Linear regression

Seminar in Criminology, Research and Analysis– Crim 7301
Week 3, 9/8/16
Andrew Wheeler

Saving EMF file

Class Overview

- How to write regression equations
- Linear regression is special
- Other tests besides $\beta = 0$

Words are Messy

We use linear regression to predict Y based on the measures of A and B, as well as their interaction. The measures are logged because they are highly right skewed.

How to write regression equations

All regression equations are just functions

$$\text{Outcome} = f(X_1, X_2, \ldots.)$$

How to write regression equations

Experimental Ideal:

$$\hat{Y}_{\mathrm{c}} = f(T = 0, X_1, X_2, \dots.)$$

$$\hat{Y}_{\mathrm{t}} = f(T = 1, X_1, X_2, \dots.)$$

$$[\hat{Y}_{\mathrm{t}} - \hat{Y}_{\mathrm{c}}] = \text{Treatment Effect}$$

How to write regression equations

Another way to write this, as the *expected value*

$$\mathbb{E}[Y] = \text{The expected value of Y}$$

$$\mathbb{E}[Y|\,T = 1] = \text{The expected value of Y when T equals 1}$$
$$\mathbb{E}[Y|\,T = 0] = \text{The expected value of Y when T equals 0}$$

So

$$\mathbb{E}[Y|\,T = 1]\; - \mathbb{E}[Y|\,T = 0] = \; \text{treatment effect}$$

How to write regression equations

Regression equations can be written as predicting the expected value, given the inputs:

$$\mathbb{E}[Y] = \beta_0 + \beta_1 T + \beta_2 X$$

Works the same later on for generalized linear models:

$$\mathbb{E}[Y] = f(\beta_0 + \beta_1 T + \beta_2 X)$$

Linear regression is special

Simple rule for expected values:

$$\mathbb{E}[a + b] = \mathbb{E}[a] + \mathbb{E}[b]$$

So you could write a linear regression equation as:

$$\mathbb{E}[Y] = \beta_0 + \beta_1(\mathbb{E}[X])$$

Linear regression is special

I'm a reductionist, so linear regression is special.

$$Y_t = y_1 + y_2$$

$$\begin{aligned} \mathbb{E}[y_1] &= \beta_{01} + \beta_{11}X \\ + \ \mathbb{E}[y_2] &= \beta_{02} + \beta_{12}X \end{aligned}$$

$$\overline{\phantom{\mathbb{E}[y_2] = \beta_{02} + \beta_{12}X \quad\quad}}$$

$$\mathbb{E}[Y_t] = (\beta_{01}+\beta_{02}) \ + (\beta_{11}+\beta_{12})X$$

Can *decompose* the effect for burglary vs. non-burglary

Can also test whether $\beta_{11} = \beta_{12}$

Linear regression is special

Decomposing the outcomes (Y) only works for linear regression ☹, but another example.....

$$X = x_1 + x_2$$

$$\mathbb{E}[Y] = \beta_0 + \beta_1(X)$$

Or

$$\mathbb{E}[Y] = \beta_0 + \beta_{11}(x_1) \ + \beta_{12}(x_2)$$

This applies to generalized linear models as well.

Linear regression is special

In general, the variance (denoted $\mathbb{V}$),

$$\mathbb{V}(a - b) = \mathbb{V}(a) + \mathbb{V}(b) - 2 \cdot \mathbb{C}\mathrm{ov}(a, b)$$

To use the central limit theorem, null needs to be equal to zero, so we rewrite null as:

$$\beta_{11} - \beta_{12} = 0$$

Variance = (Standard Error)^2

Linear regression is special

So for example, say

	Coeff.	SE
β_{11}	1.7	0.6
β_{12}	0.8	0.5

$\beta_{11} - \beta_{12} = 1.7 - 0.8 = 0.9$ (The test statistic.)

Assume the covariance is zero, the standard error will then be:

$$\sqrt{(0.6^2 + 0.5^2)} \approx 0.78$$

Homework & Next Weeks Class

Lab Assignment

Linear regression in R, Stata, or SPSS

Readings For Next Week