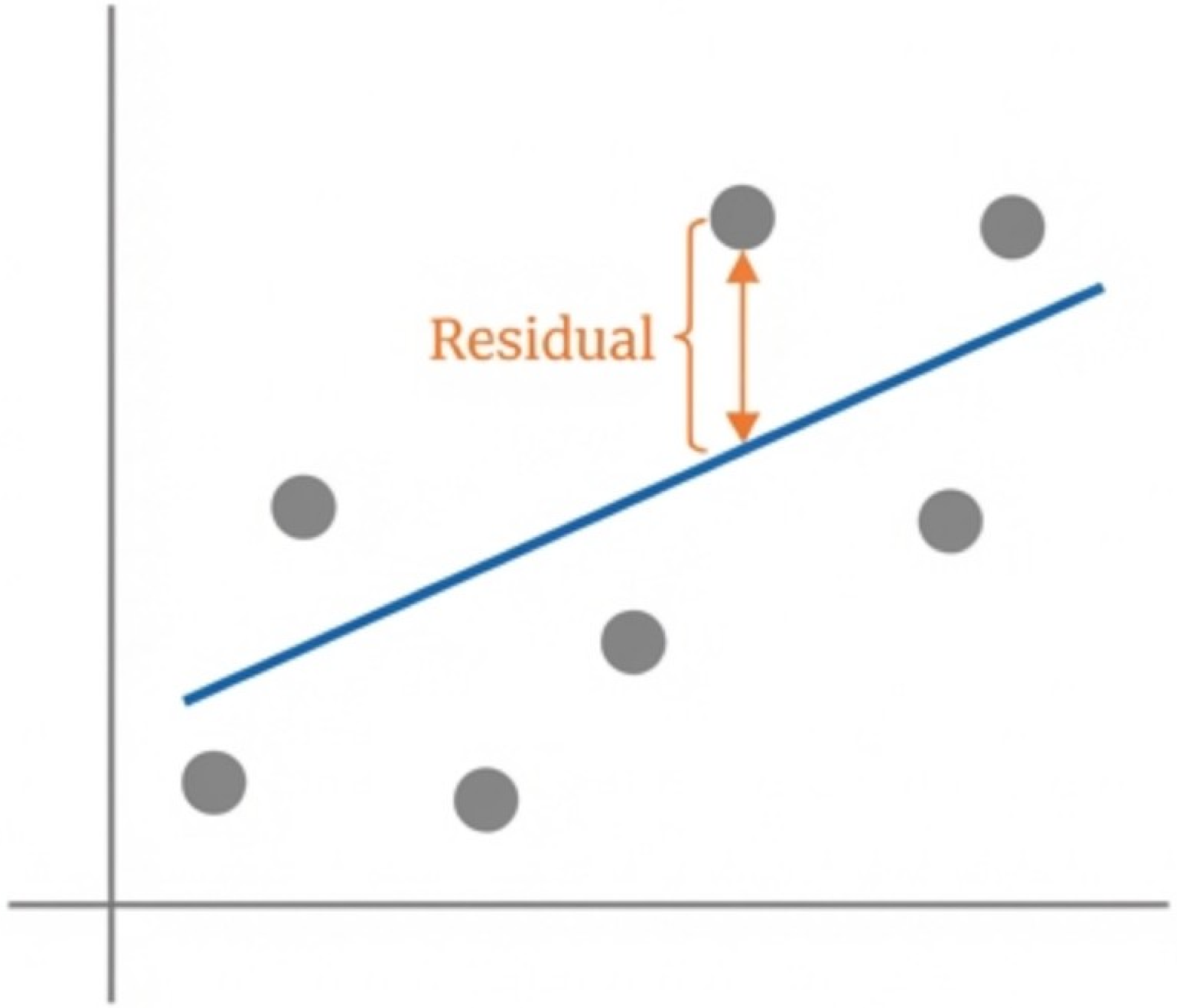
ANOVA and Standard Error of Estimate in Simple Linear Regression

A Quantitative Methods Review for UGC NET Management & Finance Aspirants



CONCEPT-DRIVEN PEDAGOGY | QUANTITATIVE ANALYSIS REVIEW

The Core Concept: Partitioning Variability

Analysis of Variance (ANOVA) partitions the total variability of a variable into components ascribed to different sources. In regression, we ask: How much of the movement in Y is due to our model (Signal), and how much is just random error (Noise)?

Total Variation (SST) = Explained Variation (RSS) + Unexplained Variation (SSE)

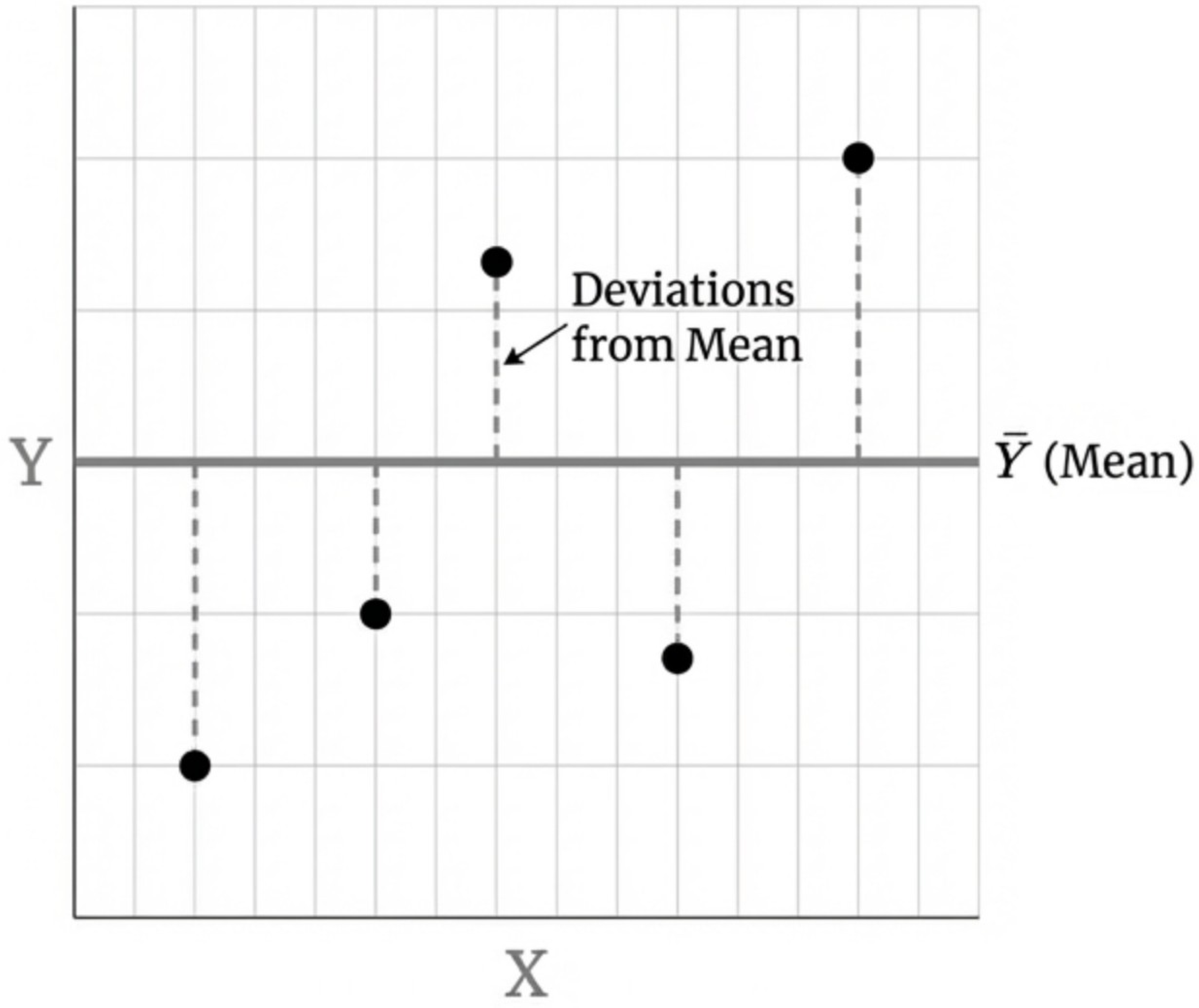
$$\mathrm{SST} = \mathrm{RSS} + \mathrm{SSE}$$

Measuring Total Variation (SST)

Total Sum of Squares (SST) measures the total variation of the dependent variable. It assumes no relationship with X, using the Mean as the best naive predictor.

$$\text{SST} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

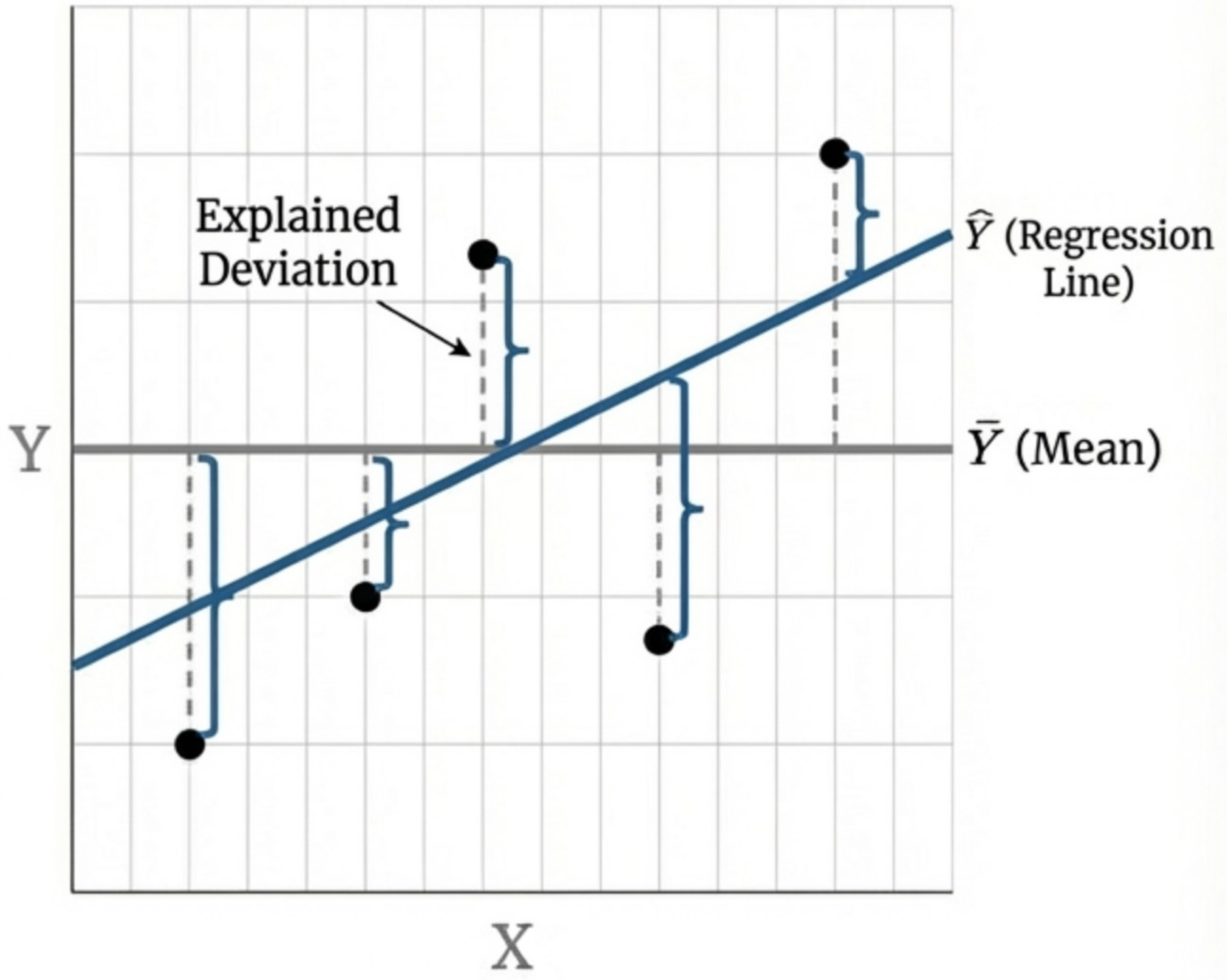
Y_i = Actual observed value
$\bar{Y}$ = Mean of y-observations



The ‘Explained’ Variation (RSS)

Regression Sum of Squares (RSS) measures the variation in the dependent variable that is explained by the independent variable. A large RSS implies the model is effective.

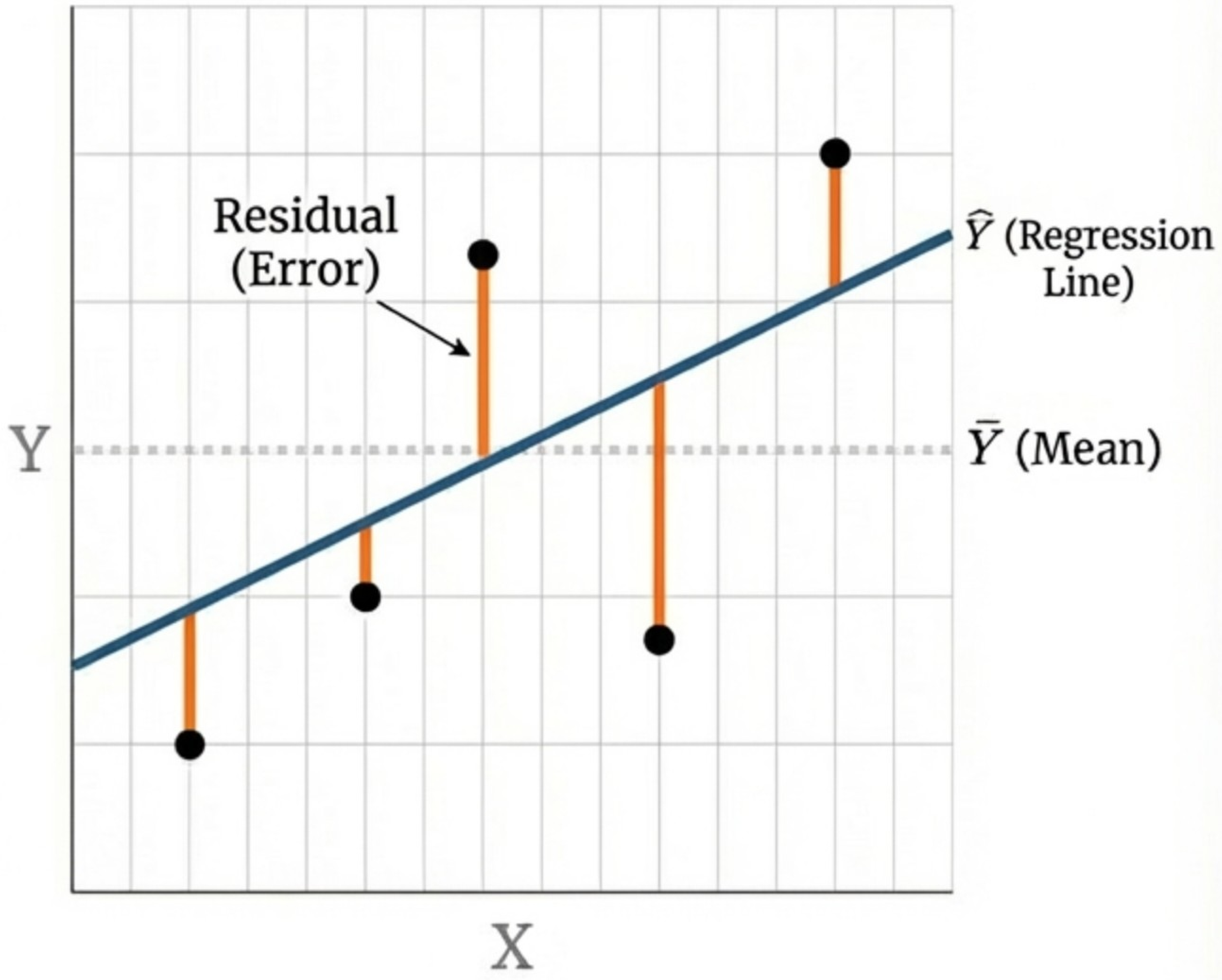
$$\mathrm{RSS} = \sum_{i=1}^{n} (\widehat{Y}_i - \bar{Y})^2$$



The 'Unexplained' Variation (SSE)

Sum of Squared Errors (SSE), or Residual Sum of Squares, measures the variation unexplained by the model. The Least Squares Method minimizes this value.

$$\mathrm{SSE} = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$



The Coefficient of Determination (R^2)

R^2 represents the proportion of total variation that is explained by the model. It ranges from 0 to 1, where values closer to 1.0 imply high predictive effectiveness.

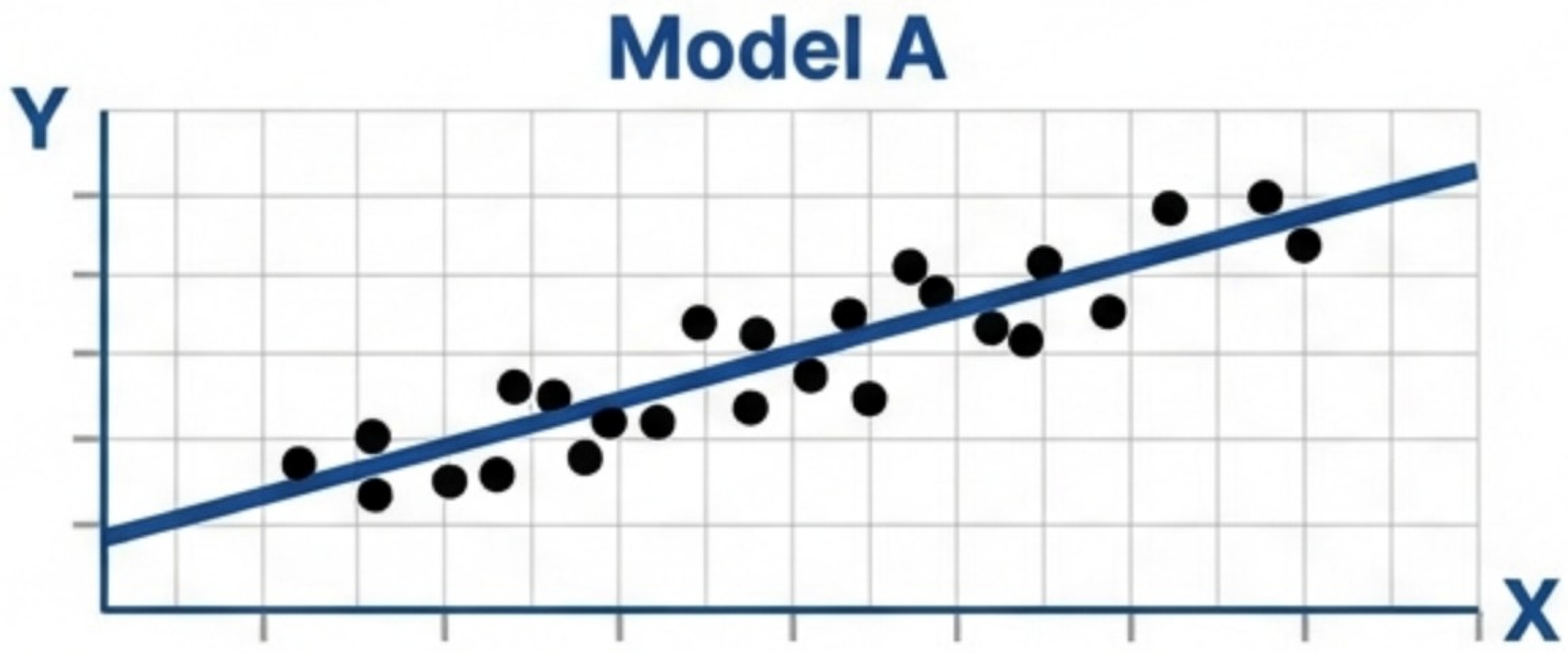
$$R^2 = \frac{\text{Explained Variation (RSS)}}{\text{Total Variation (SST)}}$$

$$R^2 = \frac{\text{SST} - \text{SSE}}{\text{SST}}$$

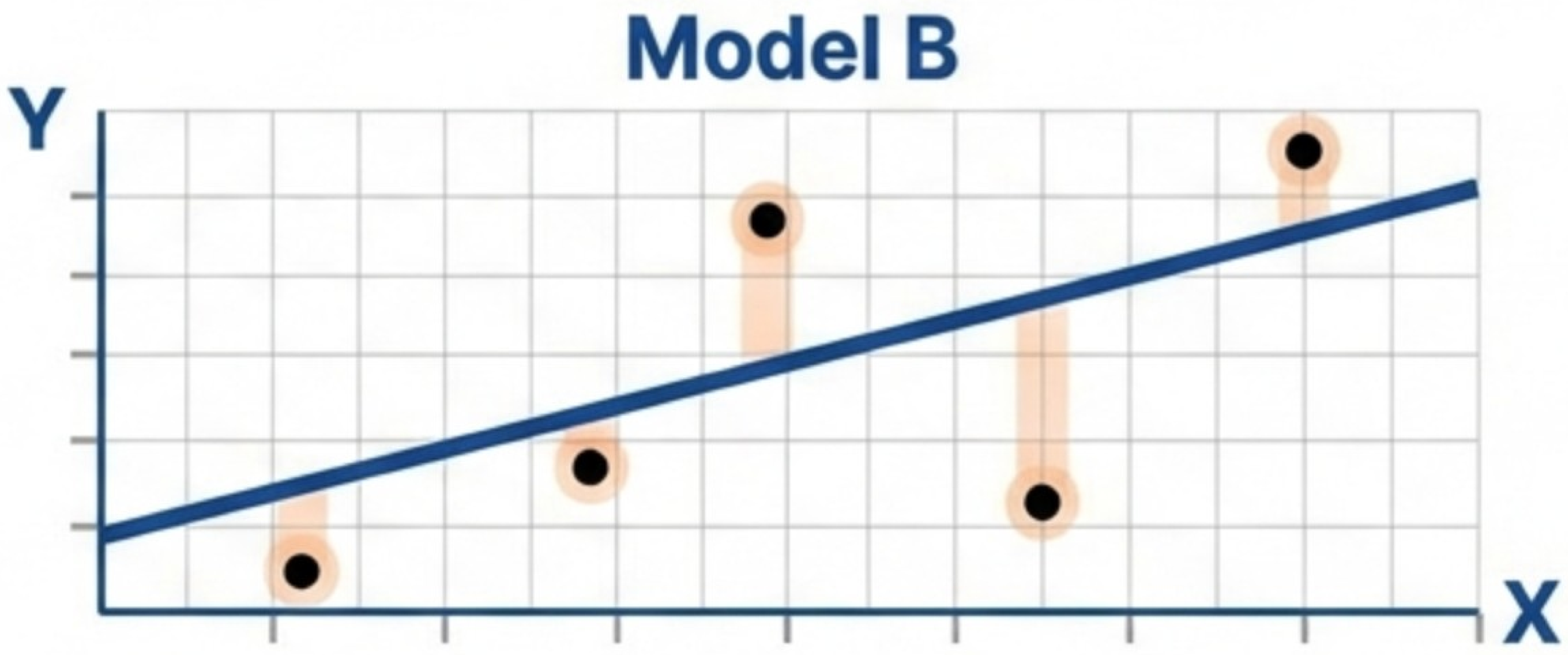
Standard Error of Estimate (S_e)

Also known as 'Standard Error of the Regression.' It measures the average distance of observed values from the regression line. Unlike R^2, this is an absolute measure in the units of the dependent variable.

$$S_e = \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}}$$



Low S_e (Better Fit)



High S_e (Poorer Fit)

The Standard ANOVA Table Architecture

Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Sum of Squares (MSS)
Regression (Explained)	1	RSS	MSR
Residual (Unexplained)	$n-2$	SSE	MSE
Total	$n-1$	SST	

n = Number of observations. k = 1 (slope parameter).

From Sums to Averages (MSR & MSE)

To fairly compare Explained vs. Unexplained variation, we must adjust for Degrees of Freedom.

$$\boxed{\begin{array}{c} RSS \\ \text{(Explained Sum)} \end{array}} \xrightarrow[\text{(df)}]{\text{Divide by 1}} \boxed{\begin{array}{c} MSR \\ \text{(Mean Regression SS)} \end{array}}$$

$$\boxed{\begin{array}{c} SSE \\ \text{(Unexplained Sum)} \end{array}} \xrightarrow[(n-2)]{\text{Divide by}} \boxed{\begin{array}{c} MSE \\ \text{(Mean Squared Error)} \end{array}}$$

MSE is effectively the variance of the residuals.

The F-Test Hypothesis

We test if the slope coefficient is statistically different from zero.

Null Hypothesis (H_0):

$$b_1 = 0$$

The independent variable explains nothing. The slope is flat.

Alternative Hypothesis (H_a):

$$b_1 \neq 0$$

The independent variable has explanatory power. The slope is significant.

If Explained Variation (MSR) $\gg$ Unexplained Variation (MSE), we reject H_0.

Calculating the F-Statistic

The F-Statistic is the ratio of Signal to Noise.

$$F = \frac{\text{Mean Regression Sum of Squares (MSR)}}{\text{Mean Squared Error (MSE)}}$$

$$F = \frac{RSS/1}{SSE/(n-2)}$$

Academic al

High F-Stat → Strong Signal → Effective Model

Alert Orange

F-Stat approx 0 → Strong Noise → Ineffective Model

Making the Decision

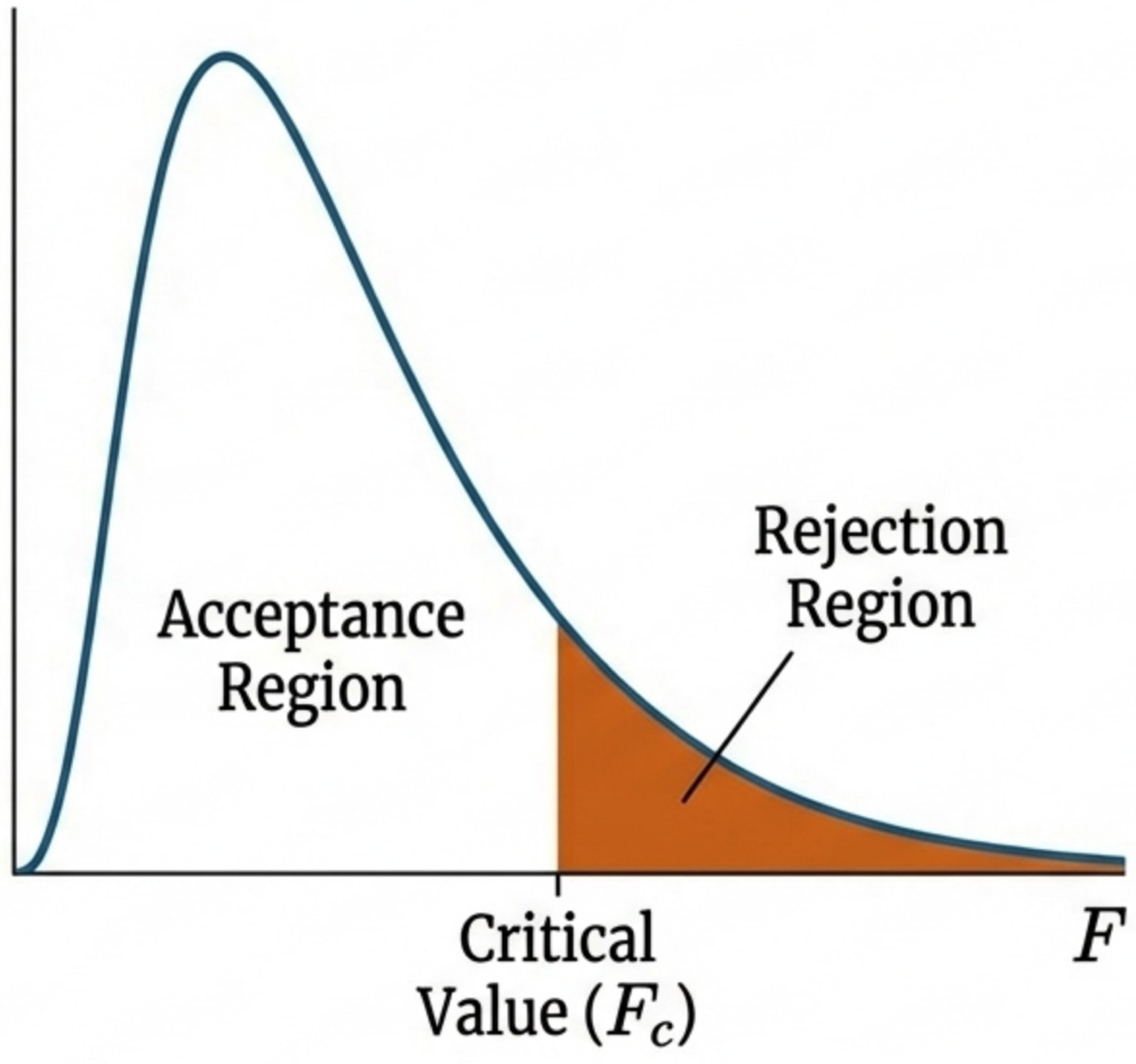
Step 1: Look up Critical F (F_c) for df $(1, n-2)$.

Step 2: Compare Calculated F vs. Critical F.

Decision Rule:

If Calculated $F >$ Critical F:
  REJECT H_0 (Significant)

If Calculated $F <$ Critical F:
  FAIL TO REJECT H_0



Applied Example: Inflation vs. Unemployment

- Observations (n) = 10
- RSS (Explained) = 0.004249
- **SSE** (Unexplained) = 0.002686

Source	df	SS	MSS
Regression	1	0.004249	0.004249
Residual	8	0.002686	0.00033575
Total	9	0.006935	

Calculating F and Conclusion

Step 1 (Calculation):

$$F = \frac{MSR}{MSE} = \frac{0.004249}{0.0001492} \approx 28.48$$

Step 2 (Comparison):

Critical F-Value (df 1, 8) at 5% Significance: 5.32

28.48 > 5.32

Step 3 (Conclusion):

Result: Reject Null Hypothesis. The relationship between inflation and unemployment is statistically significant.

Special Case: F-Test vs. t-Test

In Simple Linear Regression (one independent variable), the **F-test** and t-test provide the exact same conclusion.

$$F = t^2$$

From previous example: $t = 2.306$
$2.306^2 \approx 5.32$

Exam Note: This equality holds **ONLY** for simple regression. In Multiple Regression, F measures the overall model, while t measures individual coefficients.