Chapter 5 Inferences in the LS Estimation II

남준우·허인(2018), 제4장

Gujarati/Porter92018), 제 4 장, 5 장

1. Hypothesis Testing
2. 2-t Rule of Thumb
3. P-value
4. Implications of Hypothesis Testing
5. Statistical Significance vs. Economic Significance
6. Prediction
7. Reporting Regression Results

◎ Assumptions Revisited

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 $i \neq j$.

\+

⑦ (Normality Assumption) $\varepsilon_i \sim N(0, \sigma^2)$.

► With ass. ②, $y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$.

1. Hypothesis Testing

$$H_0 : \beta_2 = \beta_2^*, \qquad H_1 : \beta_2 \neq \beta_2^*$$

(1) Confidence Interval Approach

Since $100(1-\alpha)\%$ C.I. for β_2 is $\left[b_2 \pm t_{(n-2;\alpha/2)} s_{b_2} \right] \equiv CI$,

if $\beta_2^* \in CI$, then do not reject H_0

if $\beta_2^* \notin CI$, then reject H_0.

(2) Testing Procedure

① Set up null and alternative hypotheses: $H_0 : \beta_2 = \beta_2{*}$, $H_1 : \beta_2 \neq \beta_2{*}$.

② Configure test statistics, and its probability distribution: $\dfrac{b_2 - \beta_2}{s_{b_2}} \sim t(n-2)$.

③ Calculate the value of test statistics under H_o: $\dfrac{b_2 - \beta_2{*}}{s_{b_2}} \equiv t$.

④ At given α, find the critical value: $t_{(n-2;\alpha/2)}$.

⑤ Find rejection region: If $|t| \geq t_{(n-2;\alpha/2)}$, reject H_0.

If $|t| < t_{(n-2;\alpha/2)}$, do not reject H_0.

⑥ Decision

(Intuition)

(Example) In the consumption example,

$b_2 = 0.509, \qquad n = 10, \qquad s_{b_2} = \sqrt{0.0013} = 0.036$

95% C.I. for β_2 is $(0.509 \pm 2.306 \times 0.0357)$.

► Suppose $H_0 : \beta_2 = 0.8, \qquad H_1 : \beta_2 \neq 0.8$.

Then $t = \dfrac{b_2 - \beta_2 *}{s_{b_2}} = \dfrac{0.509 - 0.8}{0.0357} = -8.1512$ and $t_{(10-2;0.05/2)} = 2.306$.

Since $|t| = 8.1512 \geq t_{(n-2;\alpha/2)} = 2.306$, reject H_0.

(3) Type 1 Error and Type 2 Error

► There is a chance that we can make a mistake; 2 errors.

• Type 1 error: Reject H_0 when H_0 is true.

• Type 2 error: Do not reject H_0 when H_0 is false.

► P(Type 1 error)=P(reject $H_0 | H_0$ is true)=level of significance$\equiv \alpha$.

Usually, set $\alpha = 0.01, 0.05$ or 0.1 a priori.

► P(Type 2 error)=P(Do not reject $H_0 | H_0$ is false).

1-P(Type 2 error)= P(reject $H_0 | H_0$ is false)=power.

• Trade-off between P(Type 1 error) and P(Type 2 error).

► As $n\uparrow$, at given α, P(Type 2 error)$\downarrow$.

(4) One-tailed test

① $H_0 : \beta_2 = \beta_2 *$, $H_1 : \beta_2 < \beta_2 *$

Reject H_0 if $t \leq -t_{(n-2;\alpha)}$.

(Example) In the consumption function example, $H_0 : \beta_2 = 0.8, \qquad H_1 : \beta_2 < 0.8$

$t = -8.1512 \;\; \leq \;\; -t_{(n-2;\alpha)} = -1.860$, reject H_0.

② $H_0 : \beta_2 = \beta_2 *, \qquad H_1 : \beta_2 > \beta_2 *$

Reject H_0 if $t \geq t_{(n-2;\alpha)}$.

(Example) In the consumption function example, $H_0 : \beta_2 = 0.8$, $H_1 : \beta_2 > 0.8$

$t = -8.1512 \le t_{(n-2;\alpha)} = 1.860$, do not reject H_0.

2. 2-t Rule of Thumb

Consider $H_0: \beta_2 = 0, \qquad H_1: \beta_2 \neq 0$.

► $H_0: \beta_2 = 0 \quad \Rightarrow \quad X$ is NOT important variables in explaining Y.

► $H_1: \beta_2 \neq 0 \quad \Rightarrow \quad X$ is important variables in explaining Y.

• To test $H_0: \beta_2 = 0, \qquad H_1: \beta_2 \neq 0$,

► calculate $t = \dfrac{b_2 - 0}{s_{b_2}} = \dfrac{b_2}{s_{b_2}}$; t-ratio, and

► at $\alpha = 0.05$, compare with $t_{(n-2;0.05/2)} \rightarrow 1.96 \approx 2$ when $n \rightarrow \infty$.

•So, if $|t| \geq 2$, reject $H_0: \beta_2 = 0 \quad \Rightarrow \quad X$ is important variables in explaining Y.

if $|t| < 2$, do not reject $H_0: \beta_2 = 0 \quad \Rightarrow \quad X$ is NOT important variables in explaining Y.

► This is procedure is called as ‘2-t rule of thumb’.

► Note that 2-t rule of thumb is approximate decision.

• Most computer software report the value of t-ratio at output.

(Example) In the consumption function example, t-ratio=14.257.

So, <u>roughly</u>, income is important variable in explaining consumption variation.

<gretl example> Firm data file

Model 1: OLS, using observations 1-75
Dependent variable: SALES

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	496617	232942	2.132	0.0364	**
ADV	73.8115	9.93915	7.426	<0.0001	***

Mean dependent var	1161059	S.D. dependent var	2451108
Sum squared resid	2.53e+14	S.E. of regression	1862593
R-squared	0.430358	Adjusted R-squared	0.422554
F(1, 73)	55.15056	P-value(F)	1.67e-10
Log-likelihood	−1188.218	Akaike criterion	2380.436
Schwarz criterion	2385.071	Hannan-Quinn	2382.286

3. P-value(marginal significance level)

(1) What is p-value?

(Definition) Given the value of a test statistic(say t), the *p-value(marginal significance level)* is the lowest α for which one would have to reject $H_0: \beta_2 = 0$.

- p-value= $2 \cdot P\left(t(n-2) > |t|\right) = 2\left(1 - F(|t|)\right)$,

  where $F(\cdot)$: cdf of underlying distribution($t(n-2)$ our case).

(*) Underlying distribution means the distribution of the test statistics under H_0.

► So, p-value means the probability that the underlying distribution has more extreme values than the observed test-statistics.

(Examples)

(a) (When σ^2 is known) If $t=\dfrac{b_2}{\sigma_{b_2}}=1.96$, then p-value=0.05 (underlying distribution is $N(0,1)$).

(b) If $t=\dfrac{b_2}{\sigma_{b_2}}=1.87$, then p-value=0.061 (underlying distribution is $N(0,1)$).

(c) If $t=\dfrac{b_2}{\sigma_{b_2}}=-1.87$, then p-value=0.061 (underlying distribution is $N(0,1)$).

(d) (When σ^2 is UNknown) If $t=\dfrac{b_2}{s_{b_2}}=-1.87$ for $n=25$, then p-value=0.074.

(underlying distribution is $t(n-2)$).

(e) If $t=\dfrac{b_2}{s_{b_2}}=2.02$ for $n=13$, then p-value=0.068.

(underlying distribution is $t(n-2)$).

(2) Hypothesis Testing based on P-value

- If $\text{p-value} \leq \alpha$, reject $H_0: \beta_2 = 0$.
- If $\text{p-value} > \alpha$, do not reject $H_0: \beta_2 = 0$.

(Example) Consumption function example,

$\text{t-ratio} = 0.509 / 0.0357 = 14.257.$

p-value=0.

• <u>What's difference with the conventional(Neymann-Pearson) testing procedure</u>?

Consider $N(0, 1)$ distribution and suppose test statistics=1.87. P-value=0.061.

① At $\alpha = 0.05$,

Conventional method:

critical value $z_{0.05/2} = 1.96 > \text{test-statistics} = 1.87$, do not reject $H_0 : \beta_2 = 0$.

P-value method: $\text{p-value} = 0.061 > \alpha = 0.05$, do not reject $H_0 : \beta_2 = 0$.

② At $\alpha = 0.10$,

Conventional method: critical value $z_{0.05/2} = 1.645 < \text{test-statistics} = 1.87$, reject $H_0 : \beta_2 = 0$.

P-value method: $\text{p-value} = 0.061 > \alpha = 0.10$, reject $H_0 : \beta_2 = 0$.

► Same decision.

(3) <u>What's advantage of using p-value</u>?

① Sometimes, knowing the value of test-statistics does not give us any information before you look into the statistical table.

② We leave it up to the reader to pick his own α?

(Example) Consider p-value=0.061 in $N(0, 1)$ distribution and suppose test statistics=1.87.

(Example) For $X \sim N(\mu, \sigma^2)$, $\hat{\theta} = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = 1.78$,

► $\alpha = 0.05$, $z_{\alpha/2} = 1.96$, do not reject H_0.

► $\alpha = 0.10$, $z_{\alpha/2} = 1.645$, reject H_0.

What do we set for α?

► Since p-value=0.075 for $\hat{\theta} = 1.78$, we can leave it up to the reader to pick his own α.

③ The p-value method provides a measure of how decisive a test result is.

► Conventional method only tells us whether t is just inside or outside the critical region for a given α.

(Example) Suppose there are two different regression results:

$n = 16$, with $t-$distribution,

(case 1) $t_1 = 2.13,\ \text{p-value}_1 = 0.051$.

(case 2) $t_1 = 1.85,\ \text{p-value}_1 = 0.086$.

(4) P-value for one-sided test:

① $H_0 : \beta_2 = 0, \quad H_1 : \beta_2 < 0$.

► $\text{p-value} = F(t)$.

② $H_0 : \beta_2 = 0, \quad H_1 : \beta_2 > 0$

► $\text{p-value} = 1 - F(t)$.

<gretl example>

Model 1: OLS, using observations 1-75
Dependent variable: SALES

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	496617	232942	2.132	0.0364	**
ADV	73.8115	9.93915	7.426	<0.0001	***

Mean dependent var	1161059	S.D. dependent var	2451108
Sum squared resid	2.53e+14	S.E. of regression	1862593
R-squared	0.430358	Adjusted R-squared	0.422554
F(1, 73)	55.15056	P-value(F)	1.67e-10
Log-likelihood	−1188.218	Akaike criterion	2380.436
Schwarz criterion	2385.071	Hannan-Quinn	2382.286

4. Implications of Hypothesis Testing

① Accepting or rejecting.

② Forming null and alternative hypothesis.

5. Statistical Significance vs. Economic Significance

- It has been noted that the hypothesis is widely abused.

► Many researcher emphasize the statistical significance rather than economic significance.

(Example ①) For testing $H_0: \beta_2 = 1$, with test-statistics $b_2 = 1.1$.

► Suppose that the null hypothesis is rejected (a coefficient estimate may be "very significantly different with 1"). However, the difference may be economically trivial.

► Or, by the statistically testing, it may not significantly different from 1 but have an economically substantial magnitude.

(Example) $b_2 = 0.7218,$ 95% C.I.=(0.7129, 0.7306).

Then, $\beta_2 = 0.74$ is rejected?

(Example)

$$\hat{y} = 50 - 1 \cdot X_3, \qquad (\;) 안은 \text{ t-ratio.}$$
$$(-2)$$

where y: weight(g), X_3: exercise(hour).

Is X_3 important variable in reducing weight?

• There is a common practice of indiscriminately reporting all the t-statistics and it encourages rank-ordering of the independent variables with respect to their importance.

(Example)

$$\hat{y} = 50 + 2 \cdot X_2 - 1 \cdot X_3, \qquad ( )\text{안은 t-ratio.}$$
$$\qquad (4) \qquad (-2)$$

where y: weight, X_2: height, X_3: exercise.

Is X_2 more important than X_3 in reducing weight?

6. Prediction

- Suppose, for observation f, the value of $X = X_f$ is known, what is the best predictor of y_f?

► actual: $y_f = \beta_1 + \beta_2 X_f + \varepsilon_f$.

► predicted: $\hat{y}_f = b_1 + b_2 X_f$.

- prediction error: $e_f = y_f - \hat{y}_f$

► $E(e_f) = 0$.

$$V(e_f) = V(b_1) + V(b_2)X_f^{\ 2} + V(\varepsilon_f)$$

► $$+\,2X_f \cdot Cov(b_1, b_2) - Cov(b_1, \varepsilon_f) - 2X_f \cdot Cov(b_2,\, \varepsilon_f)\,.$$

- $Cov(b_1,\, \varepsilon_f) = 0, \quad Cov(b_2,\, \varepsilon_f) = 0$, why?

$$V(e_f) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right) + \frac{\sigma^2 X_f^{\ 2}}{\sum_{i=1}^{n} (X_i - \bar{X})^2} + \sigma^2 + 2X_f \frac{\sigma^2(-\bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}.$$

So, $V(e_f) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right) \equiv \sigma_{e_f}^{\ 2}.$

► Since $e_f \sim N\left(0, V(e_f)\right)$, why?

► $\dfrac{e_f}{\sqrt{V(e_f)}} = \dfrac{y_f - \hat{y}_f}{\sigma_{e_f}} \sim N(0,1)$. (*)

► 100(1-α)% C.I. for y_f is $\left(\hat{y}_f \pm z_{\alpha/2}\sigma_{e_f}\right)$.

- However, σ_{e_f} (or $\sigma_{e_f}{}^2$) is unknown.

► Estimator of $\sigma_{e_f}{}^2$, $\hat{\sigma}_{e_f}{}^2 = \hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right) \equiv s_{e_f}{}^2$,

where $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i{}^2$.

► s_{e_f} is called as <u>standard error of prediction error</u>.

- Then, (*) becomes $\frac{y_f - \hat{y}_f}{s_{e_f}} \sim t(n-2)$.

► 100(1-α)% C.I. for y_f is $\left(\hat{y}_f \pm t_{(n-2;\alpha/2)} s_{e_f} \right)$.

(Example) In the consumption example,

when $X_f = 200$, $\hat{y}_f = 24.5 + 0.51 \times 200 = 126.25$.

$$s_{e_f}{}^2 = 42.16\left(1 + \frac{1}{10} + \frac{(200-170)^2}{33000}\right) = 47.52.$$

$$s_{e_f} = 6.89$$

Therefore, 95% C.I. for y_f when $X_f = 200$ is $(126.25 \pm 2.306 \cdot 6.89)$.

(Note)

When $|X_f - \bar{X}| \uparrow$, then C.I. gets wider.

► Less informative for y_f.

► Which case is $|X_f - \bar{X}| \uparrow$?

7. Reporting Regression Results

①

$$\hat{Y}_i = -0.036 + 0.046 X_i$$
$$(0.548)\ (0.016) \qquad R^2 = 0.095$$

(　)안은 표준오차.

②

$$\hat{Y}_i = -0.036 + 0.046 X_i$$
$$(-0.066)\ (2.950) \qquad R^2 = 0.095$$

(　)안은 t-값.

► And mostly, use tables for reporting several regression results.

<표6> 회귀 분석 추정 결과

	모형(A)	모형 (B)	모형 (C)	모형 (D)
AGE			0.002*	0.002
			(0.077)	(0.116)
ORDER	0.003***		0.004***	
	(0.0088)		(0.0006)	
OrderLast		0.025***		0.028***
		(0.006)		(0.001)
ROUND=3	0.024***	0.020***	0.022***	0.018**
	(0.001)	(0.005)	(0.001)	(0.011)
Constant	5.562***	5.577***	5.480***	5.500***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.793	0.792	0.3457	0.3081

괄호안은 임의확률(p-value)을 나타냄.

*: p-value<0.10, **: p-value<0.05, ***: p-value<0.01