Section 3: RCTs, Power Analysis, Attrition

Valentine Gilbert

February 17, 2022

Overview

1. RCT example: Carter et al. (2017)
2. Power Analysis
3. Attrition

Education RCT: Setting

Carter, Greenberg, and Walker, 2017. *The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy.* Economics of Education Review

- Since mid-90s, large increase in internet access in classrooms, decrease in ratio of students to computers in classrooms
- Policy proposals to further expand internet access to students
- Competition among colleges to provide uninterrupted high-speed connectivity
- But unclear ex-ante what effect computer use in classrooms has on student performance
  - Benefits: Take faster notes, access online resources
  - Costs: Mindless note-taking, distractions, doom scrolling

Education RCT: Overview

Questions:

- What's the research question?
- What's the naïve comparison?
- What's wrong with the naïve comparison?

Education RCT: Potential Outcomes

Consider the difference in average outcomes between students who do and do not use computers in the classroom:

$$
\begin{aligned}
E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= \underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{\text{TOT}} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

- The naïve comparison gives us the TOT only if the students who used computers would have had the same average outcomes as the students who didn't use computers *if they, too, didn't use computers*

Aside: Regression and differences in means

How does the difference in group means relate to our regression estimate?

- In the simple case where we regress an outcome on a binary treatment indicator and a constant, the coefficient on treatment *is* the difference in means
- Multiple ways to see this
- Simplest is to recall that regression gives the best linear approximation of the CEF
- When $\beta_0 = E[Y_i|D_i = 0]$ and $\beta_1 = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$, the regression function is the CEF:

$$E[Y_i|D_i] = \underbrace{E[Y_i|D_i = 0]}_{\beta_0} + \underbrace{(E[Y_i|D_i = 1] - E[Y_i|D_i = 0])}_{\beta_1} D_i$$

Education RCT: Randomization

Carter et al. randomly assign introductory econ classrooms at West Point to one of three groups:

- Control group with unrestricted computer use
- Treated group with no computers allowed at all
- Treated group with tablets allowed flat on the desk

Randomization ensures that treatment status is independent of potential outcomes ($\{Y_{0i}, Y_{1i}\} \perp D_i$), which ensures that the three groups are the same, on average:

$$E[Y_{0i}|D_i = 0] = E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 2] = E[Y_{0i}]$$

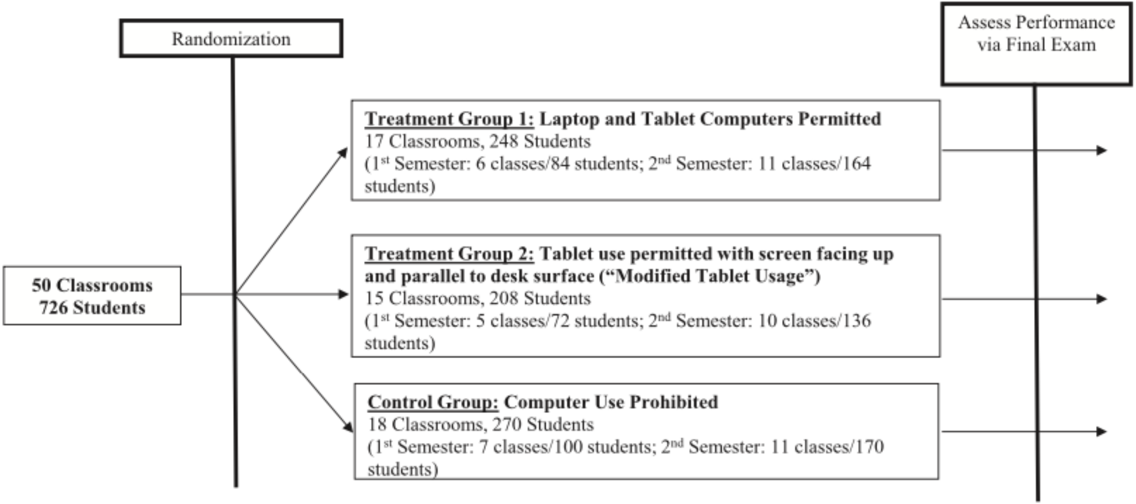


Fig. 1. Experimental design.

Estimation

To estimate the effect of computer use in the classroom, Carter et al. estimate the following regression equation:

$$Y_{ijht} = \kappa_{jt} + \lambda_{ht} + \gamma' X_i + \pi Z_{jht} + \eta_{ijht}$$

- Y_{ijht} : Outcome for student i with professor j during class hour h in semester t
- κ_{jt} : Professor-by-semester fixed effects
- λ_{ht} : Hour-by-semester fixed effects
- X_i : Student characteristics
- Z_{jht} : Treatment indicator

Education RCT: Balance Table

Table 2
Summary statistics and covariate balance.

	Control (1)	Treatment 1 (laptops/tablets) (2)	Treatment 2 (tablets, face up) (3)	Both treatments vs. control (4)	Treatment 1 vs. control (5)	Treatment 2 vs. control (6)
A. Baseline characteristics						
Female	0.17	0.20	0.19	0.03	0.06	0.00
				(0.03)	(0.04)	(0.04)
White	0.64	0.67	0.66	0.02	0.02	0.02
				(0.04)	(0.04)	(0.05)
Black	0.11	0.10	0.11	−0.02	−0.02	−0.03
				(0.03)	(0.03)	(0.04)
Hispanic	0.13	0.13	0.09	0.00	0.02	−0.03
				(0.03)	(0.03)	(0.03)
Age	20.12	20.15	20.15	0.03	0.05	0.06
	[1.06]	[1.00]	[0.96]	(0.08)	(0.09)	(0.10)
Prior military service	0.19	0.19	0.16	−0.02	0.00	−0.01
				(0.03)	(0.04)	(0.04)
Division I athlete	0.29	0.40	0.35	0.05	0.07*	0.04
				(0.04)	(0.04)	(0.05)
GPA at baseline	2.87	2.82	2.89	−0.01	−0.05	0.03
	[0.52]	[0.54]	[0.51]	(0.04)	(0.05)	(0.05)
Composite ACT	28.78	28.30	28.30	−0.34	−0.37	−0.54
	[3.21]	[3.46]	[3.27]	(0.26)	(0.31)	(0.33)
P-Val (Joint χ^2 Test)				0.610	0.532	0.361
B. Observed computer (laptop or tablet) use						
any computer use	0.00	0.81	0.39	0.62***	0.79***	0.40***
				(0.02)	(0.03)	(0.04)
Average computer use	0.00	0.57	0.22	0.42***	0.56***	0.24***
				(0.02)	(0.02)	(0.03)
Observations	270	248	208	726	518	478

Education RCT: Results

Table 3
Laptop and modified-tablet classrooms vs. non-computer classrooms.

	(1)	(2)	(3)	(4)
A. Dependent variable: Final exam multiple choice and short answer score				
Laptop/tablet class	−0.21***	−0.20***	−0.19***	−0.18***
	(0.08)	(0.07)	(0.06)	(0.06)
GPA at start of course			1.13***	1.00***
			(0.06)	(0.06)
Composite ACT				0.06***
				(0.01)
Demographic controls		X	X	X
R^2	0.05	0.24	0.52	0.54
Robust SE *P*-Val	0.010	0.005	0.001	0.002
Wild Bootstrap *P*-Val	0.000	0.000	0.000	0.000

Education RCT: Results

Table 4
Unrestricted laptop/tablet classrooms vs. non-computer classrooms.

	(1)	(2)	(3)	(4)
A. Dependent variable: Final exam multiple choice and short answer score				
Computer class	−0.28***	−0.23***	−0.19***	−0.18***
	(0.10)	(0.09)	(0.07)	(0.07)
GPA at start of course			1.09***	0.92***
			(0.07)	(0.07)
Composite ACT				0.07***
				(0.01)
Demographic controls		X	X	X
R^2	0.08	0.28	0.54	0.57
Robust SE *P*-Val	0.003	0.007	0.005	0.005
Wild Bootstrap *P*-Val	0.000	0.000	0.000	0.000

Education RCT: Results

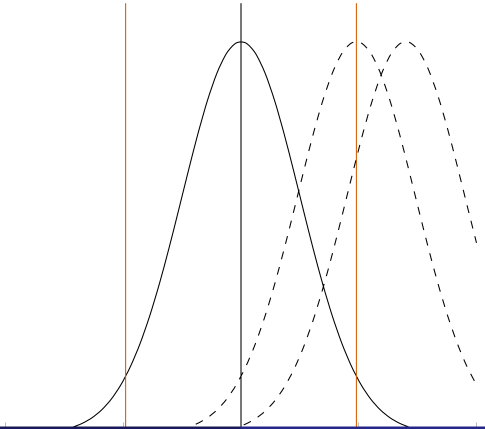
Table 5
Modified-tablet classrooms vs. non-computer classrooms.

	(1)	(2)	(3)	(4)
A. Dependent variable: Final exam multiple choice and short answer score				
Computer class	–0.17*	–0.18**	–0.20***	–0.17**
	(0.10)	(0.09)	(0.07)	(0.07)
GPA at start of course			1.12***	1.01***
			(0.07)	(0.08)
Composite ACT				0.05***
				(0.01)
Demographic controls		X	X	X
R^2	0.07	0.26	0.53	0.54
Robust SE *P*-Val	0.087	0.050	0.007	0.019
Wild Bootstrap *P*-Val	0.000	0.000	0.000	0.000

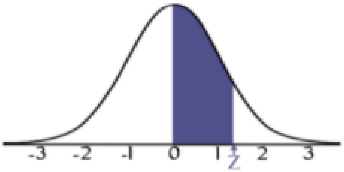
Power Analysis

- Power analyses help us determine the "minimum detectable effect" (MDE) of an RCT, which helps us to avoid the worst case scenario of an imprecise null result
- In other words, power analysis helps you know what effects you'll be able to rule out in the event that you do get an insignificant estimate
- Rule of thumb: The MDE is $2.8 \times SE(\hat{\beta})$
- Why?
- Question: Suppose the true treatment effect were 1.96 times the estimated standard error. What would your power be?

Power Analysis



Under the null, $\hat{\beta}$ has the following distribution, where the x-axis indicates standard errorsWe reject the null if $\hat{\beta}$ falls outside the 95% confidence interval, indicated by the red linesIf the true effect size is 1.96 times the standard error, the true distribution is given by the dashed pdf. What's the probability that we fail to reject the null?If the true effect size is 2.8 times the standard error, the true



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

Power Analysis

- The key estimand for power analysis is the standard error
- In the simple case where the data are i.i.d. and our estimated treatment effect is the simple difference in means, this standard error is

$$SE(\hat{\beta}) = \frac{1}{\sqrt{P(1-P)}} \frac{\sigma}{\sqrt{N}}$$

- The key unknown here is the residual variance, σ^2
- We can estimate this with baseline/pilot data, or based on historical data (e.g. from past experiments)

Attrition Review

- We discussed 3 kinds of attrition in class:
  - Random attrition
  - Non-random attrition that's unrelated to treatment status
  - Attrition that *is* related to treatment status
- Question: What are the consequences of each kind of attrition?

Differential Attrition

- Differential attrition between treatment and control groups can ruin the validity of an RCT, so super important to think about how to prevent it
- But sometimes differential attrition happens. So what do we do?
- Various approaches that allow us to bound the selection bias from differential attrition

Angrist, Bettinger, and Kremer (2006): Setup

Angrist, Bettinger, and Kremer (2006 AER) studies the long run effects of secondary school vouchers for low-income students in Colombia

- Vouchers were awarded by lottery to qualified students to subsidize private school fees starting in 6th grade
- The authors want to study the effect on high school graduation and college entrance exam scores, 7 years after the lottery
- Estimating the effect on high school graduation rates is straightforward. But comparing college entrance exam scores for lottery winners and losers is a problem – why?
- Lottery winners were more likely to take the exam
  $\rightarrow$ Selection on an endogenous variable

Angrist, Bettinger, and Kremer (2006): Problem

- Differential test-taking rates between lottery winners and losers presents the same econometric problem as differential attrition
- We can also think of this in terms of bad control:
    - Winning the lottery has an effect on test-taking
    - So conditioning on whether a student takes a test (as we do when comparing mean scores of lottery winners and losers) reintroduces selection bias
    - That is, conditional on having taken college exams, comparisons of lottery winners and losers isn't apples-to-apples
- Question: Which way do you think the selection bias goes?

Angrist, Bettinger, and Kremer (2006): Bounds

- Glennerster and Takavarasha (2013) discuss two approaches to bounding selection bias from differential attrition:
  - Manski-Horowitz bounds: Construct upper bound by assigning the most positive outcome to attritors in the treatment group and most negative outcome to attritors in control group
  - Lee bounds: Test robustness of findings by equalizing attrition rates in treatment and control groups
    - For example, drop the 5% of treatment group observations with the highest outcome if attrition rate in control group is 5% higher than in the treatment group
- Angrist, Bettinger, and Kremer take a slightly different approach based on monotonicity assumptions: Get *upper* bound on treatment effect by equalizing attrition rates and use unconditional mean comparison as lower bound
  - Assumptions: $y_{1i} \geq y_{0i}$ and $T_{1i} \geq T_{0i}$ for all i

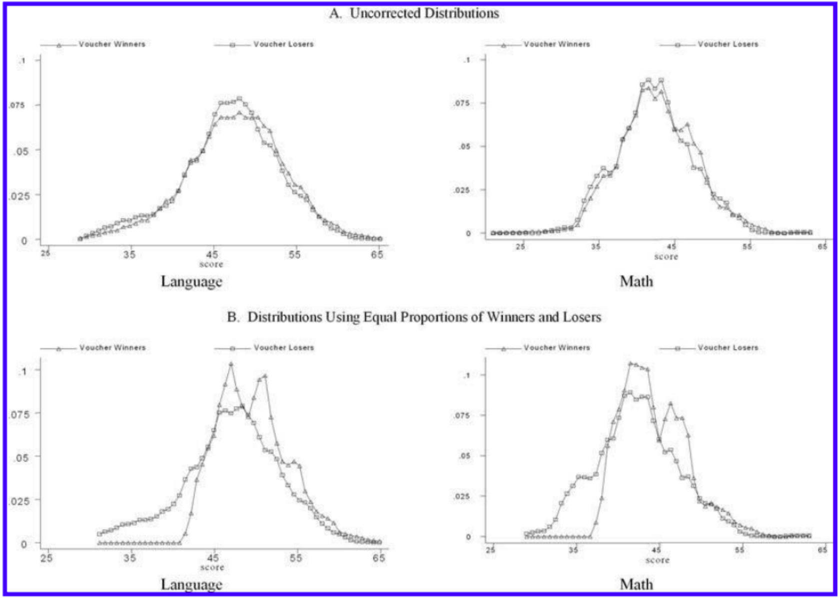


FIGURE 2. TEST-SCORE DISTRIBUTIONS