# Sampling and Bias

# Sampling and Bias

## Why take samples and how *not* to take samples

# Sampling and Bias

## Why take samples and how *not* to take samples

- When is sampling used?
- Why is sampling used?
- What makes a sample good?
- What problems arise in sampling?
- What is bias?

# Examples of Use of Sampling

- Voting polls

- Public opinion –
- Quality assurance methods for manufacturing –

- Research –

# Examples of Use of 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
- Research – monitoring virus spread; developing treatments, medications, vaccines for diseases or viruses

# 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
  - Get to know,
  - Learn about or
  - Examine, or
  - Study.

# Population

- The individual elements of the population are called the <u>units</u> ...

# Population

- The individual elements of the population are called the <u>units</u> or
  - Experimental units,
  - Records,
  - Cases,
  - Subjects,
  - Participants,
  - Respondents.

# Population

- The individual elements of the population are called the <u>units</u>
  - *A census is a special case of data collected on an entire population ...*

# Population

- The individual elements of the population are called the <u>units</u>
  - *A census is a special case of data collected on an entire population but not all members of the population are guaranteed to be included in the census ...*

# Population

- The individual elements of the population are called the <u>units</u>
  - *A census is a special case of data collected on an entire population but not all members of the population are guaranteed to be included in the census as collection of census data depends on member response and census takers.*

# Population

- The individual elements of the population are called the <u>units</u>

# Population

- The individual elements of the population are called the <u>units</u>

  - A sample is the set of units that you study.

# Population

- The individual elements of the population are called the <u>units</u>
  - The number of units is the *population size.*
  - *A* sample is the set of units that you study.

# Population

- The individual elements of the population are called the <u>units</u>
  - The number of units is the *population size.*
  - *A* sample is the set of units that you study.
    - The number of units is the *sample size.*

# Population

- The individual elements of the population are called the <u>units</u>
  - The *number of members of the population* is the *population size*.
  - A sample is the set of units that you study.
    - The *number of members of the sample* is the *sample size*.

# Sampling Used

- To collect data from/about the group to which you have access

# Sampling Used

- To collect data from/about the
  - Subjects of a study

# Sampling Used

- To collect data from/about the
  - Subjects of a study
- May involve a survey

# Problem for (Survey) data?

# Problem for (Survey) data? Bias

# Problem for (Survey) data: Bias

- *A* biased *opinion* is not objective

# Problem for (Survey) data: Bias

- *A* biased *opinion* is not objective
  - Limited view for a study
  - Not considering all possible
    - Aspects
    - Groups
  - Conscious or unconscious

# 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

# 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 represented using a disproportionately
  - Smaller number
  - Larger number

# What Makes a Sample Good?

- A sample is good if it is representative of the population

# What Makes a Sample Good?

- A sample is good if it is representative of the population
  - A *small* version of the population

# 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
  - Statistics are close in value to parameters

# What Makes a Sample Good?

- A sample is good if it is representative of the population
  - Statistics are close in value to parameters
    - Means
    - Standard deviations
    - Medians
    - IQR's

# Caution

- It is difficult to determine if a sample is representative of a population.

# Caution

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

# Sampling Bias

- Sampling bias is also known as *sample selection bias*.

# 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 low* or *too high*.

# Size Bias

- Selection of a region to survey based on its population

# Size Bias

- Selection by dropping object on paper

# Size Bias

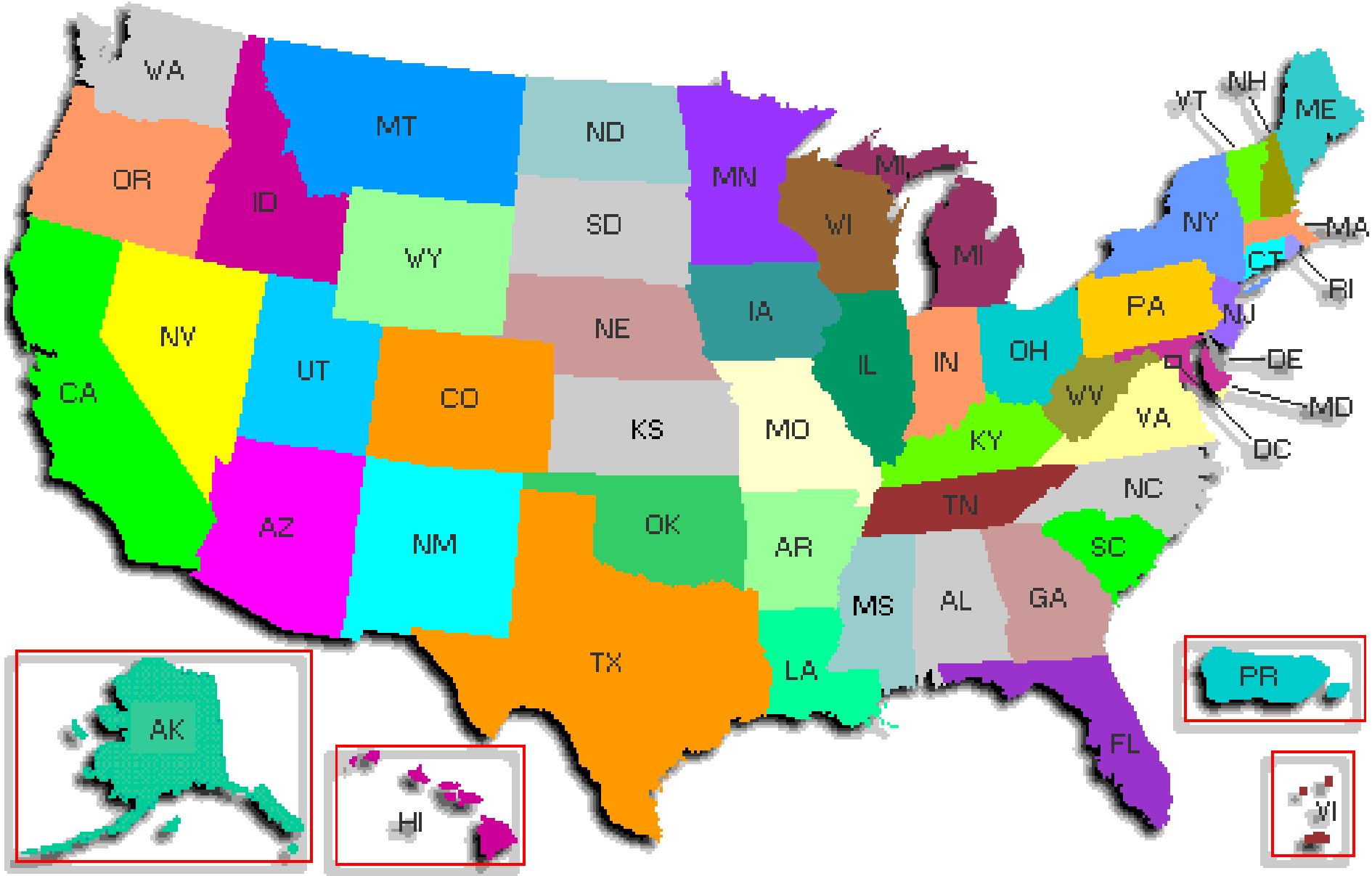
- Selection by dropping object on paper
  - Larger regions on paper more likely to be selected

# Size Bias

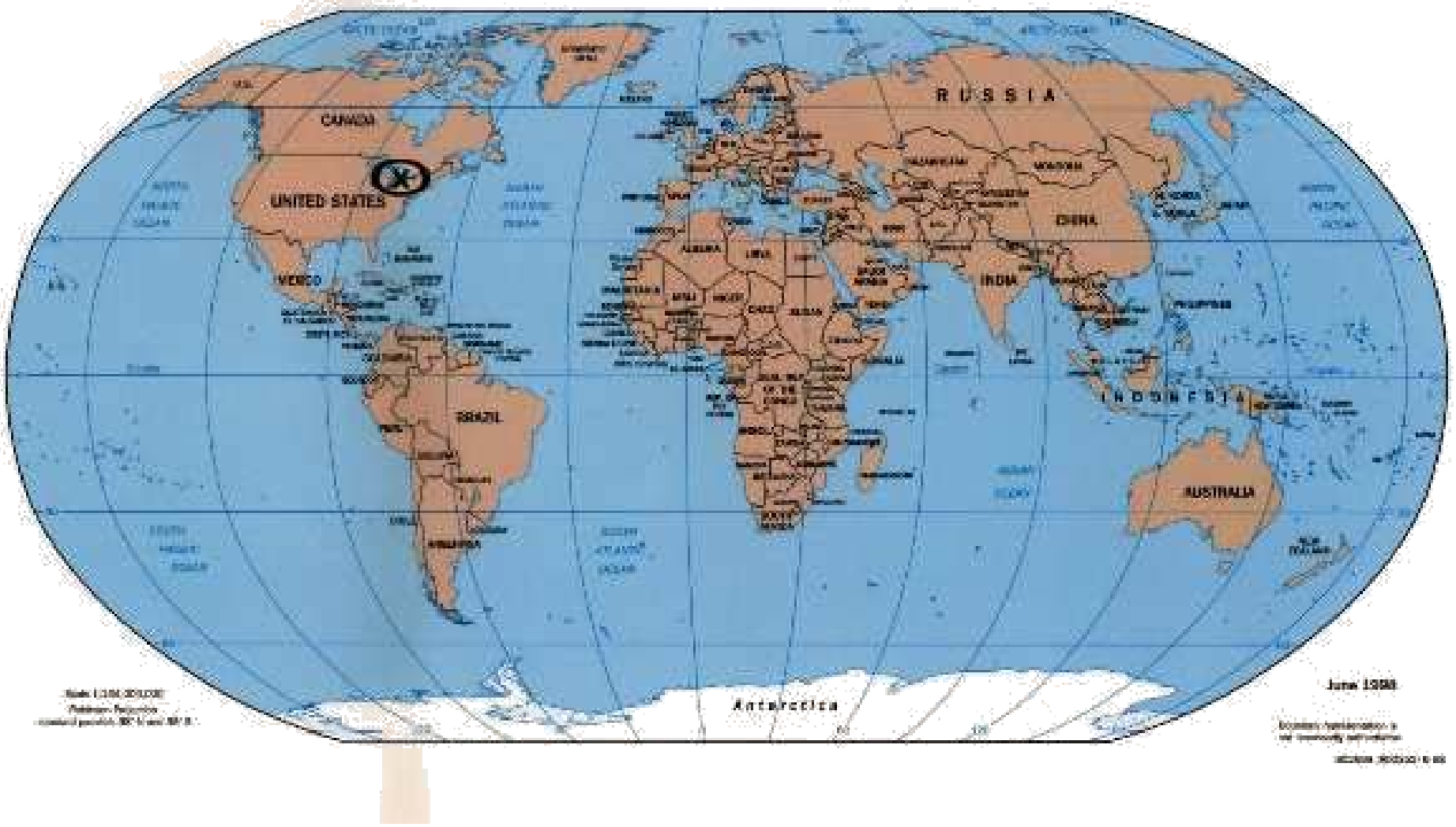
- Selection by dropping object on *a map*
  - Larger regions on paper more likely to be selected
- Examples
  - Selecting a country
  - Selecting a region of the state to survey/search

# Size Bias

- Consider the following three (3) maps.
  - What states/countries are more likely to be selected?
  - Are there any problems with these maps?


WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
VT
NH
ME
MA
CT
RI
NJ
DE
MD
DC
AK
HI
PR
VI


United States
N
Hudson Bay
YAHOO! Travel
RUS.
AK
CAN.
60°
0 500 mi
0 500 km
150°
CANADA
Pacific Ocean
HI
20°
0 150 mi
0 150 km
157°
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
MD
DE
NJ
CT
RI
MA
VT
NH
ME
Ottawa
Washington, D.C.
UNITED STATES
MEXICO
Pacific Ocean
Atlantic Ocean
Gulf of Mexico
Northern Mariana Islands
Guam
American Samoa
U.S. Virgin Islands
Puerto Rico
45°
30°
120°
100°
80°
0 250 500 mi
0 250 500 km
©maps.com


RUSSIA
CANADA
UNITED STATES
MEXICO
BRAZIL
CHINA
INDIA
INDONESIA
AUSTRALIA
Antarctica
June 1998

# Voluntary Response Bias

# Voluntary Response Bias

- Television/Radio call-in programs

# Voluntary Response Bias

- Television/Radio call-in programs
  - Who participates?

# Voluntary Response Bias

- Television/Radio call-in programs
  - Who participates?
    Those who
    - *Care* about an issue
    - *Want* to be on TV/radio

# Voluntary Response Bias

- Television/Radio call-in programs
  - Who does not participate?

# Voluntary Response Bias

- Television/Radio call-in programs
  - Who does not participate? Those who
    - *Do not care* about an issue
    - *Do not want* to be on TV/radio
    - Do not *want* to spend the time
    - *Are not watching/listening*

# Voluntary Response Bias

- Selective participation
  - Interested/willing – *participate*
  - Not interested/willing, not available – *do not participate*

# Convenience Sample

# Convenience Sample

- A sample for which the units selected are *convenient/easy* to find/include

# Convenience Sample

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

# Judgment Samples

# Judgment Samples

- Samples determined based on the *judgment of an expert*

# Judgment Samples

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

# Judgment Samples

- 1948 presidential election polls were wrong
  - Poll takers used quotas for sample selection

# Judgment Samples

- 1948 presidential election polls were wrong
  - Poll takers used quotas for sample selection
    - So many men, women, over 40, under 40, …

# Judgment Samples

- 1948 presidential election polls were wrong
  - Poll takers used quotas for sample selection
- Quota sampling is biased

# Judgment Samples

- 1948 presidential election polls were wrong
  - Poll takers used quotas for sample selection
- Quota sampling is biased
  - Proportions of groups within the population must be used to determine counts for groups

# What about Exit Polls?

# 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

- John Kerry was not elected as president in 2004.

# 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?

# What about Exit Polls?

What are some possible problems with Exit Polls?

# What about Exit Polls?

What are some possible problems with Exit Polls?

- Not everyone is counted

# What about Exit Polls?

What are some possible problems with Exit Polls?

- Not everyone is counted
  - Absentee/mail-in ballots
  - All polling places
  - All polling times
  - Selective participation

# Unbiased Sampling Methods

- All units in the population must have a known/equal chance of being selected/sampled

# Unbiased Sampling Methods

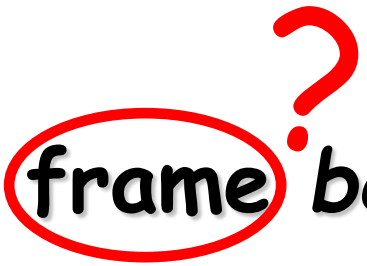
- All units in the population must have a known/equal chance of being selected/sampled
- How do this?

# Unbiased Sampling Methods

- All units in the population must have a known/equal 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/equal 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/equal 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

- To create a sampling frame, you must
  - Understand
  - Know
  - Be familiar with

  the population.

# Unbiased Sampling Methods

- If you start with a bad frame then you generate a bad sample

# Unbiased Sampling Methods

- 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 FSU
  - People living in the United States
  - Internet users
  - Population for a health care survey/study

# Unbiased Sampling Methods

- The units in the sample can represent the frame *well* but may not be representative of the actual population

# 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

# 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

# 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 – phone book???
  - Internet survey – Internet "book"???

# 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 – phone/phone book
  - Internet survey – internet access

Does everyone have these???

# 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 – phone/phone book
  - Internet survey – internet access

- In general – for any survey/study, individuals must be *willing* to participate

# Surveys

- What bias could be present in surveys of the following groups used to represent *all* FSU students?
  - Course members surveyed on 09/02/2020 regarding the percentage of students who studied at least two hours the *previous* night

# Surveys

- What bias could be present in surveys of the following groups used to represent *all* FSU students?
  - Course members surveyed on 10/30/2020 regarding the percentage of students who studied at least two hours the *previous* night

# Surveys

- What bias could be present in surveys of the following groups used to represent *all* FSU students?
  - Students enrolled in *American Writers I or American Writers II* surveyed to determine the percentage of
    - Mathematics majors
    - English majors
    - Elementary Education majors

# Response Bias

# Response Bias

- 40% of all people contacted regarding participating in surveys *refuse* to respond

# Response Bias

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

# Response Bias

- 40% of all people contacted regarding participating in surveys *refuse* to respond
- Those who agree to participate can be different from those who refuse to participate
  - Not all groups are represented

# Response Bias

- Consider this scenario
  - A Study determined that left-handed people die, on average, nine years before right-handed people

# Response Bias

- Consider this scenario
  - A Study determined that left-handed people die, on average, nine years before right-handed people
    - Problems with the study
      - Only half of the questionnaires were returned

# Response Bias

- Consider this scenario
  - A Study determined that left-handed people die, on average, nine years before right-handed people
    - Problems with the study
      - Surveys sent to everyone listed on death certificates in two LA counties to *inquire* about the handedness of the person who had died

# Response Bias

- Consider this scenario
  - A Study determined that left-handed people die, on average, nine years before right-handed people
    - Problems with the study
      - Surveys sent to everyone listed on death certificates in two LA counties to *inquire* about the handedness of the person who had died
        - Those who received the survey may not know the handedness of the person who died

# Response Bias

- Consider this scenario
  - A Study determined that left-handed people die, on average, nine years before right-handed people
    - Problems with the study
      - Surveys sent to everyone listed on death certificates in two LA counties to *inquire* about the handedness of the person who had died
        - What was happening in these LA counties at the time the survey was conducted???

# Questionnaire Bias

# Questionnaire Bias

- How you *ask* a question affects the response

# Questionnaire Bias

- How you *ask* a 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

# 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

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.*

*Respondents asked to agree or disagree with the statements*

# Questionnaire Bias

Reader's Digest survey

I. *I would be disappointed if Congress cut its funding for public television.*

*What do you notice about the wording in this question?*

# Questionnaire Bias

Reader's Digest survey

I. *I would be disappointed if Congress cut its funding for public television.*

*What do you notice about the wording in this question?*

*"I would be disappointed"*

# Questionnaire Bias

Reader's Digest survey

I. *I would be disappointed if Congress cut its funding for public television.*

*What do you notice about the wording in this question?*

*"I would be disappointed"*

- *This question is based on (your) feelings.*

# 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.*

*What do you notice about the wording in this question?*

# 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.*

*What do you notice about the wording in this question?*

*"Cuts in funding ... are justified"*

# 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.*

*What do you notice about the wording in this question?*

*"Cuts in funding ... are justified"*

- *This is a rationalization.*

# 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

How did people respond???

# Questionnaire Bias

Reader's Digest Survey Results

- Response to Statement *I*:
  - 54% agreed
  - 40% disagreed
  - 6% did not know
- Response to Statement *II*:
  - 52% agreed
  - 37% disagreed
  - 10% did not know

# Questionnaire Bias

Reader's Digest Survey Results

- Response to Statement *I*:
  - 54% agreed
  - 40% disagreed
  - 6% did not know
- Response to Statement *II*:
  - 52% agreed
  - 37% disagreed
  - 10% did not know

Notice the similarity in the percentages

# Other Bias

# Other Bias

- Some folks participating in surveys
  - Do not tell the truth

# Other Bias

- Some folks participating in surveys
  - Do not tell the truth
    - Intentional/unintentional lie

# Other Bias

- Some folks participating in surveys
  - Do not tell the truth
    - Intentional/unintentional lie
  - Give a response that
    - Sounds good
    - Seems to be desired/expected

# Other Bias

- Some folks participating in surveys
  - Do not tell the truth
    - Intentional/unintentional lie
  - Give a response that
    - Sounds good
    - Seems to be desired/expected
  - May not remember

# Other Bias

- Some folks participating in surveys
  - Do not tell the truth
    - Intentional/unintentional lie
  - Give a response that
    - Sounds good
    - Seems to be desired/expected
  - May not remember (events in their lives; previous experiences)

# Other Bias

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

# Other Bias

- Measurement bias

# 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

- 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
    - Devices not maintained/calibrated
    - Measurements not collected in a consistent manner

# Final Thoughts ...

# Avoid Bias???

# Avoid Bias???

- Understand the population
  - Determine the *proportions* of the characteristics of interest
  - Preserve these characteristics using same proportions in the sample

# Avoid Bias???

- Careful creation of sample
  - Create/use a frame
  - All units must *be equally likely* to be selected from population

# Avoid Bias???

- Careful creation of sample
  - Create/use a frame
  - All units must *have the same likelihood* to be selected from population

# Avoid Bias???

- Careful creation of sample
  - Create/use a frame
  - All units must *have the same chance* to be selected from population

# Avoid Bias???

- Do not use quotas
- Avoid "experts" who may have bias
- Do not use judgement samples

# Avoid Bias???

- Take accurate measurements
  - Collected in a consistent manner
- Use properly calibrated devices
- Use reliable means for data collection

# Survey Questions???

# Survey Questions???

- General
  - Not slanted
  - Consistent
- Be aware of words/tenses/view used

# Beware of???

# Beware of???

- Use of interviewers
  - Must be trained
    - Demeanor
    - Appearance
    - Tone of voice
  - Must be consistent for all participants
    - Interaction
    - Handling of participants

# Beware of???

- Use of interviewers
  - Must be trained
    - Demeanor
    - Appearance
    - Tone of voice
  - Must be consistent for all participants
    - Interaction
    - Handling of participants

**All participants should have the same experience.**

# Beware of???

- Response bias
- Voluntary response bias

# How Do We Create Samples???

# How Do We Create Samples???

That is what we will consider next!