Midterm Review

EC 320: Introduction to Econometrics

Amna Javed
Winter 2020

Prologue

Housekeeping

Midterm 1 on Wednesday

- **No labs this week!**
- Sichao extra office hours Tuesday 11-12pm
- My extra office hours Monday 2-3pm

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

Midterm Topics

1. Probability Theory

Random variables

Expected values

Population variance

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 (*i.e.*, compare t to $t_{\text{crit}} = 2$).

3. The Fundamental Problem of Econometrics

Causal mechanisms *v.s.* confounding factors

Ideal data on potential outcomes

- Individual *v.s.* average treatment effects.

Selection bias

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

- Linear regression.
- Know mean of control and treatment group from regression (see PS 2).

5. The Logic of Regression

Regression models and tables

- Identify outcome variable, treatment variable, constants.
- Interpret β coefficients.
- Write regressions.
- Sample versus population regression.
- Make predictions.

6. Tidy Data and R

I will not ask you to write code.

Interpret R output

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

Midterm Structure

Multiple Choice

- 6 questions
- 2 points per question (12 points total)

Short Questions

- 3 questions (10 points total)

Analytical Questions

- 3 multi-part questions with varying numbers of points (43 points total)

Midterm Protocol

Materials

- Writing utensil
- Basic or scientific calculator (no graphing or programming capabilities)
- **Nothing else**

Procedure

- 80 minutes from *"you may begin"* to *"pencils down"*
- Keep bags to side of the room or front (makes answering questions easier)
- Minimize bathroom breaks (movement can be disturbing to others)

Practice

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

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

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

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."*

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* $t > 2$ *."*

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."*

Example: 2016 election

Q: Predict Trump's victory margin if the poverty rate in a county is 50 percent.

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: $\hat{\text{Trump}}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{Poverty}_i = 42 + (-0.647)(50) = 9.65$

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?