Deriving the Least Squares Method

Loss Function: Sum of Squared Errors

$$L(\beta_0, \beta_1) = \Sigma_i (y_i - \beta_0 - \beta_1 x_i)^2$$

SSE: Sum of Squared Errors

Why Squares?

- Penalizes large errors more heavily
- Mathematically convenient (differentiable)
- No cancellation of positive/negative errors

Optimization Goal

Find $\hat{\beta}_0$, $\hat{\beta}_1$ that minimize L

Key Requirement

Unique solution exists when
X has full rank

Statistical Connection

Derivation Steps

1 **Define Loss Function**

$$L(\beta_0, \beta_1) = \Sigma(y_i - \beta_0 - \beta_1 x_i)^2$$

↓

2 **Take Partial Derivatives**

$$\partial L/\partial \beta_0 = 0$$

$$\partial L/\partial \beta_1 = 0$$

↓

3 **Solve System of Equations**

Normal Equations

↓

✓ Solution

Optimal parameters $\hat{\beta}_0$, $\hat{\beta}_1$
that minimize prediction error

Least Squares = Maximum Likelihood
under normal errors