Lecture 1: Causality and Experimental Ideal

Lingguo Cheng

2025 Fall

Business School, Nanjing University

About “Intermediate Econometrics”

- Literally, it should be located between the rudimentary and the advanced, but where?
- I feel a little bit baffled......
- I thought it over for a long while and eventually decided to emphasize two points:
  - Make it easier for students to transition to the advanced course.
  - Aims to let students know how to apply econometrics and how to write a paper (for your dissertation thesis, for summer camps)
- The middle ground will be “Micro-econometrics “ or “Empirical microeconomics ”

What is “Empirical Microeconomics”

- Empirical?
  - Versus the theoretical way
  - Does not repel economic theory:
    - In theory, there are multiple possibilities; which one is true? (education human capital, or signaling theory)
    - I know some issues might be important in real life, but how important are they?
    - Interesting, puzzling, stylized facts can challenge existing theories and push the theorists to pose new explanations
- Microeconomics?
  - The subject of study is not necessarily a human individual; it can be households, firms, states, or countries (Acemoglu, Johnson, and Robinson, 2001).
  - The subject should have a motivation, some constraints, a process of decision making, and eventually make choices (predictable consequences) – Behavioral implications.

FAQs for an Empirical microeconomic Project

FAQ I: What is the causal relationship of interest?

- As if in a randomized ~~clinical~~ controlled trial RCT
- Causality vs. correlation
  - Correlation does not mean causality 相关≠因果
    - When the Rooster crows, the sun rises. But the rooster does not cause the sun to rise.
  - No correlation does not necessarily mean no causality. 无关≠无因果
    - Sometimes, causality exists even when no discernible correlation exists at first blush. (Boat's direction, rudder, wind)
  - Correlation can be misleading, and skeptically can almost never capture the "true" causality with observational data
    - Omitted variable, reverse causality, measurement error, sample selection bias

Figure 1. No correlation doesn't mean no causality. Artwork by Seth Hahne © 2020.

Source: Cunningham(2021) P7

- Questions of cause and effect is always fascinating.
  - Class size vs. Test scores;
  - Schooling vs. Revenue
  - Colonial institutions vs. Economic growth (Acemoglu, Johnson, and Robinson, 2001)
- Counterfactual causality: what if, or if-then
- Economic explanation vs. prediction
- Big data vs. causality

Causality vs. machine learning (a subfield of AI)

- Predictive machine learning
  - Decision trees, random forests, lasso/ridge regression, boosting, support vector machines, and neural networks.
- Causal machine learning algorithms can learn in a data-driven way in which covariates affect the treatment and the outcome simultaneously to ensure that we compare "apples with apples" (in the big data context, the number of covariates tends to be very large, may be larger than the number of observations)
- Detection of important effect heterogeneities across subgroups in the data.
  - Which covariates drive economically significantly the size and heterogeneity of the effect: consumer segmentation
- Optimal policy learning
  - The optimal assignment of a treatment to specific subgroups to target groups in which the effect is largest
- Reinforcement learning:
  - Algorithms for optimally designing repeated experiments, aiming at learning over time which treatment is most effective. E.g., which advertisements generate the most revenue?

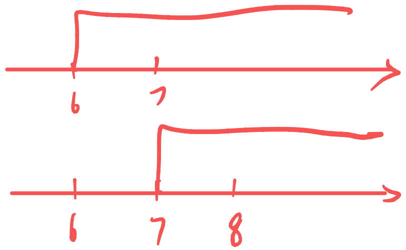
FAQ II: The experiment that could ideally be used to capture the causal effect of interest

- **Ideal experiments are most often hypothetical**. Picture yourself as a researcher with no budget constraint and no Human Subjects Committee policing your inquiry for social correctness.
- If you can’t devise an experiment that answers your question in a world where anything goes, then the odds of generating useful results with a modest budget and non-experimental survey data seem slim.
- The description of an ideal experiment also helps you formulate causal questions precisely.
- Milgram (1963): Seeking to understand the response to authority, he convinced experimental subjects to administer painful electric shocks to pitifully protesting victims.

Fundamentally Unidentified Questions (FUIQ)

- Some questions seem to be FUIQs, but it turns out that they are not
  - Labor market discrimination: The causal effect of race or gender?
  - Bertrand and Mullainathan (2004) compared employers' responses to résumés with blacker-sounding and whiter-sounding first names, such as *Lakisha* and *Emily*.
  - Switching ethnicity to gain the benefits of more quotas of giving birth or enrolling in colleges

An example of FUIQ



- Whether children do better in school by virtue of having started school a little older?
- More specifically, the effect of start age on learning in elementary school
- Randomly select some kids to start first grade at age 7, while others start at age 6, as is still typical. We are interested in whether those held back learn more in school, as evidenced by their elementary school *test scores*. To be concrete, let's look at test scores in first grade.
  - **Maturation effect** (vs test scores that 6-year-old cohort got when they were in 1^{st} score)
  - **Holding age constant instead of grade**. Suppose we wait to test those who started at age 6 until second grade and test those who started at age 7 in first grade, so that everybody is tested at age 7. **Time-in-school effects** as long as kids are still in school. (FUQ)
- Instead, investigating pure start-age effects on adult outcomes, such as earnings or highest grade completed (as in Black, Devereux, and Salvanes, 2011, RESta) is feasible.

FAQ III: What is your identification strategy?

- Angrist and Krueger (1999) “the manner in which a researcher uses observational data (i.e., data not generated by a randomized trial) to approximate a real experiment”
  - Two types of data: experimental data and non-experimental or observational data

- Trygve Haavelmo (1944, p. 14): Experiments may be grouped into two different classes,

(1) Experiments that we would like to make to see if certain real economic phenomena—when artificially isolated from “other influences”—would verify certain hypotheses

  - **Field experiments**
  - **Laboratory experiments**

(2) The stream of experiments that Nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers. (**Natural or Quasi-experiment**)

FAQ IV: What is your mode of statistical inference? (Rubin, 1991)

- The population to be studied, the sample to be used, and the assumptions made when constructing standard errors
  - Simple random sampling
  - Stratified sampling
  - Grouped or clustered data
- Usually technical, tedious, but crucial: the ultimate success of even a well-conceived and conceptually exciting project turns on the details of statistical inference.
- Keisuke Hirano

T-stat looks too good
Try clustered standard errors—
Significance gone

- *THIS IS LIFE !!!*
- Reaching after the stars / asterisks.

II. The Selection Problem

- Question: Do hospitals make people healthier?
- National Health Interview Survey (NHIS) (2005)

Group	Sample Size	Mean Health Status	Std. Error
Hospital	7,774	3.21	0.014
No hospital	90,049	3.93	0.003

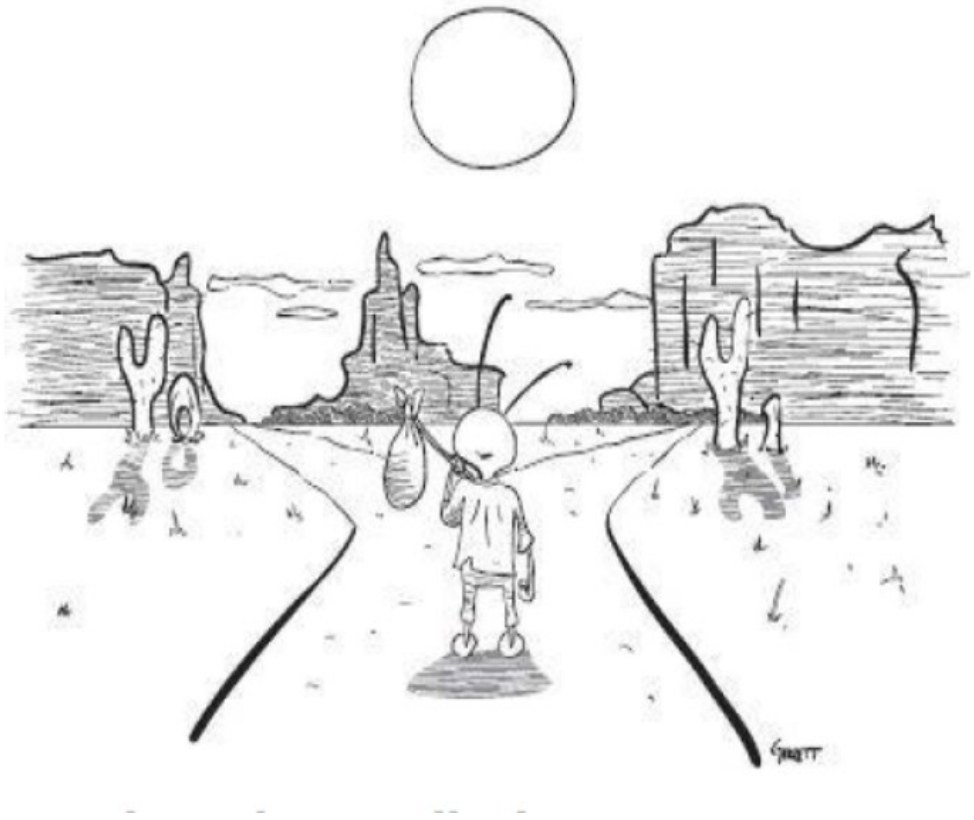
- Taken at face value, this result suggests that going to the hospital makes people sicker (-0.72, t-stat=58.9).
- What a surprise !!
  - People who go to the hospital are probably **less healthy to begin with.**
  - **Selection bias? Survival bias?** 贫困地区老人更健康？
  - **Improper benchmark!** They may well be better off than they otherwise would have been.

Rubin Causal Model (Rubin, 1974, 1977; Holland, 1986)

- Hospital treatment, described by a binary random variable, $D_i = \{0, 1\}$. The outcome of interest, a measure of health status, is denoted by Y_i. The question is whether Y_i is affected by hospital care.
- Assume we can imagine what might have happened to someone who went to the hospital if that person had not gone, and vice versa. Hence, for any individual, there are two potential health variables:

$$\textit{Potential outcome} = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

- Y_{0i}, is the health status of an individual had he not gone to the hospital, irrespective of whether he actually went, while Y_{1i} is the individual's health status if he goes.



The Road Not Taken

Robert Frost

Two roads diverged in a yellow wood,

And sorry I could not travel both

And be one traveler, long I stood

And looked down one as far as I could

To where it bent in the undergrowth;

.....

Two roads diverged in a wood, and I—

I took the one less traveled by,

And that has made all the difference.

- The observed outcome, Y_i, can be written in terms of potential outcomes as

$$Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$
$$= Y_{0i} + (Y_{1i} - Y_{0i})D_i.$$

$Y_i = Y_{0i} + \underline{(Y_{1i} - Y_{0i})} D_i$

Casual Effect

- This notation is useful because $Y_{1i} - Y_{0i}$ *is the* ***causal effect*** *of hospitalization* for an individual.
- In theory, $Y_{1i} - Y_{0i}$ *can be* different for different i , that is, the causal effect can be heterogenous 异质的
- The question is, we never see both potential outcomes for anyone individual at the same time: *"The fundamental problem of causal inference" (Holland, 1986)*
  - If Y_{1i} is observed, then Y_{0i} will be counterfactual.
  - Vice versa
- We can only have in hand the average health status of those who were and who were not hospitalized---Then how can we get the $E(Y_{1i} - Y_{0i})$

Naive Comparison and Selection Bias

- A naive comparison of averages by hospitalization status

$$\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{\text{Observed difference in average health}} = \underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=1]}_{\text{Average treatment effect on the treated}} + \underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{Selection bias}}.$$

ATT

初始健康状态差异

- The average causal effect of hospitalization on those who were hospitalized(**The Average Treatment Effect on the Treated, ATT**)

$$E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=1]=E[Y_{1i}-Y_{0i}|D_i=1]$$

$ATU = E[Y_{1i} - Y_{0i} | D_i = 0]$.
(untreated)

$ATE = P(D_i=1)\, ATT + P(D_i=0)\, ATU$.

- The nasty thing is the **selection bias term**
  - The goal of most empirical economic research is to overcome selection bias.
  - Typically, $E(Y_{0i}|D_i=1)< E(Y_{0i}|D_i=0)$, selection bias is usually negative, and the naïve comparison will underestimate the ATT

III. Random Assignment Solves the Selection Problem

此时 ATT = ATU = ATE

- Random assignment makes D_i independent of potential outcomes. Hence,

随机实验

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$
$$= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1],$$

- This simplifies further to

$$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}|D_i = 1]$$
$$= E[Y_{1i} - Y_{0i}].$$

- The effect of randomly assigned hospitalization on the hospitalized (ATT) is the same as the effect of hospitalization on a randomly chosen patient (ATE).

- The main thing, however, is that random assignment of D_i eliminates selection bias.

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{\text{Observed difference in average health}} = \underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{\text{Average treatment effect on the treated}} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{Selection bias}}.$$

$$\begin{aligned}
&\textit{Selection Bias} \\
&\quad = E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] \\
&\quad = E[Y_{0i} \mid D_i = 0] - E[Y_{0i} \mid D_i = 0] \\
&\quad = 0
\end{aligned}$$

- That is, with randomization, the selection bias term disappears!
- The observed group differences is the very ATT.

Some Iconic Examples:

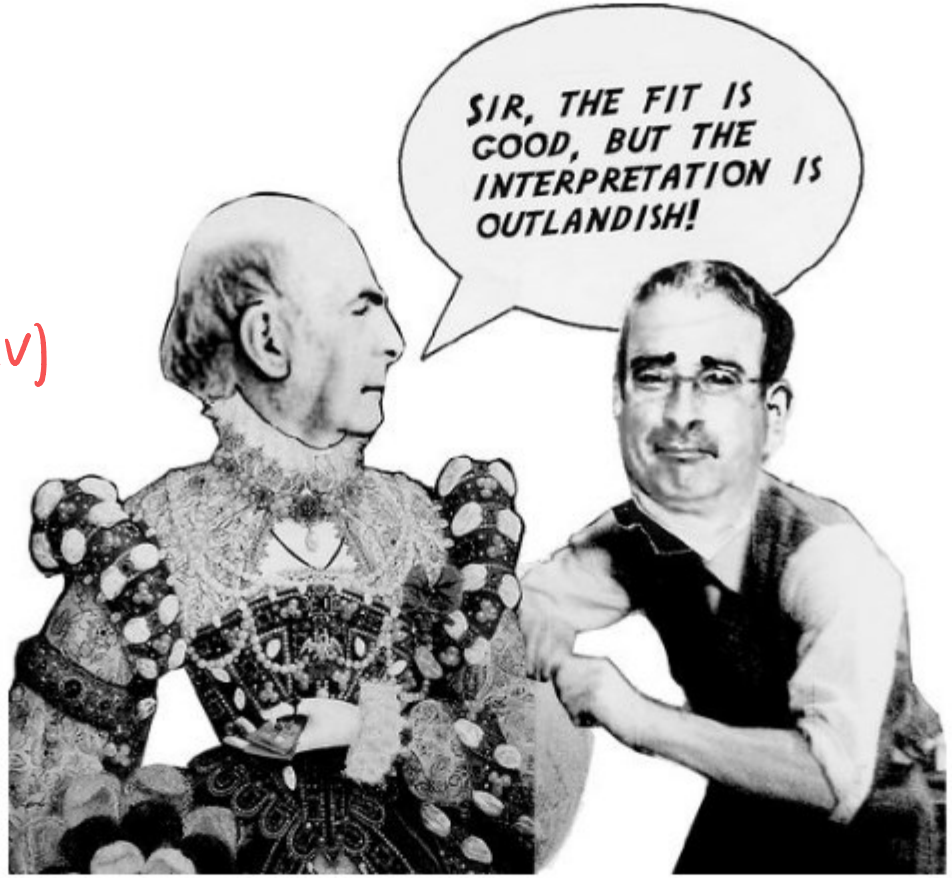
Evaluation of government-subsidized training Program 培训

- Target groups: disadvantaged workers such as long-term unemployed, drug addicts, and ex-offenders.
- Early studies based on non-experimental comparisons of participants and nonparticipants, paradoxically, often show that after training, the trainees earn less than plausible comparison groups (see, e.g., Ashenfelter, 1978; Ashenfelter and Card, 1985; Lalonde 1995). 接受培训工资更低?
- Here, too, selection bias is a natural concern: nonparticipants tend to be of higher ability, and therefore, they are predisposed not to take part in the training program
- In contrast, evidence from randomized evaluations of training programs generates mostly positive effects (see, e.g., Lalonde, 1986; Orr et al., 1996)

Which Effect Have You Identified?

- Average Treatment Effect, ATE
- Average Treatment Effect on the Treated, ATT
- Average Treatment Effect on the Untreated, ATU
- Local Average Treatment Effect on the Treated, LATE (IV)
- Intention-to-Treat Effect, ITT
- Example:
  - The impact of emigration on emigrants' income

移民对移民收入的影响



Several Examples

- Two girls in a village, named Red and Black, respectively. In the previous year, Red went to work in the city nearby and earned 20,000 yuan; Meanwhile, Black stayed in the village and earned 10, 000.
  - Black thought the earning premium of the migration (the extra 10,000) is not a big deal, just as she claimed, "If I were to work in the city, I can earn 20,000 too". Is she right? ✗
  - Actually, Red and Black are similar in almost all aspects, e.g., same age, height, similar grades in schools, etc. They both want to work in the city. But the Red made it because she won the lottery-drafting when a company came to their village to recruit new workers. However Black did not. With the new information and going back to question 1, now: "Is Black right?" ✓

random assignment

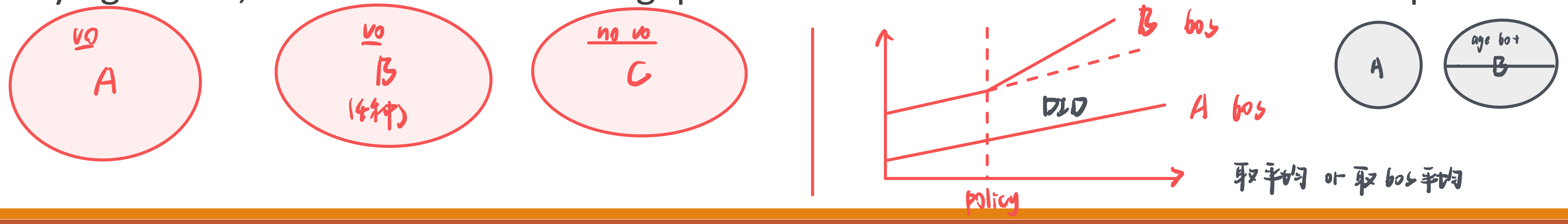
- There is another girl in the same village, named Pink. Though similar to Red and black in almost all respects, but she has never thought to work outside her village. She often thinks, “I’m not the material to work in the city. There are so many risks and uncertainties ahead. Why take the trouble for the extra 10,000 ￥?” But sometimes, she also feels sorry for the forgone 10,000￥. Do you think her regretting of losing 10,000 ￥ can be justified? risk averse.

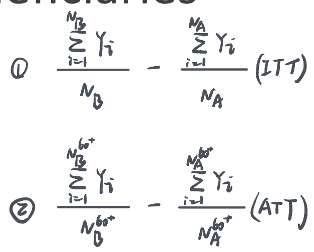
- **(The Average Treatment Effect)** One day, a volcano erupted, the lava overflowed irregularly and destroyed many houses. The villagers who suffered the loss of shelter had to move to the nearby city. Now, we gain an opportunity to compare the income between the emigrants and the villagers. What does the income gap represent? (ATU=ATT=ATE)

- **(The Local Average Treatment Effect)** In another case, e.g., another village, B, though many villagers suffered the loss of their house, not all of them were willing to go to the city. More specifically, there are roughly four groups of villagers.
  - Group one, choose to go work in the city if the catastrophe occurs, otherwise they will stay. (Compliers)
  - Group two, always go to work in the city, no matter if the catastrophe occurs (always-takers)
  - Group three, always stays in the village, no matter if the catastrophe occurs (never-takers)
  - Group four, chooses to work in the city if the catastrophe does not occur, otherwise they will stay. (defiers)

- **(Intention-to-Treat effect)** A third village, named C, was intact in this volcano eruption. Now we can compare the average income of villagers in A and C, or B and C, what does the gap mean?

  -or more realistically, Province A, Province B; B rolls out a program, e.g., a pension program, the beneficiaries is the elderly aged 60+; what does the health gap between residents in A and those in B capture?





Example I: Tennessee STAR Experiment

- Context of literature: " education production " (input-output)
  - One of the most expensive inputs is class size.
  - Whether the expense of smaller classes has a payoff in terms of higher student achievement?

- The STAR experiment cost $ 12 million and was implemented for a cohort of kindergartners in 1985-86. The study ran for four years until the original cohort of kindergartners was in third grade and involved about 11,600 children.

- The average class size in regular Tennessee classes in 1985-86 was about 22.3. The experiment assigned students to one of three treatments: *small classes with 13-17 children*, *regular classes with 22-25 children and a part-time teacher's aide (the usual arrangement)*, or *regular classes with a full-time teacher's aide.* Schools with at least three classes in each grade could choose to participate in the experiment.

Main points and pitfalls about the STAR

- The first question to ask about a randomized experiment is
  - Whether the randomization successfully balanced subjects' characteristics across the different treatment groups.
  - To assess this, it's common to compare ***pretreatment*** or ***predetermined*** outcomes or other covariates across groups.--- " Balance Test"

  (Pretreatment outcomes, pre-determined variables whose values are determined before the treatment is assigned vs. outcome variables, channel variable)

- Second question is whether the experiment has succeeded in creating the desired variation.
  - Parents might manage to lobby the teachers or principals to get their children assigned to small classes

TABLE 2.2.1

Comparison of treatment and control characteristics in the Tennessee STAR experiment

Variable	Small	Regular	Regular/Aide	across groups
Free lunch	.47	.48	.50	.09
White/Asian	.68	.67	.66	.26
Age in 1985	5.44	5.43	5.42	.32
Attrition rate	.49	.52	.53	.02
Class size in kindergarten	15.10	22.40	22.80	.00
Percentile score in kindergarten	54.70	48.90	50.00	.00

Well-balanced

F-test, 无差异.

损耗率

Post-treatment outcome

Notes: Adapted from Krueger (1999), table I. The table shows means of variables by treatment status for the sample of students who entered STAR in kindergarten. The *P*-value in the last column is for the *F*-test of equality of variable means across all three groups. The free lunch variable is the fraction receiving a free lunch. The percentile score is the average percentile score on three Stanford Achievement Tests. The attrition rate is the proportion lost to follow-up before completing third grade.

Explanatory Variable	(1)	(2)	(3)	(4)
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian	—	—	8.35	8.44
			(1.35)	(1.36)
Girl	—	—	4.48	4.39
			(.63)	(.63)
Free lunch	—	—	−13.15	−13.07
			(.77)	(.77)
White teacher	—	—	—	−.57
				(2.10)
Teacher experience	—	—	—	.26
				(.10)
Teacher Master's degree	—	—	—	−0.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

Notes: Adapted from Krueger (1999), table V. The dependent variable is the Stanford Achievement Test percentile score. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. The sample size is 5,681.

- Randomized trials are never perfect, and STAR is no exception.
  - Attrition rate (proportion of students lost to follow-up), Krueger (1999)
  - Pupils who repeated or skipped a grade left the experiment.
  - Students who entered an experimental school one grade later were added to the experiment and randomly assigned to one of the classes.
  - One unfortunate aspect of the experiment is that students in the regular and regular/aide classes were reassigned after the kindergarten year, possibly because of protests by the parents with children in the regular classrooms.
  - There was also some switching of children after the kindergarten year.
- An alternative is to find natural or quasi-experiments that mimic a randomized trial by changing the variable of interest while other factors are kept balanced.
- A notional ***"Randomized Controlled Trial (RCT)"*** is our benchmark, golden standard.

Example II: RAND Health Insurance Experiment (HIE)

- HIE ran from 1974 to 1982, was one of the most influential social experiments in research history.
- The HIE enrolled 3,958 people aged 14 to 61 from six areas of the country. The HIE sample excluded Medicare participants and most Medicaid and military health insurance subscribers.
- HIE participants were randomly assigned to one of 14 insurance plans. Participants did not have to pay insurance premiums, but the plans had a variety of provisions related to cost sharing, leading to large differences in the amount of insurance they offered.

- Four categories of plans:
  - The most generous HIE plan offered comprehensive care for free.
  - At the other end of the insurance spectrum, three "catastrophic coverage" plans required families to pay 95% of their healthcare costs, though these costs were capped as a proportion of income (or capped at $1,000 per family, if that was lower). The catastrophic plans approximate a no-insurance condition. < $1000, 5%. > $1000, 95%.
  - A second insurance scheme (the "individual deductible" plan) also required families to pay 95% of outpatient charges, but only up to $150 per person or $450 per family.
  - A group of nine other plans had a variety of coinsurance provisions, requiring participants to cover anywhere from 25% to 50% of charges, but always capped at a proportion of income or $1,000, whichever was lower.
- Participating families enrolled in the experimental plans for 3 or 5 years and agreed to give up any earlier insurance coverage in return for a fixed monthly payment unrelated to their use of medical care.

Core questions for health insurance economics

- Whether and by how much healthcare use falls when the healthcare price goes up?
- Whether more comprehensive and generous health insurance coverage lead to better health?

Demographic characteristics and baseline health in the RAND HIE

	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

- The catastrophic plans provide approximate no-insurance control, while the deductible, coinsurance, and free plans are characterized by increasing levels of coverage.
- Remember the steps:
  - Step 1: Check for balance (of predetermined or pretreatment variables):
    - (1) the difference between plan groups are mostly small and as likely to go one way as another;
    - (2) a few isolated statistically significant differences is usually also attributed to chance;
    - (3) lack of consistent patterns reinforces the notion that these gaps are due to chance
    - (4) the standard errors in this table are not very big, indicating differences are measured reasonably precisely.
  - Step 2: Check for the variation created.

Health expenditure and health outcomes in the RAND HIE

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

- Finding 1: Subjects assigned to more generous insurance plans used substantially more health care. As shown in the table left, gaps between plan groups are economically important and statistically significant.
- Did the extra care and expense make them healthier? Unfortunately, no.
- It seems to suggest that generous health insurance can have unintended and undesirable consequences, increasing health-care usage and costs, without generating a dividend in the form of better health.

pay↓, 利用率↑, 但没有转化的健康结果.

prime age.

Example III: The Oregon Trail

- *Truth is often hidden, like a shadow in darkness.*

  -Kwai Chang Caine (*Kung Fu*)

- *All those moments will be lost in time, like tears in the rain.*

  -Roy Batty (*Blade Runner*)

- Two things that might affect the generalization of HIE's findings
  - Each treatment group had at least catastrophic coverage, so the financial liability is limited under every treatment.
  - Today's uninsured Americans differ considerably from the HIE population: most of the uninsured are younger, less educated, poorer, and less likely to be working. The value of extra health care in such a group might be very different than for the middle-class families that participated in the HIE.
- Suppose we were to expand Medicaid (families on welfare, some of the disabled, other poor children, and poor pregnant women) to cover those who don't qualify under current rules. How would such an expansion affect healthcare spending?
- Would it shift treatment from costly and crowded emergency departments to possibly more effective primary care?
- Would Medicaid expansion improve health?

RCT

- Oregon's health insurance lottery: randomly selecting winners and losers from a pool of registrants, though coverage was not automatic, even for lottery winners.
- Winners won the opportunity to apply for the state-run Oregon Health Plan (OHP), the Oregon version of Medicaid.

Selective

- The state then reviewed these applications, awarding coverage to Oregon residents who were U.S. citizens or legal immigrants aged 19–64, not otherwise eligible for Medicaid, uninsured for at least 6 months, with income below the federal poverty level, and few financial assets. 资格审查，筛选
- To initiate coverage, lottery winners had to document their poverty status and submit the required paperwork within 45 days.
- Roughly 75,000 lottery applicants registered for expanded coverage through the OHP. Of these, almost 30,000 were randomly selected and invited to apply for OHP; these winners constitute the OHP treatment group. The other 45,000 constitute the OHP control sample.

Treatment: winners (intend to treat)

Control: losers

OHP effects on insurance coverage and health-care use

Outcome	Oregon: Control mean (1)	Oregon: Treatment effect (2) Δx	Portland area: Control mean (3)	Portland area: Treatment effect (4) Δx
A. Administrative data				
Ever on Medicaid	.141	.256*** (.004)	.151	.247*** (.006)
Any hospital admissions	.067	.005*** (.002)		
Any emergency department visit			.345	.017*** (.006)
Number of emergency department visits			1.02	.101*** (.029)
Sample size	74,922		24,646	
B. Survey data				
Outpatient visits (in the past 6 months)	1.91	.314*** (.054)		
Any prescriptions?	.637	.025*** (.008)		
Sample size	23,741			

- Whether OHP lottery winners were more likely to end up insured as a result of winning？
- Shift from emergency department toward less costly source of care? Frustratingly, emergency department visits, outpatient visits, and prescription drug use all increased markedly.

OHP effects on health indicators and financial health

Outcome	Oregon: Control mean (1)	Oregon: Treatment effect (2)	Portland area: Control mean (3)	Portland area: Treatment effect (4)
A. Health indicators				
Health is good	.548	.039 (.008) ***		
Physical health index			45.5	.29 (.21)
Mental health index			44.4	.47 (.24)
Cholesterol			204	.53 (.69)
Systolic blood pressure (mm Hg)			119	–.13 (.30)
B. Financial health				
Medical expenditures >30% of income			.055	–.011 (.005)
Any medical debt?			.568	–.032 (.010)
Sample size	23,741		12,229	

cognitive.

- A modest improvement in the probability they assess their health as being good or better.
- These gains stem more from improved mental rather than physical health.
- However, as in the RAND HIE, physical health indicators like cholesterol and blood pressure were largely unchanged by increased access to OHP insurance.
- Improvement in financial health; Does it account for improved mental health in the treatment group?

Implications from Rand's HIE and OHE

- The use of healthcare services increases sharply in response to insurance coverage
- Neither experiment reveals much of an insurance effect on physical health. Therefore, it appears that subsidized public health insurance should not be expected to yield a dramatic health dividend
- Anything that you want to comment on?
- For more reading see:
  - Aron-Dine, A., L. Einav and A. Finkelstein (2013). 'The RAND health insurance experiment, three decades later', Journal of Economic Perspectives, vol. 27(1), pp. 197-222.
  - Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen and K. Baicker (2012). ' The Oregon Health Insurance Experiment: Evidence From The First Year', The Quarterly journal of economics, vol. 127(3), pp. 1057-1106.

IV. Regression Analysis of Experiments

Suppose (for now) that the treatment effect is the same for everyone, say $Y_{1i} — Y_{0i} = \rho$, a constant. With constant treatment effects, we can rewrite Y_i in the form (BY DEFINITION)

$$
\begin{array}{ccccccc}
Y_i = & \alpha & + & \rho & D_i & + & \eta_i, \\
 & \shortparallel & & \shortparallel & & & \shortparallel \\
 & E(Y_{0i}) & & (Y_{1i} - Y_{0i}) & & & Y_{0i} - E(Y_{0i}),
\end{array}
$$

$$Y_i = E(Y_{0i}) + (Y_{1i} - Y_{0i}) D_i + Y_{0i} - E(Y_{0i}) = \alpha + \rho D_i + \eta_i$$

so that, the previous so-called naïve comparison can be rewritten as:

$$E[Y_i|D_i = 1] = \alpha + \rho + E[\eta_i|D_i = 1]$$
$$E[Y_i|D_i = 0] = \alpha + E[\eta_i|D_i = 0],$$

$$\Rightarrow$$

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \underbrace{\rho}_{\text{Treatment effect}} + \underbrace{E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0]}_{\text{Selection bias}}.$$

- Thus, selection bias amounts to the correlation between the regression error term, η_i, and the regressor, D_i. (ENDOGENEITY PROBLEM)
- Since

$$E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0] = E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0],$$

- (Remember that $\eta_i = Y_{0i} - E(Y_{0i})$)
- That is, if D_i is randomly assigned, then D_i and η_i will be independent of each other, then regression will obtain a consistent estimate of ρ, that is, the Treatment Effect.
- In the STAR experiment, where D_i is randomly assigned, the selection bias term disappears, and the regression of Y_i on D_i estimates the causal effect of interest, ρ.

What about the covariates X_i

- In reality, random assignment is rarely unconditional
  - As in the STAR experiment, conditional assignment was adopted. In particular, the assignment to classes of different sizes occurred within schools but not across schools. (Schools of different types, say, urban or rural, a bit more or less affect the likelihood that a child can be assigned to a small size class)
  - Contaminated by differences in achievements in schools of different types–school fixed effect

Random Assignment conditional on X_i

控制变量：性别、城市、民族，

- **Conditional Independence Assumption:** CIA：条件独立，

$$Y_{1i}, Y_{0i} \perp D_i \mid X_i$$

D_i 是否上NJU

Y potential outcome

$$Y_i = \alpha + \rho D_i + X_i'\gamma + \eta_i,$$

- The STAR experimental design used conditional random assignment. Assignment to classes of different sizes was random within schools but not across schools.
- Inclusion of the variables X_i , although not necessary in this case, may generate more precise estimates of the causal effect of interest.– They explain an important part of the variance in the dependent variable, hence, they reduce the standard error.
- If X_i is uncorrelated with the treatment D_i, then they will not affect the estimate of ρ

Final remarks

- Regression plays an exceptionally important role in empirical economic research.
- Regression is well-suited to the analysis of experimental data.
- In some cases (where CIA is satisfied), regression can also be used to approximate experiments in the absence of random assignment.

Thanks for your Attention!

Any questions or comments please write to:

chenglingguo@nju.edu.cn