Printout

Tuesday, February 20, 2018 7:01 AM

5. Interval Estimation and Hypothesis Testing

Interval Estimation

POINT ESTIMATOR Vs. INTERVAL ESTIMATOR

$\hat{\beta}_1, \hat{\beta}_2$

$$Pr(\hat{\beta}_2 - \delta \le \beta_2 \le \hat{\beta}_2 + \delta) = 1 - \alpha$$

LOWER CONFIDENCE LIMIT

UPPER CONFIDENCE LIMIT

LEVEL OF SIGNIFICANCE

Probability that random intervals, like $(\hat{\beta}_2 - \delta, \hat{\beta}_2 + \delta)$, will contain "the true β_2" is equal to $1 - \alpha$ %.

$1-\alpha$ = CONFIDENCE COEFFICIENT

Ex $\alpha = 0.05$

$1-\alpha = 0.95$ (OR 95%)

How to interpret this in statistics sense?

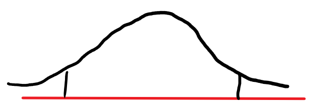
Answer: If you conduct 100 intervals (OR 100 CASES of INTERVAL)

95 OUT OF 100 cases, β_2 will {lie or be included or fall} in the intervals!!!

Page 2

Tuesday, February 20, 2018 7:01 AM

5.1 Confidence Intervals for Regression Coefficients β_1 and β_2



$$\boxed{t = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}} \cdot \sqrt{\sum x_i^2}}$$

we can establish CI for β_2 as follows:

$$\text{Prob}\left[-t_{\frac{\alpha}{2}} \leq t \leq +t_{\frac{\alpha}{2}} \right] = 1-\alpha$$

$$\text{Prob}\left[-t_{\frac{\alpha}{2}} \leq \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \leq t_{\frac{\alpha}{2}} \right] = 1-\alpha$$

$$\text{Prob}\left[-t_{\frac{\alpha}{2}} \cdot se(\hat{\beta}_2) \leq \hat{\beta}_2 - \beta_2 \leq t_{\frac{\alpha}{2}} \cdot se(\hat{\beta}_2) \right] = 1-\alpha$$

$$\text{Prob}\left[-\hat{\beta}_2 - t_{\frac{\alpha}{2}} \cdot se(\hat{\beta}_2) \leq -\beta_2 \leq -\hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot se(\hat{\beta}_2) \right] = 1-\alpha$$

$$\text{Prob}\left[\hat{\beta}_2 - t_{\frac{\alpha}{2}}\, se(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot se(\hat{\beta}_2) \right] = 1-\alpha$$

so, $100(1-\alpha)\%$ confidence interval for β_2 is...

$$\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot se(\hat{\beta}_2)$$

Page 3

Tuesday, February 20, 2018 7:01 AM

In Sum

A $100(1-\alpha)$ percent **confidence interval** for β_2 can be defined as:

$$\hat{\beta}_2 \pm t_{\alpha/2}\text{se}(\hat{\beta}_2)$$

or

$$\Pr[\hat{\beta}_2 - t_{\alpha/2}\text{se}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\text{se}(\hat{\beta}_2)] = 1-\alpha$$

Analogously, we can define $100(1-\alpha)$ percent **confidence interval** for β_1 as:

$$\hat{\beta}_1 \pm t_{\alpha/2}\text{se}(\hat{\beta}_1)$$

or

$$\Pr[\hat{\beta}_1 - t_{\alpha/2}\text{se}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}\text{se}(\hat{\beta}_1)] = 1-\alpha$$

Example

Table 5.1: Estimating the expenditure of the household with income

Family (i)	Actual Y_i	Income X_i	x_i $X-\bar{X}$	y_i $Y-\bar{Y}$	$x_i \cdot y_i$ $(X-\bar{X})(Y-\bar{Y})$	x_i^2 $(X-\bar{X})^2$
1	390	500	-250	-153.17	38291.67	62500
2	425	600	-150	-118.17	17725.00	22500
3	560	700	-50	16.83	-841.67	2500
4	575	800	50	31.83	1591.67	2500
5	630	900	150	86.83	13025.00	22500
6	679	1000	250	135.83	33958.33	62500
Sum	3259	4500	0	0	103750	175000
	$\bar{X}=?$	$\bar{Y}=?$	$\sum x_i=0$	$\sum y_i=0$	$\sum x_i y_i$	$\sum x_i^2$

Table 5.2: Estimating the expenditure of the household with income

Family (i)	Actual Y_i	Income X_i	Regression Estimate $\hat{Y}$	Residual $Y-\hat{Y}$	Residual squared $(Y-\hat{Y})^2 = \hat{u}_i^2$
1	390	500	394.95	-4.95	24.53
2	425	600	454.24	-29.24	854.87
3	560	700	513.52	46.48	2160.04
4	575	800	572.81	2.19	4.80
5	630	900	632.10	-2.10	4.39
6	679	1000	691.38	-12.38	153.29
Sum	3259	4500	0	0	3201.90 — $\sum \hat{u}_i^2$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{103750}{175000} = 0.593$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 98.524$$

$$\hat{Y}_i = 98.524 + 0.593 X_i$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{\hat{\sigma}^2}{\sum x_i^2} = \frac{800.476}{175000} = 0.0046. \Rightarrow se(\hat{\beta}_2) = \sqrt{0.0046} = 0.0678$$

σ^2 can be estimated by $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-2} = \dfrac{3201.90}{6-2} = 800.476$

Verify that $se(\hat{\beta}_1) = 52.023$ (D-I-Y)

Confidence Interval for β_2

$$\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot se(\hat{\beta}_2)$$

suppose $\alpha = 0.05$ (or 95% level of confidence)

$$\pm t_{\frac{\alpha}{2}, df} = \pm t_{\frac{0.05}{2}, n-2} = \pm t_{0.025, 4} = \pm 2.776$$

$$\rightarrow 0.593 \pm (2.776)(0.0678)$$

$$\rightarrow 0.593 \pm 0.1882$$

so $0.4048 \leq \beta_2 \leq 0.7812$ #

Interpretation?

Given the confidence level of 95%, In the long run, 95 out of 100 cases, intervals like (0.4048, 0.7812) will "contain" the true β_2

Confidence Interval for β_1

D-I-Y : $\hat{\beta}_1 \pm t_{\frac{\alpha}{2}} \cdot se(\hat{\beta}_1)$

Page 6

Tuesday, February 20, 2018 7:01 AM

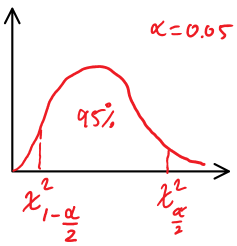
5.2 Confidence Interval for σ^2

w/ chi-square distribution

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \quad \text{with } (n-2)\text{ df}$$

we can use χ^2 to establish a confidence interval for σ^2:

$\alpha = 0.05$ suppose



$$\Pr\left[\chi^2_{1-\frac{\alpha}{2}, n-2} \le \chi^2 \le \chi^2_{\frac{\alpha}{2}, n-2}\right] = 1-\alpha$$

$$\Pr\left[\chi^2_{1-\frac{\alpha}{2}, n-2} \le (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2}, n-2}\right] = 1-\alpha \qquad 3 \le x \le 5$$

$$\Pr\left[\frac{1}{\chi^2_{\frac{\alpha}{2}, n-2}} \le \frac{\sigma^2}{(n-2)\hat{\sigma}^2} \le \frac{1}{\chi^2_{1-\frac{\alpha}{2}, n-2}}\right] = 1-\alpha \qquad \frac{1}{5} \le \frac{1}{x} \le \frac{1}{3}$$

$$\Pr\left[\frac{(n-2)\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2}, n-2}} \le \sigma^2 \le \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2}, n-2}}\right] = 1-\alpha$$

Ex: $\hat{\sigma}^2 = 800.476 \qquad \chi^2_{0.025, 4} = 11.1433$

$\chi^2_{0.975, 4} = 0.484$

$$\Pr\left[\frac{4(800.476)}{11.1433} \le \sigma^2 \le \frac{4(800.476)}{0.484}\right] = 0.95.$$

$$\Pr\left[287.347 \le \sigma^2 \le 6615.504\right] = 0.95$$

So 95% CI for $\sigma^2 \Rightarrow$ $\underline{287.347 \le \sigma^2 \le 6615.504}$ #

Given the confidence coefficient of 95%, in 95 out of 100 cases, intervals like (287.347, 6615.504) will contain the true but unknown σ^2. :)

Page 7

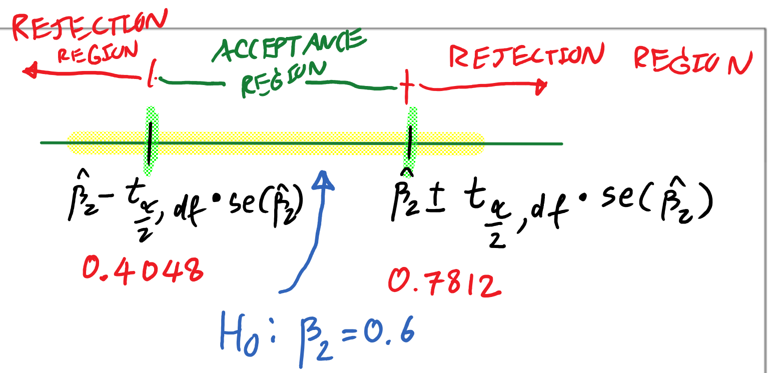
Tuesday, February 20, 2018 7:01 AM

5.3 Hypothesis Testing: The Confidence-Interval Approach — I

Based on our sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 = 0.6$$ (Null hypothesis)

$$H_1 : \beta_2 \neq 0.6$$ (Alternative hypothesis)



We cannot reject that $H_0 : \beta_2 = 0.6$ with 95% confidence interval

i.e., β_2 is not statistically different from 0.6. #

II

5.4 Hypothesis Testing: The Test of Significance Approach

5.4.1 Two-Tail Test

Based on the sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 = \cancel{0.6}\ \ 0.2$$
$$H_1 : \beta_2 \neq \cancel{0.6}\ \ 0.2$$

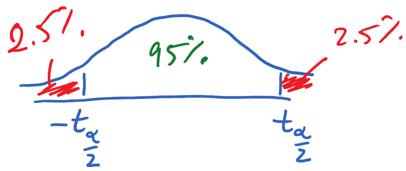
$H_0: \beta_2 = 0.2$ (true but unknown MPC' is equal to 0.2)

$H_1: \beta_2 \neq 0.2$ (true but unknown MPC' is not equal to 0.2)

① set $\alpha = 0.05$

② compute $\hat{t}$ or $t_{calculated}$ (t_{cal}) :

$$\hat{t} = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0.593 - 0.2}{0.0678} = 5.796$$



③ calculate the critical value of t : $t_{\frac{\alpha}{2}, n-2} = t_{0.025, 4} = 2.776$

④ Since $\hat{t} > t_{critical} = 2.776$, we then "REJECT" the null hypothesis $H_0: \beta_2 = 0.2$!

Therefore, with $\alpha = 0.05$, β_2 is statistically different from 0.2! (i.e., $\beta_2 \neq 0.2$)

Page 9

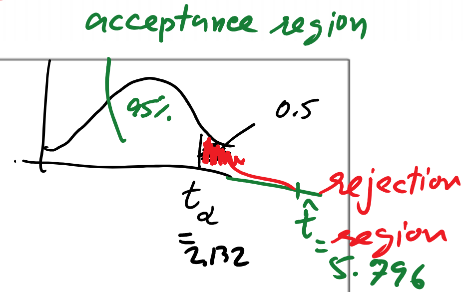
Tuesday, February 20, 2018 7:01 AM

5.4.2 One-Tail Test

Based on the sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 \le 0.2$$
$$H_1 : \beta_2 > 0.2$$

$H_0: \beta_2 \le 0.2$ [ONE-TAIL TEST]
$H_1: \beta_2 > 0.2$



① set $\alpha = 0.05$

② compute $\hat{t}$: $\hat{t} = \dfrac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = 5.796$

③ find $t_{critical}$: $t_{0.05,4} = 2.132$

④ Since $\hat{t} > t_{critical}$, then we REJECT $H_0: \beta_2 \le 0.2$ in favor of $H_1: \beta_2 > 0.2$ with 95% confidence level. #

Page 10

Tuesday, February 20, 2018 7:01 AM

We can summarize the decision rules for the t test as follow:

Figure 5.1 The t test of Significance: Decision rules

Type of hypothesis	H_0: the null hypothesis	H_1: the alternative hypothesis	Decision rule: reject H_0 if
Two-tail	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\lvert t \rvert > t_{\alpha/2,\text{df}}$
Right-tail	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_{\alpha,\text{df}}$
Left-tail	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_{\alpha,\text{df}}$

Notes: β_2^* is the hypothesized numerical value of β_2.
$|t|$ means the absolute value of t.
t_α or $t_{\alpha/2}$ means the critical t value at the α or $\alpha/2$ level of significance.
df: degrees of freedom, $(n-2)$ for the two-variable model, $(n-3)$ for the three-variable model, and so on.
The same procedure holds to test hypotheses about β_1.

5.4.3 Testing the significance of σ^2: The χ^2 test

$H_0: \sigma^2 = 750$

$H_1: \sigma^2 \neq 750$

D-I-Y

HINT: you have to calculate $\chi^2_{cal} = (n-2)\dfrac{\hat{\sigma}^2}{\sigma^2}$

SOLUTION: CANNOT REJECT $H_0: \sigma^2 = 750$ in favor of $H_1: \sigma^2 \neq 750$.

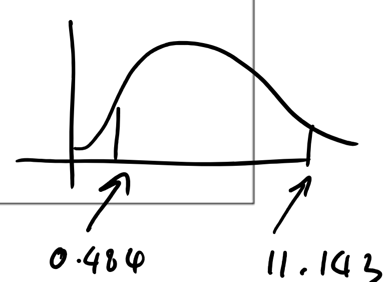


Figure 5.2 The χ^2 Test : Decision rules

H_0: the null hypothesis	H_1: the alternative hypothesis	Critical region: reject H_0 if
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha,\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha),\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2,\text{df}}$ or $< \chi^2_{(1-\alpha/2),\text{df}}$

Note: σ_0^2 is the value of σ^2 under the null hypothesis. The first subscript on χ^2 in the last column is the level of significance, and the second subscript is the degrees of freedom. These are critical chi-square values. Note that df is $(n-2)$ for the two-variable regression model, $(n-3)$ for the three-variable regression model, and so on.

Why do we say "we cannot reject the null hypothesis?" instead of "We accept the null hypothesis"