Gauss-Markov Theorem (Matrix Form)

Contents

Statement of Gauss-Markov Theorem

Theorem 1 (Gauss-Markov theorem)
Suppose there is a linear model $Y = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$*, where* X *is an* $n \times d$ *design matrix,* $\underset{\sim}{\beta}$ *is a* $d \times 1$ *matrix of weights (i.e. coefficients), and* $\underset{\sim}{\epsilon}$ *is a* $n \times 1$ *matrix of errors* ϵ_i*'s,* $i \in [n]$*.*

Under the Gauss-Markov Assumptions, i.e.:

1. ϵ_i*'s have mean zero,*

   i.e. $\mathbb{E}[\epsilon_i] = 0, \forall i \in [n]$*.*

2. ϵ_i*'s are homoscedastic,*

   i.e. $Var[\epsilon_i] = \sigma^2 < \infty, \forall i \in [n]$*.*

3. *Distinct error terms are uncorrelated,*

   i.e. $Cov[\epsilon_i, \epsilon_j] = 0, \forall i \neq j \ \& \ i, j \in [n]$*.*

or equavalently to the above assumptions, $\mathbb{E}[\underset{\sim}{\epsilon}] = \underset{\sim}{0}$*,* $Var[\underset{\sim}{\epsilon}] = \sigma^2 I$*, where* $\sigma^2 < \infty$ *and* I *is an* $n \times n$ *identity matrix, we have that the ordinary least square (OLS) estimator of the parameter* $\underset{\sim}{\beta}$*:*

$$\underset{\sim}{\hat{\beta}}_{OLS} = (X^T X)^{-1} X^T Y$$

is the best linear unbiased estimator (BLUE), i.e. it is the most efficient estimator (i.e. has lowest variance) among all linear unbiased estimators.

Proof of Gauss-Markov Theorem

Proof. The OLS estimator:

$$\hat{\underset{\sim}{\beta}}_{OLS} = (X^TX)^{-1}X^TY \tag{1}$$
$$= (X^TX)^{-1}X^T(X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}) \tag{2}$$
$$= (X^TX)^{-1}X^TX\underset{\sim}{\beta} + (X^TX)^{-1}X^T\underset{\sim}{\epsilon} \tag{3}$$
$$= \underset{\sim}{\beta} + (X^TX)^{-1}X^T\underset{\sim}{\epsilon} \tag{4}$$

Thus, the unbiasedness of $\hat{\underset{\sim}{\beta}}_{OLS}$ is given by:

$$\mathbb{E}[\hat{\underset{\sim}{\beta}}_{OLS}] = \mathbb{E}[\underset{\sim}{\beta} + (X^TX)^{-1}X^T\underset{\sim}{\epsilon}] \tag{5}$$
$$= \underset{\sim}{\beta} + (X^TX)^{-1}X^T\mathbb{E}[\underset{\sim}{\epsilon}] \tag{6}$$
$$= \underset{\sim}{\beta}, \text{ since by assumption 1, } \mathbb{E}[\underset{\sim}{\epsilon}] = \underset{\sim}{0}. \tag{7}$$

Also by assumption 1,

$$Var[\underset{\sim}{\epsilon}] = \mathbb{E}[(\underset{\sim}{\epsilon} - \mathbb{E}[\underset{\sim}{\epsilon}])(\underset{\sim}{\epsilon} - \mathbb{E}[\underset{\sim}{\epsilon}])^T] = \mathbb{E}[\underset{\sim}{\epsilon}\underset{\sim}{\epsilon}^T] = \sigma^2 I \tag{8}$$

Thus,

$$Var[Y] = Var[X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}] = Var[\underset{\sim}{\epsilon}] = \sigma^2 I \tag{9}$$

The variance of $\hat{\underset{\sim}{\beta}}_{OLS}$:

$$Var[\hat{\underset{\sim}{\beta}}_{OLS}] = \mathbb{E}[(\hat{\underset{\sim}{\beta}}_{OLS} - \underset{\sim}{\beta})(\hat{\underset{\sim}{\beta}}_{OLS} - \underset{\sim}{\beta})^T] \tag{10}$$
$$= \mathbb{E}[((X^TX)^{-1}X^T\underset{\sim}{\epsilon})((X^TX)^{-1}X^T\underset{\sim}{\epsilon})^T], \text{ by (4)} \tag{11}$$
$$= \mathbb{E}[(X^TX)^{-1}X^T\underset{\sim}{\epsilon}\underset{\sim}{\epsilon}^TX(X^TX)^{-1}] \tag{12}$$
$$= (X^TX)^{-1}X^T\mathbb{E}[\underset{\sim}{\epsilon}\underset{\sim}{\epsilon}^T]X(X^TX)^{-1} \tag{13}$$
$$= (X^TX)^{-1}X^T\sigma^2 IX(X^TX)^{-1}, \text{ by (8)} \tag{14}$$
$$= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} \tag{15}$$
$$= \sigma^2(X^TX)^{-1} \tag{16}$$

For any unbiasedness linear estimator $\hat{\underset{\sim}{\beta}}$ for $\underset{\sim}{\beta}$, it must can be expressed in the linear form by the definition of linear estimators:

$$\hat{\underset{\sim}{\beta}} = CY$$

for some constant matrix C where C is independent to $\underset{\sim}{\beta}$ which is unobservable.

Let D be a matrix so that

$$D = C - (X^TX)^{-1}X^T \iff C = D + (X^TX)^{-1}X^T. \tag{17}$$

Additionally, by the unbiasedness of $\hat{\underset{\sim}{\beta}}$:

$$\mathbb{E}[\hat{\underset{\sim}{\beta}}] = \underset{\sim}{\beta} \tag{18}$$

Thus,

$$\begin{aligned}
\mathbb{E}[\hat{\underset{\sim}{\beta}}] &= \mathbb{E}[CY] && (19)\\
&= \mathbb{E}[(D + (X^T X)^{-1} X^T)(X\underset{\sim}{\beta} + \underset{\sim}{\epsilon})] && (20)\\
&= \mathbb{E}[DX\underset{\sim}{\beta} + D\underset{\sim}{\epsilon} + (X^T X)^{-1} X^T X\underset{\sim}{\beta} + (X^T X)^{-1} X^T \underset{\sim}{\epsilon}] && (21)\\
&= DX\underset{\sim}{\beta} + D\mathbb{E}[\underset{\sim}{\epsilon}] + \underset{\sim}{\beta} + (X^T X)^{-1} X^T \mathbb{E}[\underset{\sim}{\epsilon}] && (22)\\
&= DX\underset{\sim}{\beta} + \underset{\sim}{\beta}, \text{ by assumption 1} && (23)\\
&= \underset{\sim}{\beta}, \text{ by (18)} && (24)\\
&\implies DX\underset{\sim}{\beta} = \underset{\sim}{0} && (25)\\
&\implies DX = \underset{\sim}{0} && (26)
\end{aligned}$$

Finally, the variance of any unbiased linear estimator $\hat{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$:

$$\begin{aligned}
Var[\hat{\underset{\sim}{\beta}}] &= Var[CY] && (27)\\
&= CVar[Y]C^T && (28)\\
&= C\sigma^2 I C^T, \text{ by (9)} && (29)\\
&= \sigma^2 (D + (X^T X)^{-1} X^T)(D + (X^T X)^{-1} X^T)^T, \text{ by (17)} && (30)\\
&= \sigma^2 [DD^T + DX(X^T X)^{-1} && (31)\\
&+ (X^T X)^{-1}(DX)^T + (X^T X)^{-1} X^T X (X^T X)^{-1}] && (32)\\
&= \sigma^2 DD^T + \sigma^2 (X^T X)^{-1}, \text{ by (26)} && (33)\\
&= Var[\hat{\underset{\sim}{\beta}}_{OLS}] + \sigma^2 DD^T, \text{ by (16)} && (34)
\end{aligned}$$

Notice that $\sigma^2 \geq 0$ and DD^T is always definite symmetric.

Therefore, $\hat{\underset{\sim}{\beta}}_{OLS}$ has the lowest variance among all linear unbiased estimators of $\underset{\sim}{\beta}$ so that it is the best linear unbiased estimator (BLUE).

■