EDU SERUMX COLLEGE OF COMMERCE and MANAGMENT

SYLLABUS

Class – B.B.A I year

Subject –Business Statistics

Unit	Topic	Subtopics
Unit I & II	Introduction to Statistics	- Statistical Investigations, Laws of Statistics - Scope and Limitations of Statistics
	Population and Sample	- Concept of Population and Sample - Types of Data: Primary and Secondary - Collection and Organization of Data
	Frequency Tables & Distribution	- Frequency Tables - Frequency Distribution
	Data Presentation	- Bar Diagram - Pie Diagram - Line Graph - Histogram - Frequency Polygon
Unit III	Measures of Central Tendency	- Mode - Median - Geometric Mean - Harmonic Mean
Unit IV	Measures of Dispersion	- Range - Quartile Deviation - Mean Deviation - Standard Deviation - Concept of Skewness
	Time Series Analysis	- Basic Concept of Time Series Analysis
Unit V	Correlation	- Meaning and Definition of Correlation - Types of Correlation
	Correlation Coefficients	- Karl Pearson's Coefficient of Correlation - Coefficient of Determination - Spearman's Rank Correlation
	Simple Linear Regression	- Lines of Regression - Index Numbers

UNIT — I& II STATISTICS	
Introduction	**Statistics** is a branch of mathematics that deals with collecting, analyzing, interpreting, presenting, and organizing data. Its purpose is to help us understand, summarize, and make informed decisions based on the information provided by data. Statistics has applications across almost all fields—such as economics, business, psychology, social sciences, medicine, and government—where it provides insights and supports decision-making.
Origin	The term "statistics" comes from the Latin word **"status"**, meaning "state." Historically, statistics were used for state governance, specifically for collecting information about populations, land, resources, and other assets to make administrative decisions. Early uses of statistics can be traced back to ancient civilizations, where rulers collected data on population sizes, taxation, and military strength. The modern development of statistics as a mathematical discipline began in the 17th and 18th centuries. With advancements in probability theory and data analysis methods, statistics grew to include various mathematical techniques used to analyze complex data sets. This development continued into the 19th and 20th centuries, expanding the field into inferential statistics, which allows conclusions to be drawn from sample data.
Meaning of Statistics	Statistics can have two main meanings:
	1. As Data: Statistics refers to numerical information or data collected about a particular subject. For example, data like average income, birth rates, or test scores are all examples of statistics. **2. As a Science**: Statistics is the science and methodology used to collect, analyze, interpret, and present data. It involves various techniques for making sense of large amounts of data, finding patterns, drawing conclusions, and making predictions.
Key Aspects of Statistics:	**Descriptive Statistics**: Summarizing and presenting data in a way that provides an overview, such as averages, percentages, or visualizations (e.g., bar charts, pie charts). **Inferential Statistics**: Making predictions or generalizations about a population based on a sample of data, using tools like confidence intervals, hypothesis testing, and regression analysis.

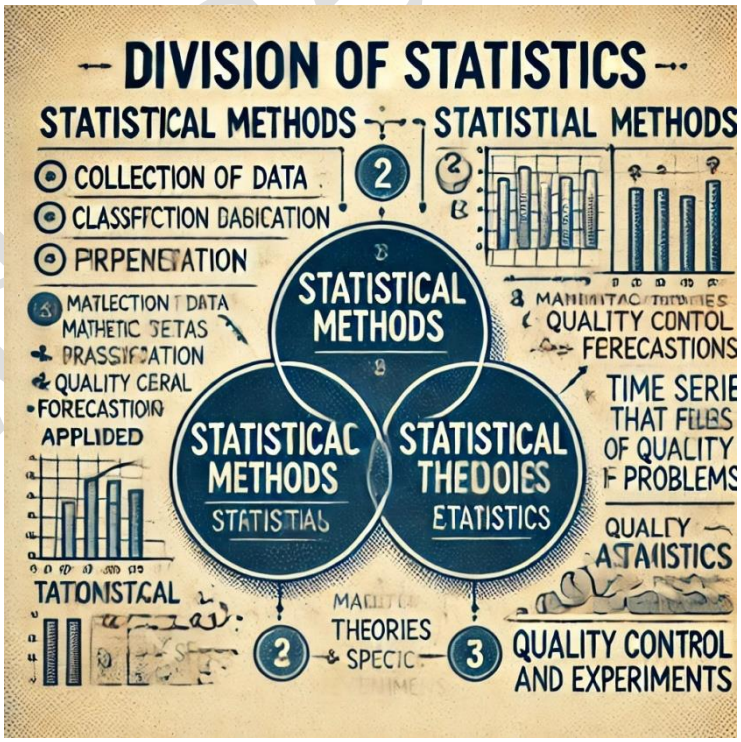
Definitions: - "The classified facts relating the condition of the people in a state specially those facts which can be stated in members or in tables of members or in any tabular or classified arrangements." -

-Webster

"Statistics may be regarded as (i) the study of population (ii) The study of variation (iii) The study of method of reduction of data"

-R.A. Fisher

Nature /Features /Characteristics of statistics
It is an aggregate of facts It is estimated according to reasonable standard of accuracy It is numerically expressed Analysis of multiplicity of causes. It is collected for pre-determined purpose It is collected in a systematic manner



Functions of Statistics:

1. **Presents Facts Clearly**: Statistics organizes data into a clear, specific form, making it easier to understand.
2. **Simplifies Large Amounts of Data**: It condenses complex data, so it's easier to handle and analyze.
3. **Facilitates Comparison**: It enables us to compare data and see patterns or trends.
4. **Helps in Prediction**: By analyzing data trends, statistics helps us make educated guesses about the future.
5. **Aids Policy Formulation**: Governments and organizations use statistics to develop effective policies and strategies.

Scope of Statistics:

1. **Statistics and Government**: Used by governments to make decisions and develop policies.
2. **Statistics in Business and Management**:
    - **Marketing**: Used to analyze customer preferences and improve marketing strategies.
    - **Production**: Helps optimize production processes and improve quality.
    - **Finance**: Assists in making financial forecasts and budgeting.
    - **Banking**: Used for credit analysis, interest rate calculations, etc.
    - **Control**: Monitors business performance and quality control.
    - **Research and Development**: Helps develop new products and improve existing ones.
3. **Statistics and Economics**:
    - **National Income Measurement**: Used to calculate national income and understand economic growth.
    - **Money Market Analysis**: Analyzes stock markets, interest rates, and inflation.
    - **Market Analysis**: Assists in studying different market structures, like competition and monopoly.
    - **Population Analysis**: Helps study population trends and demographics.
4. **Statistics and Science**: Essential for conducting scientific research and analyzing results.
5. **Statistics and Research**: Widely used in various fields to analyze research data and draw conclusions.

Limitations of Statistics:

1. **Deals with Groups, Not Individuals**: Statistics focuses on large data groups rather than individual cases.
2. Requires Expertise: Only experts can correctly interpret and apply statistical data.
3. **Not the Only Solution**: Statistics is just one tool among many for solving problems.
4. **Risk of Misuse**: Statistics can be easily manipulated to mislead people if not used ethically.

EDU SERUMX COLLEGE OF COMMERCE and MANAGEMENT

STATISTICAL INVESTIGATION

Meaning: In general it means as a statistical survey.
In brief, it is Scientific and systematic collection of data and their analysis with the help of various statistical method and their interpretation.

STAGES OF STATISTICAL INVESTIGATION:-

EDU SERUMX COLLEGE OF COMMERCE and MANAGEMENT

PROCESS OF DATA COLLECTION

Data: Data is a collection of information, like numbers, words, measurements, or observations, gathered to give us useful insights. Think of it as a bundle of facts or pieces of information about a topic.

Data Collection: This is the process of gathering information for a specific purpose, organizing it, and relating it to each other to understand something better. For example, if a company wants to understand customer satisfaction, they collect data through surveys, reviews, or interviews, then look at all the responses together to see patterns and make informed decisions.

Data

Facts and figures pertinent to the problem

Secondary Data:

Internal Data: Information collected from within the organization such as financial statements, customer letters, and sales reports.

External Data: Information collected from outside the organization like U.S. Census reports, business magazines, and internet-based reports.

Primary Data:

Observational Data: Data collected by watching people, using methods like personal or mechanical observation.

Questionnaire Data: Data collected by asking people through surveys, interviews, or focus groups. This can be done in-person, by mail, online, or over the phone.

Internal Data:

Information collected from within a company.

Examples: Financial statements, customer feedback, and sales reports.

External Data:

Information collected from outside a company.

Examples: Government reports, trade magazines, and online articles.

Observational Data:

Information collected by watching how people behave.

Examples: Watching how customers shop in a store or how they use a product.

Questioning Data:

Information collected by asking people questions.

Collection of Data: This refers to the ways we gather information needed for a study or investigation.

Methods of Data Collection (Primary Data):

1. **Direct Personal Interviews**: Talking face-to-face with people to gather information.
2. **By Observation**: Watching and recording behavior or events as they happen.
3. **By Survey**: Asking a group of people the same set of questions.
4. **By Questionnaires**: A form with a set of questions that people fill out themselves.

Preparation of Questionnaires:
This is a common method, especially for large studies. It's used by individuals, researchers, companies, and even governments. A questionnaire has a list of questions printed in a specific order that the respondents answer on their own.

Importance of Questionnaires:

1. **Low Cost and Universal**: It's affordable and can be used almost anywhere.
2. **Free from Bias**: There's less chance of influencing the answers.
3. **Adequate Time for Response**: People can answer at their own pace.
4. **Easily Accessible**: It's easy to reach many people with questionnaires.

Basis of Difference

1. **Source**:
   - **Primary Data**: Collected directly from the original source by the researcher.
   - **Secondary Data**: Already collected by someone else and available for use.
2. **Purpose**:
   - **Primary Data**: Collected for a specific purpose or research study.
   - **Secondary Data**: Collected for a different purpose but reused for current study.
3. **Examples**:
   - **Primary Data**: Interviews, surveys, observations.
   - **Secondary Data**: Research articles, government reports, historical data.
4. **Accuracy**:
   - **Primary Data**: More accurate and relevant as it's collected directly for the study.
   - **Secondary Data**: Less accurate for specific needs, as it may be outdated or not perfectly fit the study.
5. **Cost and Time**:
   - **Primary Data**: More expensive and time-consuming to collect.
   - **Secondary Data**: Cheaper and quicker, since the data already exists.

Demerits:-

(i) Low rate of return

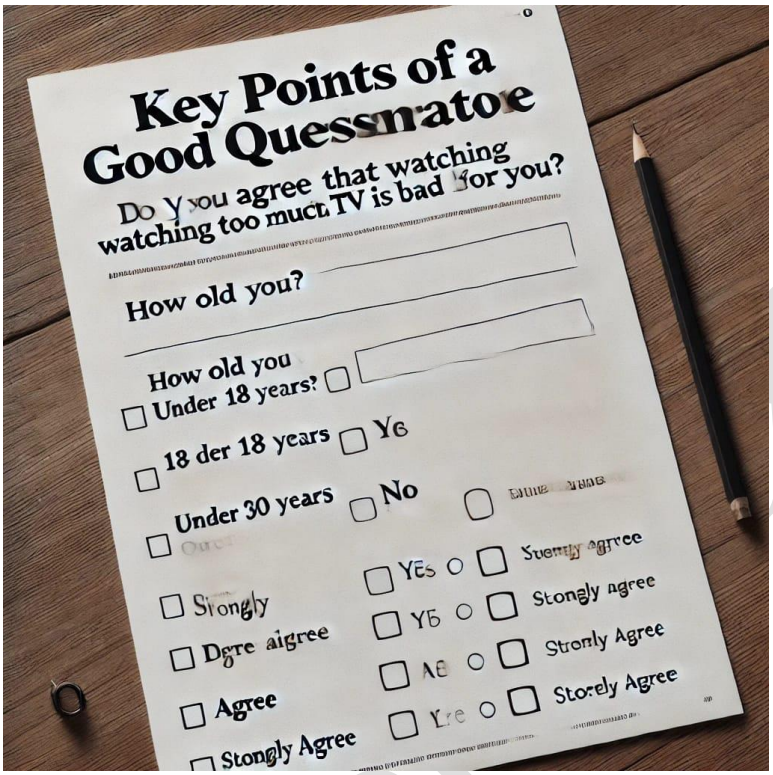
(ii) Fill on educated respondents

(iii) Slowest method of Response

Steps in construction of a questionnaire:It is considered as the heart of a survey operation. Hence it should be very carefully constructed

Prepare it in a general form	Prepare sequence of question	Emphasize on question formulation and wordings	Ask Logical and not misleading questions.	Personal questions should be left to the end.	Technical terms and vague expressions should be avoided.

Example :



Classification & Tabulation of Data

After collecting and editing of data an important step towards processing that classification. It is grouping of related facts into different classes.

Types of classification:-

i. **Geographical**:- On the basis of location difference between the various items. E.g. Sugar Cave, wheat, rice, for various states.

ii. **Chronological:-** On the basis of time

e.g.-

Year	Sales
1997	**1,84,408**
1998	**1,84,400**
1999	**1,05,000**

iii. Qualitative classification: - Data classified on the basis of some attribute or quality such as, color of hair, literacy, religion etc.

iv. Quantitative Classification: - When data is quantify on some units like height, weight, income, sales etc.

Tabulation of Data

A table is a systematic arrangement of statistical data in columns and Rows. Part of Table:-

1. Table number
2. Title of the Table
3. Caption
4. Stub
5. Body of the table
6. Head note
7. Foot Note

Types of Table:-

(i) Simple and Complex Table:-

(a) Simple or one-way table:-

Age	No of Employee
25	10
30	7
35	12
40	9
45	6

(b) Two way Table:-

Age	Male	Female	Total
25	25	15	40
30	20	25	45
35	24	20	44
40	18	10	28
45	10	8	18
Total	97	78	175

2) General Purpose and Specific Purpose Table:-General purpose table, also known as the reference table or repository tables, which provides information for general use or reference. Special purpose are also known as summary or analytical tables which provides information for one particular discussion or specific purpose.

METHODS OF SAMPLING

1. Random Sampling

- Each person or item in the population has an equal chance of being selected.
- Like drawing names from a hat where everyone has the same chance to be picked.
- **Use Case:** When you want an unbiased representation of the population.

2. Systematic Sampling

- You pick every "n-th" person or item in a list. For example, selecting every 10th person.
- **Use Case:** When the list is organized, and you need a quicker selection.

3. Stratified Sampling

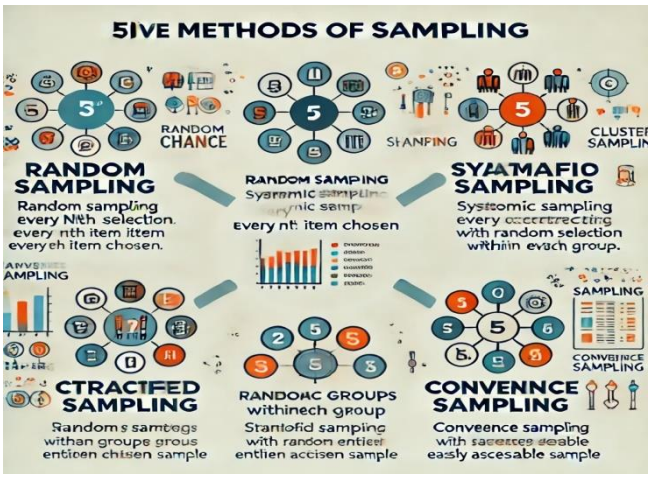
- The population is divided into groups (strata) based on a shared characteristic, like age or gender.
- Then, a random sample is taken from each group.
- **Use Case:** When certain characteristics should be represented equally.

4. Cluster Sampling

- The population is divided into clusters (e.g., schools in a city), and some clusters are chosen at random.
- Then, all individuals within selected clusters are studied.
- **Use Case:** When the population is large and spread over a wide area.

5. Convenience Sampling

- The sample is taken from a group that is easy to access, like surveying people in a mall.
- **Use Case:** When time or resources are limited, though it may not be as accurate.

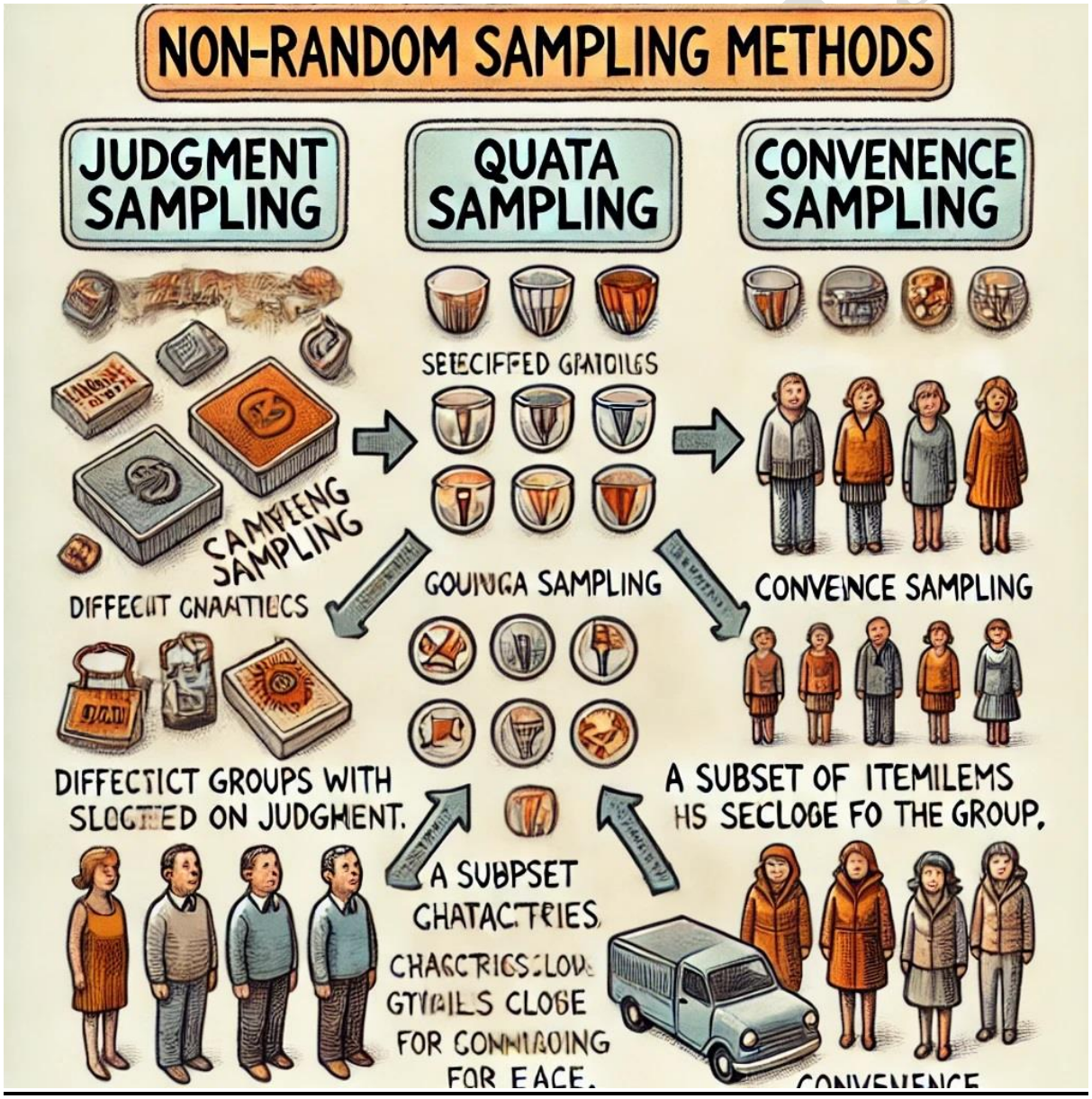


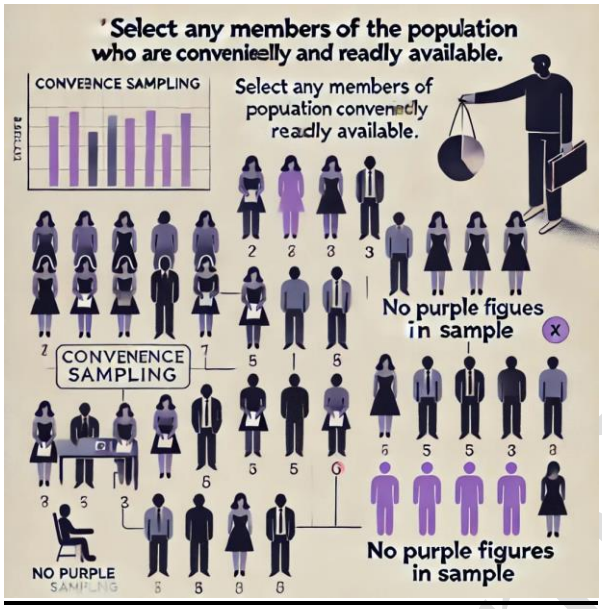
Non Random Sampling Method:-

 Judgment Sampling: Here, the researcher picks the sample items based on their own knowledge or judgment. They believe these items are the most representative or useful for the study.

 Quota Sampling: In this method, the researcher sets certain criteria (like age, gender, etc.) and picks samples to meet those quotas. Within each group, the choice of who to include is based on the researcher's judgment.

 Convenience Sampling: Also called "chunk sampling," this is simply choosing a sample that's easy to access. The sample isn't chosen for being representative or random; it's picked just because it's readily available.





Size of Sample:- It depends upon the following things:- Cost aspects.
The degree of accuracy desired. Time, etc.
Normally it is 5% or 10% of the total population.

Limitation of overall sampling Method:-
Some time result may be inaccurate and misleading due to wrong sampling.
Its always needs superiors and experts to analyze the sample. It may not give information about the overall defects. In production or any study.
It Becomes Biased due to following reason:-
(a) Faulty process of selection
(b) Faulty work during the collection of information
(c) Faulty methods of analysis etc.

UNIT-3
MEASURES OF CENTRAL TENDENCY

Central Tendency

The central tendency is the tendency for data to group around a middle value. This middle point, or "central value," is where most data points tend to concentrate.

Purpose of a Statistical Average

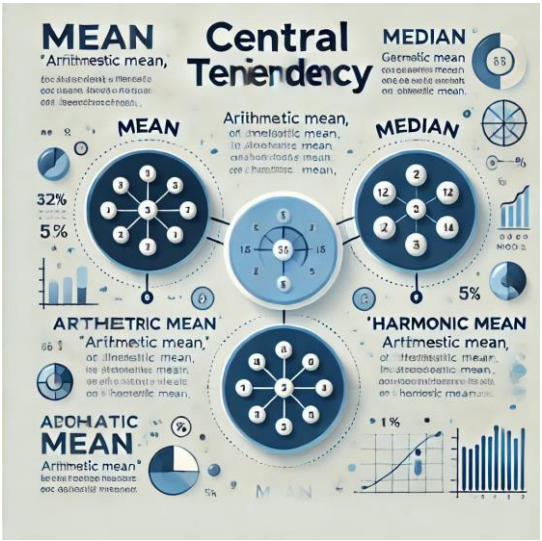
1. **Summarizes Data**: A single value represents the whole group.
2. **Allows Comparison**: Makes it easy to compare groups or data sets.

Functions of a Statistical Average

1. **Describes the Group**: Provides an overview of all data.
2. **Guides Planning**: Helps in predicting or making future decisions.
3. **Supports Analysis**: Used in calculations to understand relationships.
4. **Aids in Decision-Making**: Helps make informed choices.

Qualities of an Ideal Average

1. **Clear Definition**: It should be simple to explain.
2. **Easy to Grasp**: People should understand it easily.
3. **Quick Calculation**: Simple to compute.
4. **Includes All Data**: Takes all observations into account.
5. **Unaffected by Outliers**: Extreme values shouldn't distort it.
6. **Stable with Different Samples**: Minor variations in samples shouldn't affect it much.
7. **Algebra-Friendly**: Should work well in mathematical calculations.



ARITHMETIC MEAN (X)

Arithmetic Mean

The arithmetic mean is a common measure of central tendency. It's calculated by adding up all observations in a dataset and dividing by the number of observations. The arithmetic mean is often just called the "mean."

Merits (Advantages) of the Mean

1. **Clearly Defined**: The mean has a strict, easy-to-understand formula.
2. **Simple to Calculate**: It's quick and straightforward to compute.
3. **Uses All Data Points**: Every observation is considered in the calculation.
4. **Good for Math**: The mean works well in algebraic calculations.
5. **Less Affected by Small Variations**: It remains stable even with minor changes in data.
6. **No Order Needed**: Data doesn't need to be arranged in any specific order.

Demerits (Disadvantages) of the Mean

1. **Sensitive to Extremes**: Very high or low values can skew the mean.
2. **Not Great for Ratios or Percentages**: It doesn't handle these well.
3. **Requires Complete Data**: If data is missing, it can't be accurately calculated.
4. **May Not Match Observations**: The mean might not be an actual value in the data.
5. **Not Easy to Guess by Looking**: You can't determine the mean by just inspecting the data.
6. **Limited with Open-Ended Classes**: It's challenging to calculate if the data includes open-ended categories (e.g., income "above $100k").

Uses of the Mean

1. **Symmetrical Distributions**: Best for data that's evenly spread.
2. **Reliable Average Needed**: Ideal when we want a stable central value.
3. **Foundation for Other Calculations**: Essential for more complex statistics, like standard deviation and correlation.

MEDIAN (M)

The median is the middle value in a set of data that's been arranged in ascending or descending order. It divides the data into two equal halves, where half the values are greater and half are less than the median. It's often denoted by **M**.

Merits (Advantages) of the Median

1. **Easy to Understand and Calculate**: The median is straightforward to grasp and compute.
2. **Well-Defined**: It has a clear definition.
3. **Can Sometimes Be Found by Inspection**: In simple cases, it can be identified just by looking.
4. **Not Affected by Extremes**: Unlike the mean, very high or low values don't impact the median.
5. **Good for Qualitative Data**: Useful with data that's ranked, like satisfaction or performance.
6. **Works with Open-Ended Classes**: Only the position of values matters, not the exact numbers.

Demerits (Disadvantages) of the Median

1. **Requires Data Ordering**: Data must be sorted in ascending or descending order.
2. **Not Based on All Values**: It only considers the middle position, not each individual observation.
3. **Limited for Calculations**: The median isn't used for further mathematical analysis.
4. **Less Accurate with Large Datasets**: It can be less precise for very large datasets.
5. **Affected by Sampling Changes**: Its value can vary more across samples than the mean.

Uses of the Median

1. **When Data Has Extremes**: Ideal for data with very large or small values, which might skew other averages.
2. **Open-Ended Classes**: Useful if the data has open-ended ranges (e.g., "age above 80").
3. **Ranked Data**: Good for cases where data can be ordered but not measured precisely, like satisfaction levels.
4. **Distributions with Wealth or Income**: Useful for identifying typical values in income or wealth distributions.

Mode (Z)

The **mode** is the value that appears most often in a dataset. It represents the most "fashionable" or common item in the data. This is why it's sometimes referred to as the "typical" value. The mode is denoted by **Z**.

Geometric Mean (G.M)

The **geometric mean** is a different type of average that's useful for finding the "typical" value when you have multiple values multiplied together (like growth rates). It's found by multiplying all values in the set and then taking the nth root (where n is the number of values). The geometric mean is often used in situations involving percentages or rates and is denoted by **G.M**.

Properties of Geometric Mean

1. **Less Than Arithmetic Mean**: It's usually smaller than the arithmetic mean.
2. **Replacement of Values**: If each value is replaced by the geometric mean, the product of the values stays the same.
3. **Ratios**: For two sets of data, the ratio of their geometric means equals the geometric mean of each corresponding ratio.
4. **Product Consistency**: When items from two sets are multiplied, their geometric mean reflects the average of these products.

Merits (Advantages) of Geometric Mean

1. **Precisely Defined**: It's a clearly defined and exact figure.
2. **Uses All Observations**: Every value in the dataset is considered in the calculation.
3. **Good for Math**: Works well in mathematical calculations.
4. **Stable with Sampling**: Less affected by small changes in sample data.
5. **Not Impacted by Extremes**: Outliers (very high or low values) don't greatly affect it.

Demerits (Disadvantages) of Geometric Mean

1. **Limits with Zero or Negative Values**: It can't be used if any value is zero or negative.
2. **Complex Calculation**: It's more challenging to calculate than other averages.
3. **Harder to Understand**: Not as intuitive as the arithmetic mean.
4. **May Not Match Observed Values**: The result might not be a value that's actually in the data.

Uses of Geometric Mean

1. **When Data Includes Rates or Growth**: Good for data involving percentages, ratios, or growth rates.
2. **Stability Against Extremes**: Suitable when extreme values might distort other averages.
3. **Comparing Ratios**: Useful when comparing data that involves ratios or relative changes, like population growth.

Mode's Merits and Demerits Recap

Merits of Mode:

- **Easy to Understand**: Simple to locate in a list or chart.
- **Not Affected by Extremes**: Outliers don't impact the mode.
- **Good for Qualitative Data**: Works well with categories like popularity, ratings, etc.

Demerits of Mode:

- **Sometimes Unclear**: Some data may have no mode or multiple modes.
- **Not Based on All Data**: Only considers the most frequent values, not all of them.

Uses of Mode:

- **For Quick Estimation**: Provides a fast measure of central tendency.
- **To Find the Most Typical Value**: Shows the most commonly occurring value in a dataset.

Harmonic Mean (H.M)

The **harmonic mean** is a type of average used mainly for quantitative data, especially when dealing with rates or ratios (like speeds). It's calculated by dividing the total number of values by the sum of the reciprocals (1/value) of each value. The harmonic mean is denoted by **H.M**.

Merits (Advantages) of Harmonic Mean

1. **Uses All Values**: Every observation in the data is part of the calculation.
2. **Stable with Sampling Changes**: Less sensitive to sample variations.
3. **Works with Math**: Can be used in further algebraic calculations.
4. **Ideal for Ratios and Rates**: Best suited for situations involving ratios (like speed or prices).
5. **Balances Large and Small Values**: Gives less importance to large values, focusing more on smaller values.

Demerits (Disadvantages) of Harmonic Mean

1. **Complex Calculation**: It's harder to compute than other averages.
2. **Emphasizes Small Values**: Gives more weight to smaller numbers, which may not always be ideal.
3. **Zero Values Issue**: Can't be calculated if any value in the data is zero.
4. **Not a Data Value**: The result might not match any actual value in the dataset.

Uses of Harmonic Mean

- **Averaging Rates**: Good for finding average speed, average price, or similar rates where values vary.

UNIT - 4
DISPERSION

Dispersion (Spread or Variation)

Dispersion refers to how much data values are spread out or vary from a central point (like the mean, median, or mode). It shows the "scatter" or variation within a dataset. Measures of dispersion are considered "second-order" averages, as they describe the spread rather than the central point itself.

Importance of Dispersion

1. **Checks Reliability of Average**: Shows how well the average represents the data.
2. **Compares Datasets**: Helps to compare the spread between different datasets.
3. **Aids in Control**: Useful in quality control or assessing consistency.
4. **Supports Other Statistical Measures**: Helps in calculations like standard deviation and variance.

Characteristics of a Good Measure of Dispersion

1. **Easy to Understand**: Should be simple to explain and grasp.
2. **Easy to Calculate**: Quick to compute.
3. **Defined Clearly**: No room for ambiguity.
4. **Includes All Data Points**: Considers all values in the dataset.
5. **Stable with Sampling**: Not overly sensitive to sample changes.
6. **Unaffected by Extremes**: Less influenced by very high or low values.
7. **Good for Further Calculations**: Works well in other mathematical treatments.

Types of Dispersion Measures

1. **Range (R)**: The difference between the highest and lowest values. Only considers the two extremes, ignoring all other values.
2. **Quartile Deviation (Q.D.)**: Measures the spread of the middle 50% of values. It's half the difference between the upper quartile (Q3) and lower quartile (Q1). Quartile deviation is sensitive to sample changes.
3. **Mean Deviation (M.D.)**: This is the average of the absolute deviations (distance from a central point) of each value from the mean, median, or mode. Positive and negative deviations are treated as positive. It's also called the "first absolute moment."
4. **Standard Deviation (S.D.)**: Shows how far data points typically are from the mean. It's the square root of the average of squared deviations from the mean and is denoted by **σ (sigma)**.

Distinction between mean deviation and standard deviation

- **Definition**:

  - **Mean Deviation (M.D.)**: The average of how far each value is from a central point (usually the mean, median, or mode). It uses absolute values, so both positive and negative deviations are considered positive.
  - **Standard Deviation (S.D.)**: Shows the average spread of values around the mean, but instead of absolute values, it squares each deviation, which emphasizes larger differences.

- **Formula and Calculation**:

  - **M.D.**: Calculated by finding the absolute deviation of each value from the central point, adding them up, and dividing by the total number of values.
  - **S.D.**: Calculated by squaring each deviation from the mean, finding the average of those squared deviations, and then taking the square root of that average.

- **Sensitivity to Outliers**:

  - **M.D.**: Less sensitive to extreme values because it doesn't square deviations.
  - **S.D.**: More sensitive to large deviations due to squaring; this emphasizes larger differences.

- **Mathematical Usefulness**:

  - **M.D.**: Simpler but not as widely used in further calculations or advanced analysis.
  - **S.D.**: Commonly used in statistical analysis, especially when dealing with data distributions, as it has strong algebraic properties.

- **Interpretation**:

  - **M.D.**: Gives an average "distance" from the central value, showing overall consistency.
  - **S.D.**: Provides a more precise measure of spread, showing how tightly values cluster around the mean.

Variance

1. **Variance (σ^2)**:
   - Variance shows how much data points in a set vary from the mean. It's calculated by taking each difference from the mean, squaring it (to make everything positive), and then averaging these squares.
   - The variance is symbolized by σ^2, and it's basically the "spread" or "scattered nature" of data.

2. **Standard Deviation (σ)**:
    - Standard deviation is just the square root of variance. It tells us how spread out the data is around the mean in the same units as the data itself.
    - A lower standard deviation means data is close to the mean (more consistent), while a higher standard deviation means data is more spread out (less consistent).
3. **Coefficient of Variation**:
    - This is a way to compare the variability of two different datasets, even if they have different units or means.
    - It's calculated by dividing the standard deviation by the mean and multiplying by 100 to make it a percentage. A smaller percentage indicates greater stability in data, while a larger percentage indicates more variability.

Different Dispersion Measures

1. **Range (R)**:
    - **Definition**: Difference between the highest and lowest values in a data set. Simple but only considers the extremes.
    - **Calculation**: R=Highest Value−Lowest ValueR = \text{Highest Value} - \text{Lowest Value}R=Highest Value−Lowest Value
2. **Quartile Deviation (Q.D.)**:
    - **Definition**: Measures the spread of the middle 50% of the data. Calculated by finding the difference between the third quartile (Q3) and the first quartile (Q1), then dividing by 2.
    - **Formula**: Q.D.=Q3−Q12Q.D. = \frac{Q3 - Q1}{2}Q.D.=2Q3−Q1
3. **Mean Deviation (M.D.)**:
    - **Definition**: The average of how far each value is from the central point (mean, median, or mode), with all distances treated as positive.
    - **Formula**: M.D.=∑|x−Average|NM.D. = \frac{\sum |x - \text{Average}|}{N}M.D.=N∑|x−Average| where NNN is the number of items.
4. **Standard Deviation (S.D.)**:
    - **Definition**: A more precise measure of spread that squares each deviation, averages these squares, and then takes the square root.
    - **Formula**: S.D.=∑(x−Mean)2NS.D. = \sqrt{\frac{\sum (x - \text{Mean})^2}{N}}S.D.=N∑(x−Mean)2

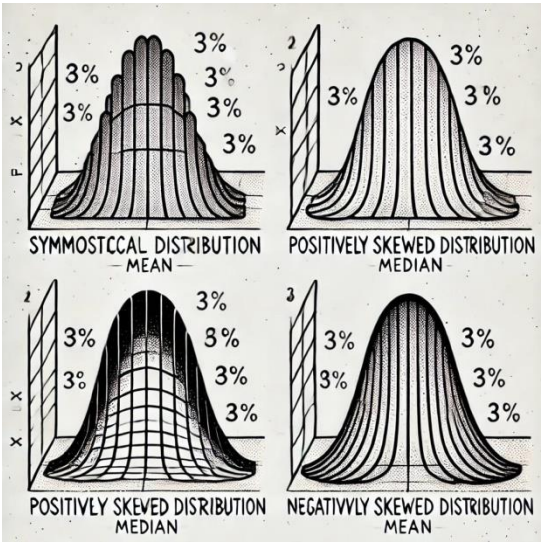
Skewness measures whether data is symmetrical or not. When data is perfectly symmetrical, it has no skewness, meaning it looks the same on both sides of the center point. When it's not symmetrical, it can either be **positively skewed** or **negatively skewed**.

1. **Positive Skewness**:
   - The tail on the right side of the distribution is longer, meaning there are a few extremely high values that pull the average up.
   - In positively skewed data, the **mean** and **median** are greater than the **mode**.
2. **Negative Skewness**:
   - The tail on the left side of the distribution is longer, meaning there are a few very low values that pull the average down.
   - In negatively skewed data, the **mean** and **median** are less than the **mode**.
3. **Zero Skewness**:
   - When data is evenly distributed around the mean, it has zero skewness, meaning the mean, median, and mode are all roughly the same.

Karl Pearson's Coefficients of Skewness

Pearson developed two methods to measure skewness:

1. **Pearson's Coefficient of Skewness #1**:
   - Formula: Skewness=Mean−ModeStandard Deviation\text{Skewness} = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}Skewness=Standard DeviationMean−Mode
   - Use when you know the **mode**. If the mean is far from the mode, skewness will be larger, showing an uneven distribution.
2. **Pearson's Coefficient of Skewness #2**:
   - Formula: Skewness=3×(Mean−Median)Standard Deviation\text{Skewness} = \frac{3 \times (\text{Mean} - \text{Median})}{\text{Standard Deviation}}Skewness=Standard Deviation3×(Mean−Median)
   - Use when you **don't know the mode**. Multiplying by 3 helps emphasize differences when using the median instead of the mode.



Time Series Analysis

A **Time Series** is a sequence of data points recorded over time, showing how something changes at specific time intervals, like every year, month, or day. This kind of data collection and analysis helps us understand trends, patterns, and behaviors of different phenomena over time. Time series data is widely used, not just by economists and business professionals, but also by scientists, sociologists, biologists, and researchers in various fields.

Examples of Time Series:

1. Tracking a country's population over the years.
2. Yearly coal production in India for the last decade.
3. Annual deposits collected by a bank.
4. Daily closing stock prices on the Bombay Stock Exchange.
5. Monthly sales of a store for the past six months.
6. Hourly temperature readings from a location over a period.

According to Patterson, "A time series consists of statistical data that is collected, recorded, or observed over regular intervals."

Importance of Time Series Analysis

Time series analysis is essential because it helps us predict and prepare for the future. Here's why it's valuable:

1. **Forecasting:** By analyzing past trends, we can make predictions, which is especially helpful for business planning. It reduces uncertainties by giving us a clearer picture of potential future outcomes.
2. **Evaluating Progress:** Time series helps in assessing current achievements by comparing them with past performance, allowing for adjustments to plans if needed.
3. **Understanding Past Patterns:** Analyzing past behaviors helps us understand what changes occurred, what factors influenced them, and the conditions under which these changes happened.
4. **Comparative Studies:** It enables comparisons over different times or places by examining and isolating various influences in a time series.
5. **Business Planning:** Separating and studying the parts of a time series is crucial for businesses, helping them to plan and make policy decisions.
6. **Setting Policies:** By analyzing economic sectors' past performances, businesses and governments can set strategies and policies to meet future needs.

Causes of Variation in Time Series

When observing time series data, certain factors often influence changes, including:

1. Changes in people's tastes, habits, and fashions.
2. Variations in customs and social norms.
3. Festivals, rituals, and cultural practices.
4. Political movements and government policies.
5. Disruptive events like wars, famines, droughts, floods, earthquakes, and epidemics.
6. Unusual weather or seasonal changes.

Components of Time Series

A **Time Series** refers to data points collected at different times, reflecting changes in certain variables like prices or population over a period. These changes are influenced by multiple factors and are broadly categorized into four main components.

1. Secular Trend (Long-Term Trend)

- This is a gradual, long-term movement in the data that shows a consistent upward or downward direction over time.
- For example, the population of a city or advancements in technology tend to increase gradually over the years.
- **Types of Secular Trends:**
  - **Linear Trends**: Show a straight line when plotted on a graph, indicating a steady rate of growth or decline.
  - **Non-Linear Trends**: These follow a curve, where growth may start slow, speed up, stabilize, and then slow down, which is common in economics and business.

2. Seasonal Variation

- Seasonal variations are short-term, recurring patterns that occur within a year, often linked to seasons, weather, holidays, or cultural events.
- **Examples:**
  - Sales of woolens increase in winter.
  - Umbrella and raincoat sales go up in the rainy season.
  - Demand for air conditioners and refrigerators rises in summer.
  - Sales peak during festivals like Christmas, Diwali, or Eid due to increased shopping.

3. Cyclical Variation

- These are repetitive changes in the data over longer periods, often lasting more than a year, associated with economic cycles.
- **Examples**: Business cycles with phases of growth, decline, depression, and recovery.
- Cyclical variations are less regular than seasonal changes but still follow a pattern where growth and decline alternate over time.

4. Irregular or Random Variation

- These are unpredictable changes that occur due to unforeseen events, like natural disasters, political upheaval, or sudden changes in consumer trends.

PHASES OF A BUSINESS CYCLE

Phases of a Business Cycle and Types of Time Series Analysis Models

Irregular or Random Variation

- **Definition**: These are unexpected, unpredictable changes in a time series due to rare or unusual events.
- **Examples**: Earthquakes, floods, droughts, epidemics, strikes, lock-outs, or wars. These variations don't follow a pattern and are difficult to predict.

Mathematical Models for Time Series Analysis

There are two main models used to break down time series data into its components: trend, seasonal, cyclical, and random variations.

1. Additive Model

- **Formula**: U=T+S+C+RU = T + S + C + RU=T+S+C+R
- **Explanation**: This model assumes that each component affects the overall time series separately and can simply be added together. Here:
  - **U** is the observed data value,
  - **T** stands for the trend,
  - **S** for seasonal variation,
  - **C** for cyclical variation, and
  - **R** for random variation.
- **Example**: If each component has its own independent effect (like separate causes), then their combined effect can be found by adding them up.

2. Multiplicative Model

- **Formula**: U=T×S×C×RU = T \times S \times C \times RU=T×S×C×R
- **Explanation**: This model assumes that the components are interdependent, meaning that each component influences the others. Here, the components are multiplied together to represent their combined effect on the time series.
- **Example**: When sales are influenced by both seasonal effects and overall growth (trend), the changes might compound each other, making multiplication more realistic.

Measurement of Trend (Secular Trend) in Time Series

To analyze the trend component in a time series, various methods are used to identify long-term directions or movements. Here are some of the most common methods:

1. Moving Average Method

- **Definition**: The moving average method smooths out fluctuations by calculating the average of successive groups of values over a specified period (e.g., 3-year, 4-year moving averages).
- **Procedure**:
  - Calculate averages of consecutive data points over a chosen period (e.g., 3 years, 5 years).
  - Center these averages by placing them between the middle years of the time intervals.
- **Example**:
  - For a 3-year moving average, add the values of three consecutive years and divide by 3. Continue this across the time series.

2. Least Squares Method

- **Definition**: This method fits a trend line to the time series data by minimizing the sum of squared deviations between observed values and predicted values, resulting in a "best-fit" line.
- **Equation of Trend Line**:
  - The trend line can be represented as y=a+bxy = a + bxy=a+bx, where:
    - yyy: Estimated value of the trend,
    - xxx: Time period deviation from an origin point,
    - aaa: Intercept (initial value of the trend line),
    - bbb: Slope (indicating rate of change).
  - If b>0b > 0b>0, the trend shows growth; if b<0b < 0b<0, it indicates a decline.

- **Normal Equations for Calculation**:

  ∑Y=na+b∑X\sum Y = na + b \sum X∑Y=na+b∑X ∑XY=a∑X+b∑X2\sum XY = a \sum X + b \sum X^2∑XY=a∑X+b∑X2

  where:

  - nnn: Total number of time points,
  - ∑Y\sum Y∑Y: Sum of actual values,
  - ∑X\sum X∑X: Sum of deviations from origin,
  - ∑XY\sum XY∑XY: Sum of product of deviations and actual values,
  - ∑X2\sum X^2∑X2: Sum of squared deviations.

Advantages of the Least Squares Method

1. Provides accurate future estimates for trend projections.
2. Useful in forecasting future values based on historical trends.
3. Minimizes personal bias in data interpretation.
4. Offers the rate of growth or decline in values over time.

Limitations of the Least Squares Method

1. Requires mathematical computations, which can be complex.
2. Lacks flexibility—adding new data requires recalculating the trend line.
3. Primarily reflects long-term trends, ignoring short-term variations like seasonal or cyclical changes.

UNIT-5
CORRELATION

Introduction

1. Correlation is a statistical tool & it enables us to measure and analyse the degree or extent to which two or more variable fluctuate/vary/change w.e.t. to each other.

2. For example – Demand is affected by price and price in turn is also affected by demand. Therefore we can say that demand and price are affected by each other & hence are correlated. the other example of correlated variable are –

3. While studying correlation between 2 variables use should make clear that there must be cause and effect relationship between these variables. for e.g. – when price of a certain commodity is changed (↑ or ↑) its demand also changed (↑ or ↑) so there is case & effect relationship between demand and price thus correlation exists between them. Take another eg. where height of students; as well as height of tree increases, then one cannot call it a case of correlation because neither height of students is affected by height of three nor height of tree is affected by height of students, so there is no cause & effect relationship between these 2 so no correlation exists between these 2 variables.

4. In correlation both the variables may be mutually influencing each other so neither can be designated as cause and the other effect for e.g. –
Price ↑→ Demand ↓
Demand ↓→ Price ↑
So, both price & demand are affected by each other therefore use cannot tell in real sense which one is cause and which one is cause and which one is effect.

DEFINITIONS OF CORRELATION

1. "If 2 or more quantities vary is sympathy, so that movements is one tend to be accompanied by corresponding movements in the other(s), then they are said to be correlated". **Connor.**

2. "Correlation means that between 2 series or groups of data there exists some casual correction".

WI King

3. "Analysis of Correlation between 2 or more variables is usually called correlation."

A.M. Turtle

4. "Correlation analysis attempts to determine the degree of relationship between variables. **YaLunchou**

Types of Correlation

Correlation measures the relationship between two or more variables. Here are the main types:

1. Positive and Negative Correlation

- **Positive Correlation**: When two variables move in the same direction. For example, if the price per unit increases, the quantity supplied also increases.
  - **Example**: Supply and Price
- **Negative Correlation**: When two variables move in opposite directions. For example, if the price of a product increases, the demand decreases.
  - **Example**: Demand and Price

2. Simple and Multiple Correlation

- **Simple Correlation**: Examines the relationship between two variables only.
  - **Example**: Examining the relationship between demand and price alone.
- **Multiple Correlation**: Studies the relationship between three or more variables, where two or more factors affect one variable.
  - **Example**: Examining how demand is influenced by both price and income.

3. Partial and Total Correlation

- **Partial Correlation**: Looks at the relationship between two variables while keeping other influencing variables constant.
  - **Example**: Studying the effect of price on demand while keeping income constant.
- **Total Correlation**: Examines the relationship between all variables, without holding any variables constant.
  - **Example**: Studying the effect of both price and income on demand without assuming any constancy.

4. Linear and Non-Linear Correlation

- **Linear Correlation**: When a change in one variable results in a consistent change in the other. The relationship graph is a straight line.
- • **Example**: Doubling the number of workers doubles production output, showing linear correlation.

• **Non-Linear (Curvilinear) Correlation**: When the change in one variable does not result in a consistent change in the other, producing a curve on the graph.

- **Example**: Spending more on advertising doesn't always result in a proportional increase in sales.

Karl Pearson's Coefficient of Correlation:

This measures the relationship (correlation) between two variables (usually labeled as x and y). It shows if an increase or decrease in one variable results in a similar change in the other.

Key Terms:

- **dx**: Deviation of each x value from the mean of x values.
- **dy**: Deviation of each y value from the mean of y values.
- $\bar{x}$: Mean of x values.
- $\bar{y}$: Mean of y values.
- $\mathbf{d^2x}$ and $\mathbf{d^2y}$: Squared deviations of x and y values.
- **dxdy**: Product of deviations.
- **Variance (σx^2, σy^2)**: Average of squared deviations for x and y.
- **r**: The correlation coefficient between x and y, showing how strong the relationship is.

Methods to Find Correlation

1. **Direct Method**: Use actual means for calculations.
2. **Assumed Mean Method**: Used when means are not whole numbers to simplify calculations.

Regression Analysis:

A tool for predicting the relationship between dependent (predicted) and independent (predicting) variables.

Key Concepts:

- **Dependent Variable (Y)**: The variable we predict.
- **Independent Variable (X)**: The variable we use to predict Y.
- **Regression Lines**: Best-fit lines that represent average relationships between X and Y. We have two:
  - **X on Y**: Predicts X from Y.
  - **Y on X**: Predicts Y from X.

Differences between Correlation and Regression:

Correlation	Regression
Measures the relationship strength between variables.	Predicts values of one variable based on another.
No cause-effect implication.	Indicates cause-effect relationship.
Cannot predict specific values.	Can predict specific values.

Regression Coefficients:

These coefficients indicate the rate of change between variables.

- **bxy and byx**: Regression coefficients. They have the same sign and are limited in range, ensuring consistency.

Index Numbers:

Index numbers track changes in a phenomenon (like prices, production) over time or between locations.

Characteristics of Index Numbers:

1. **Relative Change**: Measures only changes relative to a base point.
2. **Comparability**: Allows comparing different phenomena.

Methods to Construct Index Numbers:

1. **Unweighted Methods**:
   - Simple Average or Aggregative Method.
2. **Weighted Methods**:
   - **Laspeyres, Paasche, and Fisher's Ideal Index**: Common weighted index formulas that provide more accurate reflections of changes by assigning importance to each item.