PSC4375: Measurement and Survey Sampling Week 4: Lecture 7

Prof. Weldzius

Villanova University

Slides Updated: 2025-02-10

Group small talk

- Introduce yourselves (again)
- What did you do for the Super Bowl?
 - If you watched it: favorite moment?
 - If you didn't watch it: what did you do instead?

Social science is about understanding causal relationships

- Social science is about understanding causal relationships
 - Does minimum wage change levels of employment

- Social science is about understanding causal relationships
 - Does minimum wage change levels of employment
 - Does outgroup contact influence views on immigration?

- 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:

- 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

- 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

- 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

- 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?

- 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"?

- 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

Concept: presidential approval

- Concept: presidential approval
- Conceptual definition:

- Concept: presidential approval
- Conceptual definition:
 - Extent to which US adults support the actions and policies of the current US president

- Concept: presidential approval
- Conceptual definition:
 - Extent to which US adults support the actions and policies of the current US president
- Operational definition:

- 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: chance variation in our measurements

- Measurement error: chance variation in our measurements
 - individual measurement = exact value + chance error

- Measurement error: chance variation in our measurements
 - individual measurement = exact value + chance error
 - chance errors tend to cancel out when we take averages

- 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.

- 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

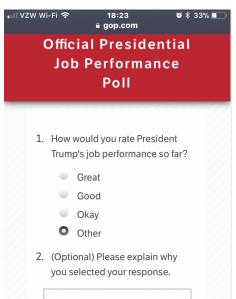
- Measurement error: chance variation in our measurements
 - individual measurement = exact value + chance error
 - chance errors tend to cancel out when we take averages
- No matter how careful we are, chance error can always affect a measurement.
 - Panel study of 19,000 respondents: 20 reported being a citizen in 2010 and then a non-citizen in 2012
 - Data entry errors

- 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.

- 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

- 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

A biased poll?



The Literary Digest

Topics of the day

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

Final Returns in The Disperts Poll of Ten Million Votes

Was the property of the Nation American State Control of the Nation American Americ

The Literary Digest

Topics of the day

LANDON, 1,293,669; ROOSEVELT, 972,897 Final Returns in The Digest's Poll of Ten Million Voters Will we be right in the current Polli

Well, the great battle of the ballots in the lican National Committee purchased Tan Pell of ten militon voters, scattered Lermany Donors' And all types and varieties when the pell of the

returned and let the people of the Nation draw their conclusions as to our accuracy So far, we have been right in every Pol dent's reelection, is in the 'lap of the gods.

Literary Digest predicted elections using mail-in polls

The Literary Digest

W YORK OCTOBER 31, 19

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 ballets in the
Foll of ten million voters, scattered
LITERANY DORSEY? And all types and varithroughout the first-right States of the etise, including: "Have the eyes preclased to

returned and let the people of the Nation draw their conclusions as to our accuracy. So far, we have been right in every Poli. Will we be right in the current Poll? That, as Mrs. Roosevelt said concerning the President's redection, is in the 'lap of the gols.' "We never make any claims before election but we respectfully refer you to the continuous of the word control of the pro-

- Literary Digest predicted elections using mail-in polls
- Source of addresses: automobile registrations, phone books, etc.

The Literary Digest

EW YORK OCTOBER 31, 1

Topics of the day

LANDON, 1,293,669; ROOSEVELT, 972,897
Final Returns in The Digast's Poll of Ten Million Voters
Well, the great hattle of the haltes in the
Bona National Countries procedured in
Hornogloon the forty-sight States of the
eins, including: "lave the deep purchased
the relation of the procedure of th

returned and let the people of the Nation draw their conclusions as to our accuracy. So far, we have been right in every Poli. Will we be right in the current Poll? That, as Mrs. Roosevelt said concerning the President's redection, is in the Tap of the gods. "We never make any claims before elections 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

The Literary Digest

YORK OCTOBER 31, 15

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 ballets in the
Poll of ten milities voters, scattered
LTERRANT DORSET' And all types and varithroughout the first-right States of the
etis-including: "Have the eves prechased to

returned and let the people of the Natios 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. Roservelt said concerning the Proident's reciection, is in the lap of the goods "We server 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

The Literary Digest

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 hattle of the ballets in the lican National Committee purchased Tau
Poll of ten million voters, scattered Lirmans Dosses?" And all types and varithroughout the forty-right States of the eries, including: "Have the eves purchased

returned and let the people of the Nation draw their conclusions as to our accuracy. So far, we have been right in every Foll. Will we be right in the current Poll That. as Mrs. Roosevelt and concerning the President's redection, is in the lap of the gold: "We never make any claims before elections 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 |



| | FDR % |
|-----------------|-------|
| Literary Digest | 43 |
| George Gallup | 56 |
| Actual Outcome | 62 |



| | FDR % |
|-----------------|-------|
| Literary Digest | 43 |
| George Gallup | 56 |
| Actual Outcome | 62 |

• Selection bias: ballots skewed toward the wealthy (with cars, phones)

8/13



| | 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



| | 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



| | FDR % |
|-----------------|-------|
| Literary Digest | 43 |
| George Gallup | 56 |
| Actual Outcome | 62 |

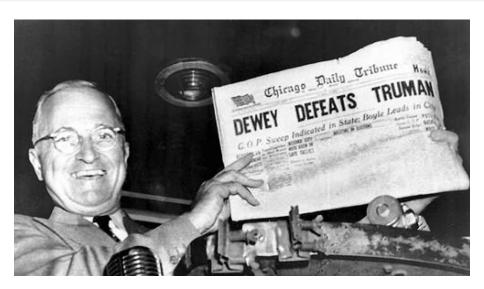
- Selection bias: ballots skewed toward the wealthy (with cars, phones)
 - Only 1 in 4 households had a phone in 1936
- Nonresponse bias: respondents differ from nonrespondents
 - \rightsquigarrow when selection procedure is biased, adding more units won't help!



| | 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



| | Truman | Dewey | ${\sf Thurmond}$ | Wallace |
|----------------|--------|-------|------------------|---------|
| Crossley | 45 | 50 | 2 | 3 |
| Gallup | 44 | 50 | 2 | 4 |
| Roper | 38 | 53 | 5 | 4 |
| Actual Outcome | 50 | 45 | 3 | 2 |

| | Truman | Dewey | ${\sf Thurmond}$ | 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

| | Truman | Dewey | ${\sf Thurmond}$ | 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

| | Truman | Dewey | ${\sf Thurmond}$ | 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

| | Truman | Dewey | ${\sf Thurmond}$ | 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 → selection bias

| | Truman | Dewey | ${\sf Thurmond}$ | 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 → selection bias
- Republicans easier to find within quotas (phones, listed addresses)

• **Probability sampling** to ensure representativeness

- Probability sampling to ensure representativeness
 - Definition: every unit in the population has a known, non-zero probability of being selected into sample

11 / 13

- 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.

- 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:

- 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.

- 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

- 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

• Target population: set of people we want to learn about

- Target population: set of people we want to learn about
 - Example: people who will vote in the next election

12 / 13

- 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

- 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!

12 / 13

- 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

- 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

- 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 ≠ respondents

- 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

- 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

- 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

PSC4375: Measurement and Survey Samplin

Problems of telephone survey

- Problems of telephone survey
 - Cell phones (double countring for the wealthy)

- Problems of telephone survey
 - Cell phones (double countring for the wealthy)
 - Caller ID screening (unit non-response)

13 / 13

- Problems of telephone survey
 - Cell phones (double countring for the wealthy)
 - Caller ID screening (unit non-response)
 - Response rates down to 9%

- 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

- 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 → non-probability sampling

13 / 13

- 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 → non-probability sampling
 - Cheaper, but non-representative

- 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 → non-probability sampling
 - Cheaper, but non-representative
 - Digital divide: rich vs. poor, young vs. old

- 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 → non-probability sampling
 - Cheaper, but non-representative
 - Digital divide: rich vs. poor, young vs. old
 - Correct for potential sampling bias via stastical methods