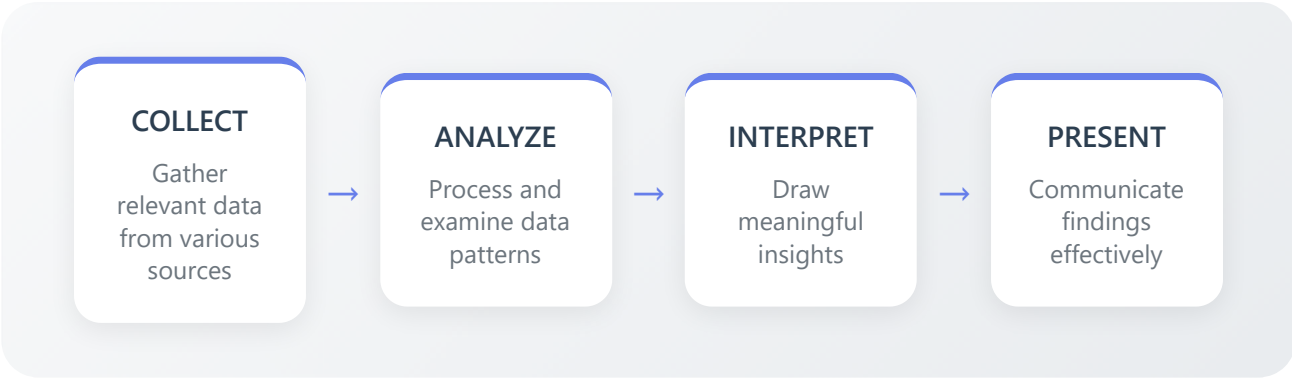
Statistics Concepts Every Data Analyst Should Know

Master these essential statistics concepts to excel in data analysis and make data-driven decisions with confidence

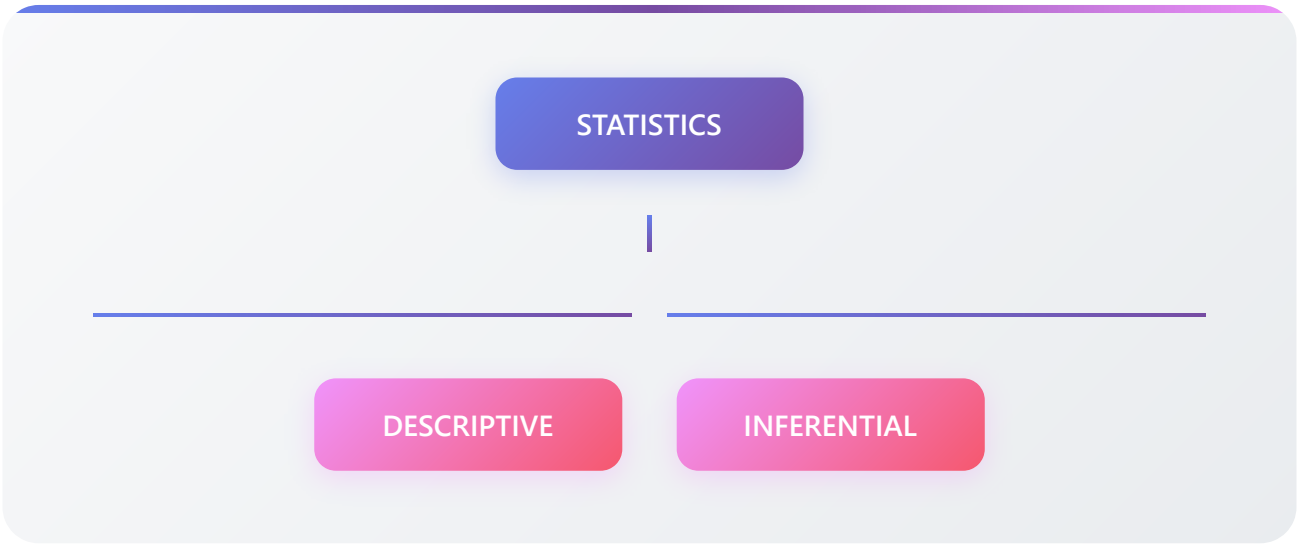
BY Vaibhav Giri

What is Statistics?

The science of collecting, analyzing, interpreting, and presenting data to make informed decisions



Types of Statistics



Descriptive Statistics

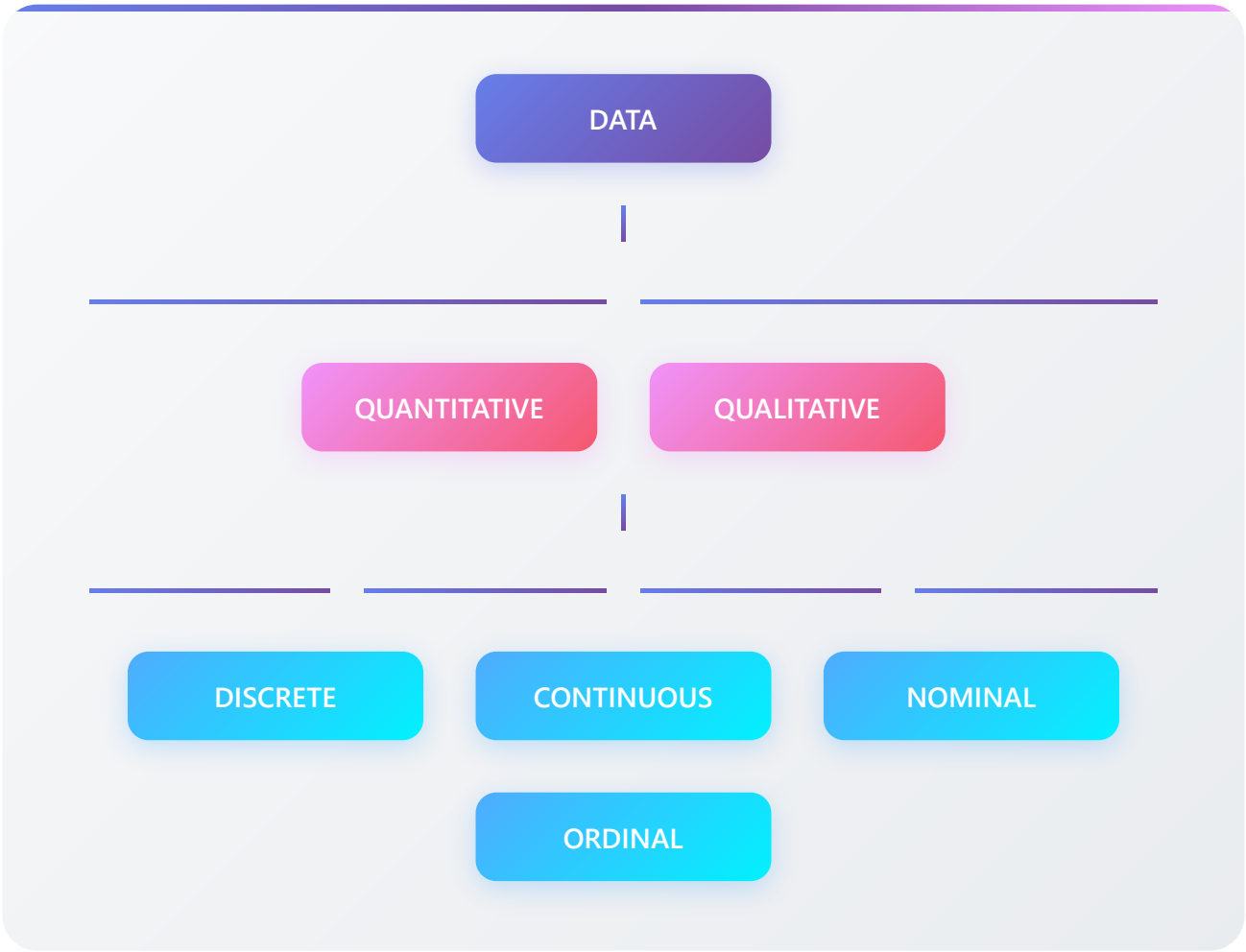
Summarizes and describes the main features of data using measures like mean, median, and standard deviation

→

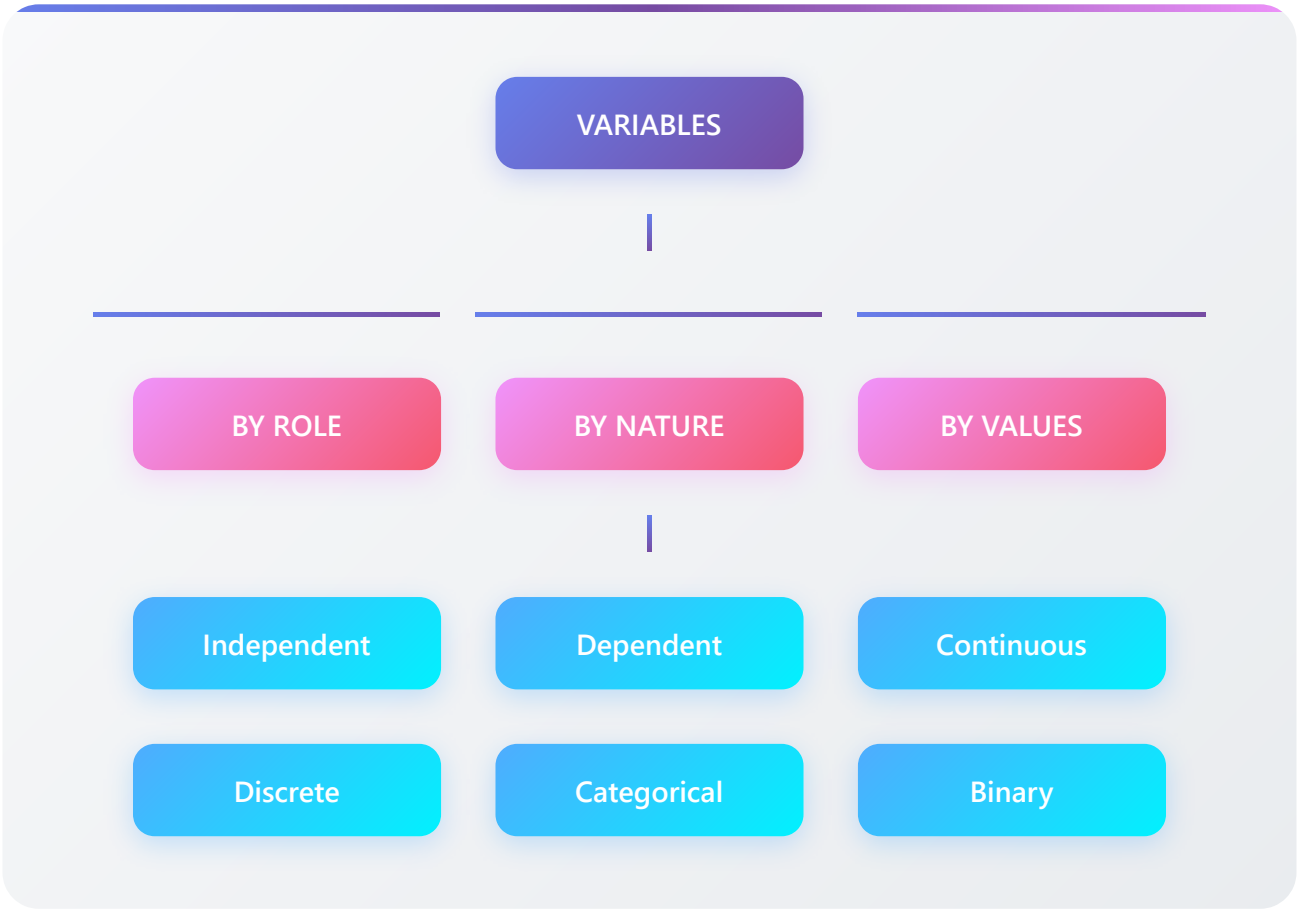
Inferential Statistics

Makes predictions and inferences about populations based on sample data using hypothesis testing

Types of Data



Types of Variables



Population & Sample

POPULATION

N

The complete group of individuals or items that we want to study and draw conclusions about

→

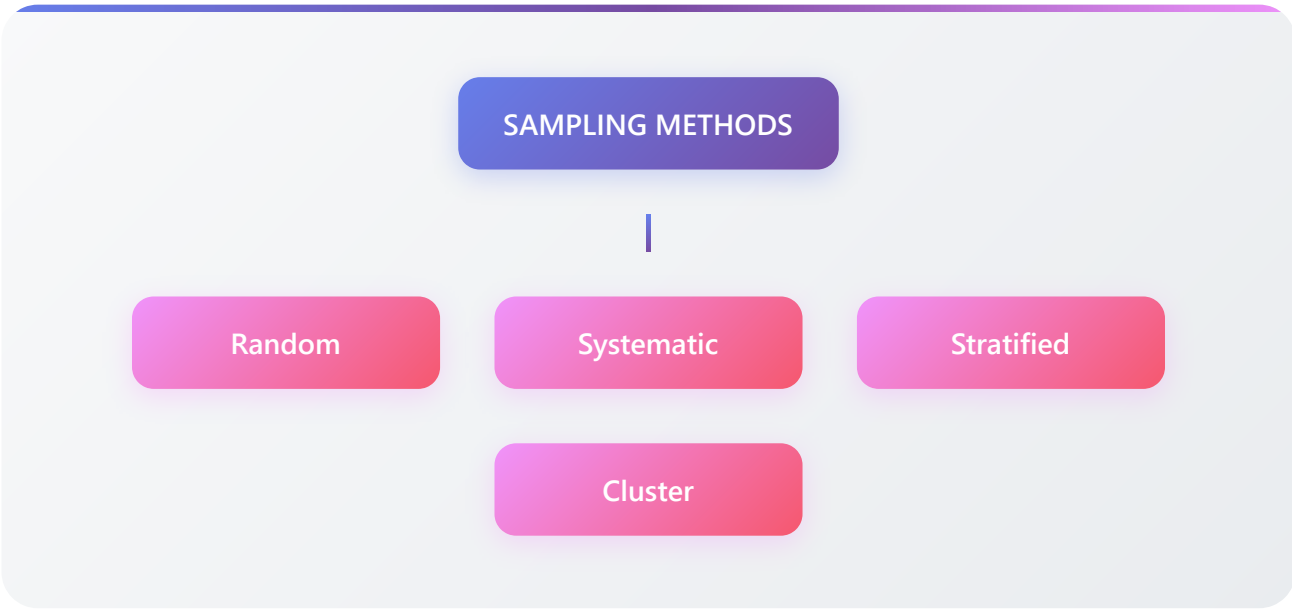
SAMPLE

n

A smaller subset of the population that is selected for analysis and research purposes

We study samples to make inferences about populations because studying entire populations is often impractical or impossible

Sampling Techniques



Random Sampling

Every member has equal chance of selection

Systematic Sampling

Select every nth member from the list

Stratified Sampling

Population divided into strata, then sampled

Measures of Central Tendency

MEAN

$\bar{x}$

Average of all values in the dataset

MEDIAN

M

Middle value when data is arranged in order

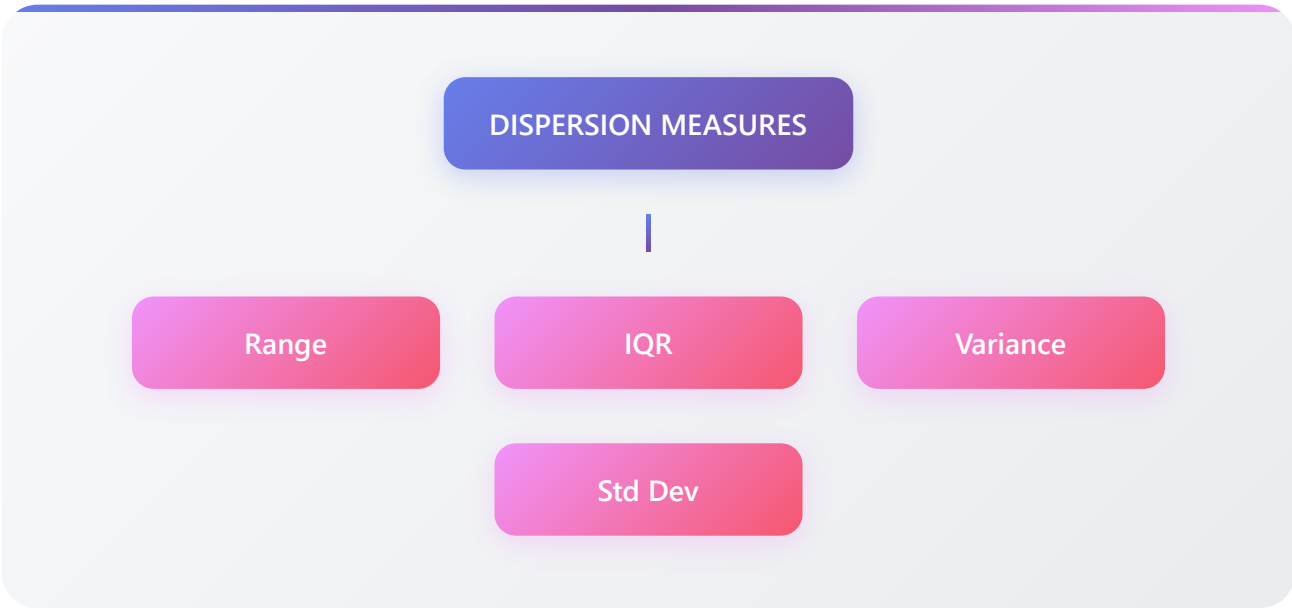
MODE

Mo

Most frequently occurring value

These measures help us understand the center or typical value of our data distribution

Measures of Dispersion



Range

Difference between max and min values

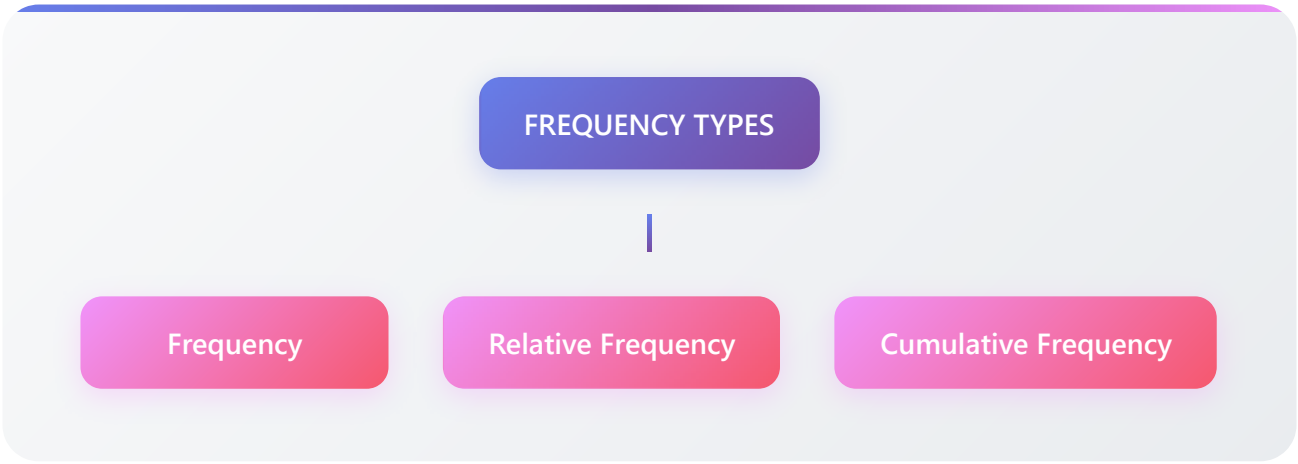
IQR

Interquartile Range: Q3 - Q1

Variance & Std Dev

Measure spread around the mean

Frequency Analysis



FREQUENCY

Count of how often each value occurs in the dataset

→

RELATIVE

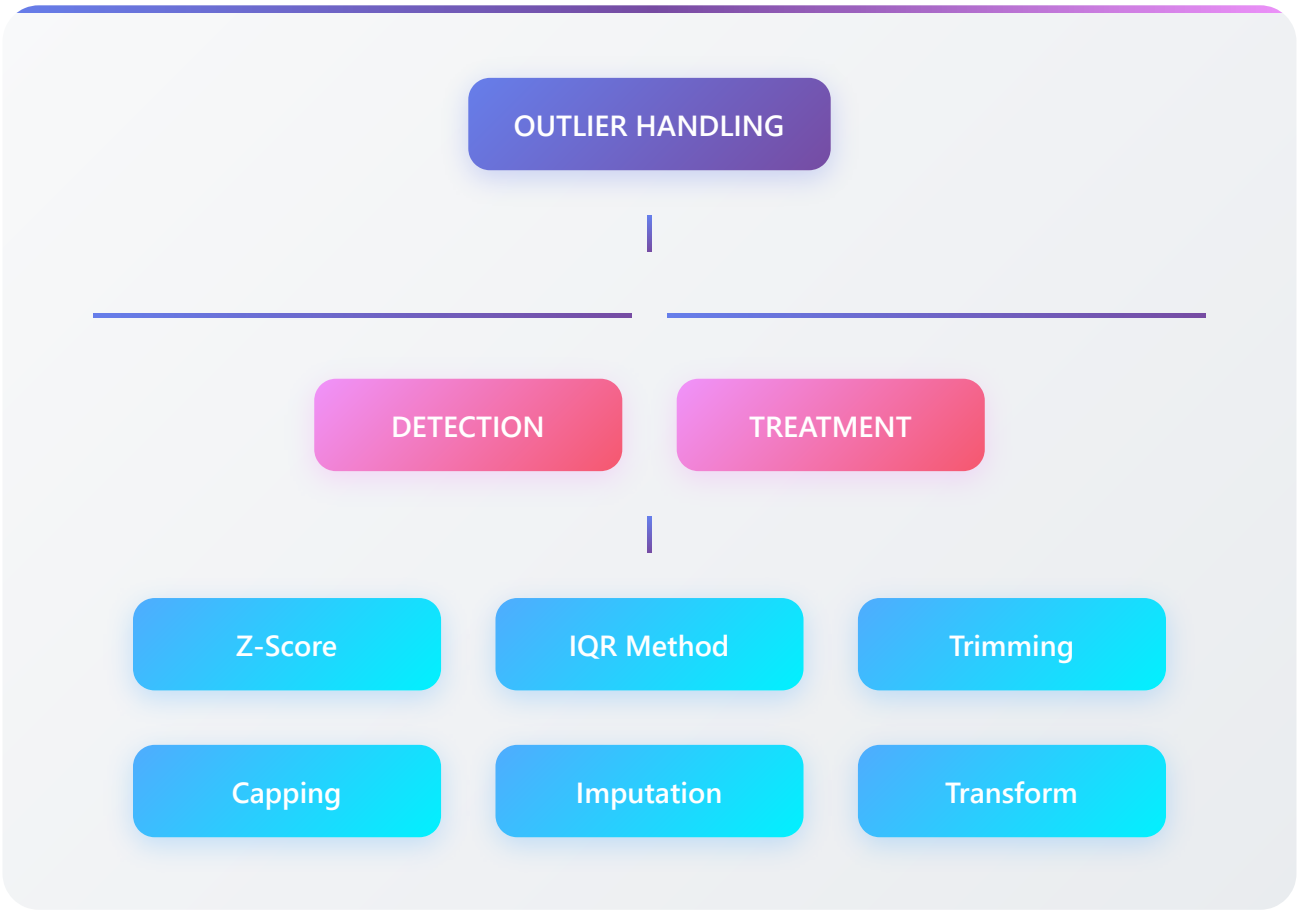
Proportion of each frequency to the total count

→

CUMULATIVE

Running total of frequencies up to each point

Outlier Detection & Treatment



Relationships in Data

RELATIONSHIP MEASURES

- Covariance
- Correlation
- Pearson

COVARIANCE

Shows direction of linear relationship between two variables

VS

CORRELATION

Shows strength and direction of relationship (-1 to +1)

Remember: Correlation does NOT imply Causation

Data Scaling Techniques

STANDARDIZATION

$$Z = (x - \mu) / \sigma$$

Mean = 0, Standard Deviation = 1

VS

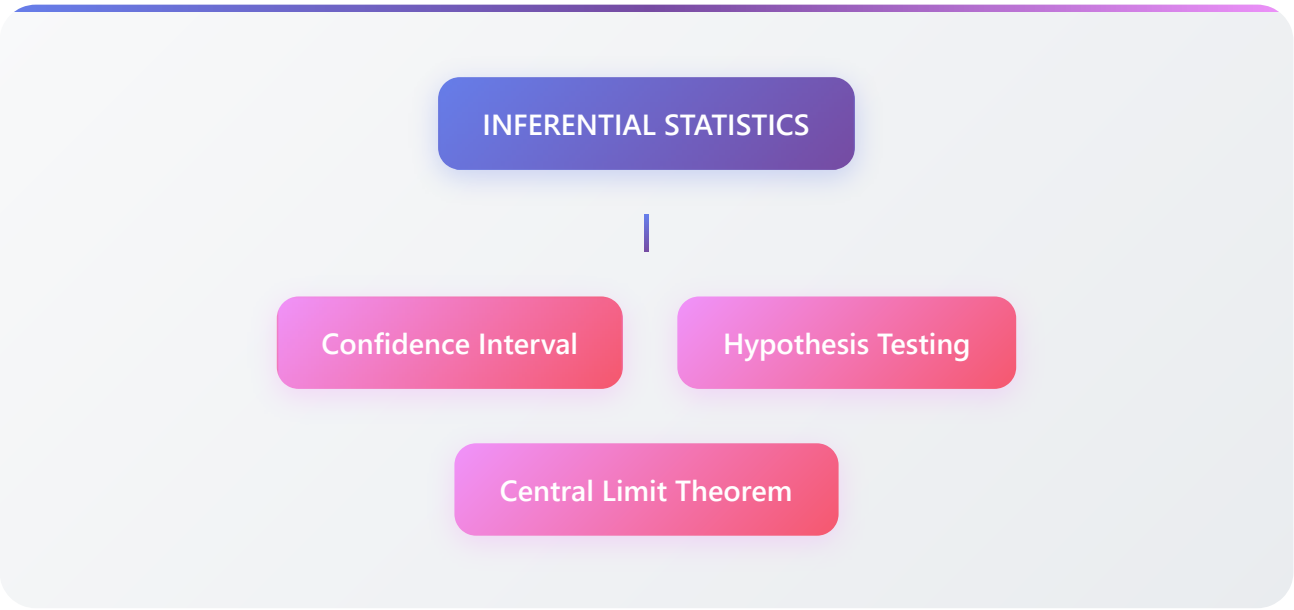
NORMALIZATION

$$(x - min) / (max - min)$$

Scale values to 0-1 range

Essential preprocessing step for machine learning algorithms

Inferential Statistics



Confidence Interval

Range of values likely to contain population parameter

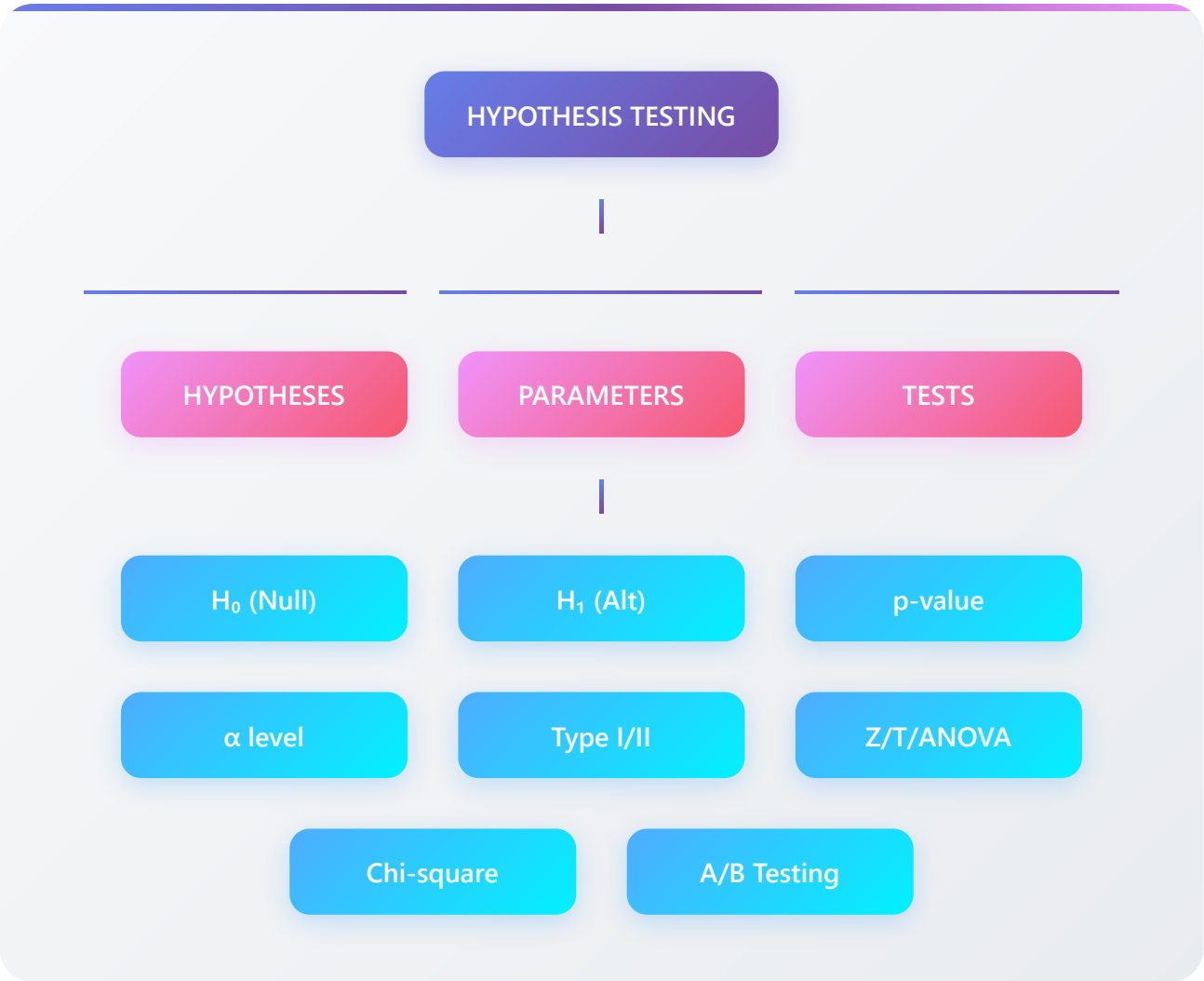
Hypothesis Testing

Statistical method to test assumptions about data

Central Limit Theorem

Foundation of statistical inference

Hypothesis Testing Framework



One-Tailed Test

Tests for effect in one specific direction

Two-Tailed Test

Tests for effect in either direction

Was This Helpful?

If you found these statistics concepts valuable for your data analysis journey, I'd love to hear from you!

👍 Like Comment Share

Thank you for learning with me!

Follow for more data analytics content and insights that will help you excel in your data career.

What statistics concept would you like me to explain next?