famine, etc.**
- **It should be relatively recent because we are more concerned with the changes with reference to the present and not with the distant past.**

IMPORTANT ISSUES IN INDEX CREATION

- ***Selection of Weights:*** **It is necessary to point out that each variable involved in composite index should have a reasonable influence on the index, i.e., due consideration should be given to the relative importance of each variable which relates to the purpose for which the index is to be used.**
- **For example, in the computation of cost of living index, sugar cannot be given the same importance as the cereals.**

IMPORTANT ISSUES IN INDEX CREATION

Use of Averages:

- **Average plays an important role in computing index numbers. The geometric mean is better in averaging relatives, but for most of the indices arithmetic mean is used because of its simplicity.**

TYPES OF Index Numbers

Price Index Numbers

Quantity Index Numbers

Value Index Numbers

RELATIVES

PRICE RELATIVES

QUANTITY RELATIVES

VALUE RELATIVES

$$\text{Price relative} = \frac{P_n}{P_o}$$

$$\text{Quantity relative} = \frac{Q_n}{Q_o}$$

$$\text{Value relative} = \frac{V_n}{V_o} = \frac{P_n Q_n}{P_o Q_o} = \left(\frac{P_n}{P_o} \times \frac{Q_n}{Q_o}\right)$$

INDEX NUMBERS

- **Relatives:** One of the simplest examples of an index number is a price relative, which is the ratio of the price of single commodity in a given period to its price in another period called the base period or the reference period. It can be indicated as follows:

$$\text{Price relative} = \frac{P_n}{P_o}$$

It has to be expressed as a percentage, it is multiplied by 100

$$\text{Price relative} = \frac{P_n}{P_o} \times 100$$

INDEX NUMBERS

$$\text{Quantity relative} = \frac{Q_n}{Q_o}$$

Similarly, there are value relatives:

$$\text{Value relative} = \frac{V_n}{V_o} = \frac{P_n Q_n}{P_o Q_o} = \left(\frac{P_n}{P_o} \times \frac{Q_n}{Q_o}\right)$$

LINK RELATIVES

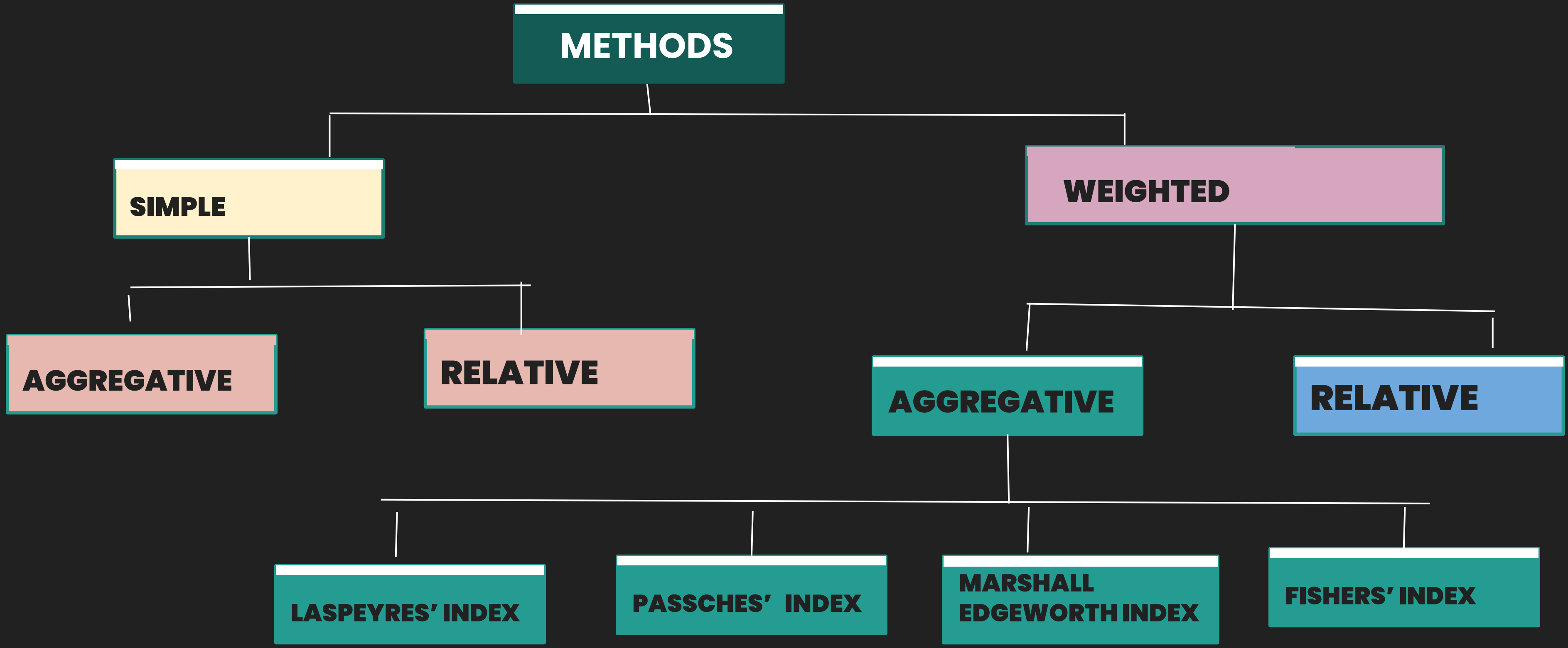
When successive prices or quantities are taken, the relatives are called the link relative,

$$\frac{P_1}{P_o}, \frac{P_2}{P_1}, \frac{P_3}{P_2}, \frac{P_n}{P_{n-1}}$$

CHAIN RELATIVES

When the above relatives are in respect to a fixed base period these are also called the chain relatives with respect to this base or the relatives chained to the fixed base. They are in the form of

$$\frac{P_1}{P_o}, \frac{P_2}{P_o}, \frac{P_3}{P_o}, \frac{P_n}{P_o}$$


METHODS
SIMPLE
WEIGHTED
AGGREGATIVE
RELATIVE
AGGREGATIVE
RELATIVE
LASPEYRES' INDEX
PASSCHES' INDEX
MARSHALL EDGEWORTH INDEX
FISHERS' INDEX

INDEX NUMBERS

SIMPLE AGGREGATIVE METHOD

- In this method of computing a price index, we express the total of commodity prices in a given year as a percentage of total commodity price in the base year. In symbols, we have

$$\text{Simple aggregative price index} = \frac{\Sigma P_n}{\Sigma P_o} \times 100$$

- where P_n is the sum of all commodity prices in the current year and P_o is the sum of all commodity prices in the base year.

INDEX NUMBERS

SIMPLE AGGREGATIVE METHOD

Commodities	1998	1999	2000
Cheese (per 100 gms)	12.00	15.00	15.60
Egg (per piece)	3.00	3.60	3.30
Potato (per kg)	5.00	6.00	5.70
Aggregrate	20.00	24.60	24.60
Index	100	123	123

Simple Aggregative Index for 1999 over 1998 = $\dfrac{\Sigma P_n}{\Sigma P_o} = \dfrac{24.60}{20.00} \times 100 = 123$

and for 2000 over 1998 = $\dfrac{\Sigma P_n}{\Sigma P_o} \times 100 = \dfrac{24.60}{20.00} \times 100 = 123$

INDEX NUMBERS

SIMPLE AGGREGATIVE METHOD

MERITS: Easy to understand

DEMERITS:

- Commodity with higher price will have greater influence in index value.
- If units are changed then the Index numbers will also change.
- Price quotations become the concealed weights which have no logical significance.

INDEX NUMBERS

SIMPLE AVERAGE OF RELATIVES

- Under this method we invert the actual price for each variable into percentage of the base period. These percentages are called relatives . The index number is the average of all such relatives.

$$\frac{\sum\left(\frac{P_n}{P_o}\times 100\right)}{N}$$

INDEX NUMBERS

SIMPLE AVERAGE OF RELATIVES

Commodities	1998	1999	2000
Cheese (per 100 gms)	12.00	15.00	15.60
Egg (per piece)	3.00	3.60	3.30
Potato (per kg)	5.00	6.00	5.70
Aggregrate	20.00	24.60	24.60
Index	100	123	123

Commodities	1998	1999	2000
A	100.0	125.0	130.0
B	100.0	120.0	110.0
C	100.0	120.0	114.0
Aggregate	300.0	365.0	354.0
Index	100.0	121.67	118.0

INDEX NUMBERS

SIMPLE AVERAGE OF RELATIVES

MERITS:

- One big advantage of price relatives is that they are pure numbers.
- Price index number computed from relatives will remain the same regardless of the units by which the prices are quoted

DEMERITS:

In Spite of some improvement, the above method has a flaw that it gives equal importance to each of the relatives

- This defect can be remedied by the introduction of an appropriate weighing system.

INDEX NUMBERS

WEIGHTED AGGREGATIVE INDEX

Under this method we weigh the price of each commodity by a suitable factor often taken as the quantity or value weight sold during the base year or the given year or an average of some years.

INDEX NUMBERS

WEIGHTED AGGREGATIVE INDEX

a. **Laspeyres' Index**: In this Index base year quantities are used as weights:

$$\text{Laspeyres Index} = \frac{\Sigma P_n Q_0}{\Sigma P_0 Q_0} \times 100$$

WEIGHTED AGGREGATIVE INDEX

Paasche's Index: In this Index current year quantities are used as weights:

$$\text{Passche's Index} = \frac{\Sigma P_n Q_n}{\Sigma P_o Q_n} \times 100$$

INDEX NUMBERS

WEIGHTED AGGREGATIVE INDEX

The Marshall-Edgeworth index uses this method by taking the average of the base year and the current year

$$\text{Marshall-Edgeworth Index} = \frac{\Sigma P_n (Q_o + Q_n)}{\Sigma P_o (Q_o + Q_n)} \times 100$$

INDEX NUMBERS

WEIGHTED AGGREGATIVE INDEX

d. **Fisher's ideal Price Index:** This index is the geometric mean of Laspeyres' and Paasche's.

$$\text{Fisher's Index} = \sqrt{\frac{\Sigma P_n Q_o}{\Sigma P_o Q_o} \times \frac{\Sigma P_n Q_n}{\Sigma P_o Q_n}} \times 100$$

INDEX NUMBERS

WEIGHTED AGGREGATIVE INDEX

BOWLEY INDEX:

$$\frac{\text{Laspeyres' Index} + \text{Paasche's Index}}{2}$$

INDEX NUMBERS

WEIGHTED AVERAGE OF RELATIVE METHOD

- To overcome the disadvantage of a simple average of relative method, we can use weighted average of relative method.
- Generally weighted arithmetic mean is used although the weighted geometric mean can also be used.
- The weighted arithmetic mean of price relatives using base year value weights is represented by

$$\frac{\Sigma \frac{P_n}{P_o} \times (P_o Q_o)}{\Sigma P_o Q_o} \times 100 = \frac{\Sigma P_n Q_o}{\Sigma P_o Q_o} \times 100$$

Same as Laspeyres' Index

CHAIN INDEX NUMBERS

Year (1)	*Price* (2)	*Link Relatives* (3)	*Chain Indices* (4)
1991	50	100	100
1992	60	$\frac{60}{50} \times 100 = 120.0$	$\frac{120}{100} \times 100 = 120.0$
1993	62	$\frac{62}{60} \times 100 = 103.3$	$\frac{103.3}{100} \times 120 = 124.0$
1994	65	$\frac{65}{62} \times 100 = 104.8$	$\frac{104.8}{100} \times 124 = 129.9$
1995	70	$\frac{70}{65} \times 100 = 107.7$	$\frac{107.7}{100} \times 129.9 = 139.9$
1996	78	$\frac{78}{70} \times 100 = 111.4$	$\frac{111.4}{100} \times 139.9 = 155.8$
1997	82	$\frac{82}{78} \times 100 = 105.1$	$\frac{105.1}{100} \times 155.8 = 163.7$
1998	84	$\frac{84}{82} \times 100 = 102.4$	$\frac{102.4}{100} \times 163.7 = 167.7$
1999	88	$\frac{88}{84} \times 100 = 104.8$	$\frac{104.8}{100} \times 167.7 = 175.7$
2000	90	$\frac{90}{88} \times 100 = 102.3$	$\frac{102.3}{100} \times 175.7 = 179.7$

$$\text{Chain Index} = \frac{\text{Link relative of current year} \times \text{Chain Index of the previous year}}{100}$$

QUANTITY INDEX NUMBERS

- To measure and compare prices, we use price index numbers.
- When we want to measure and compare quantities, we resort to Quantity Index Numbers.
- Though price indices are widely used to measure the economic strength, Quantity indices are used as indicators of the level of output in economy.
- To construct Quantity indices, we measure changes in quantities and weight them using prices or values as weights.

QUANTITY INDEX NUMBERS

- 1. Simple aggregate of quantities:

  This has the formula $\frac{\Sigma Q_n}{\Sigma Q_o}$

  2. The simple average of quantity relatives:

  This can be expressed by the formula $\frac{\frac{\Sigma Q_n}{\Sigma Q_o}}{N}$

QUANTITY INDEX NUMBERS

- 3. Weighted aggregate Quantity indices:

  (i) With base year weight : $\frac{\Sigma Q_n P_o}{\Sigma Q_o P_o}$ (Laspeyre's index)

  (ii) With current year weight : $\frac{\Sigma Q_n P_n}{\Sigma Q_o P_n}$ (Paasche's index)

  (iii) Geometric mean of (i) and (ii) : $\sqrt{\frac{\Sigma Q_n P_o}{\Sigma Q_o P_o} \times \frac{\Sigma Q_n P_n}{\Sigma Q_o P_n}}$ (Fisher's Ideal)

VALUE INDEX NUMBERS

- Value equals price multiplied by quantity. Thus a value index equals the total sum of the values of a given year divided by the sum of the values of the base year, i.e.,

$$\frac{\Sigma V_n}{\Sigma V_o} = \frac{\Sigma P_n Q_n}{\Sigma P_0 Q_0}$$

DEFLATING TIME SERIES USING INDEX NUMBERS

- $$\text{Deflated Value} = \frac{\text{Current Value}}{\text{Price Index of the current year}}$$

$$\text{or Current Value} \times \frac{\text{Base Price } (P_0)}{\text{Current Price } (P_n)}$$

$$\text{Real Wages} = \frac{\text{Actual wages}}{\text{Cost of living Index}} \times 100$$

SHIFTING PRICE INDEX

$$\text{Shifted Price Index} = \frac{\text{Original Price Index}}{\text{Price Index of the year on which it has to be shifted}} \times 100$$

Shifted Price Index

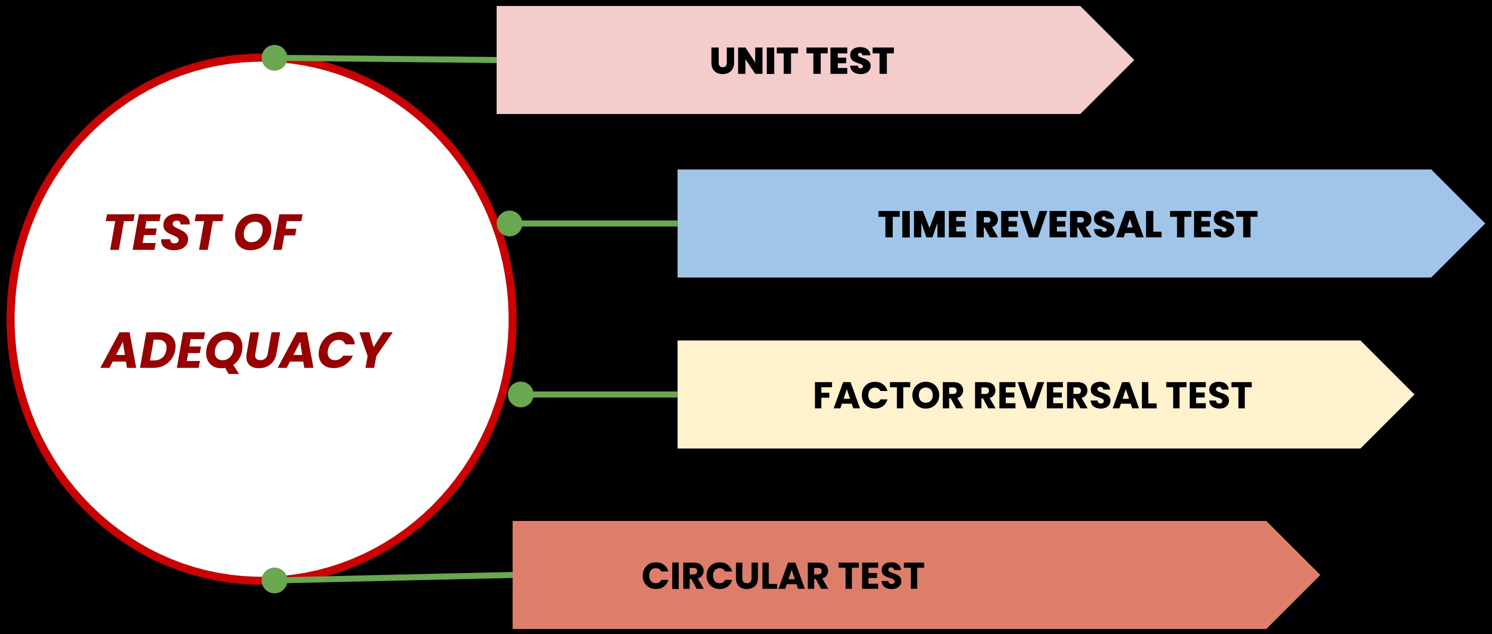
Year	*Original Price Index*	*Shifted Price Index to base 1990*
1980	100	71.4
1981	104	74.3
1982	106	75.7
1983	107	76.4
1984	110	78.6
1985	112	80.0
1986	115	82.1
1987	117	83.6
1988	125	89.3
1989	131	93.6
1990	140	100.0
1991	147	105.0

SPLICING TWO INDEX SERIES

- **two index covering different bases may be combined into single series by splicing.**

Splicing Two Index Number Series

Year	*Old Price Index [1990 = 100]*	*Revised Price Index [1995 = 100]*	*Spliced Price Index [1995 = 100]*
1990	100.0		87.6
1991	102.3		89.6
1992	105.3		92.2
1993	107.6		94.2
1994	111.9		98.0
1995	114.2	100.0	100.0
1996		102.5	102.5
1997		106.4	106.4
1998		108.3	108.3
1999		111.7	111.7
2000		117.8	117.8


TEST OF
ADEQUACY
UNIT TEST
TIME REVERSAL TEST
FACTOR REVERSAL TEST
CIRCULAR TEST

TEST OF ADEQUACY

UNIT TEST

i. This test requires that the formula should be independent of the unit in which or for which prices and quantities are quoted.
ii. Except for the simple (unweighted) aggregative index all other formulae satisfy this test.

TEST OF ADEQUACY

TIME REVERSAL TEST

- It is a test to determine whether a given method will work both ways in time, forward and backward.
- The test provides that the formula for calculating the index number should be such that two ratios, the current on the base and the base on the current should multiply into unity.
- In other words, the two indices should be reciprocals of each other.

Symbolically,

$$P_{01} \times P_{10} = 1$$

TEST OF ADEQUACY

TIME REVERSAL TEST

- where P_{01} is the index for time 1 on 0 and P_{10} is the index for time 0 on 1.
- Laspeyres' method and Paasche's method do not satisfy this test, but Fisher's Ideal Formula does.

TEST OF ADEQUACY

FACTOR REVERSAL TEST

- This holds when the product of price index and the quantity index should be equal to the corresponding value index,

$$P_{01} \times Q_{01} = V_{01}$$

- Fisher’s Index satisfies Factor Reversal test

TEST OF ADEQUACY

NOTE

- While selecting an appropriate index formula, the Time Reversal Test and the Factor Reversal test are considered necessary in testing the consistency.
- Because Fisher's Index number satisfies both the tests in , it is called an Ideal Index Number.

TEST OF ADEQUACY

CIRCULAR TEST

- As per this test , $P_{01} \times P_{12} \times P_{20} = 1$
- It is concerned with the measurement of price changes over a period of years, when it is desirable to shift the base.
- This property therefore enables us to adjust the index values from period to period without referring each time to the original base.
- The test of this shiftability of base is called the circular test.
- *This test is not met by Laspeyres, or Paasche's or the Fisher's ideal index.*
- *The simple geometric mean of price relatives and the weighted aggregative with fixed weights meet this test.*

Que. If the index numbers of prices at a place in 1994 is 250 with 1984 as base year then the prices have increase on average by .

a) 250%

b) 150%

c) 350%

d) none

b

Que. If the prices of all commodities in a place have increased 1.25 times in comparison to the base period , the index number of prices of that place now is .

a) 125

b) 150

c)225

d) none

c

30. In 1980, the net monthly income of the employee was ₹ 800/- p. m. The consumer price index number was 160 in 1980. It rises to 200 in 1984. If he has to be rightly compensated. The additional D. A. to be paid to the employee is

(a) ₹ 175/- (b) ₹ 185/- (c) ₹ 200/- (d) ₹ 125.

c

35. With the base year 1960 the C. L. I. in 1972 stood at 250. x was getting a monthly Salary of ₹ 500 in 1960 and ₹ 750 in 1972. In 1972 to maintain his standard of living in 1960 x has to receive as extra allowances of

(a) ₹ 600/-   (b) ₹ 500/-   (c) ₹ 300/-   (d) none of these.

b

Que. If the prices of all commodities in a place have decreased 35% over the the base period prices , the index number of prices of that place now is .

a) 35

b) 135

c)65

d) none

c

EXERCISE- Set (A)

Choose the most appropriate option (a) (b) (c) or (d).

Que. 1 A series of numerical figures which show the relative position is called

a) index number

b) relative number

c) absolute number

d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 2 **Index number for the base period is always taken as**

a) 200

b) 50

c) 1

d) 100

d

Choose the most appropriate option (a) (b) (c) or (d).

Que. 3 __________ play a very important part in the construction of index numbers.

a) weights

b) classes

c) estimations

d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 4 _________ is particularly suitable for the construction of index numbers.

a) H.M.

b) A.M.

c) G.M.

d) none

c

Choose the most appropriate option (a) (b) (c) or (d).

Que. 5 Index numbers show __________ changes rather than absolute amounts of change.

a) relative

b) percentage

c) both

d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 6 The _________ makes index numbers time-reversible.

a) A.M.

b) G.M.

c) H.M.

d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 7 Price relative is equal to

a) $\dfrac{\text{Price in the given year} \times 100}{\text{Price in the base year}}$

b) $\dfrac{\text{Price in the year base year} \times 100}{\text{Price in the given year}}$

c) Price in the given year × 100

d) Price in the base year × 100

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 8 **Index number is equal to**

a) sum of price relatives

b) average of the price relatives

c) product of price relative

d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 9 The _________ of group indices gives the General Index

a) H.M.

b) G.M.

c) A.M.

d) none

c

Choose the most appropriate option (a) (b) (c) or (d).

Que. 10 Circular Test is one of the tests of

a) index numbers

b) hypothesis

c) both

d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 11 ____________ is an extension of time reversal test

a) Factor Reversal test

b) Circular test

c) both

d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 12 Weighted G.M. of relative formula satisfy _________test

a) Time Reversal Test

b) Circular test

c) Factor Reversal Test

d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 13 Factor Reversal test is satisfied by

a) Fisher's Ideal Index

b) Laspeyres Index

c) Paasches Index

d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 14 Laspeyre's formula does not satisfy

a) Factor Reversal Test

b) Time Reversal Test

c) Circular Test

d) all the above

d

Choose the most appropriate option (a) (b) (c) or (d).

Que. 15 A ratio or an average of ratios expressed as a percentage is called

a) a relative number

b) an absolute number

c) an index number

d) none

C

Choose the most appropriate option (a) (b) (c) or (d).

Que. 16 The value at the base time period serves as the standard point of comparison

a) false

b) true

c) both

d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 17 An index time series is a list of _______ numbers for two or more periods of time

a) index

b) absolute

c) relative

d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 18 Index numbers are often constructed from the

a) frequency

b) class

c) sample

d) none

C

Choose the most appropriate option (a) (b) (c) or (d).

Que. 19 _________ is a point of reference in comparing various data describing individual behaviour.

a) Sample

b) Base period

c) Estimation

d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 20 The ratio of price of single commodity in a given period to its price in the preceding year price is called the

(a) base period

(b) price ratio

(c) relative price

(d) none

C

Choose the most appropriate option (a) (b) (c) or (d).

Que. 21 $\frac{\text{Sum of all commodity prices in the current year} \times 100}{\text{Sum of all commodity prices in the base year}}$ is

(a) Relative Price Index

(b) Simple Aggregative Price Index

(c) both

(d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 22 Chain index is equal to

(a) $\dfrac{\text{link relative of current year} \times \text{chain index of the current year}}{100}$

(b) $\dfrac{\text{link relative of previous year} \times \text{chain index of the current year}}{100}$

(c) $\dfrac{\text{link relative of current year} \times \text{chain index of the previous year}}{100}$

(d) $\dfrac{\text{link relative of previous year} \times \text{chain index of the previous year}}{100}$

c

Choose the most appropriate option (a) (b) (c) or (d).

Que. 23 P_{01} is the index for time

(a) 1 on 0

(b) 0 on 1

(c) 1 on 1

(d) 0 on 0

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 24 P_{10} is the index for time

(a) 1 on 0

(b) 0 on 1

(c) 1 on 1

(d) 0 on 0

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 25 When the product of price index and the quantity index is equal to the corresponding value index then the test that holds is

(a) Unit Test

(b) Time Reversal Test

(c) Factor Reversal Test

(d) none holds

c

Choose the most appropriate option (a) (b) (c) or (d).

Que. 26 **The formula should be independent of the unit in which or for which price and quantities are quoted in**

(a) Unit Test

(b) Time Reversal Test

(c) Factor Reversal Test

(d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 27 Laspeyre's method and Paasche's method do not satisfy

(a) Unit Test

(b) Time Reversal Test

(c) Factor Reversal Test

(d) b & c

d

Choose the most appropriate option (a) (b) (c) or (d).

Que. 28 The purpose determines the type of index number to use

(a) yes

(b) no

(c) may be

(d) may not be

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 29 **The index number is a special type of average**

(a) false

(b) true

(c) both

(d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 30 **The choice of suitable base period is at best temporary solution**

(a) true

(b) false

(c) both

(d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 31 Fisher's Ideal Formula for calculating index numbers satisfies the ________ tests

(a) Unit Test

(b) Factor Reversal Test

(c) both

(d) none

C

Choose the most appropriate option (a) (b) (c) or (d).

Que. 32 Fisher's Ideal Formula dose not satisfy __________ test

(a) Unit Test

(b) Circular Test

(c) Time Reversal Test

(d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 33 __________________ **satisfies circular test**

a) G.M. of price relatives or the weighted aggregate with fixed weights

b) A.M. of price relatives or the weighted aggregate with fixed weights

c) H.M. of price relatives or the weighted aggregate with fixed weights

d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 34 Laspeyre's and Paasche's method __________ time reversal test

(a) satisfy

(b) do not satisfy

(c) are

(d) are not b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 35 There is no such thing as unweighted index numbers

(a) false

(b) true

(c) both

(d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 36 Theoretically, G.M. is the best average in the construction of index numbers but in practice, mostly the A.M. is used

(a) false

(b) true

(c) both

(d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 37 Laspeyre’s or Paasche’s or the Fisher’s ideal index do not satisfy

(a) Time Reversal Test

(b) Unit Test

(c) Circular Test

(d) none

C

Choose the most appropriate option (a) (b) (c) or (d).

Que. 38 ____________ is concerned with the measurement of price changes over a period of years, when it is desirable to shift the base

(a) Unit Test

(b) Circular Test

(c) Time Reversal Test

(d) none

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 39 The test of shifting the base is called

(a) Unit Test

(b) Time Reversal Test

(c) Circular Test

(d) none

C

Choose the most appropriate option (a) (b) (c) or (d).

Que. 40 **The formula for conversion to current value**

(a) Deflated value = $\dfrac{\text{Price Index of the current year}}{\text{previous value}}$

(b) Deflated value = $\dfrac{\text{current value}}{\text{Price Index of the current year}}$

(c) Deflated value = $\dfrac{\text{Price Index of the previous year}}{\text{previous value}}$

(d) Deflated value = $\dfrac{\text{Price Index of the previous year}}{\text{previous value}}$

b

Choose the most appropriate option (a) (b) (c) or (d).

Que. 41 **Shifted price Index =** $\dfrac{\text{Original Price} \times 100}{\text{Price Index of the year on which it has to be shifted}}$

(a) True

(b) false

(c) both

(d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 42 **The number of test of Adequacy is**

(a) 2

(b) 5

(c) 3

(d) 4

d

Choose the most appropriate option (a) (b) (c) or (d).

Que. 43 **We use price index numbers**

(a) To measure and compare prices

(b) to measure prices

(c) to compare prices

(d) none

a

Choose the most appropriate option (a) (b) (c) or (d).

Que. 44 Simple aggregate of quantities is a type of

(a) Quantity control

(b) Quantity indices

(c) both

(d) none

b