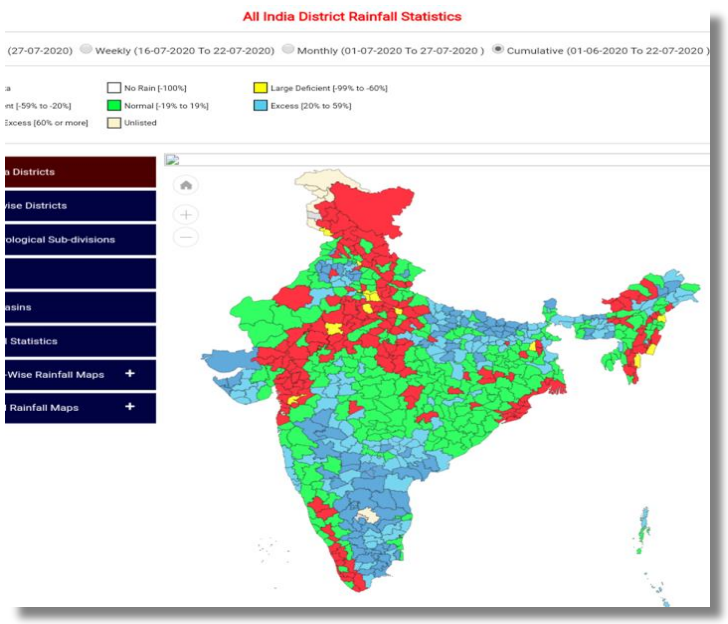
Date: 13th March,2021

VIRTUAL COACHING CLASSES
ORGANISED BY BOS (ACADEMIC), ICAI

FOUNDATION LEVEL
PAPER 3: BUSINESS MATHEMATICS AND LOGICAL REASONING & STATISTICS

Faculty: CA Rashmi Lonikar, M.Sc., FCA, DISA

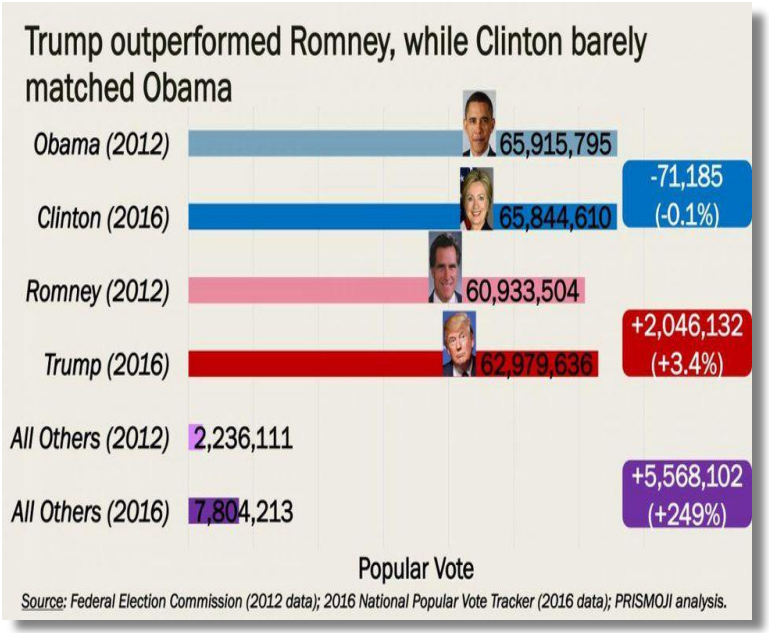
Introduction of Statistics



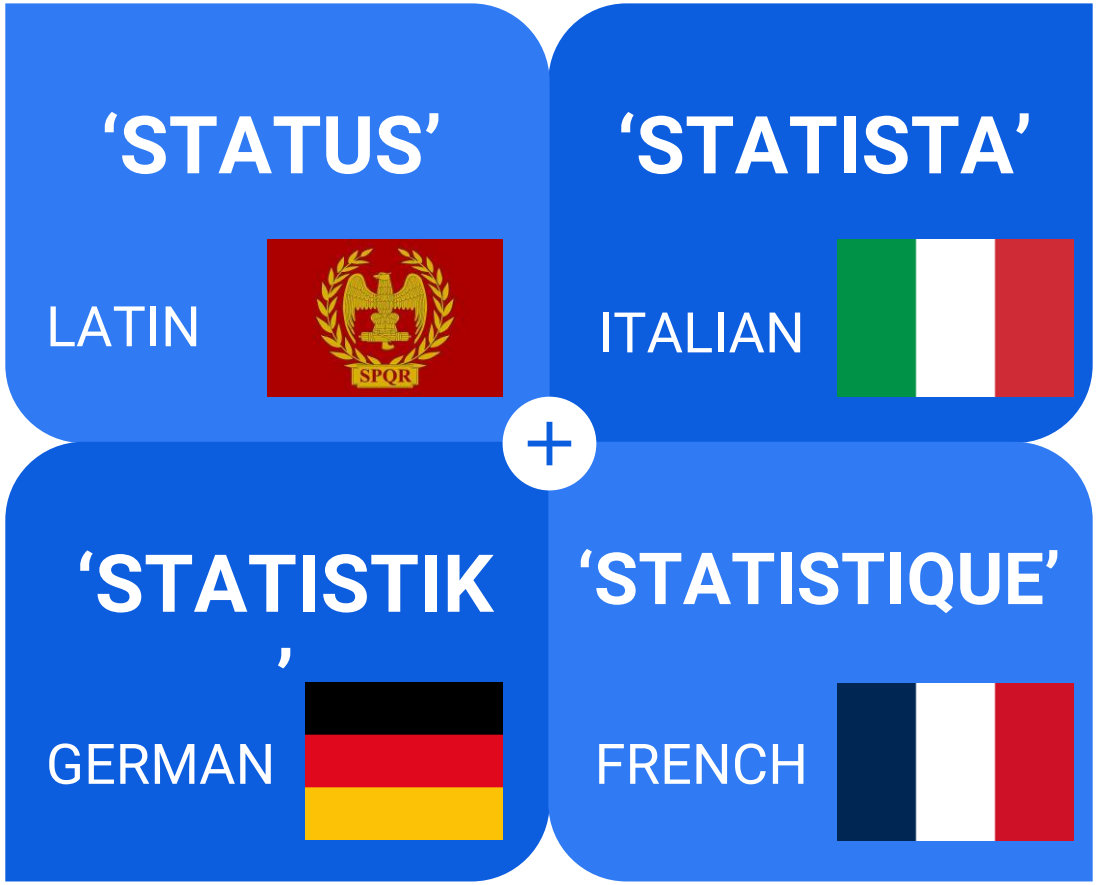
Tests	Mat	Inns	Runs	HS	Ave	100	50
Kohli Today	86	145	**7240**	**254***	53.62	27	22
Sachin at same age	126	203	**10323**	**248***	56.71	35	41
Sachin at same matches	86	139	**7112**	**217**	57.35	26	28
Sachin at same innings	91	146	**7673**	**217**	58.57	28	30

ODIs	Mat	Inns	Runs	HS	Ave	BF	SR	100	50
Kohli Today	248	239	**11867**	**183**	59.33	12726	93.25	43	58
Sachin at same age	358	349	**13909**	**186***	44.01	16190	85.91	38	71
Sachin at same matches	248	241	**9237**	**186***	42.17	10689	86.41	25	47
Sachin at same innings	246	239	**9108**	**186***	41.97	10569	86.17	25	46

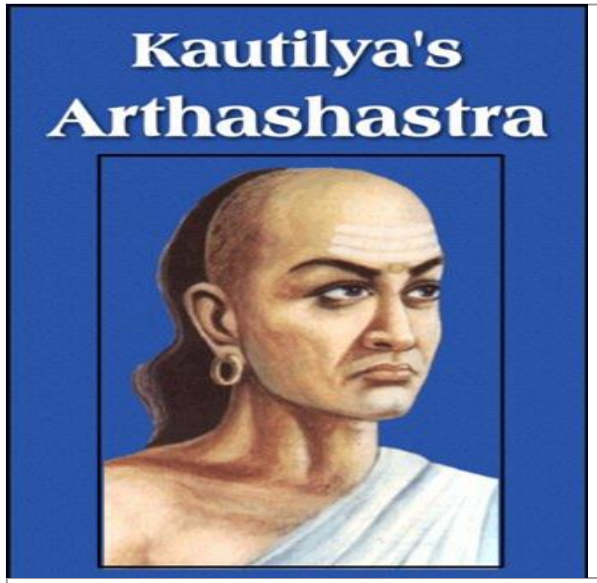
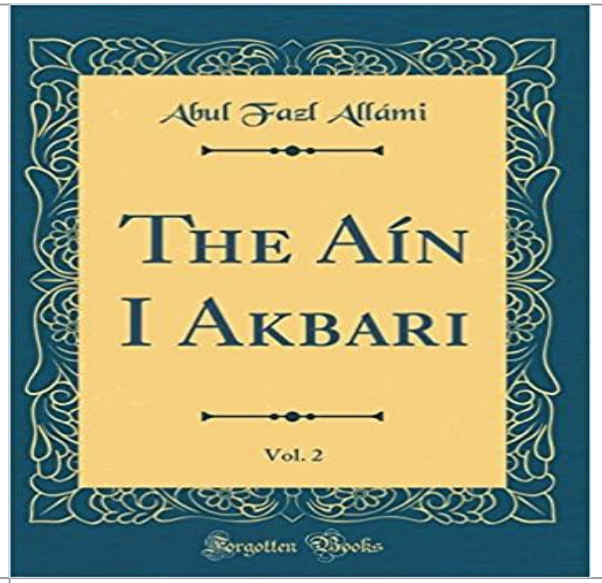
Aggregate	Mat	Inns	Runs	100	50
Kohli Today	334	384	19107	70	80
Sachin at same age	484	552	24232	73	112
Sachin at same matches	334	380	16349	51	75
Sachin at same innings	337	385	16781	53	76



Origin of the Word “STATISTICS”



History of Statistics

			
• Kept a record of births and deaths in the kingdom • Other precious statistics that helped historians map that era • Helped in the understanding of Chandagupta's reign	• Written in 16th century by Abul Fazl • Contains important agricultural records and statistics	• First recorded *census* was during the Egypt pharaoh era	• BONUS: 'Lady Tasting Tea' experiment • Helped formulate and consolidate concepts like significance, confidence interval, etc.

Definition of Statistics

Singular "STATISTICS"	Plural "STATISTICS"
• Refers to the actual scientific method of dealing with data • Includes the processes of collecting, analysing and presenting data which would lead to drawing statistical inferences • Is also called Science of Counting and Science of Averages • *Example*: "Learning statistics", "Statistics Practitioners", "Statistical Analysts"	• Refers to the actual data that is collected and analysed • Can be qualitative or quantitative • *Example*: Statistics of Tendulkar vs Kohli, Federer vs Nadal

Application of Statistics: Economics

- Economics and Statistics are closely related
- Example: Time Series Analysis, Index Numbers, Demand Analysis
- Econometrics makes heavy use of maths and stats for economic analysis
- Macro-level examples: GDP growth projections, National Budget planning, etc.

Application of Statistics: Business Management

- Business decisions are increasingly being taken by referring to numbers, data, charts, etc.
- Role of “hunches”, “intuition”, “guesses” is being complimented by insights from statistical analysis.
- Big business decisions like mergers & acquisitions, salary changes, mass hiring/firing are now taken with the help of statistical data.
- Example: McDonald’s decisions to open new stores in a place are heavily influenced by statistical factors like the residents’ average income, spending capacity, traffic density, etc.

Application of Statistics: Commerce & Industry

- Industries are broadening their horizon by including statistical analysis into decision-making
- Data of things like raw material, finished goods, salaries, wages, etc. is collected and analysed in order to maximise profits
- Accounting and financial modelling processes are also using statistics to design financial models, project future cash flows among other things

Limitations of Statistics

- **Statistics deals with aggregates**: Individual data may be insignificant and/or act as an outlier to the macro data.

Example: USA National Average Income vs Bill Gate's Income

- **Statistics considers quantitative factors only**: Qualitative factors may be missed out during analysis. However, qualitative data can be converted to quantitative using representative numbers.

Example: The impact of socio-political factors in determining country's future GDP

- **Statistical projections are based on set factors and under specific conditions**: If the conditions abruptly change then the projections may prove to be different from actuals.

Example: Projecting a region's rainfall amount vs a sudden cloudburst in that region

- **Statistical inference theory is based upon random sampling**: If sampling scheme is not proper i.e. not a random-enough or not a large-enough sample is taken, results may be skewed.

Example: Hillary Clinton's 92% winning chance stated by NY Times. More samples were taken from Democratic voters than Republican ones.

Revision

- Statistics deals with the aggregates. An individual, to a statistician has no significance except the fact that it is a part of the aggregate.
- 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.
- The theory of statistical inferences is built upon random sampling. If the rules for random sampling are not strictly adhered to, the conclusion drawn on the basis of these unrepresentative samples would be erroneous.

Collection of Data: Introduction

- What is Data? It is the quantitative information about a particular characteristic that is under consideration
- Qualitative factors are also important and need to be converted in quantitative form for analysis
- A quantitative characteristic is called a '**variable**' i.e. a variable is a measurable quantity
- When a variable has a finite or a countably infinite set of possible values, it is termed as '**discrete**'. Example: No. of petals on a flower (4, 5, 6), no. of senior citizens in a state (1 million, 2 million)
- When a variable can assume any value in a given range, it is termed as '**continuous**'. Example: Height of a person (160.5cm, 160.55cm, 160.45cm), volume of water in a bottle (700ml, 700.5ml, 700.4999ml)
- Qualitative data is called an '**attribute**'. Example: Nationality of a person, color of a rock. *Important point*: Attributes are usually not comparable i.e. 'Blue frog is more color than green frog' is not a logical statement

Collection of Data: Introduction

- Data can be classified into two types: Primary and Secondary. This classification is done on the basis of "***Method of Collection of Data***"

Primary Data	Secondary Data
• Data that is collected first-hand by the analyser is called primary data	• Data that has already been collected by another party and is being used by the analyser for further study is called secondary data
• Example: Collecting the height of all students in this lecture for calculating average height of the class • In this case, I am collecting the data first-hand, hence the data is primary	• Example: Taking national height data from govt. website to find average national height • In this case, some government body has already collected the data and I am using it for further analysis, hence the data is secondary

Collection of Primary Data

Interview Method

Personal Interview
- Investigator meets respondent directly and collects info then and there
- Used during **natural calamity** like cyclone, pandemic for quick and accurate collection
- Difficult to cover large areas

Indirect Interview
- Used when there are practical problems in reaching people
- Example: **rail accident**, where investigator collects information from people associated with the problem
- Difficult to cover large areas

Telephonic Interview
- Quick, inexpensive method to collect data over the phone
- Easy to access people over large areas with this method
- But large no. of non-responses

Mailed Questionnaire Method
- Frame an organised, sequential list of questions and mail to respondents
- Wide areas can be covered but non-responses are maximum here

Observation Method
- Data collected by direct observation and may include instruments
- Example: height of students, color of leaves in the area
- Best quality data but laborious, time-consuming and small area coverage

Questionnaires Filled and Sent by Enumerators
- Used for larger enquiries from persons surveyed
- Mixture of interview and questionnaire method
- Enumerator asks respondents questions from questionnaire, which is then mailed back to investigator

Collection of Secondary Data

- Some important sources of Secondary Data

1. International Sources: WHO, ILO, World Bank, IMF
1. Government Sources: Ministry of Food & Agriculture, Securities & Exchange Commission (USA)
1. Private and quasi-government Sources: Indian Statistical Institute, NCERT
1. Unpublished sources of various research institutes, researchers, universities

Scrutiny of Data

- Since all our analysis is based on the collected data, it is important to check if the data is **accurate** and **consistent**
- There is no set procedure for checking the validity of data; however, one must use their intelligence in order to verify it
- Errors may creep into the data from many sources, including human fault. It is equally important to both identify and remove said errors
- Example of data scrutiny: If you have collected data on the population, area and population density of a region, you must ensure that $\text{Density} = \frac{\text{Area}}{\text{Population}}$ holds before proceeding with further analysis.

Revision

- Types of Data: Primary and Secondary
- Methods of Primary Data Collection: Interview, Mailed Questionnaire, Observation, Questionnaires Filled and Sent by Enumerators
- Methods of Secondary Data Collection: International bodies, governments, private, unpublished
- Scrutiny of Data

Presentation of Data

- Once data is collected and verified, it needs to be presented in a proper format to highlight the data's essential features

- From a presentation perspective, data may be classified into the 4 types:

1. Chronological / Temporal / Time Series Data
   eg: Number of CA aspirants over the last 20 years

1. Geographical / Spatial Series Data
   eg: Number of CA aspirants in Maharashtra vs Gujarat

1. Qualitative / Ordinal Data
   eg: Data on nationality, eating habits, etc.

1. Quantitative / Cardinal Data
   eg: Data on height, weight, etc.

- 1 and 2 come under ***non-frequency*** data type while 3 and 4 come under ***frequency*** data type

Mode of Presentation of Data: Textual

- Representing data in sentences through a paragraph or multiple paragraphs
- Merits: Simple, easy to understand, exactness of data can be maintained, first step towards other forms of presentation
- Demerits: Dull, monotonous, difficult to derive insights, difficult to compare with other data

Mode of Presentation of Data: Tabular

- Representing data in a table consisting of rows and columns, reference no., title, description etc.
- Certain guidelines to be followed while representing data using this method

Mode of Presentation of Data: Tabular

- Guidelines for tabulating data:

I A statistical table should be allotted a serial number along with a self-explanatory title.

II The table under consideration should be divided into caption, Box-head, Stub and Body. Caption is the upper part of the table, describing the columns and sub-columns, if any. The Box-head is the entire upper part of the table which includes columns and sub-column numbers, unit(s) of measurement along with caption. Stub is the left part of the table providing the description of the rows. The body is the main part of the table that contains the numerical figures.

III The table should be well-balanced in length and breadth.

IV The data must be arranged in a table in such a way that comparison(s) between different figures are made possible without much labour and time. Also the row totals, column totals, the units of measurement must be shown.

V The data should be arranged intelligently in a well-balanced sequence and the presentation of data in the table should be appealing to the eyes as far as practicable.

VI Notes describing the source of the data and bringing clarity and, if necessary, about any rows or columns known as footnotes, should be shown at the bottom part of the table.

Mode of Presentation of Data: Tabular

- Example of Tabular Presentation

Table 14.1

Status of the workers of Roy Enamel factory on the basis of their trade union membership for 2009 and 2010.

Status	Member of TU			Non-member			Total		
Year	M (1)	F (2)	T (3)=(1)+ (2)	M (4)	F (5)	T (6)=(4)+ (5)	M (7)	F (8)	T (9)=(7)+ (8)
2009	3900	300	4200	300	500	800	4200	800	5000
2010	4200	840	5040	500	450	950	4700	1290	5990

Source:

Footnote: TU, M, F and T stand for trade union, male, female and total respectively.

Mode of Presentation of Data: Tabular

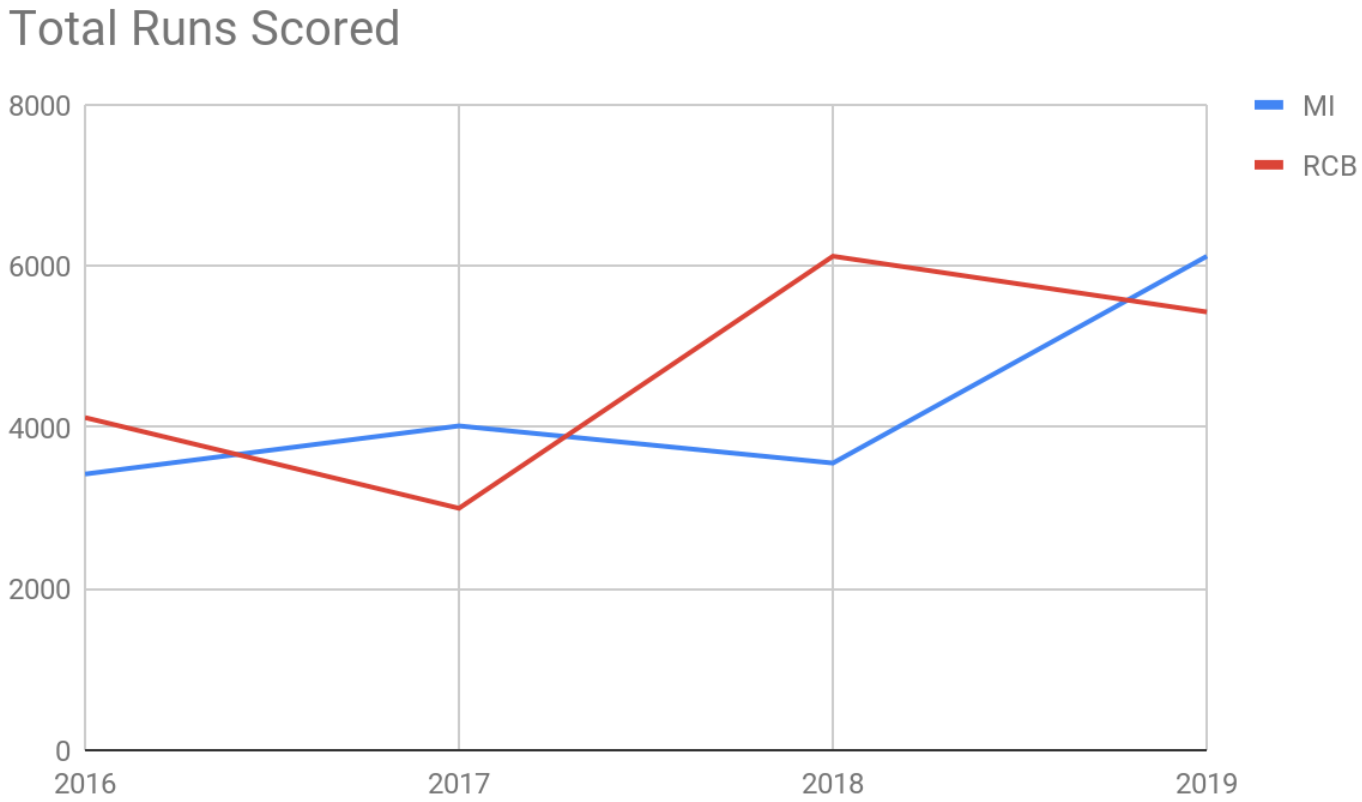
- Merits:

1. Row-Column comparison is facilitated
1. Complicated data can be represented
1. It is a precursor for diagrammatic representation
1. Statistical analysis in most cases impossible without tabular representation

Mode of Presentation of Data: Diagrammatic

I. Line Diagram:

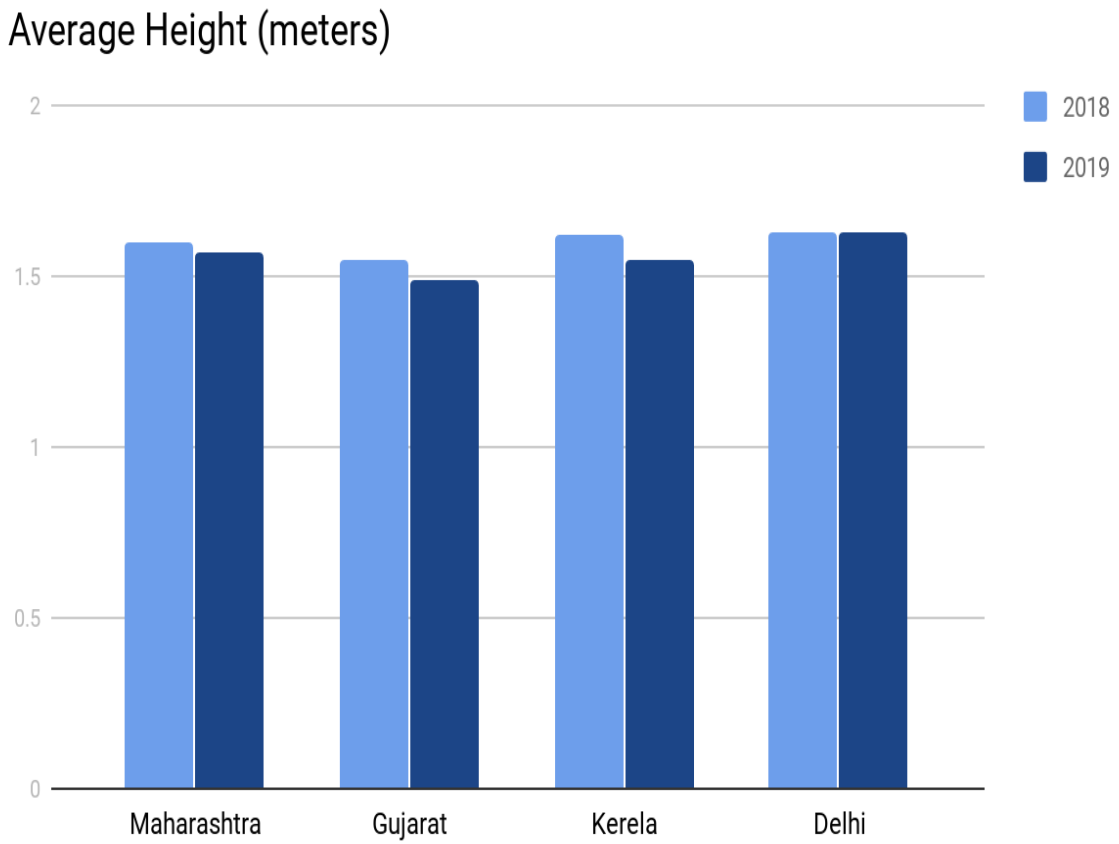
- Usually used for **time series data**
- The dependent and independent variable (usually time) are plotted on a graph and the points are sequentially connected to show a trend or lack thereof
- Multiple line charts may be used on the same graph to represent separate but related data sets



Mode of Presentation of Data: Diagrammatic

II. Bar Diagram:

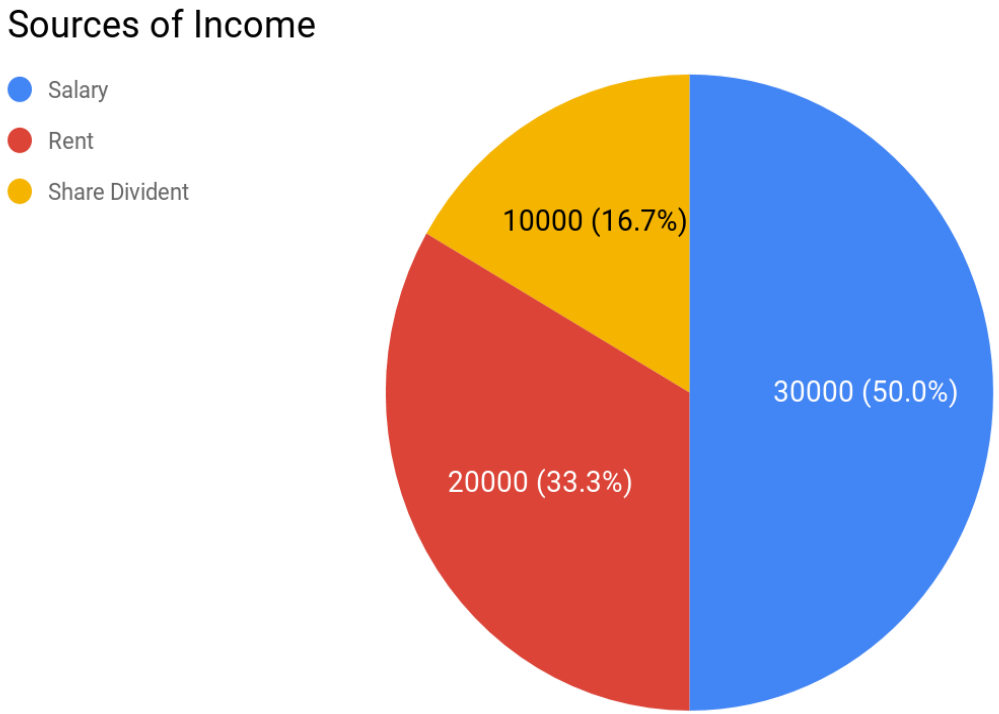
- Two types: horizontal and vertical
- Horizontal bar diagram usually used for qualitative data
- Vertical bar diagram is used in quantitative data representation, even for time series
- Consists of rectangles of equal width, equally spaced with varying length, the length being the representation of the data
- Bars may be grouped, combined or split to represent various forms of related data and its percentages



Mode of Presentation of Data: Diagrammatic

III. Pie Chart / Pie Diagram:

- Used when the percentage distribution of various components is important
- Represented as a circle or cylinder cut up into parts, with the angular size of each component representing the % weight of it among the whole
- Here, sum of all % must be 100 and sum of all angles must be 360



Revision

- Mode of Presentation of Data
  - (a) Textual presentation;
  - (b) Tabular presentation or Tabulation;
  - (c) Diagrammatic representation.
- The types of diagrams:
  - (a) Line diagram or Historiagram;
  - (b) Bar diagram;
  - (c) Pie chart.

Frequency Distribution

- Simply put: It means counting across variables and attributes
- When frequencies are counted and represented in a tabular format, it is called a frequency distribution table
- When tabulation is done for discrete random variable: **Ungrouped / Discrete / Simple** FD
- When tabulation is done for continuous variable: **Grouped** FD

Example 14.4: Following are the records of babies born in a nursing home in Bangalore during a week (B denoting Boy and G for Girl) :

B G G B G G B B G G
G G B B B G B B G B
B B G B B B G G B G

Construct a frequency distribution according to gender.

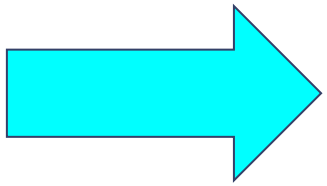


Table 14.3

Frequency distribution of babies according to Gender

Category	Number of births
Boy (B)	16
Girl (G)	14
Total	30

Frequency Distribution

- Frequency Distribution Table steps:

I Find the largest and smallest observations and obtain the difference between them, known as Range, in case of a continuous variable.

II Form a number of classes depending on the number of isolated values assumed by a discrete variable. In case of a continuous variable, find the number of class intervals using the relation, No. of class Interval X class length $\cong$ Range.

III Present the class or class interval in a table known as frequency distribution table.

IV Apply 'tally mark' i.e. a stroke against the occurrence of a particulars value in a class or class interval.

V Count the tally marks and present these numbers in the next column, known as frequency column, and finally check whether the total of all these class frequencies tally with the total number of observations.

Frequency Distribution

- Important terms:

Let us solve Example 14.5 (Page 14.15 onwards) to understand the terms like:

1. CL
2. CB
3. Mid-point
4. Class Width
5. Cumulative Frequency
6. Frequency Density of a Class Interval
7. Relative Frequency and Percentage Frequency of a Class Interval

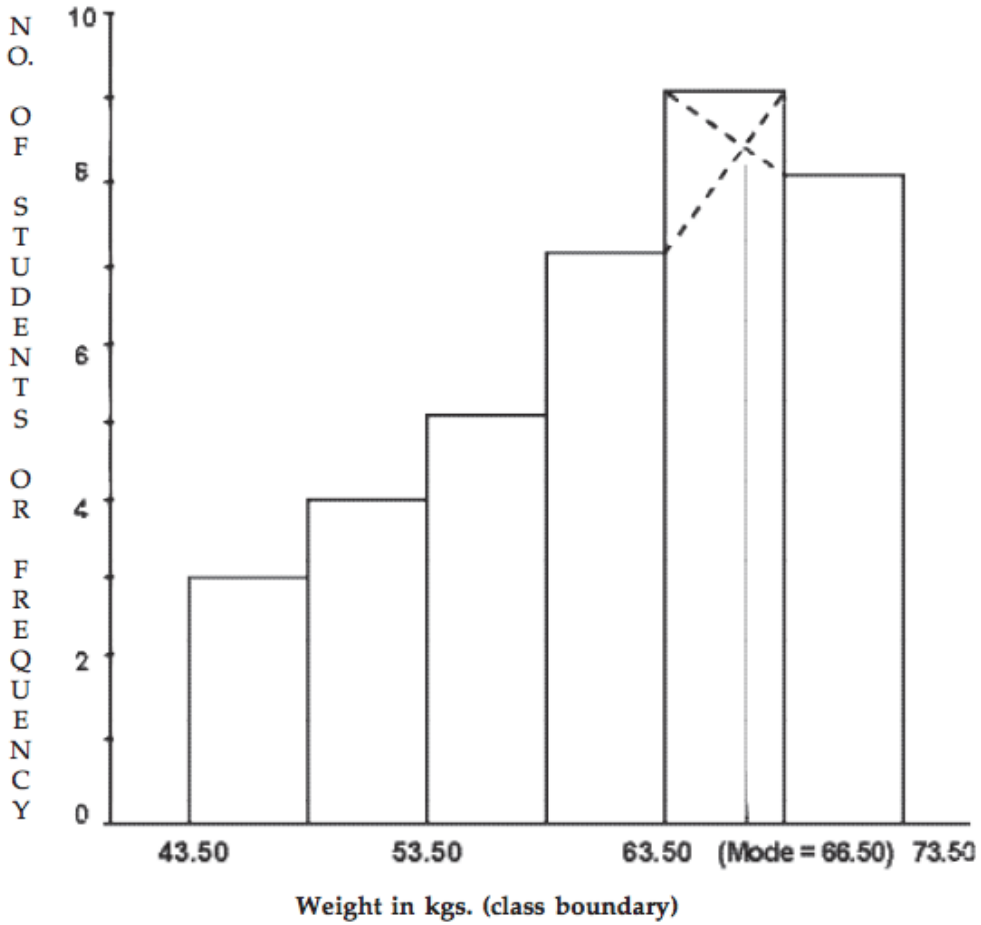
Graphical Representation of FD: Histogram

❖ **Histogram or Area Diagram**

- Convenient for FD representation
- Gives idea about the frequency curve of variable
- A comparison across different class intervals is possible in this method

Steps:

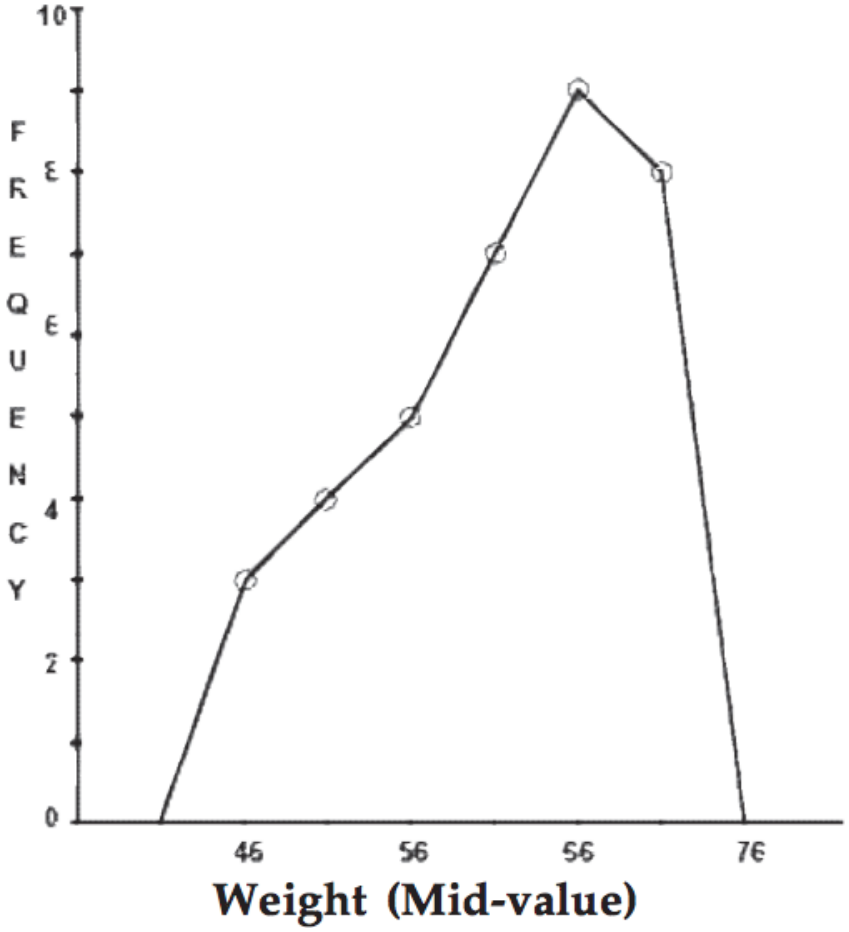
1. Convert class limits to class boundaries
2. Set rectangles adjacent to each other, with rect. width = class width
3. Make length of rec. = frequency of class



Graphical Representation of FD: Frequency Polygon

❖ **Frequency Polygon**

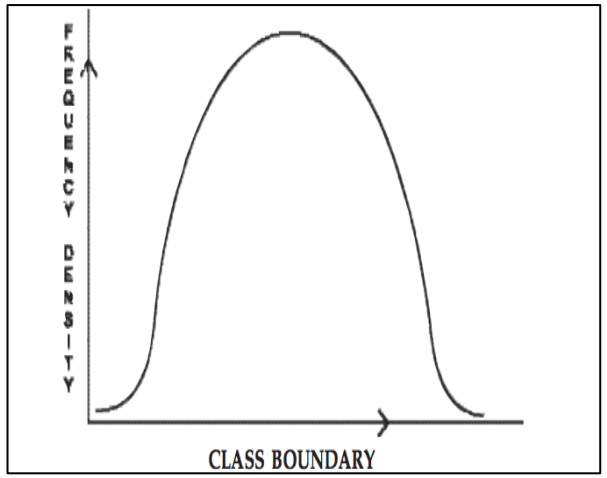
- Usually for single FD
- Can be applied to grouped with same class width
- Class mid-points and their frequencies are plotted and joined by lines
- Extreme end points with assumed 0 frequency are also connected in order to complete the polygon



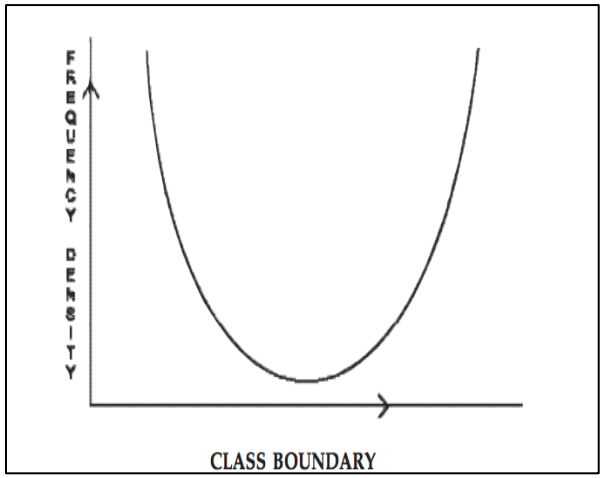
Graphical Representation of FD: Frequency Curves

❖ **Frequency Curves**

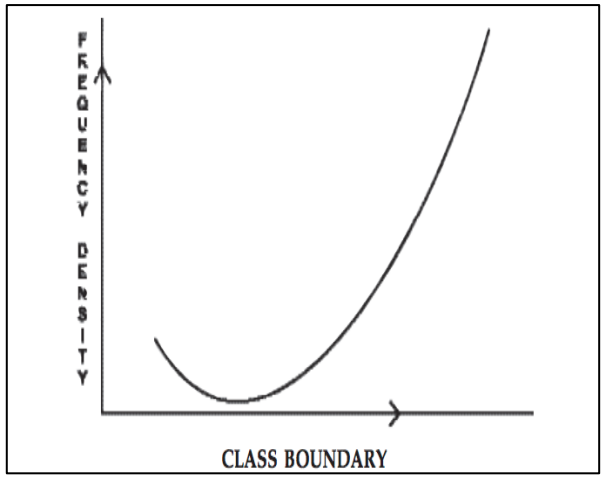
- Area under the graph is unity i.e. 1
- Smooth curve
- Similar form used heavily in probability distribution analysis, which obviously comes from the frequency data sets



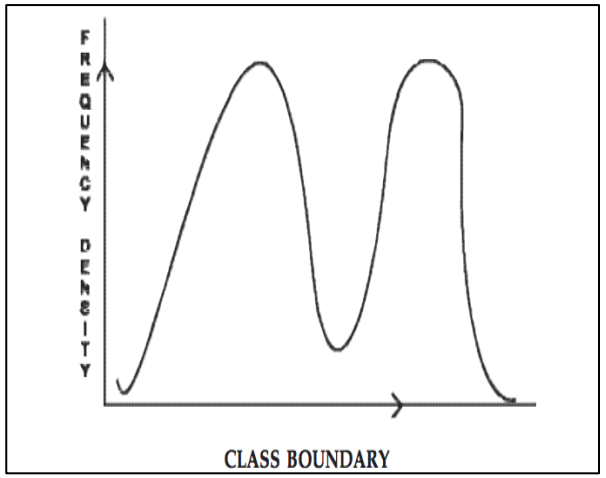
Ex: IQ Distribution



Ex: Customers in Restaurant vs Time



Ex: People entering Kolkata vs Time



Ex: Head hair distribution vs Age

Revision

- Frequency Distribution of a Variable

  (a) Find the largest and smallest observations and obtain the difference between them, known as Range, in case of a continuous variable.

  (b) Form a number of classes depending on the number of isolated values assumed by a discrete variable. In case of a continuous variable, find the number of class intervals using the relation, No. of class Interval X class length $\cong$ Range.

  (c) Present the class or class interval in a table known as frequency distribution table.

  (d) Apply 'tally mark' i.e. a stroke against the occurrence of a particulars value in a class or class interval.

  (e) Count the tally marks and present these numbers in the next column, known as frequency column, and finally check whether the total of all these class frequencies tally with thetotal number of observations.

Set B 14.34

1. Out of 1000 persons, 25 per cent were industrial workers and the rest were agricultural workers. 300 persons enjoyed world cup matches on TV. 30 per cent of the people who had not watched world cup matches were industrial workers. What is the number of agricultural workers who had enjoyed world cup matches on TV?

(a) 260 (b) 240 (c) 230 (d) 250

2. A sample study of the people of an area revealed that total number of women were 40% and the percentage of coffee drinkers were 45 as a whole and the percentage of male coffee drinkers was 20. What was the percentage of female non-coffee drinkers?

(a) 10 (b) 15 (c) 18 (d) 20

3. Cost of sugar in a month under the heads raw materials, labour, direct production and others were 12, 20, 35 and 23 units respectively. What is the difference between the central angles for the largest and smallest components of the cost of sugar?

(a) 72^{o} (b) 48^{o} (c) 56^{o} (d) 92^{o}

4. The number of accidents for seven days in a locality are given below :

No. of accidents :	0	1	2	3	4	5	6
Frequency :	15	19	22	31	9	3	2

What is the number of cases when 3 or less accidents occurred?

(a) 56 (b) 6 (c) 68 (d) 87

7. Find the number of observations between 250 and 300 from the following data :

Value	: More than 200	More than 250	More than 300	More than 350
No. of observations :	56	38	15	0

(a) 56 (b) 23 (c) 15 (d) 8

1. In a study about the male and female students of commerce and science departments of a college in 5 years, the following datas were obtained :

1995	2000
70% male students	75% male students
65% read Commerce	40% read Science
20% of female students read Science	50% of male students read Commerce
3000 total No. of students	3600 total No. of students.

After combining 1995 and 2000 if x denotes the ratio of female commerce student to female Science student and y denotes the ratio of male commerce student to male Science student, then

(a) $x = y$ (b) $x > y$ (c) $x < y$ (d) $x \geq y$

2. In a study relating to the labourers of a jute mill in West Bengal, the following information was collected.

'Twenty per cent of the total employees were females and forty per cent of them were married. Thirty female workers were not members of Trade Union. Compared to this, out of 600 male workers 500 were members of Trade Union and fifty per cent of the male workers were married. The unmarried non-member male employees were 60 which formed ten per cent of the total male employees. The unmarried non-members of the employees were 80'. On the basis of this information, the ratio of married male non-members to the married female non-members is

(a) 1 : 3 (b) 3 : 1 (c) 4 : 1 (d) 5 : 1

THANK YOU