Lecture 10

Multiple Regression V

Annoucements

- Quiz 3
- PS3 is uploaded. (Due Wednesday 3rd March).

Outline

- Level- Level model
- Log-Level model
- Level-Log model
- Log- Log model
- Interaction effects

Example:
code categorical variables into dummies

- Sometimes, you want to turn a categorical variable into dummy variables to model the relationship more flexibly
- e.g. Education: 0 if HS dropout, 1 if HS grad, and 2 if college grad. Recode it into three dummy variables:
  - less_HS:1 if education=0, 0 otherwise
  - HS_grad:1 if education=1, 0 otherwise
  - college_or_above:1 if education=2, 0 otherwise

Dummy variable trap

- If we run Health=$\beta_0 + \beta_1$Less_HS+β_2HS_grad+β_3College_or_above+ε
  - The regression will not work
  - This is called perfect multi-collinearity

Constant	less_HS	HS_grad	college	health status
1	1	0	0	3
1	0	1	0	4
1	0	1	0	4
1	0	0	1	3
1	0	0	1	2
1	1	0	0	1
1	0	1	0	4

Less_HS+HS_grad+college_or_above =constant
Cannot estimate the regression!

How to get out of dummy variable trap(1)?

- There are two ways of getting out of the dummy variable trap
- Way #1: omit one category of the dummy variables
- e.g. if we run a regression:
- Health=$\beta_0 + \beta_1$HS_grad+β_2College_or_above+ε
- We omit HS dropouts dummy and treat it as the baseline. The categories we include are compared to the category we exclude
- How do we interpret β_1?
  - The average health status of HS grads relative to HS dropouts.
  - **It is the difference in average health between HS grad and HS dropouts**
- Exercise: Interpret $\beta_{0,}\beta_2$

How to get out of dummy variable trap(2)?

- Way#2: omit the constant term
- Health=$\beta_1 \mathrm{Less}HS + \beta_2HS_grad+\beta_3$College_or_above+ε
  - How does this regression differ from the previous?
- How do we interpret β_1?
  - The average health status of HS dropouts.
- Exercise: Interpret β_2 and β_3

Level-Level

- A “Level-level” regression specification.
- $y = \beta_0 + \beta_1 x + u$
- This is called a “level-level” specification because raw values (levels) of y are being regressed on raw values of x.
- How do we interpret β_1?
- We interpret it as the increase in y when 1 unit of x increases.

Log-Level

- A “Log-level” regression specification
- $\log(y) = \beta_0 + \beta_1 x + u$
- What is log(y)? Log indicates logarithmic function.
- This is called a “log-level” specification because the natural log transformed values of y are being regressed on raw values of x.
- You might want to run this specification if you think that increases in x lead to a constant *percentage* increase in y. (e.g. wage on education)

Log-Level

- A “Log-level” regression specification
- $\log(y) = \beta_0 + \beta_1 x + u$
- How do we interpret β_1?
- One unit change in x leads to 100* β_1 percent change in Y

Level-Log

- A “Level-log” regression specification
- $y = \beta_0 + \beta_1 \log(x) + u$
- This is called a “level-log” specification because y is being regressed on natural log transformed values of x.
- You might want to run this specification if you think that *percentage* increases in x lead to a constant increase in y.

Level-Log

- A “Level-log” regression specification
- $y = \beta_0 + \beta_1 \log(x) + u$
- How do we interpret β_1?
- One percent change in x leads to $\beta_1/100$ unit change in Y

Log-Log

- A “Log-log” regression specification
- $\log(\mathrm{y}) = \beta_0 + \beta_1 \log(x) + u$
- This is called a “log-log” specification because natural log transformed values of y are being regressed on natural log transformed values of x.
- You might want to run this specification if you think that *percentage* increases in x lead to a constant *percentage* changes in y.

Log-Log

- A “Log-log” regression specification
- $\log(\mathrm{y}) = \beta_0 + \beta_1 \log(x) + u$
- How do we interpret β_1?
- One percent change in x leads to β_1 percent change in Y

Model	Equation	Interpretation
Level-Level Regression	$Y = \alpha + \beta X$	One unit change in X leads to β unit change in Y
Log-Linear Regression	$log(Y) = \alpha + \beta X$	One unit change in X leads to $100 * \beta$ percent change in Y
Linear-Log Regression	$Y = \alpha + \beta\, log(X)$	One percent change in X leads to $\beta/100$ unit change in Y
Log-Log Regression	$log(Y) = \alpha + \beta\, log(X)$	One percent change in X leads to β percent change in Y

Examples

$$\log(\widehat{earnings}) = 2.805 + 0.0087Age$$

Interpret the coefficient of age.

Examples

$$\log(\widehat{earnings}) = 2.805 + 0.0087Age$$

Interpret the coefficient of age.

Earnings are predicted to increase by 0.87 [0.0087*100]% for each additional year of age.

Examples

$$\widehat{\text{test score}} = 557.8 + 36.42 \ln(Income)$$

Interpret the coefficient of log(income).

Examples

$$\widehat{\text{test score}} = 557.8 + 36.42 \ln(Income)$$

Interpret the coefficient of log(income).

A 1% increase in income is associated with an increase in test scores of 36.42/100=0.36 point.

Examples

$$\widehat{\text{test score}} = 6.336 + 0.0554\ln(Income)$$

Interpret the coefficient of log(income).

Examples

$$\widehat{\text{test score}} = 6.336 + 0.0554\ln(Income)$$

Interpret the coefficient of log(income).

A 1% increase in income is associated with an increase in test scores of 0.0554 percent.

Percent vs. percentage point

- Most of us are comfortable with percentage increases and decreases.
- A hundred dollar skateboard goes on sale for 75 dollars and we can calculate easily enough that this is a 25 percent discount. The key feature of percentage change is that it provides a measure of change that is proportional to the original quantity (100 dollars in this case).

Percent vs. percentage point

- Unfortunately, this simple setup can become confusing when the original quantity is itself expressed as a percentage.
- For example, I heard on the news that between 1995 and 2005, the percentage of Americans without health insurance rose from 60 percent to 69 percent.
- It is tempting to call this a 9 percent increase, but this understates the size of the increase. Sixty nine percent is actually a 15 percent increase over the original sixty percent. Try it. If we start with 60, and add to it 15 percent of 60, we get 69.
- To clarify this state of affairs, we say that the percentage of uninsured Americans rose by 15 percent. Alternatively, we may say that the percentage of uninsured Americans rose by 9 *percentage points*.

Example

- Suppose you pick peaches and are paid 4 dollars per bushel. One day your boss announces that he is giving you a raise. You will now be paid 5 dollars per bushel.

- Question: what is the percentage increase in your wage?

Example

- Suppose you have a student loan with an annual interest rate of 4 percent. One day your lender announces that the interest rate will soon increase to an annual rate of 5 percent.

- Question: what is the percentage increase in your interest rate?

- Question: what is the percentage point increase in your interest rate?

Example

- Suppose you have a model: $y = \beta_0 + \beta_1 x + u$
- y is measured in %. For example y is cancer rate.
- Interpretation: when x increases by 1 unit, y increases by β_1 *percentage point*.

Interactions between independent variables

- Example: gender -> earnings
- Perhaps gender effects on earnings can be different depending on marital status
- Perhaps married women are more penalized in the labor market
- We assumed “constant” effect of being a female so far
- We want to allow for different effect of being a female depending on the marital status
- How can we do this?

Interactions between independent variables

- $y = \beta_0 + \beta_1 Female + \beta_2 Married + \beta_3 Female * Married + u$
- E(y|Female = 0, married = 0) = β_0
- E(y|Female = 1, married = 0) = β_0+ β_1
- E(y|Female = 0, married = 1) = $\beta_0 + \beta_2$
- E(y|Female = 1, married = 1) = β_0+ $\beta_1 + \beta_2 + \beta_3$

- The effect of being a female for married people:
- The effect of being a female for non-married people:

Interactions between independent variables

- $y = \beta_0 + \beta_1 Female + \beta_2 Married + \beta_3 Female * Married + u$
- E(y|Female = 0, married = 0) = β_0
- E(y|Female = 1, married = 0) = β_0+ β_1
- E(y|Female = 0, married = 1) = $\beta_0 + \beta_2$
- E(y|Female = 1, married = 1) = β_0+ $\beta_1 + \beta_2 + \beta_3$

- The effect of being a female for married people: β_1+ β_3
- The effect of being a female for non-married people: β_1
- The effect of being a female depends on marital status

Interactions between independent variables

- Example: education -> earnings
- Perhaps the effects of education on earnings can be different for different gender
- We assumed "constant" effect of education so far
- We want to allow for different effect of being education depending on gender
- How can we do this?

Interactions between independent variables

- $y = \beta_0 + \beta_1 Educ + \beta_2 Female + \beta_3 Educ * Female + u$
- E(y|Female = 0) = β_0 $+\beta_1 Educ$
- E(y|Female = 1) = β_0+ β_2 $+(\beta_1 + \beta_3)Educ$

- The effect of education can be different for males and females

- The intercept difference is β_2
- Now we also have slope difference, which is β_3

Review for quizzes

- Know how to interpret coefficients in different models (level- level; log-level; level-log, and log-log models)

- Know the difference between percentage point and percentage changes.

- Know how to interpret interaction terms.