

Descriptive Statistics.

Unit -2

1. What do you mean by statistics?

Statistics is a branch of mathematics that deals with the collection, organization, analysis, interpretation, and presentation of numerical data. It provides tools and methods to study patterns, trends, and relationships in data. Statistics can be broadly classified into two types: descriptive statistics and inferential statistics.

Descriptive statistics summarize and present data in an understandable manner using measures such as mean, median, mode, standard deviation, and graphical representations like histograms and pie charts. Inferential statistics, on the other hand, use probability theory to make predictions or inferences about a population based on a sample.

2. What are the characteristics/features of statistics?

Statistics has several key characteristics that define its role and applications. These features help in understanding data systematically and making informed decisions.

- **Quantitative Nature:** Statistics deals with numerical data rather than qualitative descriptions. It involves counting, measuring, and analyzing figures.
- **Aggregates of Facts:** A single value is not considered statistics; instead, statistics involve a collection of data over multiple observations to identify trends and

patterns.

- **Variability:** Statistical data are subject to variations due to different influencing factors such as time, place, and circumstances.
- **Collected for a Purpose:** Data is gathered with a specific objective, whether for research, analysis, or decision-making. Random data collection without purpose lacks statistical significance.
- **Systematic Collection:** Statistical data must be collected in a structured and scientific manner to ensure accuracy and reliability.
- **Comparability:** One of the essential features of statistics is the ability to compare data sets using common parameters such as percentages, ratios, or averages.
- **Subject to Errors:** Despite structured collection and analysis, statistical conclusions are often estimations and subject to errors like sampling and measurement errors.
- **Used for Decision-Making:** Statistics are widely used in various fields for planning, forecasting, and policy formation. Businesses, governments, and researchers rely on statistical methods to make evidence-based decisions.

3. What are the functions of statistics?

Statistics serves several essential functions across various fields, aiding in data-driven decision-making. Below are some key functions of statistics:

- **Data Collection and Organization:** One of the primary functions of statistics is to collect, classify, and

organize raw data systematically. This helps in structuring data for further analysis.

- **Summarization of Data:** Large volumes of data can be overwhelming. Statistics help in summarizing data using measures such as mean, median, mode, and standard deviation, making interpretation easier.
- **Analysis and Interpretation:** Statistics provide various techniques for analyzing data patterns, relationships, and trends. Methods like regression analysis, hypothesis testing, and probability distributions help in understanding the significance of data.
- **Comparison and Forecasting:** Statistics enable comparison between different sets of data using common metrics. It also allows forecasting future trends based on historical data, which is essential in business, economics, and planning.
- **Decision-Making:** Governments, businesses, and researchers use statistical methods for making informed decisions. For example, businesses use market research statistics to plan product strategies, while policymakers rely on statistical data for economic planning.
- **Quality Control:** In industries, statistical tools help maintain product quality by identifying defects and variations in manufacturing processes. Techniques like Six Sigma are widely used for process improvement.
- **Establishing Relationships:** Statistics helps in identifying correlations and dependencies between different variables, assisting in research and problem-solving.

- **Testing Hypotheses:** In scientific research, statistics provide a foundation for testing hypotheses, ensuring conclusions are drawn based on empirical evidence.

In summary, statistics plays a crucial role in data management, interpretation, and decision-making, making it an essential tool in various disciplines.

4. What are the limitations of statistics?

Despite its importance, statistics has certain limitations that should be considered while interpreting data.

- **Cannot Prove Absolute Truth:** Statistics can only provide estimates and probabilities, not absolute certainty. Conclusions drawn from statistical data are subject to errors.
- **Quality Depends on Data Accuracy:** If the collected data is incorrect or biased, statistical analysis may lead to misleading results. Proper data collection methods are crucial for accuracy.
- **Does Not Account for Individual Cases:** Statistics focus on group data and general trends rather than individual cases. This may overlook unique or rare occurrences.
- **Prone to Misinterpretation:** Statistical results can be manipulated or misrepresented, leading to incorrect conclusions. Proper understanding of statistical methods is necessary to avoid misinterpretation.
- **Affected by Sample Size:** The reliability of statistical conclusions depends on the sample size. A small or unrepresentative sample can lead to inaccurate inferences about a population.
- **Limited to Quantifiable Data:** Statistics cannot analyze

qualitative factors such as emotions, opinions, or personal experiences, which may be crucial in decision-making.

- **Influenced by Assumptions:** Many statistical methods require assumptions about data distribution, independence, or relationships. If these assumptions are incorrect, the results may be unreliable.
- **Cannot Eliminate Uncertainty:** Even with advanced statistical techniques, predictions and analyses carry a degree of uncertainty due to external factors beyond control.

In conclusion, while statistics is a powerful tool for analysis and decision-making, it must be used with caution, ensuring data quality and proper interpretation to avoid misleading conclusions.

5. What do you mean by an average?

An average is a statistical measure that represents the central value of a dataset. It provides a summary of the data by identifying a typical or representative value. Averages help in simplifying complex data and making comparisons easier.

There are three main types of averages:

- **Mean:** The arithmetic mean is the most common type of average. It is calculated by summing all values in a dataset and dividing by the number of observations.
- **Formula:** $\text{Mean} = (\text{Sum of all values}) / (\text{Number of values})$
- **Example:** The mean of 10, 20, and 30 is $(10+20+30)/3 = 20$.
- **Median:** The median is the middle value in an ordered

dataset. If the number of observations is even, the median is the average of the two middle numbers.

- Example: In the dataset {5, 10, 15, 20, 25}, the median is 15.
- **Mode:** The mode is the most frequently occurring value in a dataset. A dataset can have one mode (unimodal), multiple modes (bimodal or multimodal), or no mode if all values occur with the same frequency.
- Example: In {2, 3, 3, 4, 5, 3, 6}, the mode is 3.

Averages are used in various fields, including business (to analyze sales trends), economics (to measure income levels), and education (to determine students' performance). However, each type of average has its limitations and should be chosen based on the nature of the data.

6. What are the characteristics/features of an ideal average?

An ideal average should possess certain characteristics to ensure it accurately represents the dataset. The key features of an ideal average are:

- **Clearly Defined:** An average should have a well-defined formula that can be applied consistently across different datasets.
- **Easy to Compute and Understand:** The method of calculating an average should be simple and easy to interpret, even for non-experts.
- **Based on All Observations:** A good average should consider every value in the dataset to provide a comprehensive summary. The arithmetic mean satisfies this condition, whereas the mode may ignore certain

values.

- **Not Affected by Extreme Values:** An ideal average should not be influenced by outliers. The median is a good choice when dealing with skewed data.
- **Capable of Further Statistical Analysis:** The average should be suitable for mathematical operations and further statistical calculations. The mean is widely used in advanced statistical analysis.
- **Stability:** A good average should not change significantly with small variations in data. The mean is stable for large datasets, but the mode can change with small modifications.
- **Capable of Being Used for Comparison:** A good average should allow meaningful comparisons between different datasets. The mean is often preferred for comparing different groups.
- **Representative of the Dataset:** The chosen average should accurately reflect the central tendency of the data. The mode is useful when dealing with categorical data, while the mean is ideal for numerical data.

Each type of average has its strengths and weaknesses. The arithmetic mean is widely used but can be affected by extreme values, whereas the median is more suitable for skewed data. The mode is useful for categorical data but may not always provide a unique value.

7. What do you mean by arithmetic mean?

Write its merits and demerits.

Answer:

The **Arithmetic Mean (AM)** is the sum of all observations

divided by the total number of observations. It is a measure of central tendency used in statistics.

Merits:

- Simple and easy to calculate.
- Uses all observations, making it a comprehensive measure.
- Suitable for further statistical analysis.
- Less affected by fluctuations in data.

Demerits:

- Affected by extreme values (outliers).
- Cannot be used for qualitative data.
- Not always a representative value in skewed distributions.

8. What is median? Write its merits and demerits.

Answer:

The **Median** is the middle value in an ordered data set. If the number of observations is odd, it is the middle value. If even, it is the average of the two middle values.

Merits:

- Not affected by extreme values.
- Best for skewed distributions.
- Suitable for qualitative data like income or age groups.

Demerits:

- Ignores some data points.
- Difficult to use in large datasets.
- Not suitable for further mathematical operations.

9. What is mode? Write its merits and demerits.

Answer:

The **Mode** is the most frequently occurring value in a dataset.

Formula (Empirical Relation):

$\text{Mode} = 3 \times \text{Median} - 2 \times \text{Mean}$.

Merits:

- Easy to identify in frequency distributions.
- Not affected by extreme values.
- Useful in categorical data.

Demerits:

- May not exist or may be multiple.
- Not useful for further calculations.
- Changes in sample size can impact the mode significantly.

10. What is GM (Geometric Mean)? Write its merits and demerits.

Answer:

The **Geometric Mean (GM)** is the n th root of the product of n observations.

Merits:

- Suitable for ratios and percentages.
- Less affected by extreme values.
- Useful in economic and financial data.

Demerits:

- Cannot be used with negative values.
- Complex calculations.
- Sensitive to small variations.

11. What is HM (Harmonic Mean)? Write its

merits and demerits.

Answer:

The **Harmonic Mean (HM)** is the reciprocal of the arithmetic mean of reciprocals.

Merits:

- Best for rates like speed, efficiency, etc.
- Less affected by large values.

Demerits:

- Cannot be used if any value is zero.
- Complicated calculation.

12. What is range? Write its merits and demerits.

Answer:

The **Range** is the difference between the highest and lowest values in a dataset. $\text{Range} = X_{\text{max}} - X_{\text{min}}$

Merits:

- Easy to calculate.
- Useful for quick comparisons.

Demerits:

- Ignores data distribution.
- Affected by outliers.

13. What is Q.D (Quartile Deviation)? Write its merits and demerits.

Answer:

The **Quartile Deviation (Q.D.)** is half the difference between the upper and lower quartiles.

Merits:

- Not affected by extreme values.
- Useful in skewed distributions.

Demerits:

- Ignores data outside quartiles.
- Not useful for further statistical analysis.

14. What is S.D (Standard Deviation)? Write its merits and demerits.

Answer:

The **Standard Deviation (S.D.)** measures data dispersion around the mean.

Merits:

- Uses all data points.
- Most reliable measure of dispersion.

Demerits:

- Affected by extreme values.
- Difficult to calculate.

15. What is M.D (Mean Deviation)? Write its merits and demerits.

Answer:

The **Mean Deviation (M.D.)** is the average of absolute deviations from the mean or median.

Merits:

- Simple to understand.
- Not greatly affected by extreme values.

Demerits:

- Ignores direction of deviation.
- Cannot be used in higher statistical analysis.

16. What is variance?

Answer:

The **Variance** is the square of the standard deviation and represents data spread.

17. What is coefficient of variation?

Answer:

The **Coefficient of Variation (C.V.)** is the ratio of standard deviation to the mean, expressed as a percentage.

