POLI 30 D: Political Inquiry

Professor Umberto Mignozzetti
(Based on DSS Materials)

Lecture 11 | Causal Inference with Observational Data II

Before we start

Announcements:

- Quizzes and Participation: On Canvas.
- GitHub page: https://github.com/umbertomig/POLI30Dpublic
- Piazza forum: Not sure what the link is. Ask your TA!
- Note to self: Turn on the mic!

Before we start

Recap: We learned:

- The definitions of theory, scientific theory, and hypotheses.
- Data, datasets, variables, and how to compute means.
- Causal effect, treatments, outcomes, and randomization.
- Sampling, descriptive statistics, and descriptive plots for one variable.
- Correlation between two continuous variables.
- Prediction of a binary and a non-binary variable.
- ATE using differences-in-means and simple linear regression.

Great job!

- Do you have any questions about these contents?

Plan for Today

- How Can We Estimate Causal Effects with Observational Data?

- Multiple Linear Regression Models
  - Interpretation of Coefficients
  - Interpretation of $\widehat{\beta_1}$ When X_1 Is the Treatment Variable and the Other X Variables Are All the Potential Confounding Variables

- What is the Effect of the Death of the Leader on the Level of Democracy?

How Can We Estimate Causal Effects with Observational Data?

- Most of the time, we have no control over the assignment.
  - Then, how to eliminate potential confounders and make treatment and control groups comparable?

- First, we must identify all potentially confounding variables
  - variables that affect both (i) the likelihood of receiving the treatment and (ii) the outcome

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X \arrow[rr] & & Y \end{tikzcd}$$

- Then, we need to *control* for them by fitting a **multiple linear regression model**

Multiple Linear Regression Models

Linear models with more than one X variable

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{i1} + ... + \widehat{\beta}_p X_{ip}$$

where:

- $\widehat{Y}_i$ is the predicted value of Y for observation i
- $\widehat{\beta}_0$ is the estimated intercept coefficient
- each $\widehat{\beta}_j$ (pronounced ‘beta hat sub j‘) is the estimated coefficient for variable X_j ($j \in \{1, ..., p\}$)
- each X_{ij} is the observed value of the variable X_j for observation i ($j \in \{1, ..., p\}$ and $i \in \{1, ..., n\}$)
- p is the total number of X variables in the model.

Multiple Linear Regression Models

simple regression $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X$	multiple regression $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + ... + \widehat{\beta}_p X_p$
$\widehat{\beta}_0$: $\widehat{Y}$ when X=0	$\widehat{\beta}_0$: $\widehat{Y}$ when all X_j=0 (j=1, ..., p)
$\widehat{\beta}_1$: $\triangle\widehat{Y}$ associated with $\triangle X$=1	each $\widehat{\beta}_j$: $\triangle\widehat{Y}$ associated with $\triangle X_j$=1, while holding all other X variables constant or *ceteris paribus*

Interpretation of Coefficients in Multiple Linear Regression Models

- $\widehat{\beta}_0$ is the $\widehat{Y}$ when *all* X_j=*0*
- Because there are multiple X variables, there are multiple $\widehat{\beta}_j$ coefficients (one for each X variable)
- Each $\widehat{\beta}_j$ is the $\triangle\widehat{Y}$ associated with $\triangle X_j = 1$, **while holding all other X variables constant**

Interpretation of Coefficients in Multiple Linear Regression Models

- Adding all confounders as controls in the model makes treatment and control groups comparable **after controls**:
  - Something fancily called `conditional exogeneity`.
- As a result, if $\widehat{\beta}_1$ is the treatment of interest, we can interpret it using the **causal language**.
- $\widehat{\beta}_1$ is the $\triangle \widehat{Y}$ *caused by* the presence of the treatment ($\triangle X_1 = 1$), while holding confounders constant.
- $\widehat{\beta}_1$ should be a valid estimate of the average treatment effect if **all** confounding variables are controlled.
- Is this even possible?

Does the Death of the Leader Affect the Level of Democracy?

(Based on Jones and Olken. 2009. *Hit or Miss? The Effect of Assassinations on Institutions and War.* **American Economic Journal: Macroeconomics**, 1 (2): 55-87.)

Does the Death of the Leader Affect the Level of Democracy?

- After an assassination attempt, the death of a leader is close to random:
  - Leaders whose assassination attempts succeeded should be, on average, *comparable* to leaders whose attempts failed.
  - Do you believe that?
- If this is true, we can estimate the average causal effect of the leader's death.
- Fitting a simple linear model where the death of a leader is the "treatment" variable.

The *leaders* dataset

- Dataset on assassinations and assassination attempts against political leaders from 1875 to 2004.

variable	**meaning**
year	year of the assassination attempt
country	country where the assassination attempt took place
leadername	name of the leader
died	whether the leader died as a result of the assassination attempt
politybefore	Polity scores of the country before the assassination attempt
polityafter	Polity scores of the country after the assassination attempt

Does the Death of the Leader Affect the Level of Democracy?

1. Open RStudio
2. Open **inclass07.R** from within RStudio. The file is on Canvas and the Class GitHub.
3. Load the dataset and check the first observations.

What is the Effect of the Death of the Leader on the Level of Democracy?

```
leaders <- read.csv("https://raw.githubusercontent.com/umbertomig/POLI30Dpublic/main/data/leaders.csv")
head(leaders)
##   year     country       leadername died politybefore polityafter
## 1 1929 Afghanistan Habibullah Ghazi    0           -6   -6.000000
## 2 1933 Afghanistan       Nadir Shah    1           -6   -7.333333
## 3 1934 Afghanistan      Hashim Khan    0           -6   -8.000000
## 4 1924     Albania             Zogu    0            0   -9.000000
## 5 1931     Albania             Zogu    0           -9   -9.000000
## 6 1968     Algeria       Boumedienne    0           -9   -9.000000
```

- What is the Treatment variable (X)?
- What is the Outcome variable (Y)?

Does the Death of the Leader Affect the Level of Democracy?

- To fit the simple linear model, where $\widehat{\beta}_1$ is equivalent to the difference-in-means estimator, we run:

```
lm(polityafter ~ died, data = leaders)
##
## Call:
## lm(formula = polityafter ~ died, data = leaders)
##
## Coefficients:
## (Intercept)         died
##      -1.895        1.132
```

- Fitted model: $\widehat{polityafter}$ = -1.90 + 1.13 *died*

Does the Death of the Leader Affect the Level of Democracy?

Interpretation of $\widehat{\beta}_1$:

- $\widehat{\beta}_1$ is the $\triangle\widehat{Y}$ associated with $\triangle X = 1$
- Here: $\widehat{\beta}_1 = 1.13$ is the $\triangle\widehat{polityafter}$ associated with $\triangle died = 1$
- In words: the death of the leader is associated with a predicted increase in polity scores of 1.13 points, on average

Does the Death of the Leader Affect the Level of Democracy?

Unit of measurement of $\widehat{\beta}_1$:

- Same as $\triangle\overline{Y}$
- In here, Y is non-binary and measured in points.
- Then, as $\triangle\overline{Y}$ is measured in points, so is $\widehat{\beta}_1$

Does the Death of the Leader Affect the Level of Democracy?

Interpretation of $\widehat{\beta}_1$ (continuation):

- *died* in the attack is the "treatment" variable.
- The Polity IV score is the "outcome" of interest.
  - Thus, $\widehat{\beta}_1$ is equivalent to the difference-in-means estimator
- We may interpret $\widehat{\beta}_1$ using **causal language**:

> The death of the leader is associated with a predicted **increase** in polity scores after the assassination by 1.13 points, on average.

Does the Death of the Leader Affect the Level of Democracy?

Interpretation of $\widehat{\beta}_1$ (continuation):

- Assumption for Causal Claim be valid:
  - Assassination attempts where the leader died are comparable to those where the leader did not die
- Again, is this plausible?
  - Can you think of a confounder?

Does the Death of the Leader Affect the Level of Democracy?

- We can look at *pre-treatment characteristics*: how democratic was the country *before* the assassination attempt?
- Calculate the average *politybefore* for the two groups:

```
mean(leaders$politybefore[leaders$died==1]) # treatment
## [1] -0.7037037
mean(leaders$politybefore[leaders$died==0]) # control
## [1] -1.743197
```

- What do you see in here?

Does the Death of the Leader Affect the Level of Democracy?

- Attempts where the leader died were more democratic to begin with.
- Thus, *politybefore* might be a **confounding variable**:

$$
\begin{tikzcd}
 & \textit{politybefore} \arrow[dl] \arrow[dr] & \\
\textit{died} \arrow[rr] & & \textit{polityafter}
\end{tikzcd}
$$

Does the Death of the Leader Affect the Level of Democracy?

Estimate average causal effect while controlling for confounders:

- To control for confounders, we need to fit the following multiple regression linear model:

$$\widehat{polityafter} = \widehat{\beta}_0 + \widehat{\beta}_1 \ died + \widehat{\beta}_2 \ politybefore$$

- To fit the model, we use the function `lm()` with a formula of the type $\mathtt{Y} \sim \mathtt{X}_1 + \mathtt{X}_2$

Does the Death of the Leader Affect the Level of Democracy?

```
lm(polityafter ~ died + politybefore, data=leaders)
##
## Call:
## lm(formula = polityafter ~ died + politybefore, data = l
##
## Coefficients:
##  (Intercept)          died  politybefore
##      -0.4346        0.2616        0.8375
```

- Fitted model:

$$\widehat{polityafter} = \text{-}0.43 + 0.26\ died + 0.84\ politybefore$$

Does the Death of the Leader Affect the Level of Democracy?

Interpretation of $\widehat{\beta}_1$:

- $\widehat{\beta}_1$ is the $\triangle\widehat{Y}$ associated with $\triangle X_1$=1, while holding all other variables constant.
  - Here: $\widehat{\beta}_1 = 0.26$ is the $\triangle\widehat{\text{polityafter}}$ associated with $\triangle$*died*=1, while holding *politybefore* constant.

The death of the leader is associated with a predicted increase in polity scores after the assassination attempt of 0.26 points, on average, while holding polity scores before constant.

Does the Death of the Leader Affect the Level of Democracy?

Interpretation of $\widehat{\beta}_1$:

- Unit of measurement of $\widehat{\beta}_1$
  - Same as $\triangle\overline{Y}$
  - Here, Y is non-binary and measured in points so $\triangle\overline{Y}$ is measured in points, and so is $\widehat{\beta}_1$

Does the Death of the Leader Affect the Level of Democracy?

Interpretation of $\widehat{\beta}_1$ (continuation)

- Since here X_1 is the treatment variable, Y is the outcome variable of interest, and X_2 is the confounder we are worried about, we can interpret $\widehat{\beta}_1$ using **causal language**
- **Causal language:** We estimate that the death of the leader ***increases*** polity scores after the assassination attempt by 0.26 points, on average, when holding polity scores before the assassination attempt constant.

Does the Death of the Leader Affect the Level of Democracy?

- This should be a valid estimate of the average treatment effect if *politybefore* is the only confounder
- Note that once we control for *politybefore*, the effect size decreases substantially (it goes from 1.13 to 0.26)
- Based on this analysis, the death of the leader increases the level of democracy of a country but by a smaller amount (more on this later)

AVERAGE CAUSAL EFFECTS WITH OBSERVATIONAL DATA AND MULTIPLE LINEAR REGRESSION.

Suppose in the multiple linear regression model where X_1 is the treatment variable, we control for *all* potential confounders by including them in the model as additional X variables. In that case, we can interpret $\widehat{\beta}_1$ as a valid estimate of the average causal effect of X_1 on Y.

Summary

- **Today's Class:**
  - With Observational Data, Use Multiple Linear Regression Models to:
    1. Control for Confounders
    2. Estimate Average Treatment Effects
- **Next class**:
  - More Causality with Observational Data:
    - Internal versus External Validity.

Questions?

See you in the next class!