

제 5 강

계량경제학
5.1

단순 회귀모형에서의 추론 (Inference in the Simple Regression Model)

단순 선형회귀 모형 가정

계량경제학
5.2

1. $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$
2. $E(\varepsilon_t) = 0 \iff E(y_t) = \beta_1 + \beta_2 x_t$
3. $\text{var}(\varepsilon_t) = \sigma^2 = \text{var}(y_t)$
4. $\text{cov}(\varepsilon_i, \varepsilon_j) = \text{cov}(y_i, y_j) = 0$
5. x_t is not random and $x_{t \neq s}$ for at least one pair of (t,s).
6. $\varepsilon_t \sim N(0, \sigma^2) \iff y_t \sim N(\beta_1 + \beta_2 x_t, \sigma^2)$

최소제곱추정량의 성질

계량경제학
5.3

최소제곱추정량

1. 불편추정량 $E(b_1) = \beta_1$ $E(b_2) = \beta_2$
유효성(efficiency) ▶
2. 최소분산 (가우스-마코프 정리)

$$\text{Var}(b_1) = \frac{\sigma^2 \sum x_t^2}{T \sum (x_t - \bar{x})^2} \quad \text{Var}(b_2) = \frac{\sigma^2}{\sum (x_t - \bar{x})^2}$$

3. 일치성: T가 커짐에 따라 추정량이 더 정확해짐
4. $b_1 \sim N[\beta_1, \text{Var}(b_1)]$ $b_2 \sim N[\beta_2, \text{Var}(b_2)]$

최소제곱추정량의 성질

계량경제학
5.4

최소제곱잔차

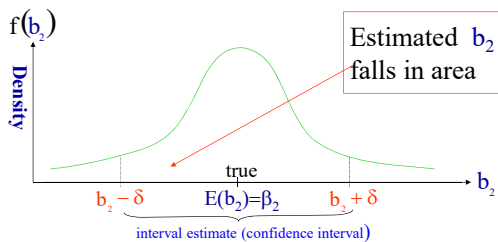
1. $\sum \hat{\varepsilon}_t = 0$
2. $\sum \hat{\varepsilon}_t x_t = 0$
3. $\sum \hat{\varepsilon}_t \hat{y}_t = 0$ ▶
4. 회귀선은 반드시 x 와 y의 표본평균 (\bar{x}, \bar{y}) 를 통과함.
5. $\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_t^2}{T-2}$: σ^2 의 불편추정량

최소제곱추정량은 얼마나 신뢰할만한가?

계량경제학
5.5

구간추정량(대칭적 분포의 경우)

- b_1 이 β_1 에 얼마나 "가까운가"?
 b_2 가 β_2 에 얼마나 "가까운가"?



최소제곱추정량은 얼마나 신뢰할만한가?

계량경제학
5.6

구간추정량(대칭적 분포의 경우)

$$\Pr(b_2 - \delta < \beta_2 < b_2 + \delta) = (1 - \alpha) \begin{cases} 0.99 \\ 0.95 \\ 0.90 \end{cases}$$

α : 유의수준(level of significance)이라 함. $\begin{cases} 0.01 \\ 0.05 \\ 0.10 \end{cases}$

$b_2 \pm \delta$ 는 β_2 에 대한 $(1 - \alpha) \times 100\%$ 구간추정(치) 또는 $(1 - \alpha) \times 100\%$ 신뢰구간이라 함

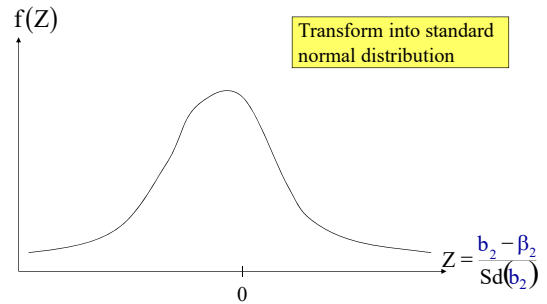
b_2 가 확률변수라면, $b_2 \pm \delta$ 는 구간추정량이라고 함

정규분포의 가정을 이용하여 σ^2 가 알려져 있다는 전제하에 b_2 에 대한 다음과 같은 확률적 명제가 제시될 수 있음

$$Z = \frac{b_2 - \beta_2}{Sd(b_2)} \sim N(0,1)$$

$$= (b_2 - \beta_2) \frac{\sqrt{\sum(x_t - \bar{x})^2}}{\sigma}$$

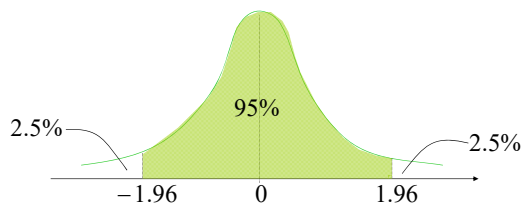
실제에 있어서 관측되지 않음



For example:

$$\Pr(Z > 1.96) = \Pr(Z < -1.96) = 0.025$$

From the table



$$\Pr(-1.96 < Z < 1.96) = 0.95$$

$$\Rightarrow \Pr\left[-1.96 < \frac{b_2 - \beta_2}{Sd(b_2)} < 1.96\right] = 0.95$$

95% interval estimator:

$$-1.96 < \frac{b_2 - \beta_2}{Sd(b_2)} < 1.96$$

$$\Rightarrow b_2 - 1.96 \times Sd(b_2) < \beta_2 < b_2 + 1.96 \times Sd(b_2)$$

$$\Rightarrow b_2 \pm 1.96 \times Sd(b_2)$$

실제로는 σ^2 가 알려지지 않으며, 따라서 그에 대한 불편추정량을 사용해야 함

$$\hat{\sigma}^2 = \frac{\sum \hat{\epsilon}_t^2}{T-2} \Rightarrow \hat{v}ar(b_2) = \frac{\hat{\sigma}^2}{\sum(x_t - \bar{x})^2}$$

$$Se(b_2) = \sqrt{\hat{v}ar(b_2)}$$

이 경우 정규분포 대신 t-분포를 사용함

$$t = \frac{(b_2 - \beta_2)}{Se(b_2)} \sim t_{(T-2)}$$

이 t는 자유도가 T-2인 t-분포(Student t-distribution)를 함

$$t = \frac{b_2 - \beta_2}{Se(b_2)}$$

$$t = \frac{\text{추정량} - \text{모수}}{\text{추정량의 표준오차}}$$

$$t = \frac{(b_2 - \beta_2) \sqrt{\sum(x_t - \bar{x})^2}}{\hat{\sigma}}$$

이 t를 이용하여 β_2 에 대한 신뢰구간을 구축할 수 있음

계량경제학 5.13

구간추정량의 도출-계속

β_2 에 대한 구간추정량을 t분포로부터의 임계값(critical value) t^c 를 이용하여 다음과 같이 구축할 수 있음

$$\Pr \left(-t_{\frac{\alpha}{2}, T-2}^c \leq t \leq t_{\frac{\alpha}{2}, T-2}^c \right) = 1 - \alpha$$

α : 유의수준 (level of significance) From the table

(T-2) : 자유도 (degree of freedom)
(2개의 모수를 추정할 경우).

계량경제학 5.14

구간추정량의 도출-계속

따라서

$$\Pr \left(-t_{0.05, T-2}^c \leq \frac{b_2 - \beta_2}{\text{Se}(b_2)} \leq t_{0.05, T-2}^c \right) = 0.90$$

$\Pr(-t_{0.025, T-2}^c \leq (b_2 - \beta_2)/\text{Se}(b_2) \leq t_{0.025, T-2}^c) = 0.95$

다시 정리하면,

$$\Pr(b_2 - t_{0.05, T-2}^c \times \text{Se}(b_2) \leq \beta_2 \leq b_2 + t_{0.05, T-2}^c \times \text{Se}(b_2)) = 0.90$$

계량경제학 5.15

구간추정량의 도출-계속

$b_2 \pm t_{0.05, T-2}^c \times \text{Se}(b_2)$: β_2 에 대한 90% 구간추정량

$t_{0.05, T-2}^c$ Check it from t-table b_2 & $\text{Se}(b_2)$ Check it from estimated result

β_2 에 대한 95% 구간추정량은

$$b_2 \pm t_{0.025, T-2}^c \times \text{Se}(b_2)$$

계량경제학 5.16

구간추정량의 도출-계속

실례: 5.1.4

식료품 지출에 대한 자료 (관측치의 수: 40)를 가지고 β_2 에 대한 95% 신뢰구간을 구축

임계값 $t_c = 2.024$ (from the table)

최소제곱추정치: $b_2 = .1283$
그리고 그 표준오차: $\text{se}(b_2) = 0.0305$

$b_2 \pm t_c \text{se}(b_2) = .1283 \pm 2.024(.0305) = [.0666, .1900]$ δ

계량경제학 5.17

가설검정 (Hypothesis Testing)

많은 경제적 결정에 있어서, 어떤 모수가 특정한 값 또는 부호를 갖는지 여부를 알고자 하는 경우가 있음

$\beta_2 \neq 0?$, $\beta_2 = 0$, $\beta_2 > 0$, $\beta_2 < 0$... : 추측들

추정된 b_2 로부터, 이 추측들을 검정(test)할 수 있음 \Rightarrow 가설검정(Hypothesis Testing)

계량경제학 5.18

가설검정 (Hypothesis Testing)

요소

1. 귀무가설(null hypothesis) H_0 .
2. 대립가설(alternative hypothesis) H_1 .
3. 검정통계량(test statistic).
4. 기각역(rejection region).

구성

1. 귀무가설과 대립가설을 결정함.
2. 검정통계량을 정하고 귀무가설이 참일 경우의 그 분포를 구체화함
3. 유의수준 α 를 정하고 그에 대응되는 기각역을 결정
4. 검정통계량과 관측치로부터 통계치를 계산함
5. 결론을 기술함

가설검정 (Hypothesis Testing) 계량경제학 5.19

양측검정

- $H_0: \beta_2 = c$
 $H_1: \beta_2 \neq c$

가설을 기술
- $t = \frac{b_2 - c}{Se(b_2)}$

검정통계량 결정 및 통계치 계산
- t 분포표로부터 임계값 계산: $t_{\frac{\alpha}{2}, T-2}^c$

가설검정 (Hypothesis Testing) 계량경제학 5.20

양측검정

- t 와 t^c 를 비교함
- 결정원칙:** 만약 $t \geq t^c$ 이거나 $t \leq -t^c$ 이면 (즉 $|t| \geq t^c$), H_0 을 기각함

가설검정 (Hypothesis Testing) 계량경제학 5.21

단측검정

- $H_0: \beta_2 = c$ ($H_0: \beta_2 = c$)
 $H_1: \beta_2 > c$ ($H_1: \beta_2 < c$)

가설을 기술함
- $t = \frac{b_2 - c}{Se(b_2)}$

통계치를 계산
- t 분포표에서 $t_{\alpha, T-2}^c$ 를 찾아 임계값을 정함
- t^c 와 t 를 비교함

가설검정 (Hypothesis Testing) 계량경제학 5.22

단측검정

Step 5: If $t \geq t^c \implies H_0$ 를 기각함 우측꼬리 Right-tail
 If $t < t^c \implies H_0$ 를 기각 않음

(If $t \leq -t^c \implies H_0$ 를 기각함) 좌측꼬리 Left-tail
 (If $t > -t^c \implies H_0$ 를 기각 않음)

가설검정 (Hypothesis Testing) 계량경제학 5.23

제 1 종 및 제 2 종 오류

- 제 1 종 오류(Type I error):**
 - 귀무가설이 옳음에도 그것을 기각하는 오류를 범함
 - $\alpha = P(\text{rejecting } H_0 \text{ when it is true}).$
- 제 2 종 오류(Type II error):**
 - 귀무가설이 그름에도 그것을 기각하지 못하는 오류를 범함
 - $\beta = P(\text{failing to reject } H_0 \text{ when it is false}).$

가설검정 (Hypothesis Testing) 계량경제학 5.24

제 2 종 오류

- 제 2 종 오류의 확률은 검정의 유의수준(제 1 종 오류의 확률) α 의 크기와 역의(inversely)관계에 있음
- 대립가설에서 가설로 제시된 모수의 값이 모수의 참값에 가까울 수록 제 2 종 오류의 확률은 커짐
- 표본의 크기 T 가 클 수록, 제 1 종 오류 α 가 주어졌을 때 제 2 종 오류의 확률은 작아짐
- 앞서 소개한 t 분포에 기반한 검정은 매우 좋은 검정방법으로 알려져 있음 (즉 주어진 유의수준에 대해 낮은 제 2 종 오류의 확률을 가짐):
 - 검정력(Power)= $1-\beta \implies$ 검정력이 높은 검정(powerful test)

가설검정 (Hypothesis Testing) 계량경제학 5.25

P-값 (p-value)

P-값 = $2 \times \Pr(t \geq |t| \text{ 검정통계량의 표본값})$
: 양측 t검정

P-값 = $\Pr(t \geq t \text{ 검정통계량의 표본값})$
: 우측꼬리 t검정

P-값 = $\Pr(t \leq t \text{ 검정통계량의 표본값})$
: 좌측꼬리 t검정

가설검정 (Hypothesis Testing) 계량경제학 5.26

유의성 검정(test of significance)

$H_0: \beta_2 = 0$
 $H_1: \beta_2 \neq 0$

이 귀무가설의 기각은 x와 y사이에 “통계적으로 유의한(Statistically Significant)” 관계가 존재함을 의미

$t^* = \frac{b_2}{\text{Se}(b_2)}$: 통계패키지의 회귀분석 결과 보고서 자동으로 계산하여 보고됨

최소제곱예측 (Least Squares Predictor) 계량경제학 5.27

x_0 가 주어졌을 때, $y_0 = \beta_1 + \beta_2 x_0 + \varepsilon_0$ 에 대한 최소제곱예측(값):

$$\hat{y}_0 = b_1 + b_2 x_0$$

예측오차: $f \equiv \hat{y}_0 - y_0$

$E(f) = E(b_1 - \beta_1 + (b_2 - \beta_2)x_0 - \varepsilon_0) = 0$

$$\text{var}(f) = \sigma^2 \left[1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum(x_t - \bar{x})^2} \right]$$

최소제곱예측 (Least Squares Predictor) 계량경제학 5.28

$$\hat{\text{var}}(f) = \hat{\sigma}^2 \left[1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum(x_t - \bar{x})^2} \right]$$

$$\text{se}(f) = \sqrt{\hat{\text{var}}(f)}$$

x_0 가 주어졌을 때, y_0 에 대한 $(1-\alpha) \times 100\%$ 예측구간(prediction interval):

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}, T-2}^c \text{se}(f)$$

유효성 (Efficiency) 계량경제학 5.29

추정량이 유효하다는 것은 추정량의 값이 추정하고자 하는 모수에 가까울 것으로 기대할 수 있음을 의미함

일반적으로 α 의 추정량 a의 유효성은 다음의 척도(Mean Squared Error: 평균제곱오차)로서 비교된다

$$E[(a - \alpha)^2] = E[(a - E(a) + E(a) - \alpha)^2]$$

$$= E[(a - E(a))^2] + E[(E(a) - \alpha)^2] + 2E[(E(a) - \alpha)(a - E(a))]$$

$$= \text{Var}(a) + (E(a) - \alpha)^2 + 2(E(a) - \alpha)E[a - E(a)]$$

$$= \text{Var}(a) + (E(a) - \alpha)^2$$

즉 분산 + 편차의 제곱의 합

잔차의 성질 계량경제학 5.30

정규방정식

$$\frac{\partial S}{\partial \beta_1} = 0 \Rightarrow \sum y_i - T b_1 - b_2 \sum x_i = 0$$

$$\frac{\partial S}{\partial \beta_2} = 0 \Rightarrow \sum y_i x_i - b_1 \sum x_i - b_2 \sum x_i^2 = 0$$

- $\sum \hat{\varepsilon}_i = \sum (y_i - b_1 - b_2 x_i) = \sum y_i - T b_1 - b_2 \sum x_i = 0$
- $\sum \hat{\varepsilon}_i x_i = \sum (y_i - b_1 - b_2 x_i) x_i = \sum (y_i x_i - b_1 x_i - b_2 x_i^2) = \sum y_i x_i - b_1 \sum x_i - b_2 \sum x_i^2 = 0$
- $\sum \hat{\varepsilon}_i \hat{y}_i = \sum \hat{\varepsilon}_i (b_1 + b_2 x_i) = b_1 \sum \hat{\varepsilon}_i + b_2 \sum \hat{\varepsilon}_i x_i = 0$

t 분포의 정의

$$V \equiv Z_1^2 + \dots + Z_m^2 \sim \chi^2(m)$$

Where Z_1, \dots, Z_m are m independent $N(0,1)$

$$t \equiv \frac{Z}{\sqrt{V/m}} \sim t_{(m)} \quad \text{where } Z \sim N(0,1) \text{ and } V \sim \chi_{(m)}^2$$

and Z and V are statistically independent

t 분포의 성질

1. t-분포나 정규분포 모두 대칭인 종모양(bell-shaped)의 분포를 함
2. t-분포의 경우 정규분포에 비해 꼬리부분이 두꺼움
3. 표본의 크기가 커질 경우 t분포는 정규분포로 수렴함
4. T분포는 자유도에 의해 그 분포가 결정됨.
5. 자유도가 30이상인 경우 정규분포는 t분포에 대해 소수점 이하 몇 자리까지의 정확성을 지니는 훌륭한 근사임

We can show the following;

$$V \equiv \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{\sigma^2} = \frac{(T-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{T-2}^2$$

Intuition?

. $\varepsilon_t \sim N(0, \sigma^2) \rightarrow (\varepsilon_t / \sigma) \sim N(0, 1)$

. Therefore, $\sum_{t=1}^T (\varepsilon_t / \sigma)^2 \sim \chi^2_{(T)}$

. ε_t is unobservable. Then $\sum_{t=1}^T (\hat{\varepsilon}_t / \sigma)^2 \sim ?$

We can show that V is statistically independent of the least squares estimators b_1, b_2

$$\text{Let } Z = \frac{(b_2 - \beta_2)}{\sqrt{\text{var}(b_2)}} \quad V = \frac{(T-2)\hat{\sigma}^2}{\sigma^2}$$

$$\text{Let } t = \frac{Z}{\sqrt{V/(T-2)}}$$

$$t = \frac{\frac{(b_2 - \beta_2)}{\sqrt{\text{var}(b_2)}}}{\sqrt{\frac{(T-2)\hat{\sigma}^2}{\sigma^2} / (T-2)}} = \frac{(b_2 - \beta_2)}{\text{Se}(b_2)}$$



	coefficient	std. error	t-ratio	p-value
const	40.7676	22.1387	1.841	0.0734 *
x	0.128289	0.0305393	4.201	0.0002 ***

Mean dependent var	130.3130	S.D. dependent var	45.15857
Sum squared resid	54311.33	S.E. of regression	37.80536
R-squared	0.317118	Adjusted R-squared	0.299148
F(1, 38)	17.64653	P-value(F)	0.000155
Log-likelihood	-201.0297	Akaike criterion	406.0594
Schwarz criterion	409.4372	Hannan-Quinn	407.2807



$f \sim N(0, \text{var}(f))$: normal error term or large T

$$\Rightarrow \frac{f}{\sqrt{\text{var}(f)}} \sim N(0,1)$$

$$\Rightarrow \frac{f}{\sqrt{\hat{\text{var}}(f)}} = \frac{f}{\text{se}(f)} \sim t_{(T-2)}$$

$$\Rightarrow P[-t_{\frac{\alpha}{2}, T-2}^c \leq \frac{\hat{y}_0 - y_0}{\text{se}(f)} \leq t_{\frac{\alpha}{2}, T-2}^c] = 1 - \alpha$$

$$\Rightarrow P[\hat{y}_0 - t_{\frac{\alpha}{2}, T-2}^c \text{se}(f) \leq y_0 \leq \hat{y}_0 + t_{\frac{\alpha}{2}, T-2}^c \text{se}(f)] = 1 - \alpha$$

$$\Rightarrow \hat{y}_0 \pm t_{\frac{\alpha}{2}, T-2}^c \text{se}(f) : 100 \times \alpha \% \text{ Prediction Interval}$$

