Mathematics 231

Lecture 12

Liam O'Brien

1

Announcements

- Reading
  - Today   M&M 2.5   148-151
            M&M 2.6   154-159

2

Stuff

- One more word on regression to the mean

- Transformations

- Association versus causation

3

Regression to the Mean

- Investors such as David Dreman and behavioural economists such as Richard Thaler and Werner De Bondt have uncovered strong evidence that regression to the mean… occurs in financial markets. It occurs at both individual and aggregate levels… Using data for the period 1926-1982, Thaler and De Bondt studied the securities of those companies whose prices over a three-year interval had either increased or decreased more than the market average. They found that "extreme returns of stocks listed on the New York Stock Exchange were subsequently followed by significant price movement in the opposite direction". If investors are either unduly optimistic or pessimistic about a particular company's securities, and if that company's fundamentals remain unchanged, then their stance will likely be reversed over time.

- From: *http://www.leithner.com.au/circulars/circular57.htm*

4

Generating Transformed Variables in Stata

- This is most easily done using the command line
- Say you have a variable *x* you want to transform using the log and square root functions:
- Type "generate logx=log(x)" and hit enter
- Type "generate sqrtx=sqrt(x)" and hit enter
- Two new variables called *logx* and *sqrtx* have been created.

5

Association and Causation

- Mother's height and daughter's height.
- A student's SAT score and first-year college GPA.
- Stock price in 1999 and stock price in 2000.
- Length of workweek and per capita GDP.
- Educational attainment and salary.

6

| Causation | Common response | Confounding |
| (a) | (b) | (c) |
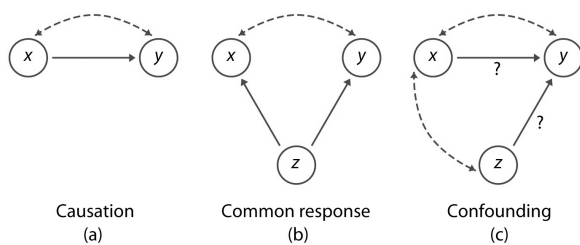
7

Example: Workweek Length and Per Capita GDP

- Data were collected on the length of the workweek and countries' wealth.
- Correlation between length of workweek and per capita GDP was r = -0.51.

  hours of work = 45.8-0.005*(per capita GDP)

- So is it true that if a country has a per capita GDP of $91600, no one would work?
- What does the correlation and regression tell us about the association?

8

2

Establishing Causation

- The best method of establishing causation is to conduct a carefully designed experiment that changes X and controls for the effects of lurking variables.
- This is easier in the natural sciences than in economics.
- So what do we do? Rely on theory. Be very careful about controlling for lurking factors.

9

"Smoking Causes Lung Cancer"

- SURGEON GENERAL'S WARNING: Smoking Causes Lung Cancer, Heart Disease, Emphysema, And May Complicate Pregnancy

- How can they say this?

10

"Smoking Causes Lung Cancer"

- The association is strong.
- The association is consistent across many studies.
- High doses are associated with stronger responses.
- The alleged cause precedes the effect in time.
- The alleged cause is plausible.

11

Confounding

- The relationship between Y and X may be **confounded** with a third variable, Z.
- In that case, Z is called a **confounder**.
- **Simpson's Paradox**: Oral Contraceptive Data and Berkeley Graduate School Admissions Data.

12

Oral Contraceptive Data

- 800 oral contraceptive users, 8.0% have high blood pressure

- 1600 not using oral contraceptive, 8.5% have high blood pressure

- Do oral contraceptives provide a protective effect against high blood pressure?

Age 18-34

	Sample Size	Number with high BP	%with high BP
Use OC	600	36	6
Don't Use OC	400	16	4

Age 35-49

	Sample Size	Number with high BP	%with high BP
Use OC	200	28	14
Don't Use OC	1200	120	10

Admissions Data for 5 Largest Majors at UC Berkeley, Fall 1973

- 8442 men applied, 44% admitted

- 4321 women applied, 35% admitted

- Is this a case of sex discrimination?

Major	Men		Women	
	# applicants	% admitted	# applicants	% admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%

Example: Race and Treatment for Heart Attacks

- Consider a study that examined the relationship between race and heart attack treatment.

- There appears to be an association between race and treatment effectiveness.

- Minority patients tend to show less improvement after treatment.

- Possible confounders?

17