

STATISTICAL DESCRIPTION OF DATA

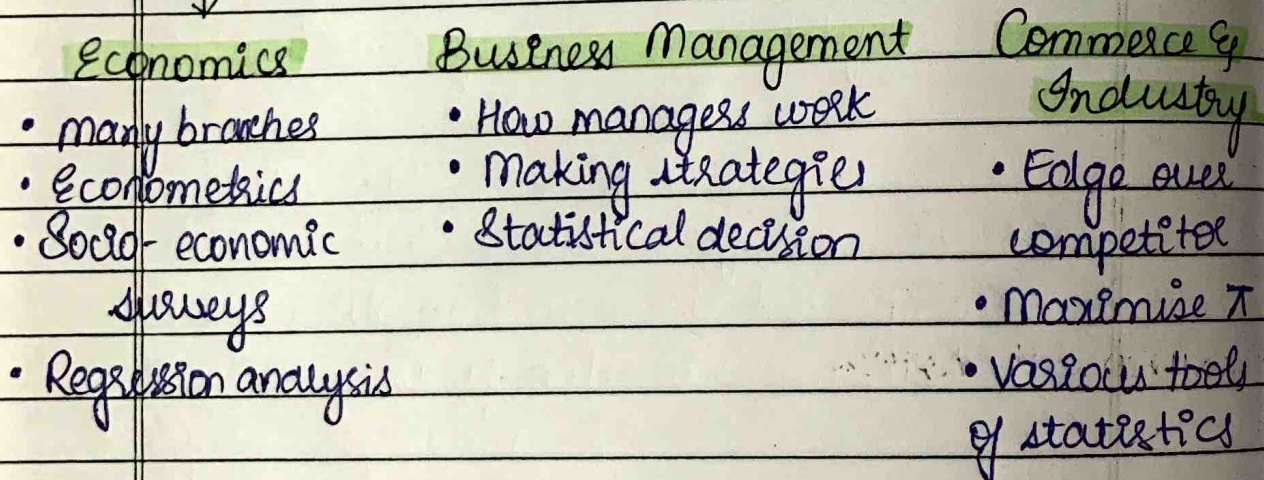
Statistics

- Latin word **Status**
- Italian word **Statista**
- German word **Statistik**
- French word **Statistique**

Plural Sense: Collection of data for some specific purpose.

Singular Sense: Science of collecting, ^{Presenting} analysing and interpretation of data.

APPLICATION OF STATISTICS



LIMITATIONS

- Can only be used by experts
- Quantitative data only.
- Study of aggregates only
- Homogeneity of data
- Results may prove to be wrong

DATA CHARACTERISTICS → aggregate of facts

DATA

Quantitative/Cardinal

- Numeric



measured as variable

Qualitative/Ordinal

- colour, honesty, love...



measured as attribute



characteristic

Discrete

expressed as whole
or countable numbers

Age, House No, No of
Books in library, days
in a month

Continuous

expressed as real
number or measurable terms.

height, weight, income, temperature,
time, distance

Sincerity
Beauty
Nationality etc

DATA

Primary

Data collected for

a specific purpose
directly from the
field of enquiry
firsthand & original

Secondary

Previously collected

as primary data by some
agency for specific purpose
but now compiled from
that source for use
in a different connection

METHODS OF COLLECTING PRIMARY DATA

Direct Observation

Mailed Questionnaire

Interview

Direct / Personal
Interview

Indirect
Interview

Telephonic
Interview

DIRECT OBSERVATION

- Best method of data collection
- Measured using instrument
- Time consuming
- Less area coverage
- High cost
- Data accuracy and reliability

MAILED QUESTIONNAIRE

- well drafted & properly sequenced covering all aspects of problem under study
- Sent to the respondents.
- Time consuming
- Non-response is maximum
- Wide area covered
- Only for literate
- Missing data.
- Data accuracy and reliability

DIRECT PERSONAL INTERVIEW

- Face to face, Oral investigation, no instrument
- More quick and accurate but not suitable for large area
- Used in case of natural calamities like earthquake, cyclone, epidemic.

INDIRECT INTERVIEW

- Used when respondents can't be reached directly
- data is collected from persons associated with the problems
- Eg: In case of accident
- More accurate but not suitable for large area

TELEPHONIC INTERVIEW

- Quick and less expensive

- Covers largest area.
- Information gathered via phone call.
- Less consistent, accurate & reliable
- Non response is maximum.

Sources of Secondary Data

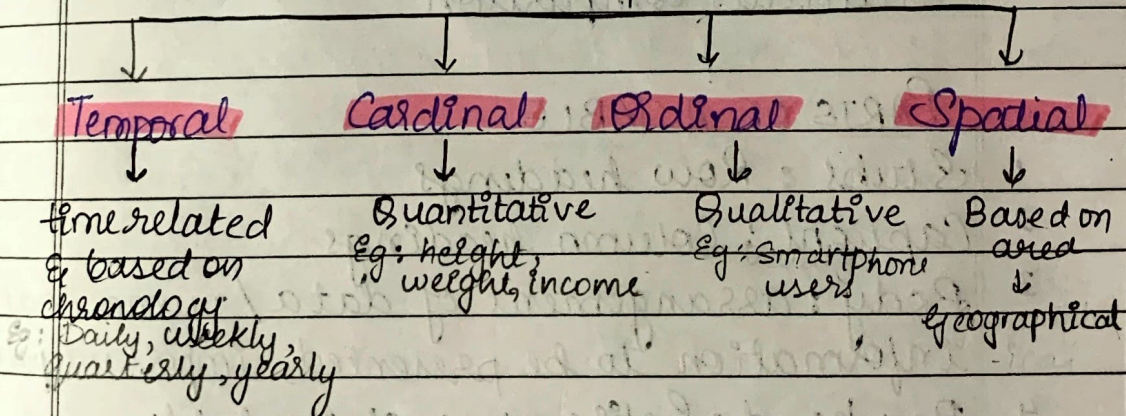
- International Sources like WHO, ILO etc.
- Government Sources like Ministry of Food etc.
- Private and Quasi Government Sources, like NCERT
- Unpublished sources of various research institutes
- Google.

CLASSIFICATION OF DATA

It is splitting, dividing or separating the data based on common characteristics.

To organise similar things together

TYPES OF CLASSIFICATION



SCRUTINY OF DATA

Checking the data to find mistakes, flaws, errors, missing data etc.

It is done by experts and trained people using tools, techniques and softwares.

The standard data is compared with available/collected data.

In short, cleaning & filtering of data.

PRESENTATION OF DATA

↓
Textual

↓
Tabular

↓
Graphical

Textual:

- In written form
- Simple but dull & monotonous
- comparison is not possible.

Tabular:

Presentation of data with the help of statistical table having rows & columns

Advantages:

- Complicated data can be represented.
- Important for diagrammatic representation
- Statistical analysis not possible without tabulation
- Facilitates comparison

PARTS OF A TABLE

1. Stubs: Row headings
2. Captions: Column headings
3. Body: arrangement of data/numerical information to be presented, important part
4. Box head: Entire upper of the table
5. Title: Brief description of contents of table, self explanatory, below or next to table number
6. Table Number
7. Head Note: units of measurement & abbreviations
8. Foot Note: additional information to be provided
9. Source Note: Source from where data is taken

TYPES - Simple and Complex

DIAGRAMMATIC REPRESENTATION

{ Graphs and diagrams }

Can get only approximate idea but emphasizes certain facts & relations more than numbers do.

Advantages:

- More attractive and informative
- Complex problems classified easily
- Reveals hidden facts
- Easily comparing two or more sets of values
- Shows relation of the parts to the whole.

TYPES

Without frequency

- Line chart/graph/diagram
- Bar Diagram/chart (1D)
- Pie chart (2D)

With frequency

- Histogram
- Frequency Polygon
- Frequency curve
- Ogive

WITHOUT FREQUENCY

LINE DIAGRAM:

Used for time series

Wide range of fluctuations = Use logarithmic or ratio charts

Multiple line chart - multiple data in one chart with same units

Multiple axis chart - multiple data in one chart with different units.

Best Presentation
Most accurate
TABULAR
Hidden trends: DIAGRAMMATIC

PAGE NO.:

DATE:

BAR DIAGRAM

1. Vertical Bar diagram.
 - Time series
 - Quantitative
2. Horizontal Bar diagram
 - Qualitative
 - Data varying over space
3. Multiple / Grouped Bar diagram
 - compare related series
4. Component / Sub-divided Bar diagram
 - Used to represent data divided into different components.
5. Percentage Bar Diagram
 - Compare different components with table value
6. Deviation Bar Diagram
 - Represents net quantities (+ve & -ve values)
7. Broken Base
 - wide variations in values
 - Done to gain space

PIE DIAGRAM

Comparing different components to total value

$$\text{Angle for a component} = \frac{\text{Value of comp}}{\text{Total value}} \times 360^\circ$$

WITH FREQUENCY

Frequency Distribution

Grouped [6]

Ungrouped (Discrete)

Continuous	Discontinuous	Equal width	Unequal width	Open end	Closed end
10-20	11-20	equal length of class	unequal lengths or different lengths	less than 10	10-20
20-30	21-30			10-20	20-30
30-40	31-40			20-30	30-40
(exclusive) overlapping	(inclusive) non-overlapping			30 and above	

FREQUENCY GRAPHS

Histogram
only for
grouped freq

Joined bar
diagram

frequency
comparison

MODE
calculation

Area diagram

frequency
bar graph

Frequency Polygon

- For ungrouped frequency data
- Joined via straight lines
- closed figure
- MODE calculation
- Frequency comparison

Frequency curve

- limiting form of polygon & histogram
- Symmetry checking left = -ve
right = +ve
- both grouped & ungrouped

QUARTILE
DECILE
PERCENTILE

Ogive

- more than & less than ogives.

Intersection

• MEDIAN calculation

• Cumulative

freq. curve

• FRACTILES

can be

calculated

graphically

• both grouped

& ungrouped

SPECIAL TYPES OF DIAGRAMS

Pictogram - for pictures

Cartogram - Map representation

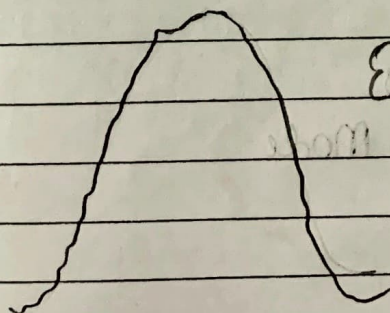
Ideogram - Pictogram representing some idea.

OTHER FREQUENCY CURVES

1. Bell Shaped (Symmetrical curve)

- Commonly used for distribution of height, weight, profit etc
- limiting form of ^{freq} polygon and histogram
- Enables to understand symmetry of distribution

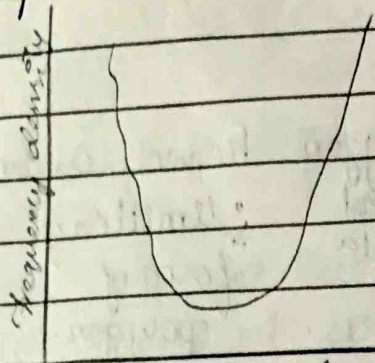
Frequency density



{ freq = max in centre }

Class boundary

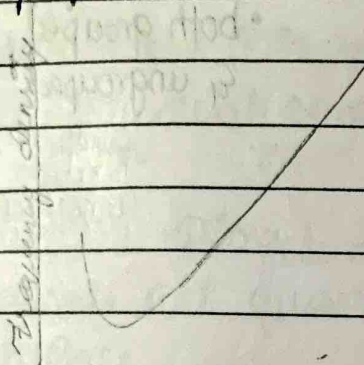
V Shaped curve



freq = min in centre

Eg: Distribution of people travelling on streets

J shaped curve

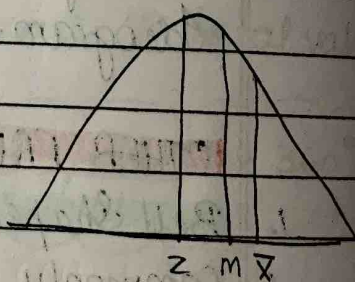


freq = starts with min gradually reaching max

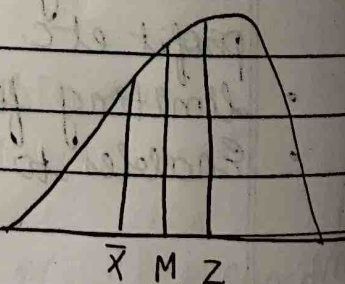
Eg: Distribution of commuters in a particular time period

Asymmetric curves

- Positively skewed
Mean > Median > Mode
- Frequency curve as a longer tail to the right



- Negatively skewed
Mean < Median < Mode
- Frequency curve is a longer tail to the left



Symmetric curves

- Mean = Median = Mode
- Skewness = 0

SOME IMPORTANT TERMS AND FORMULAE

Frequency / Tally marks: No of times a value has occurred.

Range: Largest value - Smallest value

Class interval: A large no of observations having a wide range classified into number of groups.

Class frequency: No of observations a class contains

Total frequency: Total no of observations in the frequency distribution.

Class limit: Two ends of a class interval
Smaller - LOWER class limits non-overlapping
Greater - UPPER class limits ^{or} mutually inclusive

No of class = Range

Class length $\rightarrow$ UB - LB

No of class $\times$ Class length = Range

Class Boundaries

For continuous variable; class limit and class boundaries are same and here we exclude the upper limit so it is mutually exclusive.

For discrete

UCB = UCL + m where $m = \frac{\text{UCL of class} - \text{LC of class}}{2}$

LCB = LCL + m

Discrete variable $\rightarrow$ Grouped freq distribution
Class Interval $\rightarrow$ Mutually inclusive.

$$\text{Class midpoint} = \frac{LCI + UCI}{2} \quad \text{or} \quad \frac{LCB + UCB}{2}$$

Class width: length of a class obtained by difference between upper and lower class boundaries of that class.
 $= UCB - LCB$

$$\text{Frequency Density} = \frac{\text{Class frequency}}{\text{Class width}}$$

Measures concentration of frequency of different classes

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

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

CUMULATIVE FREQUENCY

- Frequencies are cumulated.
- Prepared from grouped frequency distribution by taking class boundaries.
- No of observations less than or equal to class boundaries - LESS THAN CUMULATIVE FREQUENCY
- No of observations greater than or equal to class boundaries - MORE THAN CUMULATIVE FREQUENCY
- For discrete series - ungrouped as well as grouped data