Part I: Fundamental concepts in statistical causal inference.

Part II: PhD studies in the US and the University of Chicago.

Panagiotis (Panos) Toulis
Econometrics and Statistics
University of Chicago, Booth School of Business
panos.toulis@chicagobooth.edu
Causality

- Causality has been a fundamental intellectual human activity.

- First rigorous analysis by Aristotle (ca. 340 BC); “four causes”.

- Hume (ca. 1740) argued that our idea of causality comes from experiment and associations.

- Currently a vibrant field with contributions from statistics, physics, and computer science.
• In statistics causality is rooted in experimental design (Fisher, Yates, Neyman, Cochran, Rubin)

• Fisher (1935) argued that the fundamental problem is that we observe only the outcome under the realized experimental conditions.

• Neyman (1923) used the notation $Y(0)$ and $Y(1)$ to denote potential outcomes of control and treatment.

• The idea that causality is fundamentally a missing data problem was completed by Rubin (1974).
Basic formulation

- There are $N$ experimental units indexed by $i$.
- $Z_i$ in $\{0, 1\}$ denotes the treatment assignment of $i$.
- $Z = (Z_i)$; the distribution $p(Z)$ = design; e.g., $p(Z)$ = const. in completely randomized experiment.
- $X_i = \text{covariates of unit } i$ (e.g., age, education)
- $Y_i(Z) = \text{potential outcome of unit } i \text{ under } Z$. 
• Definition of causal effect (estimand of interest):

\[ \tau = \frac{1}{N} \sum_i Y_i(1, \cdots, 1) - Y_i(0, \cdots, 0) = \frac{1}{N} \sum_i Y_i(1) - Y_i(0). \]

• Causal effect is comparison between potential outcomes. At most one is observed; possibly none.

• Stable unit treatment value assumption (SUTVA):

\[ Y_i(Z) = Z_i \times Y_i^1 + (1 - Z_i)Y_i^0. \]

• Under SUTVA, causal effect can be written as:

\[ \tau = \frac{1}{N} \sum_i Y_i^1 - Y_i^0. \]
Causal vs non-causal approach

<table>
<thead>
<tr>
<th>unit i</th>
<th>Z_i</th>
<th>Outcome Y_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3+\varepsilon</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3-\varepsilon</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

- Consider the data in the table. Suppose we use a model

\[ Y_i = \alpha + \tau Z_i + \mathcal{N}(0, \sigma^2). \]

- Estimate of \( \tau \) is \( \hat{\tau} = \bar{Y}_{treated} - \bar{Y}_{control} \).
- Estimate of error is \( \varepsilon^2 \).

estimate of error not equal to error of experiment
Causal approach

<table>
<thead>
<tr>
<th>unit i</th>
<th>$Z_i$</th>
<th>Outcome $Y_i^1$</th>
<th>Outcome $Y_i^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>?</td>
<td>3+(\varepsilon)</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>?</td>
<td>3-(\varepsilon)</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>8</td>
<td>?</td>
</tr>
</tbody>
</table>

- First step is to write down missing data by “?”.  
- As before, estimate of $\tau$ is difference in sample means.  
- But what can we say about the error?
Causal approach

Fisher’s sharp null: suppose there is no effect from treatment:

$$H_0: Y_{i1} = Y_{i0}, \text{ for every unit } i.$$  

Now we can **impute** missing data. Table of data is full.

Idea: if we resample $$Z$$ according to the design we will get values of the causal effect estimate under the null.

Use such values to estimate error!
Causal approach

<table>
<thead>
<tr>
<th>unit i</th>
<th>$Z_i$</th>
<th>Outcome $Y_i^1$</th>
<th>Outcome $Y_i^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3+ε</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>?</td>
<td>3-ε</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>?</td>
<td>8</td>
</tr>
</tbody>
</table>

- Suppose we had sampled $Z = (1, 1, 0, 0)$. Observed data would be as in table.

- Estimate would be

$$\hat{\tau} = \frac{(8 + 3 + \varepsilon)}{2} - \frac{(8 + 3 - \varepsilon)}{2} = \varepsilon$$
• We repeat this process for **all** possible assignments.

<table>
<thead>
<tr>
<th>assignment</th>
<th>estimator value $\hat{\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0, 1, 1)</td>
<td>$-\varepsilon$</td>
</tr>
<tr>
<td>(0, 1, 0, 1)</td>
<td>$+\varepsilon$</td>
</tr>
<tr>
<td>(1, 0, 0, 1)</td>
<td>$+5$</td>
</tr>
<tr>
<td>(0, 1, 1, 0)</td>
<td>$-5$</td>
</tr>
<tr>
<td>(1, 0, 1, 0)</td>
<td>$-\varepsilon$</td>
</tr>
<tr>
<td>(1, 1, 0, 0)</td>
<td>$+\varepsilon$</td>
</tr>
</tbody>
</table>

• Comparison with model-based approach

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\tau}$</th>
<th>$\text{Var}(\hat{\tau})$</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>regression</td>
<td>5</td>
<td>$\varepsilon^2$</td>
<td>yes</td>
</tr>
<tr>
<td>Fisher</td>
<td>5</td>
<td>$.8\varepsilon^2 + 5$</td>
<td>no</td>
</tr>
</tbody>
</table>
• SUTVA assumption is strong in many cases.

• Suppose one’s treatment affects others (e.g., advertising effects on social media, politics)

• This is known as **interference** (Cox, Rubin).

• Usually there is a graph $G$ such that we can make a generalized SUTVA (Toulis & Kao, 2013):

  \[ Y_i(Z) = Y_i(Z') \text{ if } G_i^T Z = G_i^T Z' \]

• In words, outcome is the same if same #neighbors treated.
Observational studies

- When there is no experiment things get more complicated.
- Very easy to get effects wrong. Classical example by Cochran (1968):

<table>
<thead>
<tr>
<th></th>
<th>mortality rate (per 1000 people)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonsmokers</td>
<td>20.2</td>
</tr>
<tr>
<td>cigarette smokers</td>
<td>20.5</td>
</tr>
<tr>
<td>cigar/pipe smokers</td>
<td>35.5</td>
</tr>
</tbody>
</table>
Observational studies

• When there is no experiment things get more complicated.

• Very easy to get effects wrong. Classical example by Cochran (1968):

<table>
<thead>
<tr>
<th></th>
<th>mortality rate (per 1000 people)</th>
<th>mean age</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonsmokers</td>
<td>20.2</td>
<td>54.9</td>
</tr>
<tr>
<td>cigarette smokers</td>
<td>20.5</td>
<td>50.5</td>
</tr>
<tr>
<td>cigar/pipe smokers</td>
<td>35.5</td>
<td>65.9</td>
</tr>
</tbody>
</table>
• One widely used approach is based on propensity score (PS), which is defined as
\[ e(X) = p(Z = 1 \mid X) \]

• Then \( Z \) is independent of \( X \) conditional on \( e(X) \):

\[ \begin{align*}
P(X \mid Z = 1, e) & \propto p(Z = 1 \mid X, e) \cdot p(X \mid e) = e \cdot p(X \mid e) \\
& \quad (1 - e) \cdot p(X \mid e)
\end{align*} \]

• For units with same PS treatment is as if randomized. So we can group units based on estimated PS, and analyze within groups.

• Leads to unbiased estimation of causal effects under unconfoundedness assumption:
\[ Z \perp (Y^1, Y^0) \mid X. \]
Entanglement

• e.g., if treatment is new course then all students in a class are treated;

• or treatment itself is defined on pairs of units (e.g., treatment=new friendship)

• In such cases treatment assignment is entangled.
Dynamical systems

• Problems also arise when the system of units responds to treatment in a dynamic way.

• No theory or practice of classical experimental design/causal inference in such systems.

• Challenge is to control for transient effects (Toulis & Parkes, 2016);

• e.g., when we raise prices we expect increase in revenue but it may be short-term.
Part II: PhD studies in the USA
• Some things are necessary such as
  • GRE test (quantitative and verbal).
  • TOEFL for English language.
  • Statement of purpose.
  • Good grades and other certification.

• What **really** makes the difference:
  • Strong recommendation letters from good people.
  • Evidence for research potential (e.g., writeups, papers).
University of Chicago

- Founded in 1890—distinct Gothic architecture.
- Consistently ranks in top-10 universities.
- Resolute in support of free inquiry and impact.
- World-changing ideas (e.g., “Chicago School”)
Chicago Booth

• Booth is the Business school of U Chicago; world-renowned for economics/econometrics research.

• Has PhD program; 20 students/year.

• Deadline for applications is Dec 15 (hurry!)

• 5yr full financial support w/ extensions.

• Ideal for students interested in the intersection of statistics and economics; focus on impact.
A few things to know about PhD in the US

• US is far in many aspects; be ready to stand on your feet.

• US is different than Europe; often in good ways.

• Financing will not be an issue; probably you will not become rich during your PhD, but you will easily manage.

• Focus on academics; a successful PhD will be noted and rewarded in the market.

• US provides access to solid and dynamic labor markets, including academics, high-tech, medicine (e.g., pharmaceuticals), and finance (e.g., hedge-funds).
Thank you!