version: uploaded on 041819

Causal Effects / Experiments

Week 2

Yunkyu Sohn
School of Political Science and Economics
Waseda University

2019 Spring Statistics I

Logistics

- Syllabus on https://github.com/ysohn/stats/
- R Exercise (from textbook): https://github.com/kosukeimai/qss
- TA Office Hours

Contents (Book Chapter 2.1 - 2.4)

- Causality
  - Counterfactuals
  - Causal Effects
  - Fundamental Problem of Causal Inference
- Experimental Research
  - Role of Randomization
  - Sample Average Treatment Effect
- Internal and External Validities

Estimating Causal Effect Is Central Tenet of Social Science

- Essential components of causal statements
  - Treatment causes outcome.
    - Condition / Treatment
    - Effects / Outcome
- A few examples of causal statements
  - Democratic states do not conduct war with each other.
  - Increasing immigrant population facilitates nationalism.
  - Increasing income tax rate reduces inequality.
  - Democracy makes a state wealthier.
  - Having more friends makes people happier.

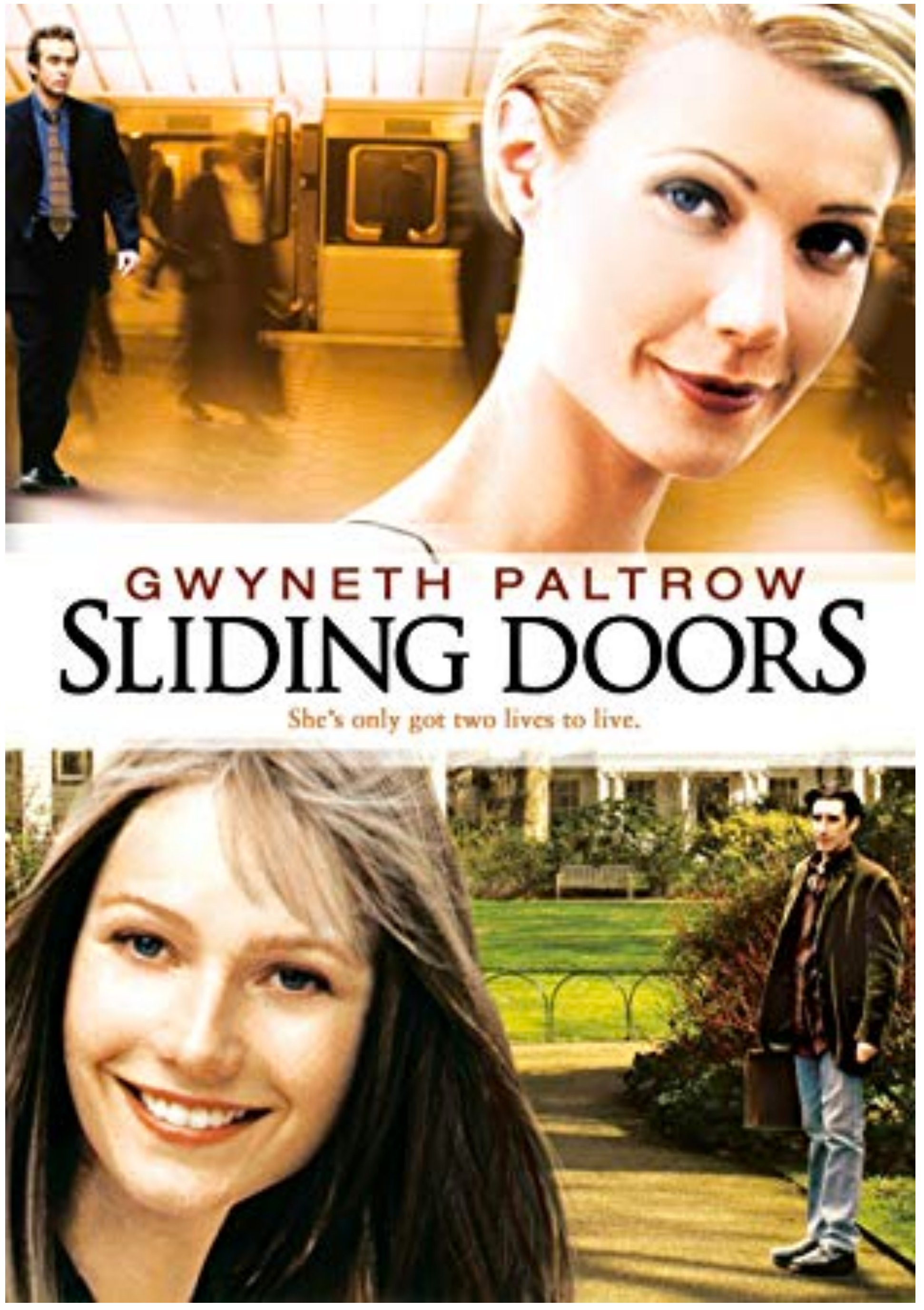
Estimating Causal Effect Is Central Tenet of Social Science

- Essential components of causal statements
  - Treatment causes outcome.
    - **Condition / Treatment**
    - Effects / Outcome
- A few examples of causal statements
  - **Democratic** states do not conduct war with each other.
  - **Increasing immigrant population** facilitates nationalism.
  - **Increasing income tax rate** reduces inequality.
  - **Democracy** makes a state wealthier.
  - **Having more friends** makes people happier.

Estimating Causal Effect Is Central Tenet of Social Science

- Essential components of causal statements
  - Treatment causes outcome.
    - **Condition / Treatment**
    - **Effects / Outcome**
- A few examples of causal statements
  - **Democratic states do not conduct war with each other**.
  - **Increasing immigrant population facilitates nationalism**.
  - **Increasing income tax rate reduces inequality**.
  - **Democracy makes a state wealthier**.
  - **Having more friends makes people happier**.

A Straightforward Example of Causal Effects



- Opening: https://youtu.be/BvUbv4iwbDs

Sliding Doors

- Can this be real?
  - outcome (Helen treated) - outcome (Helen control)
  - Unit of interest: Helen (Gwyneth Paltrow)
  - Treatment: Catching the train
  - Outcomes: Dating X or anything that follows
    - anything: e.g. wealth, happiness, divorce, and etc.

meeting James

meeting Gerry

Sliding Doors

- Can this be real? **ALL ABOUT COUNTER-FACTUALS**
  - outcome (Helen treated) - ~~outcome (Helen control)~~
  - Unit of interest: Helen (Gwyneth Paltrow)
  - Treatment: Catching the train
  - Outcomes: Dating X or anything that follows
    - anything: e.g. wealth, happiness, divorce, and etc.

meeting James

meeting Gerry

Sliding Doors

- Can this be real? **ALL ABOUT COUNTER-FACTUALS**
  - ~~outcome (Helen treated)~~ - outcome (Helen control)
  - Unit of interest: Helen (Gwyneth Paltrow)
  - Treatment: Catching the train
  - Outcomes: Dating X or anything that follows
    - anything: e.g. wealth, happiness, divorce, and etc.

meeting James

meeting Gerry

Fundamental problem of causal inference

Parallel Universe??

Fundamental problem of causal inference

Unless we find a wormhole.. If interested in physics, check Cosmos

e.g. The Resume Experiment (Book Chapter 2.1 -)

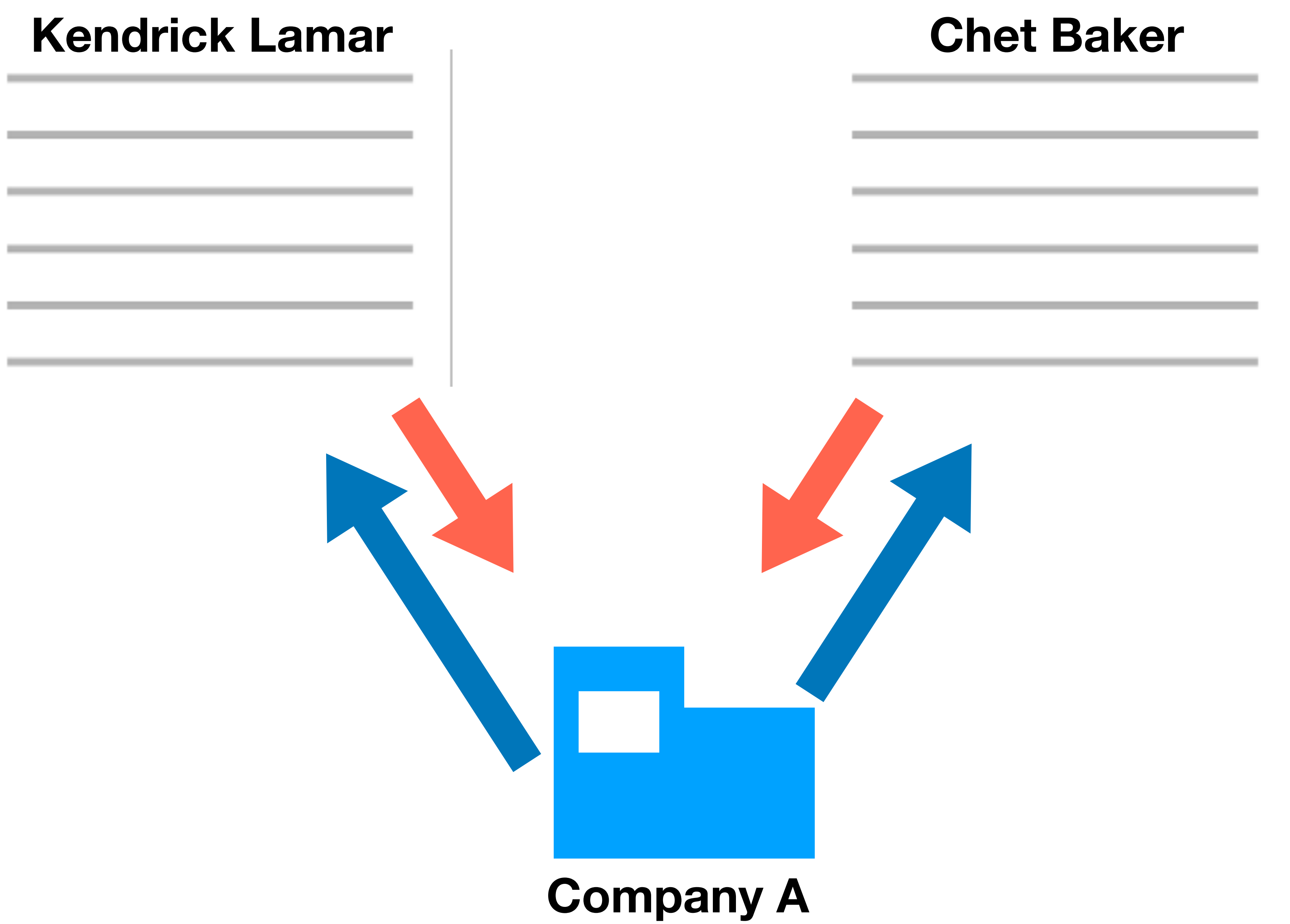
Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

By Marianne Bertrand and Sendhil Mullainathan*

e.g. The Resume Experiment (Book Chapter 2.1 -)

Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

By Marianne Bertrand and Sendhil Mullainathan*



Finding: Black names receive 50 percent less callbacks for interview

Formalizing the Design and Analysis

- Unit: fictitious job applicant (or resume)
- Unit index: i (by convention when there are multiple units)
- Treatment variable (causal variable of interest) T: black-sounding name
  - For binary treatment: T=1 treated; T=0 untreated (or controlled)
- Treatment group (treated units): resumes with black-sounding name
- Control group (untreated units): resumes with white-sounding name
- Outcome variable (response variable) Y : get callback
- Potential outcomes: $Y(1)$ and $Y(0)$
- Causal effect (for ordered outcomes): $Y(1) - Y(0)$

Formalizing the Design and Analysis

- Average treatment effect (causal effect of T on Y)
  - = quantifying the difference in outcome caused by treatment alone
  - no need to worry about (uncontrolled) characteristics
  - Sample (subset used in the study) Average Treatment Effect (SATE)

$$\textbf{SATE} = \frac{1}{n} \sum_{i=1}^{n} \{Y_i(1) - Y_i(0)\}$$

  - n: sample size
  - Not directly observable? Why?
    - Fundamental problem of causal inference
      - For unit i, only one of $Y_i(1)$ and $Y_i(0)$ is observable

Fundamental Problem of Causal Inference

- Fundamental problem of causal inference:
  - only one of the two potential outcomes is observable
  - Differences other than treatment variable prevent valid causal inference.
  - Matching on every single factors except treatment is Impossible
    - e.g. Can two people have exactly same resumes except their names?

    - Even for experiments: Can we send such pair to the same company?

Formalizing the Design and Analysis

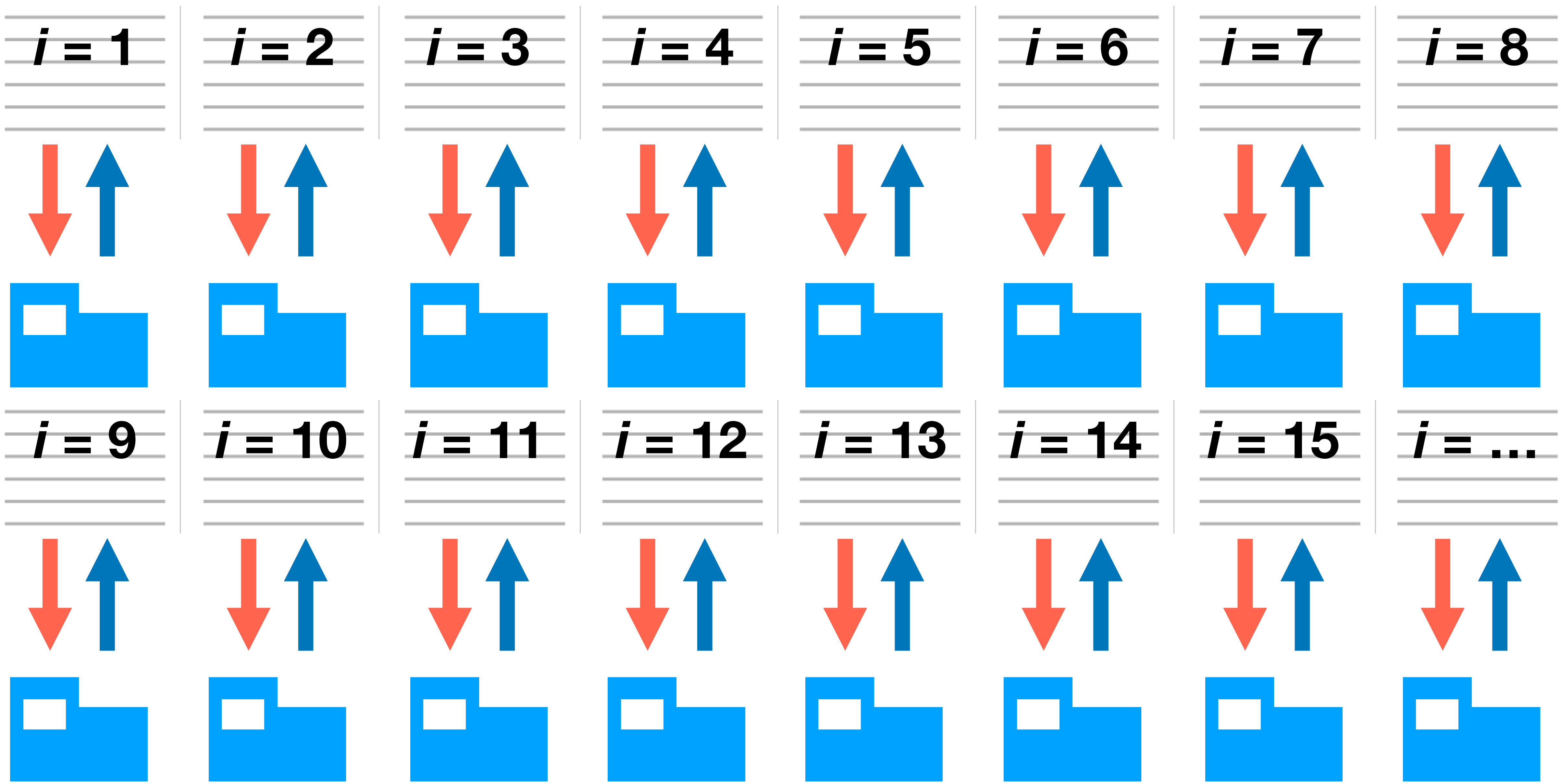
Table 2.3. Potential Outcome Framework of Causal Inference.

Résumé i	*Black-sounding name* T_i	*Callback* $Y_i(1)$	$Y_i(0)$	*Age*	*Education*
1	1	1	$\boxed{?}$	20	college
2	0	$\boxed{?}$	0	55	high school
3	0	$\boxed{?}$	1	40	graduate school
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	1	0	$\boxed{?}$	62	college

The Remedy: Randomized Controlled Trials (RCT)

Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

By MARIANNE BERTRAND AND SENDHIL MULLAINATHAN*



Randomly named 4,870 resumes sent for 1,300 employment ads

The Remedy: Randomized Controlled Trials (RCT)

- Randomly divided your sample into treatment and control groups
  - each unit is randomly assigned to the treatment or control groups
  - two groups are on average identical in all pretreatment characteristics.
  - average difference in outcome: solely due to the treatment
- Difference in the sample means estimator: Feasible alternative to SATE

$$D = \frac{1}{n/2} \sum_{i \in \{T_i = 1\}} Y_i - \frac{1}{n/2} \sum_{i \in \{T_i = 0\}} Y_i$$

  - Assuming that treatment/control groups have the same size ($n/2$)
  - The set of treated units: $\{T_i\!=\!1\}$
  - The set of controlled units: $\{T_i\!=\!0\}$
  - An example: $n = 8$ **T**=(1,0,0,1,0,1,1,0) **Y**=(1,1,0,1,0,1,0,1)
    - Calculate D

Formalizing the Design and Analysis

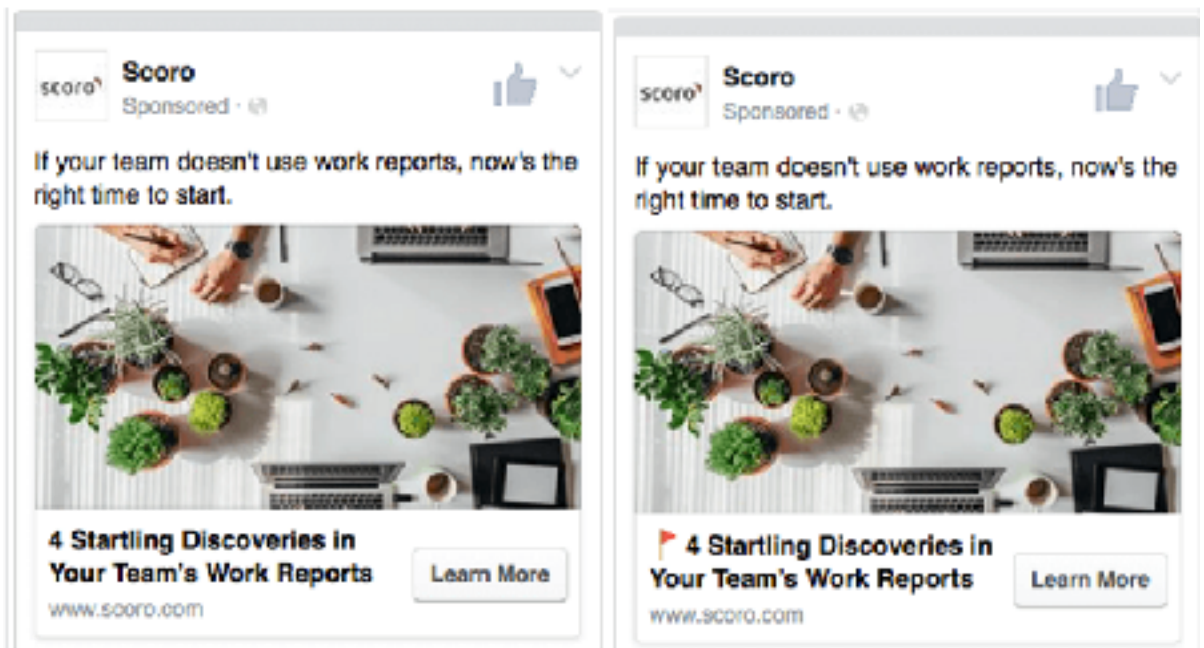
Table 2.3. Potential Outcome Framework of Causal Inference.

Résumé i	*Black-sounding name* T_i	*Callback* $Y_i(1)$	$Y_i(0)$	*Age*	*Education*
1	1	1	$\boxed{?}$	20	college
2	0	$\boxed{?}$	0	55	high school
3	0	$\boxed{?}$	1	40	graduate school
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	1	0	$\boxed{?}$	62	college

- Please follow the textbook example as an exercise at home (very easy)
  - https://github.com/kosukeimai/qss
    - https://github.com/kosukeimai/qss/tree/master/CAUSALITY

Business Applications: We all Become Subjects Everyday

- A/B Testing (bucket tests or split-run testing) HBR article
  - Service improvement protocol for any company these days
  - Multiple treatment conditions:



Control (no change) Test (change)

  - Outcome: e.g. Click-through rates
- A/B Testing ∈ RCTs: Cost-effective design & Accurate inference
  - Companies have their own platform/language (e.g. PlanOut)

Why RCTs over Observational Studies?

e.g. A Pair of Facebook Emotional Contagion Study

- Observational
- Experimental

e.g. Social Pressure Turnout Experiment

Research Question: Emotional Contagion Hypothesis

- Effect of your friends' FB wall posting on your expressed emotion

Positive post

Great foods!
I love vegetables.

Negative post

I hate vegetables..
I'll not come here again.

Your emotion

- Implication: Large-scale global synchrony/diffusion of emotion

Observational Version of Facebook Emotion Study

Detecting Emotional Contagion in Massive Social Networks

Lorenzo Coviello[1], Yunkyu Sohn[2], Adam D. I. Kramer[3], Cameron Marlow[3], Massimo Franceschetti[1], Nicholas A. Christakis[4,5], James H. Fowler[2,6]*

- Next week ➠ Observational Studies (things get **super complicated**)
  - Non-experimental study using spontaneous user activities
  - 1,180 days of observation of millions of Facebook users in US
  - Advanced statistical methods to match over
    - weather: precipitation, temperature, ...
    - user demographic characteristics
    - article length: **44** pages in total.

Experimental Version of Facebook Emotion Study

- RCT: 3 mil posts; 155,000 users
- Manipulating FB wall post content exposure probability by sentiment
- 3 page paper with a **single figure** HOW??
  - Thanks to The POWER of RCT: sole effect of treatment identified

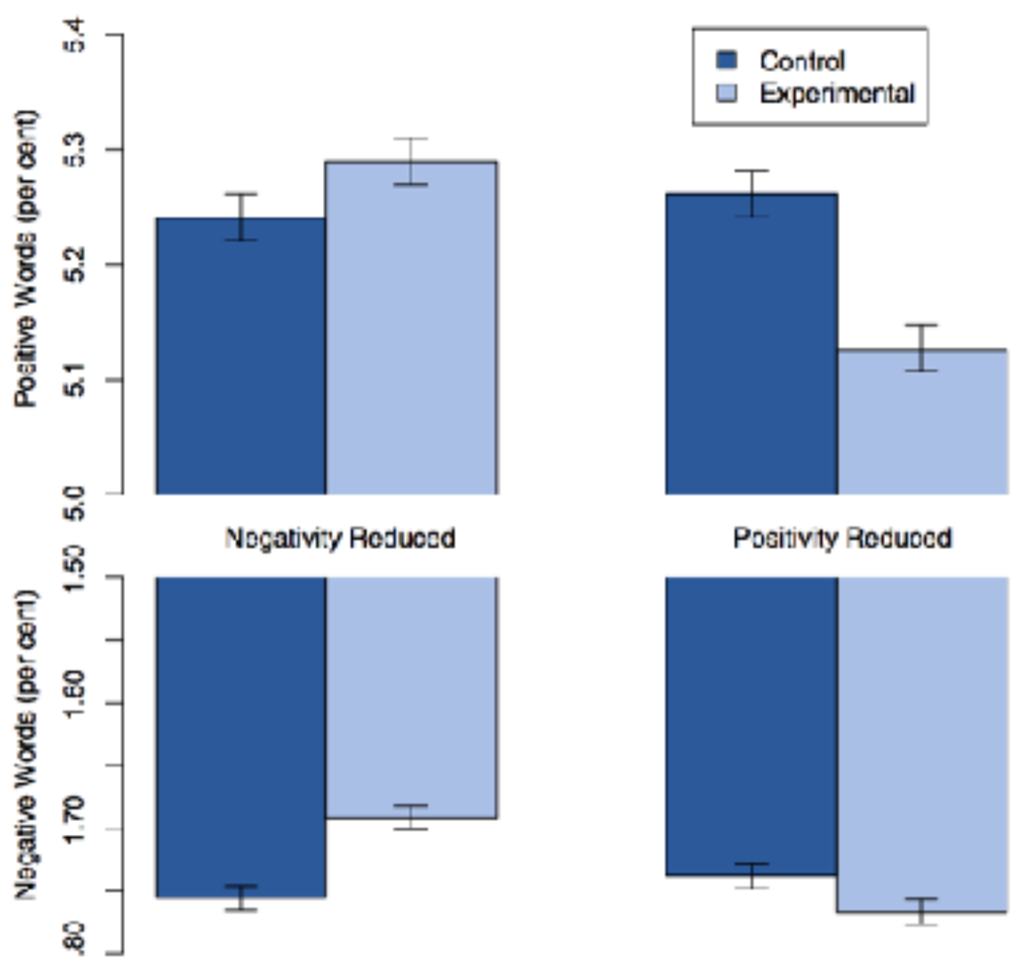


Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.

Kramer, Guillory, and Hancock (2014)

Drawbacks of Large-scale Field Experiments: Ethics

theguardian

Home | UK | World | Sport | Football | Opinion | Culture | Economy |

News › Technology › Facebook

Facebook's secret mood experiment: have you lost trust in the social network?

Secret study in which people's Facebook posts were moved to influence moods has angered users. Have you lost trust in the network?

Carmen Fishwick
theguardian.com, Monday 30 June 2014 09.06 BST

What do you think? Has the study worried you? And have you lost trust in the social network? Vote in our poll and share your thoughts in the thread below. We'll move a selection above the line

Were you surprised to learn about Facebook's study?

39% Yes

61% No

Have you lost trust in the social network?

84% Yes

16% No

Do you think you had your home feed manipulated?

23% yes

13% No

63% Don't know

Are you considering closing your account after learning of the experiment?

66% Yes

34% No

source

Social Pressure Turnout Experiment (Book Chapter 2.4.2)

- Get-out-the-vote (GOTV) messaged postcard with social pressure
  - naming and shaming strategy

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY – VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	________
995 JENNIFER KAY SMITH		Voted	________
9997 RICHARD B JACKSON		Voted	________
9999 KATHY MARIE JACKSON		Voted	________

Social Pressure Turnout Experiment (Book Chapter 2.4.2)

- Research Question: Does peer pressure facilitate turnout?
- Unit: voter in a primary election in the state of Michigan
- Treatment variable T: GOTV + social pressure messaged postcard sent
- Treatment group (treated units): voters receiving certain postcards
- Control group (untreated units): voters receiving none
- Outcome variable (response variable) Y : turnout (publicly available)
- Potential outcomes: *turnout*(received) and *turnout*(not-received)
- Causal effect: Difference in the sample means estimator
- Further details: 2 additional treatment conditions
  - GOTV w/o social pressure but civic duty messaged postcard
  - Hawthorne effect group
    - YOU ARE BEING STUDIED! w/o social pressure

Drawbacks of Large-scale Field Experiments: Ethics

The New York Times

TheUpshot

A STUDY IN CONTROVERSY

Professors' Research Project Stirs Political Outrage in Montana

By Derek Willis

Oct. 28, 2014

31

- Similar mail experiments by Stanford & Dartmouth political scientists
  - sent to 102,780 Montana registered households (15% registered voters!)
  - Nonpartisan Montana supreme court justice elections
  - Including information on ideological positions of candidates
  - Law suits: University presidents' apologies + $13k settlement

source

Things To Consider When Designing a Study: Validity

- To experiment or Observe?
  - Internal validity
    - whether causal assumptions are satisfied in the study
    - experiments: High
    - observational (non-experimental) studies: Low
  - External validity
    - the extent to which the conclusions can be generalized beyond a particular study
    - experiments: Low
    - observational (non-experimental) studies: High

Next Week

- All the obstacles for making valid causal inference
  - when conducting observational studies
  - and partial remedies
  - Textbook Chapter 2.5 ~ 2.7

See you next week.