What is Linear Regression?

The Oridnary Least Squares Estimators

Suppose Y_i are continuous and we want to model $E[Y_i|X_i]$.

A linear regression models takes

$$E[Y_i|X_i] = X_i\beta.$$

We take

$$\widehat{\beta} = (X'X)^{-1}X'Y,$$

and call these **ordinary least squares (OLS)** estimators.

OLS Estimators (Two Ways)

If $Y_i|X_i \sim N(X_i\beta, \sigma^2)$, then the OLS estimators are the **maximum likelihood estimators**.

If we take $Y_i = X_i\beta + \epsilon_i$, where ϵ_i is non-normal, then the OLS estimators are simply the best (in terms of *mean squared error*) predictor of β.

Assumptions for OLS

1. The conditional mean is **linear** (in parameters).
2. All values of Y_i have **constant variance**, denoted σ^2 (conditionally).
3. The Y_i are **independent.**

Asymptotic Analysis

As $n \to \infty$, $\widehat{\beta} \dot\sim N(\beta, \text{var}(\widehat{\beta}))$, where

$$\text{var}(\widehat{\beta}) = \sigma^2 (X'X)^{-1}.$$

We can use this result for **confidence intervals** and **hypothesis tests**.

Summary

- Linear Regression allows us to estimate a functional form for the conditional mean of a continuous outcome.
- The OLS estimators are valid MLE-type estimators when normality is assumed, and are LS estimators otherwise.
- The asymptotic analysis is valid in large samples, regardless of distributional assumptions, and can be used for Wald-type analysis.