Q-Step: Week 4 Lecture

Case Selection

Spyros Kosmidis
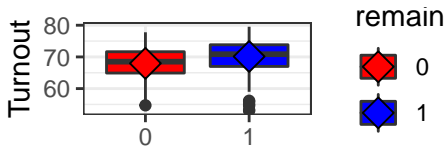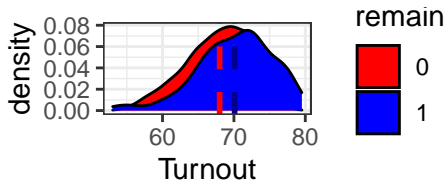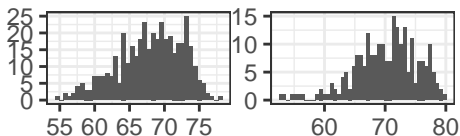
Oxford

February 5, 2022

Last week- Descriptive Statistics

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N-1}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N-1}}$$

Last week-Visualization

Roadmap

Today

- Problems with Case Selection
  - ▸ Primer on Causality
  - ▸ Selecting on the Dependent Variable
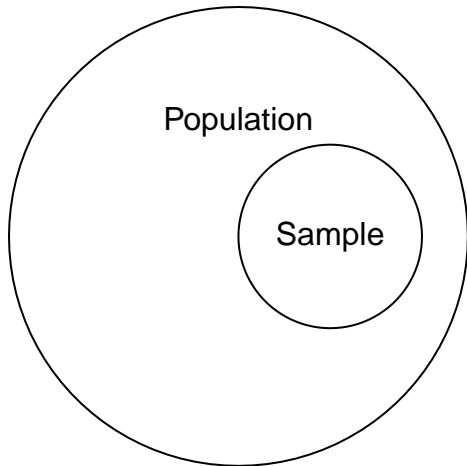- Suggestions

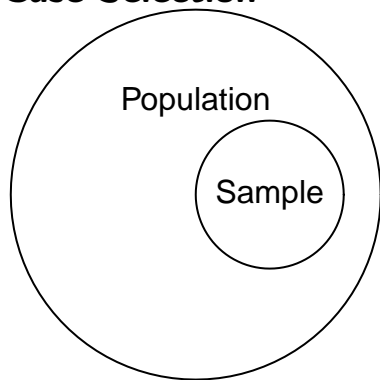**Next Week**

- Bivariate Relationships

Case Selection

- Population: The entire set of cases that our theory applies to

- Sample: Subset of cases that we analyse

Case Selection

- Population: The entire set of cases that our theory applies to

- Sample: Subset of cases that we analyse

Case Selection

Case Selection
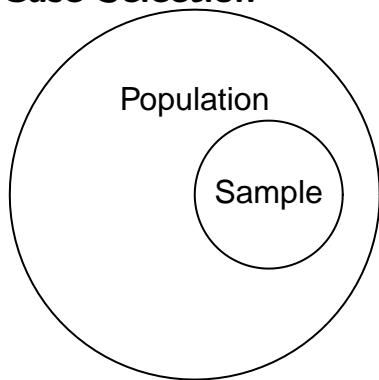
- What is Case?

Observations which are used to draw inferences about

- Why Select & How?

As analysing all cases is often impossible, cases can be selected using random sampling or purposeful/criterion-based sampling

Causality?
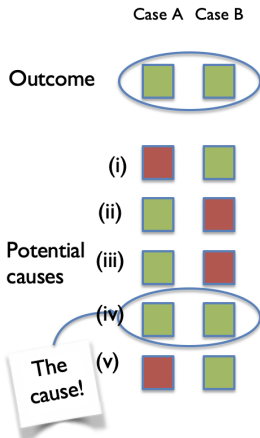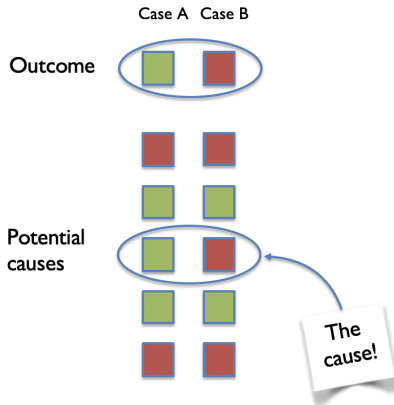
Causality?
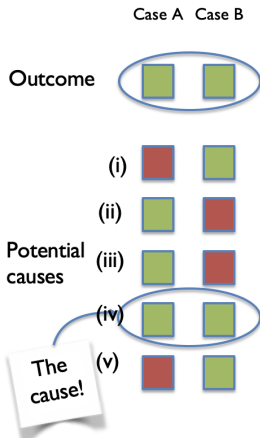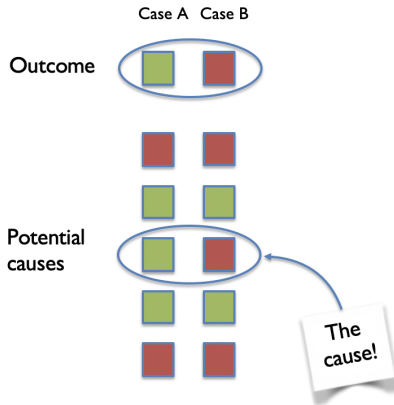
J.S. Mill's Method of (Dis)Agreement

Problems with Case Selection

Imagine a set of countries A,B,C,D,E,F,G.
And you select A & B, and only observe them.

high

A

B

What A & B share is the cause

Incumbent Vote Share

C

D

E

F

G

low

low

high

Economic Growth

You may make a flawed inference about the possible distribution of C-G based on extrapolation of relationship you found for cases A and B
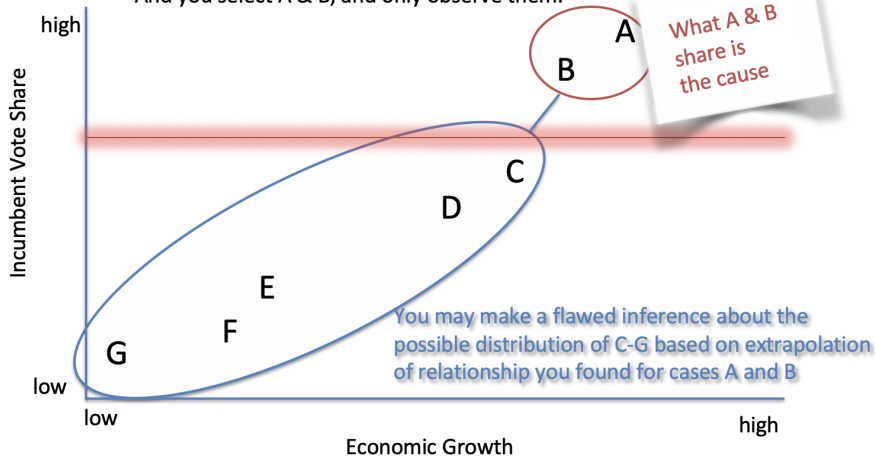
Problems with Case Selection
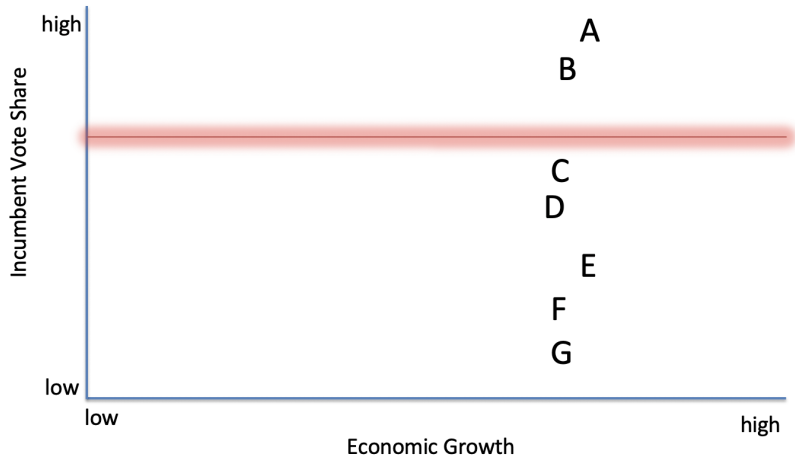
Selecting on the DV

- Can we explain the determinants of Y by exclusively looking at instances where Y is true?

- For example, we cannot infer the effect of COVID-19 vaccines if we only focus on those who took the vaccine.

- Or, are IMF economic interventions good for high debt economies?

- Hard to answer this because the IMF intervenes only in places that have high debt (and probably in the counterfactual world where they Fund had not been asked the consequences would have been worse)

- Selecting on the DV is a key concern for observational (non experimental) studies and it has serious implications for valid inferences

Example I on Selecting on the DV

From Geddes, 1990: 132

- Some scholars noticed the rapid growth of some developing countries

- At the peak of their economic growth, these countries "*exerted controls over labor and prevented most expressions of workers discontent*".

- Quick thought: Did all countries that repressed labour prospered economically?

- Where should we focus?

Example I on Selecting on the DV

From Geddes, 1990: 132

- Some scholars noticed the rapid growth of some developing countries

- At the peak of their economic growth, these countries "*exerted controls over labor and prevented most expressions of workers discontent*".

- Quick thought: Did all countries that repressed labour prospered economically?

- Where should we focus?

    ▶ Rapid Growth?

Example I on Selecting on the DV

From Geddes, 1990: 132

- Some scholars noticed the rapid growth of some developing countries

- At the peak of their economic growth, these countries "*exerted controls over labor and prevented most expressions of workers discontent*".

- Quick thought: Did all countries that repressed labour prospered economically?

- Where should we focus?

    ▸ Rapid Growth?

    ▸ Labour Repression

Example I on Selecting on the DV

From Geddes, 1990: 132

- Some scholars noticed the rapid growth of some developing countries

- At the peak of their economic growth, these countries "*exerted controls over labor and prevented most expressions of workers discontent*".

- Quick thought: Did all countries that repressed labour prospered economically?

- Where should we focus?

  ▶ Rapid Growth?

  ▶ Labour Repression

- **Looking at more cases (both fast and slow growth in NIC), shows no relationship between repression and growth.**

Example II, *States & Social Revolutions*, (1979)

- Skocpol's book is one of the most influential and well read monographs in political science

- Her research is predominantly based on the analysis of the French, Russian and Chinese revolutions

If the goal of her book was largely descriptive in nature (what is a revolution), this case selection might be defendable

- But given that the goal was to address explanatory and causal questions (why do revolution occur, or what is the effect of military defeat on revolutions) this might be problematic

  ▸ Problem of selecting on the dependent variable, DV

  ▸ Common in study of Comparative Politics (Geddes 1990, Sekhon 2004), also in IR research examining causes of war (Achen and Snidal 1989)

The Critique

- Barbara Geddes argues that Skocpol's case selection is problematic:

- Failure to provide contrasting cases for key causal claims

- Exhaustive list of all potential causes?

The cases Skocpol selects for examination confirm her argument, but would a differently selected set of cases do so?" (Geddes 1990: 143)

- No contrasting cases to back up her central claim:

"developments within the international system as such–especially defeats in wars or threats of invasion and struggles over colonial control–have directly contributed to virtually all outbreaks of revolutionary crises" (Skocpol 1979: 23)

The Problem

- "Many countries in the world have suffered foreign pressures as great as those suffered by France and yet revolutions occur infrequently." (Geddes 1990:143)

- How well do other cases fare?

The Problem

- "Many countries in the world have suffered foreign pressures as great as those suffered by France and yet revolutions occur infrequently." (Geddes 1990:143)

- How well do other cases fare?

	Revolution	*No Revolution*
Defeated Last 20 Years	*Bolivia: Defeated 1935 Revolution 1952*	*Bolivia, 1839, 1883, 1903 Mexico, 1848 Paraguay, 1869 Peru, 1839, 1883*
Not Defeated Last 20 Years	*Mexico, 1910 Nicaragua, 1979*	*All Others*

See Sekhon (2004)

- Key Consideration: Probabilistic or Deterministic Inference?

Selecting Cases

Selecting Cases-Foreign Threat and Revolution in LA

	Revolution	*No Revolution*
Defeated Last 20 Years	**Bolivia: Defeated 1935 Revolution 1952** P₁=1/8	**Bolivia, 1839, 1883, 1903 Mexico, 1848 Paraguay, 1869 Peru, 1839, 1883**
Not Defeated Last 20 Years	**Mexico, 1910 Nicaragua, 1979**	*All Others* Roughly P₂=2/34=1/18, which is considerably smaller than 1/8

Source Sekhon (2004: 286-287)

both Geddes, 1990 and Sekhon, 2004:286-7

Verdict?

- If Skocpol's analysis is deterministic (foreign threat produces revolution), then selecting on the DV is clearly problematic (albeit super-interesting and insightful)

- Adding more cases (as suggested by Geddes), and thinking about the relationship in a probabilistic way, then Skocpol's conclusions are valid. It is more likely to observe a revolution after foreign threats.

- Sekhon's (2004) analysis is illuminating and you should read it soon.

How should we select our cases?

- Your case selection will affect the answers you get for your research questions.

- Case selection vital for observational studies

- Think about type of question, if explanatory (why X leads to Y?) or causal (what is the effect of X on Y?): DON'T SELECT ON DV

Key problems due to biased case selection:

- **Guilt by Association**: Erroneous inference that any characteristic that cases with similar outcome share is a cause
- **Cherry picking**: Erroneous inference based on picking cases that support hypothesis whilst ignoring other cases
- **Overgeneralisation**: Erroneous inference that a relationship between variables in selected set of cases reflects the relationship in all cases

Summary

- Selecting your cases affects the answers you get

- This is a useful reminder when you form arguments

- But also useful when you critically assess research

Summary

- Selecting your cases affects the answers you get

- This is a useful reminder when you form arguments

- But also useful when you critically assess research

The following 4 weeks

Summary

- Selecting your cases affects the answers you get

- This is a useful reminder when you form arguments

- But also useful when you critically assess research

The following 4 weeks

- The focus of both lectures and labs will be on how to link a dependent to one or more independent variables

- Start thinking about your QStep Essay

- Those who are really keen about methods, statistics, R etc and want to learn –even– more should contact me at *spyros.kosmidis@politics.ox.ac.uk*

Thank you!