1 Random Vectors and Matrices

- The vectors $\mathbf{Y}$, $\boldsymbol{\epsilon}$, $\mathbf{b}$, $\mathbf{e}$, and $\hat{\mathbf{Y}}$ are examples of ***random vectors*** since their elements are random variables.

  We define the **expected value** of a random **vector** such as

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

  to be

$$E(\mathbf{Y}) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} \tag{1}$$

  The **expected value** of a **matrix** whose elements are random variables is defined analogously (by taking expected values of each element of the matrix).

- Let

$$\sigma^2\{Y_i\} = \text{Var}(Y_i)$$

  be the variance of Y_i, and let

$$\sigma\{Y_i, Y_j\} = E[(Y_i - E(Y_i))(Y_j - E(Y_j))] \tag{2}$$

  be the ***covariance*** between Y_i and Y_j. The **covariance** is a measure of **association** between Y_i and Y_j:

  - ▷ If Y_i and Y_j are **independent**, then $\sigma\{Y_i, Y_j\} = 0$.
  - ▷ A **positive** $\sigma\{Y_i, Y_j\}$ indicates that when Y_i is above average, Y_j tends to be above average too, and when Y_i is below average Y_j tends to be below average too.
  - ▷ A **negative** $\sigma\{Y_i, Y_j\}$ indicates that when Y_i is above average, Y_j tends to be below average, and when Y_i is below average Y_j tends to be above average.
  - ▷ It's easy to see from the definition (2) that the covariance between a random variable and itself is the variance of that random variable, i.e.

$$\sigma\{Y_i, Y_i\} = E[(Y_i - E(Y_i))^2] = \text{Var}(Y_i) = \sigma^2\{Y_i\}.$$

- We define the ***variance-covariance matrix*** of a random vector $\mathbf{Y}$ to be

$$\sigma^2\{\mathbf{Y}\} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \dots & \sigma\{Y_1, Y_n\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \dots & \sigma\{Y_2, Y_n\} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma\{Y_n, Y_1\} & \sigma\{Y_n, Y_2\} & \dots & \sigma^2\{Y_n\} \end{bmatrix}$$

Thus $\sigma^2\{\mathbf{Y}\}$ has the **variances** of the Y_i's on the diagonal and their **covariances** on the off-diagonals. The variance-covariance matrix is **symmetric** (its i, jth element equals its j, ith) since $\sigma\{Y_i, Y_j\} = \sigma\{Y_j, Y_i\}$.

- For the linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i\,, \tag{3}$$

since $\mathrm{E}(Y_i) = \beta_0 + \beta_1 X_i$ for $i = 1, 2, \dots, n$,

$$\mathrm{E}(\mathbf{Y}) = \begin{bmatrix} \mathrm{E}(Y_1) \\ \mathrm{E}(Y_2) \\ \vdots \\ \mathrm{E}(Y_n) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} \tag{4}$$

Also, since $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are independent with variances $\sigma^2\{\epsilon_i\} = \sigma^2$ for $i = 1, 2, \dots, n$, the **variance-covariance matrix** of $\boldsymbol{\epsilon}$ is the $n \times n$ diagonal matrix

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

where $\boldsymbol{I}$ is the $n \times n$ **identity matrix**

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

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

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

- We'll need the following later.

> **Fact 1.1 Some Basic Results**: Suppose $\mathbf{Y}$ is a random vector and $\mathbf{A}$ is a matrix of constants. Let
> $$\mathbf{W} = \mathbf{AY}$$
> Then:
>
> 1. $\mathrm{E}(\mathbf{A}) = \mathbf{A}$.
> 2. $\mathrm{E}(\mathbf{W}) = \mathrm{E}(\mathbf{AY}) = \mathbf{A}\mathrm{E}(\mathbf{Y})$.
> 3. $\sigma^2\{\mathbf{W}\} = \sigma^2\{\mathbf{AY}\} = \mathbf{A}\sigma^2\{\mathbf{Y}\}\mathbf{A}^T$.

1.1 Mean and Variance-Covariance Matrix of b

- Recall that the vector of estimated coefficients is

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

We know that since $\mathrm{E}(b_0) = \beta_0$ and $\mathrm{E}(b_1) = \beta_1$,

$$\mathrm{E}(\mathbf{b}) = \begin{bmatrix} \mathrm{E}(b_0) \\ \mathrm{E}(b_1) \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- Recall also that

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{5}$$

where $\mathbf{Y}$ is the $n \times 1$ response vector and $\mathbf{X}$ is the $n \times 2$ design matrix.

Letting $\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ in Result 3 of Fact 1.1 and using the fact that $\sigma^2\{\mathbf{Y}\} = \sigma^2\boldsymbol{I}$, it can be shown that the 2×2 **variance-covariance matrix** of $\mathbf{b}$ is:

$$\sigma^2\{\mathbf{b}\} = \sigma^2 \cdot (\mathbf{X}^T\mathbf{X})^{-1} \tag{6}$$

The diagonal elements of this matrix are the variances of b_0 and b_1 given in Class Notes 2, and both off-diagonal elements are the covariance between b_0 and b_1.

- Usually we don't know the value of σ^2, so we estimate it by the MSE. This gives the **(estimated)** ***variance-covariance matrix of*** $\mathbf{b}$, denoted $\boldsymbol{s^2\{\mathbf{b}\}}$:

> **(Estimated) Variance-Covariance Matrix of b:**
>
> $$s^2\{\mathbf{b}\} \;=\; \text{MSE} \cdot (\mathbf{X}^T\mathbf{X})^{-1}\,.$$

The *square roots* of the diagonal elements of $s^2\{\mathbf{b}\}$ are the (estimated) **standard errors** $s\{b_0\}$ and $s\{b_1\}$ of b_0 and b_1 given in Class Notes 2 and reported by statistical software when a regression analysis is carried out.