

# Gov 2001: Introduction

Naijia Liu

Spring 2025

# Why build a foundation?

- Sophisticated user: use or understand new methods without a strong foundation in rigorous statistics.

# Why build a foundation?

- Sophisticated user: use or understand new methods without a strong foundation in rigorous statistics.
- You will be using methods for the rest of your career—invest wisely.

# Why build a foundation?

- Sophisticated user: use or understand new methods without a strong foundation in rigorous statistics.
- You will be using methods for the rest of your career—invest wisely.
- Understanding your tools makes you better at your thinking process.

# Why build a foundation?

- Sophisticated user: use or understand new methods without a strong foundation in rigorous statistics.
- You will be using methods for the rest of your career—invest wisely.
- Understanding your tools makes you better at your thinking process.
- You should never have to abandon a project because “you don’t know how to do it.”

# What is and where and why Gov2001?

- We will focus on probability theory and linear models.
- More concrete math / stats topics, compared to the hidden curriculum.
- Prepare for more technical methods such as casual inference, Bayesian methods and etc.

# What is and where and why Gov2001?

- We will focus on probability theory and linear models.
- More concrete math / stats topics, compared to the hidden curriculum.
- Prepare for more technical methods such as casual inference, Bayesian methods and etc.
- This course is a direct replica of Prof. Matt Blackwell's Gov 2002, I also attached the course website on our syllabus page.

# Goals of this Course

- Have a solid understanding of three core topics:
  1. Probability
  2. Statistical Inference
  3. Linear Regression
- Overall goal: Be empowered to learn any new method with relative ease.



# Course Details: Staff

- Instructor: Naijia Liu
- Teaching Fellow: Hanning Luo

# Course Details: Staff

- Instructor: Naijia Liu
- Teaching Fellow: Hanning Luo
- Course website: <https://naijialiu.github.io/Gov2001/>
  - ▶ Slides
  - ▶ Slack
  - ▶ Psets, Exams and Gradescope
  - ▶ Scheduling OHs

# Course Details: Prerequisites

- Math:
  - ▶ Basic algebra and some exposure to basic statistics.
  - ▶ Calculus (limits, derivatives, integrals).
  - ▶ Linear algebra (vectors, matrices, etc.).
  - ▶ Content covered in Gov Math Prefresher (see syllabus).
- Computing:
  - ▶ Knowledge of R.

# Time Commitment

- Research, resting, health, food, friends, internship are more important than Gov2001:

# Time Commitment

- Research, resting, health, food, friends, internship are more important than Gov2001:
  - ▶ Past students of Gov 2002 spent 5–20 hours per week on homework alone.
  - ▶ This can be painful, but it is normal.
  - ▶ “Success in academia is a mix of luck, creativity, knowledge, and consistent hard work” - Blackwell, Fall 2023.

# Time Commitment

- Research, resting, health, food, friends, internship are more important than Gov2001:
  - ▶ Past students of Gov 2002 spent 5–20 hours per week on homework alone.
  - ▶ This can be painful, but it is normal.
  - ▶ “Success in academia is a mix of luck, creativity, knowledge, and consistent hard work” - Blackwell, Fall 2023.
  - ▶ Let us know if you need more time / space / help.
- Becoming “fluent” in methods will pay off in the long run.

# Teaching Resources

- Lecture: Theoretical topics, examples, etc.

# Teaching Resources

- Lecture: Theoretical topics, examples, etc.
- Sections: Review, examples, and more.



# Teaching Resources

- Lecture: Theoretical topics, examples, etc.
- Sections: Review, examples, and more.
- Course site: Contains most course materials.

# Teaching Resources

- Lecture: Theoretical topics, examples, etc.
- Sections: Review, examples, and more.
- Course site: Contains most course materials.
- Slack: Logistical and course materials discussions (please try to use general channel as much as possible).

# Teaching Resources

- Lecture: Theoretical topics, examples, etc.
- Sections: Review, examples, and more.
- Course site: Contains most course materials.
- Slack: Logistical and course materials discussions (please try to use general channel as much as possible).
- Office hours: Ask more questions.

# Textbooks

- Finish reading around lecture time.
- Peng Ding's new book offers a more technical level of things, highly recommend.
- Recommended books:
  - ▶ *Probability* by Blitzstein and Hwang (Stat 110 textbook).

# Grading

- Bi-Weekly homework assignments: 55%
- Take-home midterm exam: 15%
- Cumulative take-home final: 20%
- Participation: 10%

# Three main parts of the course

- We will go from Probability, to Inference, and end our discussion with Regression.
  - ▶ **Regression:** Core tool to estimate the relationship between variables.

# Three main parts of the course

- We will go from Probability, to Inference, and end our discussion with Regression.
  - ▶ **Regression:** Core tool to estimate the relationship between variables.
  - ▶ **Inference:** How to learn about unknowns from knowns.

# Three main parts of the course

- We will go from Probability, to Inference, and end our discussion with Regression.
  - ▶ **Regression:** Core tool to estimate the relationship between variables.
  - ▶ **Inference:** How to learn about unknowns from knowns.
  - ▶ **Probability:** What data we would expect if the truth were known.



# Deterministic versus stochastic

- **Key idea about statistics:** quantifying uncertainty
- Imagine someone comes to us and says, “what is the relationship between voter turnout and campaign spending?”

# Deterministic versus stochastic

- **Key idea about statistics:** quantifying uncertainty
- Imagine someone comes to us and says, “what is the relationship between voter turnout and campaign spending?”
- **Deterministic** account of voter turnout in a district:

$$\text{turnout}_i = f(\text{spending}_i)$$

# Deterministic versus stochastic

- **Key idea about statistics:** quantifying uncertainty
- Imagine someone comes to us and says, “what is the relationship between voter turnout and campaign spending?”
- **Deterministic** account of voter turnout in a district:

$$\text{turnout}_i = f(\text{spending}_i)$$

- What's the problem with this? Omits all other determinants:
  - ▶ Open seat, challenger quality, weather on election day
  - ▶ Whether the local college football team won the previous weekend

# Stochastic Models

- Measure everything and then add it to our model:

$$\text{turnout}_i = f(\text{spending}_i) + g(\text{stuff}_i)$$

# Stochastic Models

- Measure everything and then add it to our model:

$$\text{turnout}_i = f(\text{spending}_i) + g(\text{stuff}_i)$$

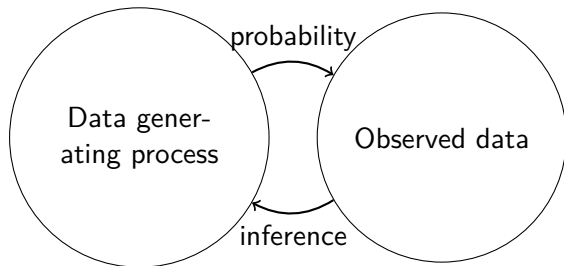
- Treat other factors as **stochastic**:
  - ▶ They affect the outcome, but are not of direct interest.
  - ▶ We think of them as part of the chance variation in turnout.
- How do we quantify chance variation: **probability**

# Stochastic Models

- Measure everything and then add it to our model:

$$\text{turnout}_i = f(\text{spending}_i) + g(\text{stuff}_i)$$

- Treat other factors as **stochastic**:
  - ▶ They affect the outcome, but are not of direct interest.
  - ▶ We think of them as part of the chance variation in turnout.
- How do we quantify chance variation: **probability**



# Why Probability?

- Next few weeks: **probability**
  - ▶ Life is uncertain.
  - ▶ Probability helps us study stochastic events.
  - ▶ Important for all of statistics.

# Why Probability?

- Next few weeks: **probability**
  - ▶ Life is uncertain.
  - ▶ Probability helps us study stochastic events.
  - ▶ Important for all of statistics.
- Statistical inference is a **thought experiment**.
- Probability is the logic of these thought experiments.



# Why Probability?

- Next few weeks: **probability**
  - ▶ Life is uncertain.
  - ▶ Probability helps us study stochastic events.
  - ▶ Important for all of statistics.
- Statistical inference is a **thought experiment**.
- Probability is the logic of these thought experiments.
- Thought experiments: assume men and women were paid the same on average, but there was chance variation from person to person.

# Why Probability?

- Next few weeks: **probability**
  - ▶ Life is uncertain.
  - ▶ Probability helps us study stochastic events.
  - ▶ Important for all of statistics.
- Statistical inference is a **thought experiment**.
- Probability is the logic of these thought experiments.
- Thought experiments: assume men and women were paid the same on average, but there was chance variation from person to person.
  - ▶ If true, how likely is the observed wage gap in this hypothetical world?

# Why Probability?

- Next few weeks: **probability**
  - ▶ Life is uncertain.
  - ▶ Probability helps us study stochastic events.
  - ▶ Important for all of statistics.
- Statistical inference is a **thought experiment**.
- Probability is the logic of these thought experiments.
- Thought experiments: assume men and women were paid the same on average, but there was chance variation from person to person.
  - ▶ If true, how likely is the observed wage gap in this hypothetical world?
  - ▶ What kinds of wage gaps would we expect to observe in this hypothetical world?

# Why Probability?

- Next few weeks: **probability**
  - ▶ Life is uncertain.
  - ▶ Probability helps us study stochastic events.
  - ▶ Important for all of statistics.
- Statistical inference is a **thought experiment**.
- Probability is the logic of these thought experiments.
- Thought experiments: assume men and women were paid the same on average, but there was chance variation from person to person.
  - ▶ If true, how likely is the observed wage gap in this hypothetical world?
  - ▶ What kinds of wage gaps would we expect to observe in this hypothetical world?

# Why Probability?

- Next few weeks: **probability**
  - ▶ Life is uncertain.
  - ▶ Probability helps us study stochastic events.
  - ▶ Important for all of statistics.
- Statistical inference is a **thought experiment**.
- Probability is the logic of these thought experiments.
- Thought experiments: assume men and women were paid the same on average, but there was chance variation from person to person.
  - ▶ If true, how likely is the observed wage gap in this hypothetical world?
  - ▶ What kinds of wage gaps would we expect to observe in this hypothetical world?
- **Probability may tell us the answer!**