Political Science 210: Introduction to Empirical Methods

Week 7: Surveys

Studying populations

To answer some research questions, we need to know the attitudes and behavior of a given population.

- “Population” can be everyone living in a geographic area (city, state, country) or the people in an area who share a common trait (race, party affiliation, registered voters)

To know the attitudes and behavior of a population, we can ask each and every member of the population…

…then we’re done!

But most of the time, we don’t have enough time or money to do this.

The purpose of (most) surveys

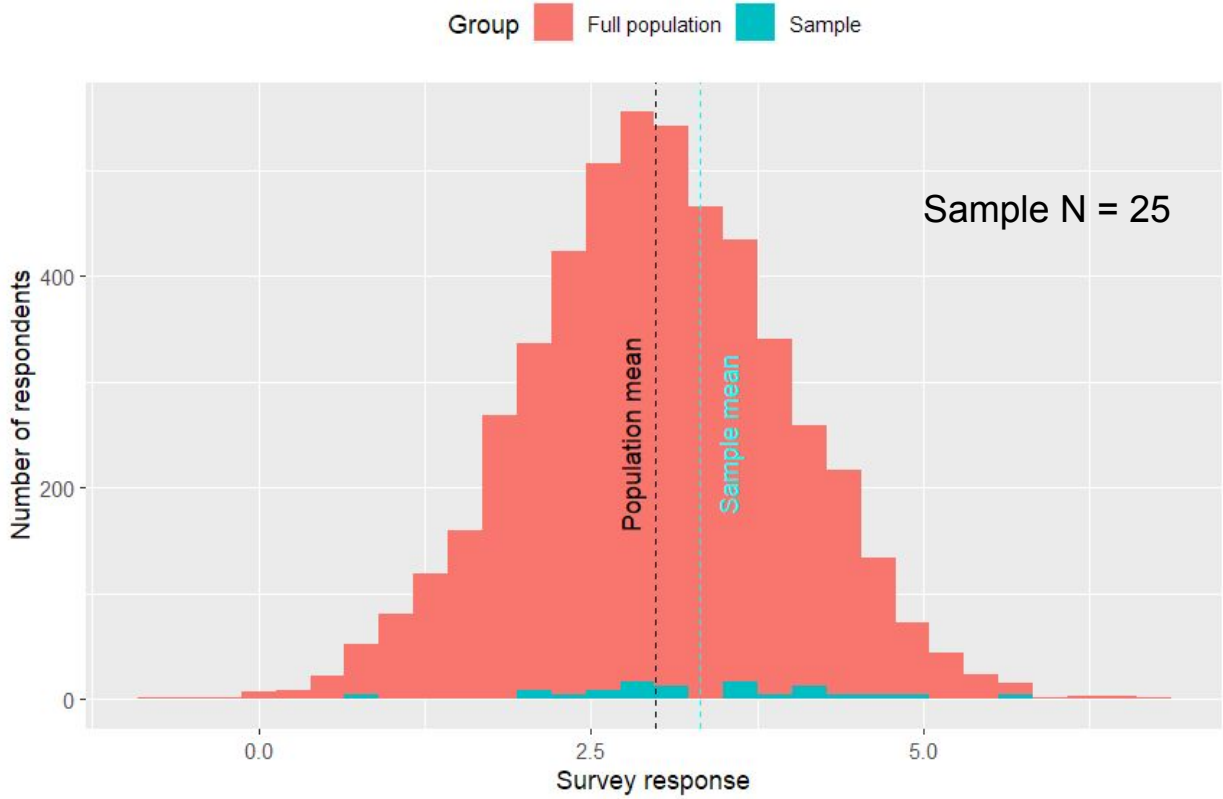
When conduct a survey, we want to select a *sample* of the population, then use that sample to make inferences about the *full* population.

But how do we know if we have a good sample?

What a good sample needs

The size of a sample should be large enough that the characteristics of the sample are representative of the characteristics of the full population.

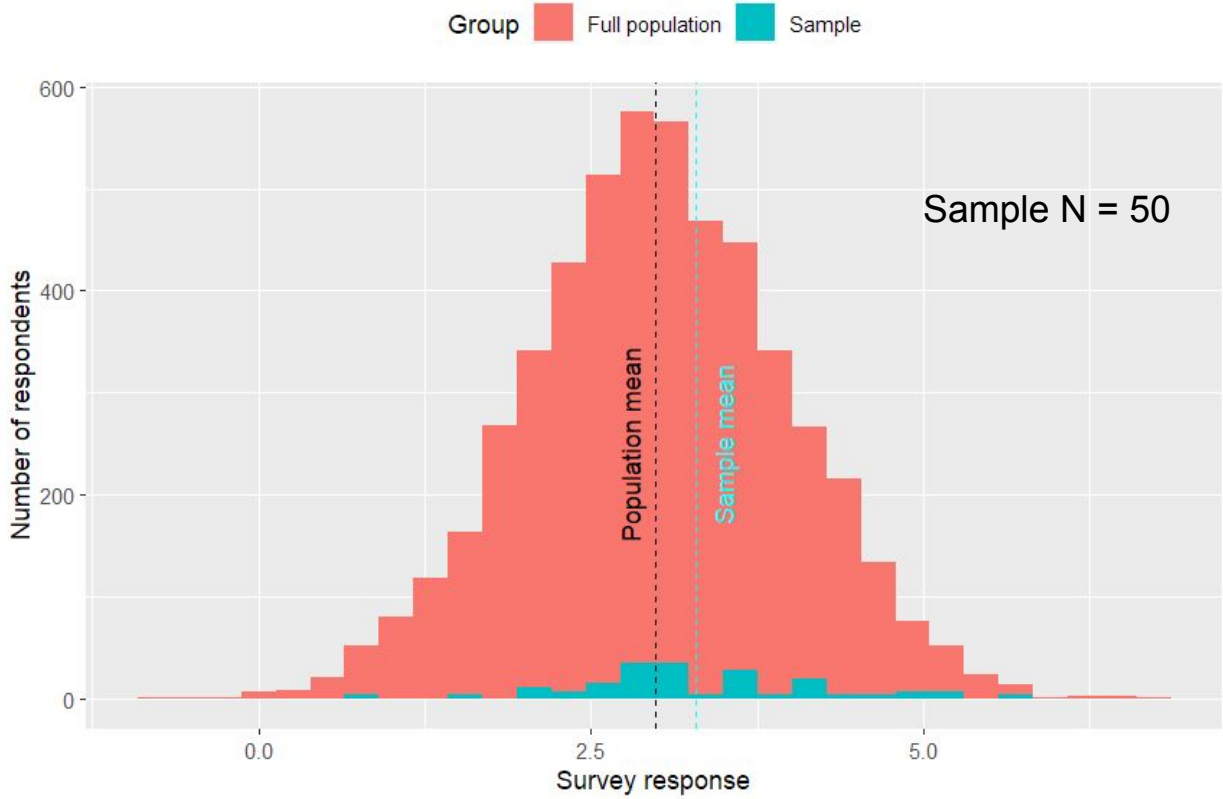
- *Law of large numbers*: The more cases that we sample from a population, the closer the sample average will be to the population average



What a good sample needs

The size of a sample should be large enough that the characteristics of the sample are representative of the characteristics of the full population.

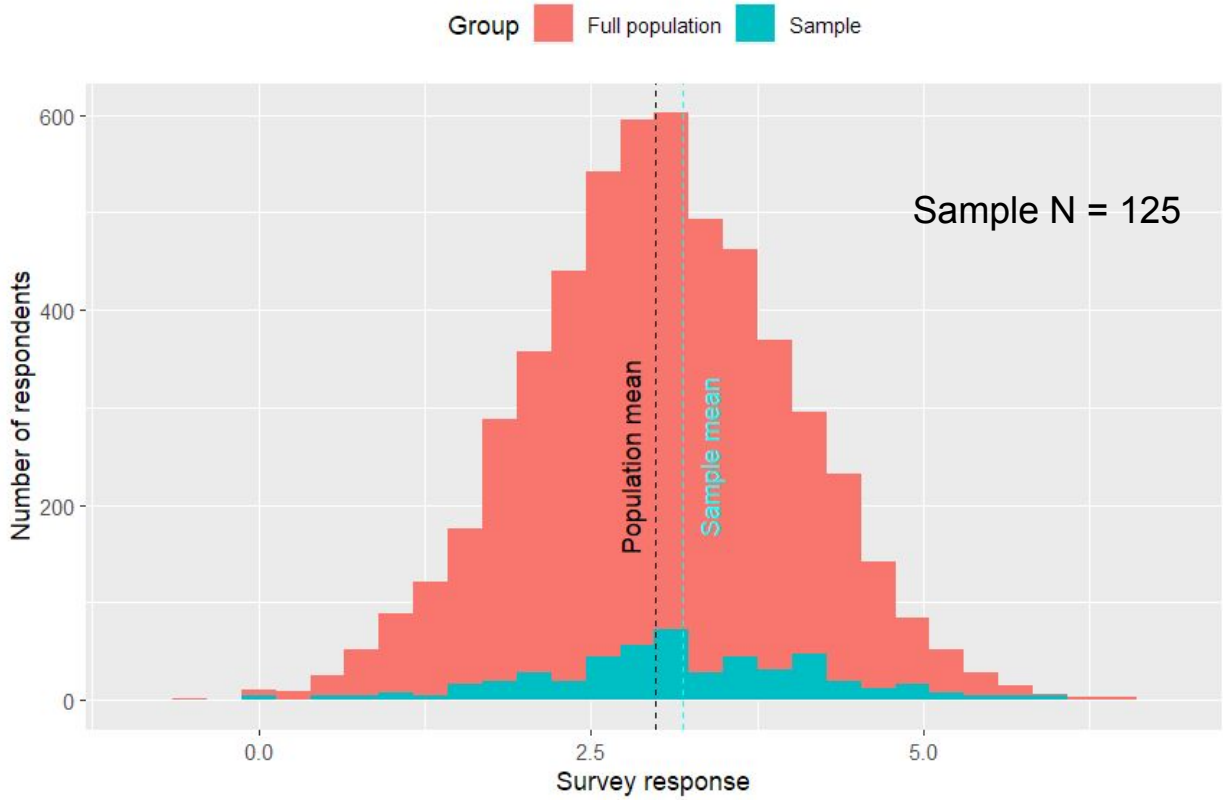
- *Law of large numbers*: The more cases that we sample from a population, the closer the sample average will be to the population average



What a good sample needs

The size of a sample should be large enough that the characteristics of the sample are representative of the characteristics of the full population.

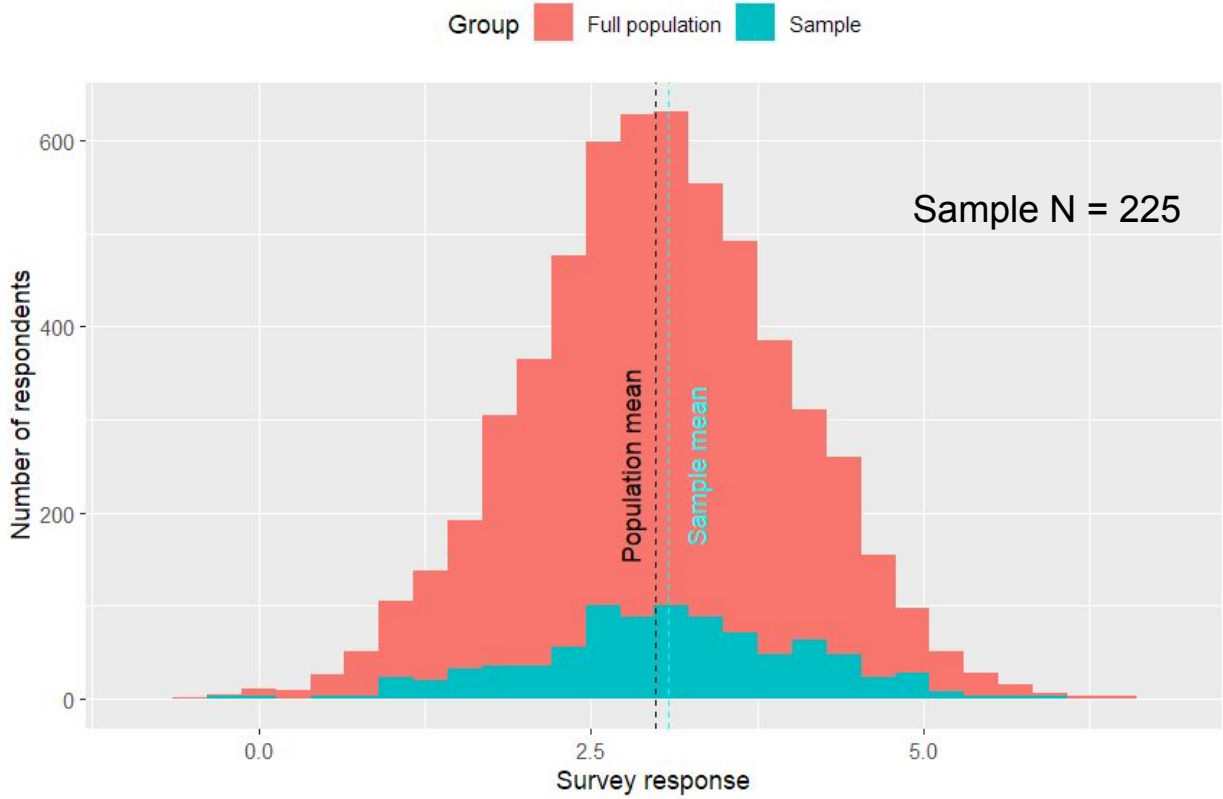
- *Law of large numbers*: The more cases that we sample from a population, the closer the sample average will be to the population average



What a good sample needs

The size of a sample should be large enough that the characteristics of the sample are representative of the characteristics of the full population.

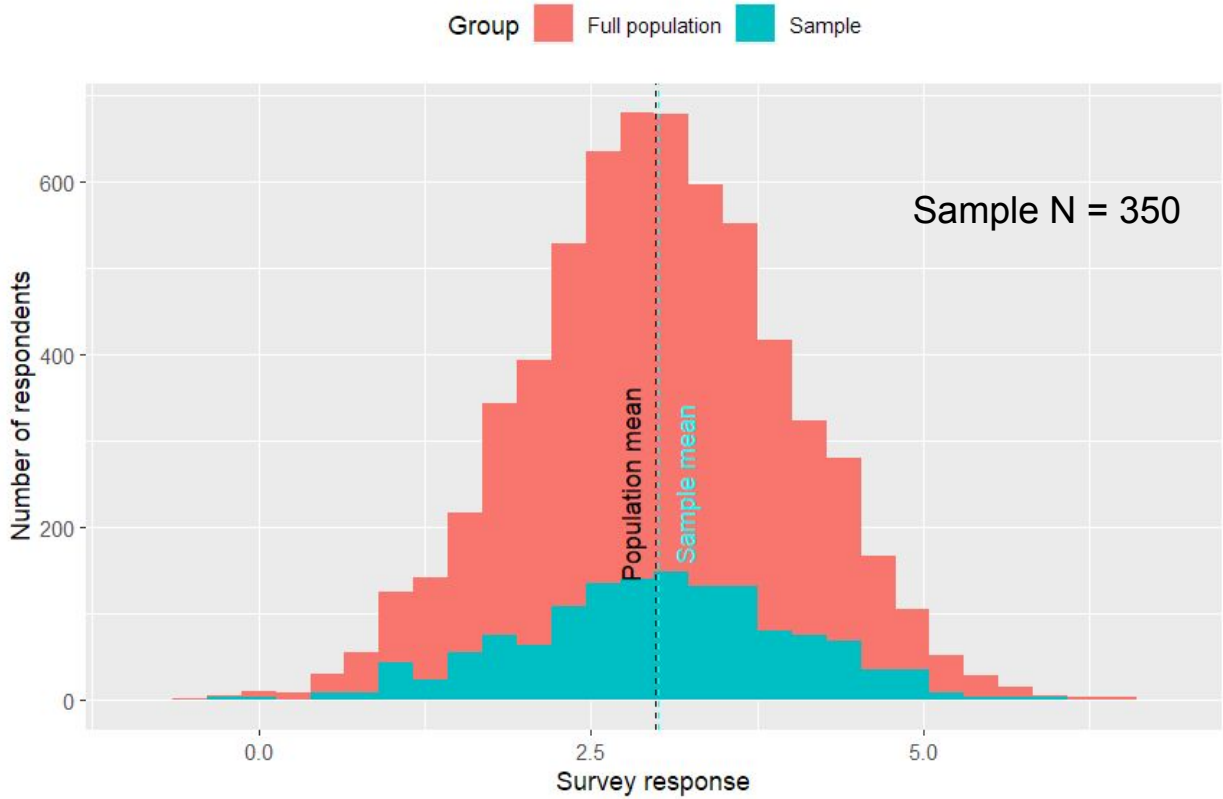
- *Law of large numbers*: The more cases that we sample from a population, the closer the sample average will be to the population average



What a good sample needs

The size of a sample should be large enough that the characteristics of the sample are representative of the characteristics of the full population.

- *Law of large numbers*: The more cases that we sample from a population, the closer the sample average will be to the population average



What a good sample needs

The sample should also be an *unbiased* measure of the population, or else sampling error from our measurement will skew our estimate of the full population.

- Coverage bias: Is the list that we're using to draw our sample (the "sampling frame") representative of the full population?
  - E.g. if we sample randomly from a telephone address book, we'll miss people who don't have landline phones.
- Selection bias: Certain segments of the population more likely to be selected for the sample than others
- Non-response bias: Certain elements of the population more likely to not respond
- Social desirability bias: Are people unwilling to answer questions truthfully if they think it will portray them in a negative light?

Sampling strategies

Non-probability sampling: Surveying anyone available without regard to their likelihood of being chosen from the overall population.

- Convenience sample: Selecting cases that are easily available
  - Necessary if short on resources or if we want to conduct a “pilot” study
  - May not create bias if the population we want to survey is narrow and specific to our geographic location
  - But usually not representative of a more general population

Sampling strategies

Probability sampling: Ensures that every individual's chance of being chosen for the survey is proportional to their share of the total population (based on age, gender, race, income, etc.).

- Random sampling: Draw a random sample from the population as a whole.
- Stratified sampling: Population is divided into subgroups ("strata") based on shared characteristics; individuals from each strata are sampled randomly
  - E.g. if we think that older people are more likely to vote for an incumbent than younger people, we divide the population by age, then sample randomly from each age cohort.
- Cluster sampling: Population is composed of subgroups ("clusters") that vary in size and composition; a random sample of clusters is taken.
  - E.g. we might divide a city into districts, then survey individuals from a random sampling of each district.

Stratified sampling

Population is divided into subgroups (“strata”) based on shared characteristics; individuals from each strata are sampled randomly.

- Example: If we want to make inferences about men and about women separately…

Stratified sampling

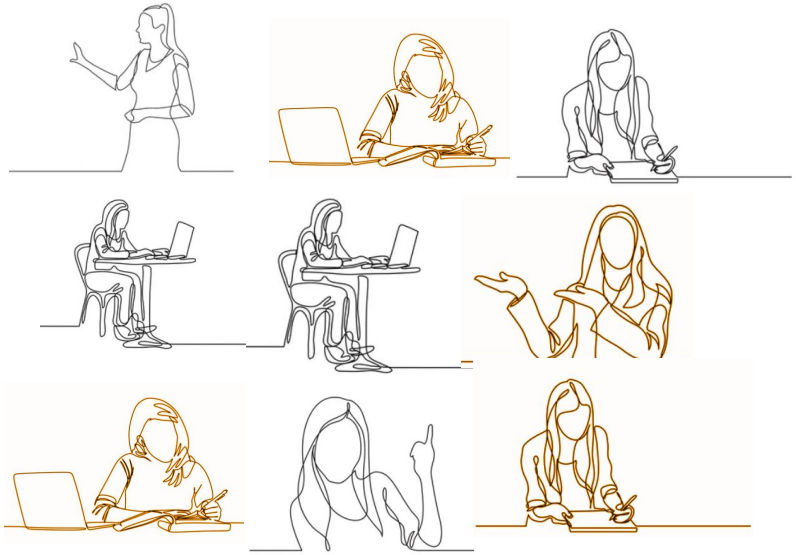
Population is divided into subgroups (“strata”) based on shared characteristics; individuals from each strata are sampled randomly.

- Example: If we want to make inferences about men and about women separately…
- …we divide the population into two different subgroups…

Stratified sampling

Population is divided into subgroups (“strata”) based on shared characteristics; individuals from each strata are sampled randomly.

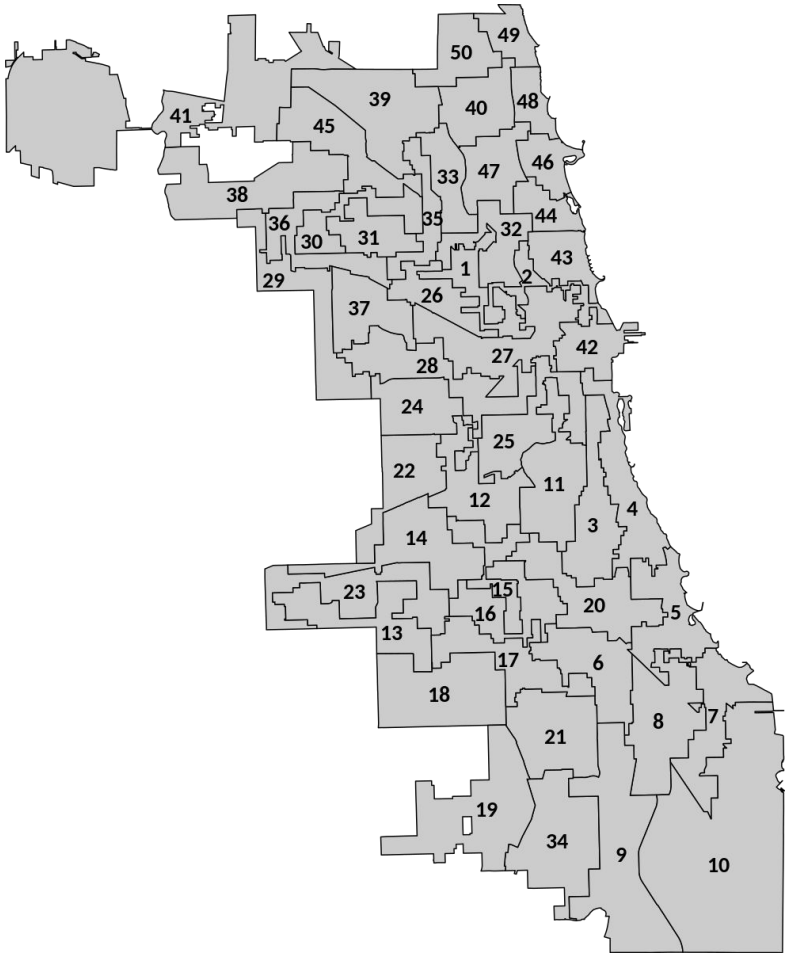
- Example: If we want to make inferences about men and about women separately…
- …we divide the population into two different subgroups…
- …then sample randomly from each subgroup.



Cluster sampling

Population is composed of subgroups (“clusters”) that vary in size and composition; a random sample of clusters is taken.

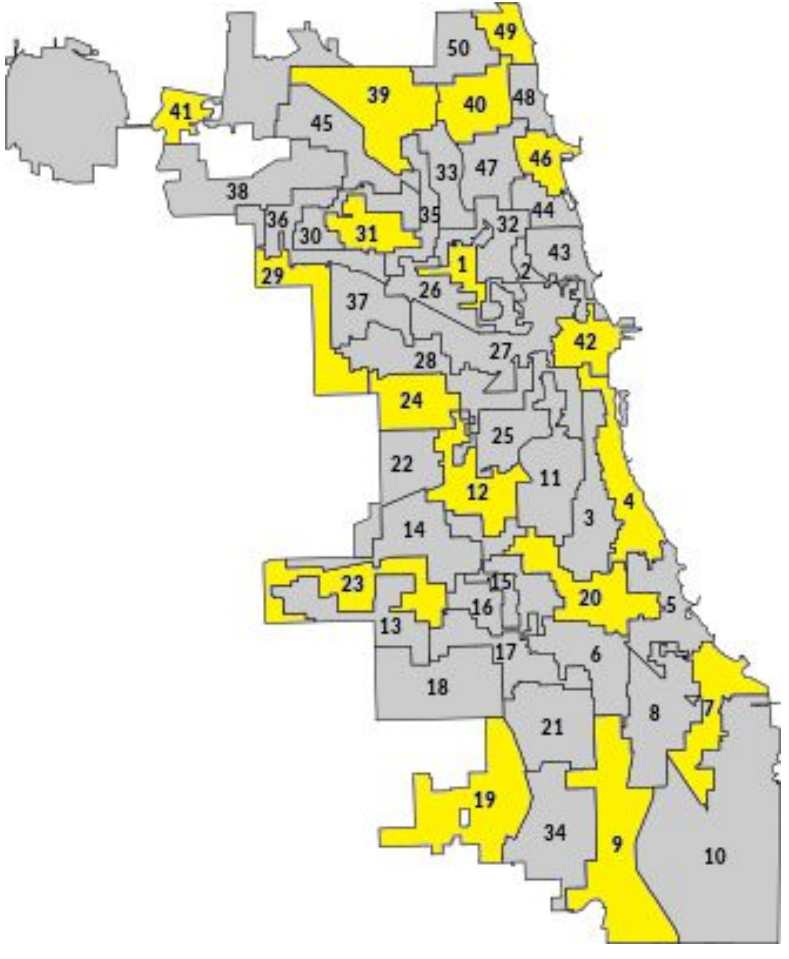
- Example: If we wanted to take a poll of voters in a large city, but couldn’t reach the entire city…



Cluster sampling

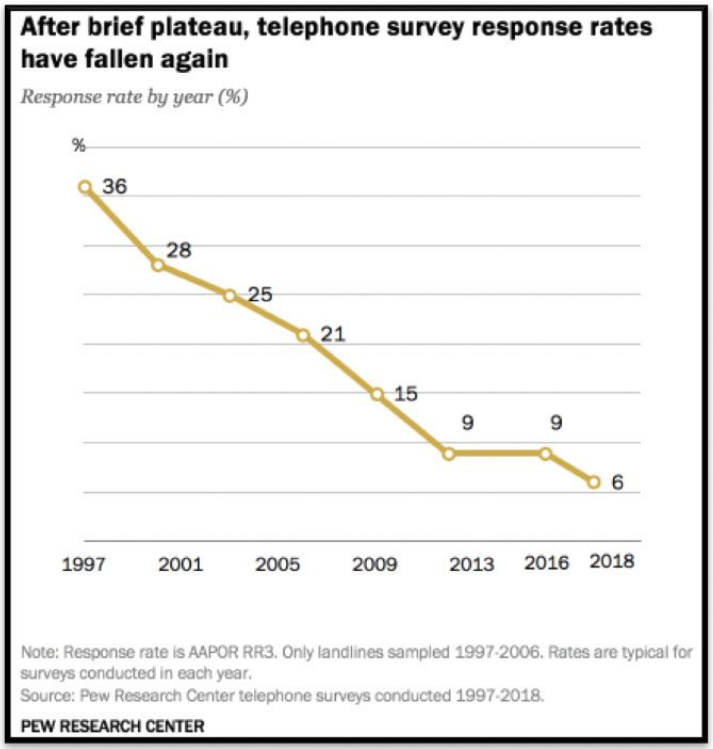
Cluster sampling: Population is composed of subgroups (“clusters”) that vary in size and composition; a random sample of clusters is taken.

- Example: If we wanted to take a poll of voters in a large city, but couldn’t reach the entire city…
- …we could divide the city into precincts, then choose a random sample of precincts to poll.
- We can also combine this method with stratified sampling.



The rapid decline of response rates

- Survey response rates have fallen dramatically in the last ~20 years
- We can still try to get estimates of the full population by weighting samples based on age, urban/rural location, and income level, but are we missing other important factors?
- 2016 US presidential election:
    - Polls predicting a Clinton win were way off in key areas of the country
    - After the election, many pointed to nonresponse as a key issue.
    - Why?
    - People who voted for Trump were ALSO less likely to respond to polls, and couldn't be identified solely by age, race, rural location, etc.



How *not* to think about surveys

What's wrong with this exchange?

- Coverage bias: Elon's followers could be a skewed sample of total Twitter users, no matter how large his follower count is.
- Response bias: Even if we assume that those who aren't among Elon's followers "tend to hear about" the polls he sends, are they as likely to respond to the polls in the same way that they would if they followed him?
- Is the wording of the poll still clear, free of leading questions, avoids potential social desirability bias?
- "Statistically significant" just means our certainty that the sample reflects the full population. There is NO point at which the sample size overcomes the potential for bias.
    - If the size of the sample is the same as the size of the population (that is, if everyone on Twitter followed Elon, then we wouldn't need statistics!).

For each of the scenarios described below, ask (1) whether the researchers used a particular sampling strategy (convenience, random, stratified, cluster) and (2) whether the strategy contains any potential for bias or mismeasurement.

To find out if students at a large private midwestern university would approve of annual fees to fund a public transportation pass, members of the student government stand in a high-traffic area of campus and ask passers-by their opinion of the proposed measure.

To find out which demographics would be most interested in their client's new television news talk show, a marketing firm buys ads on the websites of major TV news channels. Based on their browsing history, visitors to these sites are classified by their likely age, gender, and income level. Then a random selection of visitors from each subcategory are asked to complete a survey about what they enjoy most about their favorite TV news shows.

How active are legislators in responding to their constituents? To find out, researchers send each current member of the US House and Senate a survey with a list of questions about how often they listen to and engage with their constituents on a weekly basis.

A researcher wants to survey the policy preferences of the wealthiest 1% of Americans. To find out who those Americans are, the researcher obtains information from a private marketing firm that sells lists of such households to businesses that cater to ultra-wealthy clients. The researcher then randomly selects three US metropolitan areas and asks the listed households in those areas their opinion on a variety of measures to reduce poverty.

For each of the scenarios described below, ask (1) whether the researchers used a particular sampling strategy (convenience, random, stratified, cluster) and (2) whether the strategy contains any potential for bias or mismeasurement.

To find out if students at a large midwestern private university would approve of annual fees to fund a public transportation pass, members of the student government stand in a high-traffic area of campus and ask passers-by their opinion of the proposed measure.

For each of the scenarios described below, ask (1) whether the researchers used a particular sampling strategy (convenience, random, stratified, cluster) and (2) whether the strategy contains any potential for bias or mismeasurement.

To find out which demographics would be most interested in their client's new television news talk show, a marketing firm buys ads on the websites of major TV news channels. Based on their browsing history, visitors to these sites are classified by their likely age, gender, and income level. Then a random selection of visitors from each subcategory are asked to complete a survey about what they enjoy most about their favorite TV news shows.

For each of the scenarios described below, ask (1) whether the researchers used a particular sampling strategy (convenience, random, stratified, cluster) and (2) whether the strategy contains any potential for bias or mismeasurement.

How active are legislators in responding to their constituents? To find out, researchers send each current member of the US House and Senate a survey with a list of questions about how often they listen to and engage with their constituents on a weekly basis.

For each of the scenarios described below, ask (1) whether the researchers used a particular sampling strategy (convenience, random, stratified, cluster) and (2) whether the strategy contains any potential for bias or mismeasurement.

A researcher wants to survey the policy preferences of the wealthiest 1% of Americans. To find out who those Americans are, the researcher obtains information from a private marketing firm that sells lists of such households to businesses that cater to ultra-wealthy clients. The researcher then randomly selects three US metropolitan areas and asks the listed households in those areas their opinion on a variety of measures to reduce poverty.

Let's say you work for a politician located in the Chicagoland area who wants to study public support for a policy that would provide free child care to people working in Cook County, paid for by a 0.5% property tax increase.

- What is your overall population of interest? What are the different types of people and regions of the county that you might want to include in your survey?
- Let's say you had a budget of $5,000, a team of 30 staff members, and a week to complete the survey. What would your sampling strategy be? How would you compile the list from which to sample? How would you physically conduct the survey?
- What questions would you include in the survey and how would you word them? How would you word the questions? Are there ways you could avoid the potential for biases created by

Either hand your notes in to me at the end of section or email them to jack.garigliano@u.northwestern.edu. Make sure to include the names of everyone in your group!