NAME- D.NEERAJA
SUBJECT- BUSINESS STASTICTICS-1 SUBJECT CODE-302
STREAMS- BCOM ALL STREAMS
E MATERIAL
INTERNALS QUESTION BANK

UNIT-1

"Introduction to Business Statistics" is a foundational course designed to equip students with essential statistical tools and methodologies applicable in various business contexts. This course covers fundamental concepts such as descriptive statistics, probability distributions, hypothesis testing, regression analysis, and data visualization. By understanding these principles, students learn to analyze and interpret data effectively, facilitating better decision-making and strategic planning in business environments. The course emphasizes practical applications, ensuring that students can translate statistical insights into actionable business strategies.

Origin of Stats

Statistics has a rich history that begins with ancient record-keeping and evolves into a critical modern discipline:

Ancient Origins

Early Civilizations: Babylonians, Egyptians, and Chinese kept records for administration and taxes.

Greece and Rome: Conducted censuses for governance and military purposes.

Middle Ages and Renaissance

Medieval Europe: Kept agricultural and trade records.

Renaissance: Systematic data collection grew with trade; double-entry bookkeeping emerged.

17th and 18th Centuries

John Graunt: Analyzed mortality data, founding demography.

Probability Theory: Developed by Pascal and Fermat.

Bayes' Theorem: Established by Thomas Bayes.

19th Century

Adolphe Quetelet: Applied statistics to social sciences.

Francis Galton: Introduced regression and correlation.

Statistical Societies: Formation of groups like the Royal Statistical Society.

20th Century

Ronald A. Fisher: Advanced experimental design and ANOVA.

Disciplinary Expansion: Statistics became essential in economics, psychology, medicine, and engineering.

Computing Revolution: Enabled complex data analysis and advanced software development.

21st Century

Big Data: Emergence of analytics for large datasets.

Data Science: Combines statistics and computer science, emphasizing machine learning and data insights.

From ancient times to the digital age, statistics has grown into a vital tool for understanding and improving the world.

Definition of stastistics

Singular Definition

Business Statistic: A single numerical measure or value derived from data used to describe or summarize a specific aspect of business operations, such as sales figures, profit margins, or customer satisfaction scores.

Plural Definition

Business Statistics: The collection, analysis, interpretation, and presentation of data related to business activities, encompassing a wide range of techniques and methods used to inform decision-making and strategic planning in a business context.

Singular Definition

Business Statistic: A single numerical measure or value derived from data used to describe or summarize a specific aspect of business operations, such as sales figures, profit margins, or customer satisfaction scores.

Plural Definition

Business Statistics: The collection, analysis, interpretation, and presentation of data related to business activities, encompassing a wide range of techniques and methods used to inform decision-making and strategic planning in a business context.

Singular Definition

Business Statistic: A single numerical measure or value derived from data used to describe or summarize a specific aspect of business operations, such as sales figures, profit margins, or customer satisfaction scores.

Plural Definition

Business Statistics: The collection, analysis, interpretation, and presentation of data related to business activities, encompassing a wide range of techniques and methods used to inform decision-making and strategic planning in a business context.

Importance of Statistics

Decision Making: Helps businesses make informed decisions by providing data-driven insights.Trend Analysis: Identifies patterns and trends, aiding in forecasting and strategic planning.

- Quality Control: Ensures product and service quality through statistical quality control methods.
- Market Research: Assists in understanding customer preferences and market dynamics.
- Risk Management: Evaluates and mitigates risks by analyzing historical data and predicting future outcomes.

Scope of Statistics

- Descriptive Statistics: Summarizes and describes data using measures like mean, median, and standard deviation.
- Inferential Statistics: Draws conclusions and makes predictions based on sample data, using techniques like hypothesis testing and confidence intervals.
- Predictive Analytics: Uses statistical models to forecast future trends and behaviors.
- Operational Efficiency: Optimizes business processes by analyzing operational data.
- Financial Analysis: Evaluates financial performance and investment opportunities using statistical methods.

Limitations of Statistics

- Data Quality: Conclusions are only as good as the data collected; poor-quality data leads to unreliable results.
- Misinterpretation: Statistical results can be misinterpreted or manipulated to support biased viewpoints.
- Complexity: Advanced statistical methods can be complex and require specialized knowledge to apply correctly.
- Causality vs. Correlation: Statistics often show correlations but do not establish causality.
- Context Dependence: Statistical findings need to be interpreted within the context; ignoring this can lead to incorrect conclusions.
- Overreliance: Excessive dependence on statistical analysis can overlook qualitative factors and insights.

Distrust of Statistics

- Manipulation: Statistics can be manipulated or presented selectively to mislead or support biased agendas.
- Complexity: Complex methods and jargon can make statistics difficult to understand, leading to skepticism.
- Data Quality: Poor-quality or biased data can produce misleading results.
- Misuse: Incorrect application or interpretation of statistical techniques can lead to false conclusions.
- Causality Confusion: Misinterpreting correlation as causation can result in incorrect assumptions.
- Lack of Transparency: Without clear methodology and transparency, statistical results can be viewed with suspicion.

Planning a Statistical Investigation

- Define Objectives: Clearly outline the goals and purpose of the investigation.
- Identify Variables: Determine the key variables to be studied and measured.
- Design Methodology: Choose appropriate statistical methods and sampling techniques.
- Collect Data: Gather data using surveys, experiments, or existing datasets, ensuring accuracy and reliability.
- Analyze Data: Use statistical tools to process and interpret the data.
- Interpret Results: Draw conclusions based on the analysis, considering the context and limitations.
- Report Findings: Present the results in a clear, concise, and transparent manner, including any assumptions and potential biases.

Sampling Method

- Definition: A method where data is collected from a subset (sample) of the population.
- Types:
- Simple Random Sampling: Every member has an equal chance of being selected.
- Stratified Sampling: Population is divided into strata, and random samples are taken from each stratum.

- Systematic Sampling: Every nth member is selected from a list.
- Cluster Sampling: Population is divided into clusters, and entire clusters are randomly selected.
- Advantages: Faster, less expensive, and practical for large populations.
- Disadvantages: May not be as accurate as a census; risk of sampling bias

Primary Data Collection

- Definition: Data collected firsthand for a specific research purpose.
- Methods:
- Surveys and Questionnaires: Gathering responses directly from participants.
- Interviews: Conducting one-on-one or group discussions to collect detailed information.
- Observations: Recording behaviors or events as they occur.
- Experiments: Conducting controlled tests to study cause-effect relationships.
- Advantages: Highly relevant and specific to the research question; up-to-date.
- Disadvantages: Time-consuming and costly to collect.

Secondary Data Collection

- Definition: Data that has already been collected and published by others.
- Sources:
- Government Reports: Census data, economic reports, etc.
- Academic Studies: Research papers, articles, and dissertations.
- Industry Reports: Market analysis, company reports, and whitepapers.
- Online Databases: Public databases, statistical agencies, and library archives.
- Advantages: Cost-effective and time-saving; readily available.
- Disadvantages: May not be specific to the research needs; potential issues with data relevance and accuracy.

Top of Form

Statistical Errors

- Sampling Error: Differences between a sample and the population it represents due to random variation.
- Measurement Error: Inaccuracies in data collection, recording, or analysis.
- Selection Bias: Systematic differences between the sample and the population it represents due to non-random selection.
- Non-Response Bias: Occurs when individuals chosen for the sample do not respond, leading to a biased sample.
- Confounding Variables: Factors that influence both the independent and dependent variables, leading to incorrect conclusions.
- Type I Error: Incorrectly rejecting a true null hypothesis (false positive).
- Type II Error: Failing to reject a false n

Statistical Appropriation

- Data Manipulation: Selective presentation or manipulation of data to support a particular viewpoint or agenda.
- Cherry-Picking: Highlighting only data that supports a predetermined conclusion while ignoring contradictory evidence.
- Misrepresentation: Presenting statistical results in a misleading or deceptive manner.
- False Correlation: Incorrectly inferring causation based on correlation without adequate evidence.
- Statistical Gimmickry: Using statistical techniques inappropriately or without understanding their limitations.
- Overinterpretation: Drawing conclusions beyond the scope or validity of the data.

Classification of Data

1. Definition: Organizing data into categories or groups based on shared characteristics.
2. Methods:

1. Nominal: Categorizing data into distinct groups with no inherent order (e.g., colors, gender).
2. Ordinal: Ordering data into categories based on a specific criterion (e.g., ranking, satisfaction levels).
3. Interval: Dividing data into equal intervals with meaningful numerical values but no true zero point (e.g., temperature in Celsius).
4. Ratio: Similar to interval data but with a true zero point (e.g., height, weight).

3. Purpose: Simplifies data analysis and facilitates comparison between groups.

Tabulation of Data

- Definition: Presenting data in a structured format using tables.
- Components:
- Variables: Columns representing different characteristics or measurements.
- Categories: Rows representing different groups or classes.
- Frequency: Count or percentage of observations falling into each category.
- Types:
- Simple Tabulation: Basic presentation of data in tabular form.
- Complex Tabulation: Includes cross-tabulations, where data is simultaneously classified by two or more variables.
- Purpose: Provides a clear and concise summary of data, aiding in interpretation and analysis.

Frequency Distribution

- Definition: A summary of the frequency of occurrence of each value or range of values in a dataset.
- Components:
- Classes or Intervals: Ranges into which data is grouped.
- Frequency: The number of observations falling within each class.
- Cumulative Frequency: The running total of frequencies as you move through the classes.

Frequency Distribution

- Definition: A summary of the frequency of occurrence of each value or range of values in a dataset.
- Components:
- Classes or Intervals: Ranges into which data is grouped.
- Frequency: The number of observations falling within each class.
- Cumulative Frequency: The running total of frequencies as you move through the classes.

Types of Frequency Distribution

- Simple Frequency Distribution: Shows the frequency of each distinct value in a dataset.

- Grouped Frequency Distribution: Groups data into intervals or classes and displays the frequency of observations within each interval.
- Cumulative Frequency Distribution: Presents the running total of frequencies as you progress through the data, showing the cumulative frequency up to each class or interval.
- Relative Frequency Distribution: Displays the proportion or percentage of observations in each class relative to the total number of observations.
- Cumulative Relative Frequency Distribution: Similar to cumulative frequency distribution but expresses the cumulative frequency as a proportion or percentage of the total number of observations.

UNIT-2

DIAGRAMATIC PRESENTATION

A one-dimensional diagram typically represents data along a single axis. The most common example of a one-dimensional diagram is a number line. It shows numerical values placed at equal intervals along a straight line, with each value corresponding to a point on the line. It's a simple yet effective way to visualize one-dimensional data, such as measurements or quantities, in their numerical order.

A two-dimensional diagram represents data in a two-dimensional space, typically using two axes. Here are some common types:

- Scatter Plot: Displays individual data points on a Cartesian plane, with one variable represented on each axis. Useful for visualizing relationships between two variables.
- Line Graph: Connects data points with straight lines to show trends or changes over time. Often used with time-series data.
- Bar Graph: Uses rectangular bars to represent data categories on one axis and their corresponding values on the other axis. Ideal for comparing quantities between different categories.
- Histogram: Displays the distribution of continuous data by dividing it into intervals (bins) along one axis and showing the frequency or density of observations within each interval using bars.
- Heatmap: Represents data values in a grid format, with color intensity indicating the magnitude of each value. Useful for visualizing large datasets and identifying patterns or trends.

A three-dimensional diagram represents data in a three-dimensional space, adding depth to the visualization. Here are some common types:

- 3D Scatter Plot: Similar to a 2D scatter plot but with data points plotted in a three-dimensional coordinate system. It allows for the visualization of relationships between three variables.
- 3D Bar Chart: Extends the concept of a traditional bar chart into three dimensions, with rectangular bars representing data categories along three axes. Useful for comparing quantities across multiple categories in three-dimensional space.
- 3D Surface Plot: Displays a three-dimensional surface representing a function of two variables. It visualizes how the value of the function changes across the x and y axes.
- 3D Histogram: Represents the distribution of data in three dimensions, with bins or intervals along three axes showing the frequency or density of observations within each region of the space.
- 3D Pie Chart: Similar to a traditional pie chart but with a three-dimensional appearance, dividing a whole into sectors to represent the proportion of each category. However, 3D pie charts can sometimes distort proportions and make it harder to accurately compare categories.

Graphs

- Definition: Graphs, also known as charts, display data visually to show relationships, trends, or comparisons.
- Types:
- Line Graph: Connects data points with lines to illustrate trends or changes over time.
- Bar Chart: Uses rectangular bars to represent categories of data, often used for comparing discrete categories.
- Pie Chart: Divides a circle into sectors to represent proportions or percentages of a whole.
- Uses:
- Comparing data across different categories or groups.
- Showing trends or changes over time.
- Presenting proportions or percentages of a whole.

Histograms

- Definition: Histograms are graphical representations of the distribution of numerical data, divided into intervals called bins.
- Characteristics:
- Bars represent the frequency or density of observations within each bin.
- Bars are typically contiguous, with no gaps between them.

- The width of each bar represents the interval of values it covers.
- Uses:
- Visualizing the distribution of continuous data.
- Identifying patterns or trends in data distribution.
- Assessing the shape, center, and spread of the data.

Differences

- Data Type: Graphs are versatile and can represent various types of data, including categorical and numerical. Histograms specifically represent numerical data.
- Representation: Graphs display relationships or comparisons, while histograms focus on the distribution of data values.
- Visualization: Graphs emphasize trends or patterns, while histograms emphasize the shape and spread of the data distribution.

Top of Form

UNIT-3 MEASURES OF CENTRAL TENDENCY

Methods of averages

The arithmetic mean, often referred to simply as the mean, is a measure of central tendency calculated by summing up all values in a dataset and dividing by the total number of values. Here's how you can calculate it:

Arithmetic Mean Formula:

Mean=$\sum i$ =1n x i n Mean=n∑i=1nxi

Where:

- x i xi represents each individual value in the dataset.
- n n represents the total number of values in the dataset.

Example Problem:

Let's say we have the following dataset representing the daily temperatures (in degrees Celsius) recorded over a week:

20,22,24,21,23,25,2020,22,24,21,23,25,20

To find the mean temperature for the week, we'll sum up all the temperatures and divide by the total number of days:

Mean=20+22+24+21+23+25+207Mean=720+22+24+21+23+25+20

Mean=1557Mean=7155

Mean≈22.14Mean≈22.14

So, the arithmetic mean temperature for the week is approximately 22.1422.14 degrees Celsius.

The arithmetic mean, often referred to simply as the mean, is a measure of central tendency calculated by summing up all values in a dataset and dividing by the total number of values. Here's how you can calculate it:

Arithmetic Mean Formula:

Mean=$\sum i$ =1n x i n Mean=n∑i=1nxi

Where:

- x i xi represents each individual value in the dataset.
- n n represents the total number of values in the dataset.

Example Problem:

Let's say we have the following dataset representing the daily temperatures (in degrees Celsius) recorded over a week:

20,22,24,21,23,25,2020,22,24,21,23,25,20

To find the mean temperature for the week, we'll sum up all the temperatures and divide by the total number of days:

Mean=20+22+24+21+23+25+207Mean=720+22+24+21+23+25+20

Mean=1557Mean=7155

Mean≈22.14Mean≈22.14

So, the arithmetic mean temperature for the week is approximately 22.1422.14 degrees Celsius.

The harmonic mean is another measure of central tendency, particularly useful when dealing with rates or ratios. It is calculated as the reciprocal of the arithmetic mean of the reciprocals of the values in the dataset. Here's how you can calculate it:

Harmonic Mean Formula:

Harmonic Mean=n 1x 1+1x 2+...+1x n Harmonic Mean=x11+x21+...+xn1n

Where:

- x 1,x 2,...,x n x1,x2,...,xn are the individual values in the dataset.
- n n is the total number of values in the dataset.

Example Problem:

Let's say we have a dataset representing the speeds of a car during three different time intervals: 40 mph, 60 mph, and 80 mph. To find the harmonic mean speed over these three intervals, we'll use the formula:

Harmonic Mean=3140+160+180Harmonic Mean=401+601+8013

Harmonic Mean=3140+160+180Harmonic Mean=401+601+8013

Harmonic Mean=30.025+0.0167+0.0125Harmonic Mean=0.025+0.0167+0.01253

Harmonic Mean=30.0542Harmonic Mean=0.05423

Harmonic Mean≈30.0542Harmonic Mean≈0.05423

Harmonic Mean≈55.28Harmonic Mean≈55.28

So, the harmonic mean speed over the three intervals is approximately 55.28 mph.

3.5

Quartiles are statistical measures that divide a dataset into four equal parts. They provide insight into the spread and distribution of the data. There are three quartiles: the first quartile (Q1), the second quartile (Q2), and the third quartile (Q3). Here's how they are calculated:

Calculation of Quartiles:

- First Quartile (Q1): Represents the value below which 25% of the data fall.
- Second Quartile (Q2): Represents the median of the dataset, where 50% of the data fall below and 50% above.
- Third Quartile (Q3): Represents the value below which 75% of the data fall.

Steps to Calculate Quartiles:

- Order the Data: Arrange the dataset in ascending order.
- Identify the Median (Q2): If the dataset has an odd number of values, the median is the middle value. If it has an even number of values, the median is the average of the two middle values.
- Calculate Q1 and Q3:
- For Q1: Find the median of the lower half of the dataset (excluding Q2).
- For Q3: Find the median of the upper half of the dataset (excluding Q2).

Example:

Let's say we have the following dataset representing test scores: 70, 75, 80, 85, 90, 95, 100.

- Order the Data: 70, 75, 80, 85, 90, 95, 100.
- Identify the Median (Q2): 85 (the middle value).

- Calculate Q1 and Q3:
- For Q1: The lower half is 70, 75, and 80. Q1 = 75.
- For Q3: The upper half is 90, 95, and 100. Q3 = 95.

So, the quartiles for this dataset are Q1 = 75, Q2 = 85, and Q3 = 95.

Uses of Averages

- Simplification: Averages provide a single representative value that summarizes a dataset, making it easier to interpret and compare.
- Analysis: Averages help identify trends, patterns, and central tendencies in data, facilitating decision-making and problem-solving.
- Communication: Averages serve as effective communication tools to convey information concisely and clearly to stakeholders.

Limitations of Averages

- Sensitivity to Outliers: Averages can be heavily influenced by extreme values, leading to misrepresentation of the central tendency.
- Skewed Distributions: In skewed distributions, averages may not accurately reflect typical values, as they are pulled towards the skew.
- Non-Representative: In heterogeneous datasets, averages may not represent individual characteristics accurately, leading to oversimplification.
- Loss of Detail: Averages may obscure variability and nuances within the data, potentially overlooking important information.
- Context Dependency: Averages need to be interpreted within the context of the data and the specific analysis, as they may not always provide meaningful insights on their own.

UNIT-4

Measures of dispersion quantify the spread or variability of data points in a dataset. They provide insights into the extent to which individual data points deviate from the central tendency. Here are some common measures of dispersion:

Range

- Definition: The difference between the maximum and minimum values in a dataset.
- Formula: Range=Max−MinRange=Max−Min
- Characteristics: Simple to calculate but sensitive to outliers.

Interquartile Range (IQR)

- Definition: The range of the middle 50% of data, calculated as the difference between the third quartile (Q3) and the first quartile (Q1).

- Formula: IQR=Q 3−Q 1IQR=Q3−Q1
- Characteristics: Less sensitive to outliers than the range.

Variance

1. Definition: The average of the squared differences between each data point and the mean.

2. Formula: Population Variance(σ 2)=$\sum i$ =1n (x i −μ)2n Population Variance(σ2)=n∑i=1n(xi−μ)2 Sample Variance(s 2)=$\sum i$ =1n (x i −x ¯)2n −1Sample Variance(s2)=n−1∑i=1n(xi−x¯)2

3. Characteristics: Provides a measure of the average deviation from the mean, but values are squared, so it's not in the same units as the original data.

Standard Deviation

1. Definition: The square root of the variance.

2. Formula: Population Standard Deviation(σ)=$\sum i$ =1n (x i −μ)2n Population Standard Deviation(σ)=n∑i=1n(xi−μ)2 Sample Standard Deviation(s)=$\sum i$ =1n (x i −x ¯)2n −1Sample Standard Deviation(s)=n−1∑i=1n(xi−x¯)2

3. Characteristics: Provides a measure of the average deviation from the mean in the same units as the original data.

These measures help quantify the dispersion of data points, providing valuable insights into the variability and distribution of the dataset.

Top of Form

Karl Pearson's method of skewness, also known as Pearson's first coefficient of skewness, is a measure of the asymmetry of a probability distribution. It's calculated using the third moment about the mean, divided by the cube of the standard deviation. Here's the formula:

Pearson’s Coefficient of Skewness=3(Mean−Median)Standard DeviationPearson’s Coefficient of Skewness=Standard Deviation3(Mean−Median)

Where:

1. Mean: The arithmetic mean of the dataset.

2. Median: The middle value of the dataset when arranged in ascending order.

3. Standard Deviation: The standard deviation of the dataset.

This measure indicates whether the dataset is skewed to the left or right. If the coefficient is positive, it suggests a right skew (longer right tail), and if negative, a left skew (longer left tail).

BOWLEYS METHOD OF SKEWNESS.

Kelly's method of skewness, developed by John Kelly, is a statistical measure used to assess the skewness of a dataset. It's based on comparing the difference between the mode and the median to the difference between the median and the mean. The formula for Kelly's method of skewness is as follows:

Kelly’s Skewness=3(Median−Mode)Mean−MedianKelly’s Skewness=Mean−Median3(Median−Mode)

Where:

- Mean: The arithmetic mean of the dataset.
- Median: The middle value of the dataset when arranged in ascending order.
- Mode: The most frequently occurring value in the dataset.

Kelly's skewness helps determine the direction and extent of asymmetry in the distribution of the dataset. If the value is positive, it indicates a right skew (longer right tail), while a negative value suggests a left skew (longer left tail). A value close to zero suggests a symmetric distribution.

TYPES OF KURTOSIS

TYPES OF KURTOSIS

Bowley's method of skewness, also known as the quartile coefficient of skewness, is a measure of skewness that assesses the asymmetry of a dataset using quartiles. It compares the distance between the median and the lower quartile (Q1) to the distance between the upper quartile (Q3) and the median. The formula for Bowley's skewness is:

Bowley’s Skewness=Q 3−2×Median+Q 1Q 3−Q 1Bowley’s Skewness=Q3−Q1Q3−2×Median+Q1

Where:

- Q 1Q1 is the first quartile (25th percentile),
- Q 3Q3 is the third quartile (75th percentile), and
- Median is the median of the dataset.

Bowley's skewness is a dimensionless measure. If the value is negative, it suggests a left skew (longer left tail), while a positive value indicates a right skew (longer right tail). A value close to zero suggests a symmetric distribution.

TYPES OF KURTOSIS.

Mesokurtic distributions have kurtosis equal to zero, implying that their tails are similar to those of a normal distribution. In other words, they exhibit a moderate level of peakness or flatness in comparison to a normal distribution. Mesokurtic distributions are often associated with data that is neither overly peaked nor excessively flat, resembling the shape of a bell

curve. They represent a balanced distribution with tails that are neither too heavy nor too light.

A leptokurtic distribution is characterized by positive kurtosis, indicating that its tails are heavier than those of a normal distribution. In other words, it has a higher peak and thicker tails compared to a normal distribution. Leptokurtic distributions often arise when the data has more extreme values or outliers, resulting in a higher concentration of data around the mean and more frequent occurrences of extreme values in the tails. This type of distribution is sometimes referred to as "heavy-tailed" or "fat-tailed."

A platykurtic distribution is characterized by negative kurtosis, implying that its tails are lighter than those of a normal distribution. In simpler terms, it has a flatter peak and thinner tails compared to a normal distribution. Platykurtic distributions often occur when the data is more widely dispersed and lacks extreme values or outliers. In such distributions, the majority of data points are closer to the mean, resulting in a broader, shallower distribution.

UNIT- 5

CORRELATION

Correlation is a statistical measure that describes the strength and direction of a relationship between two variables. There are several types of correlation, each indicating a different kind of relationship between the variables:

Positive Correlation

- Definition: A positive correlation exists when an increase in one variable is associated with an increase in the other variable, and a decrease in one variable is associated with a decrease in the other variable.
- Graphical Representation: In a scatter plot, the data points tend to form an upward sloping pattern from left to right.

Negative Correlation

- Definition: A negative correlation occurs when an increase in one variable is associated with a decrease in the other variable, and vice versa.
- Graphical Representation: In a scatter plot, the data points tend to form a downward sloping pattern from left to right.

No Correlation (Zero Correlation)

- Definition: No correlation exists when there is no discernible pattern between the variables. Changes in one variable are not associated with changes in the other variable.
- Graphical Representation: In a scatter plot, the data points are scattered randomly without forming any particular pattern.

Perfect Correlation

- Definition: A perfect correlation occurs when all data points lie exactly on a straight line.
- Graphical Representation: In a scatter plot, all data points fall precisely along a straight line, either upward or downward.

Strong Correlation

- Definition: A strong correlation indicates a close relationship between the variables, where changes in one variable are consistently associated with changes in the other variable.
- Graphical Representation: In a scatter plot, the data points form a tight cluster around a line, either upward or downward.

Weak Correlation

- Definition: A weak correlation suggests a loose relationship between the variables, where changes in one variable are not consistently associated with changes in the other variable.
- Graphical Representation: In a scatter plot, the data points are more spread out and do not form a tight cluster around a line.

Understanding the type of correlation between variables is essential for interpreting relationships in data and making informed decisions.

Top of Form

Correlation is a statistical measure that describes the strength and direction of a relationship between two variables. There are several types of correlation, each indicating a different kind of relationship between the variables:

Positive Correlation

- Definition: A positive correlation exists when an increase in one variable is associated with an increase in the other variable, and a decrease in one variable is associated with a decrease in the other variable.
- Graphical Representation: In a scatter plot, the data points tend to form an upward sloping pattern from left to right.

Negative Correlation

- Definition: A negative correlation occurs when an increase in one variable is associated with a decrease in the other variable, and vice versa.
- Graphical Representation: In a scatter plot, the data points tend to form a downward sloping pattern from left to right

Correlation is a statistical measure that describes the strength and direction of a relationship between two variables. There are several types of correlation, each indicating a different kind of relationship between the variables:

Positive Correlation

- Definition: A positive correlation exists when an increase in one variable is associated with an increase in the other variable, and a decrease in one variable is associated with a decrease in the other variable.
- Graphical Representation: In a scatter plot, the data points tend to form an upward sloping pattern from left to right.

Negative Correlation

- Definition: A negative correlation occurs when an increase in one variable is associated with a decrease in the other variable, and vice versa.
- Graphical Representation: In a scatter plot, the data points tend to form a downward sloping pattern from left to right.

No Correlation (Zero Correlation)

- Definition: No correlation exists when there is no discernible pattern between the variables. Changes in one variable are not associated with changes in the other variable.
- Graphical Representation: In a scatter plot, the data points are scattered randomly without forming any particular pattern.

Perfect Correlation

- Definition: A perfect correlation occurs when all data points lie exactly on a straight line.
- Graphical Representation: In a scatter plot, all data points fall precisely along a straight line, either upward or downward.

Strong Correlation

- Definition: A strong correlation indicates a close relationship between the variables, where changes in one variable are consistently associated with changes in the other variable.
- Graphical Representation: In a scatter plot, the data points form a tight cluster around a line, either upward or downward.

Weak Correlation

- Definition: A weak correlation suggests a loose relationship between the variables, where changes in one variable are not consistently associated with changes in the other variable.

- Graphical Representation: In a scatter plot, the data points are more spread out and do not form a tight cluster around a line.

Understanding the type of correlation between variables is essential for interpreting relationships in data and making informed decisions

If ranks are not given, but only the raw data is provided, you need to first rank the data before calculating Spearman's rank correlation coefficient. Here's how you can proceed:

1. Rank the data in each set separately. Assign ranks based on the order of values, with the smallest value receiving rank 1, the next smallest receiving rank 2, and so on. In case of ties (i.e., identical values), assign the average rank to tied values.
2. Calculate Spearman's rank correlation coefficient

If ranks are already given for the data sets, you can directly use Spearman's rank correlation coefficient formula to calculate the correlation. Here's how you can proceed:

1. Calculate the differences between ranks for each pair of corresponding data points.
2. Square each difference.
3. Sum up the squared differences.
4. Use the formula to calculate rsr_s:

$$rs=1-6\times\sum(\text{Squared Differences})n(n2-1)r_s$$
$$=1-n(n^2-1)6\times\sum(\text{Squared Differences})$$

Where:

- rsr_s is Spearman's rank correlation coefficient.
- nn is the number of data points (or ranks) in each set.

Let's illustrate this with an example:

Example:

Suppose we have two sets of ranked data, X and Y:

Set X (Rank):2,4,6,8,10Set Y (Rank):3,6,9,12,15Set X (Rank)Set Y (Rank):2,4,6,8,10:3,6,9,12,15

Solution:

1. Calculate the differences between ranks for each pair of corresponding data points.

Differences:1,2,3,4,5Differences:1,2,3,4,5

2. Square each difference.

Squared Differences:1,4,9,16,25Squared Differences:1,4,9,16,25

3. Sum up the squared differences.

∑(Squared Differences)=1+4+9+16+25=55∑(Squared Differences)=1+4+9+16+25=55

Use the formula to calculate rsr_s=1-6
If ranks are already given for the data sets, you can directly use Spearman's rank correlation coefficient formula to calculate the correlation. Here's how you can proceed:

1. Calculate the differences between ranks for each pair of corresponding data points.
2. Square each difference.
3. Sum up the squared differences.
4. Use the formula to calculate rsr_s:

$rs=1-6\times\sum(\text{Squared Differences})n(n2-1)r_s$
$=1-n(n2-1)6\times\sum(\text{Squared Differences})$

Where:

- rsr_s is Spearman's rank correlation coefficient.
- nn is the number of data points (or ranks) in each set.

Let's illustrate this with an example:

Example:

Suppose we have two sets of ranked data, X and Y:

Set X (Rank):2,4,6,8,10Set Y (Rank):3,6,9,12,15Set X (Rank)Set Y (Rank):2,4,6,8,10:3,6,9,12,15

Solution:

1. Calculate the differences between ranks for each pair of corresponding data points.

Differences:1,2,3,4,5Differences:1,2,3,4,5

2. Square each difference.

Squared Differences:1,4,9,16,25Squared Differences:1,4,9,16,25

3. Sum up the squared differences.

∑(Squared Differences)=1+4+9+16+25=55∑(Squared Differences)=1+4+9+16+25=55

Use the formula to calculate $rsrs$ $rs=1-n(n2-1)6\times\sum(\text{Squared Differences})$ When ranks are equal, it means there are ties in the data. In such cases, ties need to be handled appropriately when calculating Spearman's rank correlation coefficient. Here's how you can proceed:

1. Assign ranks to the tied values based on their average rank.
2. Proceed with the calculation of Spearman's rank correlation coefficient as usual.

Let's illustrate this with an example:

4.

$rs=1-6\times555(52-1)rs=1-5(52-1)6\times55$ $rs=1-3305(25-1)rs=1-5(25-1)330$ $rs=1-3305(24)rs=1-5(24)330$ $rs=1-330120rs=1-120330$ $rs=1-114rs=1-411$ $rs=1-2.75rs=1-2.75$ $rs=-1.75rs=-1.75$

So, the Spearman's rank correlation coefficient $rsrs$ for the given data

When ranks are equal, it means there are ties in the data. In such cases, ties need to be handled appropriately when calculating Spearman's rank correlation coefficient. Here's how you can proceed:

1. Assign ranks to the tied values based on their average rank.

2. Proceed with the calculation of Spearman's rank correlation coefficient as usual.

Let's illustrate this with an example:

When ranks are equal, it means there are ties in the data. In such cases, ties need to be handled appropriately when calculating Spearman's rank correlation coefficient. Here's how you can proceed:

1. Assign ranks to the tied values based on their average rank.
2. Proceed with the calculation of Spearman's rank correlation coefficient as usual.

Let's illustrate this with an example:

Example:

Suppose we have two sets of ranked data, X and Y:

Set X (Rank):10,20,20,40,50Set Y (Rank):15,25,25,45,55Set X (Rank)Set Y (Rank):10,20,20,40,50:15,25,25,45,55

Soluti

1. Assign ranks to the tied values based on their average rank:

Set X (Rank):1,2.5,2.5,4,5Set Y (Rank):1,2.5,2.5,4,5Set X (Rank)Set Y (Rank):1,2.5,2.5,4,5:1,2.5,2.5,4,5

2. Calculate the differences between ranks for each pair of corresponding data points.

Differences:0,0,0,0,0Differences:0,0,0,0,0

3. Square each difference.

Squared Differences:0,0,0,0,0Squared Differences:0,0,0,0,0

4. Sum up the squared differences.

∑(Squared Differences)=0+0+0+0+0=0∑(Squared Differences)=0+0+0+0+0=0

5. Use the formula to calculate rs_r:

$rs=1-6\times05(52-1)rs=1-5(52-1)6\times0\ rs=1-05(25-1)rs=1-5(25-1)0\ rs=1-0120rs=1-1200\ rs=1rs=1$

So, the Spearman's rank correlation coefficient $rsrs$ for the given data sets is 1, indicating a perfect positive correlation.

Internal Questions

MULTIPLE CHOICE QUESTIONS

1. The word 'Statistics' seems to have been derived from the Latin word----------
   (a)Statistik (b) Status (c) Statista (d) none of these.
   2. Statistics is not commonly used in----------
   (a) Mathematics (B) Science (c) Social sciences (d) none of these
   3. Statistics is the ----------of estimates and probabilities.
   (a) Science (b) Economies (c) Soci0logy (d) none of these
   4.Statistics is essential for a ----------
   (a) City (b) State (c) Country (d) none of these
   5. Laws of ---------- science are prefect.
   (a) Physical (b) Moral (C) Social (d) Accounts
   6. The following is not a limitation of Statistics
   (a) Qualitative aspect ignored
   (b) Statistics does not deals with individuals
   © To present facts in definite form
   (d) Statistics can be misused.
   7. Statistics as a singular noun means ----------
   (a) Statistical Data (b) Statistical Methods (c) Inductive Statistics (D) Descriptive statistics
   8. Statistics as a plural noun means ----------
   (a) Statistical Methods (b) A bundle of lies (C) Statistical Data (d) Fools paradise
   9. Distrust of statistics is due to----------
   (a) Misuse of statistics (b) Insufficient statistical Methods (c) Scope of statistics is limited (d) Statistics is considered useless by many people.
   10. Statistics facilitates ----------
   (a) Comparison of Data (b) Collection of Data (c) Organisation of Data (d) Disposal of Data.
   11. Statistics is the science of analyzing
   (a) Qualitiative Data (b) Quantiative Data (c) Any kind of Data (d) Only very simple facts
   12. Data collected originally for an investigation is known as
   (a) Primary data (b) Secondary data (c) Data (d) None
   13. Primary data is preferred over secondary data because
   (a) It is concise and accurate (b) It contains no errors (d) It shows greater details
   14. A questionnaire is a set of ----------
   (a) Questions (b) Answers (c) Problems (d) Elements
   15. The source for the secondary data may be ----------
   (a) Published (b) Unpublished (c) a and b (d) None
   16. Primary data are ----------
   (a) Original (b) Already collected (c) Original and first hand information (d) None
   17. Which of the following is primary data?
   (a) Census of population (b) Whole sale price index (c) Statistics contained in R,B.I
   (d) Data collected through own field survey.

18. Collection of secondary data is ----------
(a) Cheaper (b) Faster (c) More Accurate (d) a and b
19. Classification in which upper limit of interval is same as of the lower limit of class interval is called
(a) Exclusive method (b) Inclusive method (c) Mid point method (d) Ratio method
20. The number of questions in a questionnaire should be ----------
(a) 3 (B) 15 (c) 40 (d) as small as possible keeping in view the purpose of the survey.
Answers
1(.b) 2.(C) 3 (a) 4. (c) 5. (a) 6. (c) 7. (b) 8.(c) 9,(a) 10.(a) 11. (b) 12 (a) 13. (c) 14. (a) 15. (c) 16. © 17. (d) 18, (d) 19. (a) 20.(d)

II. FILL IN THE BLANKS.
1. The term Statistics is believed to have been derived from the Latin word ----------
2. Statistics as a ---------- noun means the same thing as data.
3. Statistics are ---------- statements.
4.---------- is the first step in a statistical investigation.
5. The nature of statistical inference is----------
6. The statistics as a ---------- noun refers to statistical methods.
7.---------- includes condensation and summarization of data.
8. ---------- facilitates comparison of data.
9. In plural sense statistics means----------
10. Statistics is ---------- statement.
11. ---------- or ---------- data are influenced by a number of factors.
12. Numerical data alone constitute----------
13. A reasonable standard of accuracy is needed in both----------
14. The first step in investigation is ----------
15. Statistics is widely used in ----------
16. ---------- is one of the main function 0f statistics.
17. Statistical methods are helpful to develop----------
18. Data collected by the investigator himself from primary sources is called----------
19. ----------are original and first hand information
20. ---------- is the first step in a statistical investigation.

ANSWERS

1. statistic 2. Plural 3, Quantitative 4. Collection of data 5. Inductive 6, singular 7. Analysis 8. Statistics 9. Collection of Numerical facts. 10. Numerical 11. Quantitative , Statistical 12. Statistics 13. Enumeration , Estimation 14. The collection of data 15. Education 16. Comparison 17. New theories 18. Primary data 19. Primary data 20. Collection of data.

III. ANSWER THE FOLLOWING QUESTIONS (0ne mark each)

1. Define statistics

There are two senses of the word statistics. In a wide sense statistics denote some statistical methods and plural sense statistics refers to some numeral data or statistical data.

2. Data array

Arrangement of raw data in the form of row or column is called data array.

3. Tabulation

The process of arranging data in tabular form i.e. in an orderly manner into rows and columns is referred to as tabulation.

4. Statistical investigation.

Refers to the process of collecting the data and analyzing it.The accuracy of the data derived by conducting the investigation relies upon how effectively planning is done for a statistical investigation.

5. Unbiased errors

These errors arise on account of chance. They are an offshoot of various statistical techniques.They are not very serious and do not affect the results significantly.

6. Biased errors

These errors are caused on account of a bias or prejudice of enumerators or of measuring instruments.

7. Inferential statistics

Statistical methods that help us arrive at certain conclusions based on study of sample data are part of inferential statistics.

8. Interpretation

Refers to drawing conclusions from data analysis. Interpretation is subjective and hence this stage is most crucial in terms of efficiency of statistical research.

9. Chronological classification

When the data is classified according to the time of occurrence it is known as chronological classification.

10. Parts of a table

Number, heading, headnote , subheadings body, comparison, totals, source, footnotes, attractive form.

UNIT- II

MULTIPLE CHOICE

1. Two dimensional diagrams are ---------- diagrams.

1(a) Linear 2,(b) volume 3.(c) Area (d) None

2. Square is a ---------- diagram.

1.(a) simple bar 2.(b) two dimensional (c) Volume (d) Linear

3. Ogive is a -------

1,(a) Frequency polygon 2. (b) cumulative frequency curve (c) line diagram (d) none of these.

4. ---------- create interest in the mind of the readers.

1,(a) tables 2. (b) Diagrams 3. (c) Charts (d) All the above.

5. ---------- bar diagrams can be drawn either or horizontal or vertical bars.

1. (a) multiple 2.(b) simple (c) sub-divided (d) All the above

6. ---------- diagrams are alternative to square diagrams

1.(a) circle (b) square (c) rectangles (d) none of these.

7. ----------diagrams ranks high in understanding.

1,(a) circle (b) square (c) pie (d) all the above

8. pie diagrams is called as ---------- diagrams .

1.(a) square 2. (b) circle 3. (c) angular 4. (d) none of these

9. Component graph also known as ---------- graph.

1. (a) band 2. (b) range (c) net balance (d) all the above

10. Net balance graph is also known as ----------

(a) range (b) band (c) silhouette (d) none of these.

11.---------- determine median, quartile, percentiles

(a) ogives (b) frequency curves (c) histograms (d) all the above.

12. ---------- curve should begin and end at the base line.

(a) histogram (b) frequency (c) ogives (d) none of these.

13. ---------- graphs are also known as zero graphs

(a) range (b) band (c) angular (d) n0ne of these .

14. Pie-diagram is ---------- dimensional diagram

(a)2 (b) 3 (c)4 (d) 1

15. ----------is a special kind of graph

(a) Z chart (b) T chart (c) ogive (d) a and b

16. Graphic is a ----------0f presentation .

(a) visual form (b) oral form (c) special form (d) none of these

17. Range=----------

(a) L=S (b) L-S(c) L+S (d) none of these

18. ---------- are presented through maps.

(a) cartograms (b) pictograms (c) a and b (*d) n0ne of these.

19. ----------diagrams are not attractive.

(a) bar (b) line (c) pie (d) none of these.

20. Which of the following is not a frequency graph?

(a) histogram (b) Z curves (c) ogives (d) none of these.

ANSWERS

1. © 2. (b) 3. (a) 4. (b) 5. (B) 6(a) 7. C) 8. (c) 9. (a) 10. ((c) 11.(a) 12. (b) 13. (a) 14. (a) 15. (a) 16. (a)

17. (b) 18. (a) 19. (b) 20. (b)

II FILL IN THE BLANKS

1, ---------- diagrams are used to present only one variable.

2. Sub divided bar diagrams are also known as ----------

3. ---------- diagrams can be adopted to compare changes in two or more than two interrelated variables.

4. ---------- represent quantitative knowledge on the map

5. A----------is a visual representation of the relationship between variables.

6. ---------- is a visual form for presentation of statistical data.

7. Diagrams cannot be ---------- further.

8. A diagram plays an important role in the ----------

9. ---------- is the simplest of all the diagrams.

10. ----------are used to denote more than one phenomenon.

11.---------- bar diagrams is used to depict the net deviations in different values.

12. ---------- is a device of representing statistical data in pictures.

13 ---------- are presented through maps.

14. ---------- diagrams are not attractive.

15. ---------- presentation of statistical data gives a pictorial effect.

16. Graphic is a ---------- of presentation.

17. A grouped frequency distribution can be represented by a ----------

18. A ----------is drawn by smoothing the frequency polygon.

19. ---------- means the graph relating to a series spread over a period of time.

20. ----------is also known as Zero graph.

ANSWERS

1. simple diagram 2. Component bar diagram 3, multiple bar diagram 4. Cartograms 5. Graphs

6. diagram 7. Analysed 8. Modern advertising campaigns. 9. Line diagrams 10. Multiple bar diagrams

11. deviation 12. Pictograms 13. Cartograms 14. Line diagram 15. Graphic 16. Visual form

17. histogram 18. Frequency curve 19. Histogram 20. Range graph.

III ANSWER THE FOLLOWING QUESTIONS.

1. Diagram- is a visual presentation of statistical information. The pictorial information helps in understanding the data.

2. Pictograms- is the technique of presenting statistical data through appropriate pictures.

3. Cartograms – are used to present the data pertaining to geographical regions with the help of maps.

4.Two-dimensional diagrams- in two dimensional diagrams both length and width of the observations are represented by bars.

5.merits of one dimensional diagram

a) these are easy to construct

B) these are easy to understand

C) comparison can be made easily by this device.

6.Bar graph- is a pictorial representation of data that uses bars to compare different categories of data.

7. 0gives curves – when frequencies are added they are called cumulative frequencies. The graph of such a frequency distribution is called ogives curves.

8. Types of diagrams-

a) one-dimensional diagram

B) two-dimensional diagram

c) three dimensional diagram

D) pictograms

e)cartograms

8. Types of two dimensional diagrams-a) squares (b) rectangles (c) circles.

9. Range graph- is used to depict the variation in data with reference to each period .The highest , the lowest and the average values of the variable being studied are plotted on the graph paper.

10. Z-chart- is a special kind of graph where figures are plotted at regular intervals over a period of one year. It is slightly complicated by the fact that it incorporates three graphs on the same chart and these curves or lines taken together tend to look like the letter Z

UNIT-III(MEASURES OF CENTRAL TENDENCY)

MULTIPLE CHOICE

1.Mean, median and mode are known as----------
a) average of position (b) mathematical average (c) measurement of central tendency
(d) none of the above.
2.The most popular method of measuring the representiative value is----------
(a) arithmetic mean (b) harmonic mean (c) ge0metric mean (d) none of the above.
3. If the lower limit of the first class interval and upper limit of the last class interval are not known it is called
(a) open-end classes (b) closed –end classes (c) mid-end classes (d) all of the above
4. When the total number of observations are divided by the sum of reciprocal of the numbers it is known as
(a) harmonic mean (b) geometric mean (c) arithmetic mean (d) all the above
5. When the middle value of the group of data are arranged in an order either on an ascending or descending order it is known as----------
(a) mean (b) median (C) mode (d) mean deviation
6. When the values of total frequencies are divided into four parts then it is----------
(a) Quartiles (b) Quintiles (c) Deciles (d) Percentiles
7. When the value of total frequency is divided into 10 parts then it is----------
(a) Deciles (b) percentiles (c) Quintiles (d) Standard deviation
8. When the value occurs most frequently in a statistical distribution it is termed as----------
(a) mean (b) median (c) mode (d) all the above
9. Which of the following can be used for averaging percentages rate of change and index numbers?
(a) Geometric mean (b) Harmonic mean (c) Mode (d) None of the above
10. In a symmetric distribution the values of mean , median, and mode will----------
(a) coincide (b) differ (c) change (d) all the above
11. When the values of total frequencies are divided into eight parts then it is ----------
(a) quartiles (b) octiles (c) percentiles (d) deciles
12. If all the items of the data are of same value then the arithmetic mean would be equal to----------
(a) mean (b) mode (c) geometric and harmonic mean (d) none of the above
13. The sum of all the deviations of all the observations from arithmetic mean is equal to----------
(a) Zero (b) one (c) two (d) three
14. A short way of expressing an arithmetical result is ----------

(a) median (b) kurtosis (c) mean (d) mode
15. Number of observations are 30 and value of arithmetic mean is 15 then sum of all values is----------
(a) 15 (b) 450 (C) 200 (d) 45
16. Arithmetic mean is 25 and all sum of observations isn350 then number of observations are
(a) 25 (b) 70 (C) 14 (d) 75
17. Arithmetic mean is 12 and number of observations are 20 then sum of all values is----------
(a) 8 (b) 32 (c) 240 (d) 1.667
18. mean, median, and mode are known as----------
(a) average of position (b) mathematical average (c) measurement of central tendency (d) All the above
19. In a symmetrical distribution
(a) mean= median = mode (b) mean>median> mode (c) mean<median<mode (d) no such relationship exists
20. Which of the following averages cannot be located graphically ?
(a) arithmetic mean (b) median (c) mode (d) all the above

II FILL IN THE BLANKS
1. Arithmetic mean has an ---------- bias.
2. When a series is multi-modal, we say that mode is----------
3. ----------is the most preferred and most used average.
4. If mode is ill-defines it is ascertained by the formula----------
5. Geometric mean has a ----------bias.
6. ---------- is the mid value of all the given observations.
7. The mode of 6, 4, 5, 8, 6, 7,7, 6,2,6 is----------
8.---------- refers to the values which divide the given series into 100 equal parts.
9.The most common item in a series of observations is called----------
10. ---------- is the reciprocal of the arithmetic mean of the reciprocals of the individual observations.
11. Weighted mean should be calculated when the importance of the items in a series is ----------
12. The ---------- mean is more useful when we want to find the rate of growth.
13.---------- is the nth root of the product of n items of a series.
14.The first quartile covers ---------- percentage of the series.
15. Second quartile is same as ----------.
16. the fist decile covers ------- percentage of the series.
17. ---------- average cannot be located graphically.
18. The averages are affected by change of origin and ----------
19. mode= 3----------
20. Median is also known as ---------- average.

Multiple choice key

1)c 2)a 3) a 4) a (5) b (6) a 7(a) 8(c) 9(a) 10. (a) 11(b) 12. (b) 13. (a) 14. (c) 15. (b) 16.(c) 17. (c)

18. © 19.((a) 20.(a)

Fill in the blanks key
1. upward 2, indeterminate 3. Mean 4. Mode= 3median – 2mean 5. Downward 6. Median
7. 6 8. Percentiles 9. Mode 10. Harmonic mean 11. Not equal 12. Geometric mean
13. geometric mean 14. 25. 15. Median 16. 10percent 17. Arithmetic mean 18. Scale 19. Median –2mean 20. Positional

III ANSWER THE FOLLOWING QUESTIONS
1. Harmonic mean
It is the reciprocal of the arithmetic mean of the reciprocals of the individual observations.
2. If mode and mean are 28 and 25 respectively find value of median
Mode=3median----2mean
28=3median-2(25)
3median=78 hence median=26
3. Averages –are the values which lies between the smallest and largest observations.
4. Partition values----- are also known as fractiles. They break the series into number of equal parts.
5. Median--- is the mid value of all the given observations. It divides the distribution into two equal parts.
6. calculate arithmetic mean -- 32,35,36,37,39,41, 43
Mean=263/7=37.57
7. 2 limitations of mode
(a) it cannot be used for further mathematical calculations
(b) it is affected by fluctuations.
8. 2 merits of mean
(a) it is simple to calculate and understand
(b) it is systematically defined.
9. mode –is the value which has maximum frequency than any other value in a kurtosis set of observations.
10. Deciles --- are the partition values which break the distribution into ten equal parts.

UNIT-4

MULTIPLE CHOICE

1. ---------- tells us whether her a distribution is symmetrical or skewed.
(a) moments (B) Range (c) mesokurtic (d) a and b
2. A normal distribution is also called ---------- distribution .
(a) mesokurtic (b) kurtosis (c) flat (d) none of the above
3. Skewness means lack of ----------in a frequency distribution
(a) asymmetrical (b) symmetry (c) unity (d) none of the above
4.In leptokurtic curve is----------
(a) greater than 3 (b) less than 3 (c) equal to 3 (d) not equal to 3

5.. Coefficient of variation is proposed by

(a) Bowley (b) Kelly (c) Karl pearson (d) none of the above.

6.If frequency curve is flatter than normal curve then it is known as----------

(a) Leplokurtic (b) Mesokurtic (c) Platykurtic (d) none of these

7. standard deviation for discrete series can be calculated by using ----------

(a) Actual mean method (b) Assumed mean method (c) Step deviation method

(d) all the above

8. Which of the following ignores algebraic signs ?

(a) Mean deviation (b) Standard deviation (c) Quartile deviation (d) Coefficient of variation

9. In asymmetrical distribution the coefficient of skewness will be:

(a) 0 (b) 4 (c) 8 (d) 1

10. Departure fromsymmetry is called

(a) second moment (b) kurtosis (c) skewness (d) variation

11. The range of the values -5, -8, -10, 0, 6, 10 is

(a) 0 (b) 10 (c) -10 (d) 20

12. The measure of dispersion can never be

(a) positive (b) Zero (c) Negative (d) Equal to 2

13. The scatter in a series of values about the average is called

(a) Central tendency (b) Dispersion (c) Skewness (d) Symmetry

14. If there are many extreme scores on all examination ,the dispersion is

(a) small (b) Large (c) Normal (d) Symmetric

15. Which measure of dispersion is not impacted by extreme observations?

(a) coefficient of range (b) Quartile deviation (c) Mean deviation (d) Standard deviation

16. Given standard deviation=3,then Variance=---------

(a) 1 (b) 6 (c) 9 (d) 3

17. The value of coefficient of variance varies between ----------

(a) 0and1 (b) 1and 2 (c) -1and2 (d) 1and-1

18. Higher the variance,---------- is the variability or dispersion of the series.

(a) Greater (b) Smaller (c) Equal (d) none of these

19. Given Standard deviation=140, Arithmetic mean= 700, co-efficient of Variation =----------

(a) 20% (b) 50% (c) 10% (d) 2%

20. The measure of Dispersion which uses only two observations is called

(a) Mean (b) Median (c) Range (d) Coefficient of variation.

II. FILL IN THE BLANKS

1. ---------- is the measure of variation of items.

2. ---------- is the graphical method of measuring variation in the distribution of wages, profits, turnover etc.

3. Root mean square deviation is another name of----------

4. Statistical measures of ---------- is used to determine the extent and direction of a symmetry in a series.

5. When B curve is neither flat nor sharply peaked is known as ---------- curve.

6. ---------- enables us to measure the flatness or peakedness of the curve.

7. If the longer tail is towards the lower value or left hand side the skewness is ----------

8. ---------- is defined as the difference between the third quartile and first quartile.

9. Standard deviation is always ---------- than range.
10. Bivariate correlation is also known as ---------- correlation.
11. Variance is ---------- of standard deviation.
12. The quartile deviation includes ----------.
13. Quartile deviation is ---------- of standard Devaiation.
14. ---------- judges the differences between the central values.
15. ---------- can be used to test the reliability of an Average .
16. The standard deviation concept was introduced by---------- in 1823.
17. The word skewness refers to ----------.
18.Absolute measure of skewness =----------
19. ---------- helps in isolating the impact of different factors.
20. The Kurtosis of any univariate normal distribution is ----------.

III. ANSWER THE FOLLOWING QUESTIONS

1.Dispersion-- is the measure of variation of items.
2. Range – is the difference between the largest item and the value of smallest item in the distribution.
3. Quartile deviation – shows the average amount by which the two quartiles differ from median.
4. Standard deviation – is the square root of the arithmetic average of the squares of the deviations measured from mean.
5. Variance- is the square of standard Deviation . It helps in isolating the impact of different factors.
6. Kurtosis – is the degree of peakedness of a distribution usually taken relative to a normal distribution.
7. Demerits of mean deviation
(a) It ignores algebraic signs
(b) It does not produce 100% accurate results.
© It is not useful for sociological studies.
8. Lorenz curve- is a graphical method of measuring dispersion. It is used in measuring variations in the distribution of wages, profits, turnover and so on.
9. Merits of Range
(a) It is one of the simplest method and easiest of comparison.
(b) It takes minimum time for calculating.
10. Skewness- refers to lack of symmetry in a frequency distribution. The mean, median, mode fall at different points in the distribution.

KEY

1(a) 2(a) 3(b) 4(a) 5(c) 6(c) 7(d) 8(a) 9(a) 10(a) 11(d) 12(c) 13(b) 14(b) 15(b) 16(c) 17(d) 18(a)19(a) 20(c)

Fill in

1. Dispersion 2. Lorenz curve 3. Standard deviation 4. Skewness 5. Mesokurtic 6. Kurtosis 7. negative 8. Interquartile range 9.less 10. Simple 11. Square 12. Central 50% (13) 0.6475(14.) Skewness 15. A measure of dispersion 16. Karl Pearson 17. Lack of symmetry 18. Mean- mode 19. Variance 20) 3

UNIT-5

MULTIPLE CHOICE

1.Who is the chief exponent of the principle of correlation
(a) Bowley (b) Spearman (c) Karl Pearson (d) Connor
2. What is meant by correlation
(a) measuring relationship between variables (b) measuring changes in prices (c) measuring average correlation between the variables (d) prediction
3. Study of relationship between three or more variables is known as
(a) multiple correlation (b) simple correlation (c) both (d) none
4.What is non-mathematical method of studying the correlation between the variables
(a) scatter diagram method (b) Graphic method (c) both a and b (d) none of these
5. If Y is more than 6 times the probable error correlation is
(a) negative (b) positive (c) significant (d) none of these.
6. A process by which we estimate the value of dependent variable on the basis of one or more independent variables is called
(a) correlation (b) regression (c) residual (d) slope
7. The value we would predict for the dependent variable when the independent variable are all equal to zero is called
(a) slope (b) sum of residual (c) intercept (d) difficult to tell
8. The predicted rate of response of the dependent variable to changes in the independent variable is called
(a) slope (b) intercept (c) error (d) regression equation
9. In simple regression equation the numbers of variable involved are
(a) 0 (b) 1 (c) 2 (d) 3
10. The dependent variable is also called
(a) regress and variable (b) predict and variable (c) explained variable (d) allof these
11.The correlation between price and demand is
(a) negative (b) positive (c) 0 (d) none of these
12. The term regression was used by
(a) Edward (b) Galton (c) Connor (d) Tuttle
13.If Karl Pearson's coefficient of correlation is -123 it signifies
(a) absolute negative correlation (b) perfect negative correlation (c) correlation cannot be negative (d) correlation cannot be lower than -1
14. When price increase demand decreases. This is an example of
(a) positive correlation (b) Negative correlation (c) No correlation (d) Linear correlation
15.coefficient of Correlation lies between
(a) -1 and 1 (b) 1 and -1 (c) 2and 1 (d) N0ne of these
16. r is significant when

(a) r>6 P.E (b) r< 6 P.E (C) r=6 P.E (d) none of these
17. Which of the following methods of measuring correlation is impacted by extreme values?
(a) scatter diagram (b) Karl Pearson's method (c) Spearman's Rank Correlation (d) Concurrent deviation method
18. If the value of r is less than the probable error then
(a) r is significant (b) there is no evidence of correlation (c) there is evidence of correlation, but not much (d0 all of the above.
19. Which of the following statements is true about Spearman's Rank correlation ?
(a) It is used to study correlation between qualitative aspects.
(b) It cannot be used when two variables have the same value
© It is used to study correlation between two quantity values.
(d) none of these
20. The reason for high degree of correlation are
(a) pure chance (b) mutual dependence (c) Limiting factor (d) Both a and b

KEY
1(C) 2(A)3(A) 4(C) 5(C) 6(B) 7(C) 8(B) 9(C0 10(D) 119A0 12(B0 13((C) 14(B) 15(A) 16(A)17(B) 18(B) 19(A) 20(D)

II FILL IN THE BLANKS
1, KarlPearson's coefficient of correlation measures ----------
2, Maximum positive value of coefficient of correlation is----------
3. Correlation between price and demand is----------
4. Correlation between price and supply is----------
5. If the correlation is perfect positive, its value is----------
6. The independent variable is also called ----------
7. If the value of any regression coefficient is Zero, then two variables are----------
8. The range of regression coefficient is----------
9. The term regression was used by----------
10. The correlation coefficient between Xand Y is ----------
11. There are ---------- types of correlations.
12. ---------- is an analysis of the correlation between two or more variables.
13. High degree of correlation exists when the variables have ---------- relationship.
14. The variable whose value is influenced or is to be predicted is called----------
15. Scatter diagram portraits the relationship between two variable----------
16. The reason for high degree of correlation are pure chance and----------
18. There is a ---------- correlation between rainfall and stock prices
19. Positive correlation means that the direction of change is likely to be----------
20. Under---------- method correlation is calculated between the direction of deviation in the two series..

KEY

1. Linear correlation 2,+1 3. Negative 4. Positive 5. +1 6. Regressor 7. Independent 8, -to +

9. Galton 10. -1 (11) -2 (12) correlation 13. Cause and effect 14, dependent variable
15. Graphically 16. Mutual dependence 18. High positive 19. Same 20. Concurrent Deviation method.

III ANSWER THE FOLLOWING QUESTIONS

1. What is correlation

Correlation is the study of the linear relationship between two variables.

2. Types of correlation

(a) positive correlation (b) negative correlation (c) Linear correlation (d) N0n- linear correlation

3. positive correlation/

If the values of two variables deviate in the same direction.

4. Properties of Spearman's Rank correlation

(a) It lies between -1

(b) It is based on subjective ranking of variables.

5. Merits of Spearman's Rank correlation

(a) It is easy to understand

(b) It is not impacted by extreme values

© It facilitates comparisons between two series.

(d) It can be applied to irregular data

6. Demerits of Spearman's Rank correlation

(a) It cannot be applied to grouped data

(b) It lacks the precision of Karl Pearson's Coefficient of correlation

© The computation becomes complicated as the number of observations increase.

7. Methods of studying correlation

(a) scatter diagram (b) correlation graph (c) Karl Pearson's coefficient of correlation

(d) coefficient of correlation by rank differences. € coefficient of concurrent Deviation.

8. Examples for positive correlation

(a) sales revenue of a product and expenditure on advertising

(b) number of workers and output of a factory

© amount of rainfall and yield of a crop

9. Properties of Karl Pearson's coefficient of correlation

(a) It is based on Arithmetic Mean and Standard Deviation

(b) It lies between -1

© It takes into account all types of variables.

10. What is meant by Multiple correlation ?

It refers to the situation where in the problems have either three or more variables. These can be both dependent or independent variables.