CHAPTER

13

Statistical Description of Data and Sampling

Unit-1: Statistical Description of Data

TOPICS TO COVER

- Introduction of Statistics
- Collection of Data
  - Primary Data
  - Secondary Data
- Presentation of Data
  - Frequency Distribution
  - Cumulative Frequency
- Presentation of Data Graphically
  - Line Chart or Histogram
  - Frequency Polygon
  - Pie Chart

WHAT ARE STATISTICS AND WHY WE NEED TO STUDY IT?

Statistics refers to the collection, analysis, interpretation, presentation, and organization of data. It involves using mathematical techniques to draw meaningful conclusions and make informed decisions based on data. The study of statistics is important because it provides us with the tools and methods to analyze and understand data, which is crucial in various fields.

Where have we heard this word Statistics?

We often encounter the word "statistics" in different contexts. **E.g.:**, we may hear it when referring to the weather report, which is based on statistical analysis of historical weather data to predict future conditions. Similarly, when someone mentions that the Indian population will be the highest by 2025, this statement is likely based on statistical projections and demographic data analysis. Furthermore, when discussing business goals, such as the next year's target for a company, statistics can be used to analyze past performance and set realistic objectives.

The term "statistics" has different origins, including the Latin word "status," the Italian word "statista," the German word "statistik," and the French word "statistique." However, its exact origin remains uncertain.

MEANING OF STATISTICS

When defined in Plural sense: Statistics refer to the collection of qualitative and quantitative data in a way that allows us to draw meaningful conclusions from it.

When defined in Singular sense: Statistics represent a scientific method used to collect, organize, and analyze data, ultimately leading to drawing statistical inferences about important characteristics. It can be considered the "science of counting" or the "science of averages."

APPLICATION OF STATISTICS

Statistics is applicable in every field where understanding and analysis of data are crucial. In the context of commerce, economics, business management, and industry, statistics provides the means to analyze economic data, make informed decisions, develop theories, conduct surveys, project future outcomes, and maximize profit. It offers various tools and techniques for data analysis, allowing businesses to gain a competitive advantage and make effective strategies based on statistical insights.

ECONOMICS

1. **Many branches of economics,** such as time series analysis, index number analysis, and demand analysis, are essentially branches of statistics. These branches use statistical methods to analyze economic data and draw meaningful conclusions.
2. **Econometrics** is another important application of statistics in economics. It involves the quantitative application of statistical and mathematical models using data to develop theories or test existing hypotheses in economics.
3. **Socio–economic surveys** play a crucial role in understanding various socio–economic factors. Statistical methods are used to conduct these surveys and analyze the derived data.
4. Regression analysis is used in economics to project future outcomes, such as sales, production, and prices, based on statistical analysis of past data.

BUSINESS MANAGEMENT

1. **How do managers work today?** In modern business management, decisions are no longer based solely on instincts. Managers now rely on quantitative data analysis, combined with statistical methods and operational research techniques, to make calculative and informed decisions.
2. **Sampling helps us to create criteria for making strategy**. In complex business environments like e–commerce, statistics plays a vital role in formulating strategies at every point by analyzing relevant data.
3. **Statistical decision theory** is employed to analyze complex strategies and assist businesses in making fruitful decisions by considering the merits and demerits of different options.

STATISTICS IN COMMERCE AND INDUSTRY:

1. **Getting an Edge in a competitive business environment.** In the competitive business environment, statistics can provide a significant edge. Businesses collect, compare, and analyze data on sales, production, profit, and other relevant factors to gain insights and formulate effective strategies.

2. **To Maximize Profit:** Maximizing profit is a common goal in commerce and industry. Statistical analysis is used to examine data on previous sales, wages, raw materials, and competitor products. By comparing and analyzing this data, businesses can identify opportunities to maximize profit.
3. **Various tools used in Commerce and Industry of Statistics,** including measures of central tendency (such as mean, median, and mode), measures of dispersion (such as range and standard deviation), sampling techniques, correlation and regression analysis, and index number and time series analysis. These tools help in analyzing data, identifying patterns, and making informed decisions.

LIMITATIONS OF STATISTICS

1. **Study of Quantitative data only:** Statistics studies only such facts as can be expressed in numerical terms. It does not study qualitative phenomena like honesty, friendship, wisdom, health, patriotism, justice, etc.
2. **Study of Aggregates only:** Statistics studies only the aggregates of quantitative facts. It does not study statistical facts relating to any particular. **E.g.:** It may be a statistical fact that your class teacher earns 50,000 per month. But if you want to find statistics of average salary of different schools, you will study aggregates of salary of each school
3. **Homogeneity of Data, an essential Requirement:** To compare data, it is essential that statistics are uniform in quality. Data of diverse qualities and kinds cannot be compared. For **E.g.:** we cannot compare the statistics of banana and air pressure in tyre
4. **Results are True only on an Average:** Most statistical findings are true only as averages. They are not always valid under all conditions. For instance, if it is said that per capita income in India is 50,000 per annum, it does not mean that the income of each and every Indian is 50,000 per annum. Some may have more and some may have less.
5. **Results may Prove to be Wrong:** In order to understand the conclusions precisely, it is necessary that the circumstances and conditions under which these conclusions have been drawn are also studied. Otherwise, they may prove to be wrong.
6. **Can be used only by the Experts:** Statistics can be used only by those persons who have special knowledge of statistical methods. Those who are ignorant about these methods cannot make sensible use of statistics. In the words of Yule and Kendall, "Statistical methods are the most dangerous tools in the hands of an inexpert.

COLLECTION OF DATA

On the nature of data, we can define data as follows:

- **Quantitative Data:** This refers to data that represents numeric values of quantitative data include measurements, counts, and numerical ratings.
- **Qualitative Data:** This refers to data that represents characteristics or qualities of qualitative data include descriptive information, opinions, and categorical variables.

  If we want to analyze qualitative information using statistics, it needs to be converted into quantitative information by assigning a numeric description to the given characteristic. Quantitative data is collected in the form of variables.

WHAT IS VARIABLE?

A variable is a characteristic or attribute that can take on different values. It is the unit of measurement for collecting data.

TYPES OF VARIABLES

1. Discrete Variable
2. Continuous Variable

- **Discrete Variable:** A discrete variable is one that can only take on specific, countable values. **E.g.:** the number of children in a family, the number of students in a class, or the number of cars in a parking lot.
- **Continuous Variable:** A continuous variable is one that can take on any value within a given interval or range. It can have decimal values and is not limited to specific points. **E.g.:** height, weight, temperature, or time.

On the basis of source of data, we can define data as follows:

- **Primary Data:** Primary data refers to data collected by the investigator for their own purpose, from the beginning to the end of the research.
  - These are collected from the source of origin.
  - Primary data is considered original and specific to the research being conducted of primary data collection methods include surveys, interviews, experiments, and observations.
- **Secondary Data:** Secondary data refers to data that already exists and has been collected for some other purpose.
  - These data are considered second–hand as they have been collected by someone else.
  - Secondary data can be obtained from published or unpublished reports, articles, books, databases, or other sources of secondary data include government reports, academic studies, industry statistics, and historical records.

COLLECTION OF PRIMARY DATA CAN BE DONE BY 4 WAYS

- Interview method
- Mailed questionnaire method
- Observation method
- Questionnaires filled and sent by enumerators.

INTERVIEW METHOD

- Direct Interview Method or Personal Interview Method
- Indirect Interview Method
- Telephonic Interview Method

This method involves direct interaction between the investigator and the respondent. There are different types of interview methods, such as direct interview, indirect interview, and telephonic interview. Direct interview provides more accurate data, while telephonic interview allows for faster data collection over a wider area. However, the telephonic interview method may have a higher number of non–responses.

E.g.: Let's say a school wants to gather feedback from its students regarding their satisfaction with the school facilities and extracurricular activities. They decide to use the interview method to collect data directly from the students.

Note:

- In the first two methods data collected will be more accurate but if you have to collect data faster and at a wider area then you need to collect data by Telephonic Interview method.
- The number of non–responses is maximum for this third method of data collection

MAILED QUESTIONNAIRE METHOD

In this method, questionnaires are mailed to the informants. The questionnaire is accompanied by a letter explaining the purpose of the enquiry and ensuring the confidentiality of the information. The informants fill out the questionnaires and return them to the investigator. This method allows for covering a wide area, but it may also result in a higher number of non–responses.

E.g.: Suppose a local government wants to gather feedback from residents about their satisfaction with public transportation services in the city. They decide to use the mailed questionnaire method to reach a wide range of residents.

OBSERVATION METHOD

This method involves direct observation by the investigator to collect information. It is considered one of the best methods for data collection, but it can be time–consuming and covers only a small area.

E.g.: A researcher is interested in studying the playground behavior of children in a local park. They choose to use the observation method to gather data directly by observing the children during their playtime.

QUESTIONNAIRES FILLED AND SENT BY ENUMERATORS

Under this method, enumerators are appointed to approach the informants and fill out the questionnaires. This method is suitable for covering a wider range of respondents but can be costly.

E.g.: A market research company wants to collect data on consumer preferences for a new product. They decide to use the questionnaires filled and sent by enumerators method to gather information from a diverse group of respondents.

SOURCES OF SECONDARY DATA

Secondary data refers to data that has been previously collected by someone else for a different purpose but can be utilized for the current research or analysis. Here are some important sources of secondary data.

There are many sources of getting secondary data. Some important sources are listed below:

- **International sources:** These include organizations like the World Health Organization (WHO), International Labour Organization (ILO), International Monetary Fund (IMF), World Bank, and others. These global institutions collect and provide data on various topics such as health, labor, economy, and development.

- **Government sources:** Governments often collect extensive data on various aspects of society and the economy. Statistical agencies, such as the Central Statistical Office (CSO) in different countries, publish reports and statistical abstracts that provide valuable data. Ministries and departments related to agriculture, education, labor, and more also release official statistics.
- **Private and quasi–government sources:** Private organizations and quasi–governmental institutions also generate and maintain datasets relevant to specific sectors. These sources can include research institutes, industry associations, academic institutions, and specialized agencies.
- **Unpublished sources:** Researchers and experts often conduct studies or collect data that may not be widely available. These unpublished sources can include research papers, reports, surveys, and studies conducted by various institutions, researchers, or organizations.

SCRUTINY OF DATA

Scrutiny of data refers to the process of carefully examining the collected data to ensure its accuracy, reliability, and consistency. It involves reviewing the data for any errors, inconsistencies, or outliers that may affect the validity of the analysis.

Researchers scrutinize the data by performing data cleaning and validation procedures. This includes checking for missing values, outliers, data entry errors, and logical inconsistencies. By carefully reviewing the data, researchers can ensure the quality and integrity of the dataset before proceeding with analysis.

PRESENTATION OF DATA

- After collecting and verifying the quality of data, it should be presented in a clear and concise manner.
- Effective data presentation is important for effectively communicating findings and facilitating understanding.
- Common methods of data presentation include tables, charts, graphs, and visualizations.
- Researchers select the appropriate presentation format based on the nature of the data and the research objectives.
- The presentation should emphasize the essential features of the data to make it easier for the audience to interpret and derive meaningful insights.

CLASSIFICATION OF DATA OR ORGANIZING OF DATA

Classification of data involves organizing raw data into meaningful groups or classes based on their characteristics. This process enables researchers to draw conclusions and identify patterns or relationships within the data.

Data can be classified based on various attributes such as age groups, income brackets, geographical regions, or product categories. By categorizing data into relevant classes, researchers can analyze and interpret data more effectively.

E.g.: A market researcher conducting a survey on customer preferences for a product might classify the responses into different age groups (e.g., 18–25, 26–35, 36–45) to understand how preferences vary across different demographic segments.

Classification of data helps in summarizing and simplifying complex data sets, making it easier to identify trends, make comparisons, and draw meaningful conclusions.

OBJECTIVES OF CLASSIFICATION

1. **Simplification and Briefness:** Classification aims to simplify complex data by grouping them into categories or classes. It provides a condensed and organized representation of data, making it easier to analyze and interpret.
2. **Comparability:** Classification enables the comparison of data within and across different categories or classes. It allows for identifying patterns, trends, and relationships between different groups of data.
3. **Statistical Analysis:** Classification facilitates statistical analysis by organizing data into meaningful groups. It enables researchers to apply various statistical techniques and methods to explore the characteristics and relationships within each category or class.
4. **Makes data more understandable:** Classification makes data more understandable by presenting them in a structured manner. It provides a clear framework for data interpretation and helps in deriving insights and making informed decisions.

DATA MAY BE CLASSIFIED AS

1. **Chronological Data or Temporal or Time Series:** This classification arranges data based on their time intervals or chronological order. It helps in studying patterns and trends over time.

   E.g.: Classifying monthly sales data of a product over the past year into different time intervals such as quarters or seasons.

2. **Geographical or Spatial Series Data:** This classification groups data based on their geographical or locational differences. It helps in analyzing variations across different regions or areas.

   E.g.: Classifying population data of different cities or states into regional groups such as North, South, East, and West.

3. **Qualitative or Ordinal Data:** This classification categorizes data based on their qualities or attributes. It involves assigning data to specific categories or classes based on subjective characteristics.

   E.g.: Classifying survey responses into categories such as "Satisfied," "Neutral," and "Dissatisfied" based on customer satisfaction levels.

4. **Quantitative or Cardinal Data:** This classification organizes data into classes or groups based on their numerical values. It involves creating intervals or ranges to represent different levels or quantities.

   E.g.: Classifying test scores of students into different grade ranges such as A, B, C, etc., based on their numerical value ranges.

FREQUENCY DATA

Frequency of a particular data value is the number of times the data value occurs and data in which we can count frequency is called as frequency data

E.g.: Recording the number of times a specific word appears in a document or the number of customers who visited a store on different days of the week.

NON–FREQUENCY DATA

Non–frequency data refers to data where the individual values and their specific identities are important and need to be preserved.

E.g.: Recording the names and addresses of customers in a database for a marketing campaign, where each customer's information is unique and distinct.

THERE ARE GENERALLY THREE FORMS OF PRESENTATION OF DATA

Textual or Descriptive Presentation

In the textual presentation, data is described using paragraphs of text. This method is commonly used in official reports, where the activities, plans, or programs of a project are explained in detail, with relevant facts and figures inserted within the text. This form of presentation is suitable when the amount of data is relatively small.

However, statisticians generally prefer other methods as it can be dull, monotonous, and difficult to compare different fields of data.

E.g.: A project report describing the progress, achievements, and future plans of a company, including relevant statistical data within the narrative.

Tabular Presentation: Data is organized in rows and columns in a tabular presentation. A statistical table provides a systematic and structured representation of data in a concise format.

Considerations when creating a table:

- Designing of Table: Ensuring a clear and organized layout with appropriate column headings and row labels.
- Comparable Data: Presenting data in a format that allows for easy comparison and analysis.
- Beautiful Presentation: Enhancing the visual appeal of the table through proper formatting, font usage, and gridlines.
- Descriptive Notes: Including explanatory notes or footnotes to provide additional context or clarify any assumptions made.

Here's an example of a tabular presentation depicting the sales performance of a company across different product categories:

Product Category	Q1 Sales (in INR)	Q2 Sales (in INR)	Q3 Sales (in INR)
Electronics	5,00,000	4,50,000	5,50,000
Apparel	3,50,000	4,00,000	3,80,000
Home Goods	2,80,000	3,20,000	3,00,000
Beauty	2,00,000	1,80,000	2,20,000

DESCRIPTION OF EACH CONSIDERATION

1. **Designing of Table:** The table is designed with clear column headings ("Product Category," "Q1 Sales," "Q2 Sales," "Q3 Sales") and appropriate row labels for each product category.
2. **Comparable Data:** The data is presented in a format that allows easy comparison of sales across different quarters for each product category. This enables analysis and identification of trends or variations.

3. **Beautiful Presentation:** The table is properly formatted, with consistent alignment, appropriate font usage, and visible gridlines, making it visually appealing and easy to read.
4. **Descriptive Notes:** No specific descriptive notes are provided in this example, but they could be included below the table to explain any abbreviations, define terms, or provide additional context if needed.

FORMAT OF TABLE

Table Number: _____

Title: _____

Stub (Row Heading)	Caption (Column Heading)				Total (Rows)
	Sub-head		Sub-head		
	Column-head	Column-head	Column-head	Column-head	
Stub Entries (Row Entries)					
• • •					
• • •		BODY			
• • •					
• • •					
Total (Columns)					

Source Note:

Footnote:

Main Parts of Table:

- **Table number:** It's like an ID for the table, so you can refer to it easily.
- **Title:** Think of it as a short summary that tells you what the table is about.
- **Headnote:** Gives you some extra info or context right at the beginning.
- **Captions/Column Heading:** Labels for columns, helping you understand what each column represents.
- **Stubs/Row Heading:** These are like labels for rows, making it clear what each row is talking about.
- **Body of table:** This is where the main information is – the numbers or data you're interested in.
- **Source note:** Tells you where the information in the table comes from.
- **Footnote:** Extra details or explanations found at the bottom of the table.

The tabulation method is generally preferred over textual presentation because

1. It allows for easy comparison of data across different categories or variables.
2. It provides a structured and organized format that facilitates data analysis.

3. It presents data in a concise and condensed manner, making it easier to grasp key information.
4. It allows for the inclusion of mathematical calculations and statistical measures within the table.

DIAGRAMMATIC REPRESENTATION OF DATA

Diagrammatic presentation of data is a highly effective method of visually representing information in a concise manner. It involves the use of various types of diagrams to convey data effectively. In this context, we will focus on three commonly used diagrams.

LINE DIAGRAM OR HISTORIAGRAM

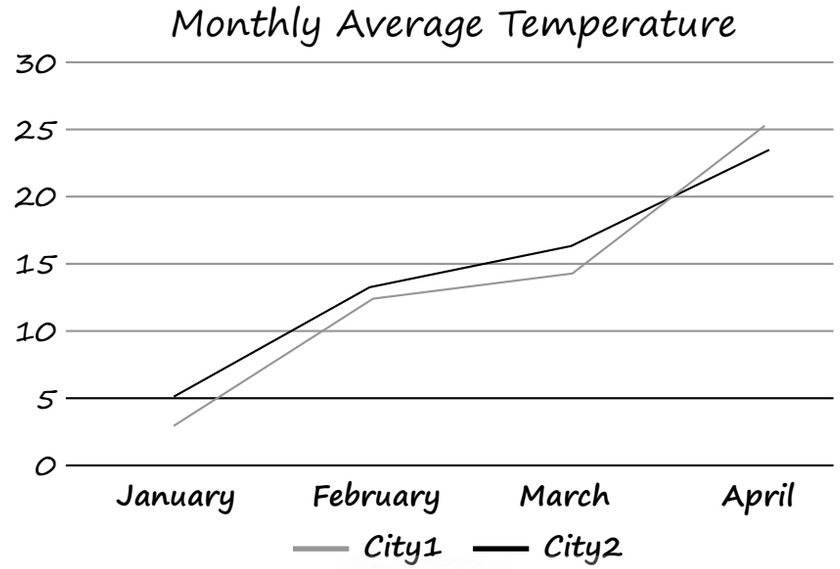
A line diagram, also known as a historiagram, is used to represent data that changes over time. It involves plotting pairs of values on a graph. By connecting these points with lines, we can visualize the trend or pattern of the data over time.

- If we have to represent multiple data (similar data in same unit) varying with time, we can use Multiple line chart.
- If we have to represent multiple data (with different unit) varying with time, we can use Multiple axis chart.

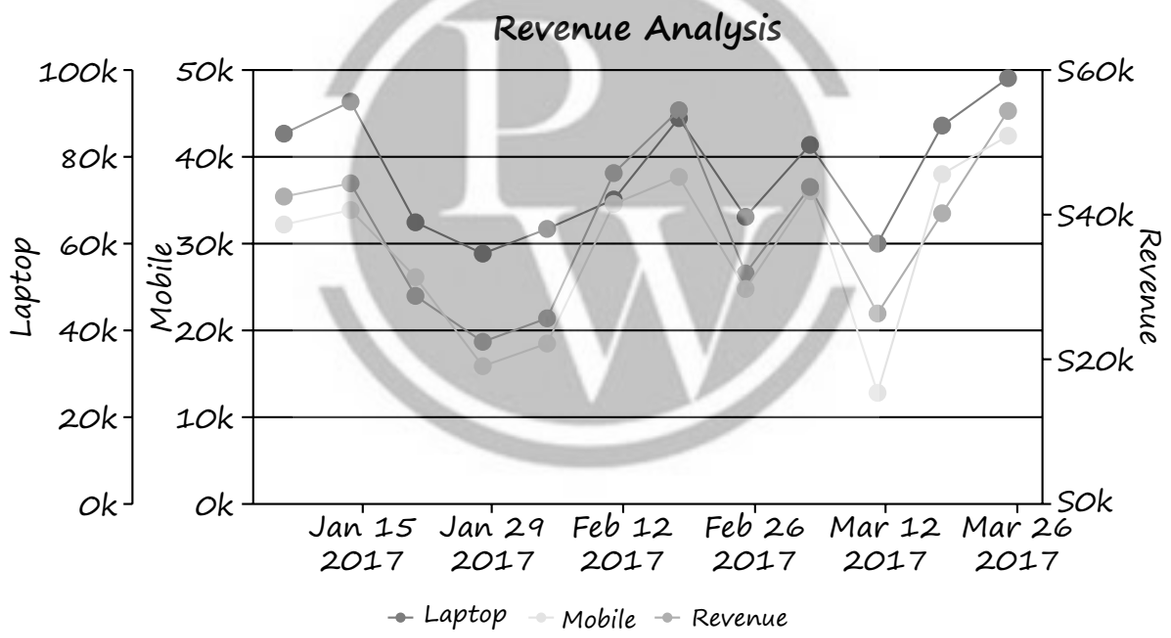
  E.g.:

  Line Chart: Suppose we have data on the monthly product (Black dress) sale in the year 2022. We can use a line diagram to plot the sale (y-axis) against the corresponding months (x-axis), allowing us to observe the sale trend over time.

Multiple Line chart:



Multiple Axis Chart:

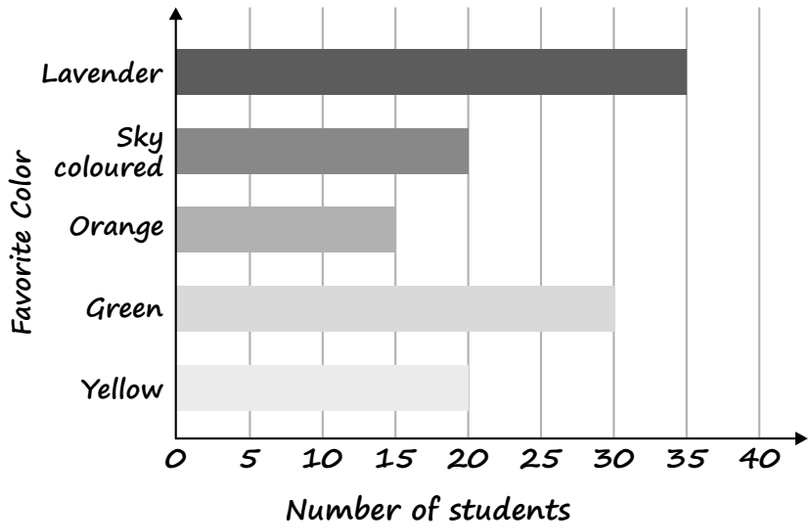


BAR DIAGRAM

A bar diagram is a popular choice for comparing different categories or variables. It involves the use of bars to represent the magnitude of each category or variable. There are different types of bar diagrams depending on their orientation and purpose.

- **Horizontal Bar Diagram:** In a horizontal bar diagram, the categories or variables are represented on the y–axis, while the corresponding values or frequencies are shown on the x–axis. The lengths of the horizontal bars represent the magnitude of the variables.

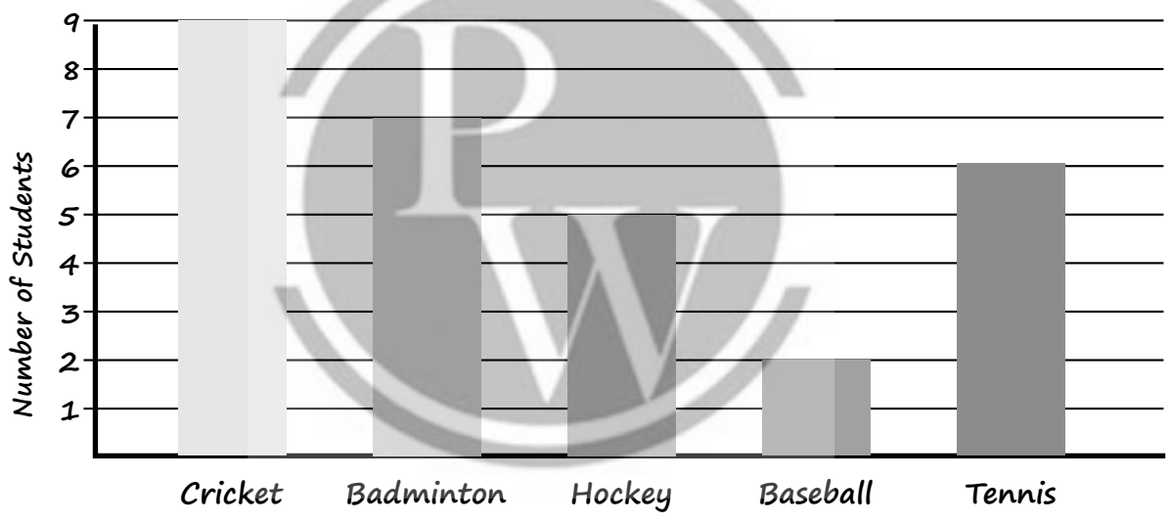
E.g.:



Note: It is used for Qualitative data or data varying over space.

- **Vertical Bar Diagram:** In a vertical bar diagram, the categories or variables are represented on the x–axis, while the corresponding values or frequencies are shown on the y–axis. The lengths of the vertical bars represent the magnitude of the variables.

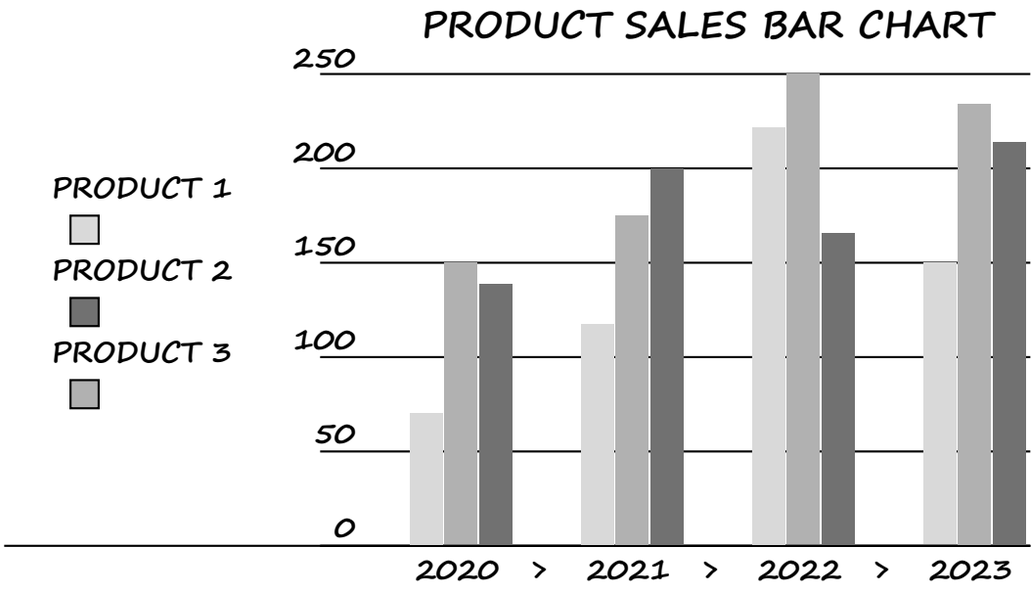
E.g.:



Note: It is used for Quantitative data or time series data.

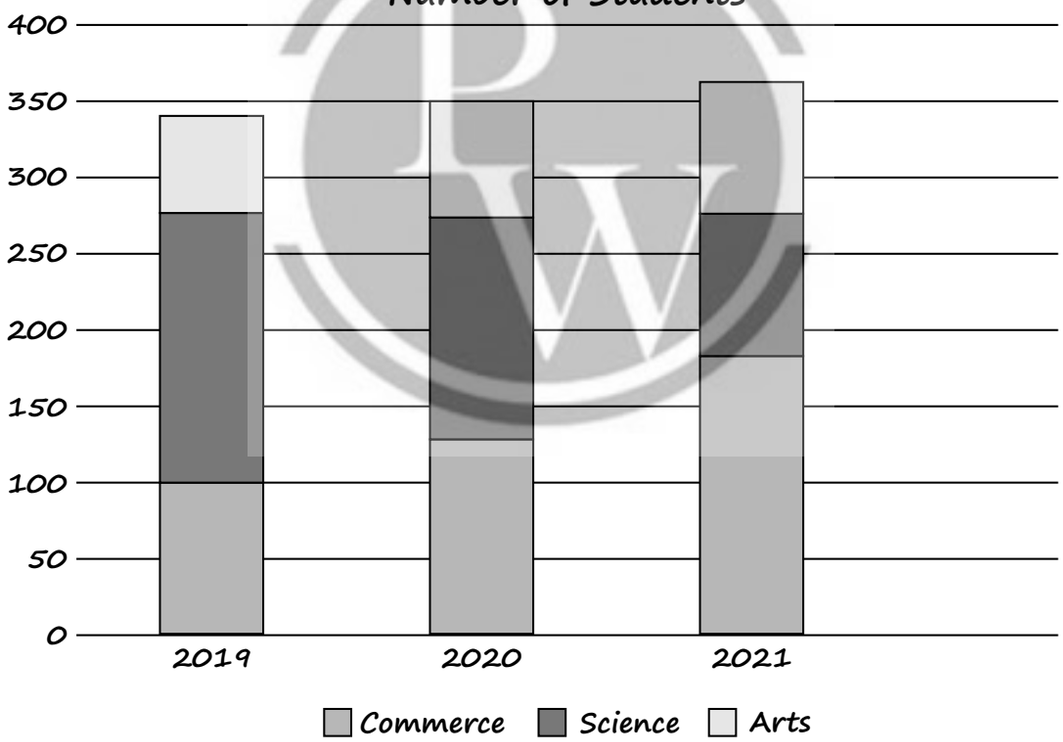
- **Multiple or Grouped Bar Diagrams:** Multiple or grouped bar diagrams are used when we need to represent multiple datasets that have the same unit and vary with time. It allows for easy comparison between the different datasets.

E.g.:

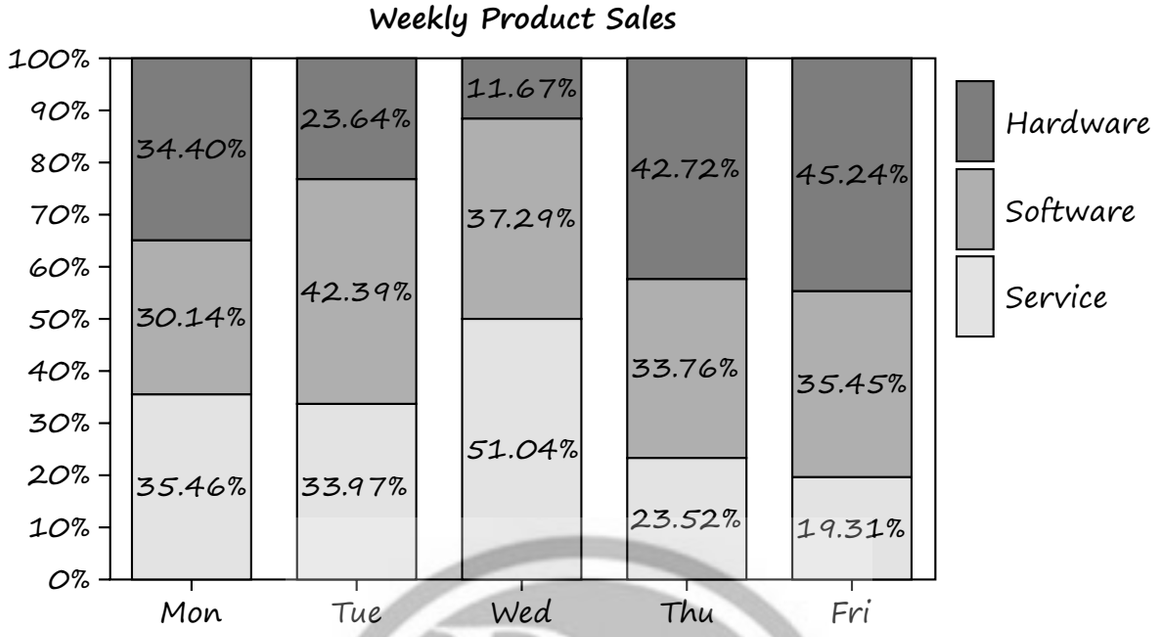


- **Component or Sub–divided Bar Diagrams:** Component or sub–divided bar diagrams are used when we need to represent multiple datasets that have different units but still vary with time. This type of diagram uses multiple axes to accommodate the different units.

E.g.:

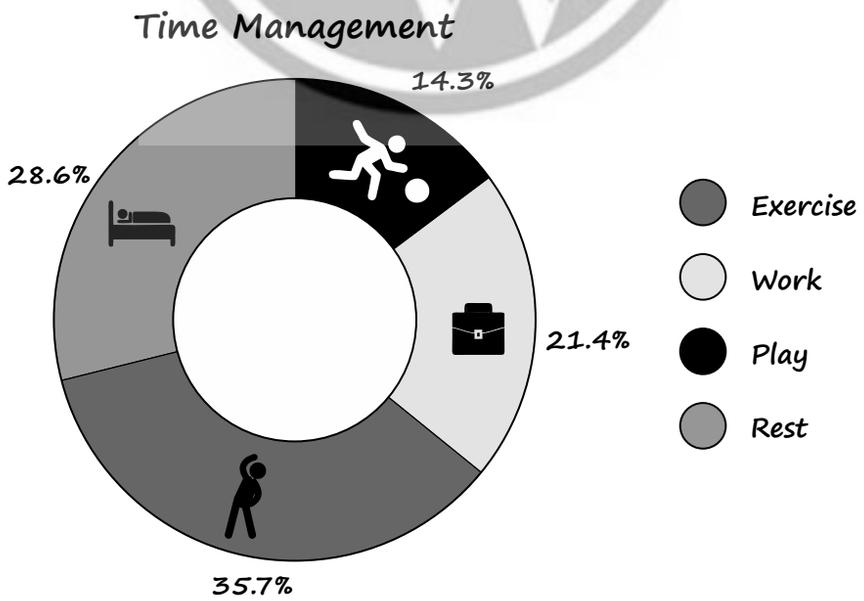


Divided Bar charts or Percentage Bar diagrams: Comparing different components of a variable and also the relating of the components to the whole.



- **Pie Chart:** A Pie chart is used to represent data as a circular chart divided into sectors. Each sector represents a specific category or variable, and its size or angle represents the proportion or percentage of that category. Pie charts are especially useful for illustrating proportions or composition.

  E.g.: The below Pie chart represents the time management. The size of each sector in the pie chart corresponds to the proportion of time dedicated to that specific category, providing a clear visual overview of how time is distributed across these four essential aspects of life.



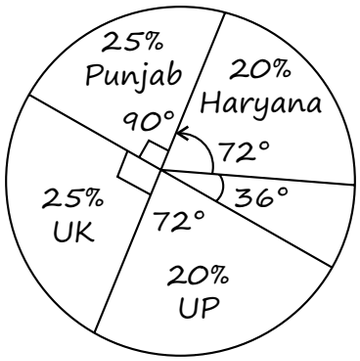
E.g.: The production of Wheat by different States of India are as shown below:

State	Production
Haryana	20%
Punjab	25%
Uttarakhand	25%
U.P	20%
Others	10%
Total	**100%**

Draw the suitable diagram to represent the information.
Here, according to the given data, we have

To find the corresponding percentage, we have $\frac{\text{Row value}}{\text{Total}} \times 360°$

State	Production	
Haryana	20%	$\frac{20}{100} \times 360° = 72°$
Punjab	25%	$\frac{25}{100} \times 360° = 90°$
Uttarakhand	25%	$\frac{25}{100} \times 360° = 90°$
U.P	20%	$\frac{20}{100} \times 360° = 72°$
Others	10%	$\frac{10}{100} \times 360° = 36°$
Total	**90**	**360°**



In summary, diagrammatic representation of data provides an effective way to present information visually. Line diagrams show trends over time, bar diagrams compare variables, and pie charts depict proportions or compositions.

Example 1. 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° (b) 48° (c) 56° (d) 92°

Sol. (d) According to the question,

Total cost of sugar = (12 + 20 + 35 + 23) units = 90 units

Now,

Value of largest component = 35 units

Value of smallest component = 12 units

$\Rightarrow$ Difference between largest and smallest components

= 35 – 12 = 23

So, the required central angle $= \frac{23}{90} \times 360° = 92°$

Hence, the correct option is (d) i.e. 92°.

Example 2. A pie diagram is used to represent the following data.

Source	Customers	Excise	Income Tax	Wealth Tax
Revenue in Millions	120	180	240	180

The central angles corresponding to Excise is

(a) 120° (b) 80° (c) 90° (d) 60°

Sol. (c) According to the question,

Total Revenue = 120 + 180 + 240 + 180 = 720

Central Angle will be given by the formula,

$= \frac{\text{Revenue of the Class}}{\text{Total Revenue}} \times 360°$

Thus, Central Angle corresponding to Excise

$= \frac{180}{720} \times 360° = 90°$

Hence, the correct answer is option (c).

Example 3. Mode of a distribution can be obtained from

(a) Histogram (b) Less than type ogives

(c) More than type ogives (d) Frequency polygon

Sol. (a) We know that,

The mode is the value that occurs most in a given data set.

The highest peak or the longest bar of the histogram represents the location of the mode of the given data set.

Example 4. The most appropriate diagram to represent the data relating to the monthly expenditure on different items by a family is

(a) Histogram (b) Pie-diagram

(c) Frequency polygon (d) Line graph

Sol. (b) Pie Diagram is the most appropriate diagram to represent the data' relating to the monthly expenditure on the different items by the family. It is divided into the different sectors which helps us to understand the numerical proportion of each item.

Hence, the correct option is (b) i.e. Pie-diagram.

Example 5. The column headings of a table are known as:

(a) Body (b) Stub (c) Box-head (d) Caption

Sol. (d) The table under consideration is divided into caption, Box-head, Stub and Body where,

Caption is the upper part of the table, describing the columns and sub-columns, if any.

Hence, the correct answer is option (d).

FREQUENCY DISTRIBUTION

- Frequency distribution is a method of organizing data to provide insights into how often certain values occur.
- It can be classified in two ways: as frequency data classified by categories or as frequency data classified by intervals.

  Frequency represents the number of times an observation occurs in the data.

  E.g.: in the list of numbers 1, 2, 3, 4, 6, 9, 9, 8, 5, 1, 1, 9, 9, 0, 6, 9, the frequency of the number 9 is 5.

  Frequency distribution involves tabulating data in a table, where the total frequency is distributed among different classes or intervals. The classes or intervals should be mutually exclusive and cover the entire range of data.

UNGROUPED FREQUENCY DISTRIBUTION

When the data consists of discrete variables, we can create an ungrouped frequency distribution table. This table presents the frequency counts for each unique value in the dataset.

E.g.: Suppose we have a dataset representing the number of pets owned by individuals: 2, 4, 1, 3, 2, 4, 2, 1, 3, 4. The ungrouped frequency distribution table would show the frequency count for each unique value:

Number of Pets	Frequency
1	2
2	3
3	2
4	3

GROUPED FREQUENCY DISTRIBUTION

When the data consists of continuous variables, we can create a grouped frequency distribution table. This table categorizes the data into intervals or classes and displays the frequency count for each interval.

E.g.: Suppose we have a dataset representing the heights (in centimeters) of a group of individuals: 156, 168, 174, 160, 162, 170, 168, 172, 158, 164. To create a grouped

frequency distribution, we can group the data into intervals (e.g., 150–160, 161–170, 171–180) and count the frequency in each interval:

Heights (in cm)	Frequency
150 – 160	2
160 – 170	5
170 – 180	3

HOW TO MAKE A FREQUENCY DISTRIBUTION TABLE FOR DISCRETE VARIABLE

I. Find the largest and smallest observations and obtain the difference between them.

II. Form a number of classes depending on the number of isolated values assumed by a discrete variable.

III. Present the class 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.

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.

E.g.: Let's say you survey a number of households and find out how many pets they own. The results are 3, 0, 1, 4, 4, 1, 2, 0, 2, 2, 0, 2, 0, 1, 3, 1, 2, 1, 1, 3. Let's distribute data in frequency table.

Number of Pets	Tally Marks	Frequency
0	|||	4
1	𝍸 |	6
2	𝍸 ||	5
3	|||	3
4	||	3

How to make a frequency distribution table for continuous variable:

Creating a Frequency Distribution Table for Discrete Variables

To create a frequency distribution table for a discrete variable, follow these steps:

1. Identify the largest and smallest observations in the dataset and calculate the range by finding the difference between them.
2. Determine the number of classes or categories to be used in the frequency distribution table. This depends on the number of distinct values assumed by the discrete variable.
3. Present the classes in a table, known as the frequency distribution table. The classes should cover the entire range of values and be mutually exclusive.

4. Use tally marks (vertical strokes) to represent the occurrence of each value within a class. Each value gets a tally mark against the corresponding class.
5. Count the tally marks and record the frequencies in the next column of the frequency distribution table. This column represents the frequency of each class.
6. Finally, verify whether the total of all class frequencies matches the total number of observations in the dataset.

Example 6. Following are the heights (in cm) of B. Com students of St. Xavier's College.

161, 150, 154, 165, 168, 161, 154, 162, 150, 151, 162, 164, 171, 165, 158, 154, 156, 172, 160, 170, 153, 159, 161, 170, 162, 165, 166, 168, 165, 164, 154, 152, 153, 156, 158, 162, 160, 161, 173, 166, 161, 159, 162, 167, 168, 159, 158, 153, 154, 159

Construct a frequency distribution of heights, taking class length as 5.

Sol. Here,

Smallest weight: 150 cm

Largest weight: 173 cm

Range: 173−150=23

$$\text{Number of classes} = \frac{\text{Range}}{\text{Class length}} = \frac{23}{5} = 4.6 \text{ (round up to 5 classes)}$$

Therefore, the frequency distribution table is as follow:

Heights (in cm)	Frequency
150 – 154	12
155 – 159	9
160 – 164	14
165 – 169	10
170 – 174	5
	50

SOME IMPORTANT TERMINOLOGIES

- **Class Limit:** The class limit refers to the minimum and maximum values that define a class interval. The lower–class limit (LCL) is the minimum value, and the upper–class limit (UCL) is the maximum value within a class interval.

  E.g.: in the class interval 10–20, 10 is the lower–class limit, and 20 is the upper–class limit.
- **Class boundary:** Class boundaries are used to establish a common boundary between adjacent class intervals when there is a difference between the upper class limit of one interval and the lower class limit of the next interval. By using class boundaries, we ensure a smooth transition and maintain consistency in data representation.

- Please note that for overlapping class interval– Class boundaries and Class limits are same.
- **Lower class boundary (LCB)** $= LCL - \frac{D}{2}$
- **Upper class boundary (UCB)** $= UCL - \frac{D}{2}$ where, D is difference between the LCL of the next class interval and UCL of the given class interval.

  E.g.: Let's consider the following class intervals for a dataset: 10–15,15–20,20–25,25–30.

  Here, the class intervals are overlapping thus Class boundaries and Class limits are same.

  E.g.: Let's consider the following class intervals for a dataset: 10–14, 15–19, 20–24, 25– 29.

  Here, the class intervals are non–overlapping.

  Thus, first class limits: 10–15

  And class boundaries of first class:

  $LCB = 10 - \frac{1}{2} = 9.5$, $HCB = 14 + \frac{1}{2} = 14.5$

  Thus, the first class boundary is 9.5–14.5.
- **Class Midpoint:** The class midpoint (or class mark) is a specific point in the center of the class interval in a frequency distribution table.

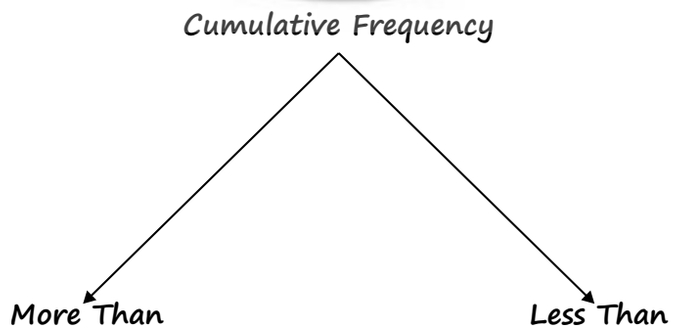
  Mid–Point = (LCL + UCL) / 2 = (LCB + UCB) / 2

  E.g.: For class interval 10–14

  $Mid\text{–}point = \frac{10 + 14}{2} = \frac{24}{2} = 12$
- **Width or Size of Class Interval:** The class width is the difference between the upper–class boundary to the lower–class boundary of consecutive classes

  E.g.: Width of the interval 10–14 is 14 – 10 = 4

Cumulative Frequency: Cumulative frequency is defined as a running total of frequencies.



LESS THAN CUMULATIVE FREQUENCY

Heights (in cm)	Frequency	Less than Cumulative frequency
150 – 154	12	12
155 – 159	9	12 + 9 = 21

Heights (in cm)	Frequency	Less than Cumulative frequency
160 – 164	14	21 + 14 = 35
165 – 169	10	35 + 10 = 45
170 – 174	5	45 + 5 = 50
	50	

MORE THAN CUMULATIVE FREQUENCY

Heights (in cm)	Frequency	More than Cumulative frequency
150 – 154	12	50
155 – 159	9	50 – 12 = 38
160 – 164	14	38 – 9 = 29
165 – 169	10	29 – 14 = 15
170 – 174	5	15 – 10 = 5
	50	

- **Frequency Density of Class Interval:** Frequency density is a measure that helps to standardize the frequencies of different class intervals. It is calculated by dividing the frequency of a class interval by its width or size. The frequency density gives us an idea of the concentration of data within each class interval, taking into account the varying widths of the intervals.

  Frequency Density = Frequency / Width
- **Relative Frequency or Percentage Frequency of Class Interval:** Relative frequency, also known as percentage frequency, is the proportion of the total frequencies that each class interval represents. It is calculated by dividing the frequency of a class interval by the total number of observations and multiplying by 100 to express it as a percentage.

  Formula: Relative Frequency = (Frequency / Total number of observations) × 100

GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTION

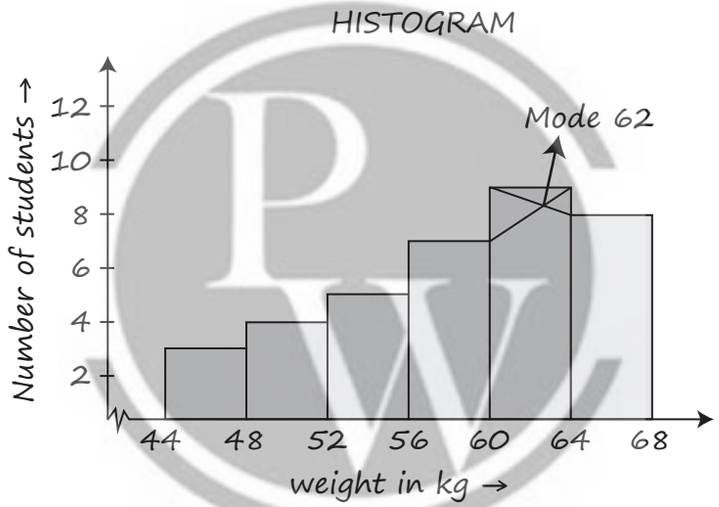
HISTOGRAM

- This is very convenient way to represent a frequency distribution. Comparison among class interval is possible in this mode of diagrammatic representation
- A two–dimensional graphical representation of a continuous frequency distribution is called a histogram.
- In histogram, the bars are placed continuously side by side with no gap between adjacent bars.
- That is, in histogram rectangles are erected on the class intervals of the distribution. The areas of rectangle are proportional to the frequencies.
- We can also find the mode from Histogram

Let's draw histogram for the following data:

FREQUENCY DISTRIBUTION OF WEIGHT OF 36 BBA STUDENT

Weight in kg (class interval)	Tally marks	No. of students (frequency)
44-48	𝍸	3
49-52	𝍸 I	4
52-56	𝍸 𝍸	5
56-60	𝍸 𝍸 II	7
60-64	𝍸 𝍸 𝍸 I	9
64-68	𝍸 𝍸 II	8
Total		36



FREQUENCY POLYGON

As name says, we have to plot frequencies on the graph. For single frequency distribution we can do it easily. By plotting points and joining the points with the line.

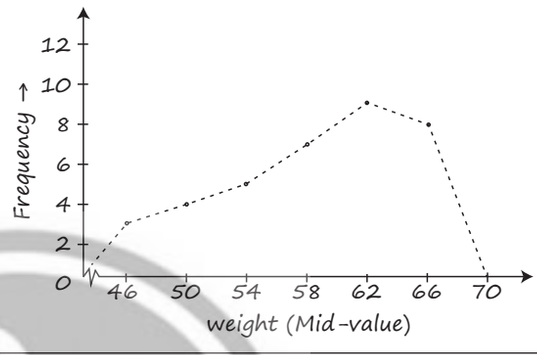
- But for grouped frequency distribution, frequency polygon can be drawn only if the class interval is uniform. Then we take the value of xi as the midpoint of the class intervals. And fi being the frequencies of the classes respectively we can draw frequency polygon by plotting the (xi, fi) and then joining it with the line.

  E.g.: Let's draw for the below table

FREQUENCY DISTRIBUTION OF WEIGHTS OF 36 BBA STUDENTS

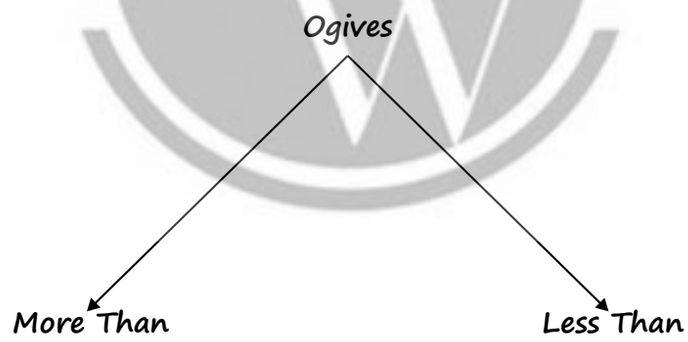
Weight in kg (class interval)	Tally marks	No. of students (frequency)
44-48	III	3
48-52	IIII	4

Weight in kg (class interval)	Tally marks	No. of students (frequency)
52-56	𝍸	5
56-60	𝍸 II	7
60-64	𝍸 IIII	9
64-68	𝍸 III	8
Total	-	36

Mid-point	Frequency	Frequency Polygon
46	3	
50	4	
54	5	
58	7	
62	9	
66	8	

OGIVES OR CUMULATIVE FREQUENCY GRAPH

By plotting cumulative frequency against the respective class boundary, we get ogives.
As Cumulative Frequency Graph can be drawn as less than type and more than type. Therefore



E.g.: Let's draw ogive for the below table
Frequency distribution of weights of 36 BBA Students

Weight in kg (class interval)	Tally marks	No. of students
44-48	III	3
49-53	IIII	4
54-58	𝍸	5
59-63	𝍸 II	7
64-68	𝍸 IIII	9

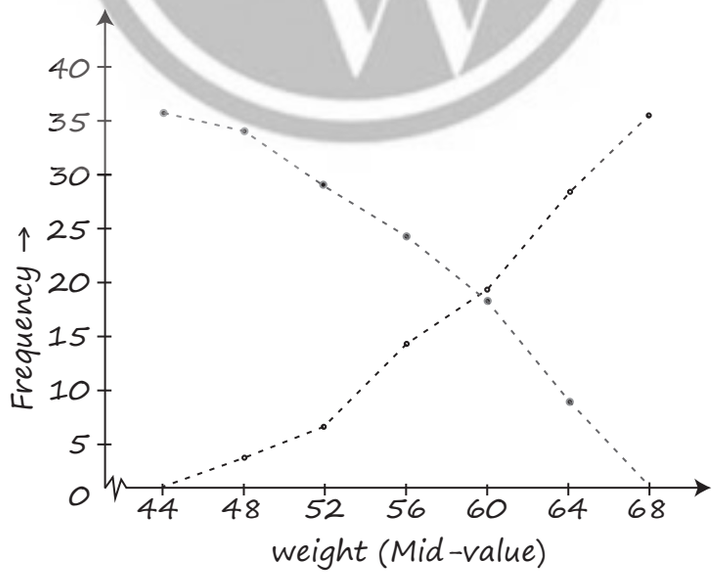
Weight in kg (class interval)	Tally marks	No. of students
69–73	𝍸 III	8
Total	–	36

Less than Ogive

Weightl)	No. of students	Commulative frequancy
Less than 48	3	3
Less than 52	4	7
Less than 56	5	12
Less than 60	7	19
Less than 64	9	28
Less than 68	8	36
Total	36	

More than Ogive

Weightl)	No. of students	Commulative frequancy
More than 44	3	36
More than 48	4	36 – 3 = 33
More than 52	5	36 – 3 – 4 = 29
More than 56	7	36 – 3 – 4 – 5 = 24
More than 60	9	36 – 3 – 4 – 5 – 7 = 17
More than 64	8	36 – 3 – 4 – 5 – 7 – 9 = 8
Total	36	



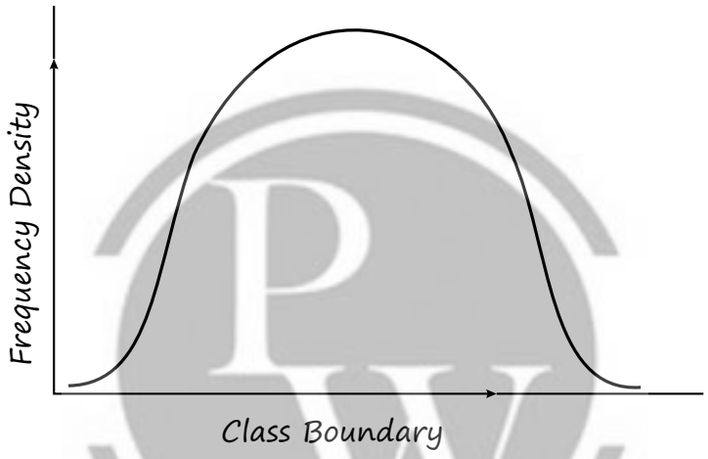
Note: Ogives help us to find out Median. If a perpendicular is drawn from the point of intersection of the two ogives on the horizontal axis, then the x–value of this point gives us the value of median

FREQUENCY CURVE

A frequency curve is a graphical representation of the frequency distribution of a dataset. It is a smooth curve obtained by joining the midpoints of the upper side of each vertical bar in a histogram. The frequency curve depicts the relationship between the frequency density on the vertical axis and the class boundary on the horizontal axis. The total area under the frequency curve is taken to be unity, meaning the graph represents the relative frequencies of the data.

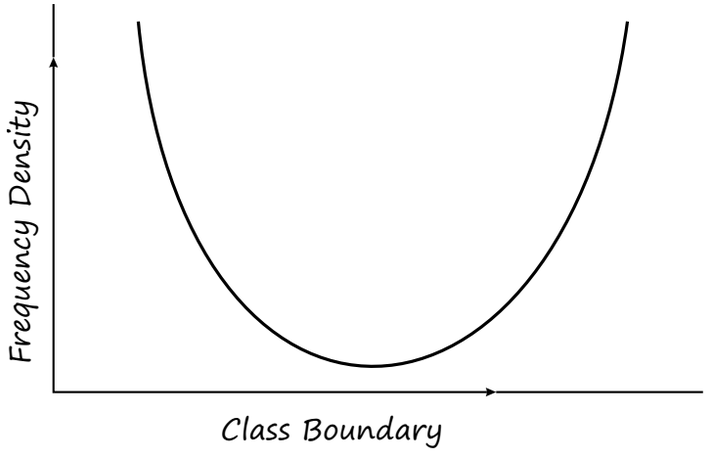
There are four main types of frequency curves:

1. **BELL SHAPED CURVE:** The bell–shaped curve, also known as the normal distribution, exhibits a symmetrical shape. It starts with a low frequency value, increases to a maximum at the center, and then decreases to a low frequency value at the other extreme. The bell–shaped curve is commonly observed in natural phenomena and human attributes, such as the distribution of heights, weights, exam scores, or profits.



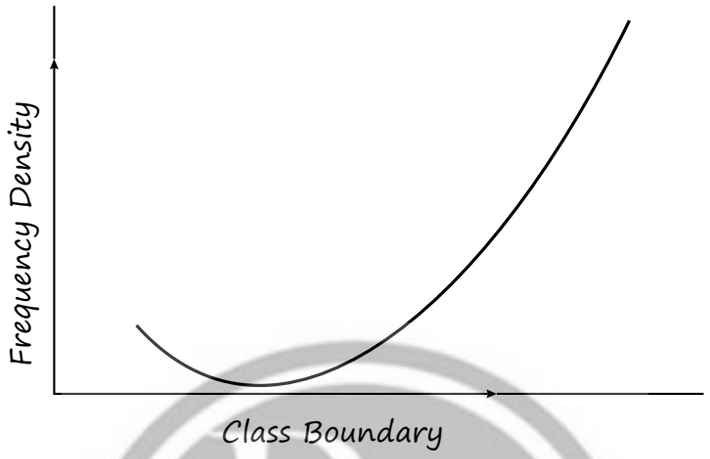
E.g.: The distribution of human heights in a population often follows a bell–shaped curve. In this distribution, the majority of individuals tend to cluster around the average height, with fewer individuals at the extreme ends (very short or very tall).

2. **U SHAPED CURVE:** The U–shaped curve represents a situation where the frequency is minimum near the central part and gradually increases to the maximum at the two extremities. This type of curve indicates a pattern where there is a higher concentration of observations at the extremes and fewer observations in the middle.



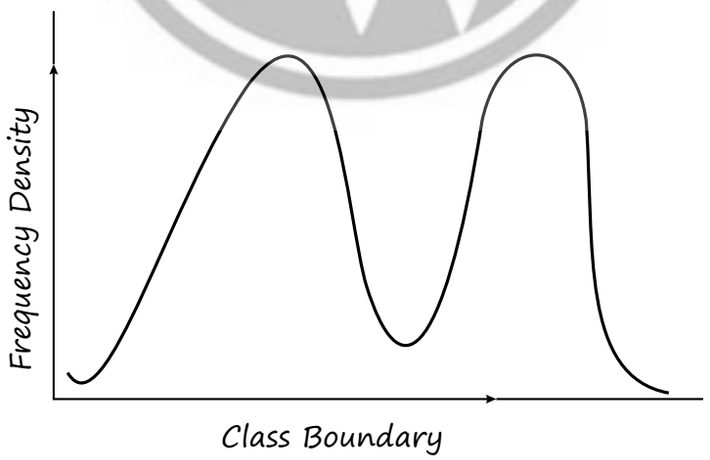
E.g.: The traffic flow in a metropolitan city during office timings often exhibits a U–shaped curve. The traffic volume is minimal during non–peak hours, increases significantly during rush hours, and then decreases again during the late evening.

3. **J–SHAPED CURVE:** The J–shaped curve starts with a minimum frequency and gradually increases to reach its maximum frequency at the other extremity. This type of curve suggests a pattern where there are initially fewer observations, but as we move towards the extreme, the frequency increases rapidly.



E.g.: The traffic on roads between 5 am to 10 am in a busy city like Delhi often follows a J–shaped curve. Initially, there is minimal traffic in the early morning, but as people start commuting to work, the traffic volume gradually increases, peaking during the morning rush hour.

4. **MIXED CURVE:** In a mixed curve, different portions of the distribution may resemble a bell–shaped curve, a U–shaped curve, a J–shaped curve, or other shapes. This indicates that the dataset is a combination of different subgroups or populations, each with its own distribution pattern.



E.g.: Let's consider a dataset representing the heights of individuals from two different populations, such as male and female. If we combine the heights of both populations in a single frequency distribution, the resulting frequency curve may exhibit characteristics of both a bell–shaped curve and a U–shaped curve. The female heights may follow a bell–shaped curve, while the male heights may show a U–shaped curve. The overall distribution would then be a mixed curve, reflecting the characteristics of both populations.

Example 7. Which of the following statements is false? **(ICAI)**

(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

Sol. (b) The correct origin of the word "Statistics" is not from the Italian word "Statista" but from the Latin word "Status" or the French word "Statistik".

Therefore, Statistics is not derived from the Italian word 'Statista'.

Hence, the correct option is (b).

Example 8. Statistics is defined in terms of numerical data in the (ICAI)

(a) Singular sense (b) Plural sense

(c) Either (a) or (b) (d) Both (a) and (b)

Sol. (c) The definition of statistics can be expressed in both the singular and plural sense.. In the singular sense, statistics refers to the discipline or field of study that deals with the collection, analysis, interpretation, presentation, and organization of data. In the plural sense, statistics refers to the numerical data or facts that are collected and analyzed for various purposes.

Therefore, the definition of statistics can be understood in either the singular sense or the plural sense.

Hence, the correct option is (c).

Example 9. Statistics is concerned with **(ICAI)**

(a) Qualitative information (b) Quantitative information

(c) (a) or (b) (d) Both (a) and (b)

Sol. (d) Statistics is a field that deals with the collection, analysis, interpretation, presentation, and organization of data. It is concerned with both qualitative and quantitative information.

Qualitative information refers to non–numerical data that describes qualities or attributes, such as gender, occupation, or type of vehicle.

Quantitative information, on the other hand, involves numerical data that can be measured or counted, such as age, height, income, or the number of items sold.

Statistics provides methods and techniques to analyze and summarize both qualitative and quantitative data, allowing for meaningful interpretations and informed decision–making.

Hence, the correct option is (d).

Example 10. Annual income of a person is **(ICAI)**

(a) An attribute (b) A discrete variable

(c) A continuous variable (d) (b) or (c)

Sol. (b) We know that,

When a variable assumes a finite or a countably infinite number of isolated values, it is known as a discrete variable.

Therefore, Annual income of a person is a discrete variable.

Hence, the correct option is (b).

Example 11. Marks of a student is an example of **(ICAI)**

(a) An attribute (b) A discrete variable

(c) A continuous variable (d) None of these

Sol. (b) A discrete variable is one that can take on specific, separate values with no intermediate values between them. In the case of marks, they are typically represented as whole numbers or specific grades, such as 85, 90, or A, B, C, etc. Each mark represents a distinct category or value, and there are no fractional or continuous values between them.

Therefore, marks of a student can be considered as a discrete variable.

Hence, the correct option is (b).

Example 12. Drinking habit of a person is **(ICAI)**

(a) An attribute (b) A variable

(c) A discrete variable (d) A continuous variable

Sol. (a) Attributes refer to qualities or characteristics that describe individuals or objects. In this case, the drinking habit is a qualitative characteristic that describes a person's behavior or preference. It does not have numerical value or magnitude associated with it, making it an attribute rather than a variable.

Therefore, the drinking habit of a person can be categorized as an attribute.

Hence, the correct option is (a).

Example 13. Which of the following is an example of qualitative data?

(a) Temperature in degrees Celsius.

(b) Number of cars in a parking lot.

(c) Rating of customer satisfaction (Excellent, Good, Average, Poor).

(d) Height in centimeters.

Sol. (c) Qualitative data refers to non-numerical information that describes qualities or characteristics. It is subjective and categorical in nature.

Therefore, Rating of customer satisfaction is an example: of qualitative data.

Hence, the correct option is (c).

Example 14. Data collected on religion from the census reports are **(ICAI)**

(a) Primary data (b) Secondary data

(c) Sample data (d) (a) or (b)

Sol. (b) Data collected on religion from census reports are considered secondary data. Secondary data refers to data that has been collected by someone else or from existing sources for a purpose other than the current research or study. In this case, the census reports have already been conducted by a designated authority, and the data collected from those reports are used for various analyses and studies.

Therefore, Data collected on religion from the census reports are Secondary data.

Hence, the correct option is (b).

Example 15. The weights of a group of individuals were recorded using a weighing scale. Based on this information, identify the nature of the data collected.

(a) Primary data (b) Secondary data

(c) Discrete data (d) Continuous data

Sol. (a) The weights recorded using a weighing scale represent measurements that are directly collected from the individuals themselves. Therefore, the data is considered "Primary data".

Hence, the correct option is (a).

Example 16. Which of the following methods is considered the most efficient for collecting primary data quickly? **(ICAI)**

(a) Personal interview (b) Indirect interview

(c) Telephone interview (d) Direct observation

Sol. (c) The quickest method to collect primary data is through a telephone interview. This method involves conducting interviews with respondents over the phone, where the interviewer asks questions and records the responses.

Hence, the correct option is (c).

Example 17. The most effective approach for data collection during a natural disaster is

(a) Personal interview (b) Indirect interview

(c) Questionnaire method (d) Direct observation method.

Sol. (a) During a natural disaster, conducting personal interviews is often the most effective method for data collection. This involves directly interacting with individuals who have experienced the disaster and gathering information through face–to–face conversations. Personal interviews allow for detailed and nuanced responses, providing valuable insights into the impact of the disaster, the needs of the affected individuals, and their perspectives on the situation. It enables researchers or relief workers to ask specific questions, probe further for details, and gain a deeper understanding of the situation on the ground.

Hence, the correct option is (a).

Example 18. When studying historical events and analyzing past records, which method of data collection is commonly used to extract information?

(a) Personal interview (b) Secondary data analysis

(c) Questionnaire method (d) None of these

Sol. (b) By conducting secondary data analysis, researchers can extract valuable information from these existing sources and analyze it to gain insights into the historical events being studied. This method saves time and resources compared to conducting personal interviews, administering questionnaires, or conducting case study analyses, especially when studying events that have already occurred.

Hence, the correct option is (b).

Example 19. Which data collection method involves gathering information from a large sample of individuals by mailing them questionnaires to be completed and returned?

(a) Telephone interview method (b) Mailed questionnaire method

(c) Direct interview method (d) All of these

Sol. (b) The mailed questionnaire method is specifically designed to reach a large number of individuals dispersed across different locations. Questionnaires are sent via mail, allowing respondents to complete them at their convenience and return them by mail. This method is often used when a wide geographic area needs to be covered, making it an effective approach for reaching a diverse population.

Hence, the correct option is (b).

Example 20. Which method of data collection involves recording information from documents, reports, or other existing sources?

(a) Personal interview (b) Secondary data analysis

(c) Questionnaire method (d) Direct observation method

Sol. (b) The method of data collection that involves recording information from documents, reports, or other existing sources is called secondary data analysis.

However, secondary data analysis is specifically focused on using existing data sources such as published reports, official records, surveys, or databases to analyze and draw conclusions.

Hence, the correct option is (b).

Example 21. Some important sources of secondary data are **(ICAI)**

(a) International and Government sources

(b) International and primary sources

(c) Private and primary sources

(d) Government sources.

Sol. (a) We know that,

International sources refer to data collected and maintained by international organizations such as the United Nations, World Bank, or World Health Organization. These organizations collect data from various countries and provide global datasets on topics such as economy, development, health, education, and more.

Government sources, as mentioned earlier, are also crucial sources of secondary data. Government agencies at national, regional, or local levels collect and publish data on a wide range of subjects, including demographics, employment, trade, crime, and public services. These datasets are often reliable and comprehensive, making them valuable for research and analysis purposes.

Therefore, International and government sources are important sources of secondary data.

Hence, the correct option is (a).

Example 22. Sweetness of sweet dish is **(Jan 2021)**

(a) An attribute (b) A discrete variable

(c) A continuous variable (d) A variable

Sol. (a) As we know, Attribute is a quality or trait that is assessed for every observation (record) and is variable among observations.

A variable's trait or property is described by an attribute. A sweet dish's characteristic is sweetness.

Therefore, Sweetness of a sweet dish is an attribute.

Hence, the correct option is (a).

Example 23. The accuracy and consistency of data can be verified by **(ICAI)**

(a) Internal checking (b) External checking

(c) Scrutiny (d) Both (a) and (b)

Sol. (c) Scrutiny refers to the careful examination and analysis of data to ensure its accuracy and consistency. It involves a thorough review of the data, checking for any errors, inconsistencies, or discrepancies. This process may involve comparing the data against predetermined criteria, conducting data validations, and performing data quality checks.

Therefore, the accuracy and consistency of data can be verified by Scrutiny.

Hence, the correct option is (c).

Example 24. There were 200 employees in an office in which 150 were married. Total male employees were 160 out of which 120 were married. What was the number of female unmarried employees? **(July 2021)**

(a) 30 (b) 40 (c) 50 (d) 10

Sol. (d) Make a data table according to the conditions given in the question,

	Male	Female	Total
Married	120	30	150
Unmarried	40	10	50
Total	160	40	200

From the data table, we see unmarried female employees are 10.

Hence, the correct option is (d).

Example 25. The distribution of scores in a test is an example of the frequency distribution of

(a) A discrete variable (b) A continuous variable

(c) An attribute (d) (a) or (c)

Sol. (a) The distribution of scores in a test is an example: of a frequency distribution, which shows the frequency or count of each score. In this case, the scores are discrete because they represent distinct categories or values.

For example, if the test scores range from 0 to 100, each score represents a separate category or value.

Therefore, the distribution of scores in a test is an example of the frequency distribution of a discrete variable.

Hence, the correct option is (a).

Example 26. Which of the following is suitable for the graphical representation of a Cumulative frequency distribution? **(Jan 2021)**

(a) Frequency polygon (b) Histogram

(c) Ogive (d) Pie chart

Sol. (c) As we know, Ogive is defined as the cumulative frequency distribution graph of a series.

The Ogive is a cumulative distribution graph that displays data values on the horizontal axis and either cumulative relative frequencies, cumulative frequencies, or cumulative percent frequencies on the vertical axis.

Thus, Ogive is suitable for the graphical representation of a Cumulative frequency distribution.

Hence, the correct option is (c).

Example 27. Which method is commonly used to present detailed information?

(a) Charts (b) Graphs (c) Tables (d) Visualizations

Sol. (c) Tables are commonly used to present detailed information in a structured and organized format. They consist of rows and columns, allowing for clear representation of data. Tables are particularly useful when presenting numerical data or when comparing multiple variables or categories. They provide a concise and systematic way to present information, making it easier for readers to interpret and analyze the data.

Hence, the correct option is (c).

Example 28. The column headings of a table are known as:

(a) Body (b) Stub (c) Box–head (d) Caption

Sol. (d) The table under consideration is divided into caption, Box–head, Stub and Body where,

Caption is the upper part of the table, describing the columns and sub–columns, if any.

Therefore, the column headings of a table are known as Caption.

Hence, the correct option is (d).

Example 29. The number of "Frequency distribution" is

(a) Two (b) One (c) Five (d) Four

Sol. (b) "Frequency distribution" refers to a tabular or graphical representation of data that shows the frequency of each value or class interval in a dataset. It presents the data in a systematic and organized manner, allowing for easy analysis and interpretation. There is typically only one frequency distribution for a given dataset, which summarizes the distribution of values or intervals and their corresponding frequencies.

Therefore, the number of "Frequency distribution" is one.

Hence, the correct option is (b).

Example 30. (class frequency)/(width of the class) is defined as

(a) Frequency density (b) Frequency Distribution

(c) Both (d) None of these

Sol. **(a)** We know that,

(Class frequency)/(width of the class) represents the frequency density of a class in a frequency distribution.

Frequency density is a measure that accounts for the width of each class interval. It is calculated by dividing the frequency of the class by the width of the class interval. Frequency density is used to compare the densities of different classes and provides a more meaningful representation of the distribution of data.

Hence, the correct option is (a).

Example 31. Tally Marks determines **(ICAI)**

(a) class width (b) class boundary

(c) class limits (d) class frequency

Sol. (d) Tally marks are used to determine the frequency or count of observations within each class of a frequency distribution. Each tally mark represents one observation, and

by counting the tally marks, we can determine the frequency or number of observations in a particular class. Tally marks are a visual representation of the frequency and help in organizing and summarizing the data in a frequency distribution.

Therefore, Tally Marks determines class frequency.

Hence, the correct option is (d).

Example 32. Most of the Commonly used distributions provide **(Jan 2021)**

(a) Bell – Shaped (b) U Shaped

(c) J – Shaped Curve (d) Mixed Curve

Sol. (a) As we know, A bell curve, commonly known as the normal distribution, is a typical type of distribution for a variable.

The graph that is used to represent a normal distribution has a symmetrical bell–shaped curve, hence the name "bell curve."

All other potential occurrences are symmetrically distributed around the mean, resulting in a downward–sloping curve on either side of the peak, while the highest point on the curve, or the top of the bell, represents the most probable event in a series of data (its mean, mode, and median in this case).

The standard deviation of the bell curve serves as a measure of its width.

Hence, the correct option is (a).

Example 33. Unequal widths of classes in the frequency distribution can create difficulties in the construction of:

(a) Ogive (b) Frequency polygon

(c) Histogram (d) None of these

Sol. (c) Unequal widths of classes in the frequency distribution can make it challenging to accurately represent the data in a histogram. Histograms require equal–width intervals to ensure the bars accurately depict the distribution of data. Unequal class widths may result in distorted representations or inaccurate interpretations of the data when constructing a histogram.

Hence, the correct option is (c).

Example 34. The number of types of cumulative Frequency is **(ICAI)**

(a) One (b) Two (c) Three (d) Four

Sol. (b) There are two types of cumulative frequency:

Less than cumulative frequency: It represents the cumulative frequency of values that are less than or equal to a given class or value.

Greater than cumulative frequency: It represents the cumulative frequency of values that are greater than or equal to a given class or value.

Hence, the correct option is (b).

Example 35. Classes with zero Frequency are called **(ICAI)**

(a) nil class (b) Empty class (c) class (d) None of these

Sol. (b) Classes with zero frequency, meaning there are no data points falling within that class interval, are referred to as empty classes. These classes are still included in the overall class structure but do not contain any data.

Hence, the correct option is (b).

Example 36. For calculating class frequencies, it is important that these classes are

(a) Mutually exclusive (b) Non–overlapping

(c) Independent (d) None of these

Sol. (b) Class intervals in a frequency distribution should be non–overlapping to ensure that each data point falls into only one class. This ensures accuracy and prevents ambiguity in determining the class frequencies.

Hence, the correct option is (b).

Example 37. The value exactly at the middle would ever be included in a class interval are called

(a) class mark (b) mid value (c) both (d) None of these

Sol. (c) The class mark or midpoint of a class interval is calculated as the average of the lower and upper limits of the interval. It represents the central value within the interval and is often used as a representative value for that interval.

E.g.: If a class interval is defined as 20–30, the class mark would be (20 + 30) / 2 = 25

Hence, the correct option is (c).

Example 38. For the non–overlapping classes 1–20,21–40,41–60,61–80,81–100, the class mark of the class 61–80 is

(a) 10.5 (b) 70.5 (c) 90.5 (d) None

Sol. (b) Class mark: Corresponding to a class interval, this may be defined as the sum of the lower class limit and upper class limit divided by 2.

Thus, the class mark of the class 61–80 is:

$$\text{Class mark} = \frac{LCL + UCL}{2} = \frac{61 + 80}{2} = 70.5$$

Hence, the correct option is (b).

Example 39.

Class	0-10	10-20	20-30	30-40	40-50
Frequency	10	12	8	15	6

For the class 20-30, cumulative frequency is

(a) 22 (b) 8 (c) 51 (d) 30

Sol. (d) According to given data,

Class:	0-10	10-20	20-30	30-40	40-50
Frequency	10	12	8	15	6
Cumulative Frequency	10	22	30	45	51

Therefore, for the class 20-30, cumulative frequency is 30.

Hence, the correct option is (d).

Example 40. The curve obtained by joining the points, whose x – coordinates are the upper limits of the class – intervals and y coordinates are corresponding cumulative frequencies is called

(a) Ogive (b) Histogram

(c) Frequency Polygon (d) Frequency curve

Sol. (a) An ogive is a graphical representation of the cumulative frequency distribution. It is obtained by plotting points using the upper limits of the class intervals on the x-axis and the corresponding cumulative frequencies on the y-axis. The curve formed by joining these points is called an ogive. It is used to observe the cumulative frequency distribution and analyze the pattern of data.

Hence, the correct option is (a).

Example 41. In Histogram, the classes are taken

(a) Overlapping (b) Non–overlapping

(c) Both (d) None of these

Sol. (a) In a histogram, the classes are typically represented by adjacent bars, where the intervals or ranges of values overlap with each other. Each data point can fall within multiple overlapping intervals, allowing for a more detailed representation of the distribution. Overlapping classes in a histogram provide a smoother and continuous representation of the data, allowing for better visualization of the frequency distribution.

Hence, the correct option is (a).

PRACTICE QUESTIONS (PART A)

1. Histogram is used for finding

(a) Mode (b) Mean (c) First quartile (d) None of these

2. Ogive graph is used for finding

(a) Mean (b) Mode (c) Median (d) None of these

3. _____ series is continuous.

(a) Open ended (b) Exclusive (c) Close ended (d) Unequal class intervals

4. Histogram can be shown as

(a) Ellipse (b) Rectangle (c) Hyperbola (d) Circle

5. Which of the following graphs is suitable for cumulative frequency distribution?

(a) Ogive (b) Histogram (c) G.M (d) A.M

6.

Class	0-10	10-20	20-30	30-40	40-50
Frequency	4	6	20	8	3

For the class 20-30, cumulative frequency is

(a) 10 (b) 26 (c) 30 (d) 41

7. An ogive is a graphical representation of

(a) Cumulative frequency distribution

(b) A frequency distribution

(c) Ungrouped data

(d) None

8. The number of times a particular item occurs in a class interval is called its:

(a) Mean (b) Frequency

(c) Cumulative Frequency (d) None of these

9. A suitable graph for representing the portioning of total into sub parts in statistics is:
   (a) A Pie chart
   (b) A pictograph
   (c) An ogive
   (d) Histogram

10. Data are said to be _____ if the investigator himself is responsible for the collection of the data.
   (a) Primary data
   (b) Secondary data
   (c) Mixed of primary and secondary data
   (d) None of these

11. Histogram is useful to determine graphically the value of _____ for the collection of the data.
   (a) Arithmetic mean
   (b) Median
   (c) Mode
   (d) None of these

12. The following Frequency distribution is classified as

X:	12	17	24	36	45
Y:	2	5	3	8	9

   (a) Continuous distribution
   (b) Discrete distribution
   (c) Cumulative frequency distribution
   (d) None of these

13. Frequency density is used in the construction of
   (a) Histogram
   (b) Ogive
   (c) Frequency polygon
   (d) None when the classes are of unequal width

14. Divided bar charts is considered for
   (a) Comparing different components of a variables
   (b) The relation of different components to the table
   (c) (a) or (b)
   (d) (a) and (b)

15. Frequency density corresponding to a class interval is the ratio of
   (a) Class frequency to total frequency
   (b) Class frequency to the class length
   (c) Class length to class frequency
   (d) Class frequency to the cumulative

16. '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

17. The point of intersection of less than ogive and greater than ogive curve gives us:
(a) Mean (b) Mode (c) Median (d) None of these

18. Find the number of observations between 250 – 300 from the following data:

Value more than:	200	250	300	500
No. of observation:	56	38	15	0

(a) 38 (b) 23 (c) 15 (d) None of the above

19. The chart that uses logarithm of variables is known as:
(a) Ratio chart (b) Line chart
(c) Multiple line chart (d) Component line chart

20. Data Collection on religion from census reports are:
(a) Primary data (b) Secondary data
(c) Sample data (d) (a) or (b)

21. Difference between the maximum & minimum value of a given data is called
(a) width (b) size (c) range (d) None of these

22. The most appropriate diagram to represent the data relating to the monthly expenditure on different items by a family is
(a) Histogram (b) Pie-diagram
(c) Frequency polygon (d) Line graph

23. What is a exclusive series ?
(a) In which both upper and lower limit are not included in class frequency.
(b) In which lower limit is not included in class frequency.
(c) In which upper limit is not included in class frequency
(d) None of the above

24. In indirect oral investigation:
(a) Data is not capable of numerical expression
(b) Not possible or desirable to approach informant directly
(c) Data is collected from the books
(d) None of these

25. If the class intervals are 10 – 14, 15 – 19, 20 – 24, ..., then the first class boundary is
(a) 9.5 – 14.5 (b) 10 – 15 (c) 9 – 15 (d) 10.5 – 15.5

26. The following data related to the marks of group of students

Marks No of Students	
More than 70%	7
More than 60%	18
More than 50%	40
More than 40%	60

Marks No of Students	
More than 30%	75
More than 20%	100

How many students have got marks less than 50%?

(a) 60 (b) 82 (c) 40 (d) 53

27. Out of 1000 persons, 25 percent were industrial workers and the rest were agricultural workers. 300 persons enjoyed world cup matches on T.V. 30 percent 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) 230 (b) 250 (c) 240 (d) 260

28. A pie diagram is used to represent the following data.

Source	Customers	Excise	Income Tax	Wealth Tax
Revenue in Millions:	120	180	240	180

The central angles corresponding to Income Tax and Wealth Tax are

(a) 90°, 120° (b) 120°, 90° (c) 60°, 120° (d) 90°, 60°

29. 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%

30. In 2000, out of total of 1,750 workers of a factory 1,200 were members of a trade union. The number of women employed was 200 of which 175 did not belong to a trade union. In 2004, there were 1,800 employees who belong to a trade union 50 who did not belong to trade union. Of all the employees in 2004, 300 were women of whom only 8 were non-trade members. On the basis of this information, the ratio of female member of the trade union in 2000 and 2004 is

(a) 292 : 25 (b) 8 : 175 (c) 175 : 8 (d) 25 : 292

Answer Key

1. (a)	**2.** (c)	**3.** (a)	**4.** (b)	**5.** (a)	**6.** (c)	**7.** (a)	**8.** (b)	**9.** (a)	**10.** (a)
11. (c)	**12.** (b)	**13.** (a)	**14.** (d)	**15.** (a)	**16.** (d)	**17.** (c)	**18.** (b)	**19.** (a)	**20.** (b)
21. (c)	**22.** (b)	**23.** (c)	**24.** (b)	**25.** (a)	**26.** (a)	**27.** (d)	**28.** (b)	**29.** (b)	**30.** (d)

SUMMARY

- 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.

- We can broadly classify data as
  - Primary
  - Secondary
- Scrutiny of Data : Data scrutiny focuses on inference, the process of deriving a conclusion based solely on what is already known by the researcher. Data needs to be curated, cleansed, enriched and translated into actionable insights before it can be put to work doing something meaningful.
- Mode of Presentation of Data
  (a) Textual presentation;
  (b) Tabular presentation or Tabulation;
  (c) Diagrammatic representation.
- The types of diagrams:
  (a) Line diagram or Histogram
  (b) Bar diagram
  (c) Pie chart
- Frequency Distribution of a Variable
  (a) Find the largest and smallest observations and obtain the difference between them, known as Range,
  (b) In the case of a discrete variable, Form a number of classes depending on the number of isolated values assumed by a discrete variable.
  (c) In case of a continuous variable, find the number of class intervals using the relation, No. of class Interval × class length = Range.
  (d) Cumulative frequency – less than and more than types
- Graphical Representation of Data
  (i) Histogram or Area diagram;
  (ii) Frequency Polygon;
  (iii) Ogives or cumulative Frequency graphs.
- Frequency Curve
  (a) Bell-shaped curve; (b) U-shaped curve;
  (c) J-shaped curve; (d) Mixed curve.
- Lower Class Boundary (LCB) = Lower Class Length $- \frac{D}{2}$,
  Upper Class Boundary = Upper Class Length $+ \frac{D}{2}$
- Mid-point $= \frac{LCB + UCB}{2}$ or $\frac{LCL + UCL}{2}$
- Frequency density of a class interval = frequency of that class interval/corresponding class length
- Relative frequency and percentage frequency of a class interval = Class frequency/total frequency

Unit-2: Nature and Scope of Business Economics

INTRODUCTION TO SAMPLING

Sampling is a fundamental technique in statistics that enables researchers to draw meaningful conclusions about a large population by studying a smaller, manageable subset. This approach acknowledges that it is often impractical, if not impossible, to collect data from an entire population. Instead, it provides a systematic way to select a representative group, known as the sample, which is carefully chosen to reflect the characteristics of the larger population.

Sampling is essential for various fields, including market research, scientific experiments, public opinion polling, and quality control, as it offers an efficient means of making inferences about a broad range of phenomena.

E.g.: When a person goes to market to buy rice, firstly, he takes some rice in his hand from a bag containing rice and then examining it, he takes the decision whether to buy it or not.

PRINCIPLES OF SAMPLE SURVEY

A sample survey is a research method where data is collected from a subset (sample) of individuals or items from a larger group (population). It's used to make inferences about the entire population by studying the characteristics and behaviors of the selected sample.

Basic principles of Sample survey comparison the components such as:

- **Law of Statistical Regularity:** This principle is based on the notion that in large and random samples, patterns and regularities emerge. It means that in sufficiently large samples, the characteristics of the sample start to mirror the characteristics of the population it's drawn from.
- **Principle of Inertia:** This principle suggests that, unless there's a good reason to the contrary, researchers often prefer not to change their methods or sample size. It helps in maintaining consistency and comparability in survey results over time.
- **Principle of Optimization:** This principle focuses on maximizing the information gained from a sample while minimizing the cost and effort of collecting data. It's about finding the right balance between sample size and data quality.
- **Principle of Validity:** Validity is all about measuring what you intend to measure. In sampling surveys, this principle emphasizes the importance of using methods and questions that accurately capture the information you're interested in. It's essential to ensure that your survey results are meaningful and truly represent the characteristics of the population.

TYPES OF ERRORS IN SAMPLE SURVEY

There are two types of errors in sample survey:

(a) Sampling errors (b) Non-sampling errors

- Sampling errors are discrepancies between the characteristics of a sample and the larger population, usually due to random chance.
  - **(a) Errors arising out of defective sampling** occur when the sampling method used is flawed, leading to an unrepresentative sample.

(b) **Errors arising out due to substitution** involve issues when intended survey participants are replaced with others, impacting data accuracy.

(c) **Errors owing to faulty demarcation of units** result from misidentifying or misclassifying survey units, affecting the overall sample.

(d) **Errors owing to the wrong choice of statistic** occur when an inappropriate statistical method is used, leading to misleading or inaccurate results.

(e) **Variability in the population** is an error related to the natural differences that exist within a population. It arises when the sampled units exhibit greater diversity than expected, leading to less precise estimates.

Non-sampling errors are inaccuracies in survey data that aren't a result of the sampling process. They encompass a wide range of errors, including coverage errors, nonresponse errors, measurement errors, and processing errors. These errors can be caused by factors such as undercoverage, nonresponse from participants, errors in data collection, and mistakes during data processing. Non-sampling errors can significantly affect the quality and reliability of survey results.

POPULATION AND SAMPLE

1. **Population:** The population is the entire group or collection of individuals, objects, or data that the researcher wants to draw conclusions about. It represents the larger entity under study and can be finite or infinite, depending on the research context.
2. **Sample:** A sample is a subset of the population selected for a research study. It is chosen to represent the population because studying the entire population may be impractical or too costly. The characteristics and behaviors of the sample are analyzed to make inferences about the entire population. The process of selecting a sample should be done systematically to ensure it is a representative and unbiased reflection of the population.

S.No.	Population	Sample
1.	Students of the school	Commerce students of grade 12
2.	All books in a library	Book

3. **Parameter:** A parameter is a characteristic of a population, determined by considering all units within that population. Statistical inferences about population parameters are made by analyzing sample observations drawn from the same population.

   Here,

   - When we are interested in '**Population Mean**' then it is represented as:

$$\mu = \frac{\sum_{a=1}^{n} x_a}{N}$$

   Where N denotes the population size and x_a represents a^{th} member of the population.

   - When we are interested in '**Population Proportion**' then it is represented as:

   Where, X represents the count possessing the required attribute.

   - Another important prameter namely the poluation vairance, to be denoted by σ^2 is given by:

$$\sigma^2 = \frac{\sum (X_a - \mu)^2}{N}$$

$$\text{Also, } SD = \sigma = \sqrt{\frac{\sum (X - \mu)}{N}}$$

STATISTICS

It is defined as a statistical measure of sample observation.

It is a function of of sample observations. If the sample observations are denoted by x_1, x_2, x_3 , ... x_n, then a statistics T may be expressed as $T = f(x_1, x_2, x_3, \ldots x_n)$.

A statistic is used to estimate a particular population parameter. The estimates of population mean, variance and population proportion are given by

$$\bar{x} = \hat{\mu} = \frac{\sum x_i}{n}$$

$$S_2 = \widehat{\sigma^2} = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$\text{and } p = p = \hat{p} = \frac{x}{n}$$

Here x (in the last formula) denotes the number of units in the sample in possession of the attribute under discussion.

UNDERSTANDING SAMPLING DISTRIBUTION AND STANDARD ERROR:

Imagine starting with a group of N items. By taking several samples of a fixed size (let's call it n), we can explore the variability in our results. If we draw samples with replacement, the possibilities are endless, but without replacement, the total number of potential samples is ${}^{N}C_{n}$.

When we calculate a statistic, like the mean, it's natural for the values to differ across samples. This variation is what we call "Sampling Fluctuations."

If we manage to gather all possible values of a statistic (let's call it T) from samples of a fixed size, along with their probabilities, we can create a probability distribution. This distribution, treating the statistic as a random variable, is the sampling distribution of the statistic. Similar to theoretical probability distributions, this sampling distribution has key characteristics. The mean of the statistic in its sampling distribution is known as the "Expectation," and the standard deviation of the statistic (T) is called the "Standard Error (SE)" of T. The SE acts as a measure of precision achieved through sampling and is inversely proportional to the square root of the sample size. This implies that as the sample size increases, the SE decreases, indicating increased precision.

SE is inversely proportional to the square root of sample size.

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}} \text{ for simple random sampling with replacement}$$

$$= \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \text{ for simple random sampling without replacement}$$

STANDARD ERROR FOR PROPORTION

$SE(p) = \sqrt{\frac{Pq}{n}}$ for simple random sampling with replacement $= \sqrt{\frac{Pq}{n}}\sqrt{\frac{N-n}{N-1}}$ for simple random sampling without replacement

The factor $\sqrt{\frac{N-n}{N-1}}$ is known as finite population correction (fpc) or finite population multiplier and may be ignored as it tends to 1 if the sample size (n) is very large or the population under consideration is infinite when the parameters are unknown, they may be replaced by the corresponding statistic.

SAMPLING

The process of selecting a sample from the given population is called Sampling.

Broadly it is of three types:

(a) Probability sampling
(b) Non-probability sampling
(c) Mixed sampling

SIMPLE RANDOM SAMPLING (SRS)

- Probability-based; each member has a fixed chance to be part of the sample.
- Units selected independently; can be with or without replacement.
- Effective for small, homogenous populations.

STRATIFIED SAMPLING

- Suitable for large, heterogeneous populations.
- Divides population into strata; minimal variation within strata.
- Enhances precision, provides representation for all sub-populations.

MULTI-STAGE SAMPLING

- Complex design with multiple stages of sampling units.
- **E.g.:** State, district, police station, household.
- Cost-effective, flexible, and covers a large population, but may be less accurate.

SYSTEMATIC SAMPLING

- Units selected at regular intervals after a random start.
- Convenient and less time-consuming; suitable with an updated sampling frame.
- Prone to bias if there is an undetected periodicity in the sampling frame.

PURPOSIVE OR JUDGMENT SAMPLING

- Non-probabilistic; dependent on the discretion of the sampler.
- Subjective, varies among individuals; unsuitable for statistical hypothesis testing.

NOTEWORTHY POINTS

MIXED SAMPLING

1. Combines probabilistic and pre-decided rules.
2. Systematic sampling is an example.

SIMPLE RANDOM SAMPLING (SRS) CONSIDERATIONS

1. Effective if the population is not very large, sample size is reasonable, and population is homogenous.
2. Free from sampler's biases; foundational for tests of significance.

STRATIFIED SAMPLING CONSIDERATIONS

1. Advisable for large, heterogeneous populations with prior information available.
2. Involves proportional allocation or Bowley's allocation for sample sizes.

MULTI-STAGE SAMPLING COVERAGE

1. Extensive coverage, computational labor savings, and cost-effectiveness.
2. Adds flexibility but may be less accurate compared to stratified sampling.

SYSTEMATIC SAMPLING DRAWBACK

1. Prone to bias if periodicity in the sampling frame exists.
2. No statistical inference about population parameters due to non-probabilistic nature.

Example 1. Sampling can be described as a statistical procedure **(ICAI)**

(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)

Sol. (c) Sampling involves drawing conclusions about an entire population based on information obtained from a randomly selected sample. This option accurately describes the purpose of statistical sampling.

Hence, the correct option is (c).

Example 2. The basis for making a statistical decision about an unknown universe is:

(a) Sample observations (b) A sampling frame

(c) Sample survey (d) Complete enumeration

Sol. (a) Sample observations. Statistical decisions about an unknown universe are typically made based on observations from a representative sample rather than examining the entire population.

Example 3. Which sampling method provides flexibility in the sampling process?

(a) Simple random sampling (b) Multistage sampling

(c) Stratified sampling (d) Systematic sampling

Sol. (c) Stratified sampling offers flexibility by dividing the population into subgroups (strata) and then sampling from each stratum separately.

Example 4. Which sampling method is most impacted if the sampling frame reveals an undetected periodicity?

(a) Simple random sampling (b) Stratified sampling

(c) Multistage sampling (d) Systematic sampling

Sol. (d) Systematic sampling is sensitive to periodic patterns in the sampling frame, as it involves selecting every kth element from a list.

Example 5. 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 **(ICAI)**

(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)

Sol. (c) Given units: 3, 6 and 1

Since, a random sample of size 2 with replacement is taken from the population thus the required samples would be (3, 3), (3, 6), (3, 1), (6, 6), (6, 3), (6, 1), (1, 1), (1, 3), (1, 6)

Example 6. Which sampling method is at the discretion of the sampler?

(a) Systematic sampling (b) Simple random sampling

(c) Purposive sampling (d) Quota sampling

Sol. (c) In purposive sampling, the sampler exercises judgment or purpose in selecting specific elements from the population.

Example 7. If from a population with 25 members, a random sample without replacement of 2 members is taken, the number of all such samples is **(ICAI)**

(a) 300 (b) 625 (c) 50 (d) 600

Sol. (a) Given,

Total members = 25

No. of sample taken = 2

Since, the number of ways to choose 2 members from a population of 25 members without replacement is $^{25}C_2 = \frac{25!}{(25-2)! \times 2!} = \frac{25!}{23! \times 2!} = 300$

Hence, the correct option is (a).

Example 8. A population comprises 5 members. The number of all possible samples of size 2 that can be drawn from it with replacement is **(ICAI)**

(a) 100 (b) 15 (c) 125 (d) 25

Sol. (d) Since, the sampling is done with replacement, each member can be chosen again so the number of possible samples is given by n^K where n is population size and k is the sample size.

According to the given problem,

The required ways = $5^2 = 25$

Hence, the correct option is (d).

PRACTICE QUESTIONS (PART A)

1. The population of roses in Salt Lake City is an example of **(ICAI)**

(a) A finite population (b) An infinite population

(c) A hypothetical population (d) An imaginary population

2. As the sample size increases, standard error **(ICAI)**

(a) Increases (b) Decreases

(c) Remains constant (d) Decreases proportionately

3. Which sampling method involves dividing the population into clusters and then randomly selecting entire clusters for the sample?

(a) Simple random sampling (b) Systematic sampling

(c) Cluster sampling (d) Stratified sampling

4. A parameter is a characteristic of **(ICAI)**

(a) Population (b) Sample

(c) Both (a) and (b) (d) (a) or (b)

5. A statistic is **(ICAI)**

(a) A function of sample observations

(b) A function of population units

(c) A characteristic of a population

(d) A part of a population

6 From a group of 30 individuals, a committee of 4 is to be formed without replacement. How many different committees can be formed?

(a) 435 (b) 27,405 (c) 27,300 (d) 29,160

7. A box contains 6 distinct items. If 3 items are randomly selected with replacement, how many different sets of items can be chosen?

(a) 216 (b) 120 (c) 729 (d) 15

8. In a survey of 50 households, a sample of 10 households is randomly chosen without replacement. How many different samples are possible?

(a) 1,102,100 (b) 102,722,500

(c) 102,722 (d) 10,000

9. A population consists of 8 distinct elements. If a sample of 3 elements is drawn with replacement, how many different samples can be formed?

(a) 512 (b) 64 (c) 1,024 (d) 125

10. If a random sample of size two is drawn without replacement from a population containing the units a, b, c, and d, what are the possible samples? **(ICAI)**

(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)

Answer Key

1. (b)	**2.** (b)	**3.** (c)	**4.** (a)	**5.** (a)	**6.** (b)	**7.** (a)	**8.** (b)	**9.** (a)	**10.** (d)

❑❑❑