Introduction to Statistics

Kosuke Imai

Department of Politics
Princeton University

Fall 2011

Descriptive and Statistical Inference

- **Descriptive inference**:
  1. Summarize the observed data
  2. Tables with statistics, Data visualization through graphs
  3. Statistic $=$ a function of data

- **Statistical inference**:
  1. Learning about unknown parameters from observed data
  2. Statistical models: All models are false but some are useful
  3. Uncertainty: How confident are you about your inference?
  4. Statistical tests: Does smoking cause cancer?

- **Research design**:
  1. What data to collect and how?
  2. Designing experiments, sample surveys, etc.
  3. Mimicking experiments in observational studies

Examples of Statistical Inference

Inference types	Parameters	Data
Forecasting	future data points	past data
Sample survey	features of population	random sample
Causal inference	counter-factuals	factuals

Causal Inference

- What does "*A* causes *B*" mean?
- Counterfactuals: "what if" questions
- What we don't observe: counterfactual outcome
- What we observe: factual outcome
- Quantity of interest: differences between the factual (observed) and counterfactual (unobserved) outcomes
- Does the minimum wage increase the unemployment rate?
  - Factual: observed unemployment rate given the enacted minimum wage increase
  - Counterfactual: (unobserved) unemployment rate that would have been realized had the minimum wage increase not occurred
- No causation without manipulation: immutable characteristics
- Does race affect your job prospect?

Statistical Framework of Causal Inference

- "Treatment": Contacted or not
- *Observed* outcome: Turnout
- *Pre-treatment* variables: e.g., Age, Party ID
- Potential outcomes:

Voters	Contact	Turnout		Age	Party ID
		Treated	Control		
John	Yes	Yes	?	20	D
Nicole	No	?	No	55	R
Mark	No	?	Yes	40	R
⋮	⋮	⋮	⋮	⋮	⋮
Amy	Yes	No	?	62	D

- Causal effect: Difference between two potential outcomes
- Fundamental problem of causal inference

How to Figure Out the Counterfactuals?

- Association is not causation
- Find a similar unit!

- Outcome: voter turnout
- Treatment: get-out-the vote phone calls
- Question: How much does a GOTV call increase turnout?

- Find an observation with same voting history, partisanship, age, gender, job, income, education, etc...
- Match on observed confounders that are systematically related to *both* treatment and outcome
- But, we cannot match on everything
- Selection bias due to unobserved confounding in observational studies

Randomized Experiments

- Randomize!
- Key idea: Randomization of the treatment makes the treatment and control groups "identical" *on average*
- The two groups are similar in terms of *all* (both observed and unobserved) characteristics
- Can attribute the average differences in outcome to the difference in the treatment
- Randomized experiments as the gold standard
- Placebo effects and double-blind experiments

- READING: FPP Chapter 1

Example 1: Get-Out-the-Vote Experiments

- What methods and messages work best for which voters?
- Selection bias: campaigns target certain voters
- Randomize the message each voter receives
- Hundreds of GOTV field experiments by political scientists
- Civic duty message:

Dear Registered Voter:

DO YOUR CIVIC DUTY AND VOTE!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

The whole point of democracy is that citizens are active participants in government; that we have a voice in government. Your voice starts with your vote. On August 8, remember your rights and responsibilities as a citizen. Remember to vote.

DO YOUR CIVIC DUTY — VOTE!

- Social pressure message:

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted i the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	_____
9995 JENNIFER KAY SMITH		Voted	_____
9997 RICHARD B JACKSON		Voted	_____
9999 KATHY MARIE JACKSON		Voted	_____

Gerber, Green, and Larimer. (2008). *American Political Science Review*.

- Turnout: 31.5% (civic duty), 37.8% (social pressure)

Example 2: Employment Discrimination

- Does race affect employment prospect?
- Selection bias: African Americans and Whites are different in terms of qualifications (e.g., education levels)
- Randomize applicant names in their resume
- Outcome: call-back rate

White female		Black female		White male		Black male	
Emily	7.9	Aisha	2.2	Todd	5.9	Rasheed	3.0
Anne	8.3	Keisha	3.8	Neil	6.6	Tremayne	4.3
Jill	8.4	Tamika	5.5	Geoffrey	6.8	Kareem	4.7
Allison	9.5	Lakisha	5.5	Brett	6.8	Darnell	4.8
Laurie	9.7	Latoya	5.8	Brendan	7.7	Tyrone	5.3

Bertrand and Mullainathan. (2004). *American Economic Review*

Experimental vs. Observational Studies

- READING: FPP Chapter 2

- Hormone-Replacement Therapy (HRT)
- Estrogen to replace the hormones
- Nurses' Health Study (1985; observational study):
  HRT reduced the risk of heart disease by two thirds
- In 2001, 15 million women were filling HRT prescriptions

- Two randomized clinical trials:
  1. Heart and Estrogen-progestin Replacement Study (1998)
  2. Women's Health Initiative (2002)
- HRT *increases* the risk of heart disease, stroke, blood clots, breast cancer, etc.
- Where did this stark difference come from?

Possible Explanations

1. Selection bias: Healthy-user bias and prescriber effects

2. Different populations:
   - RCTs: older women after menopause
   - Nurses' Study: younger women near the onset of menopause
   - New hypothesis: HRT may protect younger women against heart disease while being harmful for older women

Potential Objections to Experiments

1. Cost
   - Move-to-Opportunity experiment; $70million, 350,000 people
2. Ethical concerns
   - Experiments affecting the real world
   - Denying control units access to beneficial treatments
   - Treatments with potential harm
3. Internal vs. external validity
   - Generalizability: laboratory vs. field experiments
   - Sample selection
   - Hawthorne effects
4. Black box
   - Causal mechanisms: “why” rather than “whether”

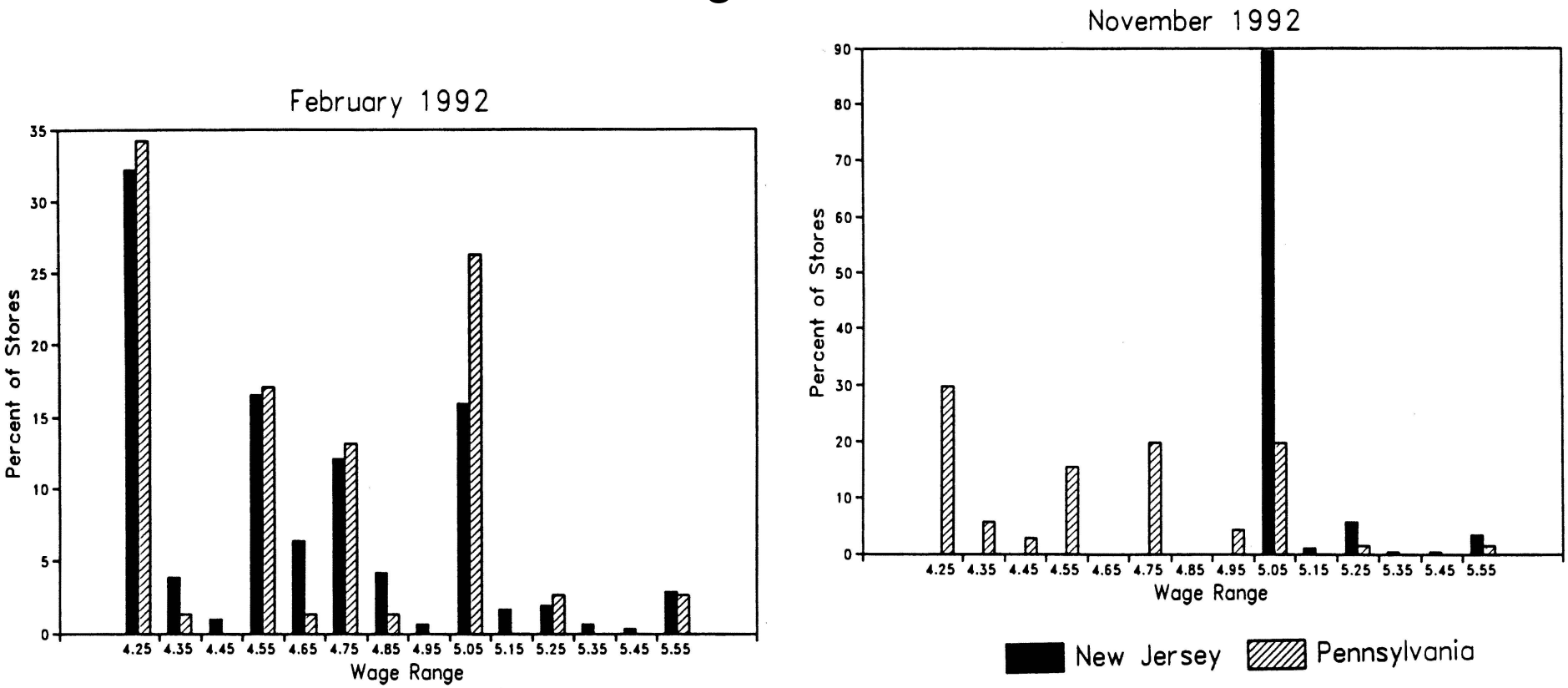
Designing Observational Studies

- We still need observational studies: we often can't do randomized experiments for ethical and practical reasons
- The main problem: internal validity
- Selection bias: lack of randomized treatments

- Strategy: Find a situation where treated units are similar to control units
- Natural experiments: treatments are haphazardly determined

Example: Effect of Raising Minimum Wage

- Before-and-after design
- In 1992, NJ raises minimum wage from \$4.25/hr to \$5.05/hr
- In (eastern) PA, the minimum wage remained at \$4.25/hr
- No evidence found for negative effects



Card and Krueger (1994). *American Economic Review*