POLI210: Political Science Research Methods

Lecture 8.2: Survey research I

Olivier Bergeron-Boutin

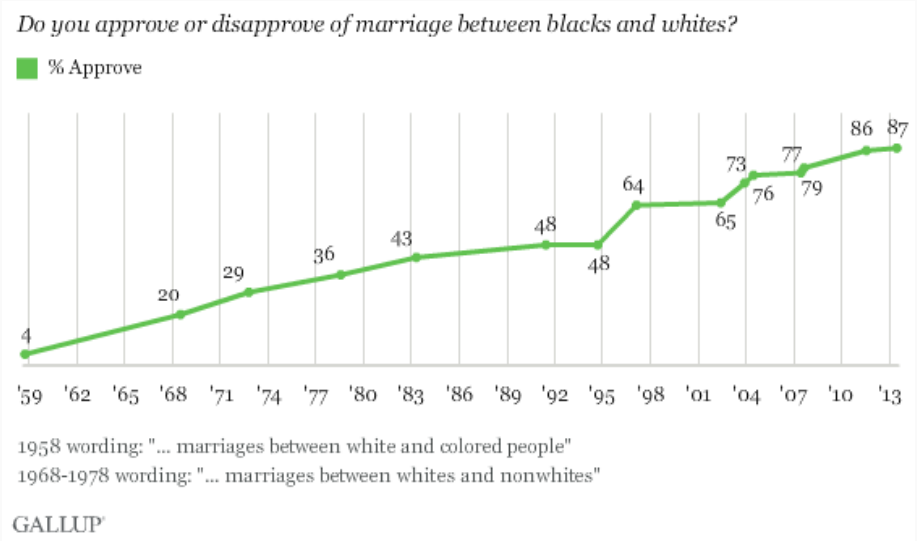
October 21st, 2021

Boring admin stuff

- Assignment 3: due next Monday
- Office hours after class
- Quiz 1 next week
- Midcourse survey
- Halloween fundraiser

What are surveys useful for?

- Gaging public opinion
  - Berinsky (2017): "in effect polls have become public opinion."
- Feedback on programs and policies
- Understanding the evolution of society



- Evaluating relationships between different attitudes/experiences

A short history of polling

First issue of Public Opinion Quarterly:

- "an essay...explaining how George Gallup's quota-controlled survey of a few thousand triumphed over the Literary Digest's straw poll of millions in correctly predicting the election outcome [of 1936]." Hillygus (2011)
- "Sampling bias": the people who are included in the survey are systematically different than the people who are not
    - Same as "selection bias": you can't solve it by increasing sample size!

The rise of survey research in political science:

- "Between 1950 and 1970, the percentage of articles in the APSR based on surveys went from 0 to almost 50 percent." (Adcock and Bevir 2010, 81)

Some basic concepts

- Population: All possible units of interest.
  - If I want to predict the election outcome, the relevant “population” is all Canadian citizens aged 18 or over
  - If I want to gage your interest in the class, the “population” is all 281 of you registered in the course
- Sample: A subset of the population that I contact
  - Key question: is it representative?
- Parameter: A number that describes a characteristic of our population
  - e.g. the mean height in this class
- Statistic: A number that describes some aspect of our sample
  - e.g. mean height of students in the sample

We use the data from our sample to make **inferences** about population parameters

The (often unattainable) gold standard

Ideally, our sample is the same as the population!

- This makes things easy: every unit in the population is in the sample, and no unit that is not in the population is in the sample
- Here, sample statistics = population parameters
- An example: the census, “big data”
- BUT resource-intensive, particularly in large populations
    - e.g. barely 30 years ago, if I wanted to survey this class, each additional respondent would require an interviewer

Instead, construct a **sampling frame** and randomly sample from it

- Sampling frame: Enumeration (listing) of all units eligible for sample selection
- e.g. list of 281 students and randomly select 20

Probability sampling: the gold standard in sampling

Probability sampling: every unit in the population has non-zero probability of being selected into sample

- If we have a perfect sampling frame...
- Get to randomly choose who to interview...
- And everyone contacted agrees to be interviewed...
- Our sample is representative (ON EXPECTATION)
- On expectation?
    - If I draw 1,000 samples, on average, the samples will be representative
    - But some samples will be a bit older than the true population parameter
    - Some younger, some more male...
    - **Sampling variance**

But, you guessed it...this is unlikely!

Sampling frame: the classlist

```
survey <- read.csv("assignments/pset1/class_survey.csv")
head(survey)
```

```
##   X gender ideology
## 1 1 Female        2
## 2 2   Male        2
## 3 3 Female        2
## 4 4 Female        1
## 5 5   Male        4
## 6 6 Female        1
```

```
# sampling one random number from 1 to 10
sample(1:10, size = 1)
```

```
## [1] 6
```

```
# sampling 10 random numbers from 1 to 100 (w/ replacement)
sample(1:100, size = 10, replace = TRUE)
```

```
##  [1] 82 43 47  5 82 69 57 42 44 48
```

Sampling frame: the classlist

```
order <- sample(1:nrow(survey), size = nrow(survey),
                replace = FALSE)
order[1:5]
```

```
## [1]   7  50 157  74  88
```

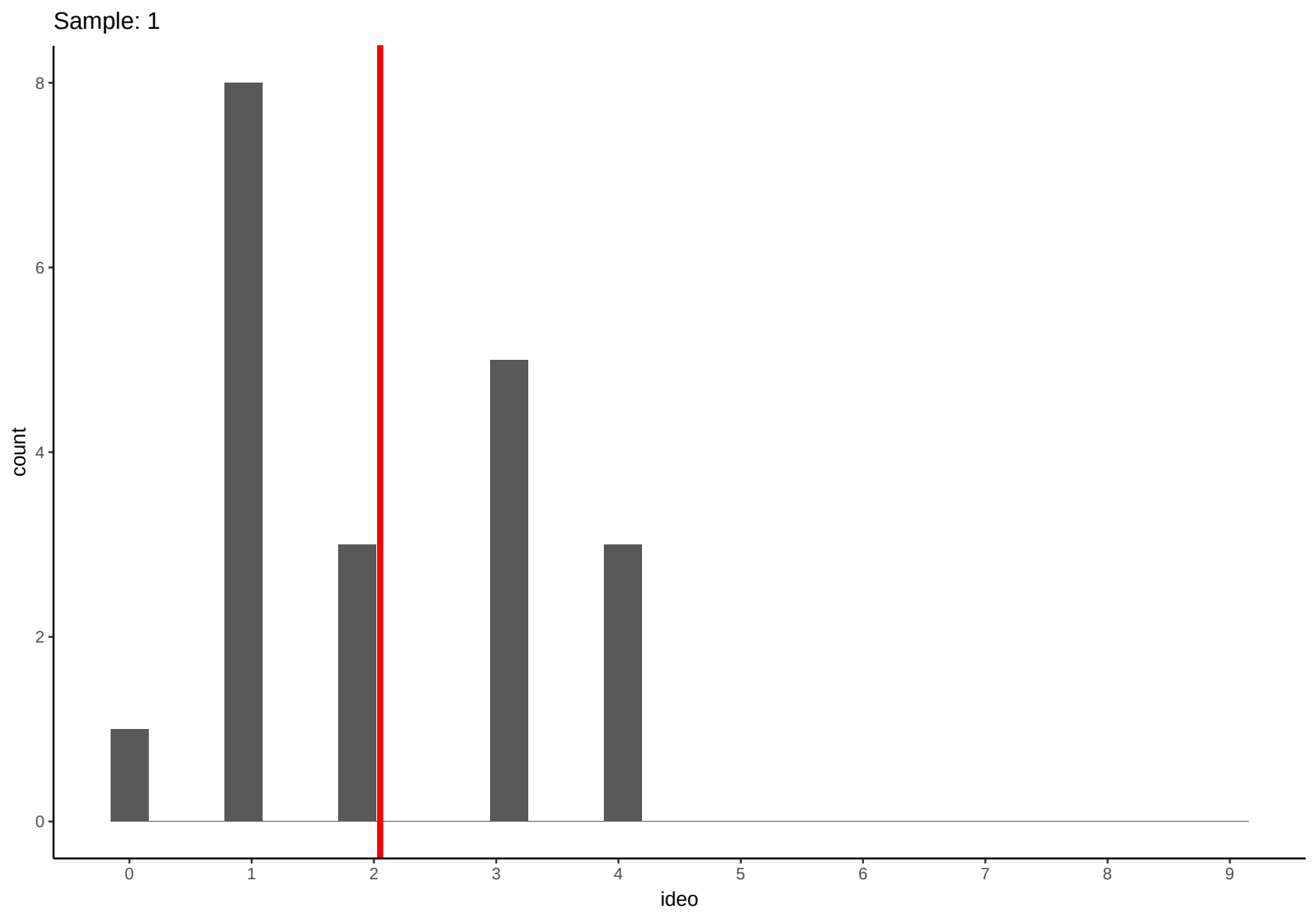
```
sample_n5 <- survey[order[1:5],]
sample_n5
```

```
##       X gender ideology
## 7     7 Female        1
## 50   50   Male        4
## 157 157   Male        1
## 74   74 Female        3
## 88   88 Female        2
```

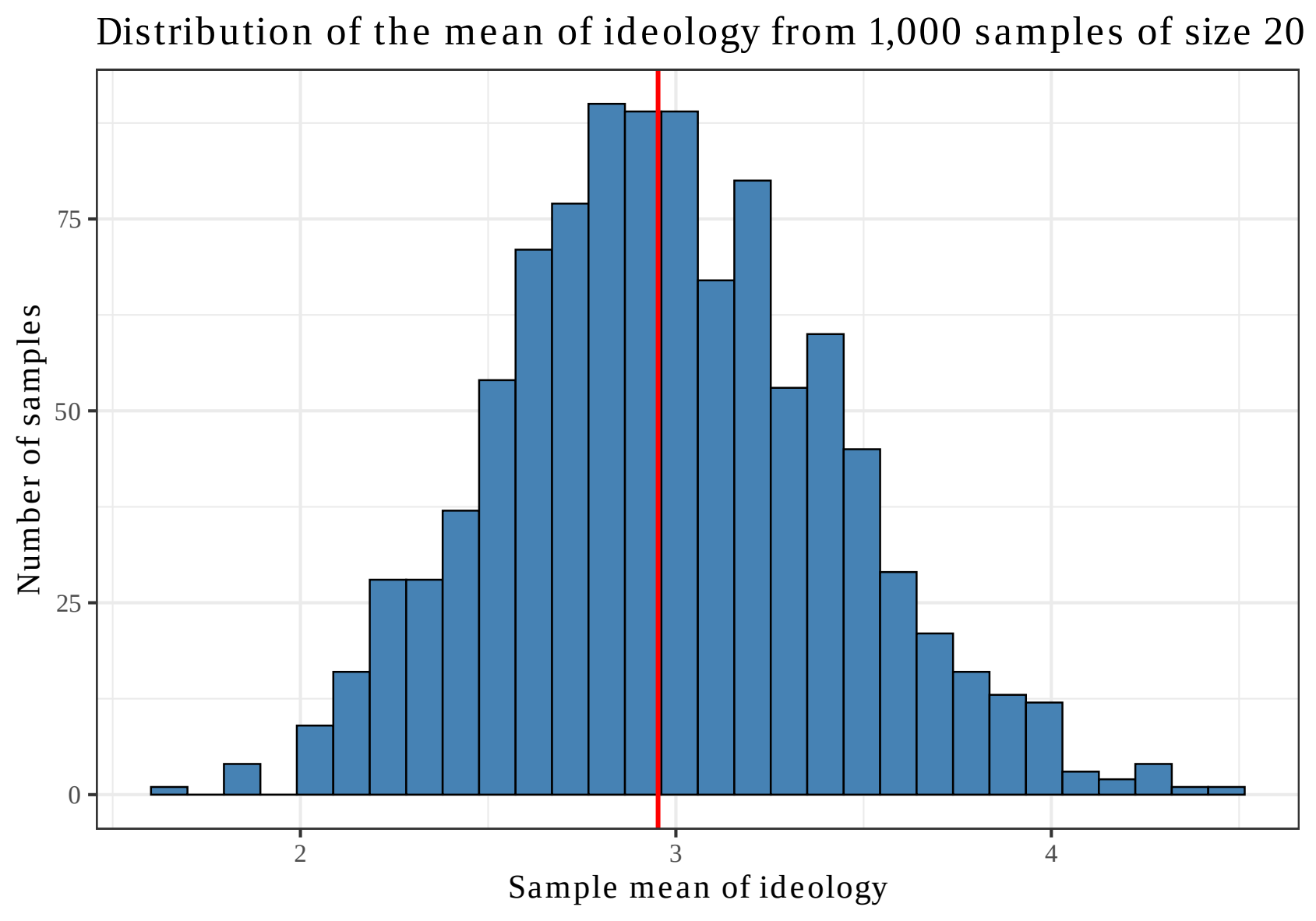
```
c(mean(sample_n5$ideology, na.rm = T), mean(survey$ideology, na.rm = T))
```

```
## [1] 2.200000 2.952555
```

Let’s draw a few samples...



Central limit theorem from class survey

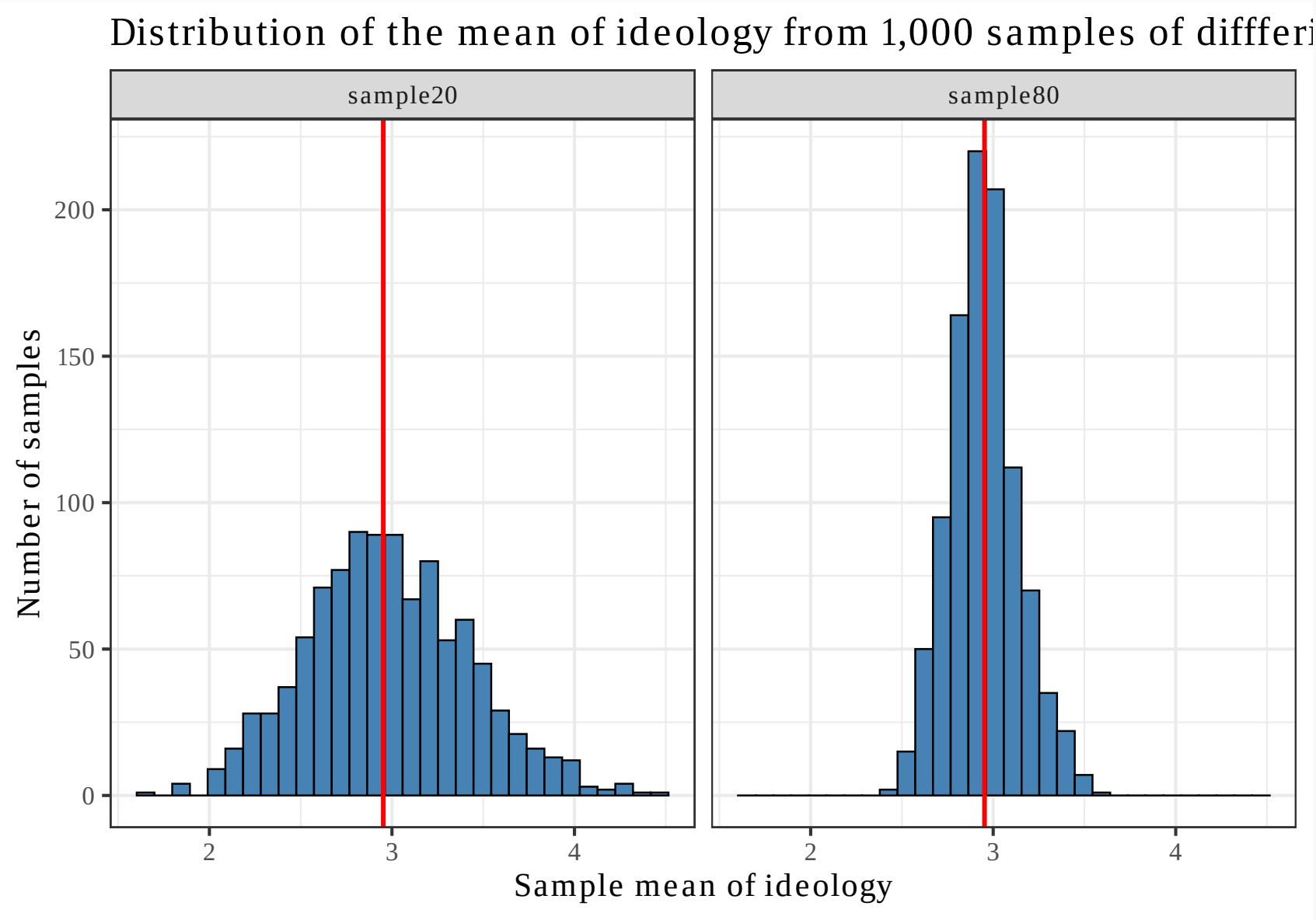


Central limit theorem

As the number of samples increases, the distribution of sample means (the “sampling distribution”) approaches a normal distribution

- Normal distribution: the “bell curve”
- No matter the underlying distribution!
  - The distribution of ideology is non-normal
  - And yet, by taking repeated samples, the distribution of samples means approximates a normal distribution centered at the true population mean
  - And, the larger the samples I take, the “tighter” the normal distribution is around the true population parameter!

CLT from class survey



Probability sampling: obstacles

Sounds good, right? Remember what probability sampling demands:

- A non-zero probability that **all** units in population are selected for sampling
- i.e. I could in theory reach everyone
- How likely is that? Depends on context
- In context of national election, very unlikely!

One problem: the sampling frame is imperfect ("frame bias")

- This class: what about people who are auditing the course?
- They don't show up in my classlist
    - Therefore: Pr(sampled) = 0
- What if they have systematically different attitudes toward the course?
    - e.g. less stressed about grades and enjoy more

Probability sampling: obstacles

Even assuming my sampling frame is perfect...

- Not everyone agrees to answer the survey once contacted
  - "Unit non-response"
  - (In contrast with "item non-response": the unit answers the survey but skips an item)
- The people who decline to answer are different $\rightsquigarrow$ non-response bias
- What sorts of people agree to answer surveys?
  - High social trust, some leisure time, some interest in public affairs...
  - If the factors that influence non-response are associated with whatever outcome we are seeking to measure (vote choice, policy support...)
  - We will have a biased estimate of the population parameter

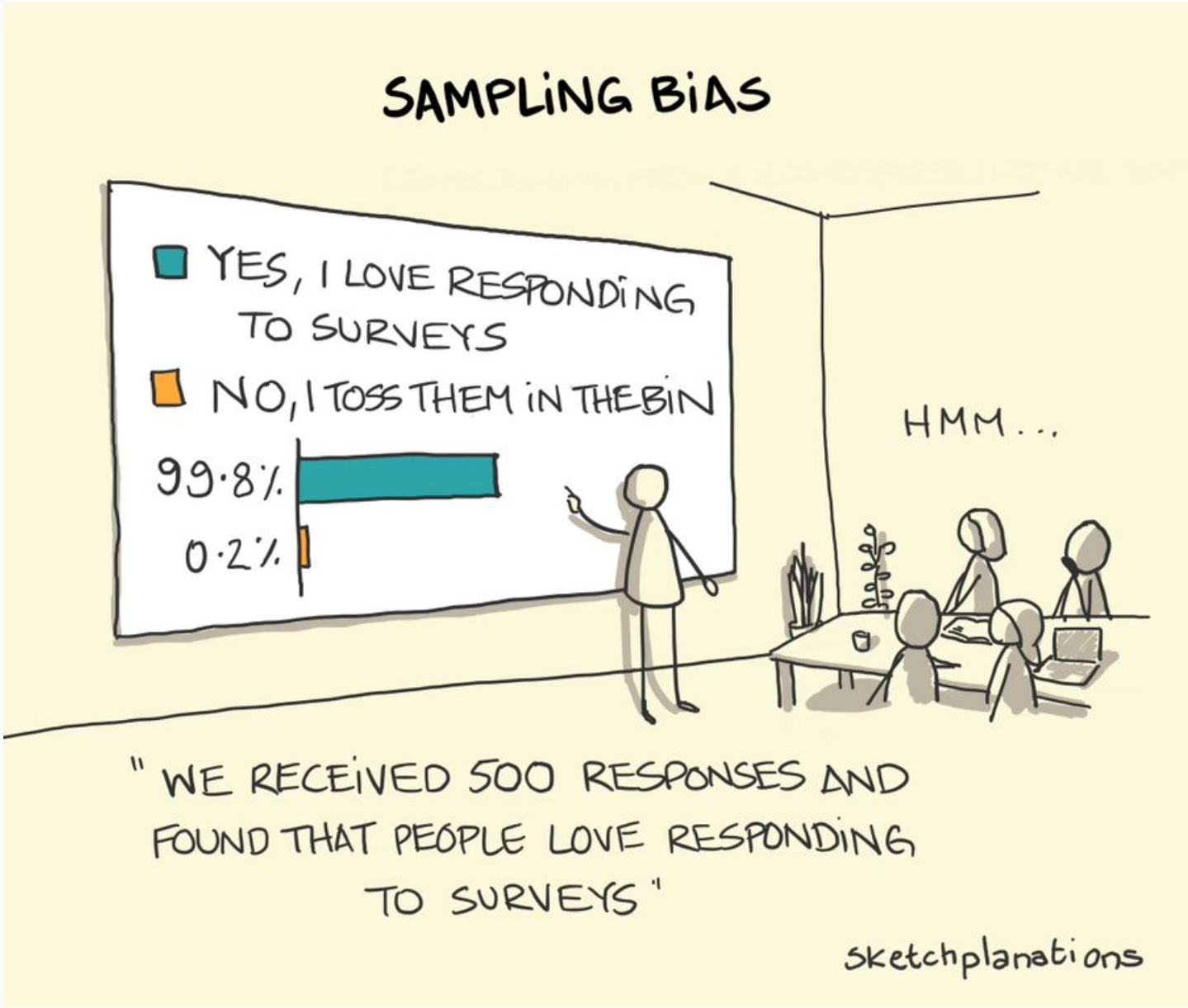
So what about that Reader's Digest poll?

Sample size of the Reader's Digest 1936 poll: 2.4 million!

- Procedure: sampling frame of 10 million (!!)
  - Constructed from telephone records, magazine subscription lists
- First, sampling bias:
  - Who is likely to end up in the sampling frame?
  - The people in the sampling frame are different
  - Sampling bias
- Second, non-response bias
  - 24% of those contacted answered the survey
  - Are they systematically different? Probably!

Survey	Roosevelt %	Sample size
Gallup	56	~50,000
Reader's Digest	43	~2.4 million
Actual election outcome	62	

Sampling bias



No comment...


Donald J. Trump
@realDonaldTrump
THANK YOU!
BREITBART IMPEACHMENT POLL
Do you stand with President Trump?
Yes
97.83%
No
2.17%
Like 16K
Share
Tweet
Return To Poll

Methods of data collection

Face-to-faces interviews

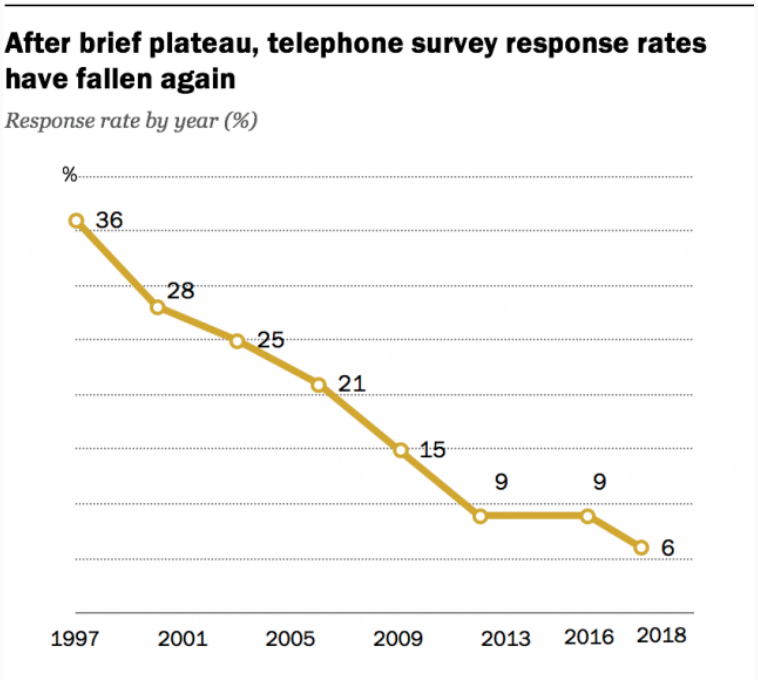
Used to be extremely common

- Advantages:
  - Higher response rates than other methods
- Disadvantages
  - High costs: travel, accommodation, salaries...
  - “Interviewer effects”: the identity of the interviewer may influence the respondent’s answers
  - Davis and Silver (2003): Black respondents perform worse on a 7-item political knowledge battery when interviewed by a white person

	Black respondent	White respondent
Black interviewer	3.42	4.00
White interviewer	2.80	3.87

Phone interviews

- Rapid completion; lower cost than face-to-face
- Sampling frame: all phone numbers
  - Random-digit-dialing: every phone number in an area (e.g. Canada) has an equal chance of being sampled
  - But who does this "sampling frame" include/exclude?
    - Excludes people without phones, e.g. the poor
- Major, recent-ish problem: declining response rates



Online surveys

- Very cheap to administer
- No interviewer/lesser risk of bias
- Flexible format: audio, video, images...
- Sampling frame?
    - There is none: there is no list of Internet users!
    - Non-probability sample
    - Survey companies specialize in maintaining “panels” of respondents
    - As a result: issues with representativeness
        - Can partially correct using statistical methods:
        - Quotas
        - Weighting
    - Also: professional survey takers!

References i

Adcock, Robert, and Mark Bevir. 2010. “Political Science.” In The History of the Social Sciences Since 1945, edited by Philippe Fontaine and Roger E. Backhouse, 71–101. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511845260.006.

Berinsky, Adam J. 2017. “Measuring Public Opinion with Surveys.” Annual Review of Political Science 20: 309–29.

Davis, Darren W., and Brian D. Silver. 2003. “Stereotype Threat and Race of Interviewer Effects in a Survey on Political Knowledge.” American Journal of Political Science 47 (1): 33–45. https://doi.org/10.1111/1540-5907.00003.

Hillygus, D. Sunshine. 2011. “The Evolution of Election Polling in the United States.” Public Opinion Quarterly 75 (5): 962–81. https://doi.org/10.1093/poq/nfr054.