Class 1: Introduction to quantitative analysis

API-201
Quantitative Analysis and Empirical Methods I

Profs. Goel and Taylor
Harvard Kennedy School

Agenda

1. Assessing claims of voter fraud
2. Discrimination in policing
3. Course overview
4. Next steps

Claims of voter fraud

Beliefs about double voting

About **25%** of the public believes that voting more than once happens either "commonly" or "occasionally".
[As opposed to "infrequently" or "never"]
[Stewart III et al. (2016)]

The statistical challenge

There is no unique national identifier that links voting records across states. Only consistently available info is **first name, last name, and date of birth**.

Estimation strategy

1. Start with complete national record of voting.
   [100M+ records]

2. Count records with matching first name, last name, DOB.
   [800k matches]

   What's going on?

The birthday paradox

In a group of 50 people, how many pairs of people have the same birthday [month & day]?

The birthday paradox

Coincidental matches are surprisingly common

In the 2012 presidential election, **141** ballots were cast by a "John Smith" born in 1970.

Among these ballots, **27** pairs had the exact same birthdate, and so would naively be flagged as double votes.
[Same name and exact date of birth]

But you would expect to see ___ matches by chance alone!

The birthday paradox

In a room full of **141** people, there are nearly **10,000** distinct *pairs* of people. **What's the likelihood two random people have the same birthday?**

Estimation strategy

1. Start with complete national record of voting.
   [100M+ records]

2. Count records with matching first name, last name, DOB.
   [800k matches]

3. Adjust estimate by matches expected to occur by chance.
   [Generalization of the *birthday paradox*]

Estimated number of double votes

After birthday paradox correction...

800k → 30k

After error correction...

30k → _____

[Based on 1% error rate in Philadelphia]

One Person, One Vote: Estimating the Prevalence of Double Voting in U.S. Presidential Elections
Goel, Meredith, Morse, Rothschild & Shirani-Mehr
American Political Science Review [2020]

Stop and Frisk

Officers can legally stop and question individuals when there is “reasonable suspicion” of criminal activity.

Until recently, 500,000 stops conducted annually in NYC
[substantially curtailed at the end of 2013]

Question: Is the policy racially discriminatory?

Stop and Frisk

Fact: 80% of stops involved Black or Hispanic individuals.

Fact: 50% of NYC population was Black or Hispanic.

Stop and Frisk

To test for discrimination, we'll look at stop *success* rates.
[An "outcome test"]

The hit rate
Proportion of stops that are "successful"

Lower hit rate for Black individuals compared to White individuals suggests a double standard, with a lower [and discriminatory] bar for stopping Black people than White people.

Outcome tests

How do you detect racial discrimination in bank lending?
[Discuss with your neighbor.]

Stop and frisk

Hit rate
Percentage of CPW stops that turn up a weapon

Stops of Black individuals
____ hit rate

Stops of White individuals
____ hit rate

Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy
Goel, Rao & Shroff
Annals of Applied Statistics [2016]

The role of statistics in reform

What is the role of statistics in public policy?
[Discuss with your neighbor.]

What you need to know about API-201

1. Teaching approach
    - **Philosophy:** Focus on deep conceptual understanding, applied to real-world problems
    - **Handouts:** take notes, work on problems, note questions
    - **Participation:** highly valued, in and out of class

What you need to know about API-201

1. Teaching approach
2. Communication

All course-related information is on your Canvas homepage. For discussion, use

- **#askforhelp**: questions and answers about course
- **#r**: R-specific questions and answers
- **#random**: articles, tweets, quotes, overheard,...
- **#section-c/d**: channel for Section C/D

What you need to know about API-201

1. Teaching approach
2. Communication
3. Resources

- **TF sessions [starting next week]**
  - Tuesdays 4:30 - 5:45 [Starr Auditorium — Belfer 200]
  - Attend each week — crucial for completing problem sets
- **Teaching Team office hours [posted next week]**
- **Faculty office hours [starting next week]**

What you need to know about API-201

1. Teaching approach
2. Communication
3. Resources
4. Expectations
    - Arrive to class **on time**, ready to engage
    - Generally no technology in class other than **Poll Everywhere**
    - Bring your **laptop** for in-class R exercises
    - Complete **pre-class exercises** the day before each class at 5 PM
    - Complete **problem sets** by Thursdays at 8am
    - Check **exam dates on syllabus** to make sure you are available

Next Steps

- ➔ Pre-Class Exercise due Monday at 5 PM
    - ◆ Available on Canvas
- ➔ Problem Set #1 due Thursday at 8am
    - ◆ Aimed at familiarizing you with Jupyter Notebooks
    - ◆ Submit as PDF on Canvas
- ➔ If you have a seating need, contact your faculty member
    - ◆ By tomorrow [Friday] at 5 PM