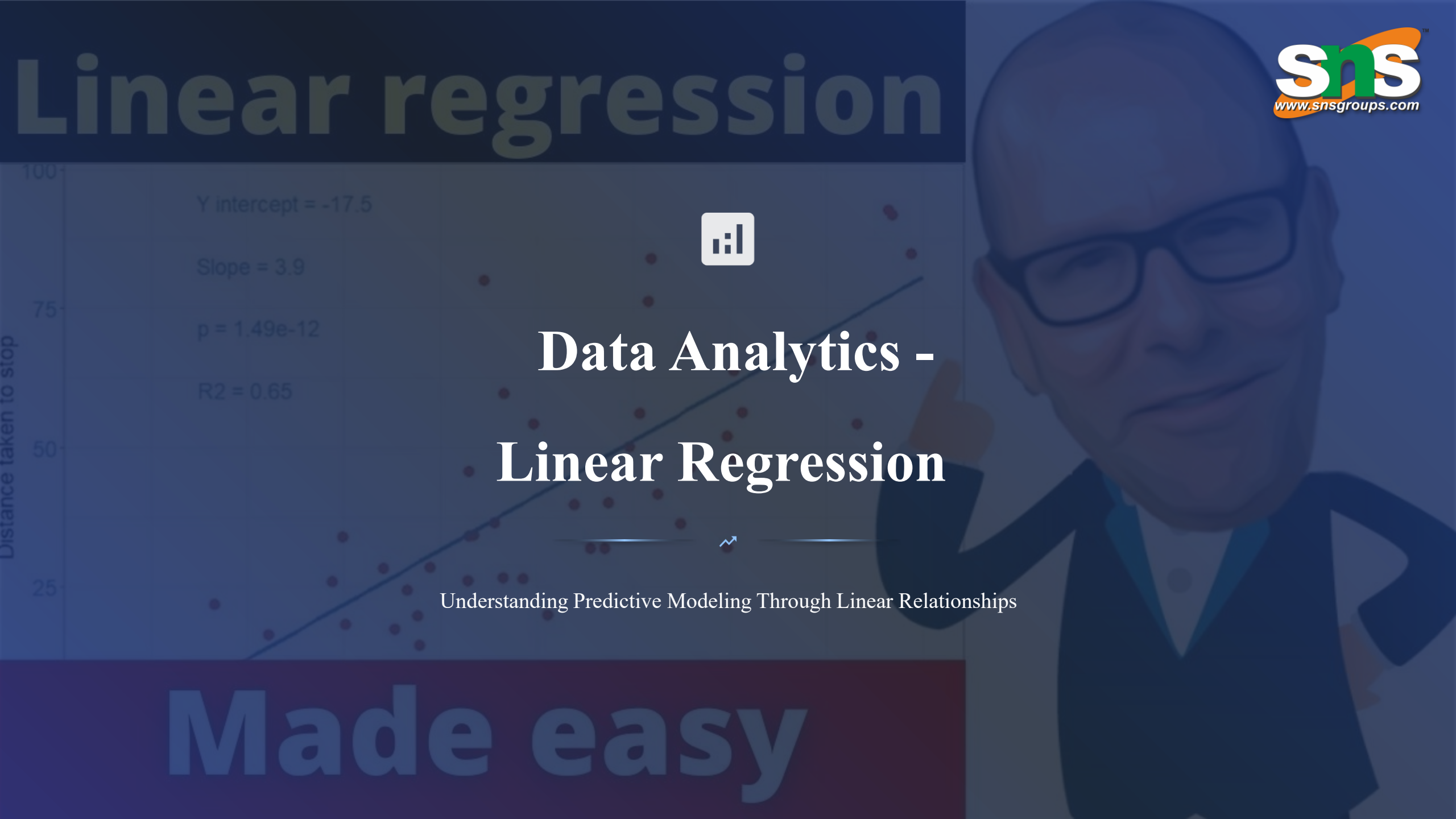


Data Analytics - Linear Regression

Understanding Predictive Modeling Through Linear Relationships

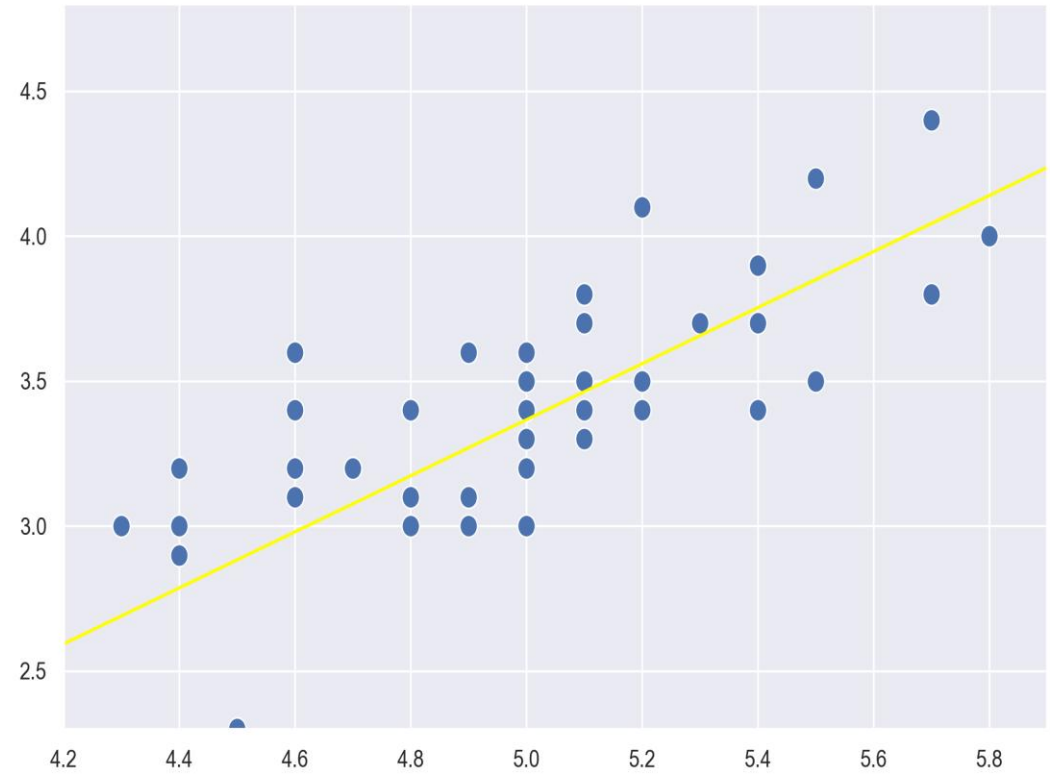
Introduction to Linear Regression

What is Linear Regression?

- Statistical method to model relationship between variables
- Predict future outcomes based on historical data

Purpose in Data Analytics

- Identify trends in data
- Make accurate predictions
- Support decision-making processes



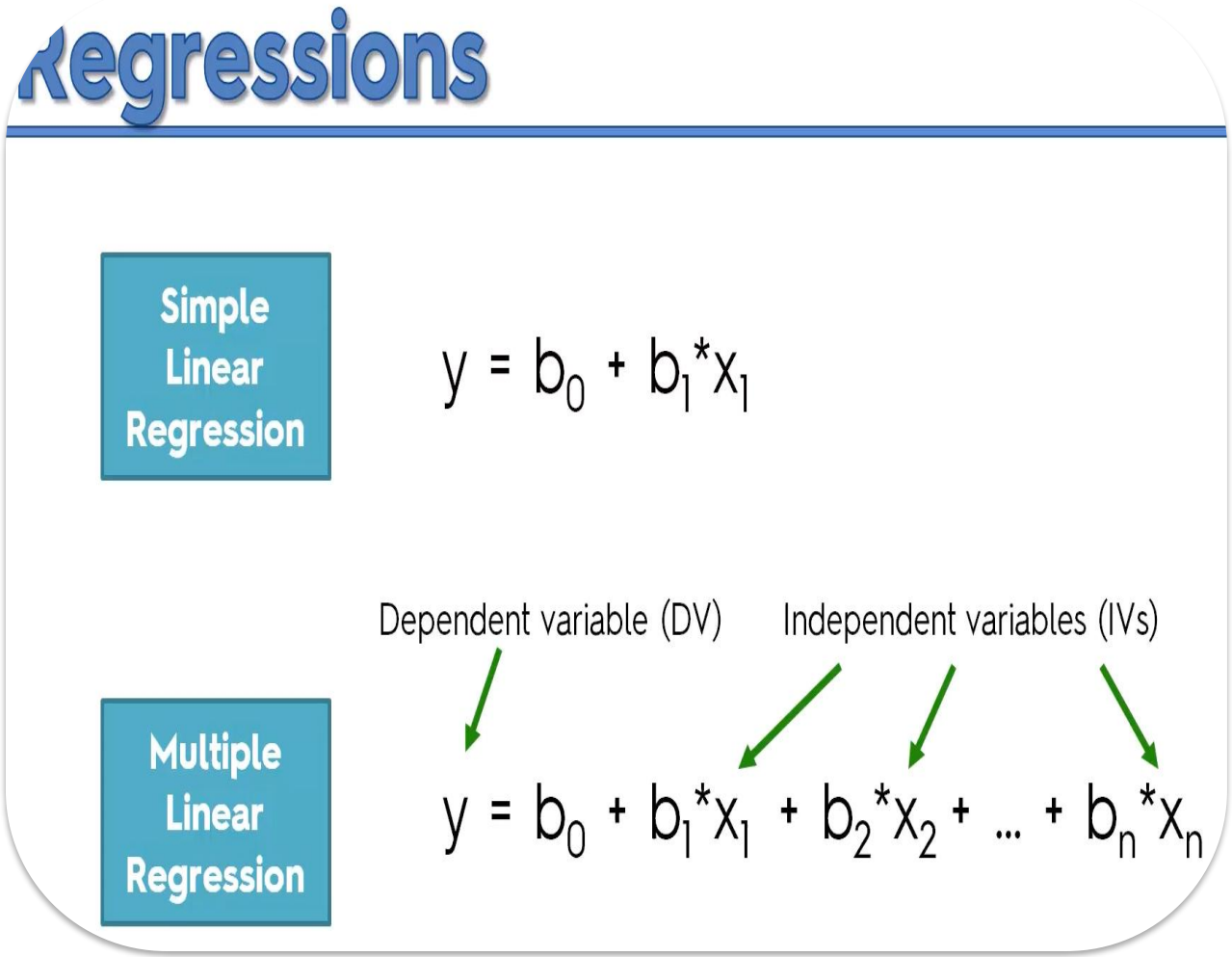
Types of Linear Regression

Simple Linear Regression

- One independent variable (X)
- Equation: $y = mx + b$

Multiple Linear Regression

- Multiple independent variables (X_1, X_2, ..., X_n)
- Equation: $y = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$



Linear Regression Equation and Components

Main Equation

$$y = mx + b$$

Components

y — **Dependent Variable**

The value we predict

m — **Slope**

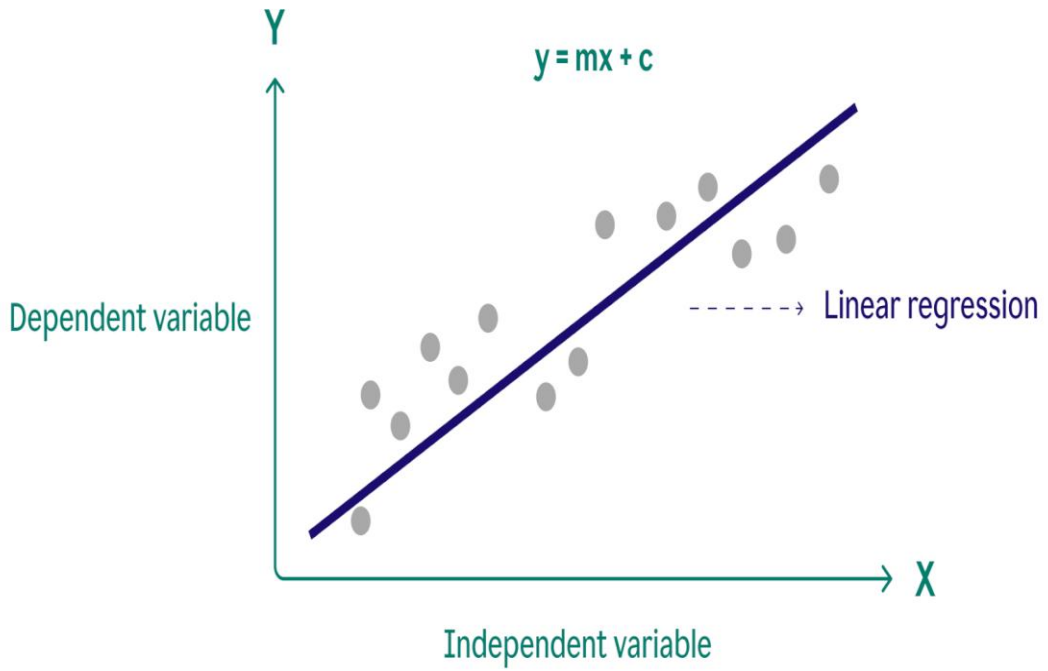
Rate of change

x — **Independent Variable**

The input value

b — **Y-Intercept**

Starting value



Applications of Linear Regression

Real-world Applications

Business

Sales forecasting, Demand prediction

Finance

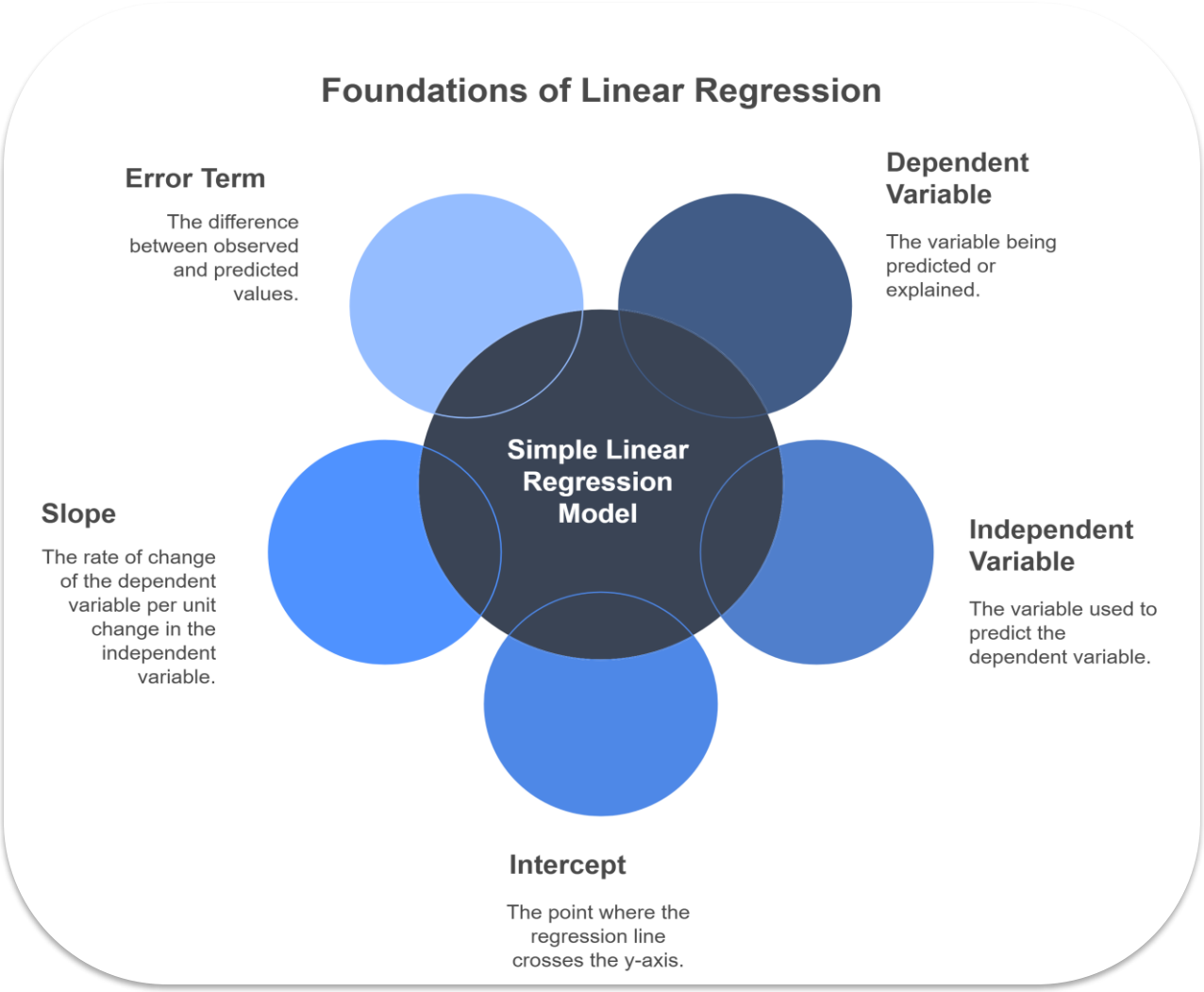
Stock price analysis, Risk assessment

Healthcare

Patient outcome prediction, Disease progression

Science

Experimental data analysis, Trend identification



Assumptions of Linear Regression

Key Assumptions

1. **Linearity**

   Relationship between variables is linear

2. **Independence**

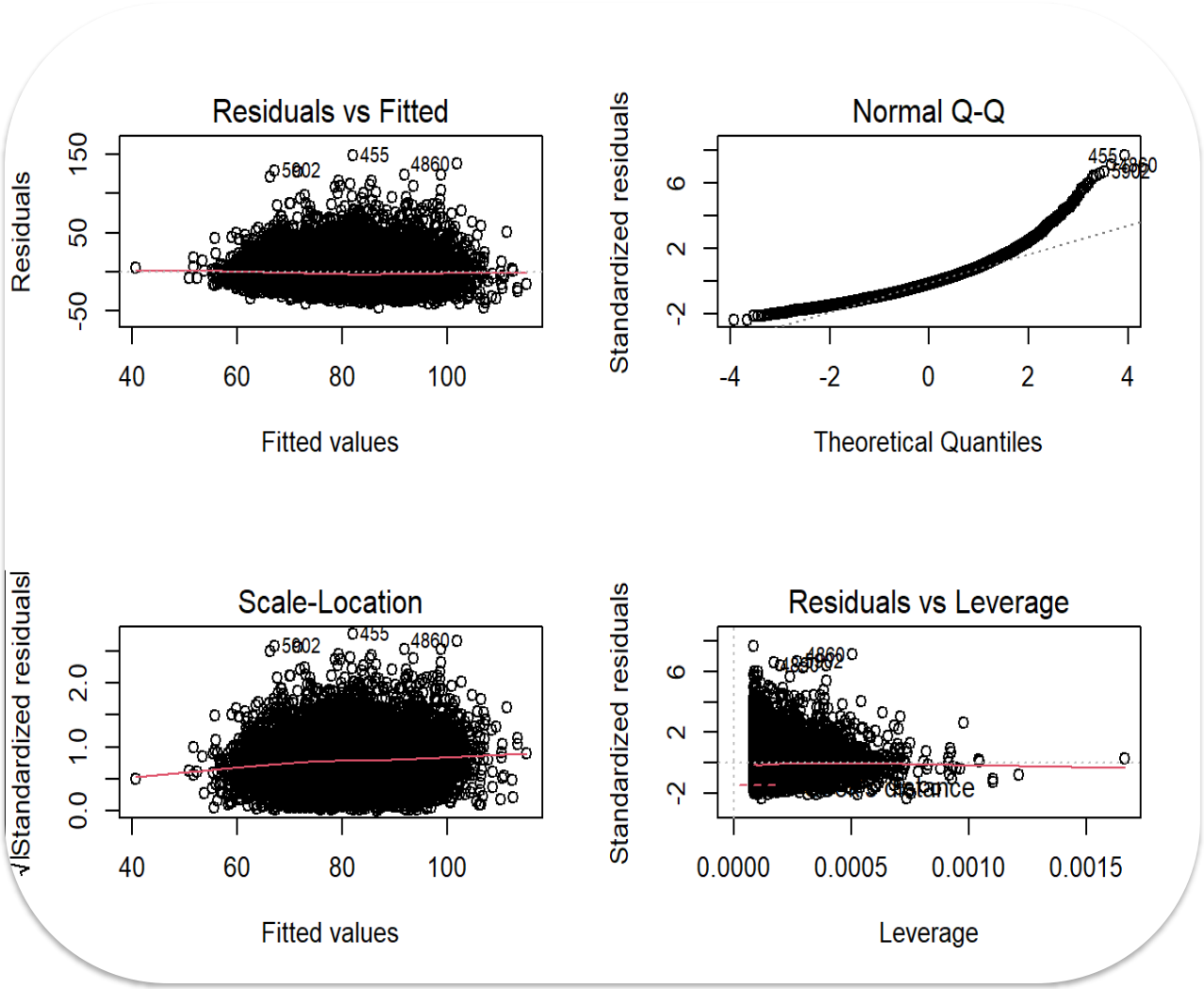
   Observations are independent of each other

3. **Homoscedasticity**

   Constant variance of residuals

4. **Normality**

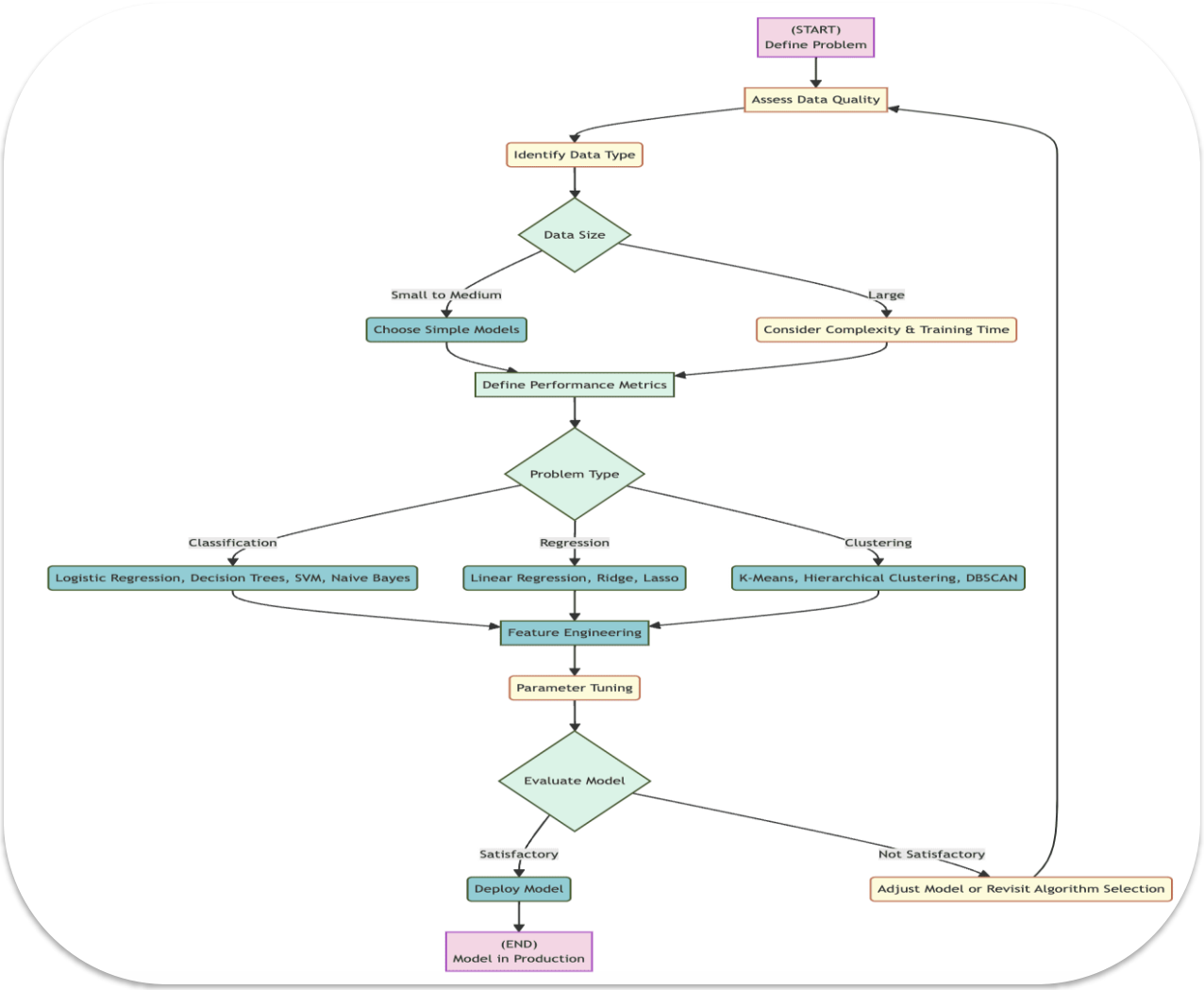
   Residuals are normally distributed



Steps in Linear Regression Analysis

Process Flow

1. **Data Collection**
   Gather relevant dataset
2. **Data Preprocessing**
   Clean and prepare data
3. **Model Training**
   Fit the regression model
4. **Model Testing**
   Validate with test data
5. **Model Evaluation**
   Assess model performance
6. **Prediction**
   Make future predictions



Evaluation Metrics

Key Metrics

R-squared (R^2)

Measures goodness of fit

- Range: 0 to 1
- Higher is better

Mean Squared Error (MSE)

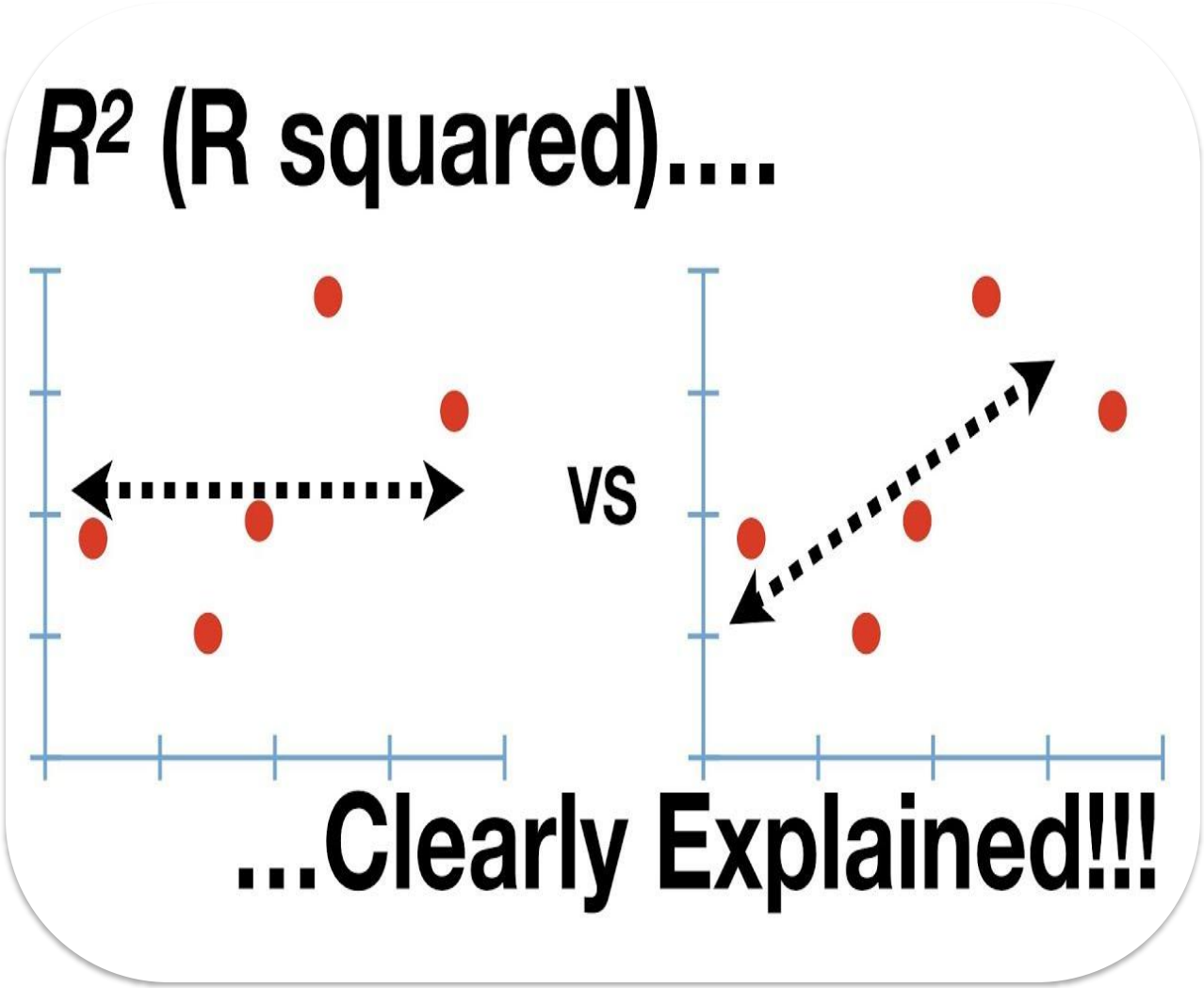
Average of squared differences

- Lower is better

Root Mean Squared Error (RMSE)

Square root of MSE

- Interpretable in target units



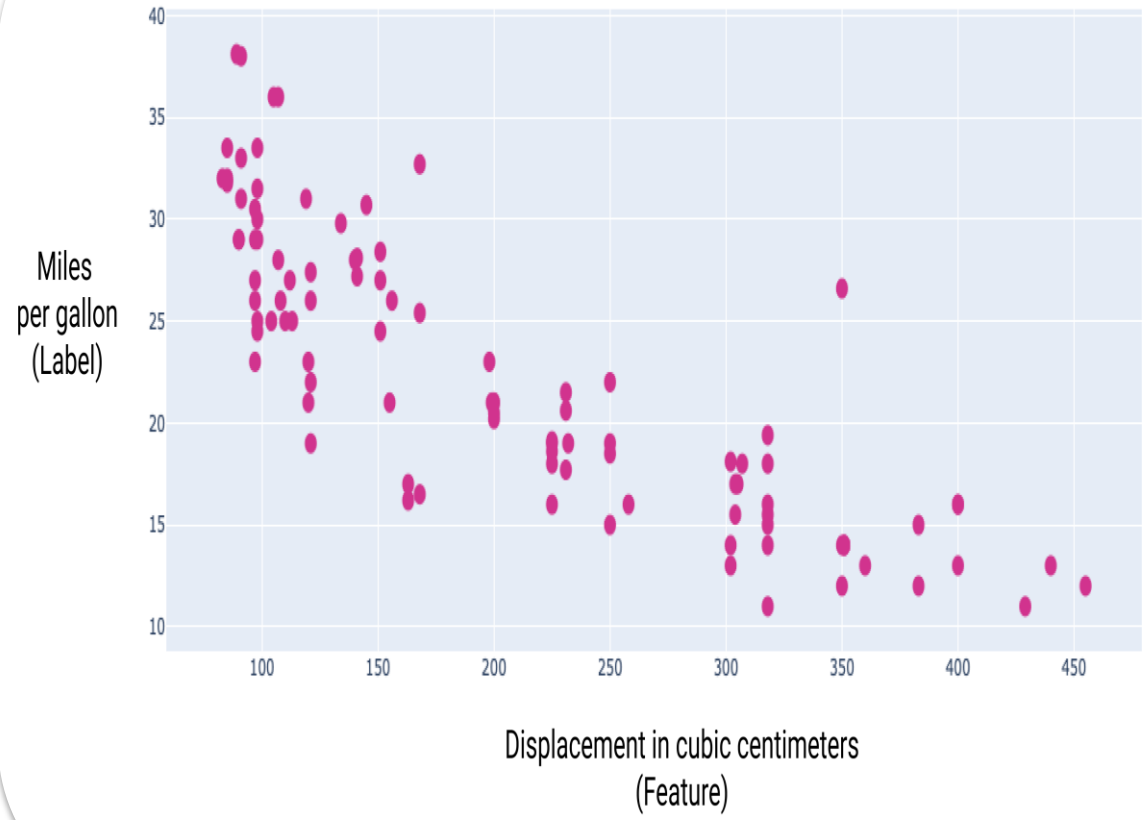
Advantages and Limitations

Advantages

- Simple to understand and interpret
- Efficient for small datasets
- Works well with linear relationships

Limitations

- Assumes linear relationship
- Sensitive to outliers
- Cannot handle non-linear patterns
- May overfit with many variables



Mind Map with Recap and Summary

Key Takeaways

- Linear regression models relationships between variables
- Equation y = mx + b forms the foundation
- Multiple types: Simple & Multiple Regression
- Widely applied across business, finance, healthcare & science
- Four key assumptions must be met for valid results
- Process: Collection → Training → Evaluation → Prediction
- R^2, MSE, RMSE measure model performance
- Balance advantages with limitations for effective use