Quantitative Social Science Methods I

Lecture 1: What Is This Course?

Scott Cunningham

Harvard University
Department of Government

January 27, 2026

A Question

Does democracy cause peace?

We observe that democracies rarely fight each other.

But is this because of democracy itself?

Or because democracies tend to be wealthy, trading partners, allies?

A Question About Voting

Does voter ID reduce turnout?

States with strict ID laws have lower turnout.

But did the law cause the lower turnout?

Or are these states different in other ways?

Scott Cunningham

A Paradox in HIV Treatment

Did antiretroviral therapy increase AIDS mortality?

Patients on HAART (antiretroviral therapy) had *higher* mortality than untreated patients.

But was the treatment actually harmful?

Or did insurance rules mean only the sickest patients qualified for coverage?

The Core Challenge

<div align="center">

We want to learn about the world from data.

</div>

But:

- Data is noisy
- Samples are finite
- Correlation is not causation
- We rarely know the true data-generating process

This course gives you tools to navigate this challenge.
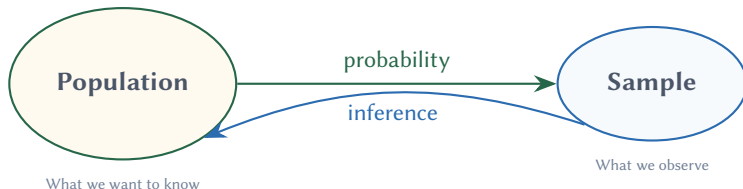
The Arc of the Course

Probability → Inference → Regression

- **Probability**: The logic of uncertainty. What data would we expect?
- **Inference**: Learning from data. What can we conclude?
- **Regression**: The workhorse tool for estimating relationships.

Why This Sequence?

Probability tells us: given the population, what samples might we see?

Inference reverses this: given a sample, what can we learn about the population?

What You Will Learn

By the end of this course, you will be able to:

1. **Read** quantitative political science papers critically

2. **Estimate** regression models and interpret the output

3. **Understand** the assumptions behind statistical claims

4. **Communicate** findings with appropriate uncertainty

5. **Diagnose** problems and know when methods break down

The "Agnostic" Philosophy

What can we learn without assuming we know the truth?

> *"I do not pretend to know where many ignorant men are sure—that is all that agnosticism means."*

—Clarence Darrow

We don't assume the world is linear, errors are normal, or that we have the "right" model.

Instead, we ask: What can we learn about a **well-defined population quantity** without assuming we know the true model?

- The **Conditional Expectation Function** (CEF) becomes our target
- **Linear regression** becomes a tool for approximating it
- **Uncertainty quantification** tells us how much to trust our estimates

Population First

Before analyzing data, ask:

What do I want to know?

- Define the **population** of interest
- Define the **quantity** you want to estimate (the "estimand")
- *Then* think about how your sample relates to that population
- *Then* choose an estimation strategy

This order matters. Getting it backwards leads to confusion.

Three Objects, Three Symbols

The parameter is fixed (we just don't know it).
The estimator is a random variable (it depends on which sample we get).
The estimate is one realization (the number from our actual sample).

Object	What it is	Notation
Parameter (Estimand)	The population quantity we want	θ, μ, β
Estimator	The formula/procedure we apply to data	$\hat{\theta}, \bar{X}, \hat{\beta}$
Estimate	The number we get from our sample	$\hat{\theta} = 2.34$

Confusing these three objects is the source of many errors in applied work.

Scott Cunningham

Course Logistics

Lectures: Monday & Wednesday, 1:30–2:45 PM

Sections: Weekly (times TBD)

Assessments: Weekly problem sets, take-home midterm, take-home final

Computing: R (install before Wednesday)

Primary Texts:

- **Blackwell**, *A User's Guide to Statistical Inference and Regression*
- **Aronow & Miller**, *Foundations of Agnostic Statistics*

See syllabus for full details, office hours, supplementary readings.

What I Expect From You

1. **Engage with the material.** Attend lectures. Do the readings. Ask questions.

2. **Struggle with problem sets.** The struggle is the learning. Don't copy solutions.

3. **Collaborate thoughtfully.** Work together, but write your own answers.

4. **Ask for help.** Office hours exist for a reason. Use them.

5. **Be patient with yourself.** This material is hard. You will be confused. That's normal.

For Wednesday

Reading:

- Aronow & Miller, Preface (pp. xiii–xv)
- Aronow & Miller, §1.1 (pp. 1–14): Probability, conditional probability, Bayes

Install R and RStudio before Wednesday if you haven't already.

Welcome to Gov 2001.