Lecture Summary: Feb. 15, 2023

• Matrix expression for multiple linear regression

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i, \quad i = 1, \ldots, n.$$

Assumptions for ϵ_i are the same as for simple linear regression.

Let

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p-1} \\ 1 & x_{2,1} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p-1} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Then, the multiple linear regression can be expressed as

$$Y = X\beta + \epsilon.$$

• Least squares (LS) estimate: Find β that minimizes $|Y - X\beta|^2$, where for a vector $v = (v_1, ..., v_n)'$, $|v|^2 = \sum_{i=1}^n v_i^2$. The solution is given by

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{pmatrix} = (X'X)^{-1}X'Y.$$

The LS estimate can be computed either by the above formula or by R.

• Inference about regression parameters

1. Test for $\mathrm{H}_0 : \beta_k = \beta_{k0}$ vs $\mathrm{H}_1 : \beta_k \neq \beta_{k0}$ ($> \beta_{k0}$, $< \beta_{k0}$), where β_{k0} is a specified value (e.g., $\beta_{k0} = 0$). Consider

$$t = \frac{\hat{\beta}_k - \beta_{k0}}{\text{s.e.}(\hat{\beta}_k)},$$

where

$$\text{s.e.}(\hat{\beta}_k) = \sqrt{\text{MSE} \times \{\text{the } k\text{th diagonal element of } (X'X)^{-1}\}}$$

$(0 \leq k \leq p-1)$. Under H_0, we have $t \sim t_{n-p}$.