

Concept of Statistical Analysis

STATATA

Section (A) —

- ① — ⑥
- ② — ⑥
- ③ — ⑨
- ④ — ⑥
- ⑤ — ③
- ⑥ — ④
- ⑦ — ⑨
- ⑧ — ⑨
- ⑨ — ④
- ⑩ — ⑥

Section - B

Statistics, Definition & Use

(B) 2

It refers to the subject of Scientific activity which deals with the theories and methods of Collection, analysis and interpretations of such data.

Biostatistics term used when tools of Statistics are applied to the data that is derived from biological sciences.

- It is aggregate of facts
- It is numerically expressed
- It is affected by multiple factors
- It must be related to some field of inquiry.
- Standard of accuracy should be maintained in statistics.

Use.

- It help in presentation of large quantity data in a simple & classified form.
- It gives the methods of comparison of data
- It enlarges individual mind.
- It helps in finding the conditions of relationship between variables
- It is very useful in varied fields like Bank, Commerce, Economics, Agriculture etc.
- It is useful in analysis of collected data
- It makes data for easily understandable
- Clear presentation & for forecasting it is important.
- To define what is normal population of variables.
- To findout Correlation between two variables.
- Comparison between variables it is used.
- Important for study of relation between variables
- By Table large data can be averaged in simple way.

Sec. 15

③ Central Tendency \Rightarrow

The word average denotes a representative of a whole set of observations. It is a single figure which describes the entire series of observations with their varying sizes. It is a typical value occupying a central position where some observations are larger and some are ~~others~~ smaller than it. Central Tendency is the central part of the distribution and therefore it is also called measure of Central Tendency. The value of Central Tendency is a middle value of data series which classify the series in two category first values are large than central value and second one values are smaller than others.

• "A measure of central tendency is a typical value around which other values are congregated." is Central Tendency

• The value of central Tendency which represents a whole series and is supposed to contain its characteristics.

\rightarrow Values of the data series are scattered around of central value. In a data series, there may be less or more variation among the data at this condition we need to observe any value which will capable to represent the value of whole data series, that value is called central value.

→ The common measure of central tendency are.

- ① Mean ② Median ③ Mode.

① Mean ⇒ In arithmetic average is known as mean in statistics. It is the value which is calculated by total sum of the items divided by No. of items in a data series.

Formula ⇒ $\frac{\sum X}{N}$

$\sum X$ = Total sum of items

N = No. of items.

⇒ 5, 10, 15, 20 mean = $\frac{50}{4} = 12.5$

② Median ⇒ Among a data group, if there are arranged in increasing or decreasing order then the middle value is called median.

⇒ 7, 2, 4, 6, 1, 5, 8,

Formula ⇒ $\frac{N+1}{2}$ N = No. of items.

→ 1, 2, 4, (5), 6, 7, 8

Median value

③ Mode ⇒ Maximum frequency of the items in a data series is called mode.

ex. ⇒ 1, 2, 4, 6, 8, 9, 10, 12, 10, 10, 2, 4, 10

No.	Frequency
1.	1
2.	2
4.	2
6.	1
8.	1
9.	1
10.	4
12.	1

Item 10 is appearing 4 times which is maximum so, Mode = 10.

Characteristics of Central Tendency:-

- # It should be rigidly defined.
- # It should be properly defined.
- # It should be based on all items.
- # Average should depend on each and every item of the series.
- # It should be easily understood.
- # It should not be unduly affected by the extreme value.
- # It should be easy to interpret.
- # It should be easily subjected to further mathematical calculations.

- ①
- ⇒ Sample is small unit of population or a group of items, which is important for data collection, for further analysis.
 - ⇒ To analyze or examine any individual among a large population its collection is required.
 - ⇒ Usually population size is large

Characters

- ① Large sample size is more accurate than small for representation of the population
 - ② Selected sample size is a balance between statistical variation
 - ③ Sample may be unit of a population
 - ④ Sampling techniques with minimum disturbance are more valued.
 - ⑤ Sample size may be small or large based on population size
 - ⑥ Sample is the best indicator of a group or population
 - ⑦ Fast & True information given by the sample under suitable sampling technique.
 - ⑧ After proper analysis of all collected samples it gives clear picture about the findings
 - ⑨ It is also important for forecasting of any events.
 - ⑩ It is a source as a small unit among the large population.
- ⑥

Sampling methods

⇒ There are many methods of sampling which are as following -

① Random sampling ⇒ ① No specific pattern of sampling.

② Each member of the population or group has an equal chance of selection

③ Each individual can be selected by this method


④ It is a purest form of sampling

② Systematic sampling ⇒ This is three types

① Point sys. sam. ⇒ Point formation is a specific distance than sampling



② Line sys. sam ⇒ Transect line formation between points than sampling is done

③ Area sys. sam. ⇒  specific direction sampling



③ Stratified sampling ⇒ It is used when the population or sampling frame is made up by sub groups or subsets

⇒ subsets should be of equal characters.

Types ← Systematic Stra. sam. ←
Random

④ Clustered Sampling

- In this method sub groups of a population are used as a sampling unit.
- Population divided into many groups/subgroups
- Subgroups are components of specific population
- Small clusters are used for random sampling for specific aim of study.
- all members of or individuals are sampled in this study.

⑤ Judgement Sampling ⇒ It is common probability method of sampling

- Selection of sample is based on judgement
ex Cloth selection ← Colour
Quality
Rate

⇒ other types are

- ① Snowball sampling
- ② Quota sampling
- ③ Convenience sampling

Scaling

- It is a unit for measurement of quantitatively variable items, and it is useful in data analysis.
- Scale is important for representation of data & graphs also.
- Scale may be small or large based on study & population size.
- Scale is important unit for statistical study.
- By selecting a suitable sampling is done that may be Simple, Random, Systematics, Clusters etc.
- Significance test is applied after proper sampling with suitable scales.
- For better Scientific or other study scale, sampling, designing of experiments are important.

Use ⇒ ① Scale provides standardization of individual items.

- ② It also focus on behavioural character
- ③ Useful for psychological Tests
- ④ for attitude examination
- ⑤ measurement of IQ
- ⑥ Important for clear presentation of data
- ⑦ It remove complexity
- ⑧ It make easy of measurement

Section B) (6)

① Poisson distribution

- It is derived by Simen denis poisson in 1837
- It is a discrete distribution and used in such case where the value ~~of~~ of λ is very small and the value of n is very large
- It does not provide theoretical distribution

Def. ⇒ A random variable X is said to follow a Poisson distribution if it assumes only non-negative values and its probability mass function is given by

$$P(X=x) = P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

λ ⇒ Parameter of the distribution

e ⇒ Constant values (2.7)

x ⇒ No. of success in a trial.

- It mostly used the notation $X \sim P(x, \lambda)$ to denote that P is Poisson variate with parameter λ
- X is always positive value / Non-negative value / zero
- It is used for rare events
- $\lambda = \text{mean} = \text{variance}$

application

- ① It is used in quality control statistics to count the no. of defects items.
- ② In Biology → To count the no. of Bacteria.
- ③ In physics → To count the no. of particles.
- ④ In Insurance problems → To count the incoming telephones calls of customers.
- ⑤ To determine the no. of death in particular area & time

Normal Distribution

- It is a continuous probability distribution
- Natural processes are closely related to the normal distribution
- These not only indicate measurement of height, weight of both person or things but also measure human characteristics like IQ.
- It is a limiting case of Binomial distribution.

Properties ① The normal curve is symmetric about the mean, there is no skewness.

② If the mean, median, mode values are equal so, in a normal distribution curve

$$\text{mean} = \text{median} = \text{mode}$$

③ The curve is symmetrical at the base.

application

① Important for theoretical distributions of many physical measurements

② Useful for analysis of data related to business

③ Important for research in social sciences.

④ Results obtained from normal distribution are highly ~~scientific~~ satisfactory. (12)

(7) Data

→ Collection of observations is known as data.

Data Types

① Primary data ⇒ It is a own or self study finding.

→ Data which are collected directly from the field of enquiry for a specific purpose

→ These are raw data or data in original nature & directly collected from population.

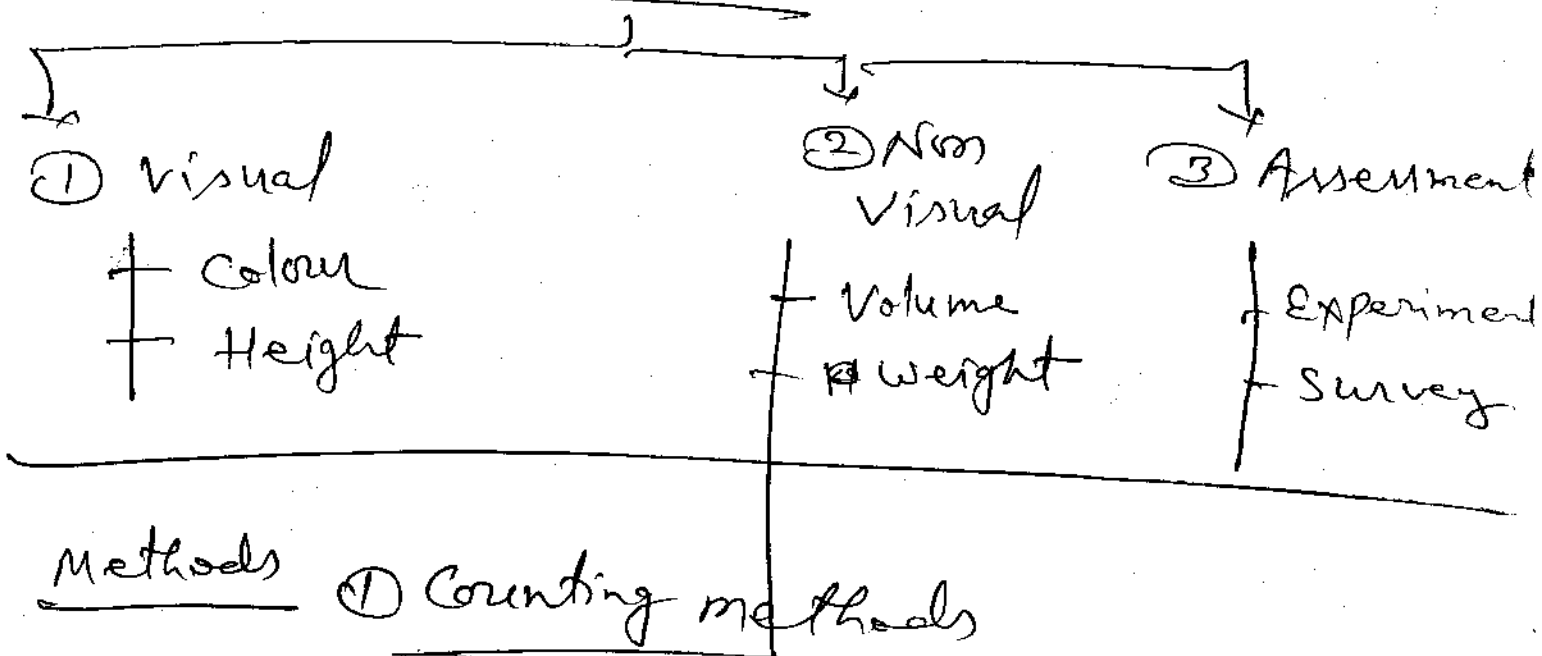
→ The collection of primary data may be made through either by complete enumeration or sampling survey methods.

② Secondary data ⇒ These are numerical information which have been already collected by someone agency for a specific purpose and the subsequently compiled from source for application in different connection.

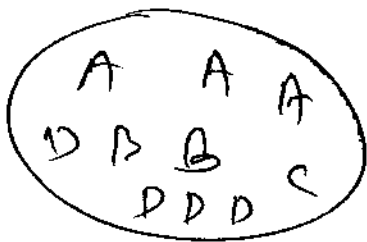
→ Data used by another agency than the collecting authority will be termed as secondary data

Collection of Data

Observations



→ estimation of objects/plants/animals by numerical data



SNo.	Plant Name	No.
①	A	3
2	B	2
③	C	1
④	D	4

② By Survey

It should be based on

- ① Time
- ② method of selection
- ③ season
- ④ sample size etc.

→ It may be Random or in a sequence

③ By experiments

→ The data which are generated by the experiments

Based on

- ① Sampling Technique
- ② Sample preparation
- ③ Standardization
- ④ observation.

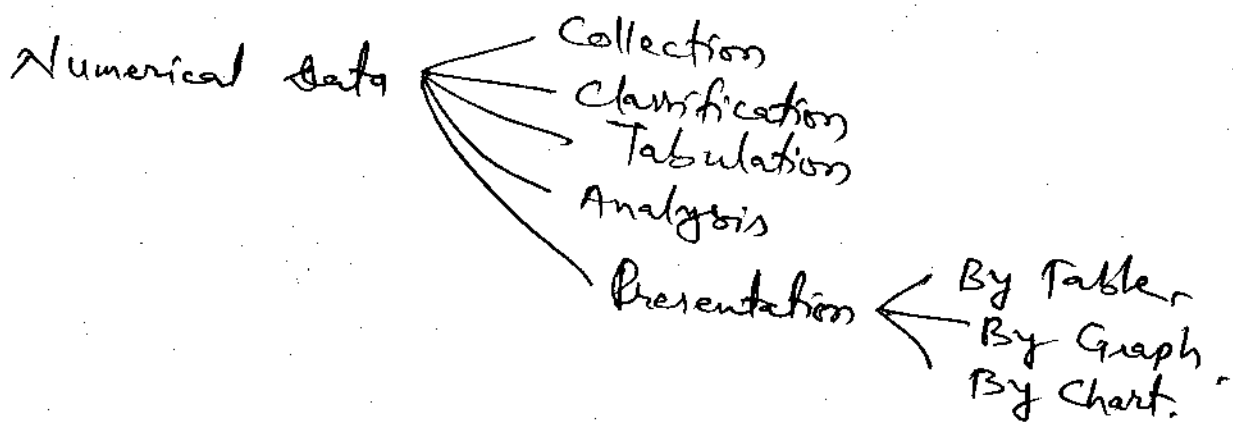
→ It is a science of collecting, analyzing & making inference of the collected data of various sources.

It is used by researchers of many fields to organize, analyze and for summarization of data.

→ Statistics word was first used by a German scientist in 1749 by G. F. Akenwal, who is called father of statistics.

→ Data is numerical description of Quantitative data is also a scientific study of collected data of variable samples.

→ It is a scientific activity of collection & interpretation of data.



Statistics

characters

- ① It should be association of facts
- ② It should be affected to a marked extent by multiplicity of causes.
- ③ It should be expressed by number
- ④ It should be estimated according to reasonable standards of accuracy
- ⑤ It should be collected in scientific manner
- ⑥ Should be collected for a predetermined purpose
- ⑦ They should be placed in relation to each other
- ⑧ These should be easy to understand.

Role of Statistics in Scientific Research.

- Validity —: It help to answer the research question
- Analysis —:
- Efficiency —: In experiment it makes best use of resources.
- Survey —: Select a sample from population & identification based on relationship
- Designing of experiments
- Comparing treatment effects
- Test of Significance
- for easily understanding of the findings.
- Important for evaluation of usefulness of informations
- Useful for measuring problems in various fields.
- Useful for analysis of samples
- It makes presentation clear.
- It also focus on correlation of variables
- Comparative study of samples
- Important for forecasting
- Arrangement of data.

(9) Calculation of mean

S.No.	Electricity Use	Family	
1	5-25	4	4
2	25-45	15	19
3	45-65	12	(31) c
4	L ₁ (65-85) L ₂	(20) f	51
5	85-105	14	65
6	105-125	6	71
7	125-145	4	75 n

$$\begin{aligned} Md &= L_1 + \frac{L_2 - L_1}{f} \left(\frac{N}{2} - c \right) \\ &= 65 + \frac{85 - 65}{20} \left(\frac{75}{2} - 31 \right) \\ &= 65 + \frac{20}{20} \left(\frac{75 - 62}{2} \right) \quad \frac{13}{2} = 6.5 \\ &= 65 + 1 (6.5) \\ &= 65 + 6.5 \\ &= 71.5 \end{aligned}$$

(10) Bogardus Type scaling

- It is a scaling related to psychological test. Created by Emory S. Bogardus to measure peoples willingness to participate in social contacts of varying degree of closeness with members of diverse social groups. Such as ethnic groups etc.
- The scale ask peoples to extent to which they would be accepting of each group (a score of 1.00 for a group is taken to indicate no social distance)
- (As close relatives by marriage (Score 1.00))
 - (As close personal friends (2.00))
 - (As neighbors on the same street (3.00))
 - (As co-workers in the same occupation (4.00))
 - (As citizens of my country (5.00))
 - (As visitors of any country (6.00))

The Bogardus scale is a social distance scale is a cumulative scale, because agreement with any item implies agreement with all preceding items.

The scale has been criticized as too

Simple because the social interactions and attitudes in close familiar or friendship may be quantitatively different from social interactions.

→ For Bogardus social distance is a function of affective distance between the members of two groups.

→ Social distance can be also measured by various parameters, like psychological test

→ It is a field characterized by use of samples of behavior in order to assess psychological constructs.

→ Sample of behaviour is called psychometrics is observation of individual performing tasks.

Principles

- ① Standardization -
- ② Objectivity -
- ③ Test norms -
- ④ Reliability -
- ⑤ Validity -

Types of Tests

- ① IQ / Achievement Tests
- ② Public safety employment tests
- ③ Attitude tests
- ④ Neuropsychological Tests
- ⑤ Personality tests.
- ⑥ Objective tests
- ⑦ Projective tests

① Scale →

→ Measurement scales are used to categorize or quantify the variables.

→ It is useful for analysis of data.

Measurement → Normally when any one here the term measurement they may think term measuring the length or measuring a quality. These represent a limited use, but in stats. The word measurement is used more broadly.

→ Scale of measurement refers for 'defining & categorization of variables/Numbers.

→ Each scale of measurement has certain properties which determine certain statistical analysis.

Properties of measurement of scales

① Identity ⇒ Each value of the measurement scale has a unique meaning.

② Magnitude ⇒ Values on the measurement scale have an order relationship as some values are large & some are small.

③ Equal Interval ⇒ Scale units along the scale are equal to one another it means the difference between 01 & 02 would be equal to difference between 19 & 20.

④ a minimum value of zero.

Types : 4 types.

① Nominal level of measurement

→ It only satisfied the identity property of measurement

→ Values assigned to variables represents a descriptive category

→ Example → Gender, ← male or female

→ Simple data category without order

→ In research Yes/No's a nominal scale measurement

② Ordinal Scale

→ It has properties of both identity + magnitude.

→ Each value of the ordinal scale has a unique meaning and it has ordered relationship to every other value on the scale.

example ⇒ Horse race reported as win, place, show.

Ranking	+	1st Division
	+	2nd Division
	+	3rd Division

- Simplest ordinal scale is ranking.
- There is no any distance between two points
- Ordinal data use non-parametric statistic & includes — median, mode, Correlation.

③ Interval Scale of measurement

- Its properties are identify magnitudes & equal interval
- It informs about bigger or smaller of different values in a data series.

S.No.	Marks Range	No. of students	
1	40-50	28	2nd position
2	50-60	10	
3	60-70	13	3rd position.
4	70-80	35	1st position

④ Ratio Scale measurement ⇒

- It is a top level of measurement
- It has free zero point

example ⇒ measurement of ~~weight~~ Length.

- It includes all properties of measurement