Day 1: Linear Regression

Review Course Schedule

- Day 1 (01/18/2023): Linear Regression Part I
- Day 2 (01/19/2023): Linear Regression Part II and ANOVA
- Day 3 (01/20/2023): GLMs for Binary Data
- Day 4 (01/23/2023): GLMs for Count Data
- Day 5 (01/24/2023): GLMs for Ordinal and Categorical Data

Overview

- Ordinary Least Squares (OLS) Regression Algebra
- OLS Assumptions
- OLS Properties
- Departures from Assumptions
- Regression Diagnostics

Regression Models

- Regression models are used to identify and quantify relationships between at least one explanatory variable $\mathbf{X}$ and a response variable, Y.

$$E[Y|X_1, X_2, ..., X_n] = \mu = \phi(x_1, x_2, ..., x_n)$$

Linear Regression

We use linear regression when we expect that ϕ is linear in terms of coefficients $\beta_1, \beta_2, \dots, \beta_n$. In the simplest case, we have

$$E[Y] = \mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n.$$

Linear Regression

Expressed in matrix form, we have

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \dots & x_{1,\ p-1} \\ x_{20} & x_{21} & x_{22} & \dots & x_{2,\ p-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ x_{n0} & x_{n1} & x_{n2} & \dots x_{n,\ p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix},$$

or

$$Y = \mathbf{X}\beta + \epsilon.$$

Least Squares Estimation

- We consider the explanatory variables, $\mathbf{X}$, given.
- Response variable Y is random.
- The coefficients β are unknown parameters that we need to estimate.
- Least squares estimation chooses the estimator $\hat{\beta}$ of β that minimizes the residual sum of squares, $\sum_{i=1}^{n} \epsilon_i^2 = \epsilon^T \epsilon$.

Assumptions

Least squares estimates require two assumptions:

- The errors have expected value of 0: $E[\epsilon] = 0$.
- The errors are uncorrelated and have constant variance: $Var[\epsilon] = \sigma^2 \mathbf{I}$.

These two assumptions are also often paired with a third assumption:

- The errors are normally distributed: $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$.

Ordinary Least Squares

To find an estimator for β, we need to minimize

$$\begin{aligned}
\epsilon^T\epsilon &= (Y - X\beta)^T(Y - X\beta) \\
&= (Y^T - \beta^T X^T)(Y - X\beta) \\
&= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta.
\end{aligned}$$

We differentiate with respect to β, giving $-2X^TY + 2X^TX\beta$.

Setting this equal to 0 gives a minimum where

$$X^TX\beta = X^TY.$$

Solving this for β yields our estimate:

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$

Properties of OLS Estimators

- $\hat{\beta}$ is unbiased

Properties of OLS Estimators

- $\hat{\beta}$ is unbiased

$$\begin{aligned} E[\hat{\beta}] &= E[(X^T X)^{-1} X^T Y] \\ &= (X^T X)^{-1} X^T E[Y] \\ &= (X^T X)^{-1} X^T X \beta \\ &= \beta \end{aligned}$$

Properties of OLS Estimators

- $Var[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$

Properties of OLS Estimators

- $Var[\hat{\beta}] = \sigma^2(X^TX)^{-1}$

$$
\begin{aligned}
Var[\hat{\beta}] &= Var[(X^TX)^{-1}X^TY] \\
&= (X^TX)^{-1}X^TVar[Y]X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T\sigma^2\mathbf{I}X(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1}
\end{aligned}
$$

Properties of OLS Estimators

- If X has full rank, then $\hat{\beta}$ has the lowest variance among unbiased linear estimators of β.
- $\hat{\beta}$ is often called the Best Linear Unbiased Estimator (BLUE).

Properties of OLS Estimators

We can estimate σ^2 with

$$S^2 = \frac{(Y - X\hat{\beta})^T(Y - X\hat{\beta})}{n-r} = \frac{RSS}{n-r},$$

where r is the rank (number of linearly independent columns) of X. This is an unbiased estimator for σ^2.