

# GAUSS-MARKOV THEOREM

(4-1)

WITHIN A CLASS OF LINEAR UNBIASED ESTIMATORS THE LEAST SQUARES ESTIMATORS HAVE THE MINIMUM VARIANCE.

THIS THEOREM ESTABLISHES THAT

- (A) OLS ESTIMATORS ARE LINEAR
- (B) OLS ESTIMATORS ARE UNBIASED
- (C) THERE IS NO OTHER LINEAR UNBIASED ESTIMATOR THAT HAS LOWER VARIANCE THAN THE OLS ESTIMATORS.

## UNBIASEDNESS

CONSIDER SOME POPULATION PARAMETER  $\theta$  & SOME SAMPLE STATISTIC  $\hat{\theta}$  WHICH IS AN ESTIMATOR OF  $\theta$ . NOTE THAT  $\theta$  IS A CONSTANT BUT  $\hat{\theta}$  TAKES DIFFERENT VALUES IN DIFFERENT SAMPLES. THE DISTRIBUTION OF  $\hat{\theta}$  ACROSS ALL POSSIBLE SAMPLES OF A GIVEN SIZE IS CALLED THE SAMPLING DISTRIBUTION OF THE ESTIMATOR  $\hat{\theta}$ . IT IS AN UNBIASED ESTIMATOR IF THE SAMPLING DISTRIBUTION HAS A MEAN EQUAL TO THE PARAMETER  $\theta$ .

EXAMPLE CONSIDER THE POPULATION MEAN ( $\mu$ ) AND THE SAMPLE MEAN:  $\bar{x} = \frac{1}{n} \sum x_i$

$$\text{THEN } E(\bar{x}) = E\left[\frac{1}{n} \sum x_i\right] = \frac{1}{n} \sum E(x_i)$$

$$\text{BUT } E(x_i) = \mu. \text{ HENCE } E(\bar{x}) = \mu$$

THE SAMPLE MEAN IS AN UNBIASED ESTIMATOR OF THE POPULATION MEAN.

(SAMPLING) VARIANCE OF AN ESTIMATOR IS (4.2)

THE VARIANCE OF THE SAMPLING DISTRIBUTION.

EXAMPLE: 
$$\text{VAR}(\bar{X}) = \text{VAR}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)$$
$$= \sum \text{VAR}\left(\frac{X_i}{n}\right) = \sum \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

WHY IS A SMALL VARIANCE DESIRABLE?

CHEBYCHEV'S INEQUALITY SHOWS THAT FOR ANY RANDOM VARIABLE  $X$  WITH MEAN  $\mu$  & VARIANCE  $\sigma^2$

$$\text{Pr}[\mu - k\sigma \leq X \leq \mu + k\sigma] \geq 1 - \frac{1}{k^2}$$

THUS A SMALL VALUE OF  $\sigma^2$  (AND HENCE  $\sigma$ ) IMPLIES THAT A RANDOMLY DRAWN VALUE OF  $X$  LIES WITHIN A SMALLER INTERVAL  $k\sigma$  EXCEEDS

$1 - \frac{1}{k^2}$ . FOR A KNOWN DISTRIBUTION, WE CAN BE MORE PRECISE.

IF  $X \sim N(\mu, \sigma^2)$ ,

$$\text{Pr}[\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma] = 0.95$$

SUPPOSE  $\mu = 10$  &  $\sigma^2 = 4$

THEN  $\text{Pr}[6.08 \leq X \leq 13.92] = 0.95$

BUT IS  $\sigma^2 = 1$

$$\text{Pr}[8.04 \leq X \leq 11.96] = 0.95$$

THUS A SMALLER VARIANCE IMPLIES THAT THE DISTRIBUTION IS MORE CONCENTRATED AROUND THE MEAN.

$$\hat{\beta} = (X'X)^{-1} X'y$$

DEFINE

$$A \equiv (X'X)^{-1} X'$$

THEN

$$\hat{\beta} \equiv Ay$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & & a_{kn} \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

THEN

$$\hat{\beta} = Ay \Rightarrow \hat{\beta}_i = \sum_j a_{ij} y_j \quad (i=1, 2, \dots, k)$$

THIS SHOWS THAT  $\hat{\beta}$  IS LINEAR IN  $y$ .

$$y = X\beta + u \Rightarrow \hat{\beta} = (X'X)^{-1} X'y$$

$$\hat{\beta} = (X'X)^{-1} X'(X\beta + u)$$

$$\hat{\beta} = (X'X)^{-1} X'X\beta + (X'X)^{-1} X'u = \beta + (X'X)^{-1} X'u$$

$$E(\hat{\beta}) = \beta + E((X'X)^{-1} X'u) = \beta + (X'X)^{-1} X'E(u)$$

THIS SHOWS UNBIASEDNESS.  $(\text{b/c } E(u) = 0)$

$$V(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

$$= E[(X'X)^{-1} X'u u' X (X'X)^{-1}]$$

$$= (X'X)^{-1} X'E(uu') X (X'X)^{-1}$$

$$= (X'X)^{-1} X'(\sigma^2 I) X (X'X)^{-1} = \sigma^2 (X'X)^{-1} X'X (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}$$

CONSIDER SOME OTHER LINEAR & UNBIASED ESTIMATOR  $\tilde{\beta}$ .

BECAUSE  $\hat{\beta}$  IS LINEAR IN  $y$  IT CAN BE WRITTEN AS  $\hat{\beta} = Cy$

(4-4)

THEN  $\hat{\beta} = C(x\beta + u) = Cx\beta + Cu$

$$E(\hat{\beta}) = Cx\beta + CE(u) = Cx\beta.$$

BUT  $\hat{\beta}$  IS UNBIASED. HENCE

$$E(\hat{\beta}) = \beta \Rightarrow Cx = I.$$

RECALL THAT  $\hat{\beta} = (x'x)^{-1}x'y \equiv Ay$

DEFINE  $D = C - A.$

NOW,  $\hat{\beta} = \beta + Cu$

$$\Rightarrow \hat{\beta} - \beta = Cu$$

$$V(\hat{\beta}) = E[Cuu'C'] = C(\sigma^2 I)C' = \sigma^2 CC'.$$

BUT  $C = A + D$

$$= D + (x'x)^{-1}x'$$

$$C' = D' + x(x'x)^{-1}$$

$$CC' = [D + (x'x)^{-1}x'] [D' + x(x'x)^{-1}]$$

$$= DD' + (x'x)^{-1}x'D' + Dx(x'x)^{-1} + (x'x)^{-1}x'x(x'x)^{-1}$$

NOW,  $D = C - (x'x)^{-1}x'$

$$Dx = Cx - (x'x)^{-1}x'x = I - I = 0$$

$$\Rightarrow x'D' = 0'$$

HENCE  $CC' = DD' + (x'x)^{-1}.$

$$V(\tilde{\beta}) - V(\hat{\beta}) \\ = \sigma^2 D D'$$

IS A POSITIVE DEFINITE  
MATRIX.

$$V(\tilde{\beta}) = V(\hat{\beta}) + \sigma^2 D D'$$