Intro to Causal Inference in Econometrics|

Lecture 1

Edward Vytlacil

Yale University, Department of Economics

June 22, 2023

About Me: Edward Vytlacil

Instructor: Professor Edward Vytlacil

- Economics Ph.D. from University of Chicago.
- Professor of Economics at Yale University,
  previously taught at Stanford, New York University, and Columbia University.
- Research focused on micro-econometrics and applied microe-econometrics, especially treatment effect estimation (causality) in education, labor and health economics. I have also published in corporate finance.
- Co-Editor of *Journal of Applied Econometrics*,
  Associate Editor of *Journal of Econometrics.*
- Founding Member and former Director of the
  International Association of Applied Econometrics.

Overview of Course: Applied Econometrics and Data Analysis

- Applied econometrics course focused on causal inference.
- Course will focus on estimation for causal analysis, including justification and implementation of estimation procedures and interpretation of results.
- Course will not focus on mathematical derivations.
- Lectures will focus on theory, while sections will focus on implementing methods using **R**.

Overview of Course: Applied Econometrics and Data Analysis

Outline of Topics:

- Intro to Causality and Randomized Controlled Trials (today),
- Linear Regression,
- Difference-in-Difference approach,
- Instrumental Variables.

Lab Sessions and Problem Sets:

- Applications for Lab Sessions:
    1. PROGRESA, effects of conditional cash transfers on school attendance in Mexico;
    2. Impact of tracking on student performance in schools in Kenya, based on "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya"
    3. Discrimination in hiring based on race and criminal background in United States, based on "Ban the Box, Criminal Records and Racial Discrimination: A Field Experiment";
    4. Effect of fertility on mother's labor supply, based on "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size".

Lab Sessions and Problem Sets:

- Applications for problem sets:
  1. Evaluate The Abecadarian Project, effect of early childhood intervention in United States,
  2. PROGRESA, effects of conditional cash transfers on school attendance in Mexico;
  3. Evaluate effect of weather insurance for Chinese farmers, based on "The Impact of Insurance Provision on Household Production and Financial Decisions."

R Coding

- I will include **R** code in lectures, your sections will focus on using **R** to implement methods from class , and your problem sets will include extensive **R** coding.
  - I will be using base-**R**, along with `ggplot` for graphics.
  - I will not be using `tidyverse` other than `ggplot`.

- If you are not already familiar with basic **R** coding, you should learn the basics on your own in order to follow the lectures and to be able to complete the problem sets.
  - Complete Problem Set 0, which is designed to get you started with **R**.
  - Work through Project 1 and Project 2 of Hands-On Programming with R, by Garrett Grolemund.
  - http://r-statistics.co/ggplot2-Tutorial-With-R.html is an excellent tutorial for `ggplot`.

Presumed Statistics Background

Course designed for students with some background in statistics, with students presumed to know such concepts as random variables, random sample versus population, expected value and variance, etc.

- Review posted preparatory materials on expected value and variance, with an application to asset diversification. (slides, handout)
- If you have no preparation in statistics, then you should additionally work through the first 12 units of Khan Academy: Statistics and Probability as soon as possible.

Course Requirements

1. Attend all lectures.
2. Three problem sets
   - Problem sets will focus on using **R** to analyze data.
   - Assigned after first, second and third classes, due one week after assigned.
   - You will work with your group to complete the problem set.
3. Two short quizzes
   - Assigned after second and fourth classes.
   - You must complete by yourself, not in a group.
4. Final project with oral presentation.
   - You will work with your group on the final project.

See your course syllabus for more details.

Course Grade

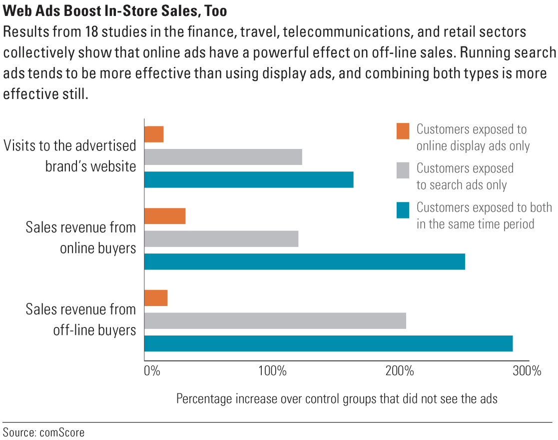
	Share of Course Grade
1. Participation	25%
2. Assignments	25%
3. Quizzes	10%
4. Final Project	40%

See your course syllabus for more details.

Overview for Today

- **Agenda: Causal Inference**
  - Counterfactual notation.
  - Treatment effects, treatment effect parameters.
  - Mean differences to estimate treatment effect parameters.
  - Selection Bias.
  - Sorting Gain.
- Stylized Example: On-Line Search Ads, from Harvard Business Review Study "The Off-Line Impact of Online Ads."

Example: HBR Study, “The Off-Line Impact of Online Ads”



Customers exposed to online search ads (compared to with no exposure):

- Spend more than twice as much on brand online,
- Spend more than three times as much on brand in store.

HBR study continued

What can we conclude from the HBR data?

What Magid Abram (CEO of comScore, author of study) says:

> *"... search ads, which appear only after a viewer has expressed interest in a subject by initiating a search, are generally more costly per impression than are display ads. Consistent with other kinds of advertising, using both types of ads in one campaign increases sales more than the two, added together, in separate campaigns."*

Is this convincing regarding the benefits of search ads?

HBR study continued

What can we conclude from the HBR data?

What Magid Abram (CEO of comScore, author of study) says:

> *"... search ads, which appear only after a viewer has expressed interest in a subject by initiating a search, are generally more costly per impression than are display ads. Consistent with other kinds of advertising, using both types of ads in one campaign increases sales more than the two, added together, in separate campaigns."*

Is this convincing regarding the benefits of search ads?

Key distinction:

1. Prediction – given that an individual was exposed to an ad, how much more do we predict they spend?
2. Causality – how much more does an individual spend because they were exposed to the ad?

Outcomes (notation)

Let:

- Y_i denote sales to consumer i,
- X_i denote an indicator for whether persion i saw a targeted ad,

$$X_i = \begin{cases} 1 & \text{if saw a targeted ad,} \\ 0 & \text{if did not see the ad.} \end{cases}$$

Outcomes (notation)

Let:

- Y_i denote sales to consumer i,
- X_i denote an indicator for whether persion i saw a targeted ad,

$$X_i = \begin{cases} 1 & \text{if saw a targeted ad,} \\ 0 & \text{if did not see the ad.} \end{cases}$$

We will refer to

- Y_i as the observed outcome;
- X_i as the treatment variable.

Prediction

From probability theory:

- Best predictor of Y_i given X_i is $E\left[Y_i \mid X_i\right]$, where
  - $E\left[Y_i \mid X_i\right]$ is defined as conditional expectation (conditional population mean) of Y_i given X_i;
  - best predictor is defined as the function of X_i that minimizes expected squared loss.
- In other words, if individual i . . .
  - was not shown the ad, then best predictor of her sales is $E\left[Y_i \mid X_i = 0\right]$;
  - was shown the ad, then best predictor of her sales is $E\left[Y_i \mid X_i = 1\right]$.
- $E\left[Y_i \mid X_i = 1\right] - E\left[Y_i \mid X_i = 0\right]$ is change in our best predictor of sales from an individual being shown the ad.

Prediction

- The HBR study estimates $E\left[Y_i \mid X_i = 1\right] - E\left[Y_i \mid X_i = 0\right]$, and interprets it as causal, but is it causal? Does it answer a question of interest to the advertiser?

- Note that it is comparing two groups of individuals, the prediction for those shown the ad and the prediction for those not shown the ad.

- To take different example:
  - Let Y_i denote i's income;
  - Let X_i denote whether i drives a luxury car;
  - Thus $E\left[Y_i \mid X_i\right]$ is best predictor of income given whether drives luxury car;
  - But we don't believe that an individual getting behind the wheel of a luxury car will *cause* their income to increase by $E\left[Y_i \mid X_i = 1\right] - E\left[Y_i \mid X_i = 0\right]$!

Stylized Example

Suppose we have data for an online store that sells football jerseys, and we wanted to know how exposure to an ad increased sales.

In this example:

- X_i is whether the person was shown a targeted ad;
- Y_i is resulting sales.

Stylized Example

i	X_i	Y_i
1	0	0
2	0	10
3	1	110
4	1	170

$$E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = \frac{110 + 170}{2} - \frac{0 + 10}{2} = 135.$$

Stylized Example

i	X_i	Y_i
1	0	0
2	0	10
3	1	110
4	1	170

$$E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = \frac{110 + 170}{2} - \frac{0 + 10}{2} = 135.$$

- Does being shown the ad *cause* spending to increase by 135? With targeted ads, perhaps those shown the ads for football jerseys are football fans, those not shown the ads are not football fans, and perhaps football fans would have spent more money than non-fans on football jerseys whether or not they had been shown the ad.

Stylized Example

i	X_i	Y_i
1	0	0
2	0	10
3	1	110
4	1	170

$$E\left[Y_i \middle| X_i = 1\right] - E\left[Y_i \middle| X_i = 0\right] = \frac{110 + 170}{2} - \frac{0 + 10}{2} = 135.$$

- Does being shown the ad *cause* spending to increase by 135? With targeted ads, perhaps those shown the ads for football jerseys are football fans, those not shown the ads are not football fans, and perhaps football fans would have spent more money than non-fans on football jerseys whether or not they had been shown the ad.
- To formalize, need notation for how much would have spent with and without being shown the ad. ..

Potential Outcomes (notation)

"potential outcomes" or "counterfactual outcomes" notation:

- $Y_{0,i}$ = sales to consumer i if she was **not shown an ad**,
- $Y_{1,i}$ = sales to consumer i if she was **shown an ad**.
- Observed outcome for individual *i*:

$$Y_i = Y_{0,i} + X_i(Y_{1,i} - Y_{0,i}) = \begin{cases} Y_{1,i} & \text{if } X_i = 1, \\ Y_{0,i} & \text{if } X_i = 0. \end{cases}$$

Potential Outcomes (notation)

“potential outcomes” or “counterfactual outcomes” notation:

- $Y_{0,i}$ = sales to consumer i if she was **not shown an ad**,
- $Y_{1,i}$ = sales to consumer i if she was **shown an ad**.
- Observed outcome for individual *i*:

$$Y_i = Y_{0,i} + X_i(Y_{1,i} - Y_{0,i}) = \begin{cases} Y_{1,i} & \text{if } X_i = 1, \\ Y_{0,i} & \text{if } X_i = 0. \end{cases}$$

Implicit in notation: No interaction across units.

- Called "Stable Unit Treatment Value Assumption” (*SUTVA*) in biostatistics.
- Plausible?

Potential Outcomes (notation)

"potential outcomes" or "counterfactual outcomes" notation:

- $Y_{0,i}$ = sales to consumer i if she was **not shown an ad**,
- $Y_{1,i}$ = sales to consumer i if she was **shown an ad**.
- Observed outcome for individual *i*:

$$Y_i = Y_{0,i} + X_i(Y_{1,i} - Y_{0,i}) = \begin{cases} Y_{1,i} & \text{if } X_i = 1, \\ Y_{0,i} & \text{if } X_i = 0. \end{cases}$$

Treatment effect for individual *i*: $Y_{1,i} - Y_{0,i}$.

- Traditional literature takes $Y_{1,i} - Y_{0,i}$ to be a constant, same for all *i*.
- Modern literature focuses on treatment effect heterogeneity, $Y_{1,i} - Y_{0,i}$ allowed to depend on *i*.

Potential Outcomes (notation)

"potential outcomes" or "counterfactual outcomes" notation:

- $Y_{0,i}$ = sales to consumer i if she was **not shown an ad**,
- $Y_{1,i}$ = sales to consumer i if she was **shown an ad**.
- Observed outcome for individual i:

$$Y_i = Y_{0,i} + X_i(Y_{1,i} - Y_{0,i}) = \begin{cases} Y_{1,i} & \text{if } X_i = 1, \\ Y_{0,i} & \text{if } X_i = 0. \end{cases}$$

We observe in our data (Y_i, X_i), we do not observe $(Y_{1,i}, Y_{0,i}, X_i)$.

- We observe either $Y_{0,i}$ (if not treated) or $Y_{1,i}$ (if treated), never both, and thus never $Y_{1,i} - Y_{0,i}$.
- Referred to as "*the fundamental problem of causal inference*" by Holland (1986).

Stylized Example

What do we observe in the HBR study?

- X_i, whether the person was shown a targeted ad;

i	X_i	Y_i
1	0	0
2	0	10
3	1	110
4	1	170

Stylized Example

What do we observe in the HBR study?

- X_i, whether the person was shown a targeted ad;
- For people who were not shown the ad, we observe $Y_i = Y_{0,i}$ but not their $Y_{1,i}$.

i	$Y_{0,i}$	$Y_{1,i}$	X_i	Y_i
1	0	0	0	0
2	10	20	0	10
3			1	110
4			1	170

Stylized Example

What do we observe in the HBR study?

- X_i, whether the person was shown a targeted ad;
- For people who were not shown the ad, we observe $Y_i = Y_{0,i}$ but not their $Y_{1,i}$.
- For people who were shown the ad, we observe $Y_i = Y_{1,i}$ but not their $Y_{0,i}$

i	$Y_{0,i}$	$Y_{1,i}$	X_i	Y_i
1	0	0	0	0
2	10	20	0	10
3	100	110	1	110
4	150	170	1	170

Stylized Example

What do we observe in the HBR study?

- X_i, whether the person was shown a targeted ad;
- For people who were not shown the ad, we observe $Y_i = Y_{0,i}$ but not their $Y_{1,i}$.
- For people who were shown the ad, we observe $Y_i = Y_{1,i}$ but not their $Y_{0,i}$
- We do not observe $Y_{1,i} - Y_{0,i}$ for any individual.

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

Question of Interest?

What is the question of interest to the advertiser?

- Effect of ad on individual i: $Y_{1,i} - Y_{0,i}$?
  - On which individual? on all individuals?
  - Typically impossible to identify/estimate unless impose that treatment effect does not vary across individuals.
  - Even if we could estimate for each individual, might still want a summary measure . . .

Question of Interest?

What is the question of interest to the advertiser?

- Average effect of the ad: $E\left[Y_{1,i} - Y_{0,i}\right]$?
  - Answers the question: *suppose we chose a random person from the population and showed them an ad. How much more revenue should we expect to get from showing them the ad?*
  - $Y_{1,i} - Y_{0,i}$ is the *treatment effect* for consumer i, and we take the expected value of this across all consumers.
  - This expression is called the **average treatment effect**.
  - When would this parameter be of interest to the advertiser?

Question of Interest?

What is the question of interest to the advertiser?

- Average effect of the ad on those shown the ad: $E\left[Y_{1,i} - Y_{0,i} \mid X_i = 1\right]$?
  - Answers the question: *suppose we chose a random person from those shown the ad. How much more revenue should we expect to get from them being shown the ad?*
  - $Y_{1,i} - Y_{0,i}$ is the *treatment effect* for consumer i, and we take the expected value of this across those consumers who were shown the ad.
  - This expression is called the **average effect of treatment on the treated**.
  - When would this parameter be of interest to the advertiser?

Question of Interest?

What is the question of interest to the advertiser?

- Average effect of the ad on those not shown the ad: $E\left[Y_{1,i} - Y_{0,i} \mid X_i = 0\right]$?
  - Answers the question: *suppose we chose a random person from those not shown the ad. How much more revenue should we expect to get from them being shown the ad?*
  - $Y_{1,i} - Y_{0,i}$ is the *treatment effect* for consumer i, and we take the expected value of this across those consumers who were *not* shown the ad.
  - This expression is called the **average effect of treatment on the untreated**.
  - When would this parameter be of interest to the advertiser?

Question of Interest?

What is the question of interest to the advertiser?

- Other parameters are sometimes considered, e.g.:
  - Quantile treatment effects, e.g., $\text{Median}\left(Y_{1,i}\right)$-$\text{Median}\left(Y_{0,i}\right)$.
  - Quantiles of the treatment effects, e.g., $\text{Median}\left(Y_{1,i} - Y_{0,i}\right)$.
  - $\Pr\left\{Y_{1,i} \geq Y_{0,i}\right\}$.

Stylized Treatment Effects Example

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

- Mean Difference: $E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = \frac{110+170}{2} - \frac{0+10}{2} = 135.$

Stylized Treatment Effects Example

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

- Mean Difference: $E[Y_i | X_i = 1] - E[Y_i | X_i = 0] = \frac{110+170}{2} - \frac{0+10}{2} = 135.$
- Average treatment effect (ATE): $E[Y_1 - Y_0] = \frac{0+10+10+20}{4} = 10.$

Stylized Treatment Effects Example

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

- Mean Difference: $E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = \frac{110+170}{2} - \frac{0+10}{2} = 135.$
- Average treatment effect (ATE): $E[Y_1 - Y_0] = \frac{0+10+10+20}{4} = 10.$
- Average effect of treatment of the treated (TT):
  $E[Y_1 - Y_0 \mid X_i = 1] = \frac{10+20}{2} = 15.$

Stylized Treatment Effects Example

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

- Mean Difference: $E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = \frac{110+170}{2} - \frac{0+10}{2} = 135.$
- Average treatment effect (ATE): $E[Y_1 - Y_0] = \frac{0+10+10+20}{4} = 10.$
- Average effect of treatment of the treated (TT): $E[Y_1 - Y_0 \mid X_i = 1] = \frac{10+20}{2} = 15.$
- Average effect of treatment of the untreated (TUT): $E[Y_1 - Y_0 \mid X_i = 0] = \frac{0+10}{2} = 5.$

Stylized Treatment Effects Example

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

- Mean Difference: $E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = \frac{110+170}{2} - \frac{0+10}{2} = 135.$
- Average treatment effect (ATE): $E[Y_1 - Y_0] = \frac{0+10+10+20}{4} = 10.$
- Average effect of treatment of the treated (TT): $E[Y_1 - Y_0 \mid X_i = 1] = \frac{10+20}{2} = 15.$
- Average effect of treatment of the untreated (TUT): $E[Y_1 - Y_0 \mid X_i = 0] = \frac{0+10}{2} = 5.$
- In this example, mean difference does not correspond to any average of the treatment effects, does not answer any causal question. why not?

Selection Bias

Can decompose mean difference into TT and selection bias:

$$\underbrace{E\big[Y_i \big| X_i = 1\big] - E\big[Y_i \big| X_i = 0\big]}_{\text{Mean Differences}}$$

$$= E\big[Y_{1,i} \big| X_i = 1\big] - E\big[Y_{0,i} \big| X_i = 0\big]$$

using that $X_i = 1 \Rightarrow Y_i = Y_{1,i}$, and $X_i = 0 \Rightarrow Y_i = Y_{0,i}$.

Selection Bias

Can decompose mean difference into TT and selection bias:

$$\underbrace{E[Y_i|X_i = 1] - E[Y_i|X_i = 0]}_{\text{Mean Differences}}$$
$$= E[Y_{1,i}|X_i = 1] - E[Y_{0,i}|X_i = 0]$$
$$= (E[Y_{1,i}|X_i = 1] - E[Y_{0,i}|X_i = 1]) + (E[Y_{0,i}|X_i = 1] - E[Y_{0,i}|X_i = 0])$$

adding and subtracting the green terms.

Selection Bias

Can decompose mean difference into TT and selection bias:

$$\underbrace{E[Y_i|X_i=1]-E[Y_i|X_i=0]}_{\text{Mean Differences}}$$
$$\begin{aligned}
&= E[Y_{1,i}|X_i=1]-E[Y_{0,i}|X_i=0] \\
&= \left(E[Y_{1,i}|X_i=1]-E[Y_{0,i}|X_i=1]\right)+\left(E[Y_{0,i}|X_i=1]-E[Y_{0,i}|X_i=0]\right) \\
&= \underbrace{E[Y_{1,i}-Y_{0,i}|X_i=1]}_{\text{Treatment on the Treated}}+\underbrace{\left(E[Y_{0,i}|X_i=1]-E[Y_{0,i}|X_i=0]\right)}_{\text{Selection Bias}}.
\end{aligned}$$

Mean Differences = Treatment on the Treated + Selection Bias.
Mean differences doesn't in general equal average effect of treatment on the treated, unless no selection bias.

Stylized Treatment Effects Example

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

Mean Differences = Treatment on the Treated + Selection Bias

Mean Differences: $\boldsymbol{E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = \frac{110+170}{2} - \frac{0+10}{2} = 135,}$

Stylized Treatment Effects Example

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

Mean Differences = **Treatment on the Treated** + Selection Bias

Mean Differences: $E\left[Y_i \middle| X_i = 1\right] - E\left[Y_i \middle| X_i = 0\right] = 135,$

Treatment on the Treated: $\boldsymbol{E\left[Y_1 - Y_0 \mid X = 1\right] = \frac{10+20}{2} = 15,}$

Stylized Treatment Effects Example

i	$Y_{0,i}$	$Y_{1,i}$	$Y_{1,i} - Y_{0,i}$	X_i	Y_i
1	0	0	0	0	0
2	10	20	10	0	10
3	100	110	10	1	110
4	150	170	20	1	170

Mean Differences = Treatment on the Treated + **Selection Bias**

Mean Differences: $E\left[Y_i | X_i = 1\right] - E\left[Y_i | X_i = 0\right] = 135,$

Treatment on the Treated: $E\left[Y_1 - Y_0 \mid X = 1\right] = 15,$

Selection Bias: $\boldsymbol{E\left[Y_0 \mid X = 1\right] - E\left[Y_0 \mid X = 0\right] = \frac{100+150}{2} - \frac{0+10}{2} = 120.}$

Bias for Average Treatment Effect

What if our target parameter is the average treatment effect, $E\left[Y_1 - Y_0\right]$?

$$\underbrace{E\left[Y_i \middle| X_i = 1\right] - E\left[Y_i \middle| X_i = 0\right]}_{\text{Mean Differences}}$$

$$= E\left[Y_{1,i} - Y_{0,i} \middle| X_i = 1\right] + \left\{E\left[Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{0,i} \middle| X_i = 0\right]\right\}$$

Bias for Average Treatment Effect

What if our target parameter is the average treatment effect, $E[Y_1 - Y_0]$?

$$\underbrace{E[Y_i|X_i = 1] - E[Y_i|X_i = 0]}_{\text{Mean Differences}}$$

$$= E[Y_{1,i} - Y_{0,i}|X_i = 1] + \{E[Y_{0,i}|X_i = 1] - E[Y_{0,i}|X_i = 0]\}$$

$$= E[Y_{1,i} - Y_{0,i}] + \{E[Y_{1,i} - Y_{0,i}|X_i = 1] - E[Y_{1,i} - Y_{0,i}]\} + \{E[Y_{0,i}|X_i = 1] - E[Y_{0,i}|X_i = 0]\}$$

adding and subtracting the green terms.

Bias for Average Treatment Effect

What if our target parameter is the average treatment effect, $E\left[Y_1 - Y_0\right]$?

$$
\begin{aligned}
&\underbrace{E\left[Y_i \middle| X_i = 1\right] - E\left[Y_i \middle| X_i = 0\right]}_{\text{Mean Differences}} \\
&= E\left[Y_{1,i} - Y_{0,i} \middle| X_i = 1\right] + \left\{E\left[Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{0,i} \middle| X_i = 0\right]\right\} \\
&= E\left[Y_{1,i} - Y_{0,i}\right] + \left\{E\left[Y_{1,i} - Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{1,i} - Y_{0,i}\right]\right\} + \left\{E\left[Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{0,i} \middle| X_i = 0\right]\right\} \\
&= \underbrace{E\left[Y_{1,i} - Y_{0,i}\right]}_{\text{Average Treatment Effect}} + \underbrace{\left\{E\left[Y_{1,i} - Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{1,i} - Y_{0,i}\right]\right\}}_{\text{Sorting Gain}} + \underbrace{\left(E\left[Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{0,i} \middle| X_i = 0\right]\right)}_{\text{Selection Bias}}.
\end{aligned}
$$

Mean Differences = Average Treatment Effect + Sorting Gain + Selection Bias.

Mean differences doesn't in general equal average treatment effect, unless no sorting gain and no selection bias.

Summary

- $E[Y_{1,i} - Y_{0,i}]$ is average treatment effect;
- $E[Y_{1,i} - Y_{0,i}|X_i = 1]$ is treatment on the treated;
- $E[Y_{1,i} - Y_{0,i}|X_i = 1] - E[Y_{1,i} - Y_{0,i}]$ is sorting gain ;
- $E[Y_{0,i}|X_i = 1] - E[Y_{0,i}|X_i = 0]$ is selection bias,

Mean differences do not generally equal average treatment effect or treatment on the treated:

- Mean differences $=$ treatment on the treated $+$ selection bias,
- Mean differences $=$ average treatment effect $+$ sorting gain $+$ selection bias.

In this stylized example:

- Football fans buy jerseys;
- Football fans are also more likely to see ads;
- This "selection" effect inflates the estimated effect of advertising
  (fans buy jerseys because they like Football already, not because of the ad).

Randomized Controlled Trials (RCTs) help solve this selection problem – in idealized RCT, treatment is assigned randomly, and, assuming full compliance, there is no selection bias.

More Generally: Causal Inference, Treatment Effects

Let:

- $Y_{0,i}$ denote potential outcome without treatment,
  which we would observe if individual i is not treated;
- $Y_{1,i}$ denote potential outcome with treatment,
  which we would observe if individual i is treated;
- X_i denote dummy variable for receiving the treatment.
- $Y_i = Y_{0,i} + X_i\left(Y_{1,i} - Y_{0,i}\right)$ is observed outcome for individual i.
- We observe $\left(Y_i, X_i\right)$, we do not observe $\left(Y_{1,i}, Y_{0,i}, X_i\right)$.

Philosophical differences between bio-stat and economics on what can be a cause/treatment, what is a valid causal question (see, e.g., Heckman (2008) and Holland (1986)).

Counterfactual Notation (cont'd)

Implicit in notation: No interaction across units.

- Called "Stable Unit Treatment Value Assumption" (*SUTVA*) in biostatistics.
- Rules out general equilibrium effects, peer-effects, etc., in economics.
- Assumption not always appropriate, for example:
  - A large scale vaccination program, where an individual being vaccinated may prevent him from infecting non-vaccinated individuals.
  - A large scale increase in college-attendance, causing an increase in the supply of skilled labor and thus a decrease in the price of skilled labor.

Counterfactual Notation (cont'd)

What if SUTVA is violated?

- Researchers sometimes redefine unit of analysis in order to plausibly satisfy this assumption.
  - For example, define unit of analysis to be a village instead of an individual when studying vaccination programs.
- Recent active literature relaxing SUTVA in context of network analysis, allowing for spillovers.

Examples:

1. From motivating example, among individuals visiting particular website:
   - X_i is a dummy variable for being shown a targeted search ad for a given product;
   - $Y_{1,i},\ Y_{0,i}$ are whether the individual would buy the product if shown or not shown the ad.
2. Among patients with a particular illness:
   - X_i is dummy variable for taking medicine;
   - $Y_{1,i},\ Y_{0,i}$ are future health outcomes with and without the medicine.
3. Among those convicted of a crime:
   - X_i is dummy variable for being imprisoned;
   - $Y_{1,i},\ Y_{0,i}$ are dummy variables for future rearrest if imprisoned or not.
4. Among mothers with two or more children:
   - X_i is a dummy variable for having a third child;
   - $Y_{1,i}, Y_{0,i}$ are hours of work per week if the mother does or does not have a third child.

Examples:

- selection bias: $E\left[Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{0,i} \middle| X_i = 0\right]$,
- sorting gain: $E\left[Y_{1,i} - Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{1,i} - Y_{0,i}\right]$.

Do you think that selection bias and sorting gain will be negative, zero, or positive in the example:

- Among those visiting the website, X_i is a dummy variable for being shown a targeted ad for a given product; $Y_{1i},\ Y_{0i}$ are whether the individual would buy the product if shown or not shown the ad.

Examples:

- selection bias: $E\left[Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{0,i} \middle| X_i = 0\right]$,
- sorting gain: $E\left[Y_{1,i} - Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{1,i} - Y_{0,i}\right]$.

Do you think that selection bias and sorting gain will be negative, zero, or positive in the example:

- Among patients, X_i is dummy variable for taking medicine;
  $Y_{1i},\ Y_{0i}$ are future health outcomes with and without the medicine.

Examples:

- selection bias: $E\left[Y_{0,i}\middle|X_i = 1\right] - E\left[Y_{0,i}\middle|X_i = 0\right]$,
- sorting gain: $E\left[Y_{1,i} - Y_{0,i}\middle|X_i = 1\right] - E\left[Y_{1,i} - Y_{0,i}\right]$.

Do you think that selection bias and sorting gain will be negative, zero, or positive in the example:

- Among those convicted of a crime, X_i is dummy variable for being imprisoned; $Y_{1i},\ Y_{0i}$ are dummy variables for future rearrest if imprisoned or not.

Examples:

- selection bias: $E\left[Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{0,i} \middle| X_i = 0\right]$,
- sorting gain: $E\left[Y_{1,i} - Y_{0,i} \middle| X_i = 1\right] - E\left[Y_{1,i} - Y_{0,i}\right]$.

Do you think that selection bias and sorting gain will be negative, zero, or positive in the example:

- Among mothers with two or more children, X_i is a dummy variable for having a third child; Y_{1i}, Y_{0i} are hours of work per week if the mother does or does not have a third child.

Solutions?

Some possible ways to address selection bias:

1. Randomized Control Trial (RCT)
    - Considered “gold standard” in causal analysis.
    - We will discuss next.

2. Multivariate regression with additional covariates to control for “ confounders”, for example, clinical and lab measurements at the time of decision to give the patient medicine or not.
    - Plausible that there is no selection bias after including additional observed covariates?
    - Will cover multiple linear regression in Lecture 2.

Solutions?

Some possible ways to address selection bias:

3. Difference-in-difference design
   - feasible? parallel trends assumption plausible?
   - Will cover difference-in-difference design in Lecture 3.
4. Instrumental Variables,
   - plausible instrument?
   - find a valid instrument?
   - Will cover instrumental variables in Lecture 4.
5. Other designs not covered in course:
   - regression discontinuity (RD) design,
   - synthetic control.

RCTs

Randomized Controlled Trials (RCTs):

- Let Z_i denote random assignment:
  - $Z_i = 1$ if randomized to treatment;
  - $Z_i = 0$ if randomized to control.
- Suppose randomization into- or out-of treatment is independent of potential outcomes:

$$Z_i \perp\!\!\!\perp (Y_{0,i}, Y_{1,i}).$$

RCTs

Randomized Controlled Trials (RCTs):

- Let Z_i denote random assignment:
  - $Z_i = 1$ if randomized to treatment;
  - $Z_i = 0$ if randomized to control.
- Suppose randomization into- or out-of treatment is independent of potential outcomes:

$$Z_i \perp\!\!\!\perp \left(Y_{0,i}, Y_{1,i}\right).$$

- Suppose full *compliance*, everyone randomly told to take the treatment does so, those told not to take the treatment do not.

$$Z_i = X_i.$$

RCTs

Randomized Controlled Trials (RCTs):

- Together, random assignment and full compliance imply no selection bias within RCT data, so that can take mean differences to estimate the treatment effect:

$$\begin{aligned}\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0] =& \mathbb{E}[Y_{1,i} \mid Z_i = 1] - \mathbb{E}[Y_{0,i} \mid Z_i = 0] \\ =& \mathbb{E}[Y_{1,i}] - \mathbb{E}[Y_{0,i}] \\ =& \mathbb{E}[Y_{1,i} - Y_{0,i}],\end{aligned}$$

  with

  1. first equality using full compliance ($Z_i = X_i$)
  2. second equality using random assignment,
  3. third equality is using rules for expectation of a sum.

RCTs

Randomized Controlled Trials (RCTs):

- Together, random assignment and full compliance imply no selection bias within RCT data, so that can take mean differences to estimate the treatment effect:

$$\begin{aligned}\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0] =& \mathbb{E}[Y_{1,i} \mid Z_i = 1] - \mathbb{E}[Y_{0,i} \mid Z_i = 0] \\ =& \mathbb{E}[Y_{1,i}] - \mathbb{E}[Y_{0,i}] \\ =& \mathbb{E}[Y_{1,i} - Y_{0,i}],\end{aligned}$$

  with

  1. first equality using full compliance ($Z_i = X_i$)
  2. second equality using random assignment,
  3. third equality is using rules for expectation of a sum.

- In this context, $\mathbb{E}[Y_{1,i} - Y_{0,i}]$ is average effect of treatment for those who select into the RCT.

RCTs (cont'd)

Potential Problem: Noncompliance, $\Pr\{Z_i \neq X_i\} > 0$.

- Some individuals assigned treatment might not actually take the treatment, while some of those randomized out of treatment might still take the treatment.
- Lack of compliance is typically addressed by either
    1. redefining object of interest to be "intention to treat", i.e., define parameter of interest as effect of being assigned to treatment instead of effect of treatment, or
    2. using non-experimental methods to correct, e.g., regression adjust.

RCTs (cont'd)

Potential Problem: ex-post lack of balance

- Random assignment guarantees no ex-ante difference between those assigned to treatment vs. control.
- However, ex-post differences can arise by random chance.
- Typically report a balance at base-line table reporting mean covariates for those assigned to treatment vs control.
  - Very often use t-test to test null hypothesis of equal means for each covariate, but doing so is not justified (Bruhn and McKenzie, 2009).

RCTs (cont'd)

Potential Problem: ex-post lack of balance

- Ways to address ex-post lack of balance?
  - Rerandomize,
  - Regression adjust.
- Alternatively, use stratified random assignment to guarantee balance on key covariates.
- Very active literature in econometrics and statistics on this issue.

RCTs (cont'd)

Some issues of running experiments:

- Ethical concerns (clinical equipoise);
- Recruitment into RCT;
- Feasibility;
- Expense;
- Length of followup, attrition;
- Answers limited question.

Summary

- Central role of selection bias and sorting gain in analyzing causality.
- Bias from mean differences depends on parameter of interest.
- Ideally, RCTs allow one to use mean differences to estimate average effect of treatment, but not always possible.

Next Section , Class

- Next Section (tomorrow):
  - Lab session focusing on coding in **R**,
  - Application: Analyzing PROGRESA, Conditional Cash Transfer (CCT) program.

- Next Class (next week): Linear regression analysis.

To do: Preparation for Coding in **R**

- By tomorrow's section:
  1. Complete (but do not turn in) Problem Set 0.
- As soon as possible (preferably by tomorrow's section):
  1. If you are new to **R**: Complete (but do not turn in) Project 1 and Project 2 of Hands-On Programming with R, by Garrett Grolemund.
  2. If you are new to `ggplot`, review http://r-statistics.co/ggplot2-Tutorial-With-R.html.

To do: Preparation for Statistics

- By next class:
    1. Review posted preparatory materials on expected value and variance, with an application to asset diversification. (slides, handout)
    2. If you have no preparation in statistics, then you should additionally work through the first 12 units of Khan Academy: Statistics and Probability.

To do: Problem Set 1