UNIVERSITY GRANTS COMMISSION

NET BUREAU

SYLLABUS

Subject: GENERAL PAPER ON TEACHING & RESEARCH APTITUDE **Code No.** : 00

PAPER-I

The main objective is to assess the teaching and research capabilities of the candidates. The test aims at assessing the teaching and research aptitude as well. Candidates are expected to possess and exhibit cognitive abilities, which include comprehension, analysis, evaluation, understanding the structure of arguments, deductive and inductive reasoning. The candidates are also expected to have a general awareness about teaching and learning processes in higher education system. Further, they should be aware of interaction between people, environment, natural resources and their impact on the quality of life.

The details of syllabi are as follows:

Unit-I **Teaching Aptitude**

- Teaching: Concept, Objectives, Levels of teaching (Memory, Understanding and Reflective), Characteristics and basic requirements.
- Learner's characteristics: Characteristics of adolescent and adult learners (Academic, Social, Emotional and Cognitive), Individual differences.
- Factors affecting teaching related to: Teacher, Learner, Support material, Instructional facilities, Learning environment and Institution.
- Methods of teaching in Institutions of higher learning: Teacher centred vs. Learner centred methods; Off-line vs. On-line methods (Swayam, Swayamprabha, MOOCs etc.).

- Teaching Support System: Traditional, Modern and ICT based.
- Evaluation Systems: Elements and Types of evaluation, Evaluation in Choice Based Credit System in Higher education, Computer based testing, Innovations in evaluation systems.

Unit-II **Research Aptitude**

- Research: Meaning, Types, and Characteristics, Positivism and Post-positivistic approach to research.
- Methods of Research: Experimental, Descriptive, Historical, Qualitative and Quantitative methods.
- Steps of Research.
- Thesis and Article writing: Format and styles of referencing.
- Application of ICT in research.
- Research ethics.

Unit-III **Comprehension**

- A passage of text be given. Questions be asked from the passage to be answered.

Unit-IV **Communication**

- Communication: Meaning, types and characteristics of communication.
- Effective communication: Verbal and Non-verbal, Inter-Cultural and group communications, Classroom communication.
- Barriers to effective communication.
- Mass-Media and Society.

Unit-V **Mathematical Reasoning and Aptitude**

- Types of reasoning.
- Number series, Letter series, Codes and Relationships.
- Mathematical Aptitude (Fraction, Time & Distance, Ratio, Proportion and Percentage, Profit and Loss, Interest and Discounting, Averages etc.).

Unit-VI Logical Reasoning

- Understanding the structure of arguments: argument forms, structure of categorical propositions, Mood and Figure, Formal and Informal fallacies, Uses of language, Connotations and denotations of terms, Classical square of opposition.
- Evaluating and distinguishing deductive and inductive reasoning.
- Analogies.
- Venn diagram: Simple and multiple use for establishing validity of arguments.
- Indian Logic: Means of knowledge.
- Pramanas: Pratyaksha (Perception), Anumana (Inference), Upamana (Comparison), Shabda (Verbal testimony), Arthapatti (Implication) and Anupalabddhi (Non-apprehension).
- Structure and kinds of Anumana (inference), Vyapti (invariable relation), Hetvabhasas (fallacies of inference).

Unit-VII Data Interpretation

- Sources, acquisition and classification of Data.
- Quantitative and Qualitative Data.
- Graphical representation (Bar-chart, Histograms, Pie-chart, Table-chart and Line-chart) and mapping of Data.
- Data Interpretation.
- Data and Governance.

Unit-VIII Information and Communication Technology (ICT)

- ICT: General abbreviations and terminology.
- Basics of Internet, Intranet, E-mail, Audio and Video-conferencing.
- Digital initiatives in higher education.
- ICT and Governance.

Unit-IX **People, Development and Environment**

- Development and environment: Millennium development and Sustainable development goals.
- Human and environment interaction: Anthropogenic activities and their impacts on environment.
- Environmental issues: Local, Regional and Global; Air pollution, Water pollution, Soil pollution, Noise pollution, Waste (solid, liquid, biomedical, hazardous, electronic), Climate change and its Socio-Economic and Political dimensions.
- Impacts of pollutants on human health.
- Natural and energy resources: Solar, Wind, Soil, Hydro, Geothermal, Biomass, Nuclear and Forests.
- Natural hazards and disasters: Mitigation strategies.
- Environmental Protection Act (1986), National Action Plan on Climate Change, International agreements/efforts -Montreal Protocol, Rio Summit, Convention on Biodiversity, Kyoto Protocol, Paris Agreement, International Solar Alliance.

Unit-X **Higher Education System**

- Institutions of higher learning and education in ancient India.
- Evolution of higher learning and research in Post Independence India.
- Oriental, Conventional and Non-conventional learning programmes in India.
- Professional, Technical and Skill Based education.
- Value education and environmental education.
- Policies, Governance, and Administration.

NOTE:

(i) Five questions each carrying 2 marks are to be set from each Module.

(ii) Whenever graphical/pictorial question(s) are set for sighted candidates, a passage followed by equal number of questions and weightage be set for visually impaired candidates.

SYLLABUS

Sub Unit – 1: Data Interpretation

SL. NO	TOPICS
1	1. Data
2	2. Sources Acquisition and Interpretation of Data
3	2.1. Primary Data
4	2.2. Secondary Data
5	2.3. Sources of Secondary Data
6	2.3.a. Published Sources of Collecting Secondary Data
7	2.3.b. Unpublished Source
8	2.3.c. Information Technology Source
9	2.3.d. Tale Source
10	2.3.e. E-mail Source
11	3. Concept of Proportion, Percentage, Average and Ratio
12	3.1. Proportions
13	3.2. Percentage
14	3.3. Average
15	4. Some concepts of statistics
16	4.1 Statistics
17	4.2. Probability
18	4.3. Sampling
19	4.4. Correlation Coefficient and Rank of Correlation
20	4.5. Dispersion
21	4.7. Normal Distribution

Sub Unit – 2: Quantitative and Qualitative Data

SL. NO	TOPICS
22	5. Quantitative Data
23	6. Qualitative Data
24	7. Difference between Quantitative and Qualitative data
25	8. Importance of Quantitative and Qualitative Data
26	9. Example
27	10. Variable
28	10.1. Discrete Variable
29	10.2 Continuous Variable

Sub Unit – 3: Graphical Representations

SL. NO	TOPICS
30	11. Data Presentation
31	11.1. Textual or Descriptive Presentation of Data
32	11.2. Diagrammatic Presentation of Data
33	11.2.a. Geometric Diagram
34	11.2.b. Frequency Diagram
35	11.2.c. Arithmetic Line Graph or Time Series Graph
36	11.2.d. Tabular Presentation
37	11.2.d.1. Qualitative Classification
38	11.2.d.2. Quantitative Classification
39	11.2.d.3. Temporal Classification
40	11.2.d.4. Spatial Classification
41	11.2.d.5. Normative Classification
42	12. Bar Diagram
43	12.1. Construction of a Bar Diagram
44	12.2. Properties of a Bar Diagram
45	13. Line Diagram
46	13.1. Properties of a Line Diagram
47	14. Pie Diagram
48	15. Histogram
49	15.1. Construction of Histogram
50	15.2. Properties of Histogram
51	16. Mapping
52	16.1. Cartography
53	16.2. Cartogram
54	16.3. Types of Maps
55	16.3.a. General vs. Thematic Cartography
56	16.3.b. Topographic vs. Topological
57	16.3.c. Counter Line

Sub Unit – 4: Data Interpretation

SL. NO	TOPICS
58	17. Concept
59	18. Importance of Data Interpretation
60	19. Different ways of Data Interpretation
61	19.1. Tables with Example:
62	19.2. Pie Diagram:
63	19.3. Bar Diagram:
64	19.4. Line Diagram:
65	19.5. Histogram:
66	20.Various Concepts Required for Solving Questions on Data Interpretation
67	21. Data Interpretation Tricks to solve Data Interpretation

Sub Unit – 5: Data and Governance

SL. NO	TOPICS
68	22. Concept
69	23. Why Data Governance is needed
70	24. What are the benefits if data a governance
71	25. Role of Data Governance

Section – 1: At a Glance

Sub Unit – 1: Data Interpretation

SOURCES OF ACQUISITION AND INTERPRETATION OF DATA: Data interpretation refers to the implementation of processes through which data is reviewed for the purpose of arriving at an informed conclusion. The interpretation of data assigns a meaning to the information analysed and determines its signification and implications.

Data acquisition is the process of sampling signals that measure real world physical conditions and converting the resulting samples into digital numeric values that can be manipulated by a computer. Data acquisition systems, abbreviated by the acronyms DAS or DAQ, typically convert analog waveforms into digital values for processing. The components of data acquisition systems include:

- Sensors, to convert physical parameters to electrical signals.
- Signal conditioning circuitry, to convert sensor signals into a form that can be converted to digital values.
- Analog-to-digital converters, to convert conditioned sensor signals to digital values.

TYPES OF DATA: Primary and secondary data there are many ways of classifying data.

A common classification is based upon who collected the data.

PRIMARY DATA: Data collected by the investigator himself/ herself for a specific purpose.

Examples: Data collected by a student for his/her thesis or research project.

(In movies) The hero is directly told by the heroine that he is her "ideal man".

SECONDARY DATA: Data collected by someone else for some other purpose (but being utilized by the investigator for another purpose).

Examples: Census data being used to analyse the impact of education on career choice and earning.

(In movies) The hero reads a fictional account of the heroine's "ideal man" (written for a course in English composition) that seems to describe him accurately. He seeks confirmation from his friends, concluding that *he* is her "ideal man". (He never asks her directly, but assumes the "facts" are correct).

PUBLISHED AND UNPUBLISHED SOURCES OF SECONDARY DATA: Printed & Published Documents. Primary source documents can be letters, diaries, newspapers, speeches, interviews, memoirs, minutes, fliers, manifestos, reports, pamphlets, handbills, government documents, or other types of texts Unpublished refers to any information source that is not officially released by an individual, publishing house, or other company, and can include both paper and electronic sources

STATISTICS AND PARAMETERS: Parameters in statistics is an important component of any statistical analysis. In simple words, a parameter is any numerical quantity that characterizes a given population or some aspect of it. This means the parameter tells us something about the whole population**.**

Sub Unit – 2: Quantitative and Qualitative Data

QUANTITAVE DATA: Variables whose values result from counting or measuring something.

Examples: height, weight, time in the 100yard dash, number of items sold to a shopper

QUALITATIVE DATA: Variables that are not measurement variables. Their values do not result from measuring or counting.

Examples: hair colour, religion, political party, profession.

DISCRETE VARIABLE: A discrete variable is a variable whose value is obtained by counting. The value is whole number and not in fractions.

Examples: Number of students present, number of red marbles in a jar, number of heads when flipping three coins.

CONTINUOUS VARIABLE: A continuous variable is a variable whose value is obtained by measuring. The value is in fractions.

Examples: height of students in class, weight of students in class, time it takes to get to school, distance travelled between classes.

Sub Unit – 3: Graphical Representation

DATA PRESENTATION: FREQUENCY DISTRIBUTION: Data can be presented in various forms depending on the type of data collected. A frequency distribution is a table showing how often each value (or set of values) of the variable in question occurs in a data set. A frequency table is used to summarize categorical or numerical data. Frequencies are also presented as relative frequencies, that is, the percentage of the total number in the sample.

EXAMPLE: Frequency distribution of peptic ulcer according to site of ulcer

Site of ulcer	Frequency	Percent
Gastric ulcer	24	30
Duodenal ulcer	50	62.5
Gastric and duodenal ulcer	6	7.5
TOTAL	80	100

GRAPHICAL METHODS: Frequency distributions and are usually illustrated graphically by plotting various types of graphs:

BAR GRAPH - A bar graph is a way of summarizing a set of categorical data. It displays the data using a number of rectangles, of the same width, each of which represents a particular category.

Bar graphs can be displayed horizontally or vertically and they are usually drawn with a gap between the bars (rectangles).

HISTOGRAM - A histogram is a way of summarizing data that are measured on an interval scale (either discrete or continuous). It is often used in exploratory data analysis to illustrate the features of the distribution of the data in a convenient form.

PIE CHART - A pie chart is used to display a set of categorical data. It is a circle, which is divided into Segments. Each segment represents a particular category. The area of each segment is Proportional to the number of cases in that category.

LINE GRAPH - A line graph is particularly useful when we want to show the trend of a variable over time. Time is displayed on the horizontal axis (x-axis) and the variable is displayed on the vertical axis (y- axis).

QUALITATIVE CLASSIFICATION: In this classification, data in a table is classified on the basis of qualitative attributes. In other words, if the data contained attributes that cannot be quantified like rural-urban, boys-girls etc. it can be identified as a qualitative classification of data

Sex	**Urban**
Boys	200
Girls	167

QUANTITATIVE CLASSIFICATION: In quantitative classification, data is classified on basis of quantitative attributes

Marks	No. of Students
0-50	29
51-100	64

TEMPORAL CLASSIFICATION: Here data is classified according to time. Thus, when data is mentioned with respect to different time frames, we term such a classification as temporal.

Year	**Sales**
2016	10,000
2017	12,500

SPATIAL CLASSIFICATION: When data is classified according to a location, it becomes a spatial classification.

Country	**No. of Teachers**
India	1,39,000
Russia	43,000

CARTOGRAPHY: Cartography is the creation, production, and study of maps. Cartographers are often geographers who specialize in the combination of art, science, and technology to make and study maps. Some cartographers teach mapmaking skills and techniques, some design and produce maps, and some are curators of map libraries**.**

CARTOGRAM: A cartogram is a map in which the geometry of regions is distorted in order to convey the information of an alternate variable. The region area will be inflated or deflated according to its numeric value. Most of the time, a cartogram is also a choropleth map where regions are coloured according to a numeric variable (not necessarily the one use to build the cartogram).

THEMATIC CARTOGRAPHY: A *thematic map* is a map that focuses on a specific *theme* or subject area such as physical phenomena like temperature variation, rainfall distribution and population density in an area. Thematic maps emphasize spatial variation of human issues like population density or prevalence of diseases.

CHOROPLETH MAPS: A choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income.

TOPOMAP: A **topographic map** is a type of map that shows heights that you can measure. A traditional topographic map will have all the same elements as a non-topographical map, such as scale, legend, and north arrow.

TOPOLOGICAL: A topological map is a map that is only concerned with relative locations of features on the map, not on exact locations. A famous example is the graph that we use to solve the Bridges of Konigsberg problem.

Sub Unit – 4: Data Interpretation.

DATA:1: factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation the *data* is plentiful and easily available.

2: information in digital form that can be transmitted or processed.

3: information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful.

DATA INTERPRETATION: Data Interpretation or DI refers to the implementation of procedures through which data is reviewed for the purpose of arriving at an in inference. Data can be obtained from multiple sources e.g. data from running of industries, census population data etc. Interpreting data requires analysing data to infer information from it in order to answer question. Data can be provided in a number of viz: Bars, tables, line graph, pie graphs.

IMPORTANCE OF DATA INTERPRETATION: The purpose of collection and interpretation is to acquire useful and usable information and to make the most informed decisions possible.

DIFFERENT WAYS OF DATA INTERPRETATION: During a study with an aim and multiple objectives, data analysis will be required to complete the required objectives and compiling or presenting the analysed data will help in overall analysis and concluding the study.

You can have a variety of data which can be used in presentations. Some of these types include: Time Series Data, Bar Charts, Combo Charts, Pie Charts, Tables, Geo Map, Scorecard, Scatter Charts, Bullet Charts, Area Chart, Text & Images.

TABLE: One of the simplest methods used to analyse the data and to display the data is in tabular form. In the tabular form, you get a systematic arrangement of rows and columns. The first column is used to indicate the titles and the first row is also used to indicate the same. It is very accurate as well as an easy method to display the data. Although this is a simple method, it can be time-consuming as well.

Sub Unit – 5: Data and Governance

WHAT IS DATA GOVERNANCE? Data governance is a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals. It establishes the processes and responsibilities that ensure the quality and security of the data used across a business or organization. Data governance defines who can take what action, upon what data, in what situations, using what methods.

WHY IS DATA GOVERNANCE IMPORTANT? Data Governance is required to ensure that an organization's information assets are formally, properly, proactively and efficiently

managed throughout the enterprise to secure its trust & accountability. Data Governance comprises the collecting of data, revising and standardizing it, and making it good for use. It makes the data consistent. Data Governance ensures that critical data is available at the right time to the right person, in a standardized and reliable form. This infers into better organization of business operations. Adopting and implementing Data Governance can result in improved productivity and efficiency of an organization.

COMMON GOALS OF DATA GOVERNANCE PROGRAMMES INCLUDE:

- To develop and communicate data strategies, policies, procedures and architecture requirements.
- To sponsor and oversee the effective delivery of data projects and strategy outcomes.
- To define and ensure compliance with data policies, standards and procedures.
- To understand and promote the value of data assets.
- To support the analysis and exchange of data insights for the benefit of the business.

DAMA INTERNATIONA: DAMA (the Data Management Association) is a not-for-profit, vendor-independent, international association of technical and business professionals dedicated to advancing the concepts and practices of information resource management (IRM) and data resource management (DRM).

DATA GOVERNANCE PROFESSIONALS ORGANIZATION (DGPO): The Data Governance Professionals Organization (DGPO) is a non-profit, vendor neutral, association of business, IT and data professionals dedicated to advancing the discipline of data governance. The objective of the DGPO is to provide a forum that fosters discussion and networking for members and to encourage, develop and advance the skills of members working in the data governance discipline.

THE DATA GOVERNANCE SOCIETY: The Data Governance Society, Inc. is dedicated to fostering a new paradigm for the effective use and protection of information in which Data is governed and leveraged as a unique corporate asset.

THE DATA GOVERNANCE COUNCIL: The Data Governance Council is an organization formed by IBM consisting of companies, institutions and technology solution providers with the stated objective to build consistency and quality control in governance, which will help companies better protect critical data.

ENTERPRISE DATA MANAGEMENT COUNCI: The EDM Council is a non-profit organization created to elevate the practice of data management as a business priority. We are advocates for both content standards and data management best practices.

Section – 2: Key Statements

Data (1), Primary Data (2), Secondary Data (3), Published Sourced of Secondary Data (4), Un published source of Secondary Data (5), Tale source (6), E-Mail Source (7), Percentage (8), Average (9), Statistics (10), Probability (11), Sampling (12), Correlation (13), Dispersion (14), Testing in Statistics (15), Normal Distribution(16), Rank of Correlation (17), Quantitative and Qualitative Data (18), Variable (19), Discrete& Continuous Variable (20), Data presentation (21), Quantitative & Qualitative Classification (22), Temporal Classification (23), Spatial classification (24), Line Diagram (25), Bar Diagram (26), Pie Diagram (27), Table Diagram (28), Histogram (29), Mapping (30), Topographic (31), Thematic Cartography (32), Topological (33), Counter Line (34), Data Interpretation (35), Data and Governance (36), Data Governance Council (37), data Governance Society (38), Data Management Council (39).

[N.B. – Numbers in parenthesis are the reference number]

Section – 3: Key Facts and Figures

Sub Unit – 1
Data Interpretation

……

1. Data

Data is a set of values of qualitative or quantitative variables; restated, pieces of data are individual pieces of information. Data is measured, collected and reported, and analysed, whereupon it can be visualized using graphs or images. Data as a general concept refers to the fact that some existing information or knowledge is represented in some form suitable for better usage or processing.

2. Sources Acquisition and Interpretation of Data

Data acquisition is the process of sampling signals that measure real world physical conditions and converting the resulting samples into digital numeric values that can be manipulated by a computer.

Interpretation is the process of making sense of numerical data that has been collected, analyzed, and presented. A common method of assessing numerical data is known as statistical analysis and the activity of analysing and interpreting data in order to make prediction is known as inferential statistics.

Data interpretation needs some mathematical and statistical skills. It needs good knowledge of concept of percentage, ratio, proportion, average etc.

Data may be of two types; primary data and secondary data.

2.1. Primary Data:

Primary data (also known as raw data) is a term for data collected from a source. Primary data has not been subjected to processing or any other manipulation, and are also referred to as raw data. Primary data is a relative term. Primary data can be input to a computer program or used

in manual procedures such as analysing statistics from a survey. The term can refer to the binary data on electronic storage devices such as hard disk drives (also referred to as low-level data).

2.2. Secondary Data:

Secondary data are such numerical information which have previously been collected by someone other than the user or by some agency for one purpose and are merely compiled from that source for use in a different connection. In fact, data collected by someone when used by another, or collected for one purpose when used for another, will be called secondary data. The same data are primary in the hands the collecting authority, but are secondary in the hands of another. For example, the Census figures published by Registrar General of India will be primary data, while the same data contained in any other publication will be called secondary data.

2.3. Sources of Secondary Data:

Secondary data can be obtained from different sources:

(i) Published Sources

(ii) Un published Sources

(iii) Information Technology Source

(iv) Table Source

(iv) E-mail Source

2.3.a. Published Sources of Collecting Secondary Data:

The various sources of published data are:

A. Reports and official publications of-

i) International bodies such as the World Bank; International labour organizations; Statistical office of the United Nations.

ii) Central and State governments such as abstract of the Indian union, economic survey of 2007-08 govt. of India, ministry of fence; Indian 2009, Publication division, ministry of information and broadcasting.

iii) Report of the committees and commission appointed by the government such as report of the committee on corporate governance. (e.g. Sixth Pay Commission)

B. Semi-official publications of various local bodies such as municipal corporations and districts boards.

C. Publications of autonomous and private institute such as-

i) Trade and professional bodies, such as the federation of Indian Chambers of Commerce and industry, the institute of Chartered accountants, the institute of foreign trade.
ii) Financial and economic journals such as Indian economic review, Reserve Bank of India bulletin, Indian finance.
iii) Annual report of joint stock companies and corporation.

2.3.b. Unpublished Source:

All statistical material is not always published. There are various sources of unpublished data such as records maintained by various government and private offices, Studies made by research institutions, Schools etc. Such sources can be used where necessary.

2.3.c. Information Technology Source:

Example: Telephone and radio equipment and switches used for voice communications. Traditional computer applications that include data storage and programme to input, process and output the data. Data, voice and video networks and all associated communications equipment and software.
Information technology means the use of hardware, software, services and supporting infrastructure to manage and deliver information using voice, data and video. To further this definition for the sake of it budgeting, ITD offers the following guidance.

2.3.d. Tale Source:

Online database that uses various data sources to produce customizable tables on AIDS, births, cancer statistics, environmental data.

2.3.e. E-mail Source:

It is a relatively inexpensive method of collecting data and one that can distribute large numbers of questionnaires in a short time. It provides the opportunity to contact hard-to-reach people and respondents are able to complete the questionnaire in their own time. Mail surveys however, do require an up-to-date list of names and address. In addition there is also the need to keep the questionnaire simple and straight forward. A major disadvantage of a mail survey is that it usually has lower response rates than other data acquisition method. This may lead to problems with data quality. Also, people with a limited ability to read or write English may experience problem.

3. Concept of Proportion, Percentage, Average and Ratio

3.1. Proportions:

A proportion is an equation that says that two or more ratios are equal.

For instance it one package of contain 20 cookies that would mean that 2 packages contain 40 cookies $\frac{20}{1} = \frac{40}{2}$

A proportion is read as "x is to y as a is to b" $\frac{X}{Y} = \frac{a}{b}$

3.2. Percentage:

Percentage means hundredths or per hundred and is written with the symbol, %. Percentage is a ratio were we compare numbers to 100 which means that 1% is $\frac{1}{100}$

Eg : 89% as a decimal

$\frac{x}{100} = \frac{89}{100} = 0.89$

0.27 in percentage

$=0.27 \rightarrow \frac{0.27}{100} \rightarrow 27\%$

3.3. Average:

The word "average" is used in everyday life to describe where the **middle number of a data set** is. It's the typical number you would expect to find in a series of numbers. In statistics, the average is called the "arithmetic mean," usually just shortened to *the mean*. Both the average and the mean use the same formula:

AVG = total sum of all the numbers / number of items in the set.

In other words, to find the average, add up all of the numbers in the set, and then divide by however many items you have. Let's say you have 5,10, and 15. Add them all up to get 5 + 10 + 15 = 30, then divide by 3(the number of items). The answer is 30 / 3 = 10.

3.4. Example:

1. Which of the following methods is best suited to show on map the types of crops grown in a region?

   (A) Choropleth (B) Chorochromatic

(C) Choro schematic (D) Isopleth

2. Which of the following sources of data is not based on primary data collection?

(A) Census of India (B) National Sample Survey

(C) Statistical Abstracts of India (D) National Family Health Survey

3. Which of the following is not a source of data?

(A) Administrative Records (B) Population Census

(C) GIS **(D) Sample Survey**

4. Which of the following is the oldest archival source of data in India?

(A) National Sample Survey **(B) Agriculture Statistic**

(C) Census (D) Vital Statistics

5. Data collected from the field by the researcher is called

(A) Primary data (B) numerical data

(C) Statistical data (D) Secondary data

6. One type of sampling that is inherently biased but is still random is

(A) Biased sampling **(B) purposive random sampling**

(C) Systematic sampling (D) Stratified sampling

7. If the total population of states and the absolute population of their respective capitals are to be represented on the same map, which of the following would be the most appropriate method?

(A) Dotes and proportionate circles (B) Dotes and choropleth methods

(C) Dotes Isopleths (D) Dotes and chorochromatic methods

Study the data in the given table and answer the questions following it.

Sl. No.	Rainfall(mm)	No. of Stations
1.	20-30	9
2.	30-40	15
3.	40-50	20
4.	50-60	10

5.	60-70	7

8. How many stations receive a rainfall of up to 50 mm?

   (A) 24 **(B) 44**

   (C) 54 (D) 35

9. How many stations receive a rainfall amounting to more than 40 mm?

   (A) 37 (B) 52

   (C) 17 (D) 30

10. How many stations receive a rainfall between 30 and 60 mm?

    (A) 45 (B) 54

    (C) 30 (D) 44

4. Some concepts of statistics

4.1 Statistics:

A statistic (singular) or sample statistic is a single measure of some attribute of a sample (e.g. its arithmetic mean value). It is calculated by applying a function (statistical algorithm) to the values of the items of the sample, which are known together as a set of data.

More formally, statistical theory defines a statistic as a function of a sample where the function itself is independent of the unknown estimator; that is, the function is strictly a function of the data. The term statistic is used both for the function and for the value of the function on a given sample.

A statistic is distinct from a statistical parameter, which is not computable in cases where the population is infinite, and therefore impossible to examine and measure all its items. However, a statistic, when used to estimate a population parameter, is called an estimator. For instance, the sample mean is a statistic that estimates the *population mean*, which is a parameter.

When a statistic (a function) is being used for a specific purpose, it may be referred to by a name indicating its purpose: in descriptive statistics, a descriptive statistic is used to describe the data; in estimation theory, an estimator is used to estimate a parameter of the distribution (population); in statistical hypothesis testing, a test statistic is used to test a hypothesis. However, a single statistic can be used for multiple purposes – for example the sample mean can be used to describe a data set, to estimate the population mean, or to test a hypothesis.

Example:

In calculating the arithmetic mean of a sample, for example, the algorithm works by summing all the data values observed in the sample and then dividing this sum by the number of data items. This single measure, the mean of the sample, is called a statistic; its value is frequently used as an estimate of the mean value of all items comprising the population from which the sample is drawn. The population mean is also a single measure; however, it is not called a statistic, because it is not obtained from a sample; instead it is called a population parameter, because it is obtained from the whole population.

Other examples of statistics include

- Sample mean discussed in the example above and sample median
- Sample variance and sample standard deviation
- Sample quantiles besides the median, e.g., quartiles and percentiles
- Test statistics, such as t statistics, chi-squared statistics, f statistics
- Order statistics, including sample maximum and minimum
- Sample moments and functions thereof, including kurtosis and skewness.

Various functionals of the empirical distribution function

Statistical properties:

Important potential properties of statistics

include completeness, consistency, sufficiency, unbiasedness, minimum mean square error, low variance, robustness, and computational convenience.

4.2. Probability:

The **probability** of an event refers to the likelihood that the event will occur.

How to Interpret Probability?

Mathematically, the probability that an event will occur is expressed as a number between 0 and 1. Notationally, the probability of event A is represented by P(A).

- If P(A) equals zero, event A will almost definitely not occur.
- If P(A) is close to zero, there is only a small chance that event A will occur.
- If P(A) equals 0.5, there is a 50-50 chance that event A will occur.
- If P(A) is close to one, there is a strong chance that event A will occur.
- If P(A) equals one, event A will almost definitely occur.

In a statistical experiment, the sum of probabilities for all possible outcomes is equal to one. This means, for example, that if an experiment can have three possible outcomes (A, B, and C), then P (A) + P (B) + P (C) = 1.

How to Compute Probability: Equally Likely Outcomes

Sometimes, a statistical experiment can have *n* possible outcomes, each of which is equally likely. Suppose a subset of *r* outcomes are classified as "successful" outcomes.

The probability that the experiment results in a successful outcome (S) is:

P(S) = (Number of successful outcomes) / (Total number of equally likely outcomes) = r / n

$$P(s) = \frac{\text{(Number of successful outcomes)}}{\text{(Total number of equally likely outcomes)}} = \frac{\text{r}}{\text{n}}$$

Consider the following experiment. An urn has 10 marbles. Two marbles are red, three are green, and five are blue. If an experimenter randomly selects 1 marble from the urn, what is the probability that it will be green?

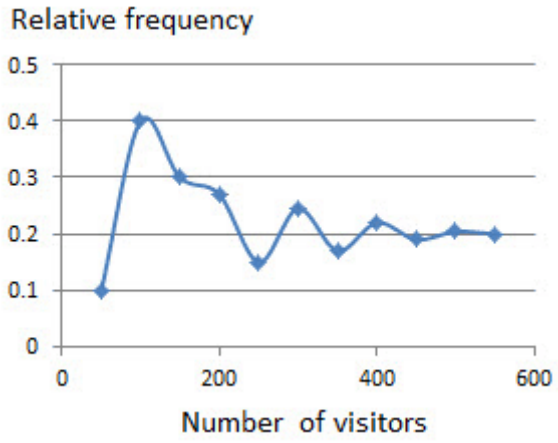
In this experiment, there are 10 equally likely outcomes, three of which are green marbles. Therefore, the probability of choosing a green marble is 3/10 or 0.30.

How to Compute Probability: Law of Large Numbers

One can also think about the probability of an event in terms of its *long-run* relative frequency. The relative frequency of an event is the number of times an event occurs, divided by the total number of trials.

P (A) = (Frequency of Event A) / (Number of Trials)

For example, a merchant notices one day that 5 out of 50 visitors to her store make a purchase. The next day, 20 out of 50 visitors make a purchase. The two relative frequencies (5/50 or 0.10 and 20/50 or 0.40) differ. However, summing results over many visitors, she might find that the probability that a visitor makes a purchase gets closer and closer to 0.20.



The scatterplot above shows the relative frequency of purchase as the number of trials (in this case, the number of visitors) increases. Over many trials, the relative frequency converges toward a stable value (0.20), which can be interpreted as the probability that a visitor to the store will make a purchase.

The idea that the relative frequency of an event will converge on the probability of the event, as the number of trials increases, is called the **law of large numbers.**

Test Your Understanding:

Problem:

1. A coin is tossed three times. What is the probability that it lands on heads *exactly* one time?

(A) 0.125

(B) 0.250

(C) 0.333

(D) 0.375

(E) 0.500

Solution:

The correct answer is (D). If you toss a coin three times, there are a total of eight possible outcomes. They are: HHH, HHT, HTH, THH, HTT, THT, TTH, and TTT. Of the eight possible outcomes, three have exactly one head. They are: HTT, THT, and TTH. Therefore, the probability that three flips of a coin will produce *exactly* one head is 3/8 or 0.375.

4.3. Sampling:

What are Sampling Methods?

In a statistical study, sampling methods refer to how we select members from the population to be in the study.

If a sample isn't randomly selected, it will probably be biased in some way and the data may not be representative of the population.

There are many ways to select a sample—some good and some bad.

Bad ways to sample

Convenience sample: The researcher chooses a sample that is readily available in some non-random way.

Example: A researcher polls people as they walk by on the street.

Why it's probably biased: The location and time of day and other factors may produce a biased sample of people.

Voluntary response sample: The researcher puts out a request for members of a population to join the sample, and people decide whether or not to be in the sample.

Example: A TV show host asks his viewers to visit his website and respond to an online poll.

Why it's probably biased: People who take the time to respond tend to have similarly strong opinions compared to the rest of the population.

PRACTICE PROBLEM 1:

A restaurant leaves comment cards on all of its tables and encourages customers to participate in a brief survey to learn about their overall experience.

Good ways to sample

Simple random sample: Every member and set of members has an equal chance of being included in the sample. Technology, random number generators, or some other sort of chance process is needed to get a simple random sample.

Example: A teachers puts students' names in a hat and chooses without looking to get a sample of students.

Why it's good: Random samples are usually fairly representative since they don't favor certain members.

Stratified random sample: The population is first split into groups. The overall sample consists of some members from every group. The members from each group are chosen randomly.

Example: A student council surveys 100100100 students by getting random samples of 252525 freshmen, 252525 sophomores, 252525 juniors, and 252525 seniors.

Why it's good: A stratified sample guarantees that members from each group will be represented in the sample, so this sampling method is good when we want some members from every group.

Cluster random sample: The population is first split into groups. The overall sample consists of every member from some of the groups. The groups are selected at random.

Example: An airline company wants to survey its customers one day, so they randomly select 555 flights that day and survey every passenger on those flights.

Why it's good: A cluster sample gets every member from some of the groups, so it's good when each group reflects the population as a whole.

Systematic random sample: Members of the population are put in some order. A starting point is selected at random, and every 9^{th} member is selected to be in the sample.

Example—A principal takes an alphabetized list of student names and picks a random starting point. Every 20th student is selected to take a survey.

4.4. Correlation Coefficient and Rank of Correlation:

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

A correlation coefficient is a statistical measure of the degree to which changes to the value of one variable predict change to the value of another. When the fluctuation of one variable reliably predicts a similar fluctuation in another variable, there's often a tendency to think that means that the change in one causes the change in the other. However, correlation does not imply causation. There may be, for example, an unknown factor that influences both variables similarly.

Here's one example: A number of studies report a positive correlation between the amount of television children watch and the likelihood that they will become bullies. Media coverage often cites such studies to suggest that watching a lot of television causes children to become bullies. However, the studies only report a correlation, not causation. It is likely that some other factor – such as a lack of parental supervision – may be the influential factor.

Rank of correlation.

In statistics, a rank correlation is any of several statistics that measure an ordinal association the relationship between rankings of different ordinal variables or different rankings of the same variable, where a "ranking" is the assignment of the ordering labels "first", "second", "third", etc. to different observations of a particular variable. A rank correlation coefficient measures the degree of similarity between two rankings, and can be used to assess the significance of the relation between them. For example, two common nonparametric methods of significance that use rank correlation are the Mann–Whitney U test and the Wilcoxon signed-rank test.

4.5. Dispersion:

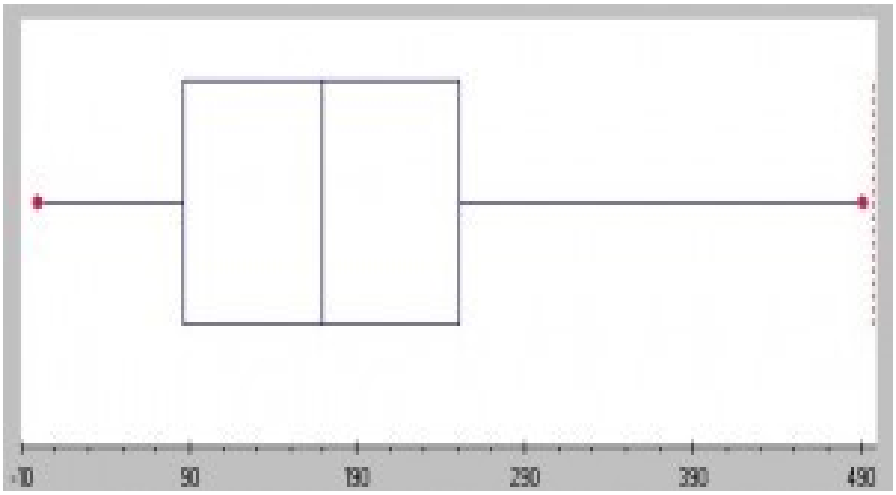
Dispersion in statistics is a way of describing how spread out a set of data is. When a data set has a large value, the values in the set are widely scattered; when it is small the items in the set are tightly clustered. Very basically, this set of data has a small value:

1, 2, 2, 3, 3, 4

…and this set has a wider one:

0, 1, 20, 30, 40, 100

The spread of a data set can be described by a range of descriptive statistics including variance, standard deviation, and interquartile range. Spread can also be shown in graphs: dot plots, boxplots, and stem and leaf plots have a greater distance with samples that have a larger dispersion and vice versa.



The larger the box, the more dispersion in a set of data. Image: Seton Hall University

Measures of Dispersion:

- **Coefficient of dispersion**: A "catch-all" term for a variety of formulas, including distance between quartiles.
- **Standard deviation**: probably the most common measure. It tells you how spread out numbers are from the mean.
- **Index of Dispersion**: a measure of dispersion commonly used with nominal variables.
- **Interquartile range (IQR)**: describes where the bulk of the data lies (the "middle fifty" percent).
- **Inter-decile range**: the difference between the first decile (10%) and the last decile (90%).
- Range: the difference between the smallest and largest number in a set of data.

- **Mean difference or difference in means:** measures the absolute difference between the mean value in two different groups in clinical trials.
- **Median absolute deviation (MAD)**: the median of the absolute deviations from a data set's median.

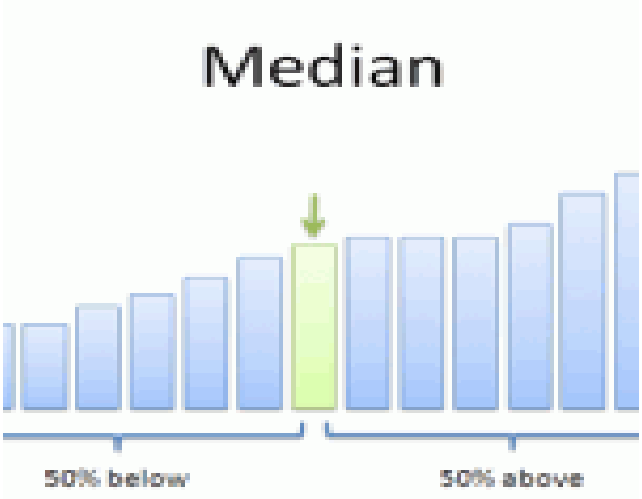
Quartiles:

Numbers that split the data into four quarters (first, second, third, and fourth quartiles).

In some processes, like manufacturing or measurement, low dispersion is associated with high precision. High dispersion is associated with low precision.

Example:

Let's say you were asked to compare measures of dispersion for two data sets. Data set A has the items 97,98,99,100,101,102,103 and data set B has items 70,80,90,100,110,120,130. By looking at the data sets you can probably tell that the means and medians are the same (100) which technically are called "measures of central tendency" in statistics.



However, the range (which gives you an idea of how spread out the entire set of data is) is much larger for data set B (60) when compared to data set A (6). In fact, nearly all measures of dispersion would be ten times greater for data set B, which makes sense as the range is ten times larger. For example, take a look at the standard deviations for the two data sets:

Standard deviation for A: 2.160246899469287.

Standard deviation for B: 21.602468994692867.

The figure for data set B is exactly ten times that of A.

4.7. Normal Distribution:

The normal distribution is the most common type of distribution assumed in technical stock market analysis and in other types of statistical analyses. The standard normal distribution has two parameters: the mean and the standard deviation. For a normal distribution, 68% of the observations are within +/- one standard deviation of the mean, 95% are within +/- two standard deviations, and 99.7% are within +- three standard deviations.

Sub Unit – 2
Quantitative and Qualitative data

5. Quantitative Data

Quantitative data, as the name suggests is one which deals with quantity or numbers. It refers to the data which computes the values and counts and can be expressed in numerical terms is called quantitative data. In statistics, most of the analyses are conducted using this data. Quantitative data may be used in computation and statistical test. It is concerned with measurements like height, weight, volume, length, size, humidity, speed, age, etc. The tabular and diagrammatic presentation of data is also possible, in the form of charts, graphs, tables, etc. Further, the quantitative data can be classified as discrete or continuous data.

6. Qualitative Data

Qualitative data refers to the data that provides insights and understanding about a particular problem. It can be approximated but cannot be computed. Hence, the researcher should possess complete knowledge about the type of characteristic, prior to the collection of data. The nature of data is descriptive and so it is a bit difficult to analyze it. This type of data can be classified into categories, on the basis of physical attributes and properties of the object. The data is interpreted as spoken or written narratives rather than numbers. It is concerned with the data that is observable in terms of smell, appearance, taste, feel, texture, gender, nationality and so on.

7. Difference between Quantitative and Qualitative data

There is some stark difference between quantitative data and qualitative data. They are

Sl. No	Quantitative	Qualitative

I	Associated with number	Associated with details
II	Implemented when data is numerical	Implemented when data can be segregated into well-defined groups
III	Collected data can be statistically analysed	Collected data can just be observed and not evaluated
IV	Examples: Height, weight, Time, Price, Temperature, etc	Example: Scents, Appearance, Beauty, Colours, Flavours, etc

8. Importance of Quantitative and Qualitative Data

Quantitative and Qualitative data provide different outcomes, and are often used together to get a full picture of a population for example, if data are collected an annual income (quantitative), occupation data on the average annual income for each type of occupation.

Quantitative and Qualitative data can be gathered from the same data unit depending on whether the variable of interest is numerical or categorical.

Example:

Data Unit	Numeric variable	= Quantitative	Categorical Variable	= Qualitative
A Person	How many children do you have	4 Children	In which country were your children born	Australia
	How much do you earn	$60,000 p.a	What is your occupation	Photography
	How many hours do you work	38 hours per week	Do you work full time or part time	Full time
A farm	How many milk cows are located on the farm	36 Cows	What is main activity of the farm?	Dairy

9. Example

1. Which of the following is the correct expression for information that indicates just the occurrence or non-occurrence of phenomenon?

(A) Numerical data
(B) Quantitative data
(C) Qualitative data
(D) Non-data information

2. When data states not just occurrence or non-occurrence of phenomena but also gives an indication of the magnitude or intensity etc. the information is called
(A) Statistical data
(B) Quantitative data
(C) Qualitative data
(D) Tangible data

3. A list of cities with their population mentioned against each one is an example of
(A) Areal data
(B) Linear data
(C) Point data
(D) Numerical data

4. Information about types of soils found in different parts of a country (telling only the types of soils found) is an example of
(A) Point specific qualitative data
(B) Point specific data
(C) Area specific quantitative data
(D) Line specific quantitative data

5. Rainfall statistics for rain-gauge stations in a region is an example of
(A) Point specific qualitative data
(B) Point specific quantitative data
(C) Area specific quantitative data
(D) Line specific quantitative data

6. Which of the following is an example of line specific quantitative data?
(A) Number of houses in a settlement
(B) Number of fields in a village
(C) Number of vehicles using a road
(D) Number of fliers using an airline

10. Variable

A variable is any characteristics, number or quantity that can be measured or contented. A variable may also be called a data item. Age, sex, business income and expenses, country of birth, capital expenditure, class grades, eye colour and vehicle type are example of variables.

10.1: Discrete Variable:

Variables that can only take on a finite number of values are called "discrete variables". All qualitative variables are discrete. Some quantitative variables are discrete, such as performance rated as 1,2,3,4 or 5 or temperature rounded to the nearest degree.

10.2 Continuous Variable:

A continuous variable is a variable that has an infinite number of possible values. In other words, any value is possible for the variable.

A continuous variable is the opposite of a discrete variable, which can only take on a certain number of values.

Continuous variables could be 1,2 and everything in between: 1.00, 1.01, 1.001, 1.0001....

Examples of continues variable are family size, intelligence, attitude etc.

Sub Unit – 3
Graphical Representations

.....

11. Data Presentation

Presentation of data refers to an exhibition or putting data in an attractive and useful manner such that it can be early interpreted. The three main forms of presentation of data are: Textual presentation, Data tables, Diagrammatic Presentation.

11.1. Textual or Descriptive Presentation of Data:

The discussion about the presentation of data starts off with its must raw and vague form which is the textual presentation, data is simply mentioned as more text that is generally in a paragraph. This is commonly used when the data is not very large.

11.2. Diagrammatic Presentation of Data:

Diagrams play an important role in statistical data presentation. Diagrams are nothing but geometrical figures like lines, bars, circles, squares etc. Diagrammatic data presentation allows us to understand the data in an easier manner.

11.2.a. Geometric Diagram:

Geometry diagrams are a distinct mode of mathematical communication. A geometric figure can be defined as a set of points that satisfy some given properties and a geometry diagram can be used as a visual representation of such a figure.

11.2.b. Frequency Diagram:

A frequency diagram, often called a line chart or a frequency polygene, shows the frequencies for different groups. The frequency charts below show the results of the table. To plot a frequency polygon of grouped data, plot the frequency at the midpoint of each group.

11.2.c. Arithmetic Line Graph or Time Series Graph:

An arithmetic line graph is also called time series graph and is a method of diagrammatic presentation of data. In it time (hour, day/date, week, month, year etc.) is plotted along x-axis and the value of the variable (time series data) along y-axis. A line graph by joining these plotted points, thus, obtained is called arithmetic line graph (time series graph). It helps in understanding the trend, periodicity etc. in a long term time series data.

11.2.d. Tabular Presentation:

Many application areas require a presentation of data where the nodes of a diagram are organized in a tabular way i.e. where each node is associated to a specific row and column in the grid-like structure of a table. Swim lane layouts are a popular example of such presentations.

11.2.d.1. Qualitative Classification:

Qualitative data occurs when the data are classified on the basis of qualitative characteristics for a phenomenon.

E.g. Unemployment in Punjab by sex and location

Location		
Sex	Rural	Urban
Male	20	10
Female	30	20
Total	50	30

11.2.d.2. Quantitative Classification:

It temporal classification data is classified on the basis of quantitative characteristics of a phenomenon.

E.g.: Marks obtained by students of class 11^{th}

Marks	No. of Students
0-10	3
10-20	7

20-30	12
30-40	22
40-50	32

11.2.d.3. Temporal Classification:

In this classification time becomes the classifying variable and data are categorized according to time. Time may be in hours, days, weeks, months, years.

Production of India in different years Production (In units)

Year	Unit
1980	580
1990	682
2000	978
2010	1215

11.2.d.4. Spatial Classification:

In spatial classification the data are classified with reference to geographical locations such as countries, states, cities, districts etc.

Country	Yield of wheat (kg/acre)
America	1925
Brazil	127
China	893
Denmark	225
India	862

11.2.d.5. Normative Classification

Normative data is **data** from a reference population that establishes a baseline distribution for a score or measurement, and against which the score or measurement can be compared. **Normative data** is typically obtained from a large, randomly selected representative sample from the wider population.

12. Bar Diagram

This is one of the simplest techniques to do the comparison for a given set of data. A bar graph is a graphical representation of the data in the form of rectangular bars or columns of equal width. It is the simplest one and easily understandable among the graphs by a group of people.

12.1. Construction of a Bar Diagram:

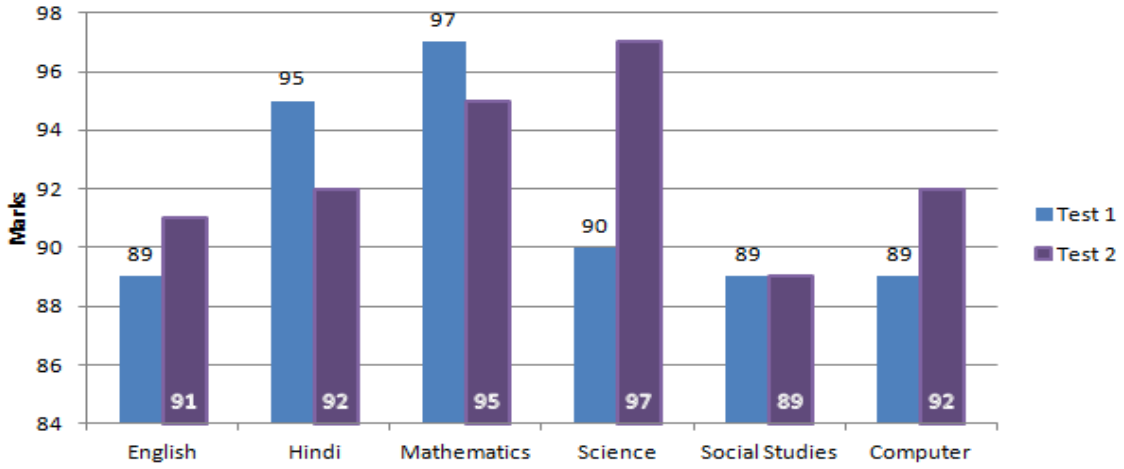
1. Draw two perpendicular lines intersecting each other at a point O. The vertical line is the y-axis and the horizontal is the x-axis.
2. Choose a suitable scale to determine the height of each bar.
3. On the horizontal line, draw the bars at equal distance with corresponding heights.
4. The space between the bars should be equal.

12.2. Properties of a Bar Diagram:

(i) Each bar or column in a bar graph is of equal width.
(ii) All bars have a common base.
(iii) The height of the bar corresponds to the value of the data.
(iv) The distance between each bar is the same.

12.3. Types of Bar Diagram:

A bar graph can be either vertical or horizontal depending upon the choice of the axis as the base. The horizontal bar diagram is used for qualitative data. The vertical bar diagram is used for the quantitative data or time series data. Let us take an example of a bar graph showing the comparison of marks of a student in all subjects out of 100 marks for two tests.



With the bar graph, we can also compare the marks of students in each subject other than the marks of one student in every subject. Also, we can draw the bar graph for every student in all subjects.

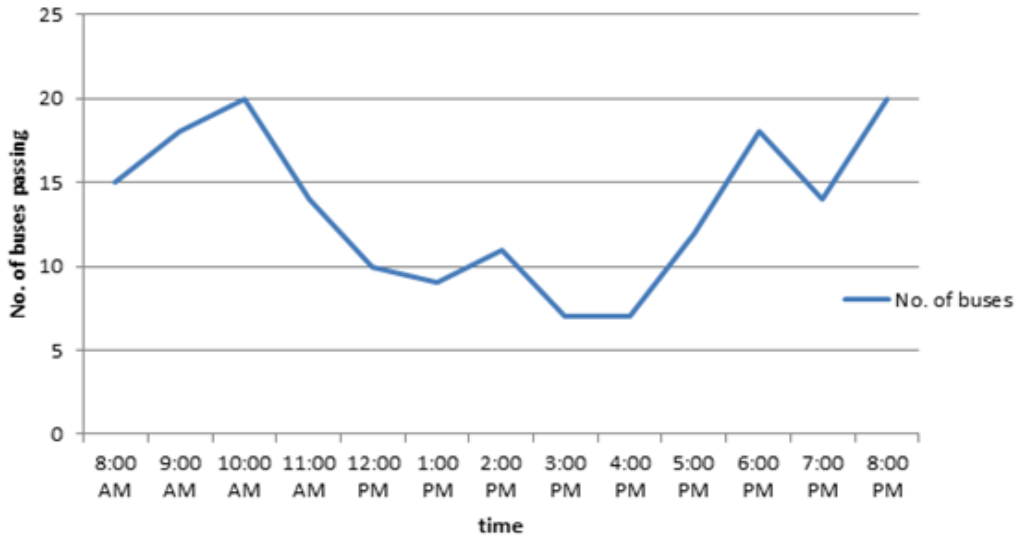
13. Line Diagram

A line graph is a type of chart or graph which shows information when a series of data is joined by a line. It shows the changes in the data over a period of time. In a simple line graph, we plot each pair of values of (x, y). Here, the x-axis denotes the various time point (t), and the y-axis denotes the observation based on the time.

13.1. Properties of a Line Diagram:

- It consists of Vertical and Horizontal scales. These scales may or may not be uniform.
- Data point corresponds to the change over a period of time.
- The line joining these data points shows the trend of change.

Below is the line graph showing the number of buses passing through a particular street over a period of time:

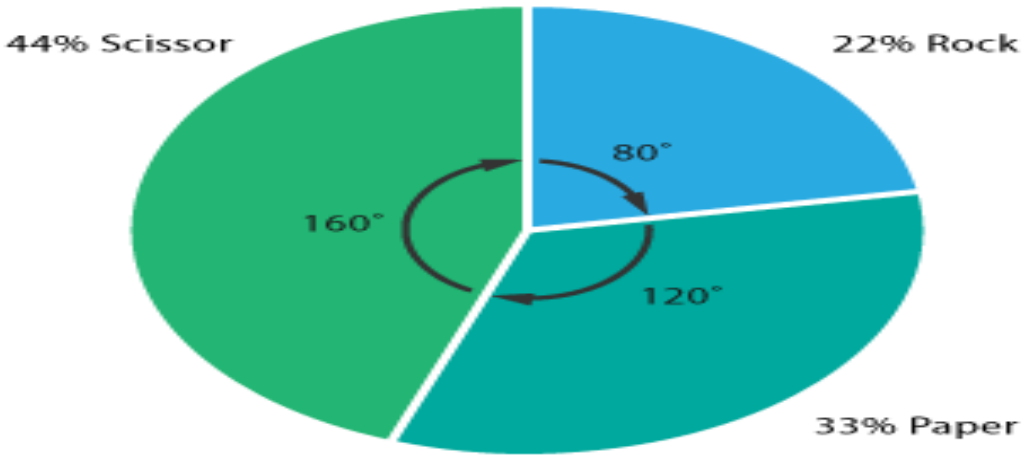


14. Pie Diagram

A pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice, is proportional to the quantity it represents.

Extensively used in presentations and offices, Pie Charts help show proportions and percentages between categories, by dividing a circle into proportional segments. Each arc length represents a proportion of each category, while the full circle represents the total sum of all the data, equal to 100%.

Anatomy:



Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
2/9=22%	3/9=33%	4/9=44%	100%
Degrees for each "pie slice"			
(2/9) x 360 = 80°	(3/9) x 360 = 120°	(4/9) x 360 = 160°	360°

15. Histogram

We can use another way of diagrammatical representation of data. If we are working with a continuous data set or grouped dataset, we can use a histogram for the representation of data.

- A histogram is similar to a bar graph except for the fact that there is no gap between the rectangular bars. The rectangular bars show the area proportional to the frequency of a variable and the width of the bars represents the class width or class interval.
- Frequency means the number of times a variable is occurring or is present. It is an area graph. The heights of the rectangles are proportional to the corresponding frequencies of similar classes.

15.1. Construction of Histogram:

(i) Draw two perpendicular lines intersecting each other at a point O. The vertical line is the y-axis and the horizontal is the x-axis.

(ii) Choose a suitable scale for both the axes to determine the height and width of each bar

(iii)On the horizontal line, draw the bars with corresponding heights

(iv)There should be no gap between two consecutive bars showing the continuity of the data

(v) If the grouped frequencies are not continuous, the first thing to do is to make them continuous

It is done by adding the average of the difference between the lower limit of the class interval and the upper limit of the preceding class width to the upper limits of all the classes. The same quantity is subtracted from the lower limits of the classes.

15.2. Properties of Histogram:

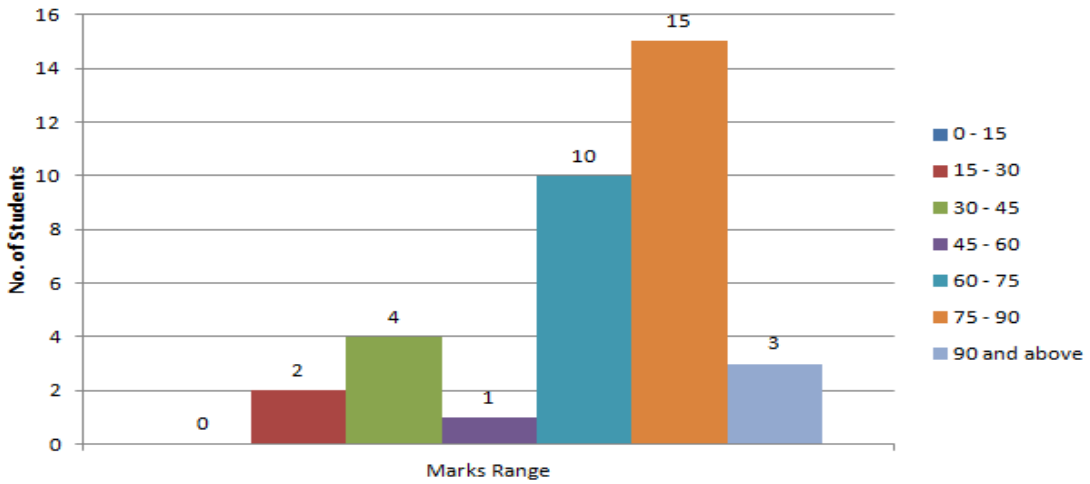
(i) Each bar or column in a bar graph is of equal width and corresponds to the equal class interval

(ii) If the classes are of unequal width then the height of the bars will be proportional to the ration of the frequencies to the width of the classes

(iii) All bars have a common base

(iv) The height of the bar corresponds to the frequency of the data

Suppose we have a data set showing the marks obtained out of 100 by a group of 35 students in statistics. We can find the number of students in the various marks category with the help of the histogram.



16. Mapping

The definition of mapping is making a map or a matching process where the points of one set are matched against the points of another set.

An example of mapping is identifying which call on one spreadsheet contains the same information as the call on another spreadsheet.

16.1. Cartography:

Cartography is the study and practice of making maps. Combining science, aesthetics and technique, cartography builds on the premise that reality can be melded in ways that communicate spatial information effectively.

16.2. Cartogram:

A cartogram is a map in which the geometry of regions is distorted in order to convey the information of an alternate variable.

For example, a world map in which the countries with a large population are drawn wider. Cartograms are particularly useful to correct.

16.3. Types of Maps:

An overview of the different types of maps including to pographic, climatological, thematic maps, weather maps and politic maps.

A map is a symbolic representation of an actual element or an area on a flat surface. Maps are useful because they illustrate specific and detailed features of a given area, region, or an object. The different types of maps are given below-

(i) Topographical Map:

A topographic map is a type of map that shows large-scale detail and quantitative information of relief by the use of contour lines and various methods. Topographic maps have various uses including geographical planning, civil engineering, large-scale architecture and for recreation such as hiking.

(ii) Geological Map:

A geological map is a map that is used to show geological features such as geologic strata and rock units. The location of these features underneath the earth's surface are shown by symbols or colours.

(iii) Political Map:

World maps are generally political or physical. A political map is a map that shows the political features of an area, country or region. It shows a region or country's different territorial borders or boundaries, the location of major areas or cities and significant land masses such as bodies of water. One of the main features of political map is the geographical boundaries.

(iv) Physical Map:

Physical maps are maps that have been designed to show the physical or natural landscape features of the earth. The maps are best known to show several geographical features such as soil type, maintains and land use including infrastructural developments such as roads and building.

(v) Road Map:

Road maps also known as route maps, indicate roads and other transport links. They are navigational maps that also include political boundaries making them part of political maps. A collection of road maps bound together in a book is referred to as road atlas. Road maps often use thin lines to indicate minor roads and thicker or bolder colours to indicate major roads.

(vi) Cadastral Map:

A cadastral map is a map showing the real estate of a country. It includes the location, area, ownership value and tenure of an individual parcel of land.

(vi) Thematic Map:

A thematic map is a map designed to highlight a theme connected with a specific geographic area. Thematic map focuses on a specific subject area and uses base data only as a point of reference for the feature being mapped.

(vii) Choropleth Map:

Choropleth maps are popular thematic maps used to represent statistical data through various shading patterns or symbols on predetermined geographic areas (i.e. countries). They are good at utilizing data to easily represent variability of the desired measurement, across a region.

16.3.a. General vs. Thematic Cartography:

In understanding basic maps the field of cartography can be divided into two general categories; general cartography and thematic cartography. General Cartography involves those maps that are constructed for a general audience and thus contain a variety of features. General maps exhibit many reference and location systems and often are produced in a series.

For example, the 1:24000 scale topographic maps of the United States Geological Survey (USGS) are a standard as compared to the 1:50000 scale Canadian maps.

16.3.b. Topographic vs. Topological:

A topographic map is primarily concerned with the topographic description of a place including the use of centaur lines showing elevation. In the present era one of the most wide spread and advanced methods used to form topographic maps is to use computer software to generate digital elevation models which show shaded relief.

A topological map is a very general type of map. The kind one might sketch on a napkin. It often disregards scale and detail in the interest of clarity of communicating specific route or relation information. Beck's London Underground map is an iconic example.

16.3.c. Counter Line:

Counter line, a line on a map representing an imaginary on the land surface, all points of which are at the same elevation above a datum plane, usually mean sea level.

Contour line are the most common method of showing relief and elevation on a standard topographic map. The elevation represented by contour lines is the vertical distance above or below sea level. The three types of contour lines used on a standard topographic map.

Sub Unit – 4
Data Interpretation

.....

17. Concept

First, let's discuss the word "data" and "interpretation" used in data interpretation.

Data: Data is based on facts and statistics collected together for reference or analysis. Data in numerical format helps us to draw conclusion by comparing the data.

Interpretation: Interpretation is the act of explaining, reframing or otherwise showing your own understanding of something.

Data Interpretation: Data interpretation is an act of analysing data with objective to gain useful information from it. It is done to draw conclusions from the given data. Different statistical tools are used to represent the data in organized structures.

18. Importance of Data Interpretation

Interpretation is essential for the simple reason that the usefulness and utility of research findings lie in proper interpretation. It is being considered a basic component of research process because of the following reasons:

1. It is through interpretation that the researcher can well understand the abstract principle that works beneath his findings. Through this he can link up his findings with those of other studies, having the same abstract principle, and thereby can predict about the concrete world of events. Fresh enquiries can test these predictions later on. This way the continuity in research can be maintained.
2. Interpretation leads to the establishment of explanatory concepts that can serve as a guide for future research studies; it opens new avenues of intellectual adventure and stimulates the quest for more knowledge.
3. Researcher can better appreciate only through interpretation why his findings are what they are and can make others to understand the real significance of his research findings.
4. The interpretation of the findings of exploratory research study often results into hypotheses for experimental research and as such interpretation is involved in the transition from exploratory to experimental research. Since an exploratory study does not have a

hypothesis to start with, the findings of such a study have to be interpreted on a post-factum basis in which case the interpretation is technically described as 'post factum' interpretation.

19. Different ways of Data Interpretation

There are different ways to organize and arrange data are as below:

19.1. Tables with Example:

Tables are the most convenient and versatile method to present data. Analyzing and drawing conclusions from tables is much easier than any other method. Tables are the fundamental method to represent data. In tables data is arranged in rows and columns which help us to scrutinize data efficiently.

Example-01:

Year	**2006**	**2007**	**2008**	**2009**	**2010**
TVs	6000	9000	13000	11000	8000
LCDs	7000	9400	9000	10000	12000

The following table presents the production of electronic items (TVs and LCDs) in a factory during the period from 2006 to 2010. Study the table carefully and answer the questions from 48 to 52:

01.In which year, the total production of electronic items is maximum?

A. 2006 B. 2007

C. 2008 D. 2010

02. What is the difference between averages of production of LCDs and TVS from 2006 to 2008?

A. 3000 B. 2867

C. 3015 **D. None of These**

03. What is the year in which production of TVs is half the production of LCDs in the year 2010?

A. 2007 **B. 2006**

C. 2009 D. 2008

04. What is the ratio of production of LCDs in the years 2008 and 2010?

A. 4:3 **B. 3:4**

C. 2:3 D. 3:2

05. What is the ratio of production of TVs in the years 2006 and 2007?

A. 6:7 B. 7:6

C. 2:3 D. 3:2

Example-02:

The following table shows the percentage profit (%) earned by two companies A and B during the year 2011-15. Answer questions 01-03 based on the data contained in the table.

Year	**Percentage Profit (%)**	
	A	**B**
2011	20	30
2012	35	40
2013	45	35
2014	40	50
2015	25	35

$$\text{Where, present (\%) Profit} = \frac{\text{Income-Expenditure}}{\text{Expenditure}} \times 100$$

01. If the total expenditure of the two companies was '9 lakh in the year 2012 and the expenditure of A and B were in the ratio 2:1, then what was the income of the company A in that year?

(1) Rs. 9.2 lakh **(2) Rs. 8.1 lakh**

(3) Rs. 7.2 lakh (4) Rs. 6.0 lakh

02. What is the average percentage profit earned by the company B?

(1) 35% (2) 42%

(3) 38% (4) 40%

03. In which year, the percentage profit earned by the company B is less than that of company A?

(1) 2012 **(2) 2013**

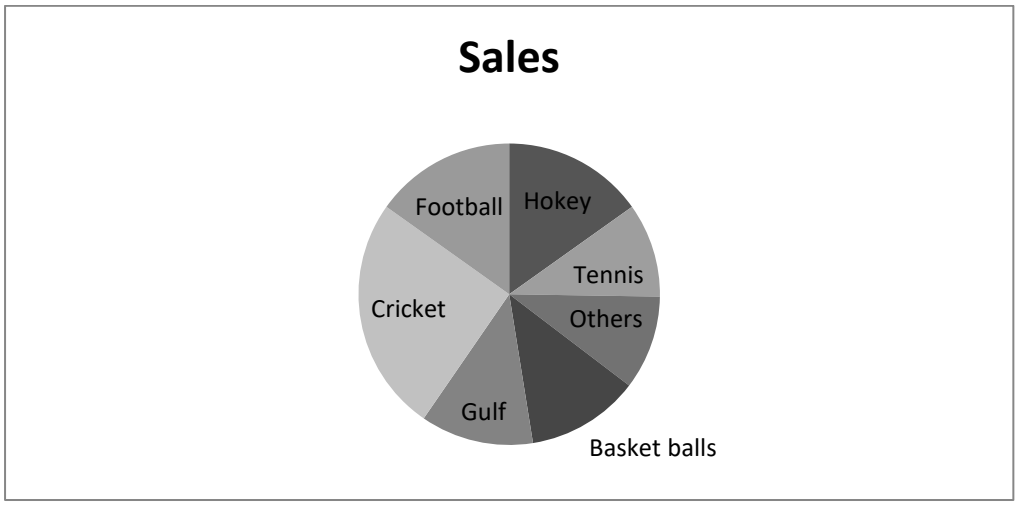
(3) 2014 (4) 2015

19.2. Pie Diagram:

A pie chart is a circular graph that shows the relative contribution that different categories contribute to an overall total. A wedge of the circle represents each category's contribution, such that the graph resembles a pie that has been cut into different sized slices. Every 1% contribution that a category contributes to the total corresponds to a slice with an angle of 3.6 degrees.

Example:

The following pie chart indicates the expenditure of a country on various sports during a particular year. Study the pie chart and answer it Question Number 01-05



01. The ratio of the total expenditure on football to that of expenditure on hockey is:

(A) 1:15 **(B) 1:1**

(C) 15:1 (D) 3:20

02. If the total expenditure on sports during the year was Rs. 1,20.000,00 how much was spent on basketball?

(A) Rs. 9,50,000 (B) Rs. 12,00,000

(C) Rs. 10,00,000 (D) Rs. 15,00,000

03. The chart shows that the most popular game of the country is:

(A) Hockey (B) Football

(C) Cricket (D) Tennis

04. Out of the following country's expenditure is the same on:

(A) Hockey and Tennis **(B) Golf and Basket ball**

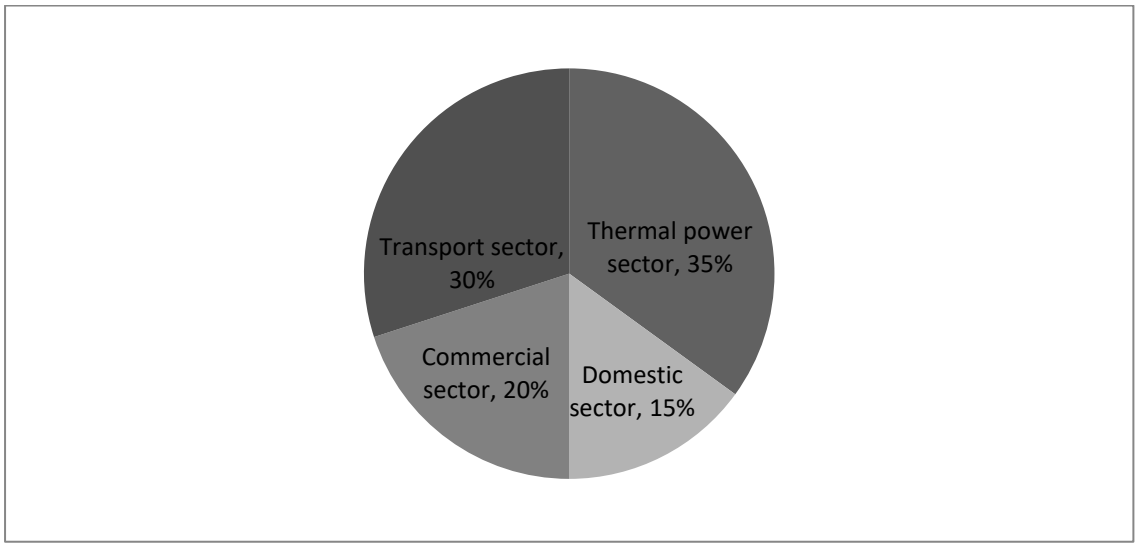
(C) Cricket and Football (D) Hockey and Golf

05. If the total expenditure on sport during the year was Rs. 1,50,00,000 the expenditure on cricket and hockey together was:

(A) Rs. 60,00,000 (B) Rs. 37,50,000

(C) Rs. 50,00,000 (D) Rs. 25,00,000

The total CO_2 emissions from various sectors are 5mmt. In the Pie Chart given below, the percentage contribution to CO_2 emissions from various sectors is indicated.



01. What is the absolute CO_2 emission from domestic sector?

(A) 1.5mmt (B) 2.5mmt

(C) 1.75mmt **(D) 0.75mmt**

02.What is the absolute CO_2 emission for combined thermal power and transport sectors?

(A) 3.25mmt (B) 1.5mmt

(C) 2.5mmt (D) 0.75mmt

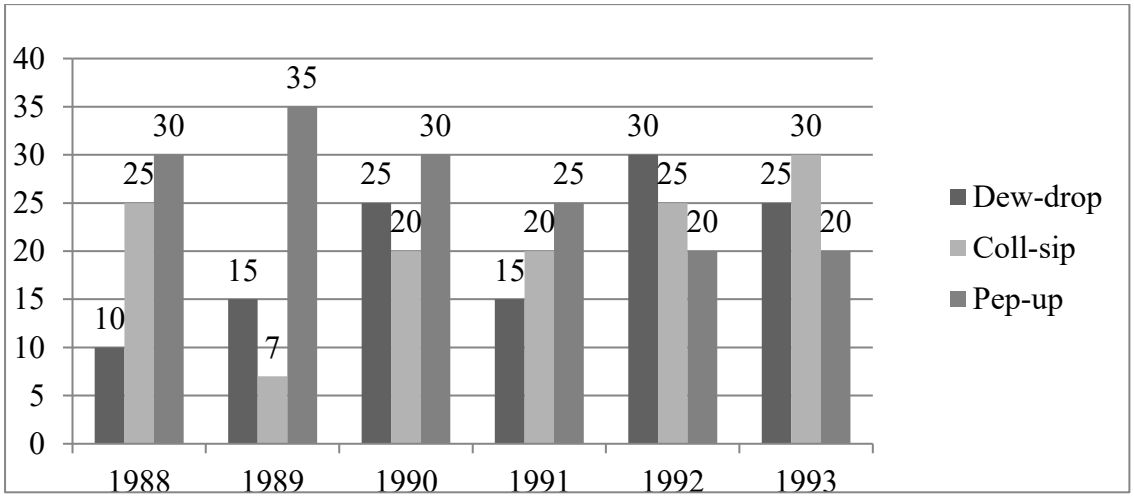
19.3. Bar Diagram:

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called line graph.

A bar graph shows comparisons among discrete categories. One axis of the chart shows the specific categories begin compared, and the other axis represents a measured value. Some bar graph present bars clustered in groups of more than one, showing the values of more than one measure variable.

Example-01:

Study the following graph and answer question Number 01, 02 and 03



01. In which year was the sale of 'Pep-p' the maximum?

(A) 1990 (B) 1992

(C) 1993 (D) None of these

02. In the case of which soft drink was the average annual sale maximum during the period 1988- 1993.

(A) Pep-p only (B) Cool-sip only

(C) Cool-sip and Dew-drop (D) Pep-up and Dew-drop

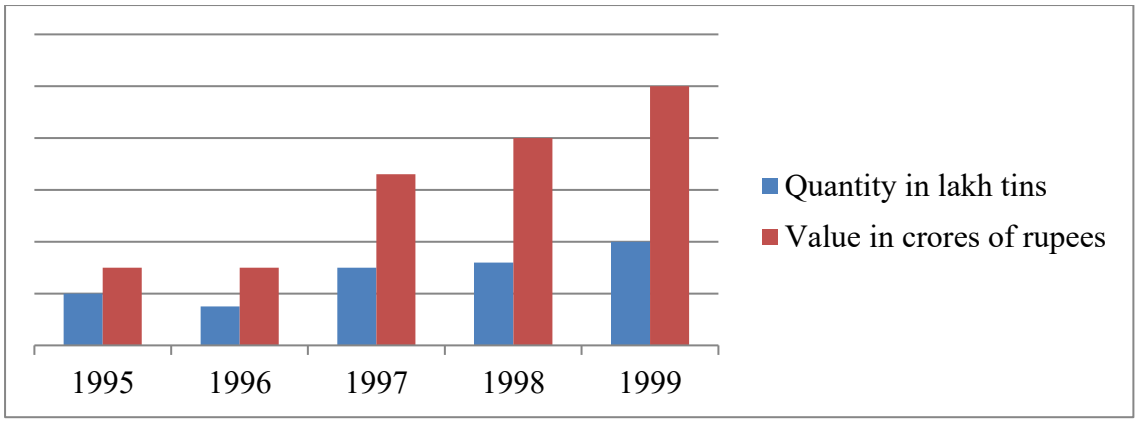
03. What was the approximate percentage of drop in the sale of Pep-up in 1990 over its sale in 1989?

(A) 5 (B) 12

(C) 14 (D) 20

Example-02:

Study the following graph carefully and answer Q.no 01-04 given below it:



01. In which year the value per tin was minimum?

(A) 1995 (B) 1996

(C) 1998 (D) 1999

02. What was the difference between the tins exported in 1997 and 1998?

(A) 10 (B) 1000

(C) 100000 (D) 1000000

03. What was the approximate percentage increase in export value from 1995 to 1999?

(A) 350 (B) 330.3

(C) 433.3 (D) None of these

03. What was the percentage drop in export quantity from 1995 to 1996?

(A) 75 (B) 50

(C) 25 (D) None of these

04. If in 1998, the tins were exported at the same rate per tin as that in 1997, what would be the value (in crores of rupees) of export in 1998?

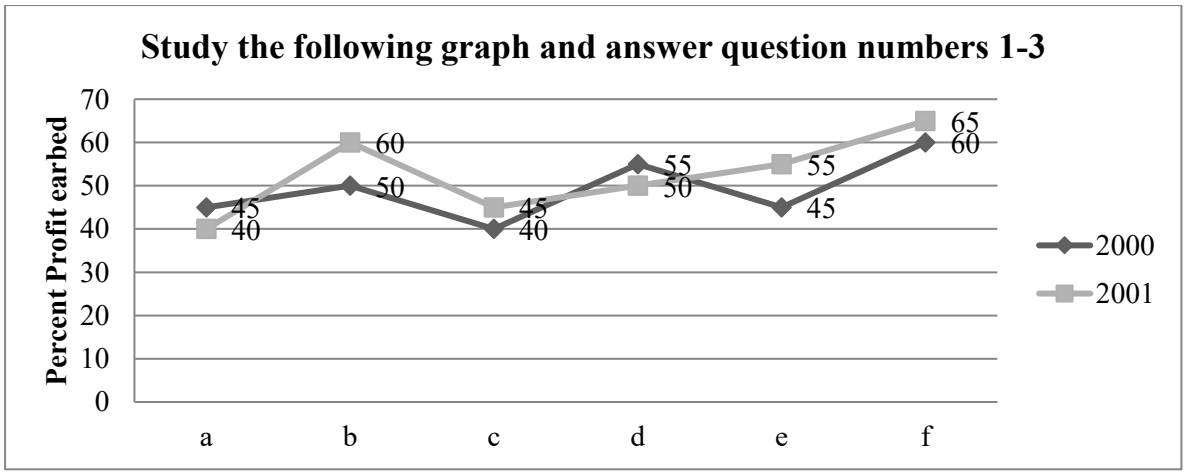
(A) 400 (B) 375

(C) 352 (D) 330

19.4. Line Diagram:

A line chart or line graph is a type of chart which display information as a series of data points called marker corrected by straight line segments. It is a basic type of chart common in many field. It is similar to a scatter plot except that the measurement points are ordered (typically by their x-axis value) and joined with straight line segments. A line chart is often used to visualize a trend in data over intervals of time a time series thus the line is often drawn chronologically. In these cases they are known as run charts.

Example-01:



01. In the year 2000, which of the following companies earned maximum percentage profit?

(A) a (B) b (C) d **(D) f**

02. In the year 2001, which of the following companies earned maximum percentage profit?

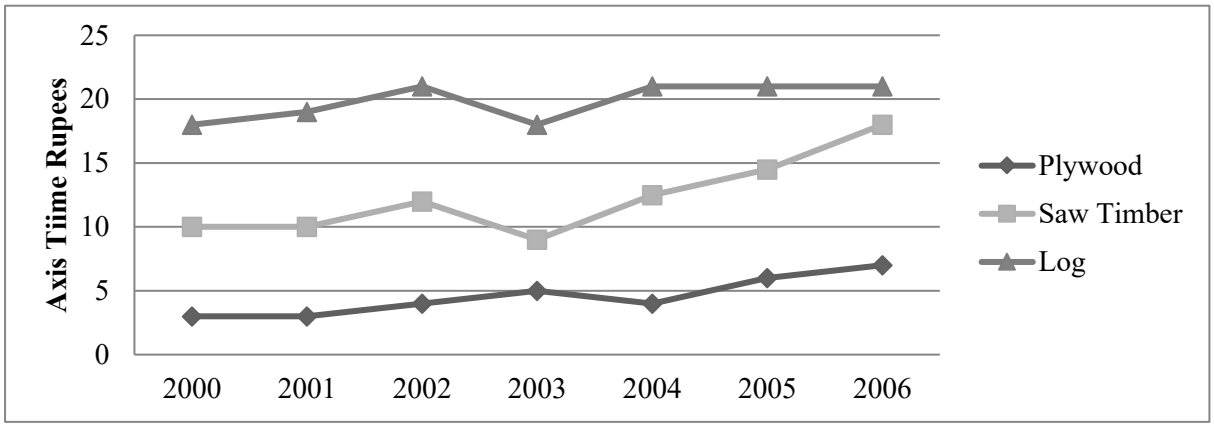
(A) a (B) c (C) d (D) e

03. In the years 2000 and 2001, which of the following companies earned maximum average percentage profit?

(A) f (B) e (C) d (D) b

Example-02:

In the following chart, the price of logs is shown in per cubic metre and that of Plywood and saw Timber in per tonnes. Study the chart and answer the following questions 10, 11 and 12



01. Which product shows the maximum percentage increase in price over the period?

(A) Saw timber (B) Plywood

(C) Log (D) None of the above

02. What is the maximum percentage increase in price per cubic metre of log?

(A) 6 **(B) 12**

(C) 18 (D) None of these

03. In which year the price of two products increased and that of third increased?

(A) 2000 **(B) 2002**

(C) 2003 (D) 2006

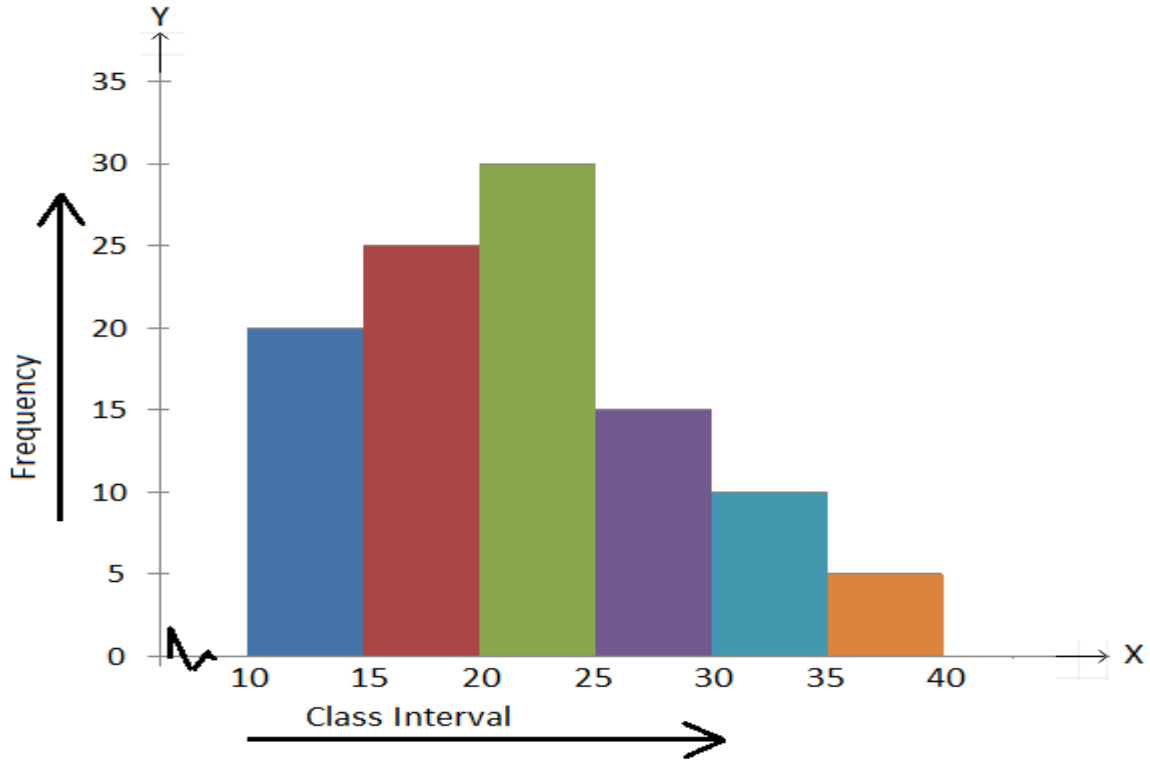
19.5. Histogram:

A histogram is a bar graph of raw data that creates a picture of the data distribution. The bars represent the frequency of occurrence by classes of data. A histogram shows basic information about the data set, such as central location, width of spread, and shape.

Example 01:

We will discuss here some of the problems on histogram.

1. The histogram for a frequency distribution is given below.



Answer the following.

(i) What is the frequency of the class interval 15 – 20?

(ii) What is the class intervals having the greatest frequency?

(iii) What is the cumulative frequency of the class interval 25 – 30?

(iv) Construct a short frequency table of the distribution.

(v) Construct a cumulative frequency table of the distribution.

Solution:

(i) 25

(ii) 20 – 25

(iii) 90

(iv)

Class Interval	**Frequency**
10 - 15	20
15 - 20	25
20 - 25	30

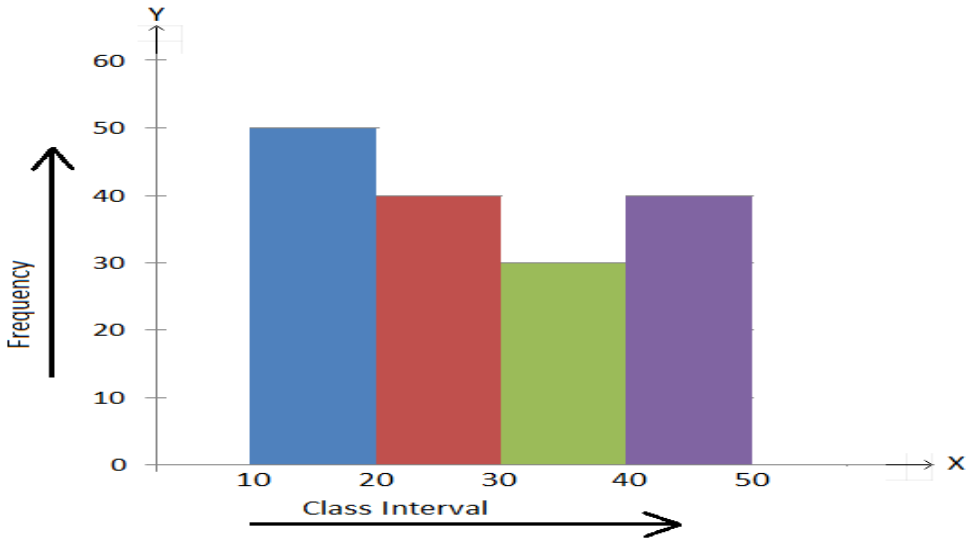
25 - 30	15
30 - 35	10
35 - 40	5

(v)

Class Interval	Frequency	Cumulative Frequency
10 - 15	20	20
15 - 20	25	45
20 - 25	30	75
25 - 30	15	90
30 - 35	10	100
35 - 40	5	105

2. The histogram for a frequency distribution is as shown below.

Example 02: The histogram for a frequency distribution is as shown below**.**



Answer the following.

(i) Find the class intervals having the greatest and the least frequencies.

(ii) Find the class interval whose frequency is 40.

(iii) What is the frequency of the class interval 30 – 40?

(iv) What is the cumulative frequency of the class interval 30 – 40?

(viii) Construct the frequency table of the distribution.

Solution:

(i) The class interval 10 – 20 has the greatest frequency as the rectangle corresponding to the interval has the greatest area, the height being the greatest.

(ii) 20 – 30

(iii) 30

(iv) 50 + 40 + 30 = 120

(v)

Class Interval	Frequency
10- 20	50
20 – 30	40
30 – 40	30
40 – 50	40

20.Various Concepts Required for Solving Questions on Data Interpretation

Average = $\frac{\text{Sum of all the values}}{\text{no.of values}}$

Ratio: Comparison of like terms in its simplest terms

Percentage: X is what percentage of Y = $\frac{X}{Y} * 100$

X is what percentage more (or) less than Y = $\frac{(X-Y)}{Y} * 100$

Percentage Change = $\frac{(\text{Final Value} - \text{Initial Value})}{\text{Initial Value}} * 100$

21. Data Interpretation Tricks to solve Data Interpretation

1) Read the question carefully.

2) Analyse the data carefully.

3) Don't worry about too much data.

4) Skip questions if they are time-consuming.

5) Avoid unnecessary Calculation.

6) Pay close attention to the units used.

7) Learn to appropriate the values.

8) Don't assume anything.

Sub Unit – 5
Data and Governance

.....

22. Concept

Data Governance is a four-way frame work comprising availability, applicability, integrity and security. It is a set of process, used by the stakeholders who use technology to ensure that the important and critical data is managed and protected. It involves a streamlined coordination of individuals (people), methods (process), and innovation (technology) in such an order that is results in realizing the value of data for any organization. It acts as a bridge between business and it for decisions and initiatives.

23. Why Data Governance is needed

Data Governance is required to ensure that an organization's information assets are formally, properly, proactively and efficiently managed throughout the enterprise to secure its trust & accountability. Adopting and implementing Data Governance can result in improved productivity and efficiency of an organization.

24. What are the benefits if data a governance

It keeps data clean. The volume of data you store is immaterial –its relevance and usability are of far greater value than its quantity. The benefit of having everyone involved in the governance process is that everyone has a reason to keep the data in good shape.

25. Role of Data Governance

These individuals are responsible compliance with data standards, successful data production resolves data related issues manage operational data Governance metrics and support enforcement of data standards. Your tactical layer roles, consists of your data governance committee.

Example – 1:

Human Development Report for ‘each’ of the year at global level has been published by

A. UNDP
B. WTO
B. IMF
D. World Bank

Example – 2:

The “Report on Currency and Finance” for each of the financial year in India is Published by:

A. Reserve bank of India
B. Ministry of Finance
C. Planning Commission
D. Central Statistical Organisation