$y = X\beta + \epsilon$, $\mu_y = E(y|x) = X\beta$ $\quad \epsilon \sim (0, \sigma^2 I_n)$

THE LINEAR MODEL

- ## The Statistical Model

In the case of the linear model it is assumed that the random vector

$$\boldsymbol{y}: n \times 1 \ \sim \ (\boldsymbol{\eta}, \sigma^2 \boldsymbol{I})$$

with

$r(X) = p \Rightarrow \hat{b} = (X^TX)^{-1}X^Ty$

$$E(\boldsymbol{y}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{b}$$

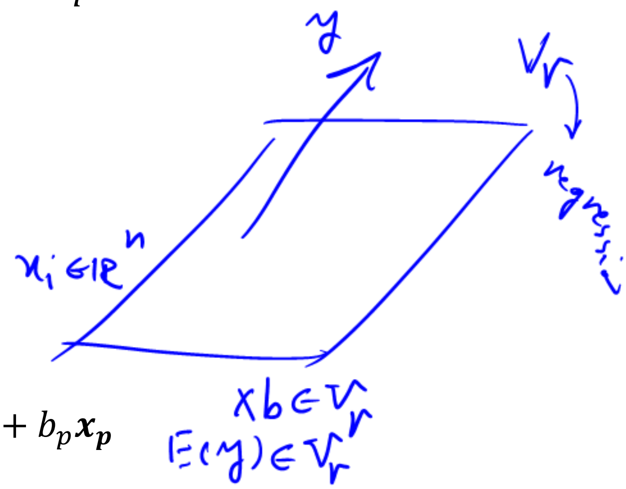
where $\boldsymbol{X}: n \times p$ is a constant matrix (with known elements) and $\boldsymbol{b}: p \times 1$ is a vector with unknown parameters. It is assumed that

$\hat{\eta} = ?$ $\xleftarrow{\text{in this case}}$

$$\text{rank}\ (\boldsymbol{X}) = r \leq p \leq n.$$

Suppose that

$$\boldsymbol{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_p}).$$



The linear model can therefore also be expressed as

$$E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{b} = b_1\boldsymbol{x_1} + b_2\boldsymbol{x_2} + \cdots + b_p\boldsymbol{x_p}$$

where the elements of $\boldsymbol{b}$ are arbitrary (unknown parameters). If V_r is the vector space generated by the columns of $\boldsymbol{X}$, then the linear model can be explicitly expressed as:

$$E(\boldsymbol{y}) \in V_r.$$

Examples

1. In the previous example (~~p.45~~), $\boldsymbol{y}: 4 \times 1 \sim (\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \alpha \\ \alpha \\ \beta \\ \beta \end{pmatrix}.$$

The linear model can be expressed as:

$$E(\boldsymbol{y}) = \overbrace{\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}}^{X} \overbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}^{b}. = \mu$$

2. Suppose that $\boldsymbol{y}: 4 \times 1 \sim (\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \alpha + 2\beta + \gamma \\ \alpha + 2\beta + \gamma \\ \alpha + \beta \\ \alpha + \beta \end{pmatrix}.$$

The linear model can be expressed as:

$$E(\boldsymbol{y}) = \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}}^{X} \overset{b}{\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}}. = \mu$$

3. Suppose that the elements of $\boldsymbol{y}: n \times 1$ are independently (μ, σ^2) distributed. The linear model can be expressed as:

$y_i \sim (\mu, \sigma^2)$

$$E(\boldsymbol{y}) = \underset{n\times 1}{\begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix}} \underset{1\times 1}{(\mu)}. = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}$$

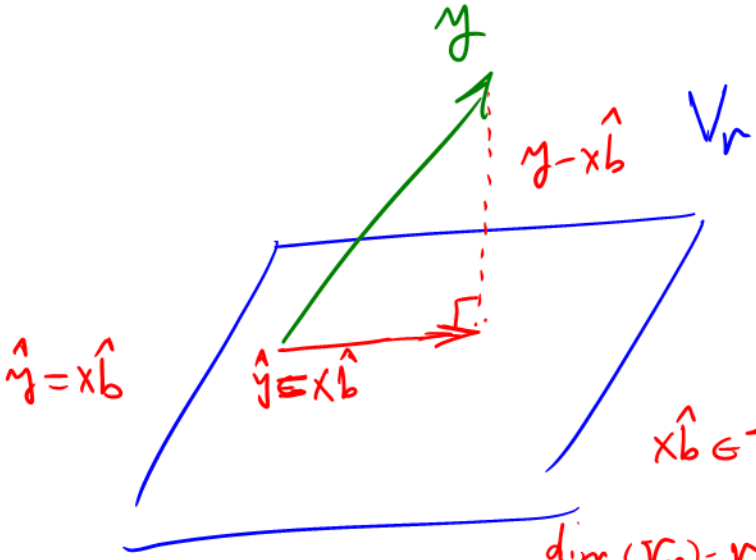
4. Suppose that $\overbrace{y_{11}, y_{12}, \cdots, y_{1n_1}}^{y_1}$ is a random sample from a (μ_1, σ^2) distribution and that $\underbrace{y_{21}, y_{22}, \cdots, y_{2n_2}}_{y_2}$ is an independent sample from a (μ_2, σ^2) distribution. The linear model can be expressed as:

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$E(\boldsymbol{y}) = E \begin{pmatrix} y_{11} \\ y_{12} \\ \cdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \cdots \\ y_{2n_2} \end{pmatrix} = \underset{n\times 2}{\overbrace{\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \cdots & \cdots \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \cdots & \cdots \\ 0 & 1 \end{pmatrix}}^{X}} \underset{2\times 1}{\overset{b}{\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}}}. = \underset{n\times 1}{\begin{pmatrix} \mu_1 \\ \mu_1 \\ \vdots \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \vdots \\ \mu_2 \end{pmatrix}}$$

5. Suppose that the elements of $\boldsymbol{y}: n \times 1$ are independently $(\alpha + \beta x_i, \sigma^2)$ distributed. As a linear model it is expressed as:

$$E(\boldsymbol{y}) = \underset{n\times 2}{\overbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{pmatrix}}^{X}} \underset{2\times 1}{\overset{b}{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}}. = \underset{n\times 1}{\begin{pmatrix} \alpha+\beta x_1 \\ \alpha+\beta x_2 \\ \vdots \\ \alpha+\beta x_n \end{pmatrix}}$$



$\mathrm{rank}(X) = r$

$E(y) \in V_r$

$V_n = V_r \oplus V_E$ error

regression

$X\hat{b} \in V_r \qquad y - X\hat{b} \in V_E$

$\dim(V_r) = r \qquad \dim(V_E) = n - r$

$L_r = \{ c^T y : c \in V_r \}$

$V = V_E \oplus V_r$

$L = L_E \oplus L_r$

$S^2 = y^T C (C^T C)^{-} C^T y$

$= S_1^2 + S_2^2$

• The Least Squares Estimators

The least squares estimates $\widehat{\boldsymbol{b}}$ of the parameters $\boldsymbol{b}$ are those values of the parameters which minimize the sum of squares

$$S^2 = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})$$

for a given value $\boldsymbol{y}$ of the random vector $\boldsymbol{y}$.

Suppose that V_e is the vector space orthogonal to V_r $(rank\ (V_e) = n - r)$. $\widehat{\boldsymbol{b}}$ must be chosen such that $\boldsymbol{X}\widehat{\boldsymbol{b}}$ is the projection of $\boldsymbol{y}$ on V_r. It then follows that

$\hat{y} - y$

$$\begin{aligned} S^2 &= \{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}}) + (\boldsymbol{X}\widehat{\boldsymbol{b}} - \boldsymbol{X}\boldsymbol{b})\}'\{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}}) + (\boldsymbol{X}\widehat{\boldsymbol{b}} - \boldsymbol{X}\boldsymbol{b})\} \\ &= \underbrace{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}})}_{S_1} + \underbrace{(\boldsymbol{X}\widehat{\boldsymbol{b}} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{X}\widehat{\boldsymbol{b}} - \boldsymbol{X}\boldsymbol{b})}_{S_2} \end{aligned}$$

since $\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}} \in V_e$ and $\boldsymbol{X}\widehat{\boldsymbol{b}} - \boldsymbol{X}\boldsymbol{b} \in V_r$. S^2 is therefore a minimum if and only if $\boldsymbol{X}\hat{\boldsymbol{b}}$ is the projection of $\boldsymbol{y}$ on V_r. Thus $\widehat{\boldsymbol{b}}$ is a set of least squares estimates if and only if

$\hat{y} = x\hat{b}$

$$\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{b}} = \boldsymbol{X}'\boldsymbol{y}.$$

These equations are called the normal equations.

The normal equations are also obtained by setting the partial derivatives of S^2 with respect to the parameters equal to zero. It follows that:

$$S^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} b_j x_{ij} \right)^2$$

$$\begin{aligned} \frac{\partial S^2}{\partial b_v} &= (-2) \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} b_j x_{ij} \right) x_{iv} \\ &= (-2)(\boldsymbol{x}_v'\boldsymbol{y} - \boldsymbol{x}_v'\boldsymbol{X}\boldsymbol{b}) \\ &= 0 \qquad v = 1, 2, \cdots, p \end{aligned}$$

which implies the normal equations.

$$y = X\beta + \epsilon \qquad \epsilon \sim N_n(0, \sigma^2 I_n)$$

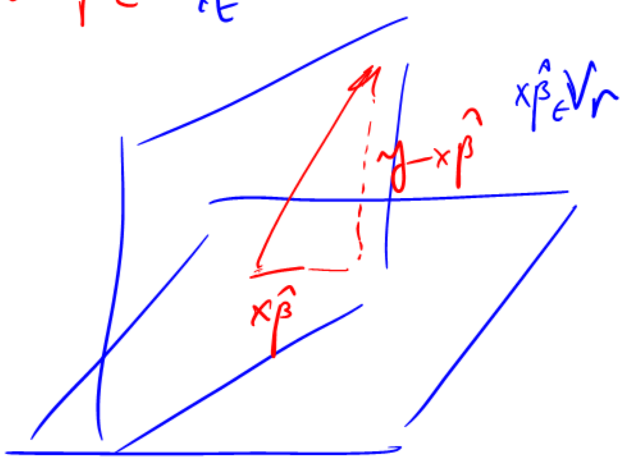
$$L(\beta) = L(\beta \mid X, y) = L(\epsilon)$$

$$= \frac{1}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2\sigma^2}(y - X\beta)^T (y - X\beta)}$$

$$l(\beta \mid X, y) \propto -\frac{1}{2\sigma^2}(y - X\beta)^T (y - X\beta)$$

$$\frac{\partial l(\beta)}{\partial \beta} = 0 \implies (X^T X)\beta = X^T y$$

$y - X\hat{\beta} \in V_E$



$X = (x_1, \ldots, x_r)$

$B = (b_1, b_2, \ldots, b_r)$

$B^T X = 0$

Consequently, if $\boldsymbol{y}$ has a normal distribution with density function

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{Xb})'(\boldsymbol{y} - \boldsymbol{Xb})\right)$$

it follows that a set of least squares estimators of $\boldsymbol{b}$ is also a set of maximum likelihood estimators.

• The Error Space and the Error Set

The vector space V_e is defined as the error space and the linear set

$$L_e = \{\boldsymbol{e}'\boldsymbol{y} \;:\; \boldsymbol{e} \in V_e\}$$

is defined as the error set. The sum of squares for L_e is called the sum of squares for errors, namely

$$\underline{SSE} = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}}) = \boldsymbol{y}'\boldsymbol{y} - \widehat{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{y}$$

$y^T y - y^T X\hat{b} - \hat{b}^T X^T y + \hat{b}^T X^T X b$

where $\widehat{\boldsymbol{b}}$ is any set of least squares estimators.

The expected value of any linear function $\boldsymbol{e}'\boldsymbol{y}$ is

$E(e^T y) \in V_e$

$$E(\boldsymbol{e}'\boldsymbol{y}) = \boldsymbol{e}'\boldsymbol{X}\boldsymbol{b}.$$

It follows that $\underline{E(\boldsymbol{e}'\boldsymbol{y}) \equiv \boldsymbol{0}}$, independent of $\boldsymbol{b}$, if and only if $\boldsymbol{e}'\boldsymbol{X} = \boldsymbol{0}'$, that is, if and only if $\boldsymbol{e}'\boldsymbol{y} \in L_e$. In other words, $E(\boldsymbol{e}'\boldsymbol{y}) \equiv \boldsymbol{0}$ if and only if $\boldsymbol{e} \in V_e$.