Midterm Review

EC 320: Introduction to Econometrics

Kyle Raze
Fall 2019

Prologue

Housekeeping

Problem Set 2

- Analytical problems due by Tuesday at 17:00 on Canvas
- Computational problems due by Friday at 17:00 on Canvas

Midterm 1 on Wednesday

- **You still have lab this week!**

Problem Set 1 grades posted

Grades

Problem sets and exams

Youssef is the person of contact

- youssefa@uoregon.edu
- PLC 506

Q: Why?

A: To ensure *consistency in grading*.

Course grades

I am the person of contact

- raze@uoregon.edu
- PLC 522

Midterm I: High Concepts

Midterm Topics

Anything from the lectures, labs, or problem sets **is fair game!**

1. Probability Theory
2. Statistics
3. The Fundamental Problem of Econometrics
4. Randomized Control Trials
5. The Logic of Regression
6. Tidy Data and R

Midterm Topics

1. Probability Theory

Random variables

- Sketch out the probability distribution of a random variable.

Expected values

Population variance

Midterm Topics

2. Statistics

Estimators

- Population *v.s.* sample distinction.
- Sample mean, sample variance, sample covariance, sample correlation coefficient.
- Unbiasedness and efficiency.

Hypothesis testing

- Statistical significance.
- t-statistics.
- Two-sided t tests using rule of thumb discussed in class (*i.e.*, compare t to $t_{\text{crit}} = 2$).

Midterm Topics

3. The Fundamental Problem of Econometrics

Causal mechanisms *v.s.* confounding factors

Ideal data on potential outcomes

- Individual *v.s.* average treatment effects.
- Difference-in-means based on observable data.

Selection bias

Midterm Topics

4. Randomized Control Trials

How do RCTs eliminate selection bias?

- Can an RCT fail to eliminate selection bias? How?

Research questions

- Identify outcome variable and treatment variable.

Estimation using experimental data

- Difference-in-means.
- Linear regression.

Midterm Topics

5. The Logic of Regression

Regression models

- Identify outcome variable, treatment variable, and control variables.
- Interpret β coefficients.

Estimation using OLS

- How does OLS adjust treatment effect estimates for confounding factors?

Midterm Topics

5. The Logic of Regression (cont.)

Omitted variables

- When does omitting a variable cause omitted-variable bias?
- Illustrate assumptions with causal diagrams.

Regression tables

- Write down model.
- Interpret coefficients.
- Calculate omitted-variable bias.

Midterm Topics

6. Tidy Data and R

What makes a dataset tidy?

Identify R functions

- What does the function do?
- I will only ask you about functions you've seen in lab or the problem sets.
- **I will not ask you to write code.**

Interpret R output

- Example: Console output of `lm()` or `cor()`.

Midterm Structure

Multiple Choice

- 5 questions
- 6 points per question (30 points total)

True or False

- 5 questions
- 6 points per question (30 points total)
- Brief explanations required for full credit

Free Response

- 5 multi-part questions with varying numbers of points (140 points total)
- Explanations required for full credit

Midterm Protocol

Materials

- Writing utensil
- 3-inch-by-5-inch note card
- Basic or scientific calculator (no graphing or programming capabilities)
- **Nothing else**

Procedure

- **Randomized** seating chart
- 80 minutes from *"you may begin"* to *"pencils down"*
- First 30 minutes: **quiet period** (no questions, no getting up)
- Last 50 minutes: ask lots of questions

Practice

Regression Table

Example: 2016 election

Q: Write down the regression model estimated in column 1.

Outcome: Trump Margin (%)

Explanatory variable	**1**	**2**	**3**
Intercept	**-40.7**	42	-65.7
	(1.95)	(1.49)	(2.99)
White (%)	**0.91**		1.05
	(0.024)		(0.027)
Poverty (%)		-0.647	0.883
		(0.087)	(0.081)

A: $\text{Trump}_i = \beta_0 + \beta_1 \text{White}_i + u_i$.

Regression Table

Example: 2016 election

Q: Write down the regression model estimated in column 2.

Outcome: Trump Margin (%)

Explanatory variable	**1**	**2**	**3**
Intercept	-40.7	**42**	-65.7
	(1.95)	**(1.49)**	(2.99)
White (%)	0.91		1.05
	(0.024)		(0.027)
Poverty (%)		**-0.647**	0.883
		(0.087)	(0.081)

A: $\text{Trump}_i = \beta_0 + \beta_1 \text{Poverty}_i + u_i$.

Regression Table

Example: 2016 election

Q: Write down the regression model estimated in column 3.

Outcome: Trump Margin (%)

Explanatory variable	**1**	**2**	**3**
Intercept	-40.7	42	**-65.7**
	(1.95)	(1.49)	**(2.99)**
White (%)	0.91		**1.05**
	(0.024)		**(0.027)**
Poverty (%)		-0.647	**0.883**
		(0.087)	**(0.081)**

A: $\text{Trump}_i = \beta_0 + \beta_1 \text{White}_i + \beta_2 \text{Poverty}_i + u_i$.

Regression Table

Example: 2016 election

Q: How do we interpret the coefficient on poverty in column 2?

Outcome: Trump Margin (%)

Explanatory variable	**1**	**2**	**3**
Intercept	-40.7	**42**	-65.7
	(1.95)	**(1.49)**	(2.99)
White (%)	0.91		1.05
	(0.024)		(0.027)
Poverty (%)		**-0.647**	0.883
		(0.087)	(0.081)

A: *"A 1 p.p. increase in the poverty rate is associated with a 0.647 p.p. decrease in Trump's margin of victory."*

Regression Table

Example: 2016 election

Q: Is the poverty coefficient in column 2 statistically significant?

Outcome: Trump Margin (%)

Explanatory variable	**1**	**2**	**3**
Intercept	-40.7	**42**	-65.7
	(1.95)	**(1.49)**	(2.99)
White (%)	0.91		1.05
	(0.024)		(0.027)
Poverty (%)		**-0.647**	0.883
		(0.087)	(0.081)

A: *"Yes. The estimate is more than twice its standard error, which corresponds to* $t > 2$*."*

Regression Table

Example: 2016 election

Q: How do we interpret the intercept in column 2?

Outcome: Trump Margin (%)

Explanatory variable	**1**	**2**	**3**
Intercept	-40.7	**42**	-65.7
	(1.95)	**(1.49)**	(2.99)
White (%)	0.91		1.05
	(0.024)		(0.027)
Poverty (%)		**-0.647**	0.883
		(0.087)	(0.081)

A: *"Counties with poverty rates of zero had a Trump margin of 42 percent, on average."*

Regression Table

Example: 2016 election

Q: Does omitting $\mathbf{White}_i$ bias the estimator of the $\mathbf{Poverty}_i$ coefficient?

Outcome: Trump Margin (%)

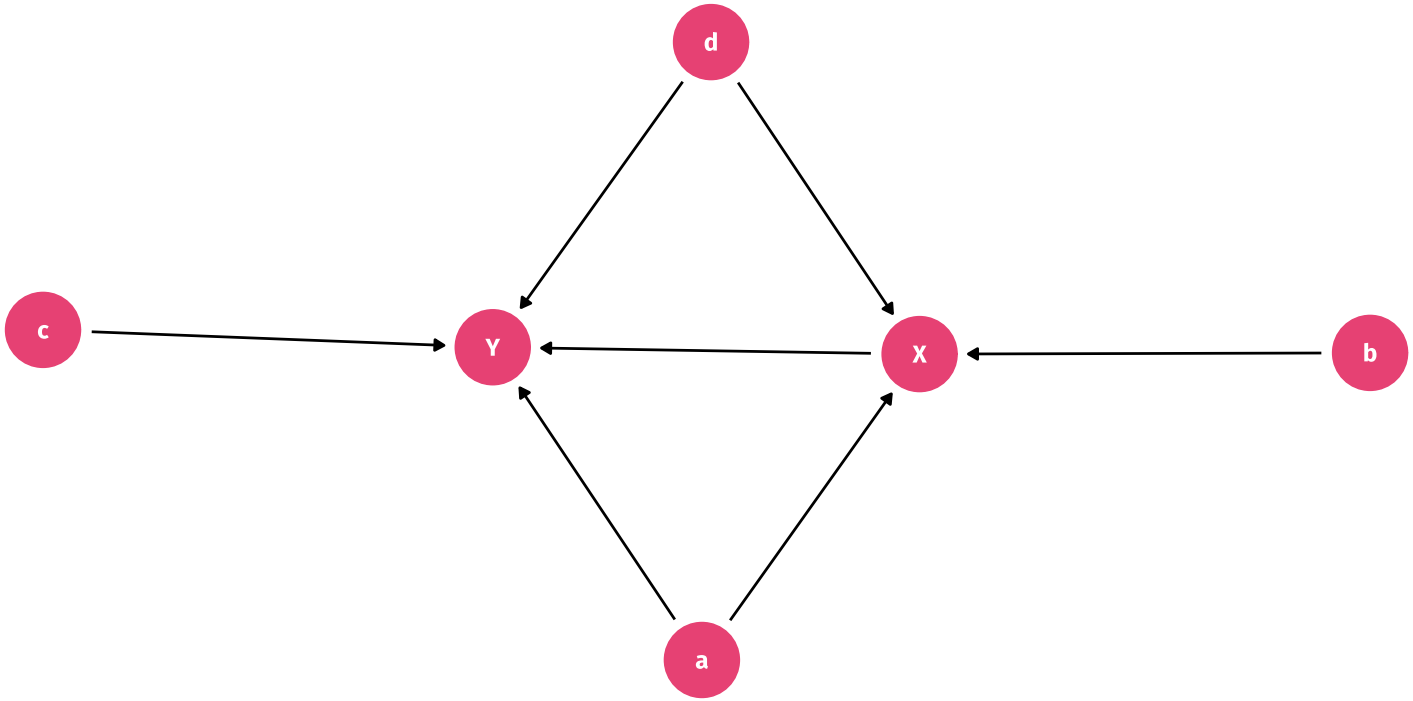
Explanatory variable	**1**	**2**	**3**
Intercept	-40.7	**42**	**-65.7**
	(1.95)	**(1.49)**	**(2.99)**
White (%)	0.91		**1.05**
	(0.024)		**(0.027)**
Poverty (%)		**-0.647**	**0.883**
		(0.087)	**(0.081)**

A: *"Omitting $\mathbf{White}_i$ appears to cause negative omitted-variable bias. The size of the bias is $\beta^{\text{Short}}_{\text{Poverty}} - \beta^{\text{Long}}_{\text{Poverty}} = -0.647 - 0.883 = -1.53$."*

Omitted Variables

Goal: Isolate the effect of **X** on **Y**.

Q: Which variables, if omitted, would cause omitted-variable bias?



A: a and d.

Regression Output

Suppose that a government agency wants to identify the causal effect of money on problems. To isolate the causal effect, the agency conducts an RCT that gives 10,000 dollars to people in the randomly selected treatment group and 0 dollars to people in the randomly selected control group.

```
reg_rct ← lm(problems ~ money, data = rct_data)
tidy(reg_rct)
```

```
#> # A tibble: 2 x 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    17.3       4.40      3.93 0.0000890
#> 2 money          -6.29      6.22     -1.01 0.313
```

Q_1: What is the regression model the agency estimated?

Q_2: What is the estimated treatment effect?

Regression Output

```
reg_rct ← lm(problems ~ money, data = rct_data)
tidy(reg_rct)
```

```
#> # A tibble: 2 x 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    17.3       4.40      3.93 0.0000890
#> 2 money          -6.29      6.22     -1.01 0.313
```

Q_3: Is the estimated treatment effect statistically distinguishable from zero?

Q_4: What is the average number of problems for those in the control group?

Q_5: What is the average number of problems for those in the treatment group?

Q_6: Does money affect problems?