1 Introduction to Multiple Regression

- ***Multiple regression*** refers to regression *more than one* predictor variable. We denote the number of predictors by $\boldsymbol{p-1}$, with $p > 2$.
- Two main uses for a multiple regression analysis:
  1. To **describe** the relationship between a response variable Y and several predictors $X_1, X_2, \ldots, X_{p-1}$, and **predict** the value of Y for a given set of predictor values.
  2. To **control** for the effects of one or more predictor variables while investigating the relationship between Y and the other predictors.

1.1 Multiple Linear Regression Models

- The (*first order* or *additive*) ***multiple linear regression model*** has $\boldsymbol{p-1}$ **predictors** and $\boldsymbol{p}$ **parameters**.

Multiple Linear Regression Model:

$$Y_i \;=\; \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{ip-1} + \epsilon_i \tag{1}$$

where for $i = 1, 2, \ldots, n$:

- ▷ Y_i is the response for the ith individual.
- ▷ $X_{i1}, X_{i2}, \ldots, X_{ip-1}$ are the values of $p-1$ (numerical) predictor variables for the ith individual.
- ▷ β_0 represents an intercept term of the true regression model mean response.
- ▷ $\beta_1, \ldots, \beta_{p-1}$ represent coefficients of the true regression model mean response.
- ▷ ϵ_i is a random ***error***, with

$$\epsilon_i \;\sim\; N(0, \sigma^2)\,,$$

and the ϵ_i's are assumed to be independent of each other, and therefore uncorrelated.

- The model (1) can be restated by saying that the responses $Y_1, Y_2, \ldots, Y_n$'s are **independent normal** random variables with **means**

$$\mu_i \;=\; \mathrm{E}(Y_i) \;=\; \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{ip-1}$$

and each with **variance**

$$\mathrm{Var}(Y_i) \;=\; \sigma^2,$$

i.e. $Y_1, Y_2, \ldots, Y_n$ are **independent** and **normally distributed**, with

$$Y_i \;\sim\; N\left(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{ip-1},\; \sigma^2\right).$$

- As a function of the X's, when $p = 3$ (i.e. *two* predictors), the **mean response**,

$$\mathrm{E}(Y) \;=\; \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1},$$

is a **plane**. See Class Notes 1.

When $p > 3$, it's a ***hyperplane***.

- $\boldsymbol{\beta_k}$, for $k = 1, 2, \ldots, p-1$, represents the **change** in the **mean response** $\mathrm{E}(Y)$ for each one-unit increase in X_k **when the other predictors are all held constant**.

In other words, β_k is the change in the mean response associated with a one-unit increase in X_k, while **controlling** for the other predictors.

2 Multiple Linear Regression Models in Matrix Terms

- Suppose we have data:

Individual Observation	1st Predictor Variable X_1	2nd Predictor Variable X_2	$\cdots$	$p-1$st Predictor Variable X_{p-1}	Response Variable Y
1	X_{11}	X_{12}	$\cdots$	X_{1p-1}	Y_1
2	X_{21}	X_{22}	$\cdots$	X_{2p-1}	Y_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	X_{n1}	X_{n2}	$\cdots$	X_{np-1}	Y_n

- We could write out the multiple linear regression model as

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_{p-1} X_{1p-1} + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_{p-1} X_{2p-1} + \epsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_{p-1} X_{np-1} + \epsilon_n \end{aligned} \tag{2}$$

- We can also write it in matrix terms. As before, define the $n \times 1$ ***response vector*** $\mathbf{Y}$ to be

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \tag{3}$$

and the $n \times 1$ ***error vector*** $\boldsymbol{\epsilon}$ to be

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{4}$$

Also, define the $p \times 1$ ***parameter vector*** $\boldsymbol{\beta}$ to be

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \tag{5}$$

and the $n \times p$ ***design matrix*** $\mathbf{X}$ to be

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np-1} \end{bmatrix} \tag{6}$$

Multiple Linear Regression Model (Matrix Approach): The regression model (2) can be written in terms of the vectors and matrix (3), (4), (5), and

(6) as

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

- Since $\mathrm{E}(\epsilon_i) = 0$ for $i = 1, 2, \ldots, n$,

$$\mathrm{E}(\boldsymbol{\epsilon}) \;=\; \mathbf{0}$$

where

$$\mathbf{0} \;=\; \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(the $n \times 1$ vector of zeros). It follows that

$$\mathrm{E}(\mathbf{Y}) \;=\; \mathrm{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \;=\; \mathbf{X}\boldsymbol{\beta} + \mathrm{E}(\boldsymbol{\epsilon}) \;=\; \mathbf{X}\boldsymbol{\beta}\,.$$

- The **variance-covariance matrix** of $\boldsymbol{\epsilon}$ is the $n \times n$ diagonal matrix

$$\sigma^2\{\boldsymbol{\epsilon}\} \;=\; \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \;=\; \sigma^2\mathbf{I}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix.

- For the multiple linear regression model, $Y_1, Y_2, \ldots, Y_n$ are independent, each with variance σ^2, so the **variance-covariance matrix** of $\mathbf{Y}$ is the $n \times n$ diagonal matrix

$$\sigma^2\{\mathbf{Y}\} \;=\; \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \;=\; \sigma^2\mathbf{I}\,.$$