Chapter 1 Introduction

Hakbae Lee

The Department of Statistics and Data Science, Yonsei University

Let's consider a linear model

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

where

$$\underbrace{\mathbf{Y}}_{n\times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad \underbrace{\boldsymbol{\beta}}_{(p+1)\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \qquad \underbrace{\boldsymbol{\epsilon}}_{n\times 1} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\underbrace{\mathbf{X}}_{n\times(p+1)} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix}$$

Introduction

- Simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- Multiple linear regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{i1} + \epsilon_i$$

- One-Way Analysis of Variance

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

- Two-Way Analysis of Variance with interaction

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

Random Vectors and Matrices

- Let $\mathbf{Y} = (y_1, \ldots, y_n)^T$ be a random vector with $\mathrm{E}(y_i) = \mu_i$, $\mathrm{Var}(y_i) = \sigma_{ii} (= \sigma_i^2)$, $\mathrm{Cov}(y_i, y_j) = \sigma_{ij}$.
- Define the expected value of $\mathbf{Y}$ elementwise as

$$\mathrm{E}(\mathbf{Y}) = (\mathrm{E}(y_1), \ldots, \mathrm{E}(y_n))^T = (\mu_1, \ldots, \mu_n)^T = \mu$$

  and the covariance matrix of $\mathbf{Y}$ as

$$\mathrm{Cov}(\mathbf{Y}) = \mathrm{E}[(\mathbf{Y} - \mu)(\mathbf{Y} - \mu)^T] = (\sigma_{ij})$$

- Note:

$$\begin{aligned} \mathrm{E}(A\mathbf{Y} + b) &= A\mu + b \\ \mathrm{Cov}(A\mathbf{Y} + b) &= A\mathrm{Cov}(\mathbf{Y})A^T \end{aligned}$$

- (Exercise 1.3) Prove or disprove that $\mathrm{Cov}(\mathbf{Y})$ is nonnegative definite.

- Covariance of $\underbrace{\mathbf{W}}_{r\times 1}$ and $\underbrace{\mathbf{Y}}_{s\times 1}$ with $\mathrm{E}(\mathbf{W}) = \gamma$ and $\mathrm{E}(\mathbf{Y}) = \mu$

$$\mathrm{Cov}(\mathbf{W}, \mathbf{Y}) = \mathrm{E}[(\mathbf{W} - \gamma)(\mathbf{Y} - \mu)^T] : \qquad r \times s$$

and

$$\mathrm{Cov}(A\mathbf{W} + a, B\mathbf{Y} + b) = A\mathrm{Cov}(\mathbf{W}, \mathbf{Y})B^T$$

Theorem 1.1.1.

$$\begin{aligned} \mathrm{Cov}(A\mathbf{W} + B\mathbf{Y}) &= A\mathrm{Cov}(\mathbf{W})A^T + B\mathrm{Cov}(\mathbf{Y})B^T \\ &\quad + A\mathrm{Cov}(\mathbf{W}, \mathbf{Y})B^T + B\mathrm{Cov}(\mathbf{Y}, \mathbf{W})A^T \end{aligned}$$

- Let $\mathbf{Z} = (z_1, \ldots, z_n)^T \sim N_n(0, I_n)$ where $z_1, \ldots, z_n$ are i.i.d $N(0, 1)$. Note that $\mathrm{E}(\mathbf{Z}) = 0$ and $\mathrm{Cov}(\mathbf{Z}) = I_n$

Definition 1.2.1. Let A be $r \times n$ and $b \in \mathbf{R}^r$. Then $\mathbf{Y}$ has an r-dimensional multivariate normal distribution :

$$\mathbf{Y} = A\mathbf{Z} + b \sim N_r(b, AA^T).$$

Theorem 1.2.2. Let $\mathbf{Y} \sim N(\mu, V)$ and $\mathbf{W} \sim N(\mu, V)$. Then $\mathbf{Y}$ and $\mathbf{W}$ have the same distribution (Proof: p.5)

Multivariate Normal Distributions

- The density of nonsingular $\mathbf{Y} \sim N(\mu, V)$ is given by

$$f(y) = (2\pi)^{-n/2}[\det(V)]^{-1/2} \exp[-(y-\mu)^T V^{-1}(y-\mu)/2]$$

Theorem 1.2.3. Let $\mathbf{Y} \sim N(\mu, V)$ and $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}$. Then

$$\mathrm{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = 0 \quad \text{if and only if} \quad \mathbf{Y}_1 \perp\!\!\!\perp \mathbf{Y}_2$$

Corollary 1.2.4. Let $\mathbf{Y} \sim N(\mu, \sigma^2 I)$ and $AB^T = 0$. Then

$$A\mathbf{Y} \perp\!\!\!\perp B\mathbf{Y}$$

Definition 1.3.1. Quadratic Form of $\mathbf{Y}$: for $n \times n$, A

$$\mathbf{Y}^T A \mathbf{Y} = \sum_{ij} a_{ij} y_i y_j$$

Theorem 1.3.2. Let $\mathrm{E}(\mathbf{Y}) = \mu$ and $\mathrm{Cov}(\mathbf{Y}) = V$. Then

$$\mathrm{E}(\mathbf{Y}^T A \mathbf{Y}) = \mathrm{tr}(\mathrm{AV}) + \mu^{\mathrm{T}} \mathrm{A} \mu$$

proof; p.8

Note: Let's consider $\mathbf{Z} \sim N_n(\mu, I_n)$. Then

$$\mathbf{Z}^T \mathbf{Z} \sim \chi^2(n, \mu^T \mu / 2)$$

where $\mu^T \mu / 2$= non-centrality parameter

Theorem 1.3.3. Let $\mathbf{Y} \sim N(\mu, I)$ and M be any orthogonal projection matrix. Then

$$\mathbf{Y}^T M \mathbf{Y} \sim \chi^2(r(M), \mu^T M \mu/2)$$

Note: Let $\mathbf{Y} \sim N(\mu, \sigma^2 I)$. Then

$$\mathbf{Y}^T M \mathbf{Y} \sim \chi^2(r(M), \mu^T M \mu/2\sigma^2)$$

Lemma 1.3.4. Let $\mathbf{Y} \sim N(\mu, M)$ with $\mu \in \mathcal{C}(M)$ and M be an orthogonal projection matrix. Then

$$\mathbf{Y}^T \mathbf{Y} \sim \chi^2(r(M), \mu^T \mu/2)$$

Lemma 1.3.5. Let $\mathrm{E}(\mathbf{Y}) = \mu$ and $\mathrm{Cov}(\mathbf{Y}) = V$. Then

$$\Pr[(\mathbf{Y} - \mu) \in \mathcal{C}(V)] = 1$$

Exercise 1.6. Let $\mathbf{Y}$ be a vector with $E(\mathbf{Y}) = 0$ and $Cov(\mathbf{Y}) = 0$. Then $\Pr(\mathbf{Y} = 0) = 1$

Theorem 1.3.6. Let $\mathbf{Y} \sim N(\mu, V)$. Then

$$\mathbf{Y}^T A\mathbf{Y} \sim \chi^2(\mathrm{tr}(\mathrm{AV}), \mu^{\mathrm{T}}\mathrm{A}\mu/2)$$

provided that (1) $VAVAV = VAV$, (2)$\mu^T AVA\mu = \mu^T A\mu$, and (3) $VAVA\mu = VA\mu$.(proof; p.10)

Exercise 1.7. (a) Show that if V is nonsingular, then the three conditions in Theorem 1.3.6 reduce to $AVA = A$. (b) Show that $\mathbf{Y}^T V^- \mathbf{Y}$ has a chi-squared distribution with $r(V)$ degrees of freedom when $\mu \in \mathcal{C}(V)$.

Theorem 1.3.7. Let $\mathbf{Y} \sim N(\mu, \sigma^2 I)$ and $BA = 0$. Then, for $A = A^T$

$$(1)\ \mathbf{Y}^T A\mathbf{Y} \perp\!\!\!\perp B\mathbf{Y} \quad \text{and} \quad (2)\ \mathbf{Y}^T A\mathbf{Y} \perp\!\!\!\perp \mathbf{Y}^T B\mathbf{Y} \text{ for } B = B^T$$

Theorem 1.3.8. Let $\mathbf{Y} \sim N(\mu, V)$, $A \geq 0, B \geq 0$ and $VAVBV = 0$. Then

$$\mathbf{Y}^T A\mathbf{Y} \perp\!\!\!\perp \mathbf{Y}^T B\mathbf{Y}$$

Theorem 1.3.9. If $\mathbf{Y} \sim N(\mu, V)$, and (1) $VAVBV = 0$, (2) $VAVB\mu = 0$, (3) $VBVA\mu = 0$, (4) $\mu^T AVB\mu = 0$, and conditions (1), (2), and (3) from Theorem 1.3.6 hold for both $\mathbf{Y}^T A\mathbf{Y}$ and $\mathbf{Y}^T B\mathbf{Y}$, then $\mathbf{Y}^T A\mathbf{Y} \perp\!\!\!\perp \mathbf{Y}^T B\mathbf{Y}$.