Quantitative Social Science Methods I

Lecture 1: What Is This Course?

Scott Cunningham

Harvard University
Department of Government

January 26, 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?

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

We want to learn about the world from data.

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.

About Me

Scott Cunningham
Visiting Professor (one year), Department of Government
Economist, Baylor University

Research: Applied microeconomics—relationships (marriage, fertility, online dating, sex work), drug policy (methamphetamine, psychedelics), suicide in corrections.

Why this matters for how I teach:

I came to statistics late. English major in college—no economics, minimal math, no statistics. My first statistics course was this one: probability, first semester of my PhD. It was rough. Very different from creative writing.

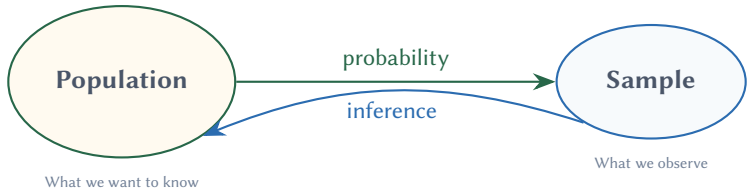
I've spent my career as an eclectic applied researcher without a natural "field." That shapes how I think about probability—as a *tool* for answering questions, not an end in itself.

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.

Course Logistics

Lectures:	Monday & Wednesday, 1:30–2:45 PM
Sections:	Weekly (times TBD)
Assessments:	Weekly problem sets, in-class midterm, in-class 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.

A Note on Proofs

I will **usually not** work through proofs in class, though sometimes I will.

But I fully expect you to do so.

In my experience, working through each theorem and proof oneself is *the only way* to learn anything mathematically technical.

- Read the proof slowly, line by line
- Ask: why does this step follow from the last?
- Try to reproduce it without looking
- If you're stuck, struggle before seeking help

This is probably one area where generative AI is not a substitute—at least not until you've exhausted other ways of learning. But even then, the goal is always to learn the proofs yourself.

The Course Website

After every class, I will update the course website with slides, readings, and announcements:

https://scunning.com/gov2001.html

On the website you'll find:

- Lecture slides (posted after class)
- Problem set assignments and due dates
- Required and supplementary readings
- Office hours and section information

Bookmark this page. Check it regularly.

Let's Begin

We're now going to move directly into the first substantive material: **probability foundations**.

Today's reading (for reference):

- Aronow & Miller, §1.1 (pp. 1–14): Probability, conditional probability, Bayes
- Blackwell, Chapter 2.1: Probability foundations

If you haven't already, install R and RStudio before section on Friday.

Welcome to Gov 2001.