PSC4375: Measurement and Survey Sampling

Week 4: Lecture 7

Prof. Weldzius

Villanova University

Slides Updated: 2025-01-25

Concepts and measurement

- Social science is about understanding **causal relationships**
  - Does minimum wage change levels of employment
  - Does outgroup contact influence views on immigration?
- Relationships are between **concepts**:
  - Minimum wage, unemployment, outgroup contact, views on immigration
  - We took these for granted when talking about causality
- Important to consider how we **measure** these concepts
  - Some straightforward: what is your age?
  - Others more complicated: what does it mean to "be liberal"?
  - **Operational definition**: mapping of concept to numbers in our data

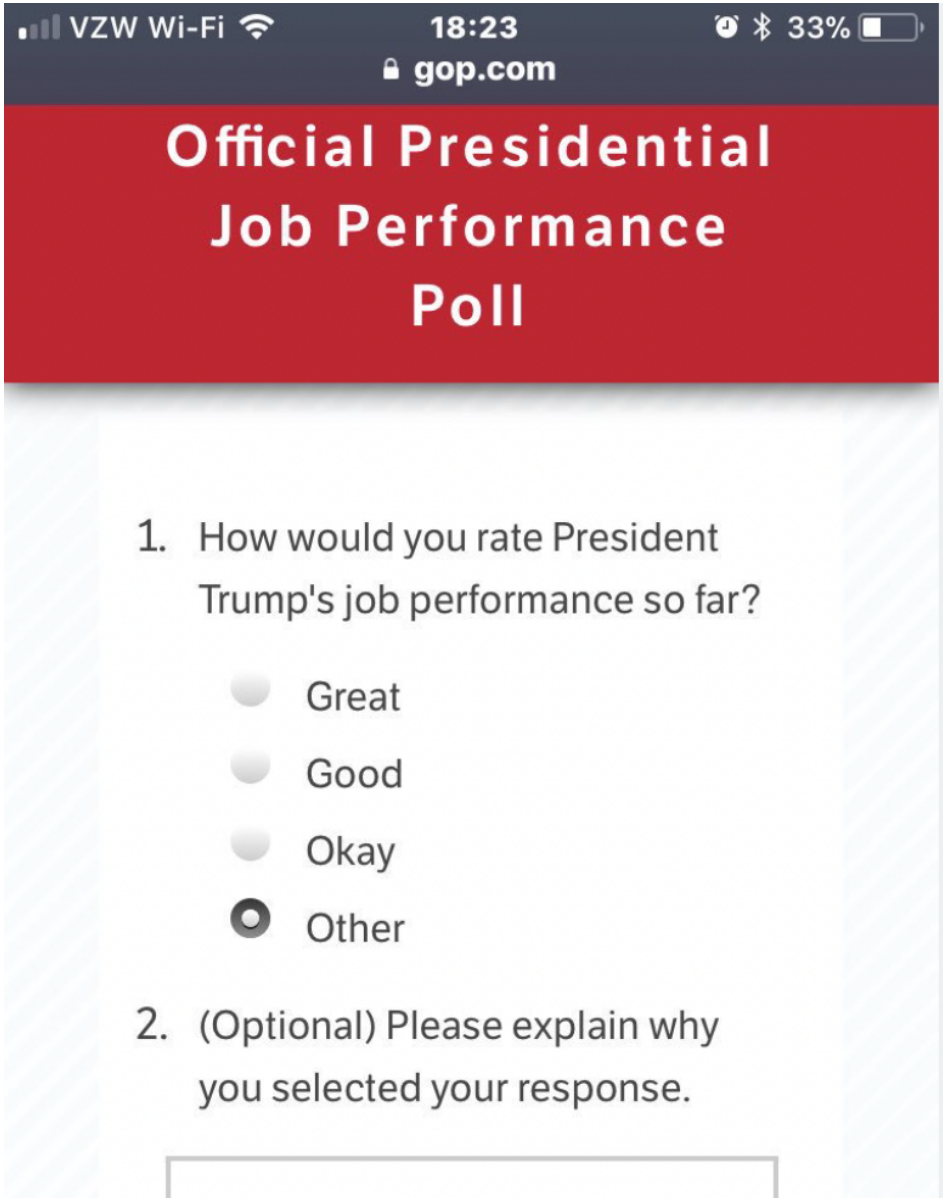
Example

- Concept: presidential approval
- Conceptual definition:
  - Extent to which US adults support the actions and policies of the current US president
- Operational definition:
  - "On a scale from 1 to 5, where 1 is least supportive and 5 is most supportive, how much would you say you support the job that Donald Trump is doing as president?"

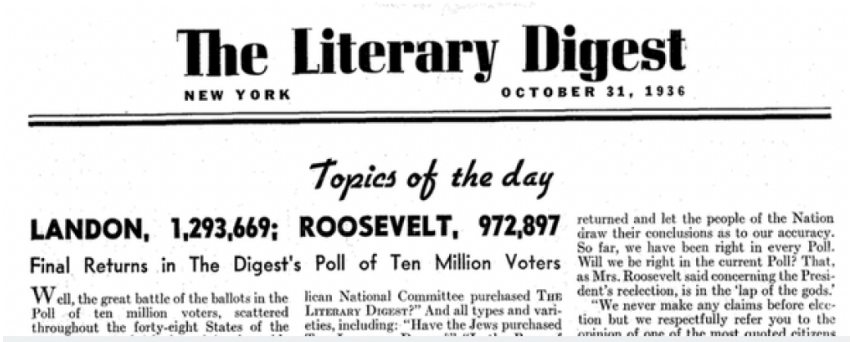
Measurement error

- **Measurement error**: chance variation in our measurements
  - individual measurement = exact value + chance error
  - chance errors tend to cancel out when we take averages
- No matter how careful we are, chance error can always affect a measurement.
  - Panel study of 19,000 respondents: 20 reported being a citizen in 2010 and then a non-citizen in 2012
  - Data entry errors
- **Bias**: systematic errors for all units in the same direction.
  - individual measurement = exact value + bias + chance error
  - "What did you eat yesterday?" $\rightsquigarrow$ underreporting

A biased poll?



1936 Literary Digest Poll



The Literary Digest

NEW YORK OCTOBER 31, 1936

Topics of the day

LANDON, 1,293,669; ROOSEVELT, 972,897

Final Returns in The Digest's Poll of Ten Million Voters

Well, the great battle of the ballots in the Poll of ten million voters, scattered throughout the forty-eight States of the

lican National Committee purchased THE LITERARY DIGEST?" And all types and varieties, including: "Have the Jews purchased

returned and let the people of the Nation draw their conclusions as to our accuracy. So far, we have been right in every Poll. Will we be right in the current Poll? That, as Mrs. Roosevelt said concerning the President's reelection, is in the 'lap of the gods.'

"We never make any claims before election but we respectfully refer you to the

- Literary Digest predicted elections using mail-in polls
- Source of addresses: automobile registrations, phone books, etc.
- In 1936, sent out 10 million ballots, over 2.3 million returned
- George Gallup used only 50,000 respondents

	FDR's vote share
Literary Digest	43
George Gallup	56

Poll fail

	FDR %
Literary Digest	43
George Gallup	56
Actual Outcome	62

- **Selection bias**: ballots skewed toward the wealthy (with cars, phones)
  - Only 1 in 4 households had a phone in 1936
- **Nonresponse bias**: respondents differ from nonrespondents
  - $\rightsquigarrow$ when selection procedure is biased, adding more units won't help!

1948 Election

The Polling Disaster

	Truman	Dewey	Thurmond	Wallace
Crossley	45	50	2	3
Gallup	44	50	2	4
Roper	38	53	5	4
Actual Outcome	50	45	3	2

- **Quota sampling**: fixed quota of certain respondents for each interviewer
  - If Black women make up 5% of the population, stop interviewing them once they make up 5% of your sample
- Sample resembles the population on these characteristics
- Potential unobserbed confounding $\rightsquigarrow$ **selection bias**
- Republicans easier to find within quotas (phones, listed addresses)

Sample surveys

- **Probability sampling** to ensure representativeness
  - Definition: every unit in the population has a known, non-zero probability of being selected into sample
- **Simple random sampling**: every unit has an equal selection probability.
- Random digit dialing:
  - Take a particular area code + exchange: 310-495-XXXX.
  - Randomly choose each digit in XXXX to call a particular phone
  - Every phone in the US has an equal chance of being included in sample

Sampling lingo

- **Target population**: set of people we want to learn about
  - Example: people who will vote in the next election
- **Sampling frame**: list of people from which we will actually sample
  - Frame bias: list of registered voters (frame) might include nonvoters!
- **Sample**: set of people contacted
- **Respondents**: subset of sample that acutally responds to the survey
  - Unit non-response: sample $\neq$ respondents
  - Not everyone picks up their phone
- **Completed items**: subset of questions that respondents answer
  - Item non-response: refusing to disclose their vote preference

Difficulties of sampling

- Problems of telephone survey
  - Cell phones (double countring for the wealthy)
  - Caller ID screening (unit non-response)
  - Response rates down to 9%
- An alternative: internet surveys
  - Opt-in panels, respondent-driven sampling $\rightsquigarrow$ **non-probability sampling**
  - Cheaper, but non-representative
  - Digital divide: rich vs. poor, young vs. old
  - Correct for potential sampling bias via stastical methods