Data Analytics

Basic Statistical Analysis

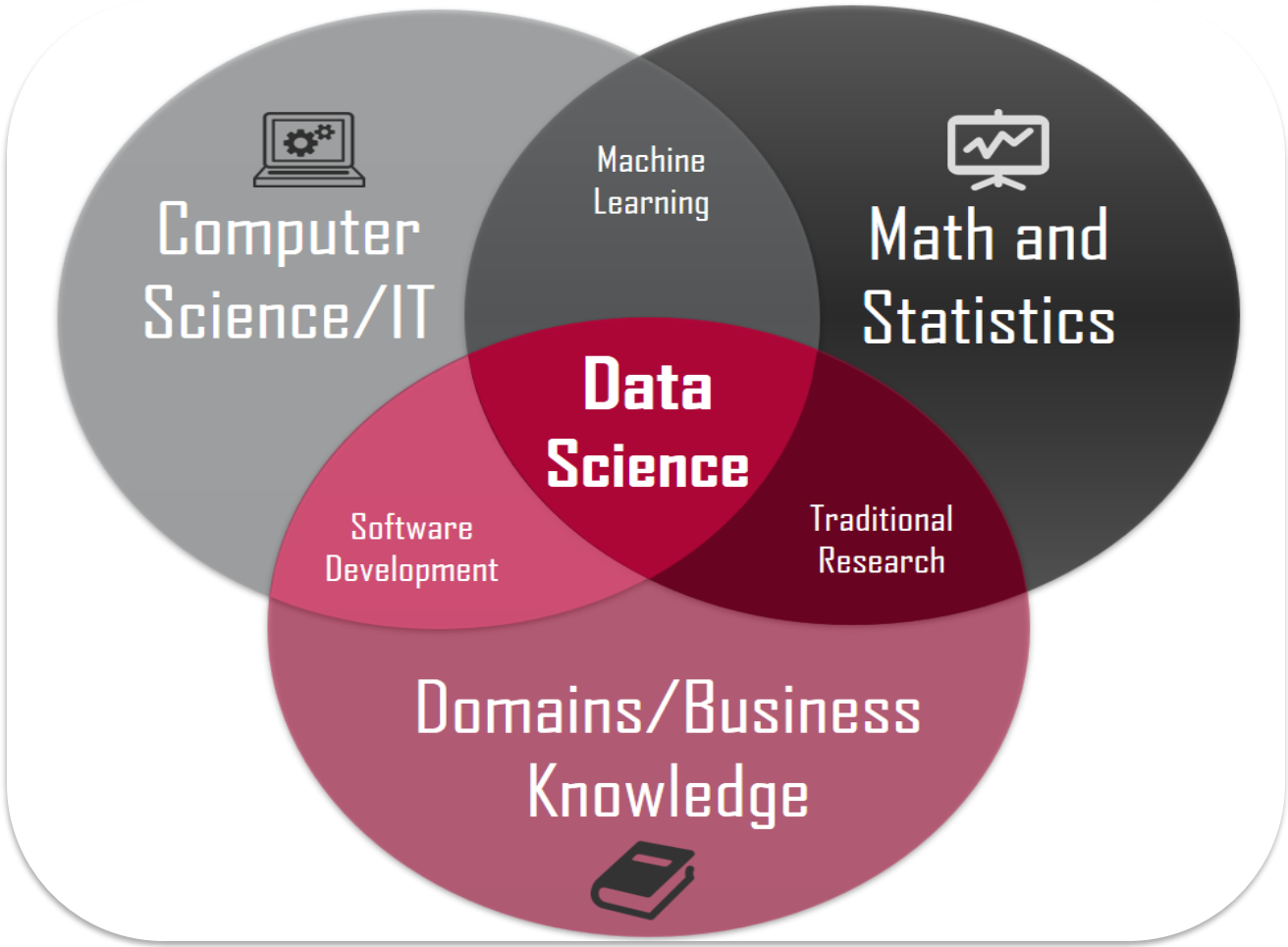
Course Name: 23MCT305 - Data Analytics in Automation System

Faculty Name: N. KARTHI, AP/MCT

Introduction to Statistical Analysis

What is Statistical Analysis?

The science of collecting, analyzing, interpreting, and presenting data to discover patterns and trends



Importance in Data Analytics

- Transform raw data into meaningful insights
- Support evidence-based decision making
- Identify patterns and relationships

Applications

- Business intelligence
- Quality control

Types of Data

Categorical Data

Qualitative information describing qualities or characteristics

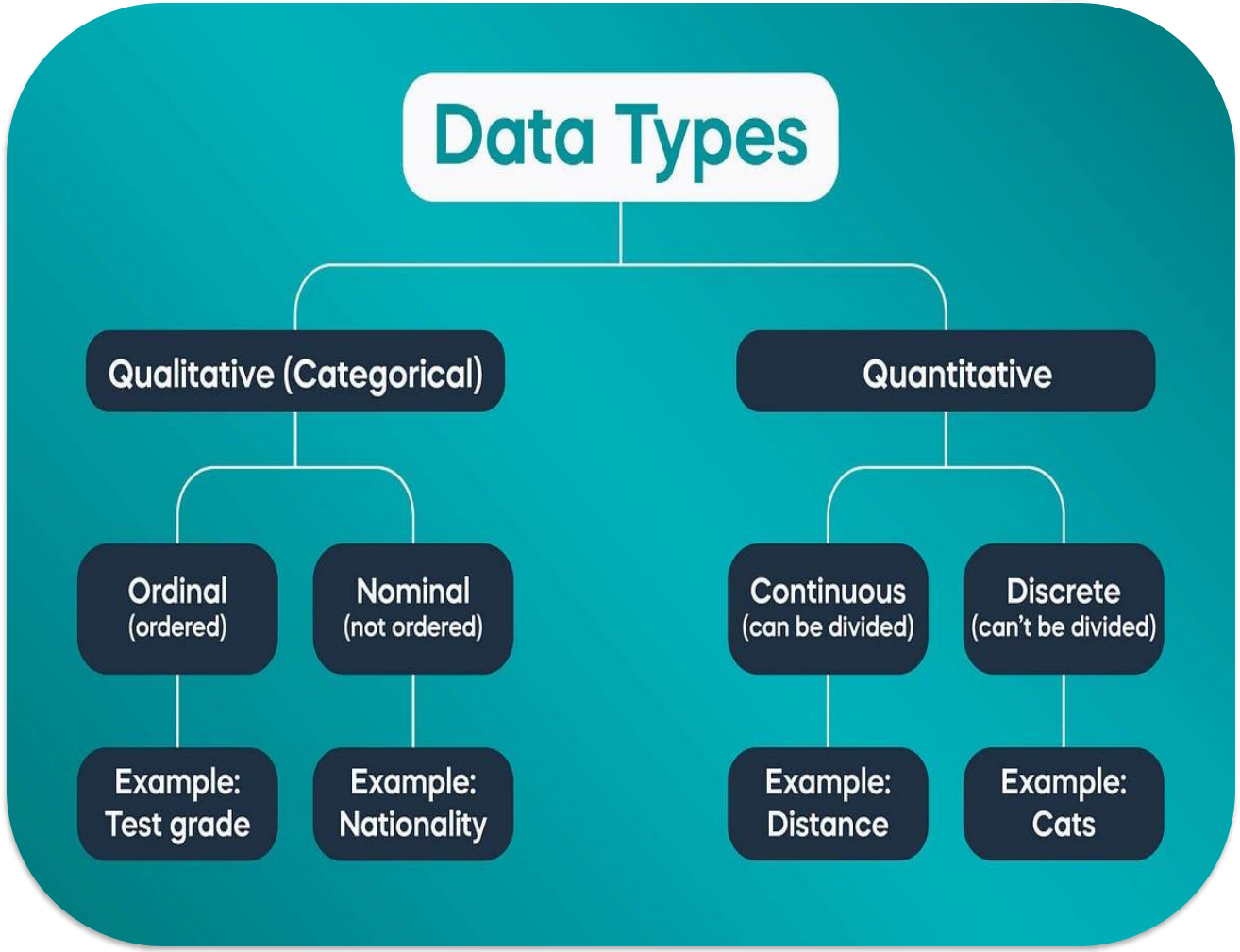
Numerical Data

Quantitative information expressed as numbers

Numerical Subtypes

Discrete: Countable values (integers)

Continuous: Measurable values (decimals)



Measures of Central Tendency

Mean

Average value: Sum of all values divided by number of observations

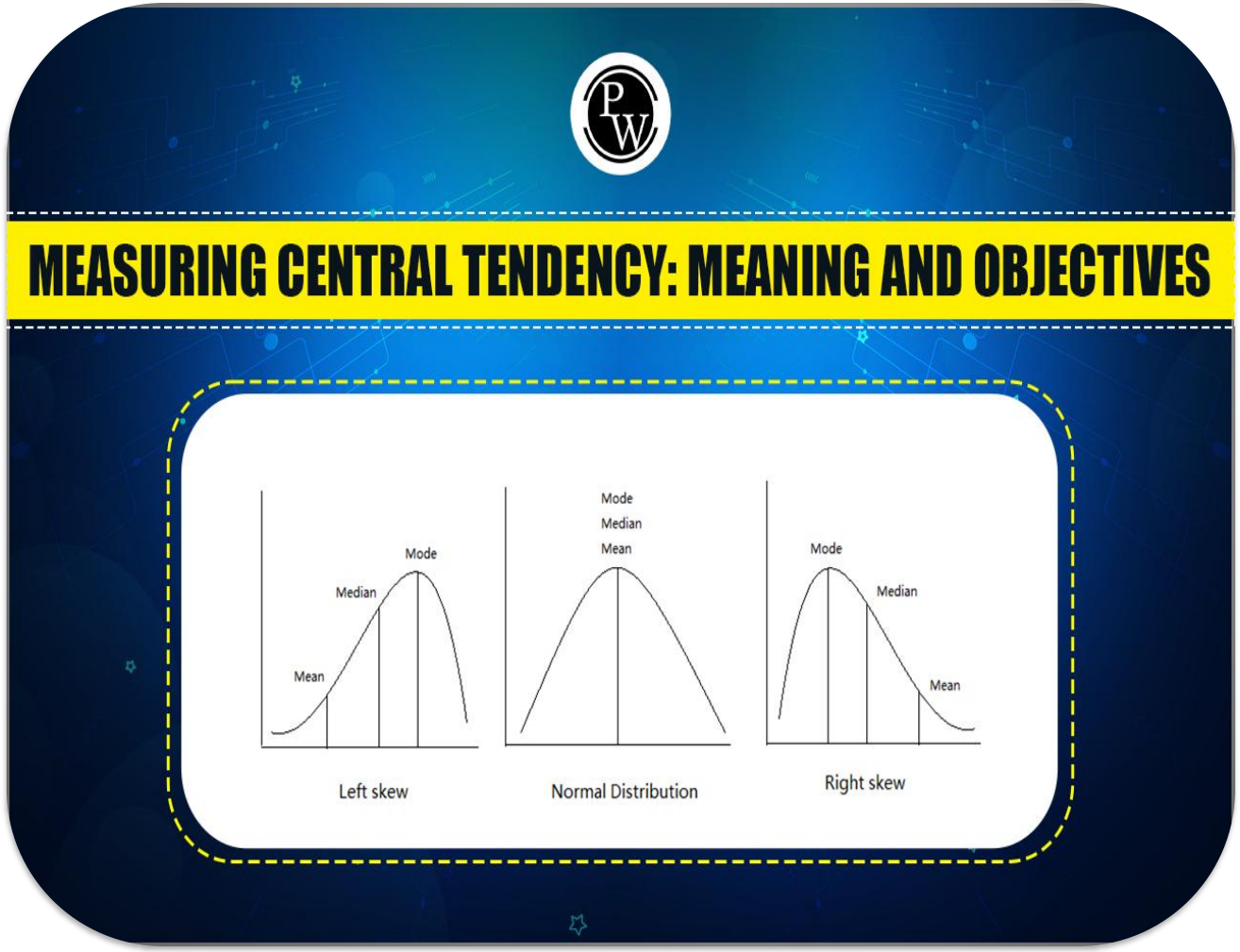
Median

Middle value when data is arranged in ascending or descending order

Mode

Most frequently occurring value in the dataset

Note: Choice depends on data type and distribution



Measures of Dispersion

Range

Difference between maximum and minimum values

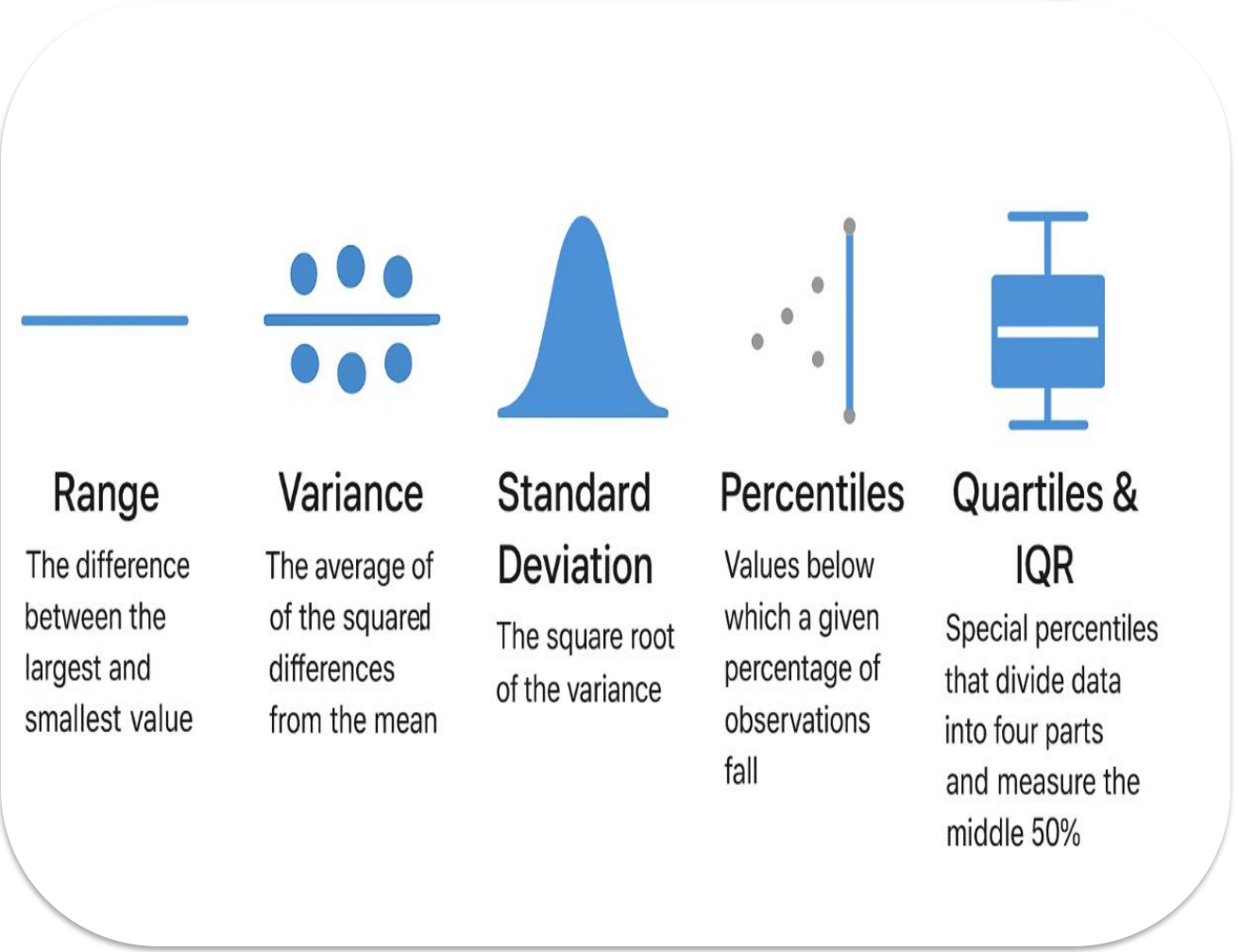
Variance

Measure of how spread data is from the mean

Standard Deviation

Square root of variance, indicates data spread in same units

Note: Higher values indicate greater data variability



Probability Distributions

Normal Distribution

Bell-shaped curve, symmetrical around the mean

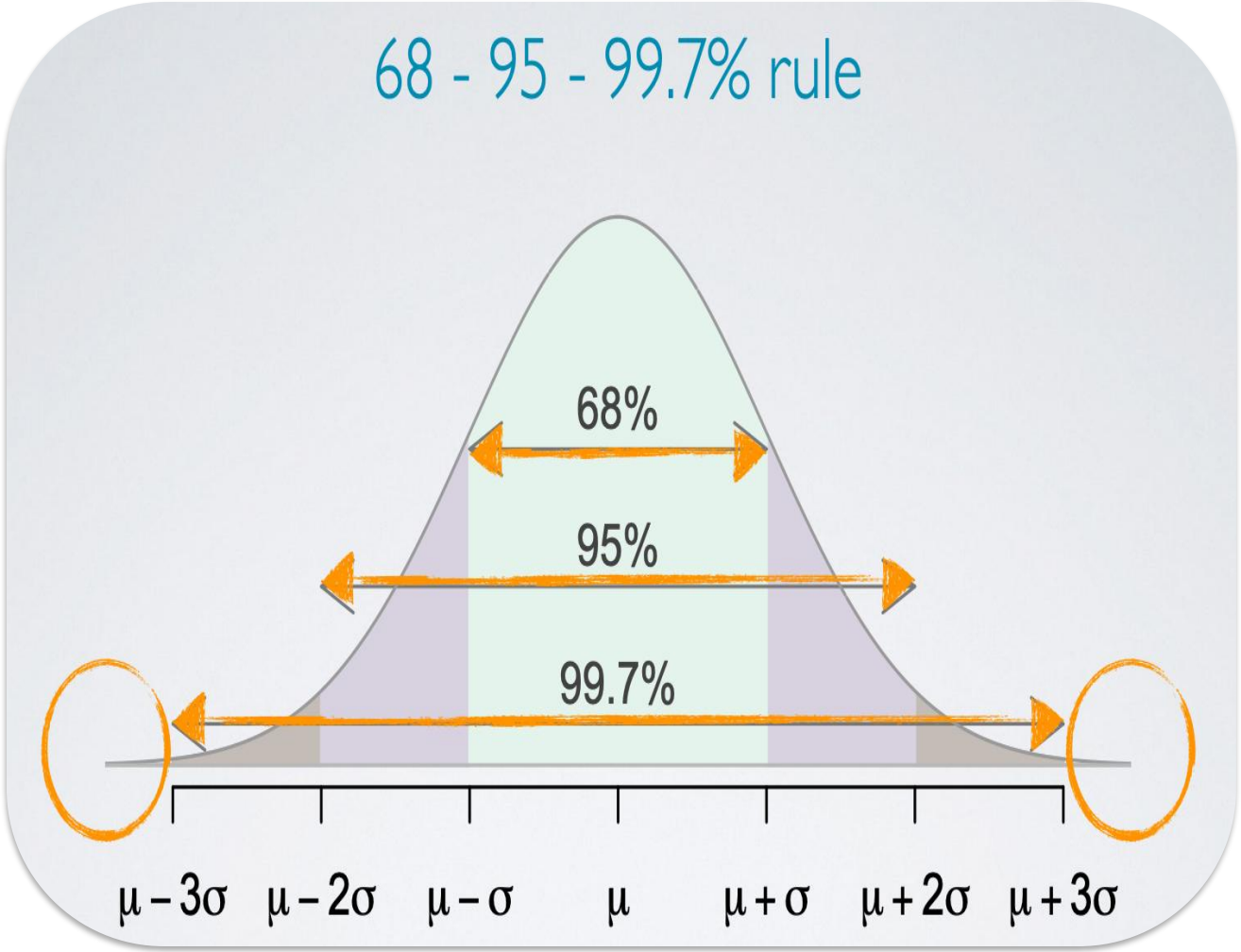
68-95-99.7 Rule

Empirical rule for normal distribution probabilities

Σ Binomial Distribution

Discrete distribution for two possible outcomes

Key: μ = mean, σ = standard deviation



Hypothesis Testing

Null Hypothesis (H_0)

Default assumption, no difference or effect

Alternative Hypothesis (H_1)

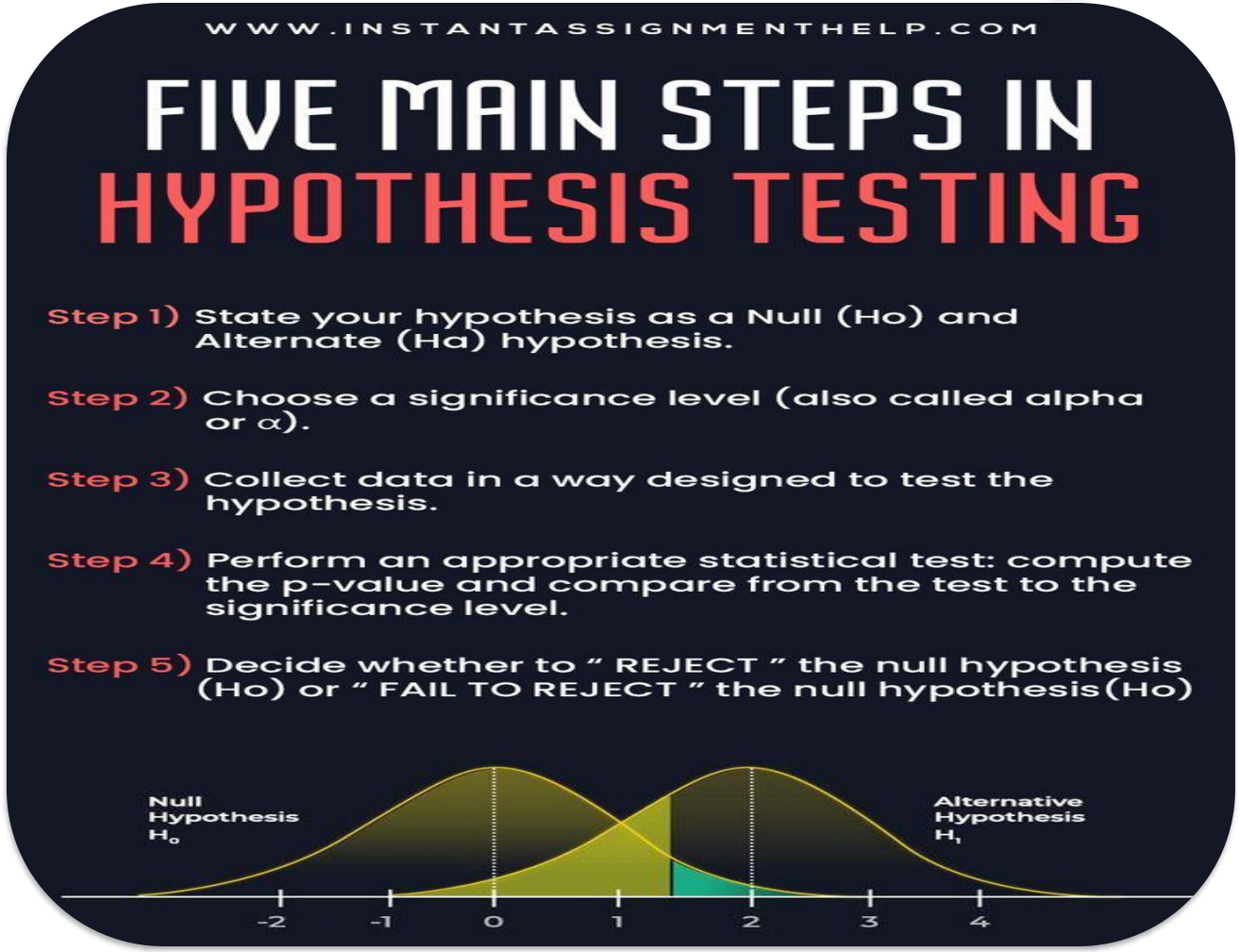
Claim to be tested, contradicts H_0

p-value

Probability of obtaining results if H_0 is true

Decision Making

Compare p-value to significance level ($\alpha = 0.05$)



Correlation and Regression

Correlation

Measures strength and direction of relationship between two variables

Correlation Types

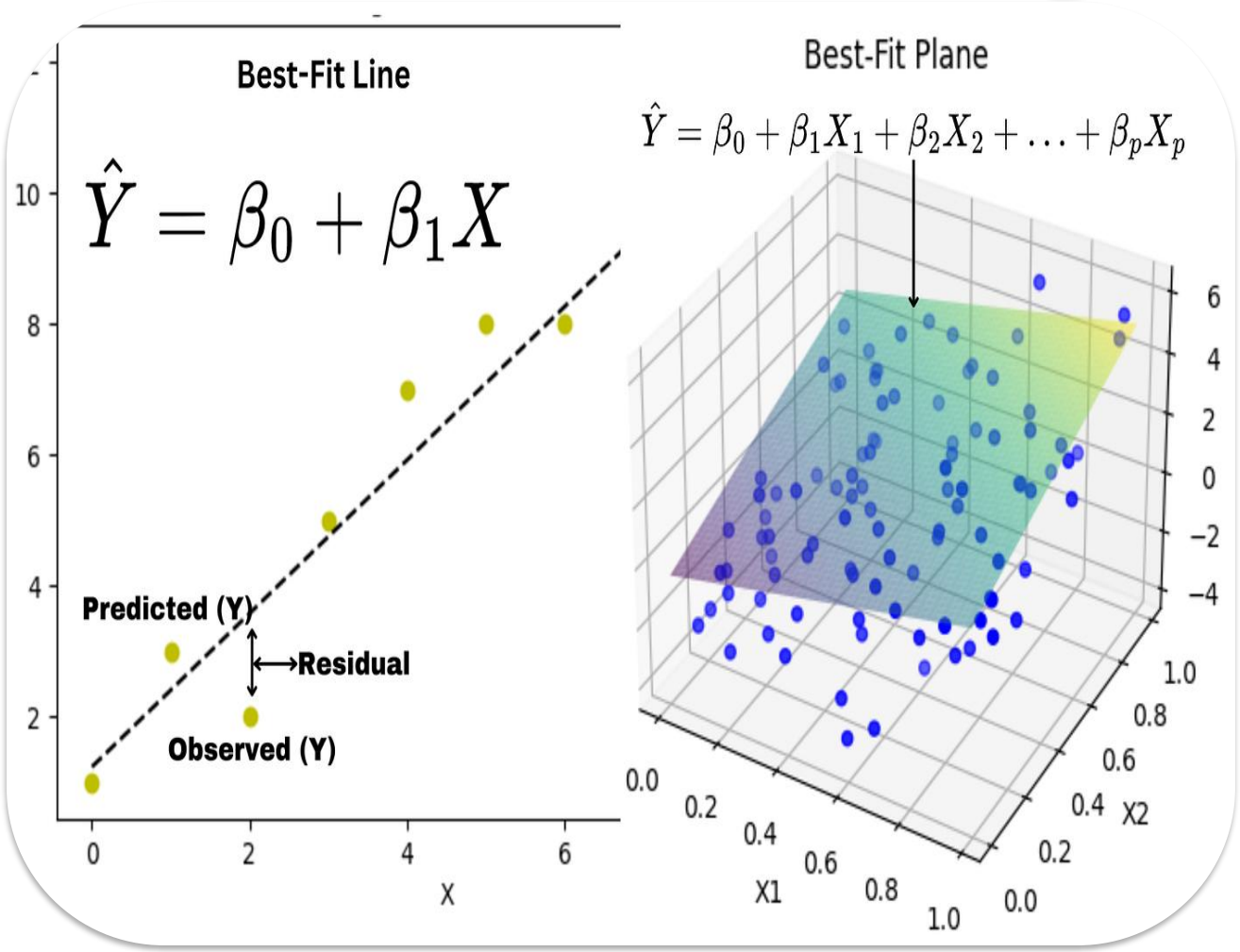
Positive: Both increase together Negative: One increases, other decreases

Regression

Predictive modeling to estimate relationships

Line of Best Fit

Straight line that best represents data points



Mind Maps

Visual Representation

Organize statistical concepts hierarchically

GenAI-Generated Content

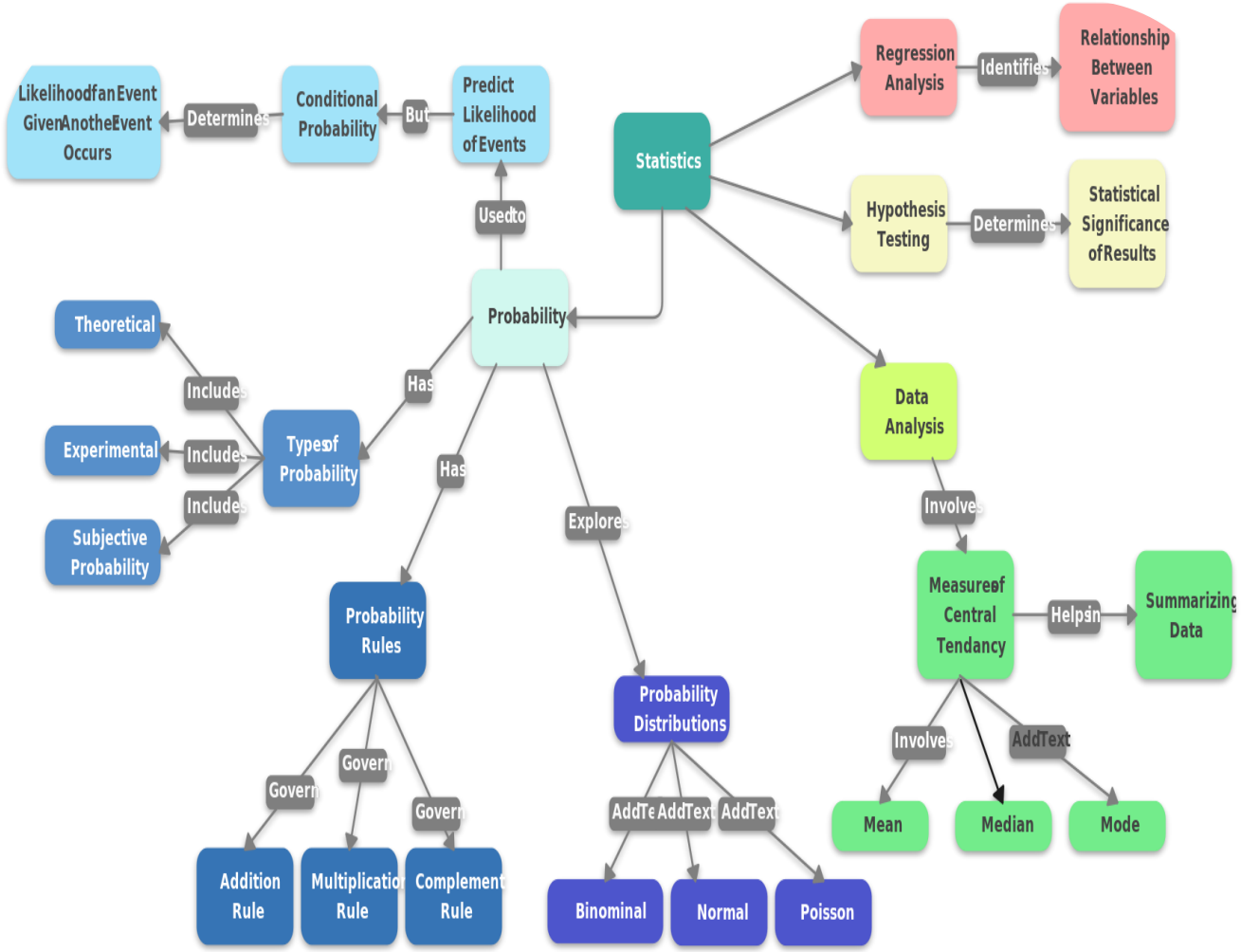
AI-powered insights and connections

Key Concepts

Central tendency, dispersion, probability, hypothesis testing

Interconnected Learning

Visualize relationships between statistical methods



Recap and Summary

Key Takeaways

Statistical analysis transforms data into meaningful insights for decision-making

Analysis Tools

Central tendency, dispersion, probability distributions, hypothesis testing, regression

Automation Applications

Quality control, predictive maintenance, process optimization, performance monitoring

Thank You!