STA3010 Regression Analysis

Feng Yin

The Chinese University of Hong Kong (Shenzhen)

March 19, 2020

General Linear Model

We consider a more general multiple linear regression model:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where we assume

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad Cov(\boldsymbol{\varepsilon}) = \sigma^2 V \tag{2}$$

with

σ^2 being unknown, V being known.

Note:

$V \in \mathbb{R}^{n \times n}$ is assumed to be a positive definite matrix, revealing the structure of variances and covariances among random errors in $\boldsymbol{\varepsilon}$.

Transformation

Since V is non-singular and positive definite, we can factorize it as

$$V = K^T K = KK, \tag{3}$$

where K is the symmetric, square-root of V, which is also positive definite.

Due to the above factorization, we transform the data as

$$\boldsymbol{z} = K^{-1}\boldsymbol{y}, \quad B = K^{-1}X, \quad \boldsymbol{g} = K^{-1}\varepsilon, \tag{4}$$

and obtain a new multiple linear regression model:

$$\boldsymbol{z} = B\boldsymbol{\beta} + \boldsymbol{g}, \tag{5}$$

for which we can prove that

$$\mathbb{E}(\boldsymbol{g}) = \mathbf{0}, \quad Cov(\boldsymbol{g}) = \sigma^2 I$$

Generalized LS (GLS) Estimator of β

The GLS model parameter estimator $\hat{\boldsymbol{\beta}}$ is obtained as

$$\begin{aligned}\hat{\boldsymbol{\beta}} = S(\boldsymbol{\beta}) &\equiv \arg\min_{\boldsymbol{\beta}} \ (\boldsymbol{z} - B\boldsymbol{\beta})^T(\boldsymbol{z} - B\boldsymbol{\beta}) \\ &= \arg\min_{\boldsymbol{\beta}} \ (\boldsymbol{y} - X\boldsymbol{\beta})^T V^{-1}(\boldsymbol{y} - X\boldsymbol{\beta}).\end{aligned} \tag{6}$$

Similarly, taking the derivative of the cost function $S(\boldsymbol{\beta})$ w.r.t $\boldsymbol{\beta}$ and setting it equal to zero, yields

$$\hat{\boldsymbol{\beta}} = (X^T V^{-1} X)^{-1} X^T V^{-1} \boldsymbol{y}. \tag{7}$$

We can derive the following properties of $\hat{\boldsymbol{\beta}}$:

- $\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$,
- $Cov(\hat{\beta}) = \sigma^2 (X^T V^{-1} X)^{-1}$.

Note:

When we assume $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 V)$, then $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (X^T V^{-1} X)^{-1})$.

GLS estimator of σ^2

We define SS_{Res} as

$$\begin{aligned} SS_{Res} &= (\boldsymbol{z} - \hat{\boldsymbol{z}})^T (\boldsymbol{z} - \hat{\boldsymbol{z}}) \\ &= \boldsymbol{y}^T A \, \boldsymbol{y} \end{aligned} \tag{8}$$

where $A \triangleq (V^{-1} - V^{-1}X(X^T V^{-1} X)^{-1} X^T V^{-1})$.

Theorem (Normal distribution is NOT assumed!)

If A is a $k \times k$ matrix of constants, and $\boldsymbol{y}$ is a $k \times 1$ random vector with mean $\boldsymbol{\mu}$ and non-singular covariance matrix Σ, then

- $\mathbb{E}(\boldsymbol{y}^T A \, \boldsymbol{y}) = \text{trace}(A\Sigma) + \boldsymbol{\mu}^T A \, \boldsymbol{\mu}$.

- We apply the above theorem and obtain

$$\begin{aligned} \mathbb{E}(SS_{Res}) &= tr(\sigma^2 V \cdot A) + (X\beta)^T A X\beta \\ &= \sigma^2 (n-p) \end{aligned} \tag{9}$$

- We let $MS_{Res} = \frac{SS_{Res}}{n-p} = \frac{\mathbf{y}^T A \ \mathbf{y}}{n-p}$ to be the GLS estimator of σ^2. It can shown that $\mathbb{E}(MS_{Res}) = \sigma^2 \rightarrow$(unbiased estimator of σ^2) .

Generalized Gauss-Markov Theorem

- $\hat{\boldsymbol{\beta}} = (X^T V^{-1} X)^{-1} X^T V^{-1} \boldsymbol{y}$ is the BLUE estimator, when $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad Cov(\boldsymbol{\varepsilon}) = \sigma^2 V$, and there is no model mismatch.
- Yet another version: $\hat{\boldsymbol{\beta}} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \boldsymbol{y}$ is the BLUE estimator, when $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad Cov(\boldsymbol{\varepsilon}) = \Sigma$, and there is no model mismatch.
- Proof can be found in our textbook, see appendix C11.

Special Cases

Example (Two spacial cases:)

1. $V = I$, which boils down to "ordinary" LS.
2. $V = \text{diag}(\frac{1}{w_1}, \frac{1}{w_2}, \ldots, \frac{1}{w_n})$, with $w_i > 0$, $\forall i = 1, 2, \ldots, n$
(uncorrelated random error terms with non-constant variance)

Note that the second case is also called "weighted LS " (WLS) in the textbook.

For this case, we could simply let $W = V^{-1} = \text{diag}(w_1, w_2, \ldots, w_n)$, then

$$\boxed{\hat{\beta} = (X^T W X)^{-1} X^T W \mathbf{y}}$$