

Recap:

Best linear unbiased ~~estimator~~ estimator for the Gauss-Markov model  $\underline{y} = \underline{X}\beta + \underline{e}$ ,  $E[\underline{e}] = \underline{0}$  and  $\text{var}(\underline{e}) = \sigma^2 \underline{I}_n$

$\underline{\lambda}^T \hat{\beta}$  is the BLUE of  $\underline{\lambda}^T \beta$ .

Let  $\underline{\lambda}_1, \dots, \underline{\lambda}_m$  are  $m$  vectors and let

$\underline{\Lambda}$  be a  $p \times m$  matrix s.t.

$$\underline{\Lambda}^T = \begin{pmatrix} \underline{\lambda}_1^T \\ \vdots \\ \underline{\lambda}_m^T \end{pmatrix} \Rightarrow \underline{\Lambda}^T \beta = \begin{pmatrix} \underline{\lambda}_1^T \beta \\ \vdots \\ \underline{\lambda}_m^T \beta \end{pmatrix}$$

The best linear unbiased estimator of  $\underline{\Lambda}^T \beta$  is  $\underline{\Lambda}^T \hat{\beta}$ .

If  $\underline{c}^T \underline{y}$  be any other unbiased estimator of  $\underline{\Lambda}^T \beta$

then  $\text{var}(\underline{c}^T \underline{y}) - \text{var}(\underline{\Lambda}^T \hat{\beta}) \geq 0$

Whenever we write  $\underline{A} - \underline{B} \geq 0$  where  $\underline{A}$  and  $\underline{B}$  are  $m \times m$  matrices, we mean  $\underline{A} - \underline{B}$  is nonnegative definite.

Consider the simple example.

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad E[e_i] = 0, \quad \text{var}(e_i) = \sigma^2$$

What is the BLUE of  ~~$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$~~   $\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ ?

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \text{ is the BLUE.} \\ = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$

$$\text{Var}\left(\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}\right) = \text{Var}\left(\begin{pmatrix} 1 \\ x_1 \end{pmatrix} \left(\beta_0 \right)\right) = \sigma^2 (\underline{X}'\underline{X})^{-1}$$

$$\underline{X}'\underline{X} = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

$$(\underline{X}'\underline{X})^{-1} = \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \frac{1}{(n \sum x_i^2 - (\sum x_i)^2)}$$

Now let's try to unbiasedly estimate  $\sigma^2$ .

$\underline{y}$  is the data vector, recall  $\hat{\underline{y}} = \underline{P}_X \underline{y} = \underline{P} \underline{y}$

$$\hat{\underline{e}} = \underline{y} - \underline{P} \underline{y} = (\underline{I} - \underline{P}) \underline{y}$$

$$\|\hat{\underline{e}}\|^2 = \hat{\underline{e}}' \hat{\underline{e}} = \underline{y}' (\underline{I} - \underline{P}) (\underline{I} - \underline{P}) \underline{y} = \underline{y}' (\underline{I} - \underline{P}) \underline{y}$$

claim: An unbiased estimator of  $\sigma^2$  is  $\frac{\|\hat{\underline{e}}\|^2}{n-k}$ .

where  $k = \text{rank}(\underline{X})$ .

$\|\hat{\underline{e}}\|^2 =$  sum of square due to errors = SSE.

$\Rightarrow$  an unbiased estimator of  $\sigma^2$  is  $\frac{\text{SSE}}{n-k}$ .

lemma: Let  $\underline{z}$  be a random vector with  $E[\underline{z}] = \underline{\mu}$

and  $\text{Cov}(\underline{z}) = \underline{\Sigma}$ , then  $E[\underline{z}' \underline{A} \underline{z}] = \underline{\mu}' \underline{A} \underline{\mu} + \text{tr}(\underline{A} \underline{\Sigma})$

$$\text{pf: } E[\underline{z}' \underline{A} \underline{z}] = E[\text{tr}(\underline{z}' \underline{A} \underline{z})] = \text{tr}(E(\underline{z}' \underline{A} \underline{z}))$$

$$= E[\text{tr}(\underline{A} \underline{z} \underline{z}' \underline{A} \underline{z})] = \text{tr}(E[\underline{A} \underline{z} \underline{z}' \underline{A} \underline{z}])$$

$$= \text{tr}(\underline{A} E[\underline{z} \underline{z}' \underline{A} \underline{z}])$$

$$\text{Cov}(\underline{z}) = E[(\underline{z} - \underline{\mu})(\underline{z} - \underline{\mu})^T]$$

$$= E[\underline{z}\underline{z}^T] - \underline{\mu}\underline{\mu}^T$$

$$\Rightarrow E[\underline{z}\underline{z}^T] = \underline{\mu}\underline{\mu}^T + \underline{\Sigma}$$

$$\Rightarrow E[\underline{z}^T \underline{A} \underline{z}] = \text{tr}(\underline{A}(\underline{\mu}\underline{\mu}^T + \underline{\Sigma}))$$

$$= \text{tr}(\underline{A}\underline{\mu}\underline{\mu}^T) + \text{tr}(\underline{A}\underline{\Sigma})$$

$$= \text{tr}(\underline{\mu}^T \underline{A} \underline{\mu}) + \text{tr}(\underline{A}\underline{\Sigma}) = \underline{\mu}^T \underline{A} \underline{\mu} + \text{tr}(\underline{A}\underline{\Sigma})$$

$$\frac{\text{SSE}}{n-k} = \frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{n-k}$$

$$E[\underline{y}^T (\underline{I} - \underline{P}) \underline{y}] = (\underline{x} \underline{\beta})^T (\underline{I} - \underline{P}) \underline{x} \underline{\beta} + \text{tr}((\underline{I} - \underline{P}) \underline{\sigma}^2)$$

$$= \underline{\beta}^T \underline{x}^T (\underline{I} - \underline{P}) \underline{x} \underline{\beta} + \underline{\sigma}^2 \text{tr}(\underline{I} - \underline{P})$$

$$= \underline{\sigma}^2 \text{tr}(\underline{I} - \underline{P}) = \underline{\sigma}^2 \text{rank}(\underline{I} - \underline{P}) \quad (\underline{I} - \underline{P} \text{ idempotent})$$

$$= \underline{\sigma}^2 (n-k)$$

$\underline{A} = \underline{Q} \underline{\Lambda} \underline{Q}^T$  for  $\underline{A}$  to be idempotent

$$\underline{A}^2 = \underline{A} \Rightarrow \underline{Q} \underline{\Lambda}^2 \underline{Q}^T = \underline{Q} \underline{\Lambda} \underline{Q}^T$$

$\Rightarrow \underline{\Lambda}^2 = \underline{\Lambda} \Rightarrow$  every eigenvalue must satisfy

$$\lambda_i(1 - \lambda_i) = 0 \Rightarrow \lambda_i = 0 \text{ or } 1$$

$\text{rank}(\underline{A}) = k \Rightarrow$  only  $k$  eigenvalues are 1, rest are zero.

$$\Rightarrow \text{tr}(\underline{A}) = \sum_{i=1}^n \lambda_i = k$$

$$\Rightarrow E\left[\frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{n-k}\right] = \underline{\sigma}^2 \Rightarrow \frac{\text{SSE}}{n-k} \text{ is unbiased for } \underline{\sigma}^2.$$

Consider a more general model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{e}, \text{ where } E[\underline{e}] = \underline{0} \text{ and } \text{Var}(\underline{e}) = \sigma^2 \underline{V}$$

Assume  $\underline{V}$  is a positive definite matrix.

~~⊗~~ If  $\underline{V}$  is positive definite, so is  $\underline{V}^{-1}$

$$\text{and } \underline{V}^{-1} = \underline{R}\underline{R}'$$

Pre-multiply the original linear model equation by  $\underline{R}$ .  $\Rightarrow$

$$\underbrace{\underline{R}\underline{y}}_{\underline{y}^*} = \underbrace{\underline{R}\underline{X}}_{\underline{X}^*} \underline{\beta} + \underbrace{\underline{R}\underline{e}}_{\underline{e}^*}$$

$$\bullet E[\underline{e}^*] = \underline{0}$$

$$\text{Var}(\underline{e}^*) = \underline{R}\sigma^2\underline{V}\underline{R}' = \sigma^2 \underline{I}$$

$$\text{N.E.S. } [(\underline{R}\underline{X})'(\underline{R}\underline{X})] \underline{\beta} = (\underline{R}\underline{X})'(\underline{R}\underline{y})$$

$$\Rightarrow [\underline{X}'\underline{R}'\underline{R}\underline{X}] \underline{\beta} = \underline{X}'\underline{R}'\underline{R}\underline{y}$$

$$\bullet A = \underline{Q}\underline{\Lambda}\underline{Q}' \quad \underline{A}^{1/2} = \underline{Q}\underline{\Lambda}^{1/2}\underline{Q}'$$

$$\Rightarrow (\underline{X}'\underline{V}^{-1}\underline{X}) \underline{\beta} = \underline{X}'\underline{V}^{-1}\underline{y}$$

~~⊗~~ If  $(\underline{X}'\underline{V}^{-1}\underline{X})$  is invertible  $\Rightarrow$  LSE of  $\underline{\beta}$

$$\text{in } \hat{\underline{\beta}} = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1} \underline{X}'\underline{V}^{-1}\underline{y}$$

This estimator is called the Generalized Least Square estimator or GLS estimator

Typically in GLS estimators  $\underline{V}$  is taken to be a diagonal matrix with unequal diagonal entries.

$\hat{\underline{\beta}}_{GLS}$  is a generalized least square estimator of  $\underline{\beta}$  iff  $\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}\underline{y} = \underline{X}\hat{\underline{\beta}}_{GLS} = \hat{\underline{y}}$ .

$\underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}$  is the projection matrix in this case.

For this projection matrix  $\underline{P}\underline{w} \neq \underline{0}$  when  $\underline{w}$  belongs to the orthogonal space of  $C(\underline{X})$ .

~~Then:~~ If  $\underline{\lambda}'\underline{\beta}$  is estimable, then the generalized least square estimate has  $\text{Var}(\underline{\lambda}'\hat{\underline{\beta}}) = \underline{V}^{-1}\underline{\lambda}'(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{\lambda}$ .

Since  $C(\underline{X}) = C(\underline{X}^*)$  and similarly  $C(\underline{X}^T) = C(\underline{X}^{*T})$

$\Rightarrow \underline{\lambda}'\underline{\beta}$  is estimable in the previous model, it is also estimable in the new model.

because  $\underline{\lambda} \in C(\underline{X}^{*T}) = C(\underline{X}^T)$ .

Estimation of  $\underline{V}^{-1}$ .

$$\begin{aligned} \hat{\underline{V}}_{GLS}^{-1} &= \frac{(\underline{y}^* - \underline{X}^*\hat{\underline{\beta}})'(\underline{y}^* - \underline{X}^*\hat{\underline{\beta}})}{n-k} = \frac{(\underline{R}\underline{y} - \underline{R}\underline{X}\hat{\underline{\beta}})'(\underline{R}\underline{y} - \underline{R}\underline{X}\hat{\underline{\beta}})}{n-k} \\ &= \frac{(\underline{y} - \underline{X}\hat{\underline{\beta}})'\underline{R}'\underline{R}(\underline{y} - \underline{X}\hat{\underline{\beta}})}{n-k} = \frac{(\underline{y} - \underline{X}\hat{\underline{\beta}})'\underline{V}^{-1}(\underline{y} - \underline{X}\hat{\underline{\beta}})}{n-k} \end{aligned}$$

Gauss-Markov model:  $E[\underline{e}] = \underline{0}$ ,  $\text{Var}(\underline{e}) = \sigma^2 \underline{I}$

Form the linear model

$$\underline{y} = \underline{X} \underline{\beta} + \underline{e}, \quad \text{we assume } \underline{e} \sim N_n(\underline{0}, \sigma^2 \underline{I}_n)$$

$$p(\underline{z}) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad m_z(t) = E[e^{t z}] = \exp\left(\frac{t^2}{2}\right)$$

$$S \sim N(\mu, \sigma^2) \Rightarrow S = \mu + \sigma z$$

$$m_S(t) = \exp\left(t\mu + \frac{t^2 \sigma^2}{2}\right)$$

$$\underline{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix} \quad \underline{z} \sim N(\underline{0}, \underline{I}_p)$$

$$p(\underline{z}) = \left(\frac{1}{2\pi}\right)^{p/2} \exp\left(-\frac{1}{2} \underline{z}^T \underline{z}\right)$$

$$m_{\underline{z}}(\underline{t}) = E\left[e^{\underline{t}^T \underline{z}}\right] = \exp\left(\frac{\underline{t}^T \underline{t}}{2}\right)$$

If  $\underline{s} \sim N(\underline{\mu}, \underline{\Sigma})$  we know

$$\underline{s} = \underline{\mu} + \underline{A} \underline{z} \quad \text{where } \underline{A} \underline{A}^T = \underline{\Sigma}$$

$$m_{\underline{s}}(\underline{t}) = \exp\left(\underline{t}^T \underline{\mu} + \frac{\underline{t}^T \underline{\Sigma} \underline{t}}{2}\right)$$

$$\text{If } \underline{y} = \underline{X} \underline{\beta} + \underline{e}, \quad \underline{e} \sim N(\underline{0}, \sigma^2 \underline{V})$$

$$\Rightarrow R \underline{y} = R \underline{X} \underline{\beta} + R \underline{e}, \quad R \underline{e} \sim N(\underline{0}, \sigma^2 \underline{I})$$

Result: If  $\underline{s} \sim N_p(\underline{\mu}, \underline{V})$  and  $\underline{y} = \underline{a} + \underline{B} \underline{s}$   
where  $\underline{a}$  is  $q \times 1$  vector  $\underline{B}$  is a  $q \times p$  matrix  
then  $\underline{y} \sim N_q(\underline{a} + \underline{B} \underline{\mu}, \underline{B} \underline{V} \underline{B}^T)$

Result: If  $\underline{x} \sim N_p(\underline{\mu}, \underline{V})$  and  $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

then  $\text{cov}(x_1, x_2) = \underline{0}$  iff  $x_1$  and  $x_2$  are independent.

Cor: Let  $\underline{x} \sim N_p(\underline{\mu}, \underline{V})$  and  $\underline{y}_1 = \underline{a}_1 + \underline{B}_1 \underline{x}$ ,

$\underline{y}_2 = \underline{a}_2 + \underline{B}_2 \underline{x}$  then  $\underline{y}_1$  and  $\underline{y}_2$  are independent

iff  $\underline{B}_1 \underline{V} \underline{B}_2^T = \underline{0}$  ( $\text{cov}(\underline{B}_1 \underline{x} + \underline{a}_1, \underline{B}_2 \underline{x} + \underline{a}_2)$   
 $= \underline{B}_1 \text{Var}(\underline{x}) \underline{B}_2^T$ )

$$\underline{y} = \underline{X} \underline{\beta} + \underline{e}, \quad \underline{e} \sim N(\underline{0}, \sigma^2 \underline{I}_n)$$

$$\hat{\underline{y}} = \underline{P} \underline{y} \quad \text{and} \quad \hat{\underline{e}} = (\underline{I} - \underline{P}) \underline{y}$$

$$\hat{\underline{y}} = \underline{P} \underline{y} \sim N(\underline{P} \underline{X} \underline{\beta}, \sigma^2 \underline{P} \underline{P}^T) = N(\underline{X} \underline{\beta}, \sigma^2 \underline{P})$$

$$\hat{\underline{e}} = (\underline{I} - \underline{P}) \underline{y} \sim N((\underline{I} - \underline{P}) \underline{X} \underline{\beta}, \sigma^2 (\underline{I} - \underline{P})) = N(\underline{0}, \sigma^2 (\underline{I} - \underline{P}))$$

Result:  $\hat{\underline{y}}$  and  $\hat{\underline{e}}$  are independent as

$$(\underline{I} - \underline{P}) \underline{P} = \underline{0} \quad (\text{Use the previous corollary})$$