Lecture \(\text{35}\)

Agenda:

1. The multivariate linear model
2. Least square estimates for the multivariate linear model

We now consider a more general situation where we want to study the relationship between a characteristic \(Y\) and another group of characteristics \(X_2, X_3, \ldots, X_p\), associated with the population/phenomenon under study. For example, \(Y\) is the income and we would like to understand its relationship with other characteristics like level of education, field of study, race, gender etc.

Data: We have \(n\) independently chosen vectors of observation, namely

\[
\begin{align*}
(Y_1, X_{11}, X_{12}, \ldots, X_{1p}) \\
(Y_2, X_{21}, X_{22}, \ldots, X_{2p}) \\
\vdots \\
(Y_n, X_{n1}, X_{n2}, \ldots, X_{np})
\end{align*}
\]
Matrix: $x_{ij}$ denotes the value of the characteristic $x_j$ for the $i^{th}$ individual/object.

Model: The linear model in this situation says that the value of the $y$ characteristic is a linear combination of the characteristics $x_1, x_2, \ldots, x_p$ plus some random error, i.e.,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i,$$

for every $i = 1, 2, \ldots, n$,

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent errors with a common distribution and $E[\varepsilon_i] = 0, \quad V(\varepsilon) = \sigma^2$.

Estimation: We need to find estimates for the parameters $\beta_0, \beta_1, \ldots, \beta_p$ and $\sigma^2$. We will look at the estimate for $\sigma^2$ next time, but let us first derive the estimates for $\beta_0, \beta_1, \ldots, \beta_p$. Let us introduce some relation to facilitate our derivation. Let

$$\beta \triangleq \left( \begin{array}{c} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{array} \right), \quad X \triangleq \left( \begin{array}{cccc} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{array} \right)$$
and

\[ Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}. \]

We want to estimate the parameter vector \( \beta \).

Consider the error sum of squares

\[ S(\beta) = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}))^2. \]

\( \text{Observed } Y \text{ value for } i^{th} \text{ individual/object} \)

\( \text{Expected } Y \text{ value under the model for the } i^{th} \text{ individual/object} \)

It is natural to choose parameter estimates which minimize the error sum of squares. Hence, the "least squares" estimate \( \hat{\beta} \) of the parameter vector \( \beta \) is defined by

\[ \hat{\beta} = \arg \min_{\beta} S(\beta). \]
Note that

\[ S(\beta) = (Y - X\beta)^T (Y - X\beta) = Y^T Y - (X\beta)^T Y - Y^T X\beta + (X\beta)^T X\beta \]

\[ = Y^T Y - 2X^T X\beta + \beta^T X^T X\beta. \]

(\because \text{Using } (\beta_0 \beta)^T = \beta^T A^{-1} \text{ and } \beta_0^T \beta = \beta^T \beta)

Hence, any minimizer of \( S(\beta) \) satisfies the gradient condition, i.e.,

\[ \nabla_\beta S(\beta) = 0. \]

\[ \implies \nabla_\beta \left\{ Y^T Y - 2X^T X\beta + \beta^T X^T X\beta \right\} = 0. \]

\[ \implies \nabla_\beta (Y^T Y) - \beta \nabla_\beta (2X^T X\beta) + \nabla_\beta (\beta^T X^T X\beta) = 0 \]

\[ \implies 0 - 2X^T Y + 2X^T X\beta = 0 \]

(\because \nabla_\beta Y^T Y = 0 \text{ and } \nabla_\beta \beta^T A^{-1} \beta = 2A\beta)

\[ \implies X^T X \beta = X^T Y \rightarrow \text{These are often known as the "normal equations".} \]

\[ \implies \beta = (X^T X)^{-1} X^T Y \]
We are inherently assuming that \((X^T X)^{-1}\) exists, i.e., 
\(\text{Rank}(X) = p+1\). This is not a very stringent condition.

One can look at the Hessian of \(S(\beta)\) at \(\hat{\beta} = (X^T X)^{-1} X^T y\) and verify that

\[ \hat{\beta} = (X^T X)^{-1} X^T y \]

is indeed the minimizer of \(S(\beta)\).