Surveys

POLI SCI 210

Introduction to Empirical Methods in Political Science

AI Prompts

- *Sampling*: Mode, frame, size, procedure, oversampling
- Independent, identically distributed (i.i.d. assumption in statistics)
- Total survey error framework
- How to ask *good* survey questions?
- What *happens* if people lie in surveys?
- What *to do* if people lie in surveys?
- *Indirect questioning techniques:* Randomized response, list experiment, others not mentioned in lecture

Last week

Ingredients for **statistical inference**

- Data summary (point estimate, test statistic)
- Uncertainty quantification (confidence intervals, p-values)
- **Keep in mind:** Statistical properties that enable inference (finite sample, asymptotic)

This week: Surveys

Survey research: Study of public opinion

Not really a *method*, more of a *data collection tool*

Tuesday: Sampling

1. Why random sampling?
2. Sources of error

Thursday: Survey design

1. Asking good questions
2. Asking sensitive questions

public opinion

Add to list | Share

The way most ordinary people feel about something, or the thing they mainly believe, is *public opinion*. If *public opinion* supports a no smoking policy, you better put down that cigar.

Politicians car a lot about public opinion — it's what gets them elected, or ultimately leads to their losing their jobs. Brave legislators and public figures will say what they really believe, rather than what they thing will please public opinion, but that's rare. The term was coined by John Locke in a 1600s essay he wrote about politics and human understanding, inspired by the French *l'opinion*, "opinion."

vocabulary.com/dictionary/public%20opinion

Public opinion

- The study of *self-reported* **attitudes** and **behaviors**
- **Self-report:** What people *say*
- **Attitude:** What people *think* or *feel*
- **Behavior:** What people *do*
- Primarily among the *general public*
- Increasingly on specific *target populations*

Goal: Mapping **self-reports** to **actual** attitudes and behaviors

Open question

Should political scientist care about **attitudes** on their own?

Do we only care about **behaviors**?

Challenge

There is a gap between *self-reports* and *actual* attitudes/behaviors

Two things get in the way:

1. Asking the right people
2. Asking the right questions

Challenge

There is a gap between *self-reports* and *actual* attitudes/behaviors

Two things get in the way:

1. **Asking the right people**
2. Asking the right questions

Key sampling decisions

1. Mode
2. Sampling frame
3. Sample size
4. Sampling procedure
5. Oversampling

Key sampling decisions

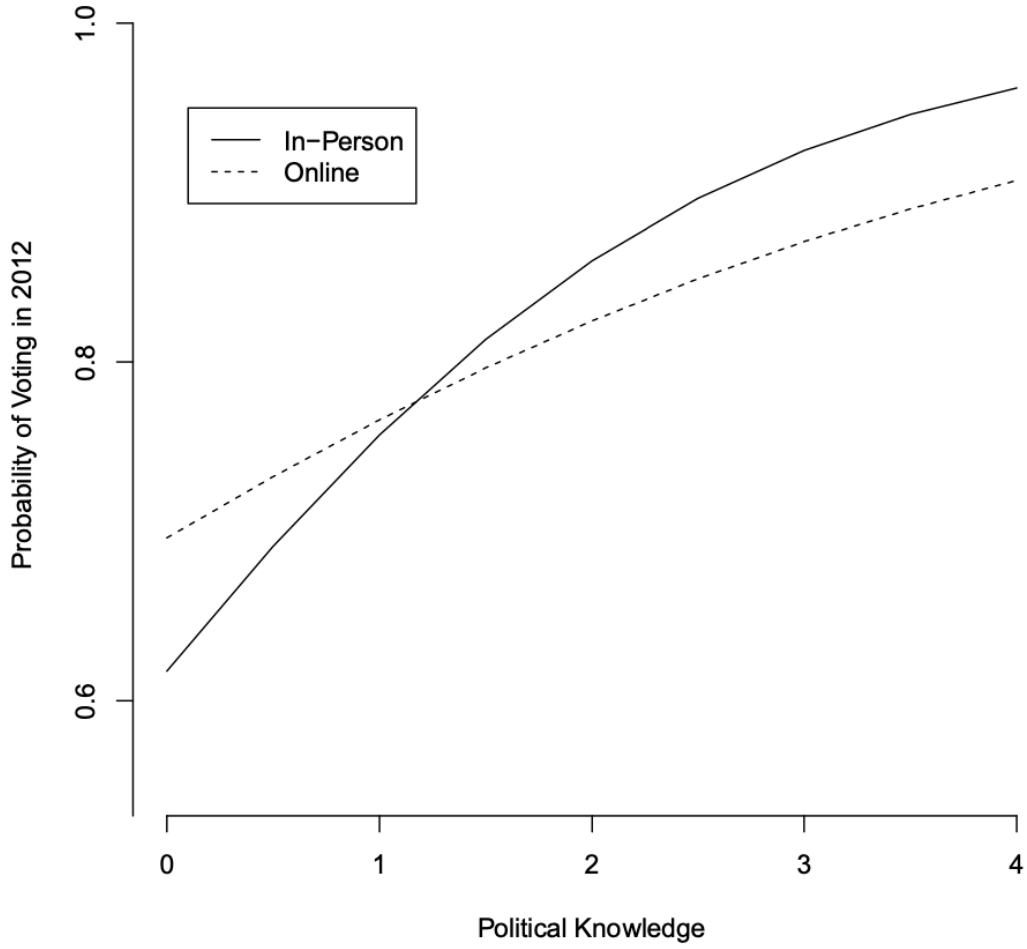
1. **Mode**
2. **Sampling frame**
3. ~~Sample size~~
4. **Sampling procedure**
5. ~~Oversampling~~

Survey mode

- **Face-to-face:** Oldest, most common method in US national and cross-national surveys
- **Laboratory:** Few people, more intricate tasks
- **Phone:** Groundbreaking at the time, but less reliable nowadays
- **Internet:** Online panels (CloudResearch Connect, Mturk, Prolific), social media ads, message app chains

National representative surveys usually combine many modes

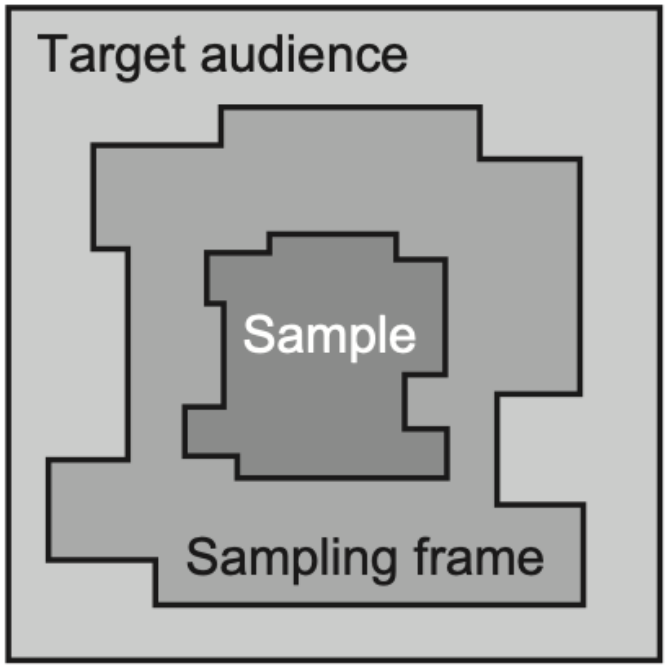
ANES evaluation of survey modes



Source: Hillygus et al (2017). "Assessing the Implications of a Mode Change."

Sampling frame

List of all the elements within the population *from which a sample is selected*



- *Common sampling frames:* Phone books, electoral rolls, employee databases, household directories
- *Ideal case:* Danish Civil Registry
- 🔥 *take:* We neve *really* know what the *target population* is

Sampling procedure

***Ideal case*: Random sampling**

Simple random sampling: All units have the same inclusion probability ($\approx$ flipping a coin for each)

Complete random sampling: *Exactly* n out of N units are selected at random ($\approx$ arrange in random order and then survey the first n)

How are these different?

Why is random sampling important?

Informally: So that our sample is (in expectation) *representative* of the target population

Formally: To justify statistical inference via CLT

We need to assume that the observations in our sample are drawn from…

Independent

Why is random sampling important?

Informally: So that our sample is (in expectation) *representative* of the target population

Formally: To justify statistical inference via CLT

We need to assume that the observations in our sample are drawn from…

Independent and **I**dentically

Why is random sampling important?

Informally: So that our sample is (in expectation) *representative* of the target population

Formally: To justify statistical inference via CLT

We need to assume that the observations in our sample are drawn from…

Independent and **I**dentically **D**istributed

… data generation processes

This is known as the **i.i.d assumption**

i.i.d. implications

Each survey respondent has an internal, unobservable **function** that draws different survey responses with some probability

Independent: How *I* come up with answers does not affect how *you* come up with answers

Identically distributed:

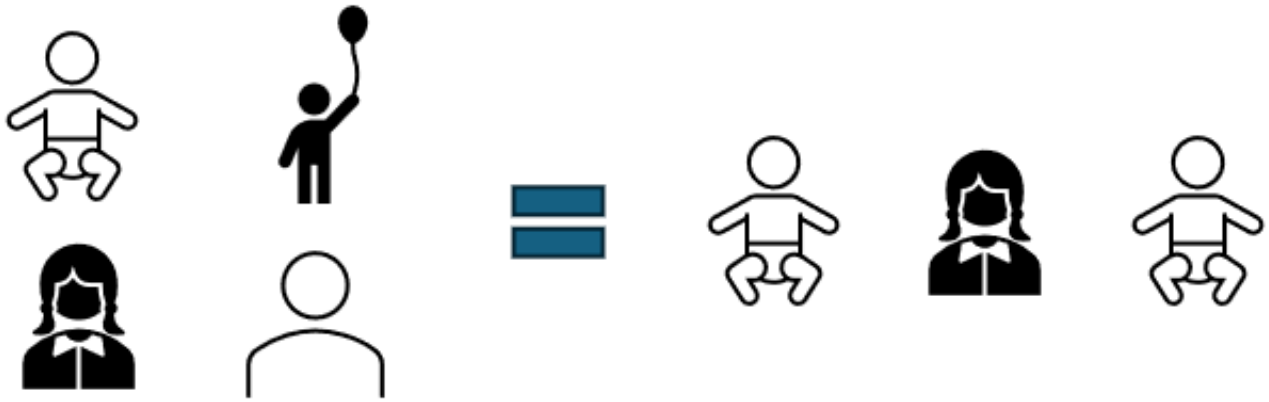
- Do not know how responses are drawn
- Assume is the same process for everyone

We call these *unobservable functions* random variables

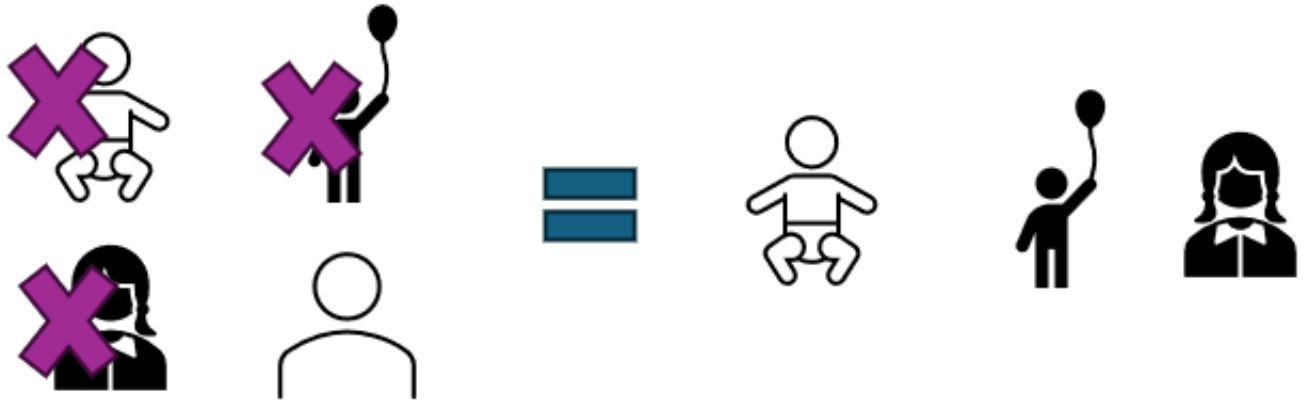
i.i.d. implications

This is equivalent to assuming that individual units are sampled at random **with replacement**

Sampling with replacement



Sampling without replacement



i.i.d. implications

This is equivalent to assuming that individual units are sampled at random **with replacement**

But we can only do random sampling *without replacement*

→ A decent approximation if sample size is large enough

> **Important**
>
> **Random sampling** lets us justify the **i.i.d. assumption**, which in turn allows using the standard formula for **confidence intervals** and **p-values**.

Problems with random sampling?

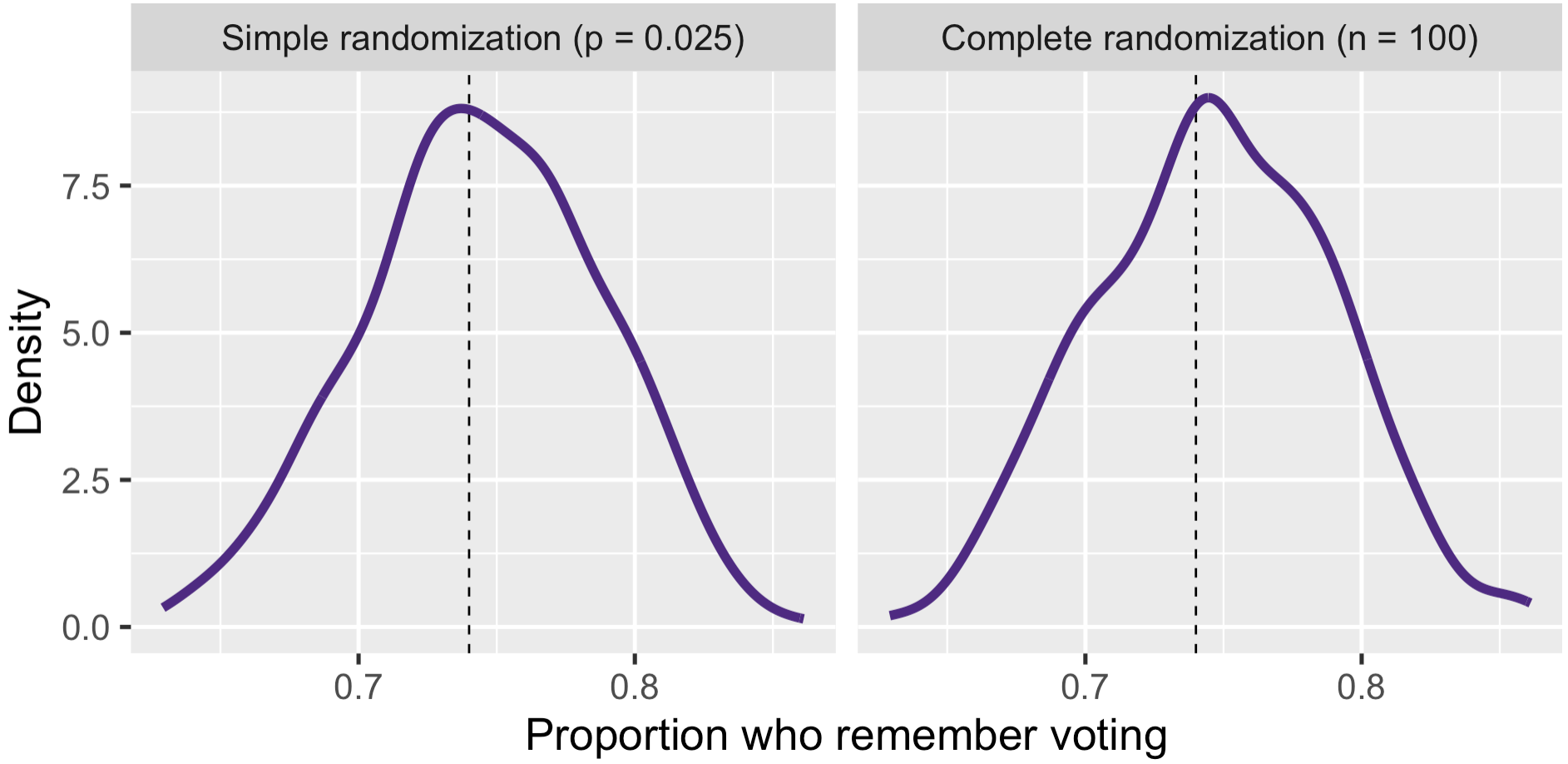
It’s expensive and can go wrong!

Back to GSS data

```
# A tibble: 3,799 × 4
     vote  sex   hispanic college
     <chr> <chr> <chr>    <chr>
   1 yes   woman no       yes
   2 yes   man   no       yes
   3 yes   woman no       no
   4 yes   woman no       yes
   5 yes   man   no       no
   6 yes   man   no       no
   7 no    woman yes      no
   8 yes   man   yes      yes
   9 yes   woman no       no
  10 yes   woman no       no
  11 yes   woman no       no
  12 yes   woman yes      yes
  13 yes   woman no       no
```

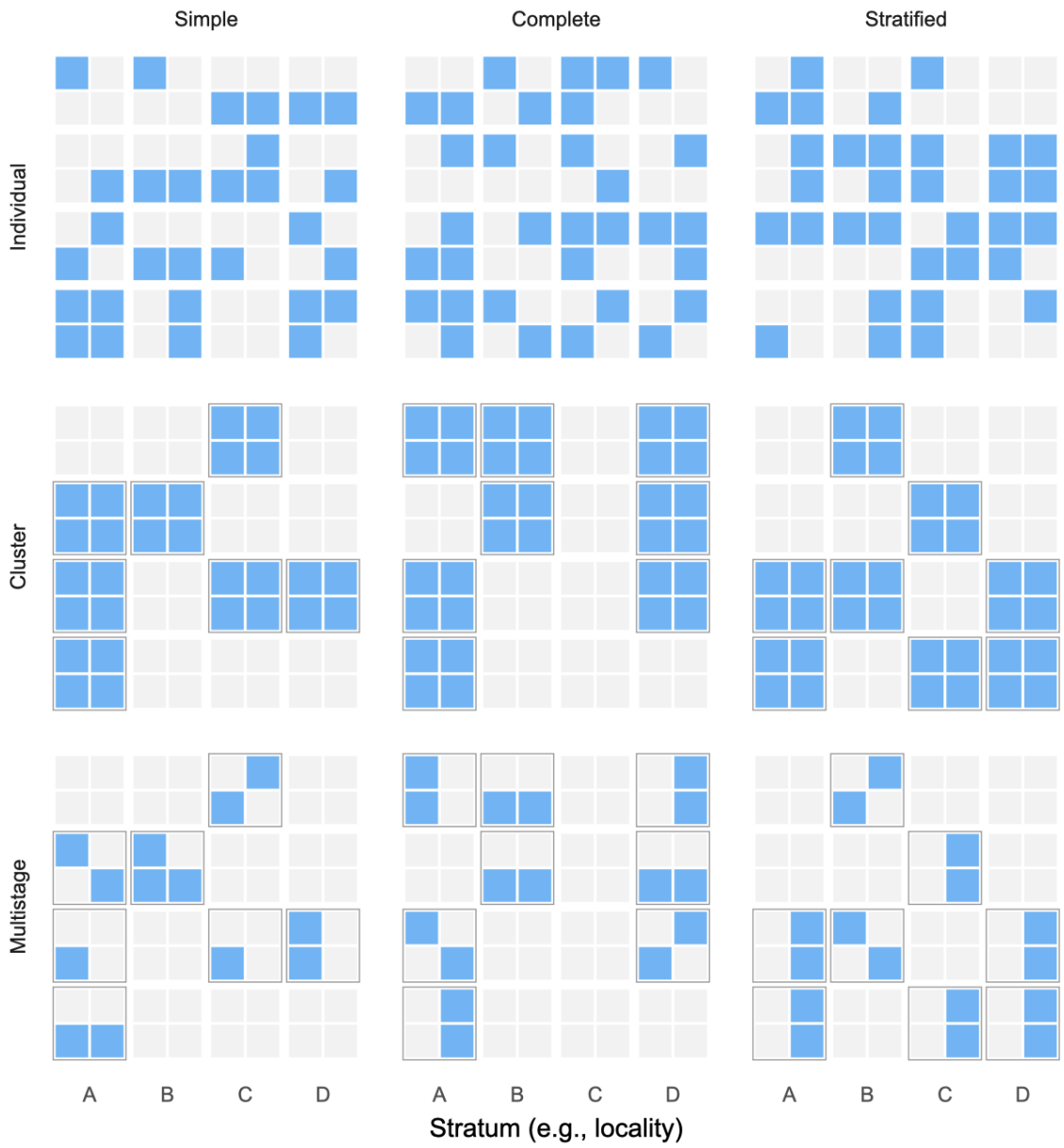
→ Take this as a population and compare different sampling procedures

500 simulations

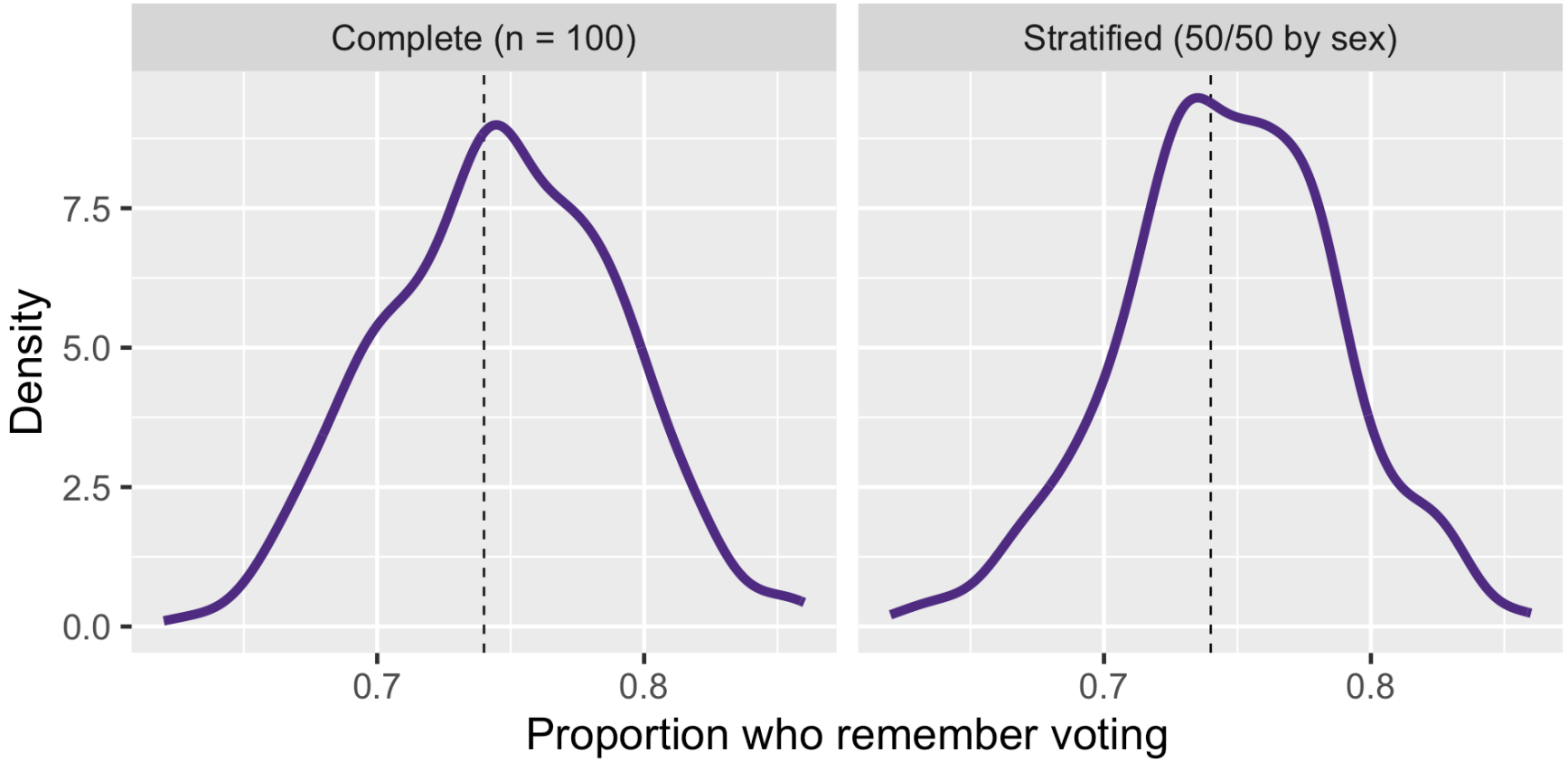


Variants of random sampling

- **Stratified random sampling:** Complete random simpling within groups/strata
- **Cluster random sampling:** Whole groups brought into sample together
- **Stratified cluster sampling:** First make groups, then sample clusters within each groups
- **Multi-stage random sampling:** First clusters, then units within clusters


Simple
Complete
Stratified
Individual
Cluster
Multistage
A
B
C
D
A
B
C
D
A
B
C
D
Stratum (e.g., locality)

What changes?



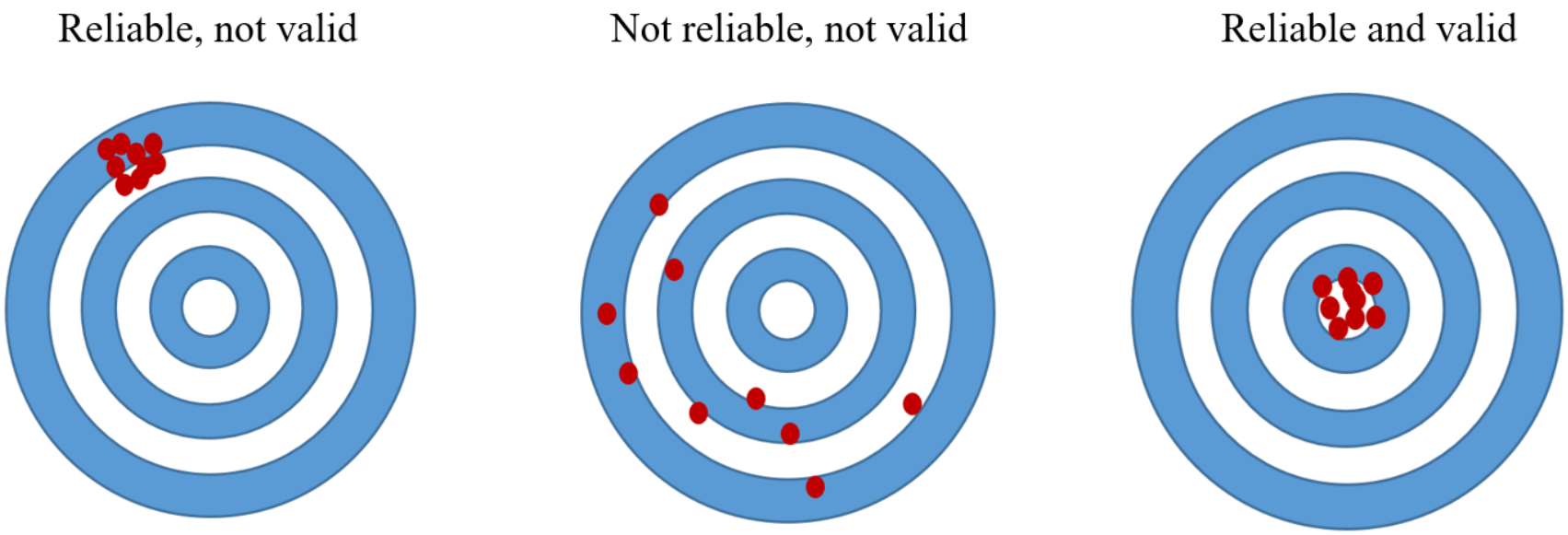
→ Less likely to get unlucky!

Non-random sampling

- **Convenience sampling:** Easy, inexpensive access
- **Quota sampling:** Search until target group count is met
- **Respondent-driven:** Ask respondents for contacts

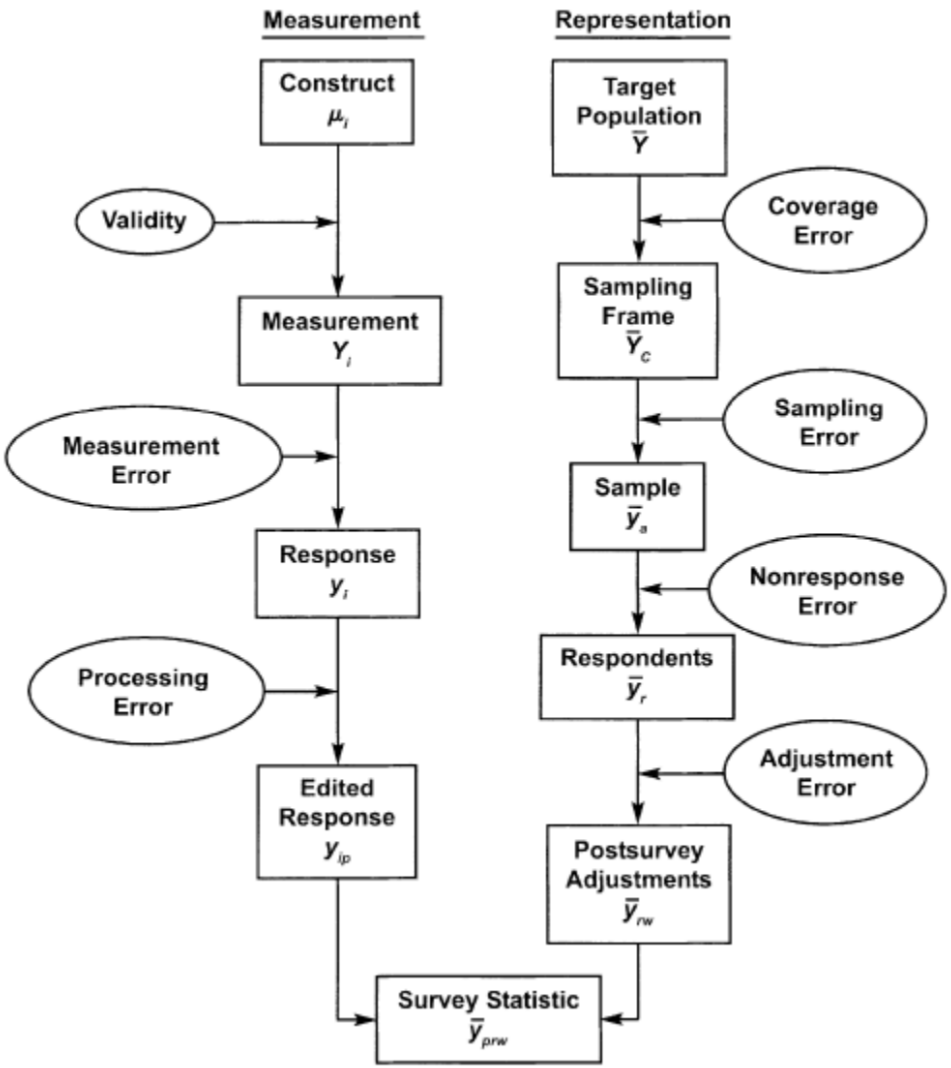
Sources of error

Return to measurement error



Many different forms of measurement error!

Total survey error framework



Summary

- We conduct surveys to understand attitudes and behaviors
- Gap between *self-reports* and *actual* attitudes behaviors opens the door for measurement error
- Many sources of error (Mode, frame, procedure)
- *Random sampling* important assumption to enable statistical inference

Surveys

POLI SCI 210

Introduction to Empirical Methods in Political Science

Last time

- Gap between *self-reports* and *actual* attitudes/behaviors opens the door for measurement error
- **Tuesday:** Sample design concerns
- **Today:** Questionnaire design concerns

1. Asking good questions
2. Asking sensitive questions

Practice

Table 1
Reported Daily TV Consumption as a Function of Response Alternatives

Low-frequency alternatives	Daily consumption	High-frequency alternatives	Daily consumption
Up to ½ hour	7.4%	Up to 2½ hours	62.5%
½ hour to 1 hour	17.7%	2½ hours to 3 hours	23.4%
1 hour to 1½ hours	26.5%	3 hours to 3½ hours	7.8%
1½ hours to 2 hours	14.7%	3½ hours to 4 hours	4.7%
2 hours to 2½ hours	17.7%	4 hours to 4½ hours	1.6%
More than 2½ hours	16.2%	More than 4½ hours	0.0%

Note. $N = 132$. From "Response Categories: Effects on Behavioral Reports and Comparative Judgments," by N. Schwarz, H. J. Hippler, B. Deutsch, & F. Strack, 1985, *Public Opinion Quarterly, 49,* p. 391. Copyright 1985 by The University of Chicago Press. Adapted with permission.

Source: Schwarz, Norbert. 1999. "Self-reports: How the questions shape the answers." *American Psychologist* 54 (2): 97

More practice

Q: *What is the most important thing to teach children to prepare them for life?*

A: **To think for themselves**

- 61.5% multiple choice
- 4.6% open-ended

What makes a good survey question?

No definitive answer!

- We want some validity and reliability, but those are hard to gauge
- Consistency over time is good

Here are some vague principles

Literal vs. pragmatic meaning

What the words mean vs. what we think the question is asking

Follow the *maxims of conversation*:

1. **Maxim of relation:** Answers are relevant *relative* to context
2. **Maxim of quantity:** Answer are informative, but no more informative than required
3. **Maxim of manner:** Most obvious meaning should be the correct one
4. **Maxim of quality:** Not to say anything false or inaccurate

Try to approximate everyday language!

Open-ended vs. closed response

Q: *What have you done today?*

What to say if **open-ended**:

- I went to class
- I took a shower
- I read the newspaper
- I wrote to my representative

Open-ended vs. closed response

Q: *What have you done today?*

What to say if **multiple choice**:

- I went to class
- I took a shower
- I read the newspaper
- I wrote to my representative

Frequency scales

Q: *How frequently have you felt really irritated?*

What does *really irritated* mean?

Maybe look at the options:

- Less than once a year
- About twice per year
- Every other month
- About once a month
- More than once a month

Reference periods

Q: *How many times have you felt really irritated **this year***

vs.

Q: *How many times have you felt really irritated **last week?***

Reference periods also convey meaning!

Rating scales

Q: *How successful would you say you have been in life?*

Scale 1

0 (Not at all successful) - 10 (extremely successful)

Scale 2

-5 (Not at all successful) - 5 (extremely succesful)

Rating scales

Q: *How successful would you say you have been in life?*

Scale 1 (13% between 0 and 5)

0 (Not at all successful) - 10 (extremely successful)

Scale 2 (34% between -5 and 0)

-5 (Not at all successful) - 5 (extremely succesful)

- 0 at the beginning is a *minimum*
- 0 in the middle is a *baseline*

Demand effects

Preamble 1

You will now be asked to consider some hypothetical (not real) online news items and to indicate which news item you would most prefer to read

Adapted from Mummolo and Peterson (2019)

Demand effects

Preamble 2

You will now be asked to consider some hypothetical (not real) online news items and to indicate which news item you would most prefer to read. ***The purpose of this exercise is so we can measure whether the news outlet offering an article influences how likely people are to read the article***

Adapted from Mummolo and Peterson (2019)

Demand effects

Preamble 3

You will now be asked to consider some hypothetical (not real) online news items and to indicate which news item you would most prefer to read. ***The purpose of this exercise is so we can measure whether the news outlet offering an article influences how likely people are to read the article.*** *We expect that people will be more likely to choose an article if the news source offering it is known to favor their preferred political party, since people tend to seek out information that is consistent with their personal views*

Adapted from Mummolo and Peterson (2019)

Demand effects

- Respondents' tendency to try use *context* clues to find the *"right" answer*

Small bits convey **researcher's epistemic interest**

- Affiliation
- Description of study
- Question-wording

Priming effects

QUESTION-ORDER EFFECTS IN SURVEYS: THE CASE OF POLITICAL INTEREST, NEWS ATTENTION, AND KNOWLEDGE

By Dominic L. Lasorsa

Subjects were exposed to one of three survey versions differing only in question order. Those who first faced difficult political knowledge questions reported significantly lower levels of both political interest and news attention than those who did not first face the knowledge test. However, when the knowledge questions and the interest and attention questions were separated by a "buffer" item that could serve as an excuse for poor knowledge, self-assessed interest and attention were less depressed. Characteristics of survey questions that may make them particularly susceptible to these types of question-order context effects are discussed and strategies for dealing with such effects are noted.

Lasorsa, Dominic L. 2003. "Question-Order Effects in Surveys: The Case of Political Interest, News Attention, and Knowledge." *Journalism & Mass Communication Quarterly* 80 (3): 499-512

Priming effects

Three question groups

1. Political interest (How often follow govt news)
2. Political knowledge (Texas politics)
3. News coverage evaluation (Excuse knowledge gaps)

Shuffle order, and compare frequency for each combination

Priming effects

TABLE 3

Crosstabulation of Political Interest by Question Order

	Question Order			
Political Interest	Interest, Knowledge, Excuse	Knowledge, Interest, Excuse	Knowledge, Excuse, Interest	Total
High	206 (174)	136 (161)	128 (135)	470
Low	89 (121)	136 (111)	100 (93)	325
Total	295	272	228	795

Chi-Square = 24.75, df = 2, $p < .001$

Note: Cell entries are observed frequencies. Numbers in parentheses are expected frequencies. "Low" political interest consists of those who reported following government and public affairs "hardly at all" and "only now and then." "High" political interest consists of those who said they did so "some of the time" and "most of the time."

Priming effects

Questions are affected by their **placement in the survey flow**

Preceding questions may influence how respondents interpret current question

What if…

The problem is *not* how to craft questions

But the questions *themselves*?

Some examples

- Have you lied about having COVID symptoms?
- Would you bribe a police officer to avoid a traffic ticket?
- Have you had sex after drinking alcohol?
- Have you been offered goods or favors for your vote?
- Do you know anyone with links to a militant organization?
- Would you oppose a black family moving next door?
- Would you allow muslim immigrants to become citizens?

What do these have in common?

- They are **sensitive questions**
- We can only learn about them through surveys
- Asking about them directly leads to **misreporting**
- This kind of *measurement error is called* **social desirability bias** or **sensitivity bias**

How to reduce sensitivity bias?

- Purposefully deviate from *maxims of conversation*
- By adding noise to the question
- Two approaches:

1. Distract from the sensitive attitude/behavior
2. Guarantee anonymity

- Different techniques vary on how they combine the two
- Broadly known as **indirect questioning techniques**

Randomized response

For this question, I want you to answer *yes* or *no*.

Randomized response

For this question, I want you to answer *yes* or *no*. But I want you to consider the number of your dice throw.

Randomized response

For this question, I want you to answer *yes* or *no*. But I want you to consider the number of your dice throw. If ⚀ shows on the dice, tell me *no*.

Randomized response

For this question, I want you to answer *yes* or *no*. But I want you to consider the number of your dice throw. If ⚀ shows on the dice, tell me *no*. If ⚅ shows, tell me *yes*.

Randomized response

For this question, I want you to answer *yes* or *no*. But I want you to consider the number of your dice throw. If ⚀ shows on the dice, tell me *no*. If ⚅ shows, tell me *yes*. But if another number shows, tell me your own opinion about the question.

[TURN AWAY FROM RESPONDENT]

Now you throw the dice so that I cannot see what comes out.

Randomized response

For this question, I want you to answer *yes* or *no*. But I want you to consider the number of your dice throw. If ⚀ shows on the dice, tell me *no*. If ⚅ shows, tell me *yes*. But if another number shows, tell me your own opinion about the question.

[TURN AWAY FROM RESPONDENT]

Have you thrown the dice?

Randomized response

For this question, I want you to answer *yes* or *no*. But I want you to consider the number of your dice throw. If  shows on the dice, tell me *no*. If  shows, tell me *yes*. But if another number shows, tell me your own opinion about the question.

[TURN AWAY FROM RESPONDENT]

Have you picked it up?

Randomized response

For this question, I want you to answer *yes* or *no*. But I want you to consider the number of your dice throw. If ⚀ shows on the dice, tell me *no*. If ⚅ shows, tell me *yes*. But if another number shows, tell me your own opinion about the question.

Now, during the height of the conflict in 2007 and 2008 *(in Afghanistan)*, did you know any militants, like a family member, a friend, or someone you talked to on a regular basis?

Please, before you answer, take note of the number you rolled on the dice.

How does it work?

- We know that about $1/6 \approx 0.17$ respondents said *yes* because they rolled a 🎲
- So if 30% in total said *yes*
- We have $\widehat{Y} = 0.3 - 0.17 = 0.13$ as our *population estimate*
- But we **do not know** who they are in our survey!

Assumptions

> ⚠ **1. Honesty given protection**
>
> People respond honestly when guaranteed anonymity.

> ⚠ **2. One-sided lying**
>
> Those who **do not** hold the sensitive trait **never** falsely claim to bear it.

- These cannot be verified with data!

List experiment

Now I am going to read you three things that make people angry or upset.

List experiment

Now I am going to read you three things that make people angry or upset. After I read all three, just tell me HOW MANY of them upset you.

List experiment

Now I am going to read you three things that make people angry or upset. After I read all three, just tell me HOW MANY of them upset you. I don't want to know which ones, just HOW MANY.

Control group

1. The federal government increasing the tax on gasoline
2. Professional athletes getting million-dollar contracts
3. Large corporations polluting the environment

List experiment

Now I am going to read you three things that make people angry or upset. After I read all three, just tell me HOW MANY of them upset you. I don't want to know which ones, just HOW MANY.

Treatment group

1. The federal government increasing the tax on gasoline
2. Professional athletes getting million-dollar contracts
3. Large corporations polluting the environment
4. **A black family moving next door**

How does it work?

- Respondents are randomly assigned to conditions
- Differences in responses can only be attributed to the presence/absence of the sensitive item (more on this in next week)
- So $\widehat{Y} = \text{Mean}(\text{treatment}) - \text{Mean}(\text{control})$ gives a *population prevalence rate*
- But we **do not know** who they are in our survey!

Assumptions

> ⚠ **1. No liars**
>
> Those who **do not** hold the sensitive item **never** falsely claim to bear it.

> ⚠ **2. No design effects**
>
> Including the sensitive item **does not change** how participants respond to the baseline items

- These can only be evaluated **indirectly**

Other techniques

- Survey administration protections
- Wisdom of the crowds
- Endorsement experiments
- Conjoint experiments
- Network scale-up method
- Variants of the randomized response and list experiment

Summary

- Asking good survey questions is challenging!
- Ideally, we want to follow the *maxims of conversation*
- But often, getting precise answers means asking *weird* questions
- **Example:** Techniques for sensitive questions