Topic 1: **Introduction**

Hanno Hilbig

Research Methods in Political Science

Let's start with the name of this course:

Research: You will be learning how to *do research*, not how to read the literature (you'll do that elsewhere).

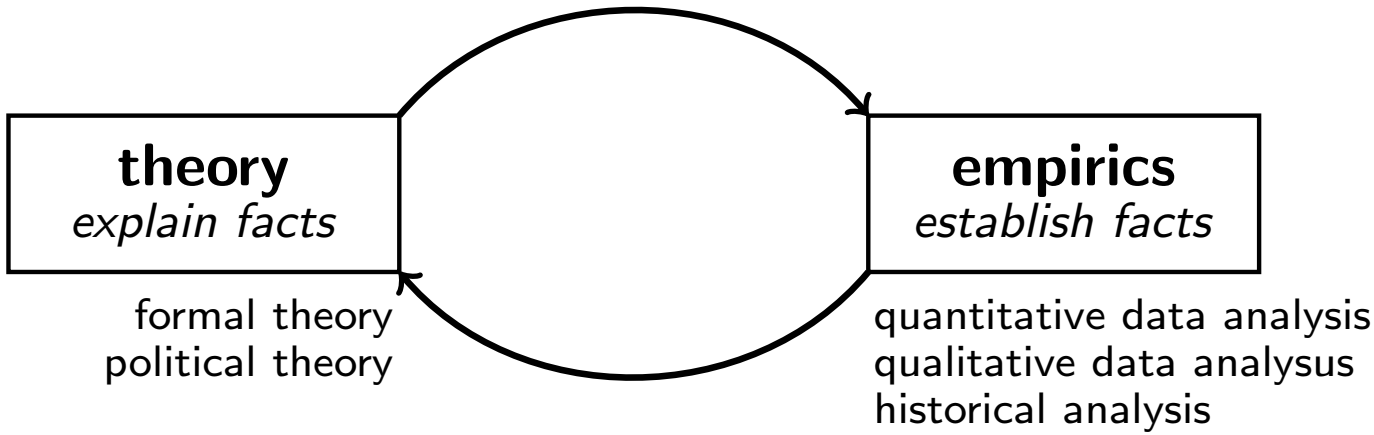
Methods: You will be learning about *methods*, which is a structured way to produce research.

Political: The research methods you learn are not specific to the study of politics, but they're very widely used to study politics.

Science: An important goal in research is *credibility*, which makes what we do something closer to a "science."

The Two Broad Categories of Research

Social science is a conversation between theory and empirics.



"Methods" in political science often means:

- *quantitative data analysis*
- *formal theory*

Purpose of This Course

Academic:

↪ Most political science research uses quantitative data analysis

↪ Evaluating empirical claims is a core skill for political scientists

Non-Academic

↪ Many fields value quantitative data analysis skills:

↪ Tech, consulting, government, etc.

↪ (Statistical) programming skills are in high demand

↪ Quantitative literacy matters even if you don't use it directly

The Required* Methods Sequence

POL 211: Foundations of quantitative data analysis (probability, statistics, and basic inference)

POL 212: Data (computational methods for data collection/analysis)

POL 213: Regression (OLS model, other kinds of regression)

POL 215: Formal theory (decision theory and game theory)

De facto (but not de jure) requirement:

POL 285: Causal inference ("natural experiments")

Requirements versus Expectations

Methods is at the core of producing good research.

Students should take as many methods courses as needed to figure out how to produce high quality, publishable research.

Advice by Ryan Hubert: (corroborated by other faculty!):

↪ Front-load methods courses in your first two years so that you can effectively use those methods in your dissertation.

↪ Take at least six quarters of methods courses, and read broadly in methods outside of your formal coursework.

Learning Methods

This course will be difficult and will require a lot of work.

↪ You have to spend a lot of time reading, thinking and doing.

↪ This is like learning a new language: you must practice!

↪ You'll struggle and that is normal. Just keep at it.

See Canvas to review the course syllabus and outline.

The Framework

When doing empirical work, social scientists are typically "interested in" using data to learn about **positive theories**.

From Gailmard (2014), ch. 1: "a *positive theory* in social science asserts that two or more concepts or events are related to each other and specifies the reason why."

↪ Note: positive theories are concerned with "what is," **normative theories** are concerned with "what ought to be."

An example from **theory of democratic peace**: two states are less likely to go to war if they are both democracies.

↪ Two distinct concepts/events: *both countries are democracies* and *both countries are at war*, related in a specific way.

The Framework

The core question: does this theory teach us about the real world?

We can turn to data to try to answer this, but it's hard:

↪ How do we measure these concepts?

↪ How do we figure out whether there is any relationship?

↪ If we find evidence of a relationship, how do we know if it is good evidence? (And what does "good evidence" mean?)

"Statistics provides a set of tools for assessing relationships in data. That is the main reason social scientists use it."

But it also has to be used well...

Thinking Clearly With Data

It's not 1980 and I don't need to convince you that data is powerful.

$\hookrightarrow$ Covid data, election forecasts, "data activism," etc.

The bigger challenge now is that data is so ubiquitous in our day to day lives that many of us have learned bad interpretative habits.

We need to train ourselves not to interpret data badly.
That's what it means to learn how to **think clearly with data**.

This is especially important for researchers and teachers.
We shouldn't put misinformation into the world!

But, this is much harder than you might think.

Thinking Clearly With Data

Consider an example, call it the **theory of hospitals**: going to the hospital improves peoples' health.

Can we use data to learn about this theory? Data from 2005:

Treatment Group	Sample Size	Mean Health Status	Standard Error
Hospital	7,774	3.21	0.014
No Hospital	90,049	3.93	0.003

Source: Angrist and Pischke (2009)

Mean health status is 0.72 higher for people who did not go to the hospital. What do we learn about the theory of hospitals?

Inferential Problems

Inference is the term we use to describe learning from data:

↪ **Statistical inference** is about what we can learn from data when we have uncertainty.

↪ **Causal inference** is about whether what we learn from data is "genuine" (in a sense that will become clearer later on).

When someone makes faulty conclusions from data, this is an **inferential mistake**. They could be due to many problems:

— You could have badly measured data.

— You could have too little data.

— You could have unrepresentative data.

— You could be comparing the wrong things.

Inferential Problems

Back to the hospital example:

Treatment Group	Sample Size	Mean Health Status	Standard Error
Hospital	7,774	3.21	0.014
No Hospital	90,049	3.93	0.003

Some initial questions:

↪ How was "mean health status" measured? Is it a good measure for what we care about?

↪ How uncertain are we about our estimates?

Inferential Problems

Back to the hospital example:

Treatment Group	Sample Size	Mean Health Status	Standard Error
Hospital	7,774	3.21	0.014
No Hospital	90,049	3.93	0.003

The major issue here: *only sick people go to the hospital!*

↪ These two groups of people are not comparable.

↪ This inferential mistake is called a **selection problem**.

On the bright side: at least there's a comparison! 😬

Inferential Problems

What's the *implied** relationship being examined in each example, and a potential inferential mistake?

↪ "Democracies don't go to war with each other."

↪ "Based on a sample of college students, we argue that people who are good at sports tend to perform worse academically."

↪ "Standardized test scores (like the GRE) are not useful for admitting students into graduate programs because those scores are uncorrelated with students' eventual outcomes."

↪ "Of the people who contracted Covid-19, those who regularly flossed developed less severe Covid-19 disease."

In your methods courses, you'll learn how to avoid these mistakes.

A Useful Crutch: Mathematical Reasoning

- ↪ Technical understanding is essential for clear thinking with data.
- ↪ Math offers clarity in representation and logical reasoning.
- ↪ Math is a practical tool, not an ideological commitment.
- ↪ Analytical transparency and complexity reduction are key reasons for quantitative analysis.