MAA OMWATI DEGREE COLLEGE HASSANPUR

NOTES

SUBJECT:- BUSINESS STATISTICS (MC)

NAME OF PROGRAM-B.COM 1ST SEM

COURSE CODE-24COM401DS02

UNIT-1

Statistics: Meaning, Definition, Needs & Objectives Collection of data types,methods, classification and tabulation of data, graphic diagrammatic presentation.

Unit 2

Measurement of Central Tendency and Variation Mathematical and fractional averages. Measures of absolute and relative variations.

Unit-3

Moments, skewness and kurtosis (with Sheppard's corrections), Index Numbers.

Unit 4

Probability and Expected Value: Meaning and Schools of thoughts, Importance of the Concept of the Probability; Calculation of Probability, Probability Theorems: Addition, Multiplication and Bayes' Theorem. Mathematical Expectations. Numerical of Probability.

UNIT-1

INTRODUCTION

Business Statistics is the application of statistical techniques to business decision-making. It involves collecting, analyzing, and interpreting data to inform business strategies and optimize operations.

ORIGIN AND DEVELOPMENTS OF STATISTICS

1 Ancient Mesopotamia (3500 BCE): Recorded census data, agricultural production, and trade transactions.
2Ancient Egypt (2500 BCE): Collected data on population, land ownership, and taxes.
3Ancient Greece (500 BCE): Philosophers like Aristotle and Plato discussed concepts like "mean" and "median".
4 Roman Empire (100 BCE): Conducted censuses and collected data on population, agriculture, and military personnel.
5Arab Empire (800 CE): Developed advanced mathematical and statistical techniques, including algebra and probability.
6 Europe (1200 CE): Governments began collecting data on population, trade, and taxes.
7 17th Century: Statistical analysis emerged as a distinct field, with pioneers like William Petty and John.
818th Century: Probability theory developed, with contributions from Pierre-Simon Laplace and Thomas Bayes.
9 19th Century: Statistical methods expanded, with the introduction of regression analysis and hypothesis testing.
10 20th Century: Statistical computing and software emerged, making data analysis more accessible.

In statistics, the terms "plural" and "singular" have specific definitions:

Plural:
In plural sense, the term statistics is used as numerical information. All statistics are numerical statements of facts but all facts numerically stated are not statistics.
Sample: A collection of individual data points (e.g., survey responses)
Population: The entire group of individuals or cases (e.g., all customers)

Singular:

The statistical information gathered must be properly organized, properly presented, Analysis and interpreted to achieve the desired objectives. Thus we require statistical methods.

Variable: A single characteristic or attribute (e.g., age)

In statistics, we often work with plural data (multiple observations) to make inferences about a population. However, we may also analyze singular data points to understand individual cases or outliers.

FUNCTION /DISADVANTAGES OF STATISTICS

Advantages of Statistics:

1. Data-driven decision-making: Statistics enables informed decisions by providing insights from data analysis.
2. Identification of trends and patterns: Statistical methods help recognize trends, patterns, and correlations within data.
3. Quantification of uncertainty: Statistics allows for the measurement of uncertainty and risk, facilitating better planning and strategy development.
4. Improved understanding of phenomena: Statistical analysis helps comprehend complex relationships and phenomena.
5. Communication of complex data: Statistical methods enable effective communication of intricate data insights to non-technical audiences.
6. Informed policy-making: Statistics informs policy decisions by providing data-driven evidence.
7. Optimization of processes: Statistical analysis improves processes and optimizes resources.
8. Prediction and forecasting: Statistical models enable prediction and forecasting of future events or trends.

Disadvantages of Statistics:

1. Data quality issues: Statistical analysis is only as good as the data quality.
2. Misinterpretation: Statistical results can be misinterpreted or misunderstood.
3. Overreliance on data: Statistics might overlook non-quantifiable factors.
4. Complexity: Statistical methods can be complex and difficult to understand.
5. Time-consuming: Data collection and analysis can be time-consuming.
6. Limited generalizability: Statistical findings may not be generalizable to all situations.
7. Ethical concerns: Statistical analysis can raise ethical concerns, such as data privacy.
8. Dependence on assumptions: Statistical models rely on assumptions that may not always hold true.

SCOPE OF STATISTICS

1. Social Sciences: Sociology, Psychology, Economics, Political Science
2. Health Sciences: Medicine, Public Health, Epidemiology, Nursing

3. Business and Finance: Marketing, Operations, Finance, Management
4. Natural Sciences: Biology, Physics, Chemistry, Environmental Science
5. Engineering: Quality Control, Reliability Engineering, Operations Research
6. Government: Policy-making, Public Administration, Census, Surveys
7. Education: Research, Assessment, Educational Psychology
8. Marketing: Market Research, Consumer Behavior, Advertising
9. Sports: Performance Analysis, Player Tracking, Game Strategy
10. Environmental Studies: Climate Change, Ecology, Conservation Biology
11. Computer Science: Data Mining, Machine Learning, Artificial Intelligence
12. Medicine and Healthcare: Clinical Trials, Public Health, Health Economics
13. Social Media: Data Analysis, User Behavior, Network Analysis
14. Quality Control: Industrial Statistics, Six Sigma, Quality Engineering
15. Economics: Econometrics, Time **Series Analysis, Forecasting**

Statistics is used in:

Data analysis and interpretation
Hypothesis testing and confidence intervals
Regression analysis and modeling
Time series analysis and forecasting
Survey research and sampling
Experimental design and optimization- Quality control and reliability engineering
Data visualization and communication

The scope of statistics is constantly expanding as data becomes increasingly important in various fields.

TYPES OF SCALES

In statistics, a scale refers to a way of measuring or categorizing data. There are four main types of scales:

1. Nominal Scale:

Nominal scale is used for measuring items of nominal categories , rocks of different sizes. The central tendency of a nominal attribute can be measured by using mode.

- Categorizes data without any numerical value or order
- Examples: Gender (Male/Female), Color (Red, Blue, Green)

2. Ordinal Scale:

In this type of scale, numbers assigned to objects or events are ranked in order of 1^{st}, 2^{nd}, 3^{rd} and so on. The central tendency of an ordinal attribute can be represented by median.

- Categorizes data with a natural order or ranking

- Examples: Education Level (High School, Bachelor's, Master's), Satisfaction (Low, Medium, High)

3. Interval Scale:

All thee quantitative attributes are measurable in interval scales. These can be measured as positive or negative numerals. Thus, the variables measured at interval level called interval variables or scaled variables as they have units of measurement. The central tendency of a variable measured at interval can be represented by mode and median or by its arithmetic mean and dispersion by range, inter-quartile range by standard deviation.

- Measures data with equal intervals between consecutive levels
- Examples: Temperature (Celsius or Fahrenheit), IQ Score

4. Ratio Scale:

Most of the measurements of physical types and engineering are done by ratio scale. The central tendency of ratio level can be represented by geometric mean, mean and harmonic mean besides by using mode, median and mean and its dispersion by coefficient of variation.

- Measures data with a true zero point and equal intervals
- Examples: Weight, Height, Age

Additionally, there are also:

1. Dichotomous Scale:

- Consists of only two categories (Yes/No, 0/1)

2. Continuous Scale:

- Measures data that can take any value within a certain range or interval (e.g., time, temperature)

3. Discrete Scale:

- Measures data that can only take specific, distinct values (e.g., number of students in a class)

PRIMARY DATA AND ITS TYPES

Primary data is original, raw data collected directly from the source, typically through experiments, surveys, observations, or other methods. There are several types of primary data: Primary data can be collected from various sources, including:

1. Surveys:

Online surveys	Paper-based surveys	Telephone	person surveys

2. Experiments:

Laboratory experiments	Field experiment		Clinical trials

3. Observations:

Participant observation   Non-participant observation   Case studies

4. Interviews:

Structured interviews   Semi-structured interviews   Unstructured interviews

5. Focus Groups:

Online focus groups   Offline focus groups

6. Sensory Data:

Sensor-enabled devices   IOT devices   Wearable technology

7. Social Media:
   - Social media platforms
   - Online forums
   - Blog comments
8. Customer Feedback:
   - Customer reviews
   - Complaints
   - Suggestions
9. Employee Data:
   - Employee surveys
   - Performance data
   - Training data
10. Secondary Research:
   - Academic research papers
   - Industry reports
   - Government statistics

SECONDARY DATA

Secondary data is pre-existing data that has already been collected by others, such as researchers, organizations, or government agencies. It is often used to:

1. Save time and resources
2. Access large datasets
3. Benefit from others' expertise
4. Identify patterns and trends
5. Inform research questions

Sources of secondary data:

1. Academic journals and papers

2. Government reports and statistics
3. Industry publications and reports
4. Online databases and archives
5. Books and book chapters
6. Conference proceedings
7. Theses and dissertations
8. Research institutes and think tanks
9. Data archives and repositories
10. Online data platforms and libraries

Types of secondary data:
1. Quantitative data:
   - Statistical data
   - Survey data
   - Experimental data
2. Qualitative data:
   - Text data
   - Interview transcripts
   - Observational data
3. Mixed-methods data:
   - Combines quantitative and qualitative data

Advantages of secondary data:
1. Time-efficient
2. Cost-effective
3. Access to large datasets
4. Allows for meta-analysis
5. Inform research questions

Limitations of secondary data:
1. Data quality issues
2. Limited control over data collection
3. May not align with research objectives

4. Potential biases
5. Requires careful evaluation and validation

1. Evaluate data quality and reliability
2. Understand the data collection methodology
3. Consider potential biases and limitations
4. Ensure data aligns with research objectives
5. Properly cite and credit original sources

DIAGRAM AND TABLE

There are many types of diagrams and tables used in statistics to visualize and summarize data. Here are a few examples:

Diagrams:

1. Bar chart: A graph that uses bars to represent categorical data.
2. Histogram: A graph that uses bars to represent continuous data.
3. Scatter plot: A graph that uses points to represent the relationship between two variables.
4. Box plot: A graph that uses boxes and whiskers to represent the distribution of data.
5. Pie chart: A circular graph that represents categorical data.

Tables:

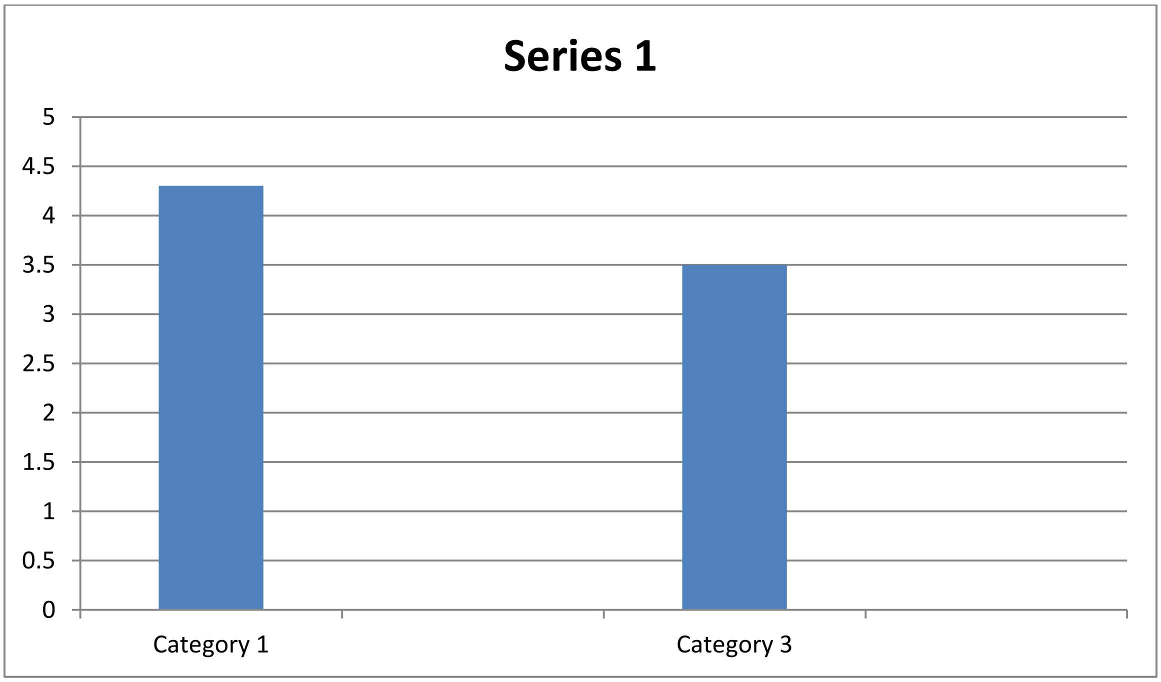
1. Frequency table: A table that displays the frequency of each category in a dataset.
2. Contingency table: A table that displays the relationship between two categorical variables.

3. Descriptive statistics table: A table that summarizes the central tendency and variability of a dataset.
4. Correlation table: A table that displays the correlation coefficients between variables.
5. Regression table: A table that displays the results of a regression analysis.

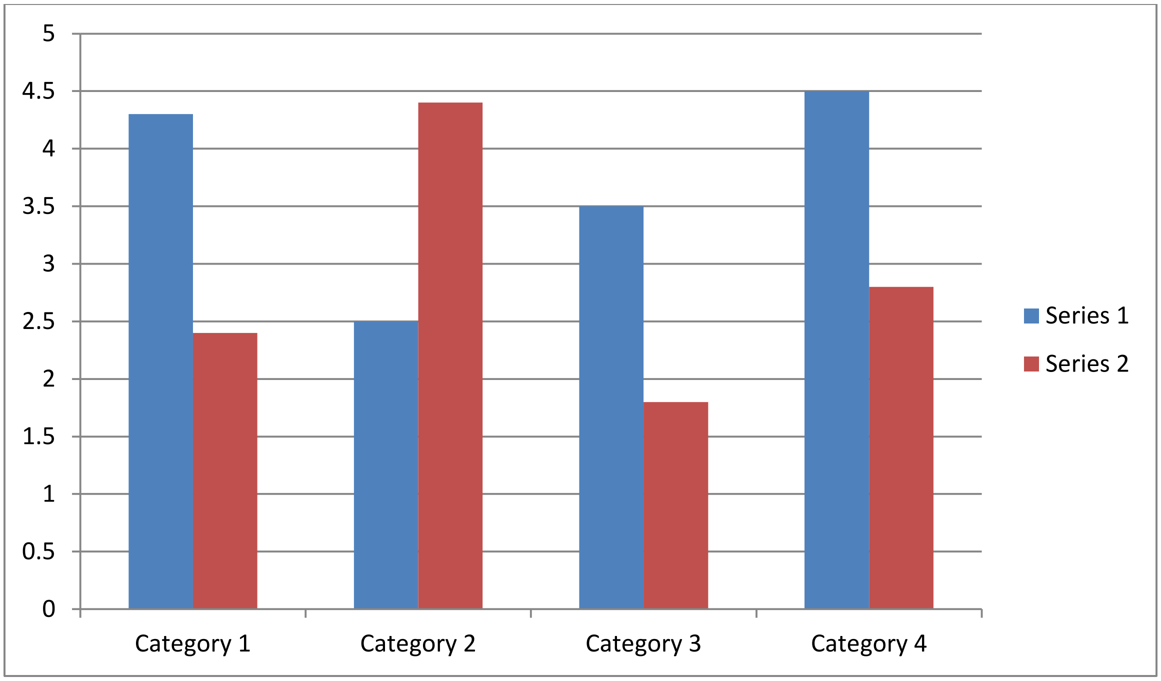
Some other types of diagrams and tables include:
- Flowcharts
- Tree diagrams
- Venn diagrams
- Pareto charts
- Q-Q plots
- Survival curves
- Heat maps
- Pivot tables

These visualizations help to:
- Summarize large datasets
- Identify patterns and trends
- Communicate complex data insights
- Facilitate data analysis and interpretation

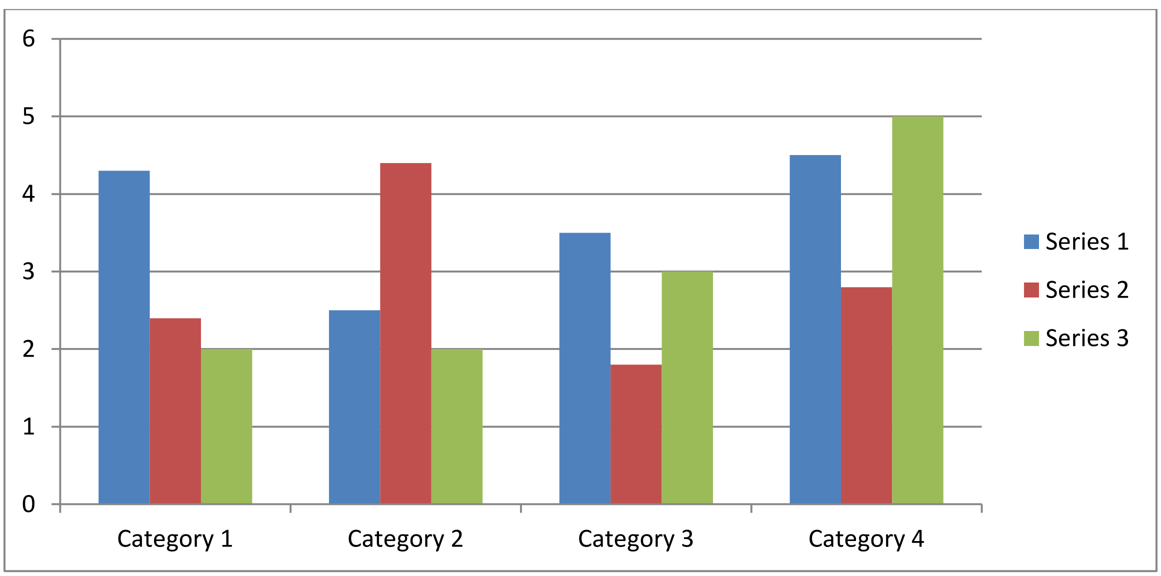
1 SIMPLE BAR DIAGRAM



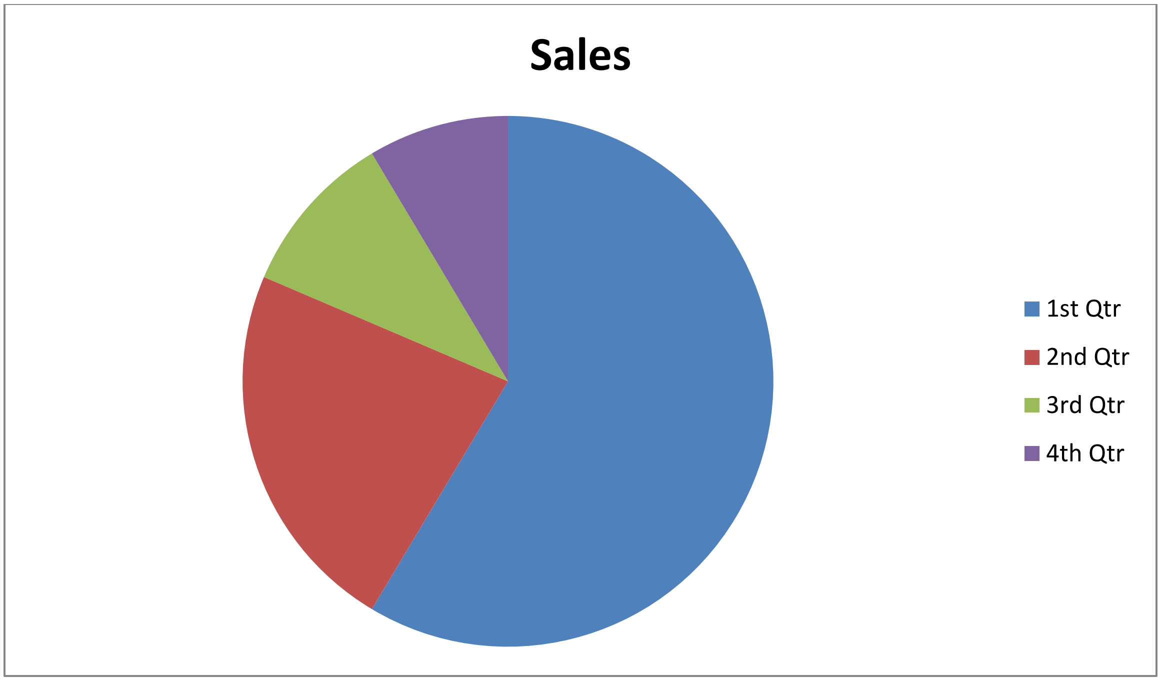
2 DOUBLE BAR DIAGRAMN



3 MULTIPLE BAR DIAGRAM



4 PIE DIAGRAM



UNIT-2

MEASURES OF CENTRAL TENDENCY

"An average is a typical value in the sense that is sometimes employed to represent all the individual values in a series or of a variable"

Measures of central tendency are statistical tools used to describe the middle or center of a distribution. They provide a single value that represents the most typical or expected value in the dataset. The three main measures of central tendency are:

केंद्रीय प्रवृत्ति के माप से अभिप्राय है कि यह सभी मूल्यों के समूह का प्रतिनिधित्व करता है और अन्य सभी मूल्य इसके इर्द-गिर्द होते हैं। यह हमें डेटा सेट के मध्य बिंदु की जानकारी देता है, जो हमें डेटा की विशेषताओं को समझने में मदद करता है।

1. Mean (Arithmetic Mean): The sum of all values divided by the number of values.

2. Median: The middle value when the data is arranged in order (50th percentile).

3. Mode: The most frequently occurring value in the dataset.

These measures provide a way to summarize and understand the central tendency of a dataset, which is essential in statistics and data analysis.

केंद्रीय प्रवृत्ति के माप से हमें पता चलता है कि डेटा सेट के मूल्य कैसे वितरित हैं, और यह हमें डेटा के मध्य मान को समझने में मदद करता है। यह हमें डेटा के विश्लेषण और व्याख्या में भी मदद करता है, जिससे हमें डेटा से संबंधित निर्णय लेने में मदद मिलती है।

Introduction to Measures of Central Tendency:

Measures of central tendency are used to describe the center of a distribution, which is a fundamental concept in statistics and data analysis. They provide a way to summarize and understand the typical value in a dataset, which is essential for making informed decisions. The three main measures of central tendency are the mean, median, and mode, each with its own strengths and weaknesses.

Understanding the measures of central tendency is crucial in various fields, such as:

Business: to analyze customer behavior, sales trends, and financial data

Healthcare: to understand patient outcomes, disease patterns, and treatment efficacy

Social Sciences: to study social phenomena, opinion polls, and demographic data

Engineering: to optimize system performance, quality control, and reliability analysis

In this tutorial, we will explore the definitions, calculations, and applications of the mean, median, and mode, as well as their strengths and limitations. We will also learn how to choose the appropriate measure of central tendency for a given dataset and how to interpret the results.

PURPOSE AND FUNCTIONS OF STATISTICAL AVERAGE

The statistical average, also known as the mean, is a measure of central tendency that represents the sum of all values in a dataset divided by the number of values. The purpose and functions of statistical average are:

Purpose:

1. The summarize and describe the central tendency of a dataset.

2. To provide a representative value for a dataset.

3. To facilitate comparison between different datasets.

Functions:

1. Descriptive statistics: The average is used to describe the central tendency of a dataset.

2. Inferential statistics: The average is used to make inferences about a population based on a sample.

3. Data analysis: The average is used to analyze and understand the distribution of data.

4. Decision making: The average is used to make informed decisions in various fields such as business, economics, and healthcare.

5. Comparison: The average is used to compare different datasets or groups.
6. Forecasting: The average is used to forecast future trends and patterns.
7. Quality control: The average is used to monitor and control the quality of products or services.
8. Research: The average is used to summarize and analyze data in research studies.

CHARACTERISTICS/PROPERTIES OF GOOD AVERAGE

1. Representativeness: The average should accurately represent the central tendency of the dataset.
2. Unbiasedness: The average should be free from any bias or distortion.
3. Consistency: The average should be consistent in its calculation and application.
4. Reliability: The average should be reliable and stable across different samples and datasets.
5. Sensitivity: The average should be sensitive to changes in the data.
6. Robustness: The average should be robust to outliers and extreme values.
7. Easy to calculate: The average should be easy to calculate and understand.
8. Easy to interpret: The average should be easy to interpret and understand.
9. Relevant: The average should be relevant to the problem or question being addressed.
10. Efficient: The average should be efficient in terms of computational **resources.**

Some other important properties of a good average include:

Linearity: The average should be linear in its calculation.
Homogeneity: The average should be homogeneous in its application.
Scalability: The average should be scalable to large datasets.

TYPES OF AVERAGE

MATHEMATICAL AVERAGE
1 ARITHMATIC MEAN($\bar{x}$)
2 GEOMETRIC MEAN
3 HARMONIC MEAN

POSITIONAL AVERAGE
1 MEDIAN(M)
2 MODE(Z)

1 ARITHMATIC MEAN

The arithmetic mean, also known as the average, is a measure of central tendency that represents the sum of all values in a dataset divided by the number of values.
Arithmetic Mean = (Sum of all values) / (Number of values)
For example, if we have the numbers 2, 4, 6, 8, 10, the arithmetic mean would be:

Arithmetic Mean = (2 + 4 + 6 + 8 + 10) / 5 = 30 / 5 = 6
So, the arithmetic mean is 6.

(A)SIMPLE ARITHMATIC MEAN

INDIVIDUAL SERIES

1 DIRECT METHOD

The simple arithmetic mean formula is:

AM = (Σx) / n Where: - AM is the arithmetic mean
- Σx is the sum of all values - n is the number of values

For example, if we have the numbers 1, 2, 3, 4, 5, the formula would be:
AM = (1 + 2 + 3 + 4 + 5) / 5
AM = 15 / 5
AM = 3
So, the arithmetic mean is 3.

2 SHORTCUT METHOD

$$(\bar{x}) = A + \frac{\Sigma d}{N}$$

(B) DISCRETE METHOD

1 DIRECT METHOD

$$(\bar{x}) = \frac{\Sigma fx}{N}$$

2 shortcut method

$$(\bar{x}) = A + \frac{\Sigma fd}{\quad}$$

$$N$$

(c)continuous series

1 DIRECT METHOD

$$(\bar{x}) = \frac{\Sigma fm}{N}$$

2 shortcut method

$$(\bar{x}) = A + \frac{\Sigma fd}{N}$$

3 step DEVIATION METHOD

$$(\bar{x}) = A + \frac{\Sigma fd' \ (i)}{N}$$

Combined ARITHMETIC MEAN

The combined arithmetic mean, also known as the combined mean, is a way to calculate the average of two or more sets of data. It's a useful technique when you have multiple groups of data and you want to find the overall average.

The formula for combined arithmetic mean is:

Combined Arithmetic Mean = (n1x1 + n2x2 + ... + nmxm) / (n1 + n2 + ... + nm)

Where:- n1, n2, ..., nm are the sizes of each data set - x1, x2, ..., xm are the means of each data set

For example, let's say you have two sets of exam scores:

Set 1: 70, 80, 90 (mean = 80) Set 2: 60, 70, 80 (mean = 70)

To calculate the combined arithmetic mean, you would:

1. Multiply the mean of each set by its size: 80(3) + 70(3) = 240 + 210 = 450
2. Add up the sizes of the sets: 3 + 3 = 6
3. Divide the sum by the total size: 450 ÷ 6 = 75

So, the combined arithmetic mean is 75.

Combined Arithmetic Mean

$$\bar{X}_{12} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}$$

$$\bar{X}_{123} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2 + N_3 \bar{X}_3}{N_1 + N_2 + N_3}$$

Weighted Arithmetic Mean

$$\bar{X}_w = \frac{\Sigma wX}{\Sigma w}$$

$\bar{X}_w$ → Weighted Arithmetic Mean
Σwx → weight x value
Σw → Sum of weight

2. Median

Individual series

M = Size of $\left(\frac{N+1}{2}\right)$th item

Discreate Series

M = Size of $\left(\frac{N+1}{2}\right)$th item

Continuous Series

$\left(\frac{N}{2}\right)$ th item

$$M = L_1 + \frac{\frac{N}{2} - c.f}{f} \times i$$

Partition

There are other measures besides median which divide the data into equal parts. The important measures among them are quartiles, deciles and percentiles.

value is a concept in business statistics that refers to the process of dividing a dataset or a population into smaller, distinct groups or partitions, based on certain characteristics or criteria. This is also known as segmentation or clustering.

In business statistics, partition value is used to:

1. Identify patterns and relationships: By partitioning data, businesses can identify patterns and relationships that may not be apparent in the overall dataset.
2. Target specific groups: Partitioning allows businesses to target specific groups with tailored marketing strategies, products, or services.
3. Analyze and compare: Partitioning enables businesses to analyze and compare the characteristics and behaviors of different groups.

Common partitioning methods in business statistics include:

1. Demographic partitioning (age, gender, income.)
2. Geographic partitioning (region, country, city.)
3. Behavioral partitioning (purchase history.)

By applying partition value, businesses can gain a deeper understanding of their customers, optimize resources, and drive more effective decision-making.

3. Partition value

Quartiles

Deciles

Percentiles

Individual/ Discrete	Continuous	Formula continuous series
$Q_1 = \frac{N+1}{4}$	$Q_1 = \frac{N}{4}$	$Q_1 = L_1 + \frac{\frac{N}{4} - c.f}{f} \times i$
$Q_3 = \frac{3(N+1)}{4}$	$Q_3 = \frac{3N}{4}$	$Q_3 = L_1 + \frac{\frac{3N}{4} - c.f}{f} \times i$
$D_1 = \frac{N+1}{10}$	$D_1 = \frac{N}{10}$	$D_1 = L_1 + \frac{\frac{N}{10} - c.f}{f} \times i$
$D_9 = \frac{9(N+1)}{10}$	$D_9 = \frac{9N}{10}$	$D_9 = L_1 + \frac{\frac{9N}{10} - c.f}{f} \times i$
$P_1 = \frac{N+1}{100}$	$P_1 = \frac{N}{100}$	$P_1 = L_1 + \frac{\frac{N}{100} - c.f}{f} \times i$
$P_{99} = \frac{99(N+1)}{100}$	$P_{99} = \frac{99N}{100}$	$P_{99} = L_1 + \frac{\frac{99N}{100} - cf}{f} \times i$

4 MODE

Mode is a fundamental concept in business statistics that refers to the most frequently occurring value or category in a dataset or distribution. It's a measure of central tendency that helps businesses understand the typical or most common value in a dataset.

In business statistics, mode is used to:

1. Identify popular products or services: By analyzing sales data, businesses can identify the most popular products or services, which can inform inventory management, marketing strategies, and product development.
2. Understand customer behavior: Mode can help businesses understand the most common customer behaviors, preferences, or demographics, which can inform customer segmentation and targeting strategies.
3. Analyze quality control: In quality control, mode can help identify the most common defects or issues, which can inform process improvements and quality assurance initiatives.

Types of mode:

1. Unimodal: One mode in the dataset (e.g., most common product)
2. Bimodal: Two modes in the dataset (e.g., two equally popular products)
3. Multimodal: More than two modes in the dataset (e.g., multiple popular products)

Notable aspects of mode:

1. A dataset can have multiple modes (bimodal, multimodal)
2. Mode may not always be the same as the mean (average) or median (middle value)
3. Mode is sensitive to the dataset's distribution and sample size

By understanding mode, businesses can gain insights into customer preferences, optimize operations, and drive informed decision-making.

4. Mode (Z)

Individual series

(I) Inspection method

Item in series mostly repeated

5, 10, 15, 5, 10, 5, 15, 20, 5

Z = 5 is mode

(II) By changing individual to discrete

Frequency is mostly repeated.

Discrete Series

(I) Inspection

Frequancy repeated is mode

(II) Grouping method

(i) Grouping Table

(ii) Analytical Table

Continuous series

$$Z = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

OR

$$Z = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

RELATIONSHIP BETWEEN ARITHMATIC MEAN , MODE AND MEDIAN.

Relationship between $\bar{X}$, m and Z

$$\bar{X} - Z = 3(\bar{X} - M)$$

(a) $Z = 3M - 2\bar{X}$

(b) $M = \frac{1}{3}(2\bar{X} - Z)$

(c) $\bar{X} = \frac{1}{2}(3M - Z)$

5 GEOMETRIC MEAN

The geometric mean It's a powerful tool in business statistics, and I'm happy to share its importance and applications

Importance:

1. Average growth rate: Geometric mean is used to calculate the average growth rate of a dataset, which is essential for understanding compound interest, population growth, and investment returns.
2. Non-linear data: Geometric mean is more suitable for non-linear data, such as exponential growth or decay, where the arithmetic mean would be misleading.
3. Ratio data: Geometric mean is used with ratio data, where the relative change is more important than the absolute change.

Applications:

1. Finance: Calculate average returns on investments, portfolio performance, and compound interest.
2. Economics: Analyze economic growth rates, inflation, and population growth.
3. Biology: Study population growth, disease spread, and chemical reactions.
4. Engineering: Design and optimize systems, such as electronic circuits and mechanical systems.

5. Data analysis: Use geometric mean to summarize and compare non-linear data, like website engagement metrics or customer satisfaction scores.
Some real-world examples:
1. Calculating the average annual return on a stock portfolio
2. Analyzing the growth rate of a startup's user base
3. Determining the average rate of inflation over a period
4. Optimizing the performance of a chemical reaction
5. Evaluating the effectiveness of a marketing campaign
In summary, the geometric mean is a vital tool for analyzing and understanding non-linear data, growth rates, and ratio data in various fields. Its applications are diverse, and it helps drive informed decision-making in business, finance, economics, and more.

Relationship between $\bar{X}$, m and z

$$\bar{X} - Z = 3(\bar{X} - M)$$

(a) $Z = 3M - 2\bar{X}$

(b) $m = \frac{1}{3}(2\bar{X} - Z)$

(c) $\bar{X} = \frac{1}{2}(3M - Z)$

AVERAGE RATE OF GROWTH OF POPULATION

Average Rate of Growth of Population

$$P_n = P_0 (1+r)^n$$

$$r = \sqrt[n]{\frac{P_n}{P_0}} - 1$$

weighted G.M.

$$W.G.M = \text{Antilog}\left[\frac{\Sigma W \log X}{\Sigma W}\right]$$

5 HARMONIC MEAN

The harmonic mean It's a specialized average that's particularly useful in business statistics when dealing with rates, ratios, and proportions. Let's dive into its importance and applications:

Importance:

1. Rates and ratios: Harmonic mean is ideal for averaging rates, such as production rates, response rates, or failure rates.
2. Proportions: It's used to average proportions, like market share, employee turnover, or defect rates.
3. Non-uniform data: Harmonic mean is suitable for data with varying units or scales.

Applications:

1. Finance: Calculate average cost of capital, portfolio returns, or credit scores.
2. Operations: Analyze production rates, capacity utilization, or supply chain efficiency.
3. Marketing: Study customer acquisition rates, conversion rates, or customer retention.
4. Quality control: Monitor defect rates, failure rates, or quality metrics.

Key aspects:

1. Sensitivity to small values: Harmonic mean is heavily influenced by small values, making it suitable for datasets with varying magnitudes.
2. Not sensitive to extreme values: Unlike arithmetic mean, harmonic mean is less affected by outliers.
3. Not always intuitive: Harmonic mean can produce counterintuitive results, so interpretation requires care.

Real-world examples:

1. Calculating the average cost of capital for a company

2. Analyzing the production rate of a manufacturing line
3. Studying the customer acquisition rate for a new product
4. Monitoring the defect rate in a quality control process
5. Examining the average credit score for a portfolio of loans

5. Harmonic Mean

Individual Series → $\frac{N}{\frac{1}{X_1}+\frac{1}{X_2}\ \frac{1}{X_3}+\frac{1}{X_n}}$

Discrete Series

$H.M = \frac{N}{\Sigma f \times \frac{1}{X}}$

Continuous Series

$H.M. = \frac{N}{\Sigma(f/M)}$

Weighted H.M.

$= \frac{\Sigma w}{\Sigma\left(\frac{w}{x}\right)}$

MEASURE OF DISPERSON

Measures of Dispersion

Dispersion is the measure of the variation of the items.

Common measures of dispersion:

1. Range: Range may be defined as the difference between largest and smallest item of the distribution

RANGE = L - S

2. Quartile deviation: the average amount by which the upper and lower quartile differ from median.

Q_3-median= median-Q_1

Coefficient of Q.D.

COEFFICIENT OF Q.D. = Q_3-Q_1 / Q_3+Q_1

3. Variance: The average of the squared differences from the mean.
4. Standard Deviation: The square root of the variance.

Features and Applications:

1. Risk analysis: Measures of dispersion help assess risk in finance, insurance, and engineering.
2. Quality control: Dispersion measures monitor consistency in manufacturing and production.
3. Data analysis: Dispersion helps identify patterns, outliers, and anomalies in data.
4. Investment analysis: Dispersion measures evaluate portfolio risk and diversification.
5. Scientific research: Dispersion measures study variability in experimental data.

Key aspects:

1. Higher dispersion: Indicates more variability in the data.
2. Lower dispersion: Indicates less variability in the data.
3. Comparison: Dispersion measures allow comparison across different datasets.

Real-world examples:

1. Analyzing stock price volatility
2. Monitoring product quality control
3. Studying patient recovery times in medical research
4. Evaluating student performance in education
5. Assessing credit risk in lending

Dispersion rules

Dispersion rules are guidelines for interpreting and applying measures of dispersion in various contexts. Here are some key rules:

1. Rule of Thumb: A common rule for understanding standard deviation:
   - 1 σ (sigma) = 68% of data points within 1 standard deviation of the mean
   - 2 σ = 95% of data points within 2 standard deviations of the mean
   - 3 σ = 99.7% of data points within 3 standard deviations of the mean
2. Interquartile Range (IQR) Rule: IQR is typically used to identify outliers:
   - Data points below Q1 - 1.5 * IQR or above Q3 + 1.5 * IQR are considered outliers
3. Variance Rule: Variance is sensitive to extreme values:
   - A single extreme value can greatly inflate the variance
4. Standard Deviation Rule: Standard deviation is a more stable measure than variance:
   - Use standard deviation for most applications, reserving variance for specific cases (e.g., hypothesis testing)
5. Coefficient of Variation (CV) Rule: CV helps compare dispersion across datasets:
   - CV = (Standard Deviation / Mean) * 100
   - Lower CV indicates less relative dispersion
6. Rule: A general guideline for estimating the proportion of data points within a certain number of standard deviations:
   - At least 75% of data points within 2 σ
   - At least 89% of data points within 3 σ

Measure of Disperson

1. Range

$R = L - S$

L → Large value

S → Smallest value

~~Coeff~~ Coefficiant of Range

$= \frac{L-S}{L+S}$

2. Inter quartile Range and quartile Deviation

$I.Q.R \rightarrow Q_3 - Q_1$

$Q.D = \frac{Q_3 - Q_1}{2}$

Coefficient of Q.D. $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

3. Mean Deviation

M.D from Median $= \frac{\Sigma X - M}{N}$ or $\frac{\Sigma d_m}{N}$

M.D from Mean $= \frac{\Sigma X - \bar{X}}{N}$ or $\frac{\Sigma d\bar{x}}{N}$

$$\text{Coefficient of } M.D._m = \frac{M.D._m}{\text{Median}}$$

$$\text{Coefficient of } M.D._{\bar{x}} = \frac{M.D._{\bar{x}}}{\text{Mean}}$$

4. Standard Deviation (Karl Pearson)

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}} \text{ or } \sqrt{\frac{\Sigma x^2}{N}}$$

$$\text{Coefficient of } S.D = \frac{\sigma}{\bar{x}}$$

Step Deviation

$$\sigma = \sqrt{\frac{\Sigma f a'^2}{N} - \frac{(\Sigma f a')^2}{N}} \times i$$

$$\text{Variance} = (s.d)^2 = \sigma^2$$

$$\sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}} \text{ OR}$$

$$\frac{\Sigma f a^2}{N} - \left(\frac{\Sigma f a}{N}\right)^2$$

UNIT-3

SKEWNESS

Skewness in business statistics

When a series is not symmetrical it is said to be skewed.

Skewness refers to the asymmetry or lack of symmetry in a distribution. It's a crucial concept in business statistics, as it helps us understand the shape of the data and make informed decisions.

Types of Skewness:

1. Positive Skewness (Right-skewed): The tail is longer on the right side.
2. Negative Skewness (Left-skewed): The tail is longer on the left side.
3. Zero Skewness (Symmetric): The distribution is perfectly symmetrical.

Causes of Skewness:

1. Data transformation: Changing the scale or units of measurement.
2. Outliers: Presence of extreme values.
3. Non-normality: Data doesn't follow a normal distribution.
4. Unequal variance: Variance changes across different groups.

Effects of Skewness:

1. Biased estimates: Skewness can lead to inaccurate estimates of mean and standard deviation.
2. Misleading conclusions: Ignoring skewness can result in incorrect interpretations.
3. Model misspecification: Skewness can indicate a need for alternative models or transformations.

Business Applications:

1. Risk analysis: Skewness helps assess potential losses or gains.
2. Marketing: Skewness informs customer behavior and preference analysis.
3. Finance: Skewness is crucial in portfolio optimization and risk management.

Key Statistics for Skewness:

1. Skewness coefficient (γ): Measures the degree of skewness.
2. Kurtosis (κ): Related to skewness, measures "tailedness".

Measure of Skewness

1 Karl Pearson's Measures

$$Sk = \bar{X} - Z$$

when mode (Z) is ill defined then

$$Sk = 3(\bar{X} - M)$$

$$\text{Coefficient of } Sk = \frac{\bar{X} - Z}{\sigma}$$

$$\text{Coeff. of } Sk = \frac{3(\bar{X} - M)}{\sigma} \quad \text{(when Z is defined)}$$

2. Bowley's measures

$$Sk = Q_3 + Q_1 - 2M$$

$$\text{Coefficient of } Sk = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

3. Kelly's measures

$$Sk = P_{90} + P_{10} - 2M \text{ or } D_9 + D_1 - 2M$$

$$\text{Coeff. } Sk = \frac{P_{90} + P_{10} - 2M}{P_{90} - P_{10}} \text{ or } \frac{D_9 + D_1 - 2M}{D_9 - D_1}$$

MOMENTS

Moments are mathematical expectations that help describe the distribution of a random variable. They're crucial in business statistics for understanding data properties and making informed decisions.

Types of Moments:

1. First Moment: The mean (average value)
2. Second Moment: The variance (spread or dispersion)
3. Third Moment: The skewness (asymmetry)
4. Fourth Moment: The kurtosis (tailedness)

Key Concepts:

1. Expected Value: The long-term average value of a random variable
2. Central Moments: Moments calculated around the mean (e.g., variance, skewness)
3. Raw Moments: Moments calculated around zero (e.g., mean, second moment)

Applications in Business Statistics:

1. Risk Analysis: Moments help assess potential losses or gains
2. Quality Control: Moments monitor process stability and defect rates
3. Finance: Moments inform portfolio optimization and risk management
4. Marketing: Moments analyze customer behavior and preference

Interpretation:

1. Mean: Central tendency
2. Variance: Spread or dispersion
3. Skewness: Asymmetry (positive or negative)
4. Kurtosis: Tailedness (leptokurtic, mesokurtic, or platykurtic)

moments / kurtosis

moments about mean

$$\mu_1 = \frac{\Sigma(x-\bar{x})^1}{N} = 0$$

$$\mu_2 = \frac{\Sigma(x-\bar{x})^2}{N}$$

$$\mu_3 = \frac{\Sigma(x-\bar{x})^3}{N}$$

$$\mu_4 = \frac{\Sigma(x-\bar{x})^4}{N}$$

FOR A FREQUENCY Distribution

$$\mu_1 = \frac{\Sigma f(x-\bar{x})^1}{N}$$

$$\mu_2 = \frac{\Sigma f(x-\bar{x})^2}{N}$$

moments about Arbitary origin 'A'

$$\mu_1 = \frac{\Sigma(x-A)^1}{N} \quad \mu_2 = \frac{\Sigma(x-A)^2}{N}$$

$$\mu_3 = \frac{\Sigma(x-A)^3}{N} \quad \mu_4 = \frac{\Sigma(x-A)^4}{N}$$

for a Frequancy Distribution

$\mu_1 = \frac{\Sigma f(X-A)^1}{N} y_i$ $\quad \mu_2 = \frac{\Sigma f(X-A)^2}{N} x_i$

moments about zero

$v_1 = \frac{\Sigma X^1}{N}$ $\quad v_2 = \frac{\Sigma X^2}{N}$

$v_3 = \frac{\Sigma X^3}{N}$ $\quad v_4 = \frac{\Sigma X^4}{N}$

Relationship between central and non-central moments

$\mu_1 = 0$

$\mu_2 = \mu_2 - \mu_1^2$

$\mu_3 = \mu_3 - 3\mu_2 \mu_1 + 2\mu_1^3$

$\mu_4 = \mu_4 - 4\mu_3 \mu_1 + 6\mu_2 \mu_1^2 - 3(\mu_1)^4$

Relationship Between central moments and moments about origin

$v_1 = \bar{x}_1$ $\quad v_2 = \mu_2 + v_1^2$ $\quad v_3 = \mu_3 + 3\mu_2 \cdot v_1 + (v_1)^3$

$v_4 = \mu_4 + 4\mu_3 \cdot v_1 + 6\mu_2 \cdot (v_1)^2 + (v_1)^4$

Skewness and kurtosis

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ $\quad \beta_2 = \frac{\mu_4}{\mu_2^2}$

$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}}$ $\quad \gamma_2 = \beta_2 - 3$

INDEX NUMBER

Index numbers are statistical devices used to measure changes in the value of a group of items or commodities over time. They're essential in business statistics for analyzing and understanding economic and business trends.

Types of Index Numbers:

1. Price Index: Measures changes in prices of a basket of goods and services (e.g., Consumer Price Index, CPI)
2. Quantity Index: Measures changes in quantities of production, sales, or other economic variables
3. Value Index: Measures changes in the value of a group of items or commodities

Common Index Numbers:

1. Consumer Price Index (CPI): Measures changes in prices of a basket of goods and services for consumers
2. Producer Price Index (PPI): Measures changes in prices of goods and services for producers
3. Dow Jones Industrial Average (DJIA): A stock market index that measures the performance of 30 large-cap companies
4. GDP Deflator: Measures changes in the value of all goods and services produced within an economy

Applications in Business Statistics:

1. Inflation analysis: Index numbers help understand inflation rates and trends
2. Economic growth analysis: Index numbers help analyze changes in economic output and growth
3. Business performance analysis: Index numbers help evaluate changes in sales, revenue, and productivity
4. Purchasing power analysis: Index numbers help understand changes in the value of money

Key Concepts:

1. Base year: The reference year used for comparison
2. Index number formula: (Current year value / Base year value) x 100
3. Chain base indexing: A method of linking index numbers to form a continuous time series

INDEX Number

1. Simple Index Number

(i) Simple Aggregative

$$P_{01} = \frac{\Sigma P_1}{\Sigma P_0} \times 100$$

(ii) Simple Average of Price Relative

(a) $$P_{01} = \frac{\Sigma \frac{P_1}{P_0} \times 100}{N}$$ (using A.M)

(b) $$P_{01} \text{ Antilog}\left[\frac{\Sigma \text{Log}(P_1/P_0)}{N} \times 100\right]$$ using G.M.

2. Weighted Aggregate

(i) Weighted Aggregate

(a) Laspeyre $$P_{01} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100$$

(b) Paasche

$$P_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times 100$$

Fisher

(c) $$P_{01} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}} \times 100$$

(d) Dorbish - Bowley

$$P_{01} = \frac{L+P}{2} \text{ or } \frac{1}{2}\left[\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} + \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}\right]$$

(e) Marshall - edgeworth

$$P_{01} = \frac{\Sigma P_1 (q_0 + q_1)}{\Sigma P_0 (q_0 + q_1)} \times 100$$

(f) Kelly's method

$$P_{01} = \frac{\Sigma P_1 q}{\Sigma P_0 q} \times 100$$

(ii) Weighted Average of Price Relative

(a) If A.M is used $P_{01} = \frac{\Sigma P w}{\Sigma w}$

(b) If G.M is used $P_{01} = \text{Antilog}\left[\frac{\Sigma (w \log P)}{\Sigma w}\right]$

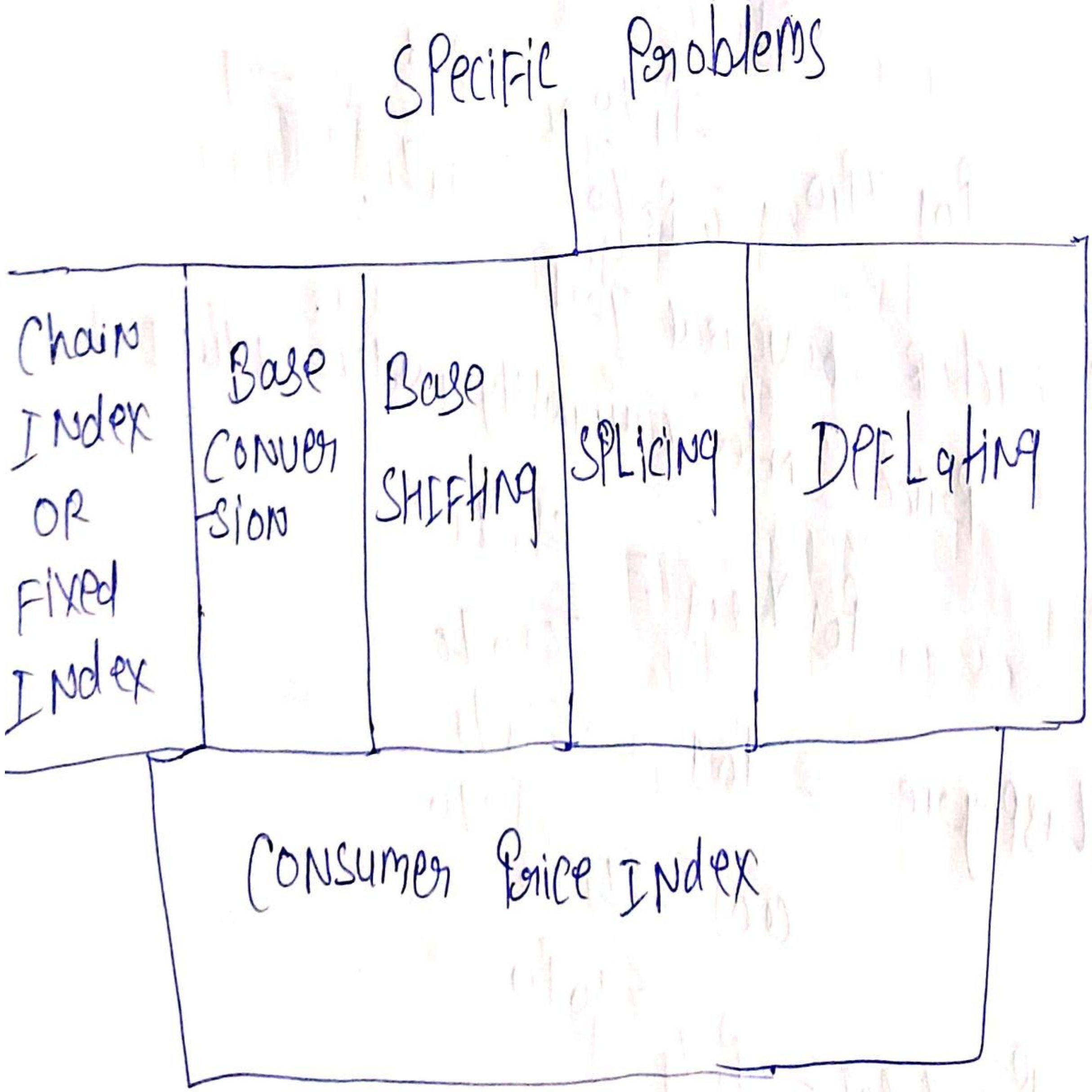
3. Tests of Adequacy

Time Reversal Test = $P_{01} \times P_{10} = 1$

Factor Reversal test = $P_{01} \times Q_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$

Circular test i satisfied when

$$P_{01} \times P_{12} \times P_{20} = 1$$


Specific Problems
Chain Index OR Fixed Index
Base Conversion
Base Shifting
Splicing
Deflating
Consumer Price Index

1. CHAIN INdex

$$\text{LINK Relatives} = \frac{\text{current year's}}{\text{Previous year}} \times 100$$

$$\text{CHAIN Base INdex} = \frac{\dfrac{\text{LINK Relatives of current year} \times \text{CHAIN INdex of Previous year}}{}}{100}$$

1 (i) $$\text{Price Relatives} = \frac{\text{current year's Price}}{\text{Base year's Price}} \times 100$$

2. Base - CONVERSION

(i) $$\text{current year's FBI} = \frac{\text{Current year CBI} \times \text{Previous year FBI}}{100}$$

(ii) Fixed Base INdex to Chain Base INdex

$$\text{current year CBI} = \frac{\text{current year}}{\text{Previous year}} \times 100$$

(3) Base Shifting

$$\text{INdex NO with New Base} = \frac{\text{INdex NO. with old Base}}{\text{INdex NO. OF New Base year}} \times 100$$

4. Splicing

(i) Spliced New Index series to old Index series

$$= \frac{\text{New Index} \times \text{old Index of overlapping year}}{100}$$

(ii) Speced old Index series with new Index series = $\frac{\text{old Index} \times 100}{\text{old Index of overlapping year}}$

(5) Deflating

$$\text{Real wages} = \frac{\text{Money wages}}{\text{Price Index}} \times 100$$

$$\text{Real Income} = \frac{\text{Money Income}}{\text{Cost of Living Index}} \times 100$$

Purchasing Power of Money

$$= \frac{1}{\text{Life Index}}$$

6. Consumer Price Index

or

Cost of Living Index

(i) Aggregate expenditure

$$P_{01} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100$$

(ii) Family Budget

$$P_{01} = \frac{\Sigma PW}{\Sigma W}$$

$$P_{01} = AL\left[\frac{\Sigma W \log P}{\Sigma W}\right]$$

UNIT-4

DEVELOPMENT OF THEORY OF PROBABILITY

The development of the theory of probability has a rich history that spans thousands of years. Here's a brief overview:

Ancient Civilizations (3000 BCE - 500 CE) Evidence of probabilistic thinking in ancient cultures, such as astragali (animal bones) used for divination Games of chance and gambling in ancient Egypt, Greece, and Rome

Medieval Period (500 - 1500 CE):

Study of probability in Islamic mathematics and philosophy (e.g., Al-Khwarizmi, Ibn Adlan)

Introduction of the concept of "chance" in scholastic philosophy TAquinas)Renaissance and Enlightenment (1500 - 1800 CE) Emergence of probability theory as a distinct field (e.g., Girolamo Cardano, Pierre de Fermat)Development of probability calculus (e.g., Blaise Pascal, Christiaan Huygens)

Classical Probability (1600 - 1900 CE):

Major contributions by Pierre-Simon Laplace, Thomas Bayes, and Carl Friedrich Gauss

Development of probability distributions (e.g., Bernoulli, Poisson)

Modern Probability (1900 - present):

Andrey Kolmogorov's axiomatic approach (1933)

Development of stochastic processes (e.g., Wiener process, Markov chains)

Applications in statistics, physics, engineering, economics, and finance

Some notable figures in the development of probability theory include:

- **Pierre-Simon Laplace (1749-1827)**
- **Thomas Bayes (1702-1761)**
- **Carl Friedrich Gauss (1777-1855)**
- **Andrey Kolmogorov (1903-1987)**
- **Ronald Fisher (1890-1962)**

THEORY OF PROBABILITY

The Theory of Probability is a branch of mathematics that deals with the study of chance events and their likelihood of occurrence. It provides a mathematical framework for analyzing and modeling random phenomena, making predictions, and estimating probabilities.

Key Concepts:

1. Experiment: An action or situation that can produce a set of outcomes.

2. Sample Space: The set of all possible outcomes of an experiment.
3. Event: A set of one or more outcomes of an experiment.
4. Probability: A number between 0 and 1 that represents the likelihood of an event occurring.
5. Random Variable: A variable whose possible values are determined by chance.

Basic Principles:

1. The probability of an event is always between 0 and 1: $P(A) \geq 0$ and $P(A) \leq 1$
2. The probability of the sample space is 1: $P(S) = 1$
3. The probability of the empty set is 0: $P(\emptyset) = 0$
4. The probability of an event is equal to the sum of the probabilities of its disjoint subsets: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$

Types of Probability:

1. Theoretical Probability: Based on the number of favorable outcomes divided by the total number of possible outcomes.
2. Empirical Probability: Based on experimental data and relative frequencies.
3. Conditional Probability: The probability of an event occurring given that another event has occurred.
4. Independent Probability: The probability of an event occurring regardless of other events.

Probability Distributions:

1. Bernoulli Distribution: Models binary outcomes (success/failure)
2. Binomial Distribution: Models the number of successes in a fixed number of trials
3. Poisson Distribution: Models the number of events in a fixed interval
4. Normal Distribution: Models continuous data with a symmetric bell-shaped curve

Applications of Probability:

1. Statistics: Hypothesis testing, confidence intervals, and statistical inference
2. Engineering: Reliability, quality control, and signal processing
3. Economics: Risk analysis, finance, and econometrics
4. Computer Science: Machine learning, data mining, and artificial intelligence

some important words in probability:

1. Experiment: An action or situation that can produce a set of outcomes.
2. Sample Space: The set of all possible outcomes of an experiment.
3. Event: A set of one or more outcomes of an experiment.

4. Probability: A number between 0 and 1 that represents the likelihood of an event occurring.
5. Random Variable: A variable whose possible values are determined by chance.
6. Outcome: A specific result of an experiment.
7. Trial: A single execution of an experiment.
8. Independence: The occurrence of one event does not affect the probability of another event.
9. Mutually Exclusive: Events that cannot occur simultaneously.
10. Conditional Probability: The probability of an event occurring given that another event has occurred.
11. Binomial: Relating to a probability experiment with two possible outcomes (success/failure).
12. Normal: A continuous probability distribution with a symmetric bell-shaped curve.
13. Variance: A measure of the spread or dispersion of a probability distribution.
14. Expectation: The long-run average value of a random variable.
15. Bayes' Theorem: A formula for updating the probability of a hypothesis based on new data.
16. Stochastic: Involving randomness or chance.
17. Distribution: A function that describes the probability of each possible value of a random variable.
18. Mean: The expected value of a random variable.
19. Median: The middle value of a probability distribution.
20. Mode: The most frequently occurring value in a probability distribution.

1 EMPIRICAL /RELATIVE FREQUENCY THEORY OF PROBABILITY

Empirical probability is a theoretical approach that focuses on the frequency of events based on experimental data and observation. It emphasizes the use of real-world data to estimate probabilities, rather than relying solely on theoretical models.
Key aspects of empirical probability:
1. Frequency interpretation: Probability is defined as the relative frequency of an event occurring in a large number of trials.
2. Experimental data: Probabilities are estimated based on data collected from experiments or observations.
3. Empirical distribution: A probability distribution is constructed from the data, such as a histogram or an empirical cumulative distribution function (ECDF).
4. Estimation: Probabilities are estimated from the data using statistical methods, such as maximum likelihood estimation (MLE).

Empirical probability is used in:

1. Statistics: Hypothesis testing, confidence intervals, and statistical inference.

2. Machine learning: Training models on data to make predictions and estimate probabilities.

3. Data science: Analyzing and visualizing data to understand patterns and trends.

4. Engineering: Reliability engineering, quality control, and signal processing.

5. Economics: Econometrics, financial modeling, and risk analysis.

Advantages of empirical probability:

1. Data-driven: Based on real-world data, making it more representative of actual outcomes.

2. Flexibility: Can be used with various data types and distributions.

3. Easy to understand: Probabilities are estimated from data, making it more intuitive.

Limitations:

1. Data quality: Requires high-quality data to produce accurate estimates.

2. Sample size: Large enough samples are needed to ensure reliable estimates.

3. Model assumptions: May require assumptions about the underlying distribution.

2. empirical OR Relative Frequency

$$P(A) = \text{Limit}_{n \to} \frac{r}{N}$$

$$P(E) = \frac{\text{No. of Favourable Cases of } E}{\text{Total No. of Equally likely Cases}} = \frac{m}{N}$$

$$P(\bar{E}) = 1 - P(E)$$

2 CLASSICAL APPROACH

Classical probability theory, also known as a priori probability, is a theoretical approach that defines probability based on the number of favorable outcomes divided by the total number of possible outcomes. This approach assumes that all outcomes are equally likely and that the probability of an event can be determined through logical reasoning and mathematical calculations.

Key aspects of classical probability theory:

1. Equally likely outcomes: Assumes that all outcomes have the same probability of occurring.

2. Finite sample space: The number of possible outcomes is finite.
3. Logical reasoning: Probabilities are determined through logical arguments and mathematical calculations.
4. Theoretical probability: Probabilities are calculated based on theoretical models rather than empirical data.

Classical probability theory is used in:

1. Coin tossing: Calculating the probability of heads or tails.
2. Dice rolling: Calculating the probability of different outcomes.
3. Card games: Calculating the probability of different hands.
4. Random number generation: Generating truly random numbers.

Advantages of classical probability theory:

1. Simple and intuitive: Easy to understand and apply.
2. Logical and consistent: Based on logical reasoning and mathematical calculations.
3. Applicable to idealized situations: Useful for analyzing idealized situations, such as coin tossing.

Limitations:

1. Assumes equally likely outcomes: May not hold true in real-world situations.
2. Limited to finite sample spaces: Not applicable to continuous or infinite sample spaces.

Probability

1. Classical or mathematical
2. Statistical or Relative Frequency
3. Subjective

1. Classical

$$P(A) = p = \frac{\text{Number of Favourable cases}}{\text{Total number of Equally likely cases}}$$

$$= \frac{m}{N}$$

$$p = \frac{m}{m+N}$$

$$q = \frac{m}{m+N}$$

3 SUBJECTIVE APPROACH

The subjective approach to probability, also known as the personalistic or Bayesian approach, is a theoretical framework that views probability as a measure of an individual's degree of belief or uncertainty about an event. This approach emphasizes the role of personal judgment and experience in assigning probabilities.

Key aspects of the subjective approach:

1. Personal probability: Probability is a measure of an individual's degree of belief or uncertainty.
2. Subjective judgment: Probabilities are assigned based on personal experience, expertise, and judgment.
3. Bayesian inference: Probabilities are updated based on new information or data using Bayes' theorem.
4. Degree of belief: Probability is seen as a measure of an individual's strength of belief.

Subjective approach is used in:

1. Decision-making: Making decisions under uncertainty.
2. Risk analysis: Assessing and managing risks.
3. Expert systems: Eliciting probabilities from experts.
4. Artificial intelligence: Modeling uncertainty in AI systems.

Advantages:

1. Flexible: Allows for personal judgment and experience.
2. Adaptable: Probabilities can be updated based on new information.
3. Realistic: Recognizes uncertainty and ambiguity.

Limitations:

1. Subjective: Probabilities are dependent on individual perspectives.
2. Arbitrary: Probabilities can be assigned arbitrarily.
3. Difficult to quantify: Degree of belief can be hard to measure.

USE OF COMBINATION IN THEORY OF PROBABILITY

use of combinations in theory of Probability

$$^nC_r = \frac{n!}{n-r!\ r!}$$

$$A, B, C = {}^3C_2 = \frac{3!}{3-2!\ 2!} = \frac{3\times2\times1}{1\times2\times1}$$

$= 3$ ways

1 ADDITION THEOREM

The Addition Theorem, also known as the Sum Rule, is a fundamental concept in probability theory. It states that the probability of two or more mutually exclusive events occurring is the sum of their individual probabilities.

Mathematically, the Addition Theorem can be represented as:

$P(A \cup B) = P(A) + P(B)$

where:

- $P(A \cup B)$ is the probability of either event A or event B occurring
- $P(A)$ is the probability of event A occurring

- P(B) is the probability of event B occurring
For example, if we have two events:
- Event A: Rolling a die and getting a 1 or 2
- Event B: Rolling a die and getting a 3 or 4
The probability of either event A or event B occurring is:
$P(A \cup B) = P(A) + P(B) = 2/6 + 2/6 = 4/6 = 2/3$

The Addition Theorem can be extended to more than two events, as long as they are all mutually exclusive:
$P(A \cup B \cup C) = P(A) + P(B) + P(C)$
This theorem is useful in many real-world applications, such as:
- Calculating the probability of winning a game
- Estimating the likelihood of a natural disaster
- Determining the probability of a medical diagnosis

(1) Addition Theorem

Addition Theorem for mutually exclusive events

$$P(A \text{ or } B) = P(A) + P(B)$$

or

$$P(A+B) = P(A) + P(B)$$

Proof of the Theorem

$$P(A) = \frac{m_1}{n} \qquad P(B) = \frac{m_2}{n}$$

$$P(A \text{ or } B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B)$$

$$P(A) + P(B) = P(A \text{ or } B)$$

Generalisation

$$P(A+B+C) = P(A) + P(B) + P(C)$$

ADDITION THEOREM FOR NOT MTUALLY EXCLUSIVE EVENTS

The Addition Theorem for not mutually exclusive events is also known as the Inclusion-Exclusion Principle. It states that the probability of two or more events occurring is the sum of their individual probabilities, minus the probability of their intersection.

Mathematically, it can be represented as:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

where:

- $P(A \cup B)$ is the probability of either event A or event B occurring
- $P(A)$ is the probability of event A occurring
- $P(B)$ is the probability of event B occurring
- $P(A \cap B)$ is the probability of both event A and event B occurring (their intersection)

For example, if we have two events:

- Event A: Rolling a die and getting a 1, 2, or 3
- Event B: Rolling a die and getting a 2, 3, or 4

The probability of either event A or event B occurring is:

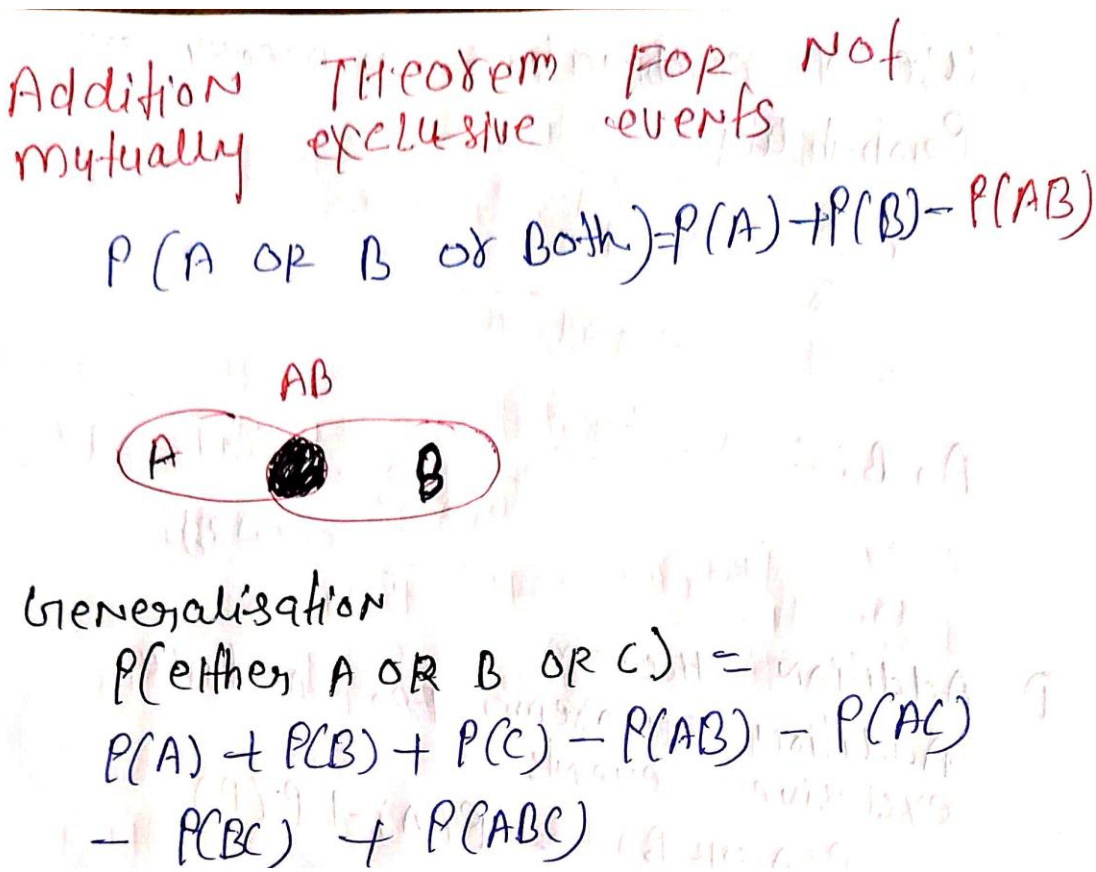
P(A ∪ B) = P(A) + P(B) - P(A ∩ B) = 3/6 + 3/6 - 2/6 = 4/6 = 2/3

The Inclusion-Exclusion Principle can be extended to more than two events, as long as they are not mutually exclusive:

P(A ∪ B ∪ C) = P(A) + P(B) + P(C) - P(A ∩ B) - P(A ∩ C) - P(B ∩ C) + P(A ∩ B ∩ C)

This theorem is useful in many real-world applications, such as:

- Calculating the probability of winning a game
- Estimating the likelihood of a natural disaster
- Determining the probability of a medical diagnosis



2 MULTIPLICATION THEOREM FOR INDEPENDENTS EVENTS

2. MULTIPLICATION THEOREM

multiplication THEOREM FOR INdependents events

$$P(AB) = P(A) \times P(B)$$

Proof: $P(A) = \frac{m_1}{n_1}$ $P(B) = \frac{m_2}{n_2}$

$$P(AB) = \frac{m_1 m_2}{n_1 n_2} = \left(\frac{m_1}{n_2}\right)\left(\frac{m_2}{n_2}\right)$$

$$P(A) \quad P(B)$$

$$P(AB) = P(A) \cdot P(B)$$

PROBABILITY OF HAPPENING OF ATLEAST ONE EVENT OF n INDEPENDENT

Probability of Happening of atleast one event of n Independent events

$$P = 1 - P$$

$$= 1 - [(1-P_1).(1-P_2).(1-P_3).(1-P_n)]$$

Conditional Probability

$$P(B|A) = \frac{P(AB)}{PA} = P(A|B) = \frac{P(AB)}{P(B)}$$

multiplication Theorem in case of conditional probability

$$P(AB) = P(A).P(B|A)$$

or

$$P(AB) = P(B).P(A|B)$$

USE OF BERNOULLI THEOREM IN THEORY OF PROBABILITY

use of Bernoulli Theorem In Theory of Probability

$$P(r) = {}^{n}C_{r}\, p^{r} . q^{n-r}$$

$$r = 1, 2, \ldots n.$$

mathematical expectation

$$m.E. = p \times m$$

p -> Probability of Happing (sucess)

m -> money to be Received on the Happening of the event

3 BAYES THEOREM Bayes' Theorem, also known as Bayes' Rule or Bayes' Formula, is a fundamental concept in probability theory and statistics. It describes how to update the probability of a hypothesis (H) given new evidence (E).

Mathematically, Bayes' Theorem is expressed as:
P(H|E) = P(E|H) × P(H) / P(E)
Where:
- **P(H|E) is the posterior probability of the hypothesis (H) given the evidence (E)**
- **P(E|H) is the likelihood of the evidence (E) given the hypothesis (H)**
- **P(H) is the prior probability of the hypothesis (H)**
- **P(E) is the marginal probability of the evidence (E)**

Bayes' Theorem is used in various applications, such as:

- Hypothesis testing
- Decision-making under uncertainty
- Machine learning
- Artificial intelligence
- Data analysis

The theorem is named after Thomas Bayes (1702-1761), who first proposed it. It's a powerful tool for updating beliefs and making predictions based on new data.

3. Bayes's Theorem

Thomas Bayes (1763)

also called → Inverse Probability

$$P(A_1|B) = \frac{P(A_1) \cdot P(B|A_1)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2)}$$

$$P(A_2|B) = \frac{P(A_2) \cdot P(B|A_2)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2)}$$

Proof of the Theorem