Causal analysis

Chapter 20: Experiments

Álmos Telegdy

Corvinus University of Budapest

Controlled experiments

- In experimental data, subjects are assigned to a treated and an untreated ("control") group by the experimenter $\Rightarrow$ **controlled assignment**
  - No self-selection: the value of x that subjects "receive" is not affected by their state or decisions
  - No reverse causality: the outcome y does not affect x in any way

Controlled experiments

- Well-controlled assignment of x results in very similar features for subjects with $x = 1$ and $x = 0$
- Observations in the treatment group and observations in the control group are expected to have (on average)
  - the same values of $E[y|x = 0]$
  - same values of the treatment effect $E[y|x = 1] - E[y|x = 0]$
- Independence assumption of the potential outcomes framework is met
- ATE will be identified and we can estimate it

Controlled Experiment types

- **Field experiments**: aim is to have control and treated groups as similar as possible in real-world decision situations
  - Test the impact of small loans in rural areas
- **Lab experiments**: carried out in an artificial environment, usually a computer lab
  - Test how people play games, react to incentives
- **A/B tests:** aim to evaluate different versions of the same product
  - Online advertisements or websites

Randomization in controlled experiments

- Controlled assignment involves **randomization**
- Randomization is an assignment rule: it assigns different values of x to subjects
- Random assignment rule is independent of all potential confounders
- This independence is **by design**
    - A well-executed randomization guarantees that the observations are very similar in groups with different values of the causal variable $x \rightarrow$ treated and control groups are similar

The Experimental Setup

- Well-controlled experiments – average difference in y identifies the average effect of x
  - With a binary treatment, the average difference in y between the treated and untreated group identifies ATE
- Simple regression of $y^E = \alpha + \beta x$ plus a well-controlled experiment: Estimated ATE $= \beta$

Random assignment and checking balance

- Random assignment should make the distribution of all variables identical in assignment groups
  - Binary: in the treated and control groups
  - Quantitative intervention: at all values of x
- This is called **balance**. The variables are said to be balanced across groups if they have the same distribution

- Must check it - process of random assignment may be imperfect
- A random rule leads to independence of potential outcomes **in expectation**
  - However, with some tiny probability, the actual assignment may lead to groups that differ in potential outcomes
  - Matters when groups are small

Field experiments

Subjects are studied in their own environment

- Subjects: individuals, firms, townships
- Treatments assigned randomly

Case study: Working from home

- Question: what is the effect of working from home four days a week on worker performance and quit rates?
- Data: Chinese travel agency. The experiment took place in its call center in Shanghai that dealt with booking hotels and airfare
- Background: commuting time for employees 80 minutes per day. Employees work in cubicles

Case study: Working from home

- This is a field experiment
- About half of the subjects **order takers** – answer calls from customers and administer those calls
- Intervention: part of employees work from home four days a week
- Outcome variables:
  - quit firm (yes/no)
  - performance (number of phone calls)

Case study: Working from home

- 503 people volunteered for the experiment (external validity?)
- 249 qualified for the experiment (tenure > 6 months, had broadband internet access, had independent workspace at home)
- Selection based on birthday
  - even birth date assigned to work from home (131 workers)
  - odd birth date assigned to work from the office (118 workers)

Case study:balancing

	Treatment mean	Control mean	Std.dev.	p-value
Number of observations	131	118	249	
Prior performance z-score	-0.03	-0.04	0.58	0.87
Age	24	24	4	0.85
Male	0.47	0.47	0.50	0.99
Secondary technical school	0.46	0.47	0.50	0.80
High school	0.18	0.14	0.36	0.38
Tertiary	0.35	0.36	0.48	0.94
University	0.02	0.03	0.15	0.91
Prior experience (months)	19	17	26	0.48
Tenure (months)	26	28	22	0.45
Married	0.22	0.32	0.44	0.07
Children	0.11	0.24	0.38	0.01
Age of youngest child	4.60	3.00	3.35	0.14
Rent apartment	0.24	0.20	0.42	0.44
Cost of commute (yuan)	7.89	8.34	6.96	0.63
Own bedroom	0.99	1.00	0.06	0.13
Bonus (yuan monthly)	1031	1093	625	0.43
Gross wage (yuan monthly)	2950	3003	790	0.59
Proportion of order takers	0.52	0.56	0.50	0.53

Balancing contd.

- Difference in means typically small ✓
  - But: there are some larger differences in means – watch for them
- p-value usually large ✓
  - But: some differences have a p-value smaller than 10% – watch for them
- The productivity of the two groups is very similar *before treatment* (this is important) ✓

Compliance

Subjects can self-select even if they are randomly assigned to treatment

- Some subjects assigned to the treatment group do not get the treatment
  - Some unemployed decide to stop attending training
- Some subjects assigned to the control group get the treatment
  - Some unemployed decide to take the course

Compliance

Subjects can self-select even if they are randomly assigned to treatment

- Some subjects assigned to the treatment group do not get the treatment
  - Some unemployed decide to stop attending training
- Some subjects assigned to the control group get the treatment
  - Some unemployed decide to take the course

Compliance is rarely random

Intent to treat

If compliance is not complete, two treatment effects:

1. The "normal" treatment effect $\rightarrow$ ATE
2. The effect of being assigned to the treated group $\rightarrow$ **average intent-to-treat effect (AITTE)**
   - If compliance is perfect, ATE $=$ AITTE
   - If there is noncompliance, ATE $<$ AITTE in general $\rightarrow$ typically those comply who find it worth
   - AITTE is important $\rightarrow$ one wants to know the effect of a policy intervention, BUT the policy may also be affected by noncompliance

Noncompliance

Noncompliance can bias the estimation even if subjects are randomly assigned to treatment $\Rightarrow$ **Noncompliance is typically not random**

- E.g., those treated subjects fall out of treatment who think that the intervention is not useful
- They will have smaller ITE than the average subject

We can measure $AITTE$ in an RCT, but not ATE

What to do?

- If noncompliance rate is high, acknowledge
- Instead of ATE, estimate the **local average treatment effect (**$LATE$**)**

$LATE$: intuition

We can correct $AITTE$ if we know the proportion of treated (assumption: without the treatment, no effect)

$$AITTE = ATE^{'} \cdot Prop(\text{assigned, treated}) + 0 \cdot Prop(\text{assigned, untreated})$$
$$- \text{ATE}^{'} \cdot Prop(\text{not assigned, treated})$$

$ATE^{'} = LATE \rightarrow$ ATE on the treated group

LATE: estimation

Local *ATE*: the average effect of the treatment on those subjects which **behaved according to their assignment**

- $w =$ assignment ($w = 1$ – assigned; $w = 0$ – not assigned)
- $x =$ treatment ($x = 1$ – treated; $x = 0$ – untreated)
- $p =$ share of treated among those who were assigned to the treated group ("well behaved"): $p = \Pr[x = 1|w = 1]$
- $q =$ share of treated among those who were assigned to the untreated group ("deserters"): $q = \Pr[x = 1|w = 0]$
- $AITTE = E[y|w = 1] - E[y|w = 0]$

$$LATE = \frac{AITTE}{p - q} \tag{1}$$

$LATE$: example

$p = 0.5,\ q = 0,\ AITTE = 0.1,\ E[y|x = 0] = 0$

- $AITTE = p \cdot LATE + 1 - p \cdot 0$
- $LATE = \frac{AITTE}{p} = 0.1/0.5 = 0.2$

Case study

- Some workers assigned to the treated group went back to the office to work ($< 20\%$) – internet problems, working conditions, etc.
- The control group did comply perfectly
- Treatment is incomplete $\Rightarrow$ we can measure $AITTE$
- But: treatment was close to complete $\Rightarrow$ $AITTE$ should be close to ATE

Estimation

Endogeneity is solved by random assignment

- Computation of the mean difference (+ t-test) is okay
- It is still better, to run a simple OLS regression: $y^E = \alpha + \beta x$
  - We get the same result than with mean difference: $\alpha = E[y^0]$, $\beta = ATE$ or $\beta = AITTE$
  - The regression computes the SD of the $\beta \Rightarrow p$-value, confidence interval
  - One can add control variables to test the balance of the samples

OLS regression

$$y_i^E = \alpha + \beta W_i + (\gamma_1 z_{i1} + \gamma_2 z_{i2} + ...) \tag{2}$$

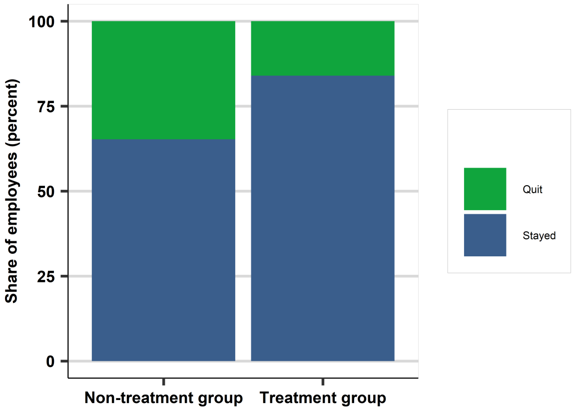
- Unassigned: $W_i = 0$
- Assigned: $W_i = 1$
- Control variables: $Z_{i1,2...}$

- The average of the outcome in the untreated group: $\alpha = \overline{y_i}[W_i = 0]$
- Average difference: $\beta = \overline{y_i}[W_i = 1] - \bar{y}_i[W_i = 0]$,

Case study continued

- Outcomes: quit rate, productivity
- 16% of treated, 35% of untreated quits. Difference is 19 pp.
- Number of phone calls 14 000 in treated group, 10 000 in control group. Difference is 4 000 calls

Case study: quits



- Stayed: did not leave the firm for at least 8 months
- Quit: left the firm during the 8 months

Case study: results

	(1)	(2)
VARIABLES	Quit job	Phone calls (thousand)
Treatment group	-0.19**	4.0**
	(0.055)	(0.99)
Constant	0.35**	10.1**
	(0.044)	(0.75)
Observations	249	134
R-squared	0.047	0.113

Source: `working-from-home` dataset.

Case study: results

VARIABLES	(1) Quit job	(2) Phone calls (thousand)
Treatment group	-0.19**	4.1**
	(0.056)	(0.96)
Married	-0.13	-5.4*
	(0.074)	(2.17)
Children	0.11	3.9
	(0.097)	(2.41)
Internet at home	0.18**	
	(0.036)	
Constant	0.19**	10.7**
	(0.056)	(0.76)
Observations	249	134
R-squared	0.055	0.168

Source: `working-from-home`

Case study: internal consistency

- Based on all the information, we can judge internal validity well.

Case study: internal consistency

- Based on all the information, we can judge internal validity well.
- Assignment was random.
- Compliance was imperfect, but only in the treatment group, and even here more than 80% of the subjects complied with the treatment.
- Spillovers are unlikely to be important in this experiment

Case study: external validity

- Based on all the information, we can somewhat judge external validity

Case study: external validity

- Based on all the information, we can somewhat judge external validity
- It had an actual impact, management changed practices
- Would it work for other employees? Yes for those who are like the ones in the experiment (those who applied)
- Not necessary for all
- What can we say about other companies?

Problem 1: spillover effects

The intervention has spillover (or external) effect: the treatment of individual i affects individual j $\rightarrow$ the estimated effect different from the true effect

Example: Online ads

Effect of advertising: individual i seeing the ad alters the spending behavior of individual j

- Communication $\rightarrow$ seeing the ad may make individual i convince a friend j to purchase the product
- Imitation $\rightarrow$ the ad drives individual i purchase the product, and this example may be followed by a neighbor j
- Substitution if supply is fixed $\rightarrow$ if individual i purchases the product, individual j may be prevented from purchasing it

Problem 2: the experimental setting may alter the behavior of subjects

- **Hawthorne effect:** both treated and untreated individuals change behavior due to being observed
  - How light affects worker productivity? – Western Electrics Hawthorne factory
    - Later proved that the effect was not there at all
- **Placebo effect:** treated individuals change behavior just because they are in the treatment group $\Rightarrow$ Important in medical interventions
- **John Henry effect:** untreated individuals change behavior because they are in the comparison group
  - John Henry, a legendary US steel driver in the 1870s. When heard his output was being compared with that of a steam drill, worked so hard to outperform the machine that he died

Problem 3: Compliance (recap)

- Noncompliance
  - Not all units assigned to treatment actually finish treatment
  - Some units assigned to non-treatment end up being treated
- The estimation can be biased even if assignment was random

Problem 4: heterogeneity of the effect

RCTs report average treatment effects $\rightarrow$ BUT treatment effects are usually heterogeneous

- Small sample size: hard to look at subgroups
- Example: education interventions $\rightarrow$ average test score improvement might hide that most students see little change, while a small group sees a large gain

Problem 5: external validity

RCT usually based on small groups of observations

- Narrow focus → hard to generalize
- RCTs are expensive → hard to replicate
  - one result is not enough
- Hard to scale up the program → a program may work in small groups, but not for a whole region
  - general equilibrium effects