Chapter 3 Sampling Properties of LS Estimator

남준우·허인(2021), 제4장

Gujarati/Porter(2018), 제 3장

(1) Assumptions

(2) Discussions

(3) Mean and Variance of LS Estimators

(4) Gauss-Markov Theorem

(5) Coefficient of Determination

(1) Assumptions

① (Linear model) $y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$

② (Nonstochastic Independent variables) X is nonstochastic.

③ (Identification condition) Support of X is rich.

④ (Zero mean of error term) $E(\varepsilon_i) = 0$.

⑤ (Equal variance of error term) $V(\varepsilon_i) = \sigma^2$ for all i.

⑥ (No autocorrelation) $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.

(2) Discussions

① (Linear model) $y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$

► CEF is linear.

② (<u>Nonstochastic Independent variables</u>) X is nonstochastic.

► Stratified sampling(층화표본)

► NOT a strong assumption. Can be relaxed.

► In general, actually, we require $Cov(X_i, \varepsilon_i) = 0$ for unbiasedness of LS estimator.

Assumption ② automatically fulfills this condition.

③ (<u>Identification Condition</u>) Support of X is rich.

► X has at least two different values.

► $S_X{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \neq 0$

► Excludes the case of $X_i = c$ for all i.

④ (<u>Zero mean of error term</u>) $E(\varepsilon_i) = 0$.

► With ②, $E(y_i) = \beta_1 + \beta_2 X_i$.

► What if $E(\varepsilon_i) \neq 0$?

⑤ (<u>Equal variance of error term</u>) $V(\varepsilon_i) = \sigma^2$.

• Homoscedasticity assumption.

► With ass. ②, $V(y_i) = \sigma^2$ for all i.

► Will be relaxed.

⑥ (No autocorrelation) $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.

- No systematic change across error terms.

► No serial correlation.

► With ass. ②, $Cov(y_i, y_j) = 0$ for all $i \neq j$.

► Too strong assumption for time series data. Will be relaxed.

(2) Mean and Variance of LS Estimator

► We only focus on slope coefficient estimator b_2, since the same manipulation can be easily applied to b_1.

① Note that (b_1, b_2) are random variables.

② b_2 is linear in y_i.

③ $E(b_2) = \beta_2$; unbiasedness.

- Is $b_2 = \beta_2$?

► In repeated sampling, center of distribution is β_2.

④ $V(b_2) = \dfrac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$, $V(b_1) = \sigma^2\left[\dfrac{1}{n} + \dfrac{\bar{X}^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right]$.

► The meaning of variance? Precision.

► As $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \uparrow$, $V(b_2)\downarrow$.

► As $n\uparrow$, $V(b_2)\downarrow$.

⑤ $Cov(b_1, b_2) = \dfrac{-\bar{X}\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$.

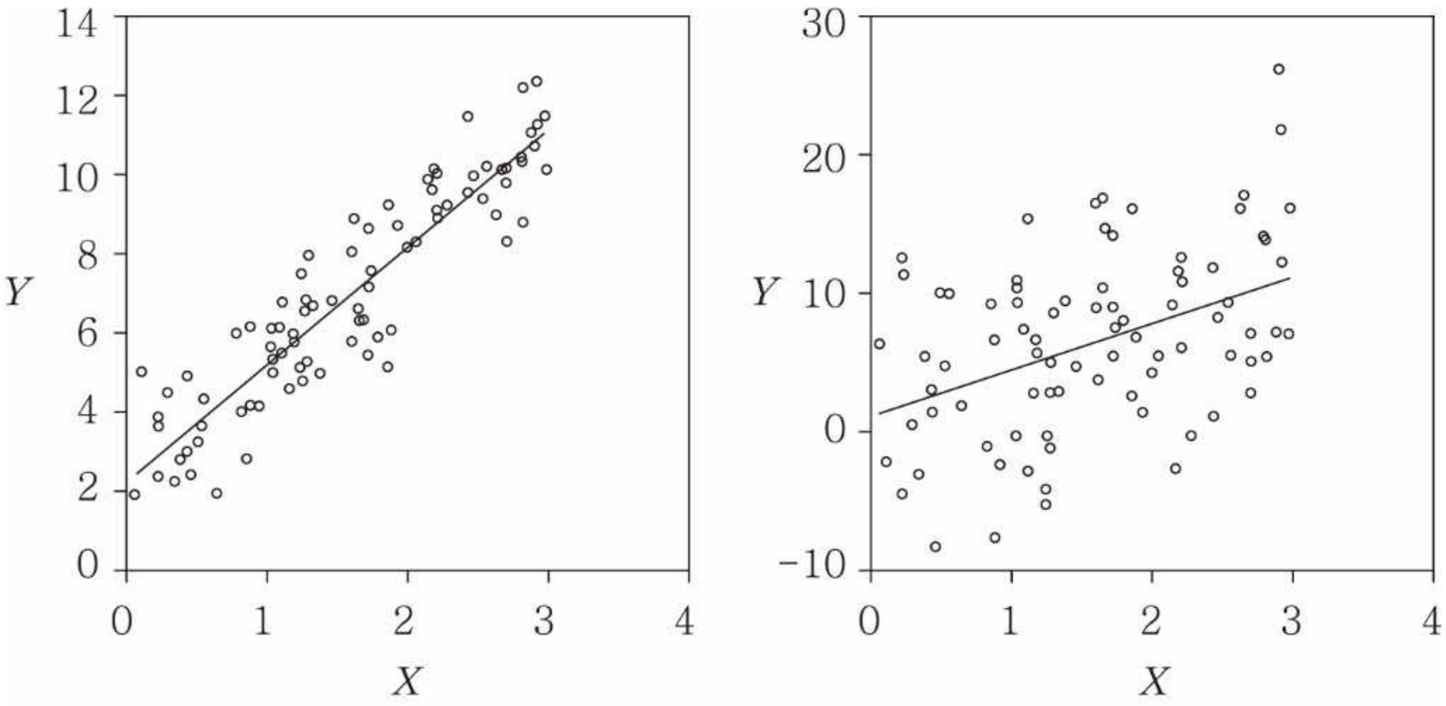
(3) <u>Gauss-Markov Theorem</u>

LS estimator is BLUE(MVLUE).

(4) The Coefficient of Determination

• How explain the variation in y_i with variation in X_i in estimated regression.

► How good is the fit?



(a) 결정계수 값이 큰 경우

(b) 결정계수 값이 작은 경우

- $y_i = \hat{y}_i + e_i$

$$\overline{y} = \overline{\hat{y}}$$

$$\Rightarrow \ (y_i - \overline{y}) = (\hat{y}_i - \overline{\hat{y}}) + e_i$$

$$\Rightarrow \ \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})^2 + \sum_{i=1}^{n} e_i{}^2, \qquad \text{why?}$$

$$TSS \quad = \quad ESS \quad + RSS$$

- $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}$

$$\Rightarrow \quad R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^{\,2}}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

; proportion of variation in y explained by X.

- $0 \le R^2 \le 1$.

► What makes $0 \le R^2 \le 1$?

► $R^2 = 1$ in which case?

(Example) In consumption example,

(Example)

- y:소비, X: 소득.

obs	Y_i	X_i	$(X_i - \bar{X})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	e_i	e_i^2	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})X_i$	$(X_i - \bar{X})Y_i$
1	70	80	8100	3690	65.28	4.82	23.21	1681		
2	65	100	4900	3220	75.36	-10.36	107.41	2116		
3	90	120	2500	1050	85.55	4.45	19.84	441		
4	95	140	900	480	95.73	-0.73	0.53	256		
5	110	160	100	10	105.91	4.09	16.74	1		
6	115	180	100	40	116.09	-1.09	1.19	16		
7	120	200	900	270	126.27	-6.27	39.35	81		
8	155	240	4900	3080	146.64	8.36	69.95	1936		
9	150	260	8100	3510	156.82	-6.82	46.49	1521		
10	140	220	2500	1450	136.45	3.55	12.57	841		
합	1110	1700	33000	16800		0	337.27	8890		

$$R^2 = 1 - \frac{337.27}{8890} = 0.96.$$