Lecture 13

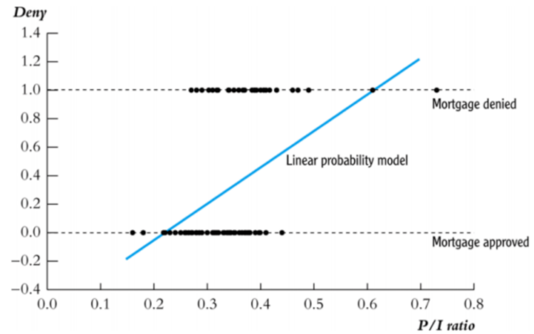
Annoucements

- PS3 due today
- Midterm is on 003/18/2020 (Wednesday)

The linear probability model: shortcomings

- Predicted probability can be above 1 or below 0!

Example: linear probability model, HMDA data
Mortgage denial v. ratio of debt payments to income (P/I ratio) in a subset of the HMDA data set (*n* = 127)



Outline

- Critique of Multiple Regression
  - How to check threats to internal validity and external validity of multiple regression models

- Randomized Control Trial (RCT)
  - Balance test
  - Result table
  - Attrition

Assessing Studies Based on Multiple Regression

- Let's step back and take a broader look at regression:
- Is there a systematic way to assess (critique) regression studies? We know the strengths – but what are the pitfalls of multiple regression?
- When we put all this together, what have we learned about the effect on test scores of class size reduction?

Is there a systematic way to assess regression studies?

- Multiple regression has some key virtues:
  - It provides an estimate of the effect on *Y* of arbitrary changes ΔX.
  - It resolves the problem of omitted variable bias, if an omitted variable can be measured and included.
  - It can handle nonlinear relations (effects that vary with the *X*'s)

- Still, OLS might yield a *biased* estimator of the true *causal* effect – it might not yield "valid" inferences…

A Framework for Assessing Statistical Studies: Internal and External Validity

- ***Internal validity***: the statistical inferences about causal effects are valid for the population being studied.

- ***External validity***: the statistical inferences can be generalized from the population and setting studied to other populations and settings, where the "setting" refers to the legal, policy, and physical environment and related salient features.

Threats to External Validity of Multiple Regression Studies

- How far can we generalize class size results from California school districts?
- Differences in populations
  - California in 2005?
  - Massachusetts in 2005?
  - Mexico in 2005?
- Differences in settings
  - different legal requirements concerning special education
  - different treatment of bilingual education
  - differences in teacher characteristics

Threats to Internal Validity of Multiple Regression Analysis

- ***Internal validity***: the statistical inferences about causal effects are valid for the population being studied.
-
- *Five threats to the internal validity of regression studies:*
  1. Omitted variable bias
  2. Wrong functional form
  3. Errors-in-variables bias
  4. Sample selection bias
  5. Simultaneous causality bias

1. Omitted variable bias

- Omitted variable bias arises if an omitted variable is ***both***:

  (i) a determinant of *Y* and

  (ii) correlated with at least one included regressor.

- We first discussed omitted variable bias in regression with a single *X*, but OV bias will arise when there are multiple *X*'s as well, if the omitted variable satisfies conditions (i) and (ii) above.

Potential solutions to omitted variable bias

- If the variable can be measured, include it as an additional regressor in multiple regression

2. Wrong functional form

- Arises if the functional form is incorrect – for example, an interaction term is incorrectly omitted; then inferences on causal effects will be biased.

- **Potential solutions to functional form misspecification**
- Continuous dependent variable: use the "appropriate" nonlinear specifications in X (logarithms, interactions, etc.)

3. Errors-in-variables bias

- So far we have assumed that *X* is measured without error.
- In reality, economic data often have measurement error
  - Data entry errors in administrative data
  - Recollection errors in surveys (when did you start your current job?)
  - Ambiguous questions problems (what was your income last year?)
  - Intentionally false response problems with surveys (What is the current value of your financial assets? How often do you drink and drive?)
- In general, measurement error in a regressor results in "errors-in-variables" bias.

Potential solutions to errors-in-variables bias

- Obtain better data.
- Develop a specific model of the measurement error process.
- This is only possible if a lot is known about the nature of the measurement error – for example a subsample of the data are cross-checked using administrative records and the discrepancies are analyzed and modeled.

4. Sample selection bias

- So far we have assumed simple random sampling of the population. In some cases, simple random sampling is thwarted because the sample, in effect, “selects itself.”

- *Sample selection bias* arises when a selection process:

  (i) influences the availability of data and

  (ii) that process is related to the dependent variable.

Example #1: Mutual funds

- Do some mutual funds consistently beat other funds and the market?
- Empirical strategy:
  - Sampling scheme: simple random sampling of mutual funds available to the public on a given date.
  - Is there sample selection bias?

Example #1: Mutual funds

- Do some mutual funds consistently beat other funds and the market?
- Empirical strategy:
  - Sampling scheme: simple random sampling of mutual funds available to the public on a given date.
  - Is there sample selection bias?

- Yes, the most poorly performing funds are omitted from the data set because they went out of business or were merged into other funds. This will overstate the mean return of all funds.

Example #2: returns to education

- What is the return to an additional year of education?
- Empirical strategy:
  - Sampling scheme: simple random sample of employed college grads (employed, so we have wage data)
  - Data: earnings and years of education
  - Estimator: regress ln(*earnings*) on *years_education*
  - Ignore issues of omitted variable bias and measurement error – is there sample selection bias?

Example #2: returns to education

- What is the return to an additional year of education?
- Empirical strategy:
  - Sampling scheme: simple random sample of employed college grads (employed, so we have wage data)
  - Data: earnings and years of education
  - Estimator: regress ln(*earnings*) on *years_education*
  - Ignore issues of omitted variable bias and measurement error – is there sample selection bias?
- Yes, “employed” are already selected samples.

Potential solutions to sample selection bias

- Collect the sample in a way that avoids sample selection.
  - *Mutual funds example*: change the sample population from those available at the *end* of the ten-year period, to those available at the *beginning* of the period (include failed funds)
  - *Returns to education example*: sample college graduates, not workers (include the unemployed)

5. Simultaneous causality bias

- So far we have assumed that *X* causes *Y*.
- What if *Y* causes *X*, too?

- *Example*: Class size effect
- Low *STR* results in better test scores
- But suppose districts with low test scores are given extra resources: as a result of a political process they also have low *STR*
- What does this mean for a regression of *TestScore* on *STR*?
- The estimate is likely to be biased.

Potential solutions to simultaneous causality bias

- Randomized controlled experiment. Because X_i is chosen at random by the experimenter, there is no feedback from the outcome variable to Y_i.
- Develop and estimate a complete model of both directions of causality. This is the idea behind many large macro models (e.g. Federal Reserve Bank-US). *This is extremely difficult in practice*.

Randomized control trial (RCT)

RCT Introduction video

- Randomized evaluations & the power of evidence by health economist Amy Finkelstein
- https://www.youtube.com/watch?v=N8rD844McrA

Randomized Control Trial (RCT)

- Randomized control trial is also called random assignment or simply randomization or experiment
- They really mean the same thing: randomly assign people to treatment and control (comparison) group
- It is the best method to eliminate OVB/selection bias
    - If health insurance is randomly assigned, then People in treatment and control groups are not only similar in observables but also in unobservables
    - We are comparing apples to apples-No selection bias.
- The downside: too costly; ethical concerns.

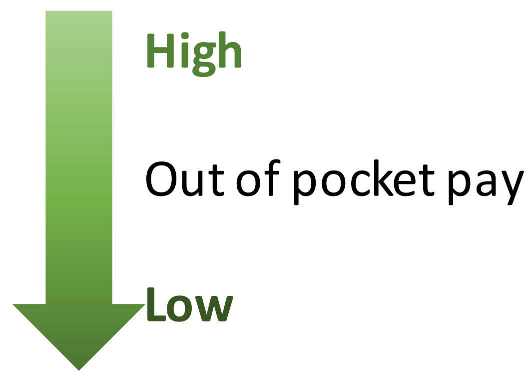
Example: RAND Health Insurance Experiment (HIE)

- Is it a good thing to make health insurance free?
- What unintended consequences might happen?
  - Cons: Moral hazard (cost)
  - Pros: Improved health? (benefit)
- RAND HIE
  - Families were **randomly assigned** to one of four health insurance plans
  - Don't pay premium, but out-of-pocket copay varies (you may think these different plans have different prices)
- Objective:
  - How does cost-sharing affect usage of health care?
  - Does increased usage of healthcare lead to better health outcomes?

Treatment and control group

- Control Group: Catastrophic Plan-(mimic no insurance)

- Treatment Group 1: Deductible
- Treatment Group 2: Coinsurance
- Treatment Group 3: Free



How do we evaluate randomized control trial?

- Is this truly a randomized experiment?
  - The meaning of random: Treatment and control group should look similar in terms of demographic characteristics and outcome variable before the experiment
  - How do you test it?
    - Show a table containing the means of demographic variables for treatment and control group before the experiment
    - Test if their difference is statistically different from 0
    - This is called a **balance test (balance table)**
    - The function of balance table is to test whether there is omitted variable/selection bias problem before the experiment

RCT: Balance table and Result Table

- Balance table: to test whether **demographic variables** are similar in treatment and control groups **before the experiment**

- Result table: to see if there is any difference in the the **outcome variable after the experiment**

Exercise:

- Is this a balance table or result table?
- Is the age difference between coinsurance and catastrophic (control group) economically big?
- Is the difference statistically significant?

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
		A. Demographic characteristics			
Female	.560	−.023 (.016)	−.025 (.015)	−.038 (.015)	−.030 (.013)
Nonwhite	.172	−.019 (.027)	−.027 (.025)	−.028 (.025)	−.025 (.022)
Age	32.4 [12.9]	.56 (.68)	.97 (.65)	.43 (.61)	.64 (.54)
Education	12.1 [2.9]	−.16 (.19)	−.06 (.19)	−.26 (.18)	−.17 (.16)
Family income	31,603 [18,148]	−2,104 (1,384)	970 (1,389)	−976 (1,345)	−654 (1,181)
Hospitalized last year	.115	.004 (.016)	−.002 (.015)	.001 (.015)	.001 (.013)
		B. Baseline health variables			
General health index	70.9 [14.9]	−1.44 (.95)	.21 (.92)	−1.31 (.87)	−.93 (.77)
Cholesterol (mg/dl)	207 [40]	−1.42 (2.99)	−1.93 (2.76)	−5.25 (2.70)	−3.19 (2.29)
Systolic blood pressure (mm Hg)	122 [17]	2.32 (1.15)	.91 (1.08)	1.12 (1.01)	1.39 (.90)
Mental health index	73.8 [14.3]	−.12 (.82)	1.19 (.81)	.89 (.77)	.71 (.68)
Number enrolled	759	881	1,022	1,295	3,198

Before the experiment:

"Balance test"-a test for randomization

-0.023: $(\text{Mean}_{\text{treat}}-\text{Mean}_{\text{control}})$=difference in proportion of female between treatment and control

0.016:standard error

Example: conduct hypothesis testing at α=5%

H_0: diff=0

$t^*=(-0.023-0)/0.016=1.43$
$t_{\alpha/2}$ for n=500 and α=5% is 1.97
$t^* < t_{\alpha/2}$. Fail to reject H_0.

Conclusion: Treatment and control groups have similar proportion of females

Randomization is good (for this variable)!

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
		A. Demographic characteristics			
Female	.560	−.023 (.016)	−.025 (.015)	−.038 (.015)	−.030 (.013)
Nonwhite	.172	−.019 (.027)	−.027 (.025)	−.028 (.025)	−.025 (.022)
Age	32.4 [12.9]	.56 (.68)	.97 (.65)	.43 (.61)	.64 (.54)
Education	12.1 [2.9]	−.16 (.19)	−.06 (.19)	−.26 (.18)	−.17 (.16)
Family income	31,603 [18,148]	−2,104 (1,384)	970 (1,389)	−976 (1,345)	−654 (1,181)
Hospitalized last year	.115	.004 (.016)	−.002 (.015)	.001 (.015)	.001 (.013)
		B. Baseline health variables			
General health index	70.9 [14.9]	−1.44 (.95)	.21 (.92)	−1.31 (.87)	−.93 (.77)
Cholesterol (mg/dl)	207 [40]	−1.42 (2.99)	−1.93 (2.76)	−5.25 (2.70)	−3.19 (2.29)
Systolic blood pressure (mm Hg)	122 [17]	2.32 (1.15)	.91 (1.08)	1.12 (1.01)	1.39 (.90)
Mental health index	73.8 [14.3]	−.12 (.82)	1.19 (.81)	.89 (.77)	.71 (.68)
Number enrolled	759	881	1,022	1,295	3,198

Exercise

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
		A. Demographic characteristics			
Female	.560	−.023	−.025	−.038	−.030
		(.016)	(.015)	(.015)	(.013)
Nonwhite	.172	−.019	−.027	−.028	−.025
		(.027)	(.025)	(.025)	(.022)
Age	32.4	.56	.97	.43	.64
	[12.9]	(.68)	(.65)	(.61)	(.54)
Education	12.1	−.16	−.06	−.26	−.17
	[2.9]	(.19)	(.19)	(.18)	(.16)

1. Which group is the control group? How many treatment groups do we have?
2. Look at the coefficient in the red circle. What does it mean?
3. What is the standard error for that coefficient?
4. To conduct statistical inference for that coefficient. What is the null hypothesis? What is the t-stat?
5. Use d.f.=500 and a=0.05 to conclude whether you reject null hypothesis
6. **What does reject the null means in this example?**
7. **Is the result what you would like to see (the purpose of randomization)?**

Results of RCT

- If the randomization is successful, a simple mean comparison (naïve comparison) will provide the causal impact of treatment on the output
- In regression framework, the naive mean comparison is provided by simple regression:
- Health=β_0+β_1Insured+ε
- $\widehat{\beta_1}$ are our main results. We have different measures for health, so we end up with different $\widehat{\beta_1}$

- What are the two main sets of results?
- Does giving people free insurance increase medical care usage?
- Does it lead to better health?

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
		A. Health-care use			
Face-to-face visits	2.78	.19	.48	1.66	.90
	[5.50]	(.25)	(.24)	(.25)	(.20)
Outpatient expenses	248	42	60	169	101
	[488]	(21)	(21)	(20)	(17)
Hospital admissions	.099	.016	.002	.029	.017
	[.379]	(.011)	(.011)	(.010)	(.009)
Inpatient expenses	388	72	93	116	97
	[2,308]	(69)	(73)	(60)	(53)
Total expenses	636	114	152	285	198
	[2,535]	(79)	(85)	(72)	(63)
		B. Health outcomes			
General health index	68.5	−.87	.61	−.78	−.36
	[15.9]	(.96)	(.90)	(.87)	(.77)
Cholesterol (mg/dl)	203	.69	−2.31	−1.83	−1.32
	[42]	(2.57)	(2.47)	(2.39)	(2.08)
Systolic blood pressure (mm Hg)	122	1.17	−1.39	−.52	−.36
	[19]	(1.06)	(.99)	(.93)	(.85)
Mental health index	75.5	.45	1.07	.43	.64
	[14.8]	(.91)	(.87)	(.83)	(.75)
Number enrolled	759	881	1,022	1,295	3,198

After the experiment:
Result of RCT

There are two types of main results

- usage of medical care
- health status

What does the coefficient 0.19 mean?

- The average face-to-face visits for the deductible group is 0.19 unit higher than the average of the control group (catastrophic plan)

- Can you conduct statistical inference based on its standard error? Use d.f.=500 and α=0.05

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
A. Health-care use					
Face-to-face visits	2.78	$\boxed{.19}$	.48	1.66	.90
	[5.50]	(.25)	(.24)	(.25)	(.20)
Outpatient expenses	248	42	60	169	101
	[488]	(21)	(21)	(20)	(17)
Hospital admissions	.099	.016	.002	.029	.017
	[.379]	(.011)	(.011)	(.010)	(.009)
Inpatient expenses	388	72	93	116	97
	[2,308]	(69)	(73)	(60)	(53)
Total expenses	636	114	152	285	198
	[2,535]	(79)	(85)	(72)	(63)
B. Health outcomes					
General health index	68.5	−.87	.61	−.78	−.36
	[15.9]	(.96)	(.90)	(.87)	(.77)
Cholesterol (mg/dl)	203	.69	−2.31	−1.83	−1.32
	[42]	(2.57)	(2.47)	(2.39)	(2.08)
Systolic blood pressure (mm Hg)	122	1.17	−1.39	−.52	−.36
	[19]	(1.06)	(.99)	(.93)	(.85)
Mental health index	75.5	.45	1.07	.43	.64
	[14.8]	(.91)	(.87)	(.83)	(.75)
Number enrolled	759	881	1,022	1,295	3,198

Exercise

1. Explain what the coefficient in the red circle means

2. Conduct statistical inference for this coef.

3. What is your conclusion? Did you identify a statistically significant difference between the treatment and control?

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
		A. Health-care use			
Face-to-face visits	2.78	.19	.48	1.66	.90
	[5.50]	(.25)	(.24)	(.25)	(.20)
Outpatient expenses	248	42	60	169	101
	[488]	(21)	(21)	(20)	(17)
Hospital admissions	.099	.016	.002	.029	.017
	[.379]	(.011)	(.011)	(.010)	(.009)
Inpatient expenses	388	72	93	116	97
	[2,308]	(69)	(73)	(60)	(53)
Total expenses	636	114	152	285	198
	[2,535]	(79)	(85)	(72)	(63)

Evaluate RCT-attrition problem

- Attrition: participants drop out of the program before experiment is finished
- Attrition is a problem only when attribution is not balanced across the treatment and control group
- It is not a problem if attribution is random
- e.g. If more poor people in the **control** group drop out of the experiment than treatment group, health status for control group after the experiment will appear to be higher than it is without attrition (Because poor people tend to have poor health, and these poor people drop out in the control)
- **Treatment effect=health of treatment-health of control** (after the experiment)
  - Health of control increases, therefore treatment effect decreases
- Downward bias the effect by inflating health status of the control group

Exercise:

- In the previous example, what happens to the treatment effect if more educated (with a college degree) people in the treatment group drop out than the control? (Maybe they have better choices than free insurance). Will this attribution problem cause biased estimate? Upward or downward bias? Explain.

Review for quizzes

- Know how to check for threats to internal validity and external validity of multiple regression models

- How to read balance table and result table of RCT
- Be able to perform statistical inference for the balance table and result table
- Understand the attrition problem in RCT