Random Experiments

Prof.Mike Aguilar

https:/www.linkedin.com/in/mike-aguilar-econ

Random Experiments - Summary

- Random Experiments are void of selection bias, so a simple Difference in Means estimator between treated and untreated groups can elicit the treatment effect
- However, RCTs are the sample counterpart to Random Experiments, and as such selection bias might be present in a simple Difference in Means estimator
- For RCTs, care must be taken before the treatments (e.g. experiment design, assessing randomness) and after treatment (e.g. inference)

Housekeeping

- Throughout the slide deck you will see "Q", which indicates a question to you, the reader
- You will also see "A", which indicates the associated answer
- It is generally most efficient to learn this material through active participation. Whenever you encounter a "Q", be sure to try and develop your answer before turning the page to the provided "A" answer
- I recommend reading through these slides before engaging in associated coding exercises
- This slide deck is part of a broader curriculum in causal estimation and inference
- Suggested pre-requisite: Potential outcomes framework

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

Does Insurance Improve Health?



Image created with Bing Image Creator
The following is adapted from the book Mastering 'Metrics

Meet Khuzdar



Meet Khuzdar



- He does have health insurance
- He is known to be relatively weak (e.g. poor constitution, gets sick often, etc..)
- We want to know the impact health insurance has on his health as measured by NHIS health index score (higher is better)

Fundamental Problem of Causal Inference

Q: Why is it difficult to estimate the impact of insurance for Khuzdar?

Fundamental Problem of Causal Inference

Q: Why is it difficult to estimate the impact of insurance for Khuzdar?

A: We don't see him with and without insurance.







Earth 1



Earth 0

Meet Maria



- She does not have health insurance
- She is known to be relatively strong (e.g. good constitution, doesn't gets sick often, etc..)

Comparing Khuzdar and Maria

Q: Why shouldn't we use Maria to determine impact of insurance on Khuzdar's health?

Comparing Khuzdar and Maria

Q: Why shouldn't we use Maria to determine impact of insurance on Khuzdar's health?

Khuzdar



VS

Maria



Comparing Khuzdar and Maria

Q: Why shouldn't we use Maria to determine impact of insurance on Khuzdar's health?



A: Maria is stronger than Khuzdar. Comparing their health isn't fair. We don't know if we are capturing the impact of insurance or inherent strength.

Randomization

Suppose we have the ability to randomly assign (no)insurance to Khuzdar, Maria, and many others.



Assigned to Treated Group

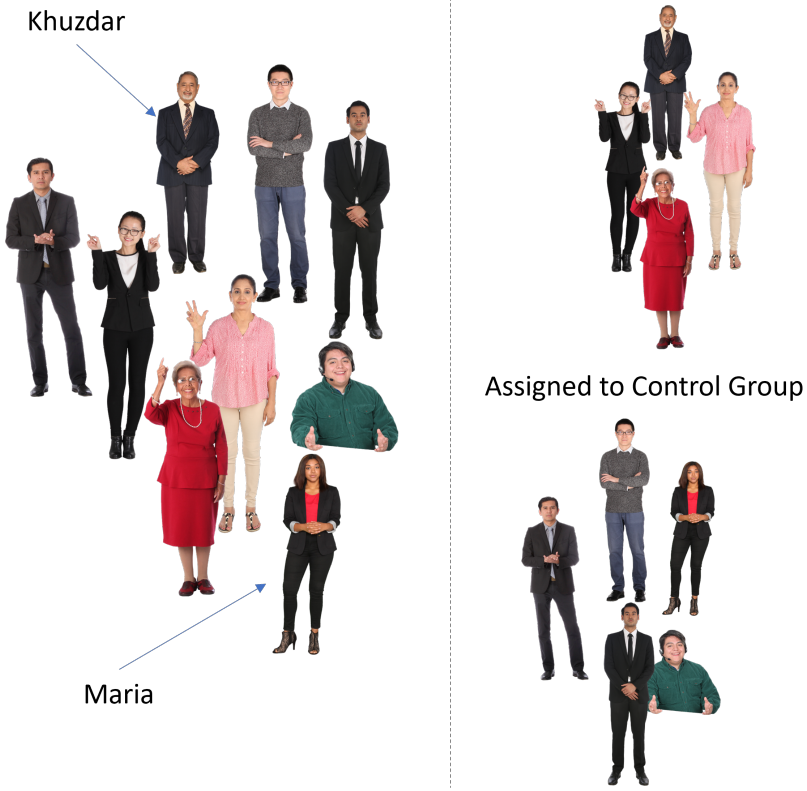


Assigned to Control Group



Randomization

Suppose we have the ability to randomly assign (no)insurance to Khuzdar, Maria, and many others.



Then we can compare the treated and control group's average health to determine the treatment effect.

Hierarchy of Causality

Recall the hierarchy of various settings to explore causality

- Controlled Experiments
- Random Experiments
- Natural Experiments
- Observational Studies

Hierarchy of Causality

Recall the hierarchy of various settings to explore causality

- Controlled Experiments
- Random Experiments
- Natural Experiments
- Observational Studies

Our focus here is on Random Experiments.

We will distinguish Random Experiments from Random Control Trials (RCTs) in order to make an important point.

Hierarchy of Causality

Recall the hierarchy of various settings to explore causality

- Controlled Experiments
- Random Experiments
- Natural Experiments
- Observational Studies

Our focus here is on Random Experiments.

We will distinguish Random Experiments from Random Control Trials (RCTs) in order to make an important point.

RCTs are the sample counterpart of Random Experiments.

Hierarchy of Causality

Recall the hierarchy of various settings to explore causality

- Controlled Experiments
- Random Experiments
- Natural Experiments
- Observational Studies

Our focus here is on Random Experiments.

We will distinguish Random Experiments from Random Control Trials (RCTs) in order to make an important point.

RCTs are the sample counterpart of Random Experiments. Though powerful, RCTs don't carry all the benefits of a Random Experiment.

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

Reminder on Potential Outcomes Framework

Recall the notation for the Potential Outcomes Framework for causal modeling

Y_{1i} : Potential outcome of Y for individual i in treated state

Y_{0i} : Potential outcome of Y for individual i in untreated state

$D_i = 1$: if individual i is observed as treated; 0 if untreated (aka control)

$Y_{0i}|D_i = 1$: Potential outcome of Y for individual i in the untreated state, given that i is observed to be treated

$Y_{0i}|D_i = 0$: Potential outcome of Y for individual i in the untreated state, given that i is observed to be untreated

$Y_{1i}|D_i = 1$: Potential outcome of Y for individual i in the treated state, given that i is observed to be treated

$Y_{1i}|D_i = 0$: Potential outcome of Y for individual i in the treated state, given that i is observed to be untreated

Reminder on Treatment Effects

Q: Can you define ATT,ATC,ATE,CATE?

Reminder on Treatment Effects

Q: Can you define ATT,ATC,ATE,CATE?

$$ATT:$$

$$ATC:$$

$$ATE:$$

$$CATE:$$

Reminder on Treatment Effects

Q: Can you define ATT,ATC,ATE,CATE?

$$
\begin{aligned}
ATT &: E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\
ATC &: \\
ATE &: \\
CATE &:
\end{aligned}
$$

Reminder on Treatment Effects

Q: Can you define ATT,ATC,ATE,CATE?

$$
\begin{aligned}
ATT &: E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\
ATC &: E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0] \\
ATE &: \\
CATE &:
\end{aligned}
$$

Reminder on Treatment Effects

Q: Can you define ATT,ATC,ATE,CATE?

$$
\begin{aligned}
ATT &: E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\
ATC &: E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0] \\
ATE &: E[Y_{1i} - Y_{0i}] \\
CATE &:
\end{aligned}
$$

Reminder on Treatment Effects

Q: Can you define ATT,ATC,ATE,CATE?

$$
\begin{aligned}
ATT &: E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\
ATC &: E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0] \\
ATE &: E[Y_{1i} - Y_{0i}] \\
CATE &: E[Y_{1i} - Y_{0i}|X_i = x]
\end{aligned}
$$

Independent Assignment

If truly random, then treatment should not be related to potential outcomes

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

Independent Assignment

If truly random, then treatment should not be related to potential outcomes

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

Q: What is the implication of independent assignment upon ATE,ATT,ATC?

Independent Assignment

If truly random, then treatment should not be related to potential outcomes

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

Q: What is the implication of independent assignment upon ATE,ATT,ATC?

A:

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

The implication is that

Independent Assignment

If truly random, then treatment should not be related to potential outcomes

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

Q: What is the implication of independent assignment upon ATE,ATT,ATC?

A:

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

The implication is that $ATT = ATE = ATC$.

Independent Assignment

If truly random, then treatment should not be related to potential outcomes

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

Q: What is the implication of independent assignment upon ATE,ATT,ATC?

A:

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

The implication is that $ATT = ATE = ATC$.

Whether or not you are actually treated doesn't impact your potential outcomes. So the treatment effect for the "Khuzdars" is the same as for the "Marias".

Conditional Independence

If truly random, then treatment effect should not covary with confounders

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

Conditional Independence

If truly random, then treatment effect should not covary with confounders

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

Q: What is the implication of conditional independence upon ATE,ATT,ATC,CATE?

Conditional Independence

If truly random, then treatment effect should not covary with confounders

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

Q: What is the implication of conditional independence upon ATE,ATT,ATC,CATE?

A:

$E[Y_{1i}|D_i, X_i] = E[Y_{1i}]$ and $E[Y_{0i}|D_i, X_i] = E[Y_{0i}]$

The implication is that

Conditional Independence

If truly random, then treatment effect should not covary with confounders

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

Q: What is the implication of conditional independence upon ATE,ATT,ATC,CATE?

A:

$E[Y_{1i}|D_i, X_i] = E[Y_{1i}]$ and $E[Y_{0i}|D_i, X_i] = E[Y_{0i}]$

The implication is that $ATT = ATE = ATC = CATE$

Conditional Independence

If truly random, then treatment effect should not covary with confounders

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

Q: What is the implication of conditional independence upon ATE,ATT,ATC,CATE?

A:

$E[Y_{1i}|D_i, X_i] = E[Y_{1i}]$ and $E[Y_{0i}|D_i, X_i] = E[Y_{0i}]$

The implication is that $ATT = ATE = ATC = CATE$

The treatment effects are the same for the "old" and the "young".

Calculating Treatment Effects

If truly random, we can use a simple difference to compute TE

Suppose we have perfect estimators of $\hat{Y}_1 \approx E[Y_1]$ and $\hat{Y}_0 \approx E[Y_0]$

Calculating Treatment Effects

If truly random, we can use a simple difference to compute TE

Suppose we have perfect estimators of $\hat{Y}_1 \approx E[Y_1]$ and $\hat{Y}_0 \approx E[Y_0]$

Then the TE can be calculated via a simple differencing

$$TE = \hat{Y}_1 - \hat{Y}_0$$

Calculating Treatment Effects

If truly random, we can use a simple difference to compute TE

Suppose we have perfect estimators of $\hat{Y}_1 \approx E[Y_1]$ and $\hat{Y}_0 \approx E[Y_0]$

Then the TE can be calculated via a simple differencing

$$TE = \hat{Y}_1 - \hat{Y}_0$$

There is no selection bias in this estimate due to the nature of the randomness.

Calculating Treatment Effects

If truly random, we can use a simple difference to compute TE

Suppose we have perfect estimators of $\hat{Y}_1 \approx E[Y_1]$ and $\hat{Y}_0 \approx E[Y_0]$

Then the TE can be calculated via a simple differencing

$$TE = \hat{Y}_1 - \hat{Y}_0$$

There is no selection bias in this estimate due to the nature of the randomness.

RCTs and observational settings cannot leverage this simple differencing approach without bias concerns.

The Magic of Random Experiments

- By virtue of their perfect randomness, Random Experiments i) have independent assignment of treatment and ii) treatment assignment is conditionally independent of outcomes
- This is in contrast to RCTs (and other settings like Regression Adjustment), which assume these conditions hold
- Implication is that DAG is not needed within a random experiment in order to elicit treatment effects

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

RCT as a Proxy for Random Experiments

- RCTs are the sample counterpart of Random Experiments
- There are myriad implementation considerations when running RCTs
- We will very briefly touch upon such considerations, but not delve into any detail. Our goal will be to analyze the results of an RCT, not implement an RCT in the field.

Randomization

Q: How could we manufacture randomness in our health insurance case study?

Randomization

Q: How could we manufacture randomness in our health insurance case study?

A: Here are a few examples

- Simple - flip a coin

Randomization

Q: How could we manufacture randomness in our health insurance case study?

A: Here are a few examples

- Simple - flip a coin
- Block - equal number treat/control

Randomization

Q: How could we manufacture randomness in our health insurance case study?

A: Here are a few examples

- Simple - flip a coin
- Block - equal number treat/control
- Stratified - randomize within strata of population

Randomization

Q: How could we manufacture randomness in our health insurance case study?

A: Here are a few examples

- Simple - flip a coin
- Block - equal number treat/control
- Stratified - randomize within strata of population
- Cluster - randomize across families/towns/etc..

Randomization

Q: How could we manufacture randomness in our health insurance case study?

A: Here are a few examples

- Simple - flip a coin
- Block - equal number treat/control
- Stratified - randomize within strata of population
- Cluster - randomize across families/towns/etc..
- Adaptive - assignment probability varies with prior participants

Don't dwell here. Our goal will be to analyze the results of an RCT, not implement an RCT in the field.

Compliance Design

Q: How could we enforce compliance in our case?

Compliance Design

Q: How could we enforce compliance in our case?

A:

- Brute Force – forced into compliance within treat/control groups
- Self Selection – randomize after individuals express interest in treatment
- Eligibility – i) determine who is eligible 2) randomly select treat and control 3) permit treated (only) to self select into treatment
- Encouragement – i) determine who is eligible 2) randomly select treat and control 3) encourage treated to participate 3) permit treated and control to self select into treatment

Don't dwell here. Our goal will be to analyze the results of an RCT, not implement an RCT in the field.

Other considerations

- Cadence: In a crossover design participants receive (no-)treatment at random time intervals
- Number of Treatments: In a factorial design Multiple treatments tested at once
- Blinding (masking): Withholding assignment from participants
- Concealment: Withholding assignment from researcher before treatment
- Pre-register: Can publicly document research plan and hypotheses before implementation to mitigate "look ahead bias" and increase transparency

Don't dwell here. Our goal will be to analyze the results of an RCT, not implement an RCT in the field.

Advantages and Disadvantages of RCTs

Advantages

- Non parametric, implying few assumptions needed
- Unbiased if done well

Advantages and Disadvantages of RCTs

Advantages

- Non parametric, implying few assumptions needed
- Unbiased if done well

Disadvantages

Advantages and Disadvantages of RCTs

Advantages

- Non parametric, implying few assumptions needed
- Unbiased if done well

Disadvantages

- Costs: can be costly to implement
- Lengthy: can take long time to see results, analyze, and make decisions
- Legal: may not be legal to withhold/impose treatment
- Ethical: may not be ethical to withhold/impose treatment
- Feasible: may not be feasible to withhold/impose treatment
- Extrapolation: dangerous if extrapolate results to another different population

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

Diagnostics for Covariate Balance

We want to ensure that those we intend to treat "look like" those we intend not to treat. During the design phase we can deploy some diagnostics to compare the intended treat/control groups.

Diagnostics for Covariate Balance

We want to ensure that those we intend to treat "look like" those we intend not to treat. During the design phase we can deploy some diagnostics to compare the intended treat/control groups.
For each covariate we can

Diagnostics for Covariate Balance

We want to ensure that those we intend to treat "look like" those we intend not to treat. During the design phase we can deploy some diagnostics to compare the intended treat/control groups.
For each covariate we can

- Run t-test of equal means of X for treat vs control
- Compare density plots of X for treat vs control
- Compare QQ plots of X for treat vs control
- Run density similarity tests (e.g. KS test) of X for treat vs control

Diagnostics for Covariate Balance

We want to ensure that those we intend to treat "look like" those we intend not to treat. During the design phase we can deploy some diagnostics to compare the intended treat/control groups.
For each covariate we can

- Run t-test of equal means of X for treat vs control
- Compare density plots of X for treat vs control
- Compare QQ plots of X for treat vs control
- Run density similarity tests (e.g. KS test) of X for treat vs control

Any differences in the treat/control groups along these measures signals covariate imbalance, non random assignment and trouble for our RCT.

Diagnostics for Covariate Balance

We want to ensure that those we intend to treat "look like" those we intend not to treat. During the design phase we can deploy some diagnostics to compare the intended treat/control groups.
For each covariate we can

- Run t-test of equal means of X for treat vs control
- Compare density plots of X for treat vs control
- Compare QQ plots of X for treat vs control
- Run density similarity tests (e.g. KS test) of X for treat vs control

Any differences in the treat/control groups along these measures signals covariate imbalance, non random assignment and trouble for our RCT.
Q: Why wouldn't you run need to these tests in a Random Experiment?

Diagnostics for Covariate Balance

We want to ensure that those we intend to treat "look like" those we intend not to treat. During the design phase we can deploy some diagnostics to compare the intended treat/control groups.
For each covariate we can

- Run t-test of equal means of X for treat vs control
- Compare density plots of X for treat vs control
- Compare QQ plots of X for treat vs control
- Run density similarity tests (e.g. KS test) of X for treat vs control

Any differences in the treat/control groups along these measures signals covariate imbalance, non random assignment and trouble for our RCT.
Q: Why wouldn't you run need to these tests in a Random Experiment?
A: We wouldn't need to because treat and control are identical (aside from treatment) by design.

Resolve Imbalance

Q: What should you do if covariate imbalance is found?

A: (partial list)

Resolve Imbalance

Q: What should you do if covariate imbalance is found?

A: (partial list)

- Rethink your randomization scheme
- Rerandomize
- Stratify along the covariate(s) that is causing the imbalance and rerandomize within each strata

Resolve Imbalance

Q: What should you do if covariate imbalance is found?

A: (partial list)

- Rethink your randomization scheme
- Rerandomize
- Stratify along the covariate(s) that is causing the imbalance and rerandomize within each strata

RCTs can be expensive to implement. It is paramount that these diagnostics are completed BEFORE treatment is provided.

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

Difference in Means

Similar to the Random Experiment approach to estimating treatment effects, we can use DiM (Difference in Means) on observed outcomes

$$(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 0) = \big\{(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 1)\big\} + \big\{(\bar{Y}_0|D_i = 1) - (\bar{Y}_0|D_i = 0)\big\}$$

Difference in Means

Similar to the Random Experiment approach to estimating treatment effects, we can use DiM (Difference in Means) on observed outcomes

$$(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 0) = \left\{(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 1)\right\} + \left\{(\bar{Y}_0|D_i = 1) - (\bar{Y}_0|D_i = 0)\right\}$$

Q: Can you interpret the expression in the second set of curly brackets?

Difference in Means

Similar to the Random Experiment approach to estimating treatment effects, we can use DiM (Difference in Means) on observed outcomes

$$(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 0) = \{(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 1)\} + \{(\bar{Y}_0|D_i = 1) - (\bar{Y}_0|D_i = 0)\}$$

Q: Can you interpret the expression in the second set of curly brackets?

A:

- $(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 1) = \widehat{ATT}$
- $(\bar{Y}_0|D_i = 1) - (\bar{Y}_0|D_i = 0) =$ Selection Bias

Difference in Means

Similar to the Random Experiment approach to estimating treatment effects, we can use DiM (Difference in Means) on observed outcomes

$$(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 0) = \{(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 1)\} + \{(\bar{Y}_0|D_i = 1) - (\bar{Y}_0|D_i = 0)\}$$

Q: Can you interpret the expression in the second set of curly brackets?

A:

- $(\bar{Y}_1|D_i = 1) - (\bar{Y}_0|D_i = 1) = \widehat{ATT}$
- $(\bar{Y}_0|D_i = 1) - (\bar{Y}_0|D_i = 0) =$ Selection Bias

Selection bias wasn't a concern in a Random Experiment, but may be in an RCT.

The Damage of Selection Bias

Suppose the DGP is $Y = \alpha + \beta D + \gamma X$

$$\begin{aligned}\bar{Y}_1 &= \alpha + \beta + \gamma \bar{X}|_{D=1} \\ \bar{Y}_0 &= \alpha + \gamma \bar{X}|_{D=0}\end{aligned}$$

The Damage of Selection Bias

Suppose the DGP is $Y = \alpha + \beta D + \gamma X$

$$\begin{aligned}\bar{Y}_1 &= \alpha + \beta + \gamma \bar{X}|_{D=1} \\ \bar{Y}_0 &= \alpha + \gamma \bar{X}|_{D=0}\end{aligned}$$

Then, our DiM estimate $\bar{Y}_1 - \bar{Y}_0 = \beta + \gamma(\bar{X}|_{D=1} - \bar{X}|_{D=0})$.

The Damage of Selection Bias

Suppose the DGP is $Y = \alpha + \beta D + \gamma X$

$$\begin{aligned}\bar{Y}_1 &= \alpha + \beta + \gamma \bar{X}|_{D=1} \\ \bar{Y}_0 &= \alpha + \gamma \bar{X}|_{D=0}\end{aligned}$$

Then, our DiM estimate $\bar{Y}_1 - \bar{Y}_0 = \beta + \gamma(\bar{X}|_{D=1} - \bar{X}|_{D=0})$.

Q: What does $(\bar{X}|_{D=1} - \bar{X}|_{D=0})$ mean and what is the implication?

The Damage of Selection Bias

Suppose the DGP is $Y = \alpha + \beta D + \gamma X$

$$\begin{aligned}\bar{Y}_1 &= \alpha + \beta + \gamma \bar{X}|_{D=1} \\ \bar{Y}_0 &= \alpha + \gamma \bar{X}|_{D=0}\end{aligned}$$

Then, our DiM estimate $\bar{Y}_1 - \bar{Y}_0 = \beta + \gamma(\bar{X}|_{D=1} - \bar{X}|_{D=0})$.

Q: What does $(\bar{X}|_{D=1} - \bar{X}|_{D=0})$ mean and what is the implication?
A: The average values of X within the treated group might be different than in the control group.

The Damage of Selection Bias

Suppose the DGP is $Y = \alpha + \beta D + \gamma X$

$$\begin{aligned}\bar{Y}_1 &= \alpha + \beta + \gamma \bar{X}|_{D=1} \\ \bar{Y}_0 &= \alpha + \gamma \bar{X}|_{D=0}\end{aligned}$$

Then, our DiM estimate $\bar{Y}_1 - \bar{Y}_0 = \beta + \gamma(\bar{X}|_{D=1} - \bar{X}|_{D=0})$.

Q: What does $(\bar{X}|_{D=1} - \bar{X}|_{D=0})$ mean and what is the implication?
A: The average values of X within the treated group might be different than in the control group. Our DiM is then a biased estimate of the TE.

Randomization and Selection Bias

Q: How does randomization alleviate selection bias?

Randomization and Selection Bias

Q: How does randomization alleviate selection bias?

A: Under random assignment $D_i \perp \{Y_1, Y_0, X_i\}$ the treated and control groups come from the same underlying distribution.

Randomization and Selection Bias

Q: How does randomization alleviate selection bias?

A: Under random assignment $D_i \perp \{Y_1, Y_0, X_i\}$ the treated and control groups come from the same underlying distribution.

If our outcome is built from X, eg $Y = \alpha + \beta D + \gamma X$, then $D_i \perp X_i$, suggesting $(\bar{X}|_{D=1} - \bar{X}|_{D=0}) = 0$.

Randomization and Selection Bias

Q: How does randomization alleviate selection bias?

A: Under random assignment $D_i \perp \{Y_1, Y_0, X_i\}$ the treated and control groups come from the same underlying distribution.

If our outcome is built from X, eg $Y = \alpha + \beta D + \gamma X$, then $D_i \perp X_i$, suggesting $(\bar{X}|_{D=1} - \bar{X}|_{D=0}) = 0$.

In our case study, if X was age, then this condition means that since we randomly assigned treatment, the average age of those with insurance is the same as those without.

Randomization and Selection Bias

Q: How does randomization alleviate selection bias?

A: Under random assignment $D_i \perp \{Y_1, Y_0, X_i\}$ the treated and control groups come from the same underlying distribution.

If our outcome is built from X, eg $Y = \alpha + \beta D + \gamma X$, then $D_i \perp X_i$, suggesting $(\bar{X}|_{D=1} - \bar{X}|_{D=0}) = 0$.

In our case study, if X was age, then this condition means that since we randomly assigned treatment, the average age of those with insurance is the same as those without.

This random assignment is an assumption of RCTs.

RCT and OLS

We can use OLS, rather than DiM to compute the TE within an RCT

Set up a regression model with typical assumptions on the error

$$Y = \alpha + \beta D + e$$

RCT and OLS

We can use OLS, rather than DiM to compute the TE within an RCT

Set up a regression model with typical assumptions on the error

$$Y = \alpha + \beta D + e$$

Q: How derive the TE?

RCT and OLS

We can use OLS, rather than DiM to compute the TE within an RCT

Set up a regression model with typical assumptions on the error

$$Y = \alpha + \beta D + e$$

Q: How derive the TE?
A: $ATE = E[Y|D = 1] - E[Y|D = 0] = \beta$

RCT and OLS

We can use OLS, rather than DiM to compute the TE within an RCT

Set up a regression model with typical assumptions on the error

$$Y = \alpha + \beta D + e$$

Q: How derive the TE?
A: $ATE = E[Y|D = 1] - E[Y|D = 0] = \beta$

Q: Why would anyone bother with OLS if DiM is equivalent?

RCT and OLS

We can use OLS, rather than DiM to compute the TE within an RCT

Set up a regression model with typical assumptions on the error

$$Y = \alpha + \beta D + e$$

Q: How derive the TE?
A: $ATE = E[Y|D = 1] - E[Y|D = 0] = \beta$

Q: Why would anyone bother with OLS if DiM is equivalent?
A: We could include covariates in the regression (ala Regression Adjustment), which tends to reduce the sampling noise and thus increase precision.

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

Conventional Significance of TE

DiM (Difference in Means)

Conventional Significance of TE

DiM (Difference in Means)

- We can test the significance of TE via a two sample t-test
- $H_0 : \bar{Y}_1 - \bar{Y}_0 = 0$ vs $H_a : \bar{Y}_1 - \bar{Y}_0 \neq 0$
- If reject, then TE is significant
- Given randomization, the densities of the treated and control groups should be similar and so we can use equal variances
- Prudence suggests also running with unequal variances

OLS

Conventional Significance of TE

DiM (Difference in Means)

- We can test the significance of TE via a two sample t-test
- $H_0 : \bar{Y}_1 - \bar{Y}_0 = 0$ vs $H_a : \bar{Y}_1 - \bar{Y}_0 \neq 0$
- If reject, then TE is significant
- Given randomization, the densities of the treated and control groups should be similar and so we can use equal variances
- Prudence suggests also running with unequal variances

OLS

- We can test the significance of TE via t-test on treatment indicator
- $H_0 : \beta = 0$ vs $H_a : \beta \neq 0$
- If reject, the TE is significant
- Given randomization, we should not need to worry about issues like heteroscedasticity
- Prudence suggests we probe the existence of such issues

Outline

1. Motivation - Case Study
2. Random Experiments
3. Randomized Control Trials
   - Overview
   - Assessing Randomness
   - Selection Bias
   - Inference
     - Randomization Inference

Randomization Inference

- Aforementioned tests require various parametric assumptions (e.g. Gaussianity, independence, etc..), but RCTs are inherently nonparametric
- Aforementioned tests also rely upon asymptotic properties
- Randomization inherent in RCTs permits an alternative

Randomization Inference

- Aforementioned tests require various parametric assumptions (e.g. Gaussianity, independence, etc..), but RCTs are inherently nonparametric
- Aforementioned tests also rely upon asymptotic properties
- Randomization inherent in RCTs permits an alternative
- Fisher (1935) and Rosenbaum (2002) exploit knowledge of the random assignment to build an exact reference distribution for the observed RCT trial
- Sometimes called placebo test, permutation test, or Fisherian inference

Randomization Inference

Fisher's Null

The conventional modes of inference considered averages across individuals

$$H_0 : \bar{Y}_1 - \bar{Y}_0 = 0 \ vs \ H_a : \bar{Y}_1 - \bar{Y}_0 \neq 0$$

Randomization Inference

Fisher's Null

The conventional modes of inference considered averages across individuals

$$H_0 : \bar{Y}_1 - \bar{Y}_0 = 0 \ vs \ H_a : \bar{Y}_1 - \bar{Y}_0 \neq 0$$

Fisher's approach searches for TE on any unit in the population, not just averages

$$H_0 : Y_{1i} = Y_{0i}, \forall i = 1, \cdots, N \ vs \ H_a : \exists i \ s.t. Y_{1i} \neq Y_{0i}$$

Observed Data

Q: Can you compute the ATE from this RCT?

ID	D	Y
Khuzdar	1	10
Maria	0	11
John	1	12
Sangeeta	0	13

Observed Data

Q: Can you compute the ATE from this RCT?

ID	D	Y
Khuzdar	1	10
Maria	0	11
John	1	12
Sangeeta	0	13

A: $\widehat{ATE} = \bar{Y}_1 - \bar{Y}_0 = avg(10, 12) - avg(11, 13) = -1$

Observed Data

Q: Can you compute the ATE from this RCT?

ID	D	Y
Khuzdar	1	10
Maria	0	11
John	1	12
Sangeeta	0	13

A: $\widehat{ATE} = \bar{Y}_1 - \bar{Y}_0 = avg(10, 12) - avg(11, 13) = -1$

The goal now is to create a confidence interval around this estimate.

Create a pseduo sample by redesignating who was treated

Q: Can you compute the ATE from this pseudo RCT trial?

Notice the Treatment dummy D is different

ID	D	Y
Khuzdar	0	10
Maria	1	11
John	0	12
Sangeeta	1	13

Create a pseduo sample by redesignating who was treated

Q: Can you compute the ATE from this pseudo RCT trial?

Notice the Treatment dummy D is different

ID	D	Y
Khuzdar	0	10
Maria	1	11
John	0	12
Sangeeta	1	13

A: $\widehat{ATE} = \bar{Y}_1 - \bar{Y}_0 = avg(11, 13) - avg(10, 12) = 1$

Created another pseduo sample

Q: Can you compute the ATE from this pseudo RCT trial?

ID	D	Y
Khuzdar	0	10
Maria	0	11
John	0	12
Sangeeta	1	13

Created another pseduo sample

Q: Can you compute the ATE from this pseudo RCT trial?

ID	D	Y
Khuzdar	0	10
Maria	0	11
John	0	12
Sangeeta	1	13

A: $\widehat{ATE} = \bar{Y}_1 - \bar{Y}_0 = avg(13) - avg(10, 11, 12) = 2$

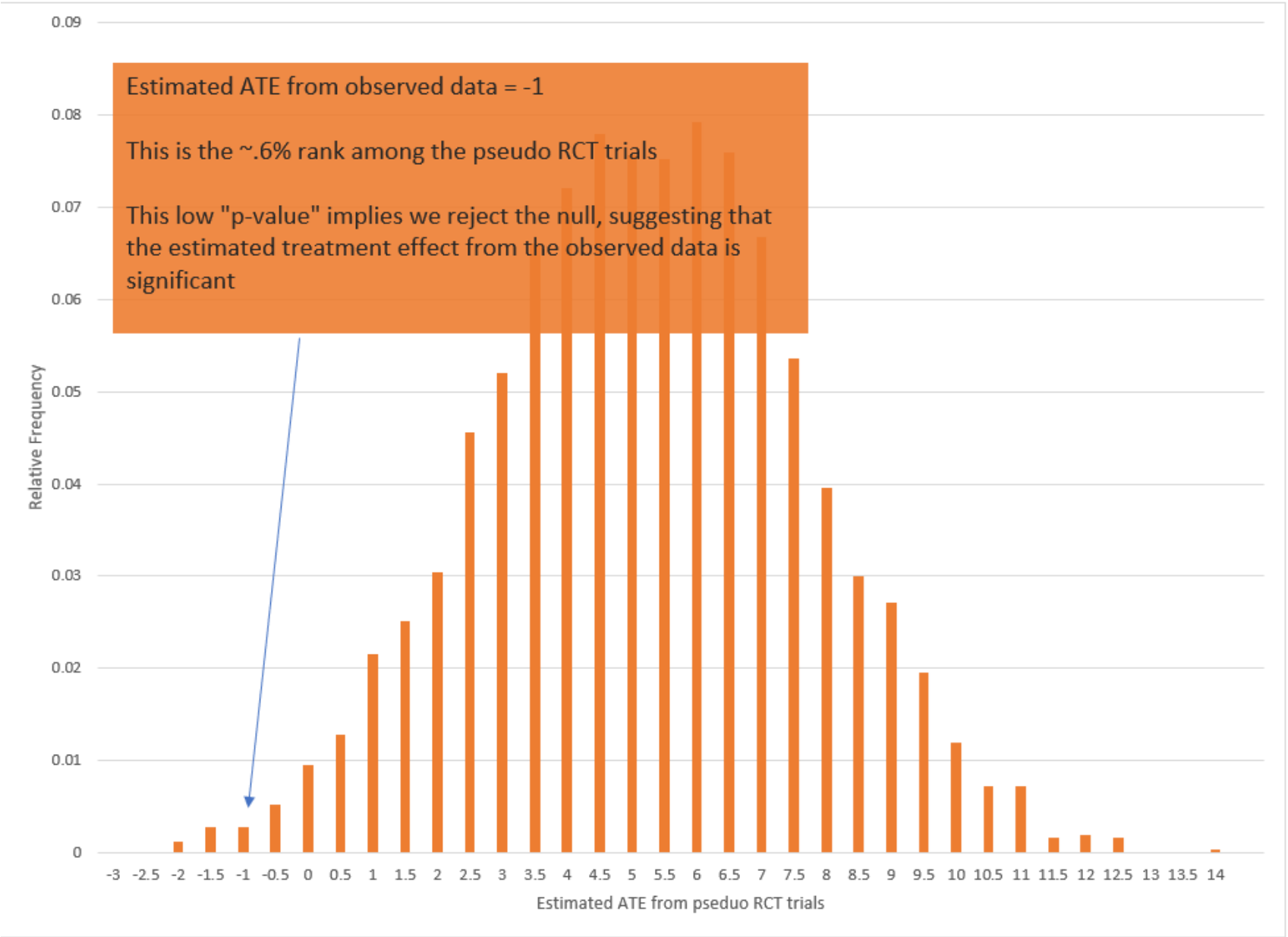
Fisher's Exact pValue

Algorithm

- Continue generating many pseudo samples
- Compute the percentile rank of the estimated TE from the observed RCT trial within the density of pseudo trials
- Use this percentile rank as the p-value for Fisher's test. e.g. If p-value $< 5\%$ then reject, implying that TE is significant

Fisher's Exact pValue

Applied to our case study



Random Experiments - Summary

- Random Experiments are void of selection bias, so a simple Difference in Means estimator between treated and untreated groups can elicit the treatment effect
- However, RCTs are the sample counterpart to Random Experiments, and as such selection bias might be present in a simple Difference in Means estimator
- For RCTs, care must be taken before the treatments (e.g. experiment design, assessing randomness) and after treatment (e.g. inference)