Regression Analysis Fundamentals

Welcome to our comprehensive exploration of regression analysis. This powerful statistical technique helps us understand relationships between variables and make predictions from data.

by S MM

Purpose and Applications

Scientific Research

Identifies causal relationships between variables. Used in medicine, psychology, and environmental studies.

Business Forecasting

Predicts future sales, stock prices, and market trends. Essential for strategic planning.

Quality Control

Monitors manufacturing processes. Detects defects and optimizes production variables.

Simple Linear Regression

Key Components

- Dependent variable (y)
- Independent variable (x)
- Slope (β_1)
- Intercept (β_0)

The Equation

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Where ε represents random error in the model. The line minimizes squared distances from data points.

Line of Best Fit

Ordinary Least Squares (OLS)

Minimizes the sum of squared residuals. The standard method for fitting regression lines.

Geometric Interpretation

The line passes through the centroid ($\bar{x}$,$\bar{y}$) of the data points.

Outlier Sensitivity

Extreme points can significantly pull the line. Requires careful data examination.

Hands-On Implementation

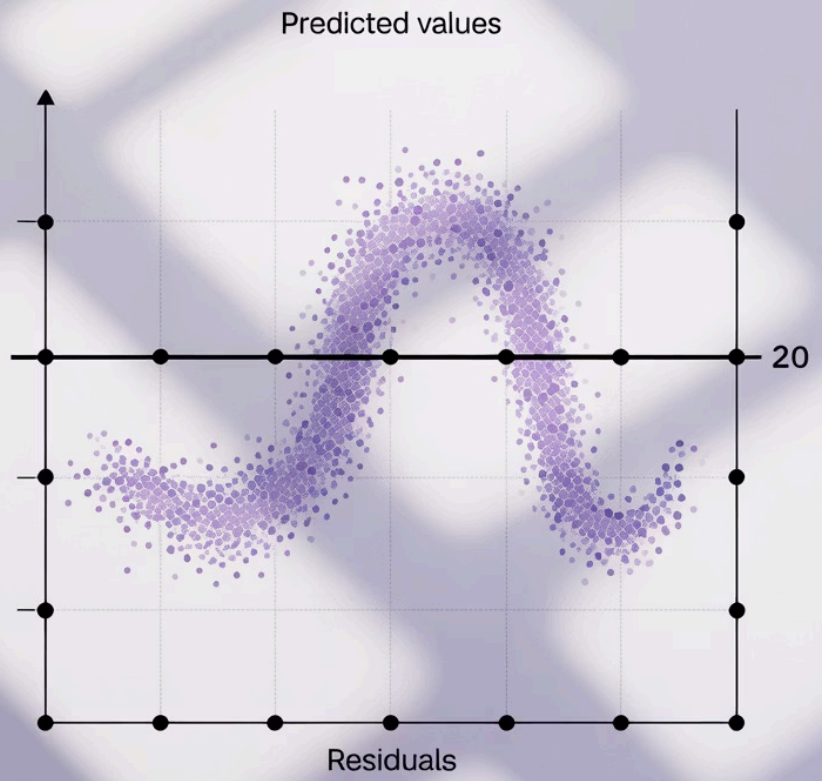
Excel Implementation

1. Enter data in two columns
2. Select Data Analysis add-in
3. Choose Regression tool
4. Specify input ranges
5. Interpret results table

Python Implementation

```
import numpy as np
import statsmodels.api as sm

X = sm.add_constant(x)
model = sm.OLS(y, X)
results = model.fit()
print(results.summary())
```



Evaluating Model Accuracy

R^2

Coefficient of Determination

Measures proportion of variance explained by the model. Ranges from 0 to 1.

RMSE

Root Mean Square Error

Average magnitude of residuals. Lower values indicate better fit.

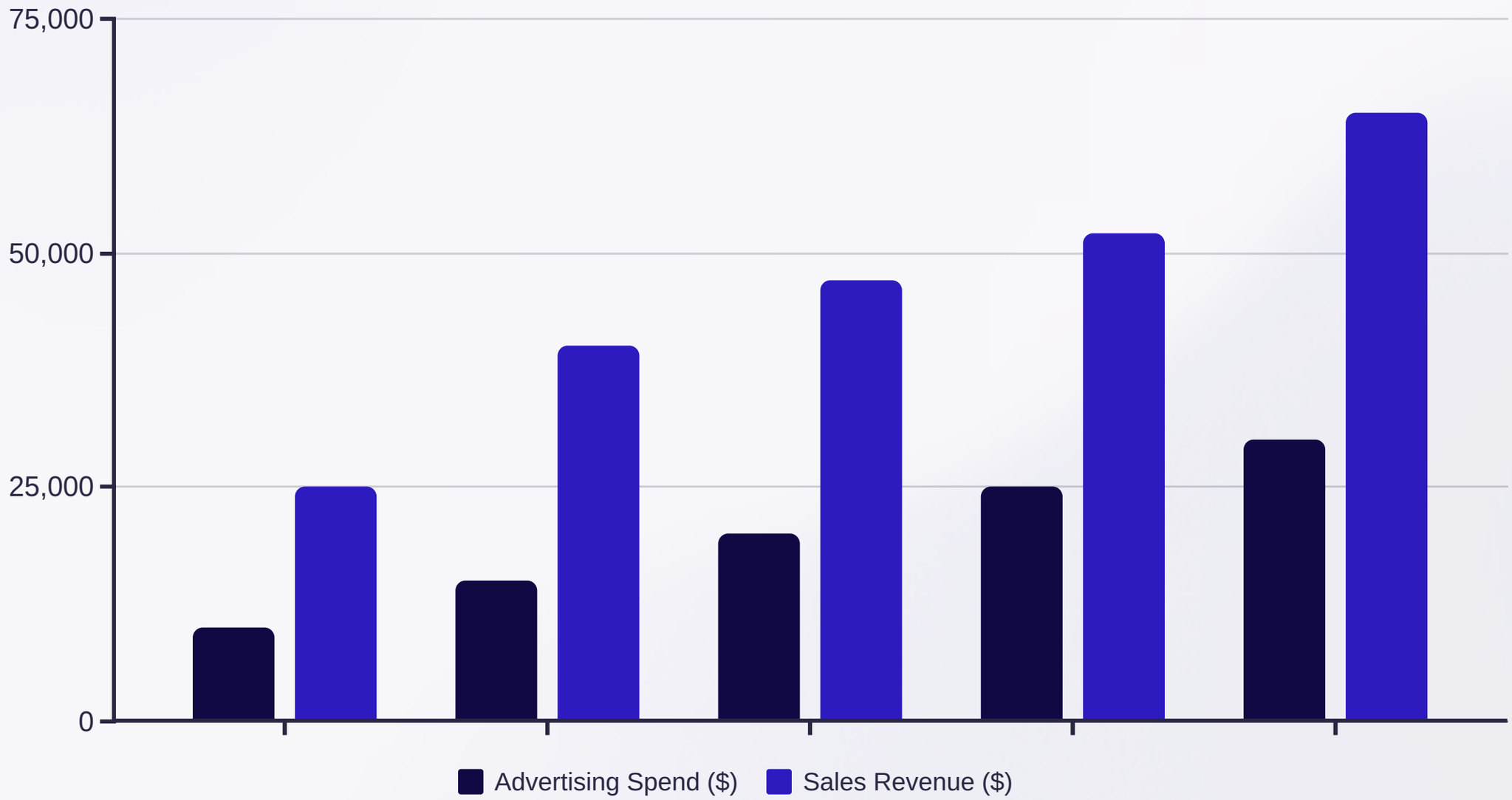
p

p-value

Tests statistical significance. Values below 0.05 suggest significant relationships.

Residual analysis reveals patterns that might indicate model inadequacies or violations of assumptions.

Mini Project: Advertising & Sales



Project goal: Build a model predicting sales revenue based on advertising expenditure. Assess diminishing returns on ad spend.

Key Takeaways & Next Steps

Apply What You've Learned

Complete the advertising-sales mini project. Submit your regression analysis and interpretation.

Prepare for Multiple Regression

Next week, we'll extend our models to include multiple predictors. Review matrix notation.

Explore Further

Examine assumptions behind linear regression. Consider logistic regression for categorical outcomes.