Lecture 1: Causality and Experiments
Modeling Social Data, Spring 2017
Columbia University

Min-heng Wang

April 23, 2017

1 Problems of Experiments

Random assignment is the gold standard for causal inference, but it has some limitations:
1. Small sample size
2. Researchers degree of freedom
3. Publication bias: only those who reach statistically significant could be published.
4. P hacking

Factors influence power:
N = sample size
Alpha = significance level
Effect size (how strong is the effect? i.e. Cohen's D)
P(significance — no effect) = "False Alarm"
Power = 1- Beta = chance of detecting a real effect if one exists.
P(significance — effect) = "Hit"

How to explain 95 percent confidence level:
if we replicate the experiments infinite time, 95 percent results would contain the parameter (true value).

2 Caveat and Limitations

1.randomization often is not feasible or ethical
2.experiments are costly in terms of time and money
3.its difficult to create convincing parallel worlds
4.inevitably people deviate from their random assignment

3 Natural Experiments

Sometimes we get lucky and nature effectively runs experiments for us e.g.:
1. As-if random: people are randomly exposed to water sources
2. Instrumental variables: a lottery influences military services:
IV influences treatment, but no association with the errors.
3. Regression-discontinuities:
idea: things change around an arbitrary chosen threshold.

4. Difference in differences:
idea: compare difference after a sudden change with trends in a control group

T C : Compliers
T T : Always treats
C C : Never treats

ATE = pcATEc + paATEa + pnATEn

(Fraction accept treatment in treatment group) (Fraction accept treatment in control group) = (Pc + Pa)
Pa = Pc

4 Natural Experiments limitations

1. Good natural experiments are hard to find
2. They rely on naturally-occurring event rather than controlled manipulation
3. Difficult to control for possible alternative explanations
4. Limited sample size

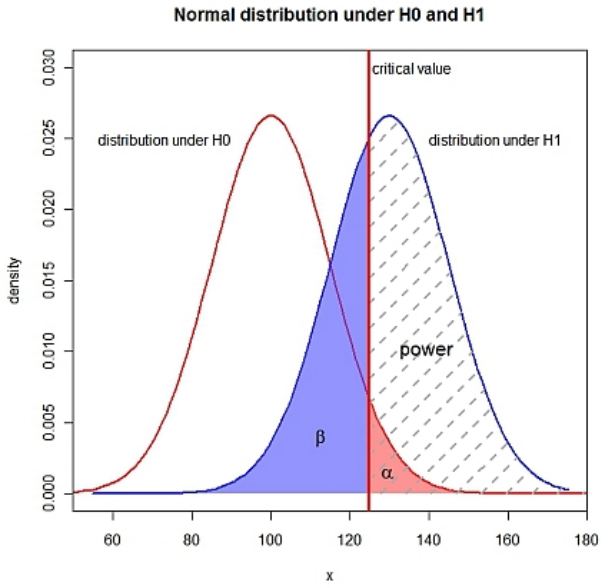

Figure 1: This figure's source is as below:
http://jamescaldwell.info/optimization/the-maths-behind-statistically-significant-sample-sizes/.

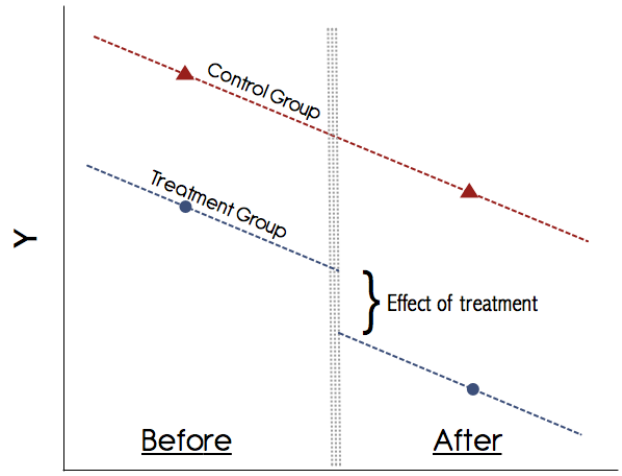

Figure 2: Example plot for Difference in Differences.source: https://i.stack.imgur.com/J7P3p.png

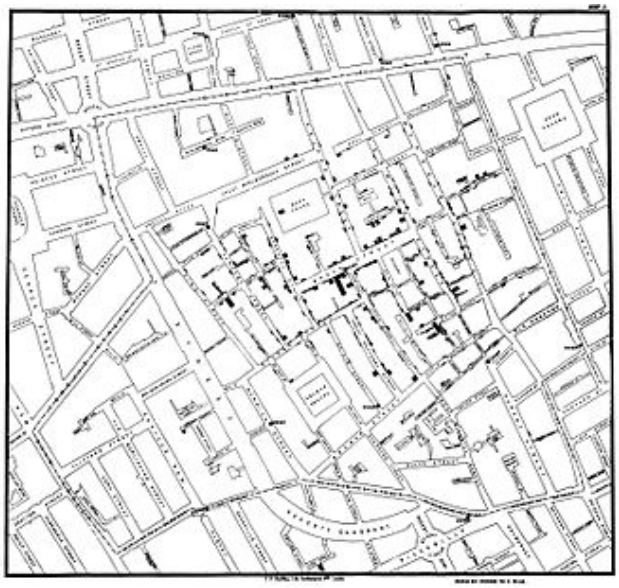

Figure 3: Example plot for natural experiment.
source: http://thelancet.com/journals/lancet/article/PIIS0140-6736(13)60830-2/fulltext?rss