Model

Suppose we conduct an experiment that gives rise to the random variables $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$.

1. $\mathbb{E}(y_i)$ is a linear function of the parameters $\beta_1, \ldots, \beta_p$ with known coefficients. In matrix notation

$$\mathbb{E}(\mathbf{y}) = \mathrm{X}\boldsymbol{\beta}, \tag{1}$$

   (a) $\mathbf{y}$ is a $n \times 1$ vector with components $y_1, \ldots, y_n$.
   (b) X is a known nonrandom matrix of order $n \times p$.
   (c) $\boldsymbol{\beta}$ is the $p \times 1$ vector of parameters $\beta_1, \ldots, \beta_p$.

Model

2. $y_i' s$ are uncorrelated i.e. $\text{Cov}(y_i, y_j) = 0$ for all $i \neq j; i, j \in \{1, \ldots, n\}$.

3. y_i's are homoscedastic that is $\text{Var}(y_i) = \sigma^2$ for all $i = 1, \ldots, n$.

Thus,

$$\text{Var}(\mathbf{y}) = \mathrm{D}(\mathbf{y}) = \sigma^2 \mathrm{I}. \tag{2}$$

Model

Another way to write the model is

$$\mathbf{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3}$$

where the vector $\boldsymbol{\epsilon}$ satisfies $\mathbb{E}(\boldsymbol{\epsilon}) = 0$ and $\mathrm{D}(\boldsymbol{\epsilon}) = \sigma^2\mathrm{I}$.

Estimability

Consider the linear model

$$\mathbb{E}(\mathbf{y}) = \mathrm{X}\boldsymbol{\beta},\ \mathrm{D}(\mathbf{y}) = \sigma^2\mathrm{I}. \tag{4}$$

Definition 1

Estimability The linear parametric function $\mathbf{l}'\boldsymbol{\beta}$ is said to be estimable if there exists a linear function $\mathbf{c}'\mathbf{y}$ of the observations such that $\mathbb{E}(\mathbf{c}'\mathbf{y}) = \mathbf{l}'\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$.

Estimability

Result 1: The condition $\mathbb{E}(\mathbf{c}'\mathbf{y}) = \mathbf{l}'\boldsymbol{\beta}$ is equivalent to $\mathbf{c}'\mathrm{X}\boldsymbol{\beta} = \mathbf{l}'\boldsymbol{\beta}$, and since this must hold for all $\boldsymbol{\beta} \in \mathbb{R}^p$, we must have $\mathbf{c}'\mathrm{X} = \mathbf{l}'$. Thus $\mathbf{l}'\boldsymbol{\beta}$ is estimable if and only if $\mathbf{l}' \in \mathcal{R}(\mathrm{X})$.

Estimability

Example: Consider the following model

$$\mathbb{E}(y_{ij}) = \alpha_i + \beta_j, \; i = 1, 2; \; j = 1, 2.$$

In matrix form

$$\mathbb{E}\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Estimability

Let S be the set of all vectors $\begin{pmatrix} l_1 & l_2 & m_1 & m_2 \end{pmatrix}$ such that $l_1 + l_2 = m_1 + m_2$. Note that if $\mathbf{x} \in \mathcal{R}(\mathrm{X})$, then $\mathbf{x} \in S$. Thus $\mathcal{R}(\mathrm{X}) \subset S$. Also, $\dim(S) = 3$, and the rank of X is 3 as well. Therefore, $\mathcal{R}(\mathrm{X}) = S$, and we conclude that $l_1\alpha_1 + l_2\alpha_2 + m_1\beta_1 + m_2\beta_2$ is estimable if and only if $l_1 + l_2 = m_1 + m_2$.

Results

Result 1: The following facts concerning generalized inverse are frequently used. For any matrix X, $\mathcal{R}(\mathrm{X}) = \mathcal{R}(\mathrm{X}'\mathrm{X})$.

Proof: Following Theorem is needed to prove this result.

Theorem 2

Let A be an $m \times n$ matrix of rank r. Then $dim(\mathcal{N}(A)) = n - r$, where $\mathcal{N}(A)$ is the null space of A.

Results

Clearly, $\mathcal{R}(\mathrm{X}'\mathrm{X}) \subset \mathcal{R}(\mathrm{X})$. Now observe that $\mathrm{X}\mathsf{z} = 0$ implies $\mathrm{X}'\mathrm{X}\mathsf{z} = 0$ for any vector $\mathsf{z} \in \mathbb{R}^n$. Next, if $\mathrm{X}'\mathrm{X}\mathsf{z} = 0$ then $\mathsf{z}'\mathrm{X}'\mathrm{X}\mathsf{z} = (\mathrm{X}\mathsf{z})'(\mathrm{X}\mathsf{z}) = 0$ implying that $\mathrm{X}\mathsf{z} = 0$. Thus, by Theorem 2, X and $\mathrm{X}'\mathrm{X}$ have the same rank. and therefore their row spaces have the same dimension. This implies that the spaces must be equal. As a consequence we can write $\mathrm{X} = \mathrm{M}\mathrm{X}'\mathrm{X}$ for some matrix M.

Results

Result 2: The matrix $\mathrm{AC^{-}B}$ is invariant under the choice of the g–inverse C^{-} of C if $\mathcal{C}(\mathrm{B}) \subset \mathcal{C}(\mathrm{C})$ and $\mathcal{R}(\mathrm{A}) \subset \mathcal{R}(\mathrm{C})$.

Proof: We can write $\mathrm{B} = \mathrm{CU}$ and $\mathrm{A} = \mathrm{VC}$ for some matrices U, V. Then

$$\mathrm{AC^{-}B} = \mathrm{VCC^{-}CU} = \mathrm{VCU},$$

which does not depend on the choice of the g–inverse.

Results

Result 3: The matrix $\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{X}$ is invariant under the choice of the g–inverse. This is immediate from **Result 2**since $\mathcal{R}(\mathrm{X}) = \mathcal{R}(\mathrm{X}'\mathrm{X})$.

Result 4: $\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{X} = \mathrm{X}$ and $\mathrm{X}'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'$. Than can be shown by writing $\mathrm{X} = \mathrm{M}\mathrm{X}'\mathrm{X}$.

BLUE

Theorem 3

Let $\mathbf{l}'\boldsymbol{\beta}$ be an estimable function and let G be a least squares g–inverse of X. Then $\mathbf{l}'\mathrm{G}\mathbf{y}$ is an unbiased linear estimate of $\mathbf{l}'\boldsymbol{\beta}$ with minimum variance among all unbiased linear estimates of $\mathbf{l}'\boldsymbol{\beta}$. $\mathbf{l}'\mathrm{G}\mathbf{y}$ is said to be BLUE (best linear unbiased estimate) of $\mathbf{l}'\boldsymbol{\beta}$. The variance of $\mathbf{l}'\mathrm{G}\mathbf{y}$ is $\sigma^2\mathbf{l}'(\mathrm{X}'\mathrm{X})^{-}\mathbf{l}$.

BLUE

Proof: Since $\mathbf{l}'\beta$ is estimable, $\mathbf{l}' = \mathbf{u}'\mathrm{X}$ for some $\mathbf{u}$. Then

$$\mathbb{E}(\mathbf{l}'\mathrm{G}\mathbf{y}) = \mathbf{u}'\mathrm{XGX}\boldsymbol{\beta} = \mathbf{u}'\mathrm{X}\boldsymbol{\beta} = \mathbf{l}'\boldsymbol{\beta},$$

and hence $\mathbf{l}'\mathrm{G}\mathbf{y}$ is unbiased for $\mathbf{l}'\boldsymbol{\beta}$. Any other linear unbiased estimate is of the form $(\mathbf{l}'\mathrm{G} + \mathrm{w}')\mathbf{y}$, where $\mathrm{w}'\mathrm{X} = 0$ (why?).

BLUE

Now

$$\begin{aligned}\mathsf{Var}\{(\mathbf{l}'\mathrm{G}+\mathsf{w}')\mathbf{y}\} &= \sigma^2(\mathbf{l}'\mathrm{G}+\mathsf{w}')((\mathrm{G}'\mathbf{l}+\mathsf{w})) \\ &= \sigma^2(\mathbf{u}'\mathrm{X}\mathrm{G}+\mathsf{w}')((\mathrm{G}'\mathrm{X}'\mathbf{u}+\mathsf{w}))\end{aligned}$$

Since G is a least square g–inverse of X,

$$\mathbf{u}'\mathrm{X}\mathrm{G}\mathsf{w} = \mathbf{u}'\mathrm{G}'\mathrm{X}'\mathsf{w} = 0.$$

BLUE

Therefore,

$$\begin{aligned} \mathrm{Var}\{(\mathbf{l}'\mathrm{G} + \mathrm{w}')\mathbf{y}\} &= \sigma^2(\mathbf{u}'(\mathrm{XG})(\mathrm{XG})'\mathbf{u} + \mathrm{w}'\mathrm{w}) \\ &\geq \sigma^2\mathbf{u}'(\mathrm{XG})(\mathrm{XG})'\mathbf{u} \\ &= \mathrm{Var}(\mathbf{l}'\mathrm{G}\mathbf{y}). \end{aligned}$$

Hence, $\mathbf{l}'\mathrm{G}\mathbf{y}$ is BLUE of $\mathbf{l}'\boldsymbol{\beta}$.

BLUE

The variance of $\mathbf{l}'\mathrm{G}\mathbf{y}$ is $\sigma^2\mathbf{l}\mathrm{G}\mathrm{G}'\mathbf{l}$. It is easily seen that for any choice of g–inverse, $(\mathrm{X}'\mathrm{X})^-\mathrm{X}'$ is a least squares g–inverse of X. Now,

$$\begin{aligned}
\mathbf{l}\mathrm{G}\mathrm{G}'\mathbf{l} &= \mathbf{l}'(\mathrm{X}'\mathrm{X})^-\mathrm{X}'\mathrm{X}\{(\mathrm{X}'\mathrm{X})^-\}'\mathbf{l} \\
&= \mathbf{u}'\mathrm{X}(\mathrm{X}'\mathrm{X})^-\mathrm{X}'\mathrm{X}\{(\mathrm{X}'\mathrm{X})^-\}'\mathrm{X}'\mathbf{u} \\
&= \mathbf{u}'\mathrm{P}_\mathrm{X}\mathrm{P}_\mathrm{X}'\mathbf{u} \\
&= \mathbf{u}'\mathrm{P}_\mathrm{X}\mathbf{u} \\
&= \mathbf{u}'\mathrm{X}(\mathrm{X}'\mathrm{X})^-\mathrm{X}'\mathbf{u} \\
&= \mathbf{l}'(\mathrm{X}'\mathrm{X})^-\mathbf{l}.
\end{aligned}$$

Thus,

$$\mathrm{Var}(\mathbf{l}'\mathrm{G}\mathbf{y}) = \sigma^2\mathbf{l}'(\mathrm{X}'\mathrm{X})^-\mathbf{l}.$$

g–inverse

In particular, if we choose $\mathrm{G} = (\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'$ then

- G is a g–inverse of X:

$$\mathrm{XGX} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{X} = \mathrm{X}$$

- G is a least square g–inverse.

$$(\mathrm{XG})' = (\mathrm{X}(\mathrm{X}'\mathrm{X})^{-}\mathrm{X}')' = \mathrm{P}_{\mathrm{X}}' = \mathrm{P}_{\mathrm{X}} = \mathrm{XG}.$$

g–inverse

Consider the model (1).

- A least square estimator minimizes $S(\boldsymbol{\beta}) = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})$.
- Differentiating $S(\boldsymbol{\beta})$ wrt $\boldsymbol{\beta}$ result in the normal equations $\mathrm{X}'\mathrm{X}\boldsymbol{\beta} = \mathrm{X}'\mathbf{y}$.
- Let $\widehat{\boldsymbol{\beta}}$ is any solution to the normal equations. Then $\widehat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-}\mathrm{X}\mathbf{y}$ for some choice of g–inverse.
- $\widehat{\boldsymbol{\beta}}$ depends on the choice of g–inverse. Thus $\widehat{\boldsymbol{\beta}}$ is not unique and does not admit any statistical interpretation.

Estimability

Example: Let us revisit the example we discussed earlier. Consider the following model

$$\mathbb{E}(y_{ij}) = \alpha_i + \beta_j,\ i = 1, 2;\ j = 1, 2.$$

In matrix form

$$\mathbb{E}\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Estimability

$$X'X = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}$$

and

$$(X'X)^{-} = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & -2 & -2 \\ 0 & -2 & 3 & 1 \\ 0 & -2 & 1 & 3 \end{bmatrix}$$

in one possible g–inverse.

Estimability

$$X(X'X)^{-}X' = \frac{1}{4}\begin{bmatrix} 3 & 1 & 1 & -1 \\ 1 & 3 & -1 & 1 \\ 1 & -1 & 3 & 1 \\ -1 & 1 & 1 & 3 \end{bmatrix}$$

Now we can compute the BLUE of any estimable function $\mathbf{u}'X\boldsymbol{\beta}$ as $\mathbf{u}'X(X'X)^{-}X'\mathbf{y}$.

Estimability

For example, if $\mathbf{u}' = (1, 0, 0, 0)$, then we get the BLUE of $\alpha_1 + \beta_1$ as

$$\frac{1}{4}(3y_{11} + y_{12} + y_{21} - y_{22}).$$

Full–rank model

The model (2) is said to be a *full–rank model* if X has full column rank, i.e. $\mathsf{Rank}(\mathrm{X}) = p$. For a full–rank model the following results can be easily verified.

(i) $\mathcal{R}(\mathrm{X}) = \mathbb{R}^p$, and therefore every function $\mathbf{l}'\boldsymbol{\beta}$ is estimable.

(ii) $\mathrm{X}'\mathrm{X}$ is nonsingular.

(iii) The BLUE of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathbf{y}$ and the variance matrix of $\widehat{\boldsymbol{\beta}}$ is $\mathsf{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathrm{X}'\mathrm{X})^{-1}$.

(iv) The BLUE of $\mathbf{l}'\boldsymbol{\beta}$ is $\mathbf{l}'\widehat{\boldsymbol{\beta}}$ with variance $\sigma^2\mathbf{l}'(\mathrm{X}'\mathrm{X})^{-1}\mathbf{l}$.

Parts (iii) and (iv) constitute the Gauss–Markov theorem.

RSS

For a linear regression model $\mathbf{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, the aim is to minimize the error i.e. $\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) = S(\boldsymbol{\beta})$.

Theorem 4

The minimum of $(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})$ *is attained at* $\widehat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathbf{y}$.