Lecture 6 - Economics 113

Todays Topics

- Simple Regression Model
- Simple Regression Assumptions
- Simple Regression Derivation
- Announcements
  - Book chapters
  - Chapter 1 is a good review for today's lecture.
  - Start reading Chapter 2
  - Two books have been placed on reserve at the Science and Engineering Library.

Simple Regression Model

The Example

- Example: Class attendance and grades

$$grade_i = \beta_0 + \beta_1 Attend_i + u_i$$

  - What is β_0 ?
  - What is β_1 ?
  - What is u?
- Suppose we estimate:

$$\widehat{grade}_i = 22.769 + 0.121 Attend_i$$

  - Each additional class attended is associated with a higher grade of 0.121.
  - Is this causal?
  - When does β_1 summarize a causal relationship between *Attend* and *grade*?

Simple Regression Model

The Assumptions

- General framework:
$$y_i = \beta_0 + \beta_1 x_i + u_i$$
- Assumption 1:
$$E(u) = 0$$
- We assume that the unobserved disturbance term is 0 on average.
  - This is not a problem if we have an intercept in the model
- Assumption 2:
$$E(u|x) = E(u)$$
- Combined with assumption 1 this gets us
$$E(u|x) = 0$$
- This means that **given any** x, the value of u we expect will be 0.
  - This is not necessarily realistic.
  - This is the hard assumption to satisfy.

Simple Regression Model

The Example

- Example: Class attendance and grades

$$grade_i = \beta_0 + \beta_1 Attend_i + u_i$$

- The key: *u contains all the variables, other than Attend, that help determine your grade!!!!*
- What are these?
  - Motivation
  - Hours studying
  - Going to section
  - Prior experience
  - Work commitments
  - Drugs
- For example, for A2 to hold, we would need

$$E(u|Attend = 32) = E(u|Attend = 10)$$

  - What does this mean?
  - Is this likely?

Least Squares Regression

The Derivation

- How do we estimate β_0 and β_1?
- Predicted Value:
$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$$
- Residual:
$$\widehat{u}_i = y_i - \widehat{y}_i$$
- Suppose that we choose to minimize the sum of squared errors
$$\min_{\widehat{\beta}_0 \ \widehat{\beta}_1} \sum_{i=1}^{n} \widehat{u}_i^2$$
- Thus:
$$\min_{\widehat{\beta}_0 \ \widehat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2 \tag{1}$$
- Take derivatives!

Least Squares Regression

The Derivation

- Differentiate (1) with respect to $\widehat{\beta}_0$:

$$\sum_{i=1}^{n} -2\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0$$

- Divide by -2, divide by n

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0}$$

- Differentiate (1) with respect to $\widehat{\beta}_1$:

$$\sum_{i=1}^{n} -2x_i\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0$$

- Divide by -2, divide by n

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} x_i\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0}$$

Least Squares Regression

The Derivation

- After lots of algebra, we get:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \widehat{\mu}_x)(y_i - \widehat{\mu}_y)}{\sum_{i=1}^{n} (x_i - \widehat{\mu}_x)^2}$$

- Another way to write this

$$\widehat{\beta}_1 = \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}$$

- To solve for $\widehat{\beta}_0$, take means of $y_i = \beta_0 + \beta_1 x_i + u_i$ and rearrange:

$$\widehat{\beta}_0 = \widehat{\mu}_y - \widehat{\beta}_1 \widehat{\mu}_x$$

- We can also solve for the residuals:

$$\begin{aligned} \widehat{u}_i &= y_i - \widehat{y}_i \\ \widehat{u}_i &= y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i) \end{aligned}$$

Simple Regression Model

The Derivation

- Another way to derive $\widehat{\beta}_0$ and $\widehat{\beta}_1$ is by using our assumptions.
- Assumption 1:

  $$E(u) = 0$$

  Assumption 2:

  $$Cov(x, u) = 0$$

- Assumptions 1 and 2 imply the following:

  $$E(xu) = 0$$

Simple Regression Model

The Derivation

- We have two equations to estimate two unknowns, β_0 and β_1:

$$\begin{aligned} E(u) &= 0 \\ E(xu) &= 0 \end{aligned}$$

- We will be using the "sample analogue" of these equations

$$\frac{1}{n}\sum_{i=1}^{n}(u_i) = 0 \qquad (2)$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i u_i) = 0 \qquad (3)$$

- The goal is to substitute $y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + u_i$ into (2) and (3) and solve for $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

Simple Regression Model

The Derivation

- If we substitute $y_i = \beta_0 + \beta_1 x_i + u_i$ into $\frac{1}{n}\sum_{i=1}^{n}(u_i) = 0$:

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0}$$

- This is equivalent to our first derivative (with respect to $\widehat{\beta}_0$)
- Substitute $y_i = \beta_0 + \beta_1 x_i + u_i$ into $\frac{1}{n}\sum_{i=1}^{n}(x_i u_i)$:

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} x_i\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0}$$

- This is equivalent to the derivative with respect to $\widehat{\beta}_1$

Simple Regression Model

Diagnostic Measures

- SST: Total sum of squares

$$SST = \sum_{i=1}^{n} \left(y_i - \widehat{\mu}_y \right)^2$$

  Measures the total amount of variability in the data

- SSR: Sum of squared residuals

$$SSR = \sum_{i=1}^{n} \left(\widehat{u}_i \right)^2$$

- R-Squared: R^2

$$R^2 = 1 - \frac{SSR}{SST}$$

  Measures the variation "explained" by the model
- Often misinterpreted as "goodness of fit"