Quantitative Methods

Human Sciences, 2020–21

Elias Nosrati

Lecture 10: 22 February 2021

Today

- Review and discussion of causal inference.
- Presenting statistical results.
- Finishing touches.

Defining causal effects

- Comparison of outcomes between scenarios where an action (treatment, intervention, exposure) is taken versus when the action is withheld.
- If the two outcomes differ, we say that the action had a causal effect on the outcome. Otherwise, the action has no causal effect.
- Formally: treatment variable T and outcome variable Y.
- Two *potential* or *counterfactual outcomes*: Y_1 and Y_0.
- Y_1 is the outcome under treatment: the value of outcome variable when the individual has treatment value $T = 1$.
- Y_0 is the outcome under non-treatment (or *control*): the value of outcome variable when the individual has treatment value $T = 0$.
- The treatment T has a *causal effect* on an individual's outcome Y if $Y_1 \neq Y_0$ for that individual.
- The causal effect itself is defined as $Y_1 - Y_0$.

Causal inference

- Only one counterfactual outcome is observed for each individual: impossible to reconstruct the contrast between values of counterfactual outcomes.
- An *average* causal effect in the population is present if

$$\mathbb{E}(Y_1) \neq \mathbb{E}(Y_0).$$

- The *population average treatment effect* (PATE) is defined as a contrast between expected values of counterfactual outcomes:

$$\mathbb{E}(Y_1 - Y_0) = \mathbb{E}(Y_1) - \mathbb{E}(Y_0).$$

- Is it enough to compute $\mathbb{E}(Y \mid T = 1) - \mathbb{E}(Y \mid T = 0)$?

Causal inference

- Only one counterfactual outcome is observed for each individual: impossible to reconstruct the contrast between values of counterfactual outcomes.
- An *average* causal effect in the population is present if

$$\mathbb{E}(Y_1) \neq \mathbb{E}(Y_0).$$

- The *population average treatment effect* (PATE) is defined as a contrast between expected values of counterfactual outcomes:

$$\mathbb{E}(Y_1 - Y_0) = \mathbb{E}(Y_1) - \mathbb{E}(Y_0).$$

- Is it enough to compute $\mathbb{E}(Y \mid T = 1) - \mathbb{E}(Y \mid T = 0)$?

Three forms of systematic bias

- A study suffers from *systematic bias* when the data are insufficient to compute a causal effect even with an infinite sample size.
- Any structural association between treatment and outcome that does not arise from the causal effect of treatment on outcome.
- Three forms of systematic bias:
    1. Confounding: not conditioning on common causes.
    2. Selection: conditioning on common effects.
    3. Measurement: information error.
- These biases may arise in observational studies *and* in randomised experiments.

Visualising systematic bias

Confounding:

$$\begin{tikzcd}
X & & Y \\
& C \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Selection:

$$\begin{tikzcd}
X \arrow[rr, bend left] & Y \arrow[r] & \boxed{C}
\end{tikzcd}$$

Measurement error:

$$\begin{tikzcd}
C^* & & \\
C \arrow[u] \arrow[r] \arrow[rr, bend left] & X \arrow[r] & Y
\end{tikzcd}$$

Three forms of systematic bias

- A study suffers from *systematic bias* when the data are insufficient to compute a causal effect even with an infinite sample size.
- Any structural association between treatment and outcome that does not arise from the causal effect of treatment on outcome.
- Three forms of systematic bias:
  1. Confounding: not conditioning on common causes.
  2. Selection: conditioning on common effects.
  3. Measurement: information error.

Are RCTs a panacea?

- Randomised controlled trials (RCTs) are generally viewed as the 'gold standard' for causal inference. Why?
- Randomisation ensures exchangeability of treatment and control groups $\rightarrow$ identification of causal effects.
- Internal validity: validity of conclusions within the context of a particular study.
- External validity: extrapolation of conclusions beyond a particular study.
- Sample versus (target) population: trial sample may not be representative of the population that is ultimately of interest.
- *Average* causal effects as difference in means: how about other features of a distribution (median, percentiles, variances, etc.)?
- Heterogeneous causal effects: a treatment can have differential impacts that depend on factors other than the treatment itself.
- Black box: we might know *that* a treatment has an effect, but not *how*. What is the underlying mechanism?

RCTs and systematic bias

- Randomisation in and of itself does not ensure exchangeability: in any one trial, there may be ‘random confounding’.
- Sample size and replication across trials matter.
- Attrition: loss of participants during an experiment (after randomisation) $\rightarrow$ selection bias.
- Information error: endpoints and measurement.

Causal structures and causal reasoning

- Russell's chicken: the bird infers, on the basis of repeated evidence, that when the farmer arrives, he feeds her. Christmas morning arrives: he wrings her neck and serves her for dinner.
- An RCT demonstrates positive causal effect of a new fertiliser on crop yields. A farmer who adopts the new fertiliser sees increase in income. However, if procedure is adopted by all farmers, the price drops and incomes decline: the scaled-up effect is opposite to the trial effect.
- Takeaway: meaningful causal inference requires substantive knowledge and careful reasoning. Randomisation does not warrant blind application of quantitative methods.
- For more, see special issue of *Social Science & Medicine*: 'Randomized Controlled Trials and Evidence-based Policy: A Multidisciplinary Dialogue'.

Presenting statistical results

- Descriptive statistics: describe data and sources.
- Parameter estimate: meaningful scale.
- Uncertainty: standard error, confidence interval.
- Model fit: absolute versus relative model fit.
- Do not overload with data: carefully select from set of all findings.
- Do not present 'tests' or other statistics just because your software prints them out: avoid redundant information.
- Visualisation or other easy way of conveying key quantities of interest.
- Discuss assumptions, strengths, and limitations.
- Useful resources: `Zelig` and `tidyverse`.

Assignment and exam

- Coursework: analyse data and present findings.
- Exam: show theoretical understanding, discuss and interpret models.
- Preparation: solve problems.
- If you have questions, get in touch!