Recap: ① Some basic definitions of linear model

② Definitions of simple and multiple linear regression.

③ Some concepts of design of experiment.

$\cancel{Y = X\beta}$ $\quad y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$

predictor

$\beta_1, \ldots, \beta_p$ are called predictor coefficients.

Example:
y_{ij} is the nitrogen concentration of ith plot that received jth level of the fertilizer.

$i = 1, \ldots, n$; $j = 1, \ldots, a$.

$$y_{ij} = \mu + \alpha_j + e_{ij} \leftarrow \text{error}$$

mean level $\quad$ effect due to jth level of the fertilizer.

$$\left.\begin{aligned}
y_{11} &= \mu + \alpha_1 + e_{11}\\
&\vdots\\
y_{n1} &= \mu + \alpha_1 + e_{n1}\\
y_{12} &= \mu + \alpha_2 + e_{12}\\
&\vdots\\
y_{n2} &= \mu + \alpha_2 + e_{n2}\\
&\vdots\\
y_{1a} &= \mu + \alpha_a + e_{1a}\\
&\vdots\\
y_{na} &= \mu + \alpha_a + e_{na}
\end{aligned}\right\}
\begin{pmatrix} y_{11}\\ \vdots\\ \vdots\\ \vdots\\ y_{na}\end{pmatrix}
=
\begin{bmatrix}
1 & 1 & 0 & & 0\\
\vdots & \vdots & \vdots & & \vdots\\
\vdots & 1 & 0 & & \vdots\\
\vdots & 0 & 1 & \cdots & \vdots\\
\vdots & \vdots & \vdots & & \vdots\\
\vdots & \vdots & 1 & & \vdots\\
\vdots & \vdots & 0 & & \vdots\\
\vdots & \vdots & \vdots & & 0\\
\vdots & \vdots & \vdots & & 1\\
1 & 0 & 0 & \cdots & 1
\end{bmatrix}
\begin{pmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \vdots\\ \alpha_a\end{pmatrix}
+
\begin{pmatrix} e_{11}\\ \vdots\\ \vdots\\ \vdots\\ e_{na}\end{pmatrix}$$

$$\underline{y} = \begin{pmatrix} y_{11} \\ \vdots \\ \vdots \\ y_{na} \end{pmatrix}$$

$$\underline{y} = \begin{bmatrix} \underline{1}_n & \underline{1}_n & \cdots & \underline{0}_n \\ \vdots & \underline{0}_n & & \vdots \\ \vdots & \vdots & & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \cdots & \underline{1}_n \end{bmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{pmatrix} + \begin{pmatrix} e_{11} \\ \vdots \\ \vdots \\ e_{na} \end{pmatrix}$$

$na \times (a+1)$

$$\underline{1}_n = \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{n \text{ times}}, \quad \underline{0}_n = \left.\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}\right\} n \text{ times}$$

$$\underline{X} = \begin{bmatrix} \underline{1}_n & \underline{1}_n & \cdots & \underline{0}_n \\ \vdots & \underline{0}_n & \cdots & \vdots \\ \vdots & \vdots & & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \cdots & \underline{1}_n \end{bmatrix}, \quad \underline{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{pmatrix},$$

$$\underline{e} = \begin{pmatrix} e_{11} \\ \vdots \\ \vdots \\ e_{na} \end{pmatrix} \implies \underline{y} = \underline{X}\,\underline{\beta} + \underline{e}$$

Definition: Kronecker product between two matrices $\underline{A}_{n \times m}$ and $\underline{B}_{k \times l}$ is ~~given~~ denoted by $\underline{A} \otimes \underline{B}$ and the entries of $\underline{A} \otimes \underline{B}$ is given by the following

$$\begin{pmatrix} a_{11} \underset{k \times l}{\underline{B}} & a_{12} \underset{k \times l}{\underline{B}} & \cdots & a_{1m} \underset{k \times l}{\underline{B}} \\ \vdots & \vdots & & \vdots \\ a_{n1} \underset{k \times l}{\underline{B}} & \cdots & \cdots & a_{nm} \underset{k \times l}{\underline{B}} \end{pmatrix}_{nk \times ml}$$

$$\underline{X} = \text{a times} \left\{ \begin{bmatrix} \underline{1}_n & \underline{1}_n & \underline{0}_n & \cdots & \underline{0}_n \\ \vdots & \underline{0}_n & \underline{1}_n & & \underline{0}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \underline{0}_n & \cdots & \underline{1}_n \end{bmatrix} \right. = \left[\underline{1}_a \otimes \underline{1}_n \;\vdots\; \underline{I}_a \otimes \underline{1}_n \right]$$

$$\underline{I}_a \otimes \underline{1}_n = \begin{pmatrix} 1 \cdot \underline{1}_n & 0 \cdot \underline{1}_n & \cdots & 0 \cdot \underline{1}_n \\ 0 \cdot \underline{1}_n & 1 \cdot \underline{1}_n & \cdots & 0 \cdot \underline{1}_n \\ \vdots & & & \\ 0 \cdot \underline{1}_n & \cdots & \cdots & 1 \cdot \underline{1}_n \end{pmatrix}$$

$$\underline{1}_a \otimes \underline{1}_n = \begin{pmatrix} 1 \cdot \underline{1}_n \\ 1 \cdot \underline{1}_n \\ \vdots \\ 1 \cdot \underline{1}_n \end{pmatrix}$$

$$\underline{X} = \left[\underline{1}_a \otimes \underline{1}_n \; \vdots \; \underline{I}_a \otimes \underline{1}_n \right]$$

$$\underline{y} = \underline{X} \, \underline{\beta} + \underline{e} \quad \text{(One way Anova model)}.$$

__Example__: There are two different fertilizers and the first one can be applied at a different levels, the second one can be applied at b different levels. Each combination of fertilizers is applied to n different agricultural plots.

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$$

y_{ijk} = Nitrogen conc. of the kth plot that received ith level of fertilizer 1 and jth level of fertilizer 2.

$k = 1, \ldots, n$.

$$y_{111} = \mu + \alpha_1 + \beta_1 + e_{111}$$

$$\vdots$$

$$y_{11n} = \mu + \alpha_1 + \beta_1 + e_{11n}$$

$$\vdots$$

$$y_{ab1} = \mu + \alpha_a + \beta_b + e_{ab1}$$

$$\vdots$$

$$y_{abn} = \mu + \alpha_a + \beta_b + e_{abn}$$

$$\begin{pmatrix} y_{111} \\ \vdots \\ y_{11n} \\ \vdots \\ y_{ab1} \\ \vdots \\ y_{abn} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & 1 & 0 & & \vdots \\ \vdots & 0 & 1 & & \vdots \\ \vdots & \vdots & 0 & & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \\ \beta_1 \\ \vdots \\ \beta_b \end{pmatrix} + \begin{pmatrix} e_{111} \\ \vdots \\ e_{11n} \\ \vdots \\ e_{ab1} \\ \vdots \\ e_{abn} \end{pmatrix}$$

$$\underbrace{\qquad\qquad}_{a \text{ times}}$$

$$\begin{pmatrix} y_{111} \\ \vdots \\ y_{11n} \\ y_{121} \\ \vdots \\ y_{12n} \\ \vdots \\ y_{ab1} \\ \vdots \\ y_{abn} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & 1 & & \vdots & 0 & & 0 \\ \vdots & \vdots & & \vdots & 0 & & 0 \\ \vdots & \vdots & & \vdots & 0 & & 0 \\ \vdots & 0 & & 1 & 0 & & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 1 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \\ \beta_1 \\ \vdots \\ \beta_b \end{pmatrix} + \begin{pmatrix} e_{111} \\ \vdots \\ e_{11n} \\ \vdots \\ e_{ab1} \\ \vdots \\ e_{abn} \end{pmatrix}$$

Column corresponding to β_1

$$\begin{array}{r} n \\ \\ \\ n \\ (b-1)n \quad n \\ \\ n \\ \\ n \\ \\ (b-1)n \\ \\ \vdots \\ n \\ \\ (b-1)n \\ \\ \end{array} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \cancel{\underset{a\times 1}{\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}} \otimes \mathbf{1}_n}$$

$$= \underset{ab\times 1}{\left.\begin{pmatrix} b\left\{\begin{matrix} 1 \\ 0 \\ \vdots \\ 0 \end{matrix}\right. \\ b\left\{\begin{matrix} 1 \\ 0 \\ \vdots \\ 0 \end{matrix}\right. \\ \vdots \\ b\left\{\begin{matrix} 1 \\ 0 \\ \vdots \\ 0 \end{matrix}\right. \end{pmatrix}\right]}_{a \text{ times}} \otimes \underset{n\times 1}{\mathbf{1}_n}$$

a times

$y_{111} = \mu + \alpha_1 + \beta_1 + e_{111}$

$\vdots$

$y_{11n} = \mu + \alpha_1 + \beta_1 + e_{11n}$

$y_{121} = \mu + \alpha_1 + \beta_2 + e_{121}$

$\vdots$

$y_{12n} = \mu + \alpha_1 + \beta_2 + e_{12n}$

$\vdots$

$y_{1b1} = \mu + \alpha_1 + \beta_b + e_{1b1}$

$\vdots$

$y_{1bn} = \mu + \alpha_1 + \beta_b + e_{1bn}$

$y_{211} = \mu + \alpha_2 + \beta_1 + e_{211}$

$\vdots$

$y_{21n} = \mu + \alpha_2 + \beta_1 + e_{21n}$

$\vdots$

$y_{ab1} = \mu + \alpha_a + \beta_b + e_{ab1}$

$\vdots$

$y_{abn} = \mu + \alpha_a + \beta_b + e_{abn}$

$$
\underline{X} = \begin{bmatrix}
\underline{1}_n & \underline{1}_n & \underline{0}_n & \cdots & \underline{0}_n & \underline{1}_n & \cdots & \underline{0}_n \\
\underline{1}_n & \vdots & \vdots & & \vdots & \underline{0}_n & & \underline{0}_n \\
\vdots & \underline{1}_n & \underline{0}_n & & \vdots & \underline{0}_n & & \underline{1}_n \\
\vdots & \underline{0}_n & \underline{1}_n & & \vdots & \underline{1}_n & & \underline{0}_n \\
& & & & & \underline{0}_n & & \\
\underline{1}_n & \underline{0}_n & \underline{1}_n & & \vdots & \vdots & & \underline{0}_n \\
\underline{1}_n & \vdots & \underline{0}_n & & \vdots & \underline{0}_n & & \underline{1}_n \\
\vdots & & \vdots & & \underline{0}_n & \vdots & & \vdots \\
\vdots & \underline{0}_n & \underline{0}_n & & \underline{1}_n & \underline{1}_n & & \underline{0}_n \\
& & & & & \underline{0}_n & & \vdots \\
\underline{1}_n & \underline{0}_n & \underline{0}_n & & \underline{1}_n & \underline{0}_n & \cdots & \underline{0}_n \\
& & & & & & & \underline{1}_n
\end{bmatrix}
$$

$$
= \left[\underline{1}_{ab} \otimes \underline{1}_n \;\vdots\; \underline{I}_a \otimes \underline{1}_{nb} \;\vdots\; \begin{pmatrix} \underline{I}_b \otimes \underline{1}_n \\ \underline{I}_b \otimes \underline{1}_n \\ \vdots \\ \underline{I}_b \otimes \underline{1}_n \end{pmatrix} \right]
$$

$$
= \left[\underline{1}_{ab} \otimes \underline{1}_n \;\vdots\; \underline{I}_a \otimes \underline{1}_{nb} \;\vdots\; \underline{1}_a \otimes \underline{I}_b \otimes \underline{1}_n \right]
$$

$$
\underline{y} = X \underline{\beta} + \underline{e} \qquad \underline{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \\ \beta_1 \\ \vdots \\ \beta_b \end{pmatrix}
$$

How to estimate the coefficient vector $\underline{\beta}$

$$\underline{y} = X\underline{\beta} + \underline{e}$$

Least square estimator

Def: Let $\underline{h}$ be a vector of dimension n.

$$\|\underline{h}\| = \sqrt{\underline{h}^T\underline{h}} \qquad \underline{h} = (h_1, \ldots, h_n)^T$$

$$= \sqrt{\sum_{i=1}^{n} h_i^2}$$

Distance between two vectors $\underline{x}$, $\underline{y}$ both of dimension $n \times 1$ is given by

$$\|\underline{x} - \underline{y}\| = \sqrt{(\underline{x}-\underline{y})^T(\underline{x}-\underline{y})}$$

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

then $\|\underline{x} - \underline{y}\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$

Least square estimator of $\underline{\beta}$ is obtained by minimizing $Q(\underline{\beta}) = \|\underline{y} - X\underline{\beta}\|^2 = (\underline{y} - X\underline{\beta})^T(\underline{y} - X\underline{\beta})$

w.r.t. $\underline{\beta}$.

Write $\dfrac{\partial Q(\underline{\beta})}{\partial \underline{\beta}} = \underline{0}$

Derivative w.r.t. a vector

Let $\underline{a}$, $\underline{b}$ be $p \times 1$ vectors and $\underline{A}$ be a $p \times p$ matrix. Then,

(1) $$\frac{\partial(\underline{b}^T\underline{a})}{\partial \underline{b}} = \frac{\partial(\underline{a}^T\underline{b})}{\partial \underline{b}} = \underline{a}$$

(2) $$\frac{\partial(\underline{b}^T\underline{A}\,\underline{b})}{\partial \underline{b}} = (\underline{A} + \underline{A}^T)\,\underline{b}$$

$$Q(\underline{\beta}) = (\underline{y} - \underline{X}\,\underline{\beta})^T(\underline{y} - \underline{X}\,\underline{\beta})$$
$$= \underline{y}^T\underline{y} - \underline{y}^T\underline{X}\,\underline{\beta} - \underline{\beta}^T\underline{X}^T\underline{y} + \underline{\beta}^T\underline{X}^T\underline{X}\,\underline{\beta}$$

$$\frac{\partial Q(\underline{\beta})}{\partial \underline{\beta}} = \underline{0} - \underline{X}^T\underline{y} - \underline{X}^T\underline{y} + \left(\underline{X}^T\underline{X} + (\underline{X}^T\underline{X})^T\right)\underline{\beta}$$
$$= -2\,\underline{X}^T\underline{y} + 2\,\underline{X}^T\underline{X}\,\underline{\beta}$$

$$\frac{\partial Q(\underline{\beta})}{\partial \underline{\beta}} = \underline{0} \iff \underline{X}^T\underline{X}\,\underline{\beta} = \underline{X}^T\underline{y}$$ (Normal Equations)

Let $\underline{X}$ be of dimension $n \times p$, $\underline{y}$ be of dimension $n \times 1$, $\underline{\beta}$ is of dimension $p \times 1$

$$\Rightarrow \underset{p\times n}{\underline{X}^T}\,\underset{n\times p}{\underline{X}}\,\underset{p\times 1}{\underline{\beta}} = \underset{p\times n}{\underline{X}^T}\,\underset{n\times 1}{\underline{y}}$$

When $(\underline{X}^T\underline{X})$ can be inverted these equations has a unique solution $\underline{\beta} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$

But when $\underline{X}^T\underline{X}$ is not invertible, then these equations have infinitely many solutions.

Example: $y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \ldots, n$

$$\underline{y} = \underline{X}\,\underline{\beta} + \underline{e} \qquad \underline{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\underline{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}, \quad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\underline{X}^T \underline{X}\, \underline{\beta} = \underline{X}^T \underline{y}$$

$$\underline{X}^T \underline{X} = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

we will see later that $(\underline{X}^T \underline{X})^{-1}$ exists if and only if not all x_i's are the same.

Thus the Normal Equations have a unique solution if not all x_i's are equal.