

Sampling and Bias - Why take samples and how *not* to take samples

Sampling

- Voting polls - which candidate is in the lead
- Public opinion - Gallup polls
- Quality assurance methods for manufacturing - limited number of items checked for quality/defect

Why Sample?

- Voting polls -
- Public opinion -
- Quality assurance methods for manufacturing -

Why Sample?

- Voting polls - could you survey everyone?
- Public opinion - Gallup polls - can everyone be surveyed?
- Quality assurance methods for manufacturing - can all items be examined for quality/defect?

Population

- A population is the set of people or things that you want to know/learn about or examine/study
 - The population is the *set*.

Population

- The individual elements of the population are called the units
 - The number of units is the *population size*.
 - A sample is the set of units that you study.
 - A census is a special case of data collected on an entire population.

Problem for survey data: Bias

- A biased *opinion* is not objective
- A sampling method is *biased* if it produces samples for which some characteristic of a population is *underrepresented* or *overrepresented*.

What makes a sample good?

- A sample is good if it is representative of the population
 - A *small* version of the population
 - *Proportions* determined *from* the *sample* are close to corresponding *proportions* for the entire *population*

What makes a sample good?

- A sample is good if it is representative of the population
 - Sample statistics are close to those for the entire population
 - ◉ Means
 - ◉ Standard deviations
 - ◉ Medians
 - ◉ IQR's

Caution

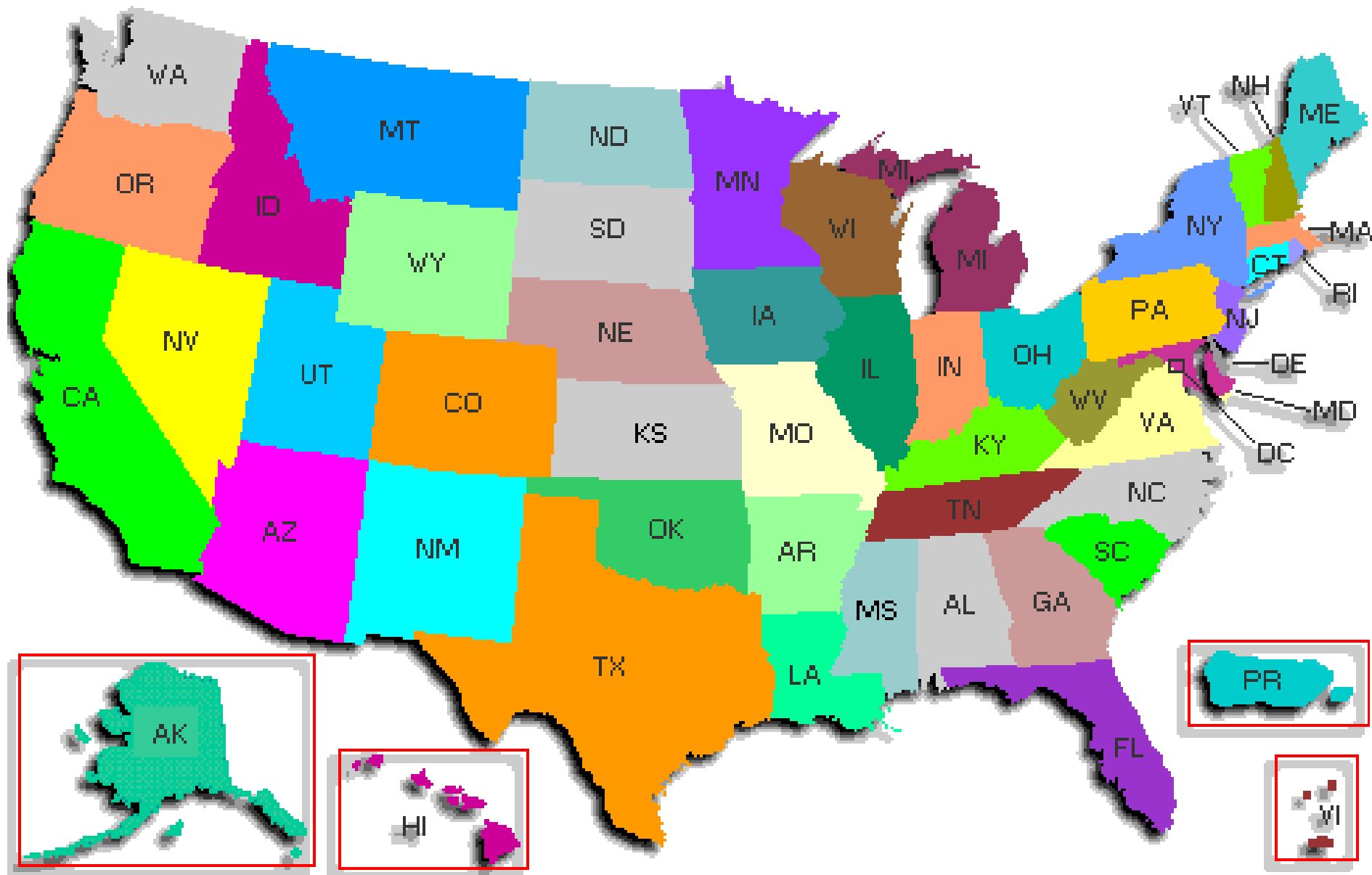
- It is difficult to tell if a sample is representative of a population.
- **HOWEVER**, *we can determine if the method used to determine the units in the sample is biased.*

Sampling Bias

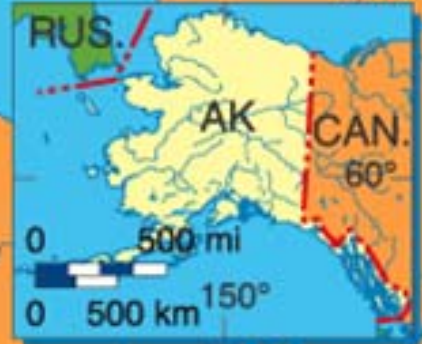
- Sampling bias is also known as *sample selection bias*.
- Sampling bias is present in a sampling procedure if samples result in numerical summaries that are systematically *too high* or *too low*.

Size bias

- Selection by dropping object on paper
 - Larger regions on paper more likely to be selected
- Examples
 - Selecting a country
 - Selecting a region of the state to survey/search
 - Selecting a region to survey based on region population



N United States





Scale 1:100,000,000
 Robinson Projection
 standard parallels 30° N and 30° S

June 1998

Country boundaries as
 of January 1998

© 1998 WOODGATE INC.

Voluntary Response Bias

- Television/Radio call-in programs
 - Those who *care* about an issue or those who *want* to call in are represented but those who do not care about the issue, those who do not *want* to spend the time, and those *not watching/listening* are not represented

Convenience Sample

- A sample for which the units that are chosen are *convenient/easy* to include
 - Examples
 - ◉ Survey of one's friends
 - ◉ Survey at a mall

Judgment Samples

- Samples determined based on the *judgment of an expert*
 - Problems
 - ◉ Expert may be biased
 - ◉ Expert may ignore certain characteristics of the population

Judgment Samples

- 1948 presidential election polls were wrong
 - Poll takers used quotas for sample selection
 - ◉ So many men, women, over 40, under 40, ...
- Quota sampling is biased

What about Exit Polls?

- 2004 presidential election Exit Polls were wrong
 - Poll information indicated that John Kerry would win

What about Exit Polls?

- 2004 presidential election Exit Polls were wrong
 - Poll information indicated that John Kerry would win
- What are some possible problems with Exit Polls?

Unbiased Sampling Methods

- All units in the population must have a known chance of being selected/sampled
- How do this?

Unbiased Sampling Methods

- All units in the population must have a known chance of being selected/sampled
- Create a sampling frame *before* the sample is created

Unbiased Sampling Methods

- All units in the population must have a known chance of being selected/sampled
- Create a sampling **frame** *before* the sample is created

Unbiased Sampling Methods

- All units in the population must have a known chance of being selected/sampled
- Create a sampling frame *before* the sample is created
- *Sampling Frame* or *frame* - a "list" of population units

Unbiased Sampling Methods

- Creating the sampling frame cannot be taken for granted
- If you start with a bad frame then you generate a bad sample
 - Non-representative list for the units in the population will lead to a sample containing units that are not representative of the population that it is suppose to represent

Unbiased Sampling Methods

- What should be considered when creating a frame for the following populations?
 - Students of FSC
 - People living in the United States
 - Internet users
 - Population for a health care survey

Unbiased Sampling Methods

- The units in the sample can represent the frame well but may not be representative of the actual population
- Examples
 - Telephone survey
 - Internet survey
 - In general

Unbiased Sampling Methods

- The units in the sample can represent the frame well but may not be representative of the actual population
- Examples
 - Telephone survey - must have phone
 - Internet survey - must have internet access
 - In general - must be *willing* to participate in survey

Surveys

- What bias could be in the surveys of the following groups were used to represent all FSC students?
 - Students surveyed on 12/1/04 - a survey for percentage of students studying at least two hours the previous night
 - Students *American Writers I/II* - the percentage of mathematics majors

Surveys

- What bias could be in the surveys of the following groups were used to represent all FSC students?
 - Students surveyed on 03/31/05 - a survey for percentage of students studying at least two hours the previous night

Response Bias

- 40% of all people contacted for surveys *refuse* to respond
- Those who agree to participate in the survey can be different from those who refuse to participate in the survey

Response Bias

- Good example from text
 - Study that found that left-handed people die, on average, nine years before right-handed people
 - Problems
 - ▣ Surveys sent to everyone listed on death certificates in two LA counties to inquire about the handedness of the person who had died
 - ▣ Only half of the questionnaires were returned

Questionnaire Bias

- How you *ask* the question affects the response
 - Tone of voice of the interviewer
 - Demeanor of the interviewer
 - Wording of questions
 - Vocabulary
 - Negative/positive

Questionnaire Bias

- Reader's Digest survey
 - I.* *I would be disappointed if Congress cut its funding for public television.*

Questionnaire Bias

- Reader's Digest survey

II. *Cuts in funding for public television are justified as part of an overall effort to reduce federal spending.*

Questionnaire Bias

- Reader's Digest survey
 - I. I would be disappointed if Congress cut its funding for public television.*
 - II. Cuts in funding for public television are justified as part of an overall effort to reduce federal spending.*

Questionnaire Bias

- Reader's Digest Survey
 - Agreeing with statement *I* is almost the same as disagreeing with statement *II*.

Questionnaire Bias

- Reader's Digest Survey results
 - Statement *I*:
 - ◉ 54% agreed
 - ◉ 40% disagreed
 - ◉ 6% did not know
 - Statement *II*:
 - ◉ 52% agreed
 - ◉ 37% disagreed
 - ◉ 10% did not know

Other Bias

- Some folks who participate in a survey
 - Do not tell the truth
 - ◉ Intentional/unintentional lie
 - Give a response that sounds good
 - Give a response that they feel is desired
 - Do not remember

Other Bias

- **Measurement bias**
 - Patients in a study often over state or over estimate how well they are following the doctor's orders
 - Measurements taken with a faulty measuring device
 - Ruler with worn end
 - Measurements not taken at the *actual beginning*

Other Bias

- **Intentional/unintentional lie or desired response**
 - **Weight**
 - **Grade point average**
 - **Television viewing**
 - **Favored candidate**
 - **income**