Chapter 4 Inferences in the LS Estimation I

남준우·허인(2018), 제4장

Gujarati/Porter(2018), 제4장, 제5장

(1) Probability Distribution of LS Estimator when σ^2 is known

① Probability Distribution

② Confidence Interval for β_2

(2) Probability Distribution of LS Estimator when σ^2 is UNKNOWN

① Estimator of σ^2 and $V(b_2)$

② Sampling distribution when σ^2 is unknown

(3) Review: Normal distribution, Chi-square distribution and t-distribution

(4) Review: Sampling Distributions in (경제통계학 vs. 계량경제학)

(5) Confidence Interval for Regression Coefficient

◎ <u>Assumptions Revisited</u>

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) Probability Distribution of LS Estimator when σ^2 is known

① Probability Distribution

- Note that b_2 is a linear function of y_i.

Since $b_2 \sim N\left(\beta_2, V(b_2) \equiv \sigma_{b_2}{}^2\right)$, where $V(b_2) = \dfrac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$,

$$\boxed{\frac{b_2 - \beta_2}{\sigma_{b_2}} \sim N(0,1)}$$

where $\sigma_{b_2} = \sqrt{V(b_2)} = \dfrac{\sigma}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}$.

② <u>Confidence Interval for β_2</u>

- So, $100(1-\alpha)\%$ C.I. for β_2 is $\left[b_2 \pm z_{\alpha/2}\sigma_{b_2}\right]$,

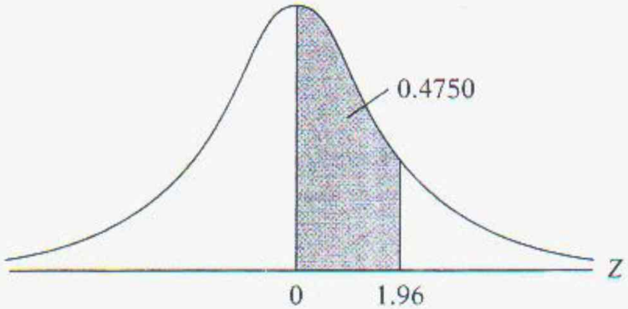
where $z_{\alpha/2}$ is critical value of $N(0,1)$ where right tail area is of $\alpha/2$.

TABLE D.1
Areas Under the Standardized Normal Distribution

Example

$\Pr(0 \le Z \le 1.96) = 0.4750$

$\Pr(Z \ge 1.96) = 0.5 - 0.4750 = 0.025$



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4454	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4454	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Note: This table gives the area in the right-hand tail of the distribution (i.e., $Z \geq 0$). But since the normal distribution is symmetrical about $Z = 0$, the area in the left-hand tail is the same as the area in the corresponding right-hand tail. For example, $P(-1.96 \leq Z \leq 0) = 0.4750$. Therefore, $P(-1.96 \leq Z \leq 1.96) = 2(0.4750) = 0.95$.

(2) Probability Distribution of LS Estimator when σ^2 is unknown

Since σ^2 is unknown,

$V(b_2) = \dfrac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$ or σ_{b_2} is unknown.

► We have to estimate σ^2.

① Estimator of σ^2 and $V(b_2)$

- Use $\widehat{\sigma^2} = s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$ then $E(s^2) = \sigma^2$.

► Also, use $\hat{V}(b_2) = s_{b_2}{}^2 = \frac{s^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$; estimated variance of b_2 (or estimator of $V(b_2)$)

then $E\left(s_{b_2}{}^2\right) = V(b_2)$.

► $s_{b_2} \equiv \sqrt{\hat{V}(b_2)} = \dfrac{s}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}$; standard error of b_2.

(Example) In the consumption 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		

- $s^2 = \dfrac{\sum_{i=1}^{n} e_i^2}{n-2} = \dfrac{337.27}{10-2} = 42.16, \qquad s = \sqrt{42.16} = 6.49,$
- $s_{b_2}{}^2 = \dfrac{42.16}{33000} = 0.0013, \qquad s_{b_2} = \sqrt{0.0013} = 0.036$

② Sampling distribution when σ^2 is unknown

$$\boxed{t = \frac{b_2 - \beta_2}{s_{b_2}} \sim t(n-2)}$$

(proof)

(claim) $\dfrac{\sum_{i=1}^{n} e_i^2}{\sigma^2} \sim \chi^2(n-2)$.

(proof of claim)

(3) Review: Normal distribution, Chi-square distribution and t-distribution

If $Y_i \sim N(\mu, \sigma^2)$, (note that Y_i's are independent)

① $Z_i = \dfrac{Y_i - \mu}{\sigma} \sim N(0, 1)$.

② $W = \sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$.

► W is right skewed with $E(W) = k, V(W) = 2k$.

③ $t = \dfrac{Z_j}{\sqrt{W/k}} \sim t(k)$.

► t has same shape as $N(0, 1)$ with $E(t) = 0,\ V(t) = \dfrac{k}{k-2}$.

(4) Review: Sampling Distributions in 경제통계학, 계량경제학

경제통계학	계량경제학
$X_i \sim N(\mu, \sigma^2)$, μ is unknown.	$y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$, (β_1, β_2) is unknown.

경제통계학	계량경제학

경제통계학	계량경제학

(5) Confidence Interval for Regression Coefficient

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

where $t_{(n-2;\alpha/2)}$ is critical value of $t(n-2)$ distribution where right tail area is of $\alpha/2$.

(Example) In the consumption example,

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

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

TABLE D.2
Percentage Points of the *t* Distribution

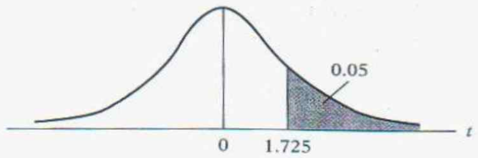
Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.

Example

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ for df = 20

$\Pr(|t| > 1.725) = 0.10$



df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

<<u>Interpretation and Note</u>>

① Meaning of C.I.

► One sample vs. repeated sampling

② What if ass. ⑦ $\varepsilon_i \sim N(0, \sigma^2)$ does NOT hold?

<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