Data Science Essentials: Identifying Variable Relationships in Six Steps

Define the Problem and Hypotheses

- Objective and hypotheses
- Example: "Is there a relationship between study hours and exam scores?"

Collect and Prepare Data

- Data collection and cleaning
- Example: "Collect student study hours and exam scores, handle missing data"

Perform Exploratory Data Analysis (EDA)

- Summary statistics and visualizations
- Example: "Plot scatter plot of study hours vs. exam scores, calculate correlation"

Choose and Apply Appropriate Statistical Tests

- Statistical tests and measures
- Example: "Apply Pearson correlation test, calculate correlation coefficient"

Build and Evaluate Models

- Model building and evaluation
- Example: "Develop linear regression model, assess R-squared value"

Interpret Results and Draw Conclusions

- Analysis, interpretation, and recommendations
- Example: "High positive correlation found, recommend study hour guidelines"