Lecture 13 - Economics 113

Professor Spearot

- Agenda
  1. Confidence Intervals
  2. P-values
- Homework #1 is posted
- Exam #3 in one week!
- Quote of the day:
  "People who are trillions of dollars in debt, yelling at people who are billions of dollars in debt."
  - Jay Leno

Confidence intervals

A more informative technique

- Confidence intervals give us the region of most likely β_j.
- Recall that we cannot reject the null of a two sided t-test if:

$$\left| \frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \right| < t_{crit}$$

- This is satisfied if:

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} < t_{crit} \quad \text{and} \quad -\left(\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \right) < t_{crit}$$

- Manipulating the first, we get:

$$\widehat{\beta}_j - t_{crit} \cdot se(\widehat{\beta}_j) < \beta_j$$

- The second:

$$\beta_j < \widehat{\beta}_j + t_{crit} \cdot se(\widehat{\beta}_j)$$

Confidence intervals

A more informative technique

- Combining the two results, we get a "confidence interval" for β_j

$$\underbrace{\widehat{\beta}_j - t_{crit} \cdot se(\widehat{\beta}_j)}_{lower} < \beta_j < \underbrace{\widehat{\beta}_j + t_{crit} \cdot se(\widehat{\beta}_j)}_{upper}$$

- The range of values **which cannot be rejected** using a hypothesis test.
- If t_{crit} is at calculated at 95% confidence, this is called the 95% confidence interval.
- Why are these more informative?
- Effect of $se(\widehat{\beta}_j)$?
  - $se(\widehat{\beta}_j) \uparrow \implies$ wider confidence interval.
- Effect of t_{crit}?
  - $t_{crit} \uparrow$ implies a higher level of confidence
  - $\implies$ wider confidence interval
  - What if I desire 100% confidence?
    - $-\infty < \beta_j < \infty$

Confidence intervals

A more informative technique

- Back to the crime example:

$$\widehat{\log(Crime)} = \underset{(0.104)}{-6.63} + \underset{(0.11)}{1.27} \log(Enroll)$$
$$obs = 197,\ R^2 = 0.585$$

- 95% confidence interval for the coefficient on $\log(Enroll)$?
- Interpret?

$$1.27 - 1.96 \cdot 0.11 < \beta_{enroll} < 1.27 + 1.96 \cdot 0.11$$
$$1.054 < \beta_{enroll} < 1.485$$

- 99% Confidence interval?
- $T_{stat} = 2.58$
- Use Formula:

$$1.27 - 2.58 \cdot 0.11 < \beta_{enroll} < 1.27 + 2.58 \cdot 0.11$$
$$0.9862 < \beta_{enroll} < 1.5538$$

Confidence intervals

Housing prices and cancer risk

- Housing prices and cancer risk (Davis, 2004)
- Housing Data - two similar Nevada Counties
  1. Churchill county
     - 31 new cases of Pediatric Leukemia (PL), 1997-2002
     - Similarly sized counties should expect 1 new case.
  2. Lyon county
     - A similar county, located directly to the west.
- Specification

$$\log(Price) = \beta_1 Risk + Other + u$$

  1. $\log(Price)$: log value of housing prices
  2. $Risk$: Index of perceived PL risk
  3. $Other$: Other controls (lot size, floor space, etc.)
- Housing data from 1990-2002

TABLE 3—THE EFFECT OF HEALTH RISK ON HOUSING VALUES

	OLS (1)	OLS (2)	FE
Leukemia risk (linear spline)	−0.123	−0.156	−0.140
	(0.013)	(0.017)	(0.015)
Lot size (acres)	0.011	0.012	—
	(0.002)	(0.002)	
Lot size squared	−1.88E-05	−2.02E-05	—
	(3.20E-06)	(3.18E-06)	
Floor space (square feet, 100s)	0.054	0.044	—
	(0.001)	(0.001)	
Building age (years)	−0.009	−0.006	—
	(0.001)	(0.001)	
Building age squared	3.57E-05	1.20E-05	—
	(8.61E-06)	(8.42E-06)	
Churchill County dummy	—	0.068	—
		(0.009)	
Class dummies	no	yes	—
Year dummies	no	yes	yes
Month dummies	no	yes	yes
n	10204	10204	4922
R^2	0.60	0.63	0.05

Notes: The sample consists of sales of single-family residences from 1990 to 2002 from both counties. The dependent variable is sales price in logs. The linear spline is zero through 1999, rises by 1/24 each month during 2000 and 2001, and then takes the value of one. For the control county the linear spline is equal to zero for all periods. Standard errors are corrected for heteroskedasticity and correlated errors within county-month groups.

Confidence intervals

Housing prices and cancer risk

- Estimation (other effects suppressed for brevity)

$$\log(Price) = \underset{(0.017)}{-0.156} Risk$$
$$n = 10204,\ R^2 = 0.63$$

- What is the 99% confidence interval β_1?
- $t_{crit} = 2.576$
- Use the formula:

$$\begin{aligned} \widehat{\beta}_1 - t_{crit} \cdot se(\widehat{\beta}_1) &< \beta_1 < \widehat{\beta}_1 + t_{crit} \cdot se(\widehat{\beta}_1) \\ -0.156 - 2.576 \cdot 0.017 &< \beta_1 < -0.156 + 2.576 \cdot 0.017 \\ -0.200 &< \beta_1 < -0.112 \end{aligned}$$

- At the 99% level, cancer risk negatively affects housing prices.

P-values

A different approach

- How likely is it that I falsely reject the null?
- P-value:

The probability that the null hypothesis is falsely rejected

- In the crime and enrollment example

$$H_0 : \beta_{Enroll} = 1,\ H_A : \beta_{Enroll} \neq 1$$

- T-statistic

$$tstat = \frac{1.27 - 1}{0.11} = 2.45$$

- The p-value is the value of the following

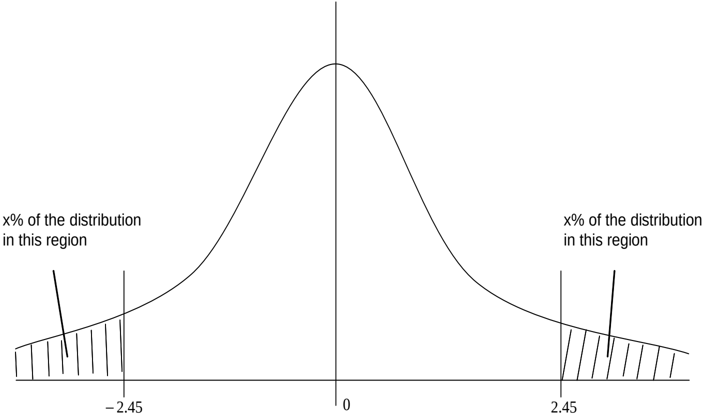
$$pvalue = \Pr(|T| > 2.45)$$

- The p-value is the probability that I randomly draw a value from the t-distribution that is larger (in absolute terms) than the estimated T-statistic

P-values

An illustration

- P-value: The smallest significance level such that the null would be rejected.



- The p-value for this is the probability of being in the shaded region.

P-values

A different approach

- To find the p-value
  1. Calculate the t-statistic
  2. Find the closest match to your t-statistic on the normal table
  3. The p-value is the significance level that generates this t-statistic.
- In the crime and enrollment example
  1. $t_{stat} = 2.45$
  2. $\Pr(|T| > 2.45)$?

$$
\begin{aligned}
\Pr(|T| &> 2.45) = \Pr(T > 2.45 \bigcup T < -2.45) \\
&= \Pr(T > 2.45 \bigcup T < -2.45) \\
&= \Pr(T > 2.45) + \Pr(T < -2.45) = 2 * \Pr(T > 2.45) \\
&= 2 * (1 - 0.9929) = 0.0142
\end{aligned}
$$

- Thus, we will falsely reject the null less than 1.42% of the time.