Lesson 48
Sampling Methods and Bias

NAME:

Start by navigating to the Online Lesson for instructions.

Objectives

- Determine the type of probability sampling method being used.
- Identify the purpose of a study.
- Describe the bias or misuse of statistics.
- Write an unbiased statistical question.

Why?

The integrity of research data is only as good as the integrity applied to selecting sample populations. Similarly, considering bias in reports is an important part of evaluating statistics. While recognizing potential bias improves with experience, practicing will help you begin to identify it and build your inner healthy skepticism.

Warm Up

1) What is the difference between the population and the sample?

2) Why are samples more common than a census?

To continue, return to the Online Lesson.

Explore

Sampling Methods

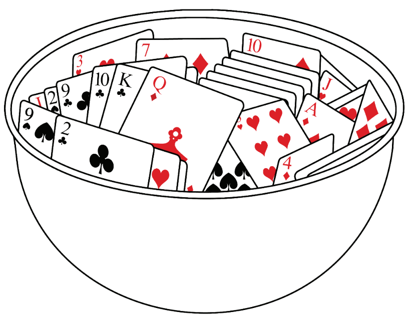
Fill in the notes as you watch the video in the Online Lesson.

- Probability sampling is the concept of statistical sampling in which every member of the population has ______________________________.
- Selecting a probability sampling method is determined by:
  - ______________________________
  - ______________________________
  - ______________________________

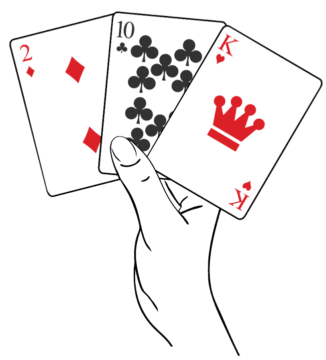
- Random sampling is one category of probability sampling that is used to help ensure that the sample accurately:
  - ____
  - ____
  - ____

Random Sampling Techniques

- ________________________: Any individual in the population has an equal chance of being selected at random from the population.

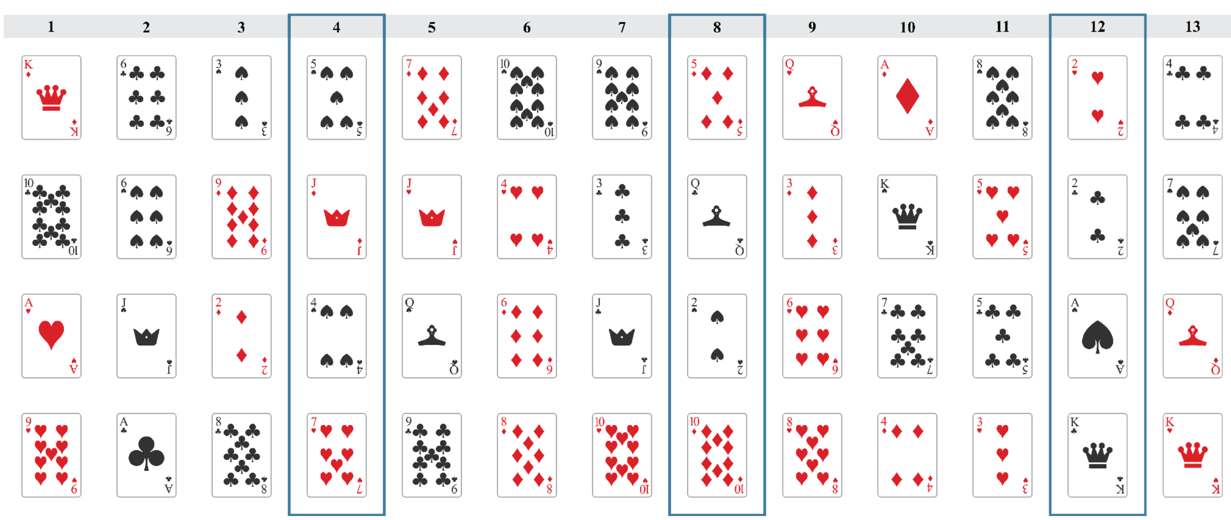


population

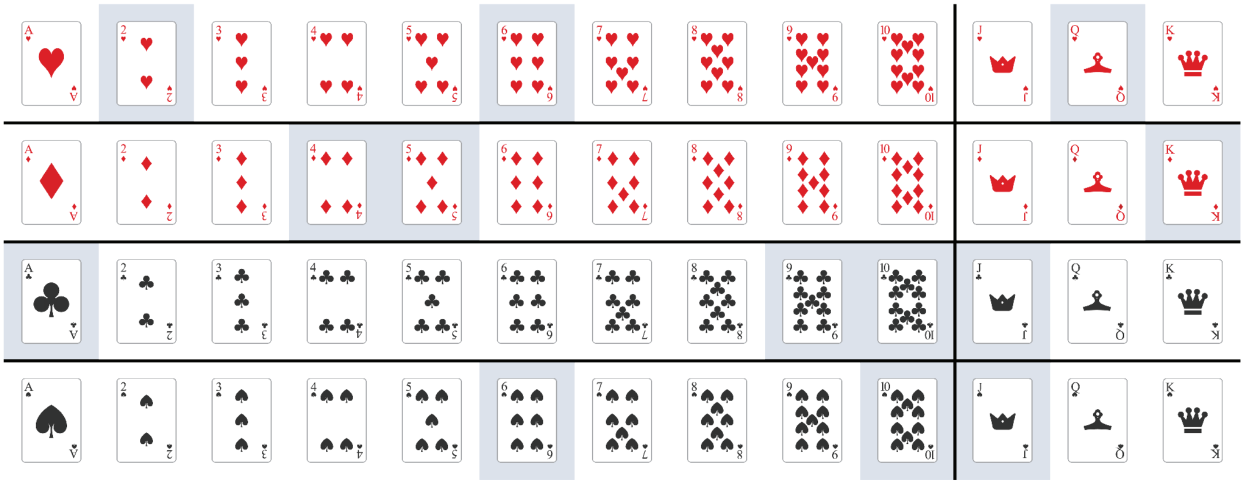


sample

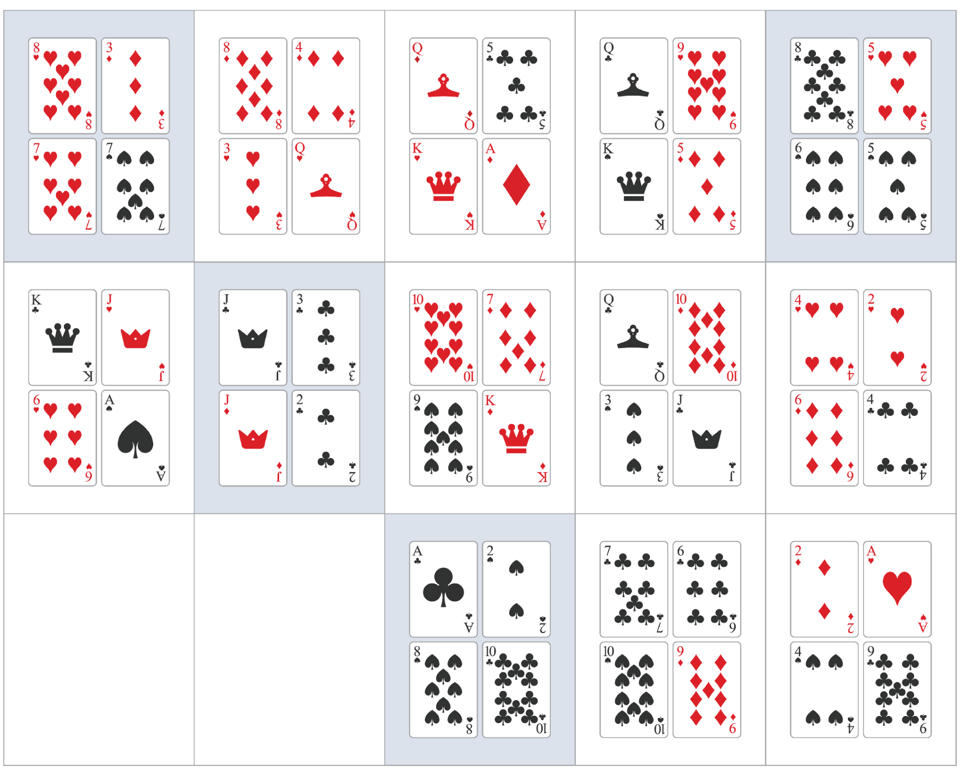
- ________________________: After the population is randomized to avoid bias, a starting point is determined from the population, and then every n^{th} individual is selected.



- ________________________________: The researcher first divides the population into smaller groups with similar population characteristics that do not overlap. Then the researcher takes a random sample from each group in proportion to the population.



- ________________________________: The researcher first groups the population by regions (or clusters), then entire clusters are randomly selected to represent the population.



Example 1

Complete the example as you watch the video in the Online Lesson.

Name the probability sampling method and the population.

A marketing company wants to test a new car advertisement. They divide their target audience into age groups $(18-25,\ 26-40,\ 41-60,\ 61+)$. From each age group, they randomly select a number of individuals, proportional to the group's size, to watch the ad and provide feedback.

Example 2

Complete the example as you watch the video in the Online Lesson.

Name the probability sampling method and the population.

A researcher wants to study the health habits of high school students in Los Angeles, CA. Instead of trying to survey individual students from every school, they randomly select a cluster of five high schools in LA, and then survey every student within those selected high schools.

Checkpoint: Sampling Methods

Name the probability sampling method and the population.

Ms. Guthrie wants to know how much time students spend doing homework each night. She makes a list of all 150 students in her classes, arranges them alphabetically, picks a random starting point, and then selects every tenth student on the list to survey.

To continue, return to the Online Lesson.

Statistical Questions and Bias

Fill in the notes as you watch the video in the Online Lesson.

- A statistical question has an answer that ______________________________.
- In the study of statistics, ______________________________________ in data collections.

- However, the results of data collection ______________________________ to show there is no bias (or little bias).
- If there is little to no variability, it is possible that the sample:
  - ______________________________
  - ______________________________
- Statistical bias is an error that causes the results of a study to ______________________________ ______________________________.
- When bias is present, the collected data is ______________________________ ______________________________ (or sometimes even the sample).
- Bias can occur when questions:
  - ______________________________
  - ______________________________
  - ______________________________
- Therefore, statistical questions should not:
  - ______________________________
  - ______________________________

    ____
- A ______________________________ is a simple statement that is clear and without presumption that explains the goal of the research.
  - When writing statistical questions, a purpose of study can help you write questions that ______________________________.
  - When participating in a study, identifying the purpose of study can help you be a healthy skeptic and ______________________________.

Example 3

Complete the example as you watch the video in the Online Lesson.

Describe the bias in each question.

A) How tall is the oak tree beside the playground at the park?

B) Everyone knows that schools with mascots are better than schools without. What are your reasons for agreeing?

Example 4

Complete the example as you watch the video in the Online Lesson.

Rewrite the question to remove bias.

Given the obvious benefits of raising the age at which you can get your license, what are your thoughts?

Checkpoint: Statistical Questions and Bias

Describe the bias in the question.

Do you agree that the beautiful new park is a fantastic addition to the community?

To continue, return to the Online Lesson.

The Misuse of Statistics

Fill in the notes as you watch the video in the Online Lesson.

If you cannot properly represent a population, you should not make claims about the population.

- A **non-representative sampling** occurs when the sampling methods:
  - Do not match ________________________________
  - Do not allow everyone in a population ________________________________

    ____
- Misuse of statistics can occur when a researcher provides a non-representative sampling that:
  - ________________________________
  - ________________________________

Types of Non-Representative Sampling

- Some **non-probability sampling methods** include:
  - ____________________: Researchers only sample people closest/most convenient to them.
  - ____________________: Members of the population choose to participate in the study.
  - ____________________: The sample is too small to draw any valid conclusions about the population.
- Some types of **non-statistical reporting** include:
  - ____________________: Only the "good" information is presented to represent what the researcher wants to show.
  - ____________________: The statistic lacks a point of reference or basis for comparison, i.e., the "compared to what" part of the sentence is missing.

- ______________________________: A number or graph is shown in a way to "prove" a point, but is not representative of the population.
- ______________________________ questions.

Example 5

Complete the example as you watch the video in the Online Lesson.

A) Name the intended purpose of study.

B) Identify and explain the misuse of statistics.

The principal at East Middle School wanted to prove that the school's After-the-Bell tutoring program causes an increase in student test scores. Signing up for After-the-Bell is voluntary. At the end of the school year, the principal reported: "The average test scores of After-the-Bell students increased 15%. This proves the tutoring program effectively boosts student achievement."

A) Purpose of Study: Determine if After-the-Bell's tutoring program ______________________ ______________________.

B) Misuse:

Identify	Explain
Self-selected	
Detached	

Example 6

Complete the example as you watch the video in the Online Lesson.

Identify and explain the misuse of statistics and suggest how to improve it.

The Start-Up Tech firm was hiring and wanted to attract more applicants. The company's founder earns a $2 million annual salary. The three managers earn $150,000 annually. The other 30 employees earn $60,000 annually each. Start-Up Tech posted: "Come join our team! The average salary at our firm is $125,000!"

Checkpoint: The Misuse of Statistics

A) Name the intended purpose of the study.

B) Identify and explain the misuse of statistics.

A local radio show conducts a phone-in poll on its morning show, saying, "Call in to vote for the city to invest in new, dedicated bike lanes." The radio show then reported, "An overwhelming 85% of callers support investing in new bike lanes!"

To continue, return to the Online Lesson.

Practice 1

Complete problems on a separate sheet of paper.

For problems 1–6, select a sampling method, then name the population.

A) Simple random **B)** Systematic **C)** Stratified **D)** Cluster

1) A park ranger wants to estimate the deer population in Coventry Woods. She divides the forest into a grid, randomly selects a starting grid square, and then surveys every fifth grid square to count the number of deer.

2) The Marshalton School District wants to understand how parents feel about online learning. They divide parents into groups based on the grade level of their oldest child (elementary, middle, high school). Then, they randomly select a proportional number of parents from each group to participate in a survey.

3) A statewide grocery store chain wants to assess customer satisfaction in the geographic regions in which the stores are located. The chain randomly selects 3 of the 10 geographic regions and surveys customers entering those regional stores on a specific day.

4) The principal wants feedback from students on the new lunch menu. She puts all 1,200 student names into a hat and draws out 50 names without looking. These 50 students will be surveyed.

5) A librarian wants to survey library patrons about their satisfaction with the new digital resources. Every hour, on the hour, for one day, the librarian approaches every fifth person who checks out a book and asks them to complete a short survey.

6) A wildlife biologist is studying the health of the crayfish population in a lake. They net a large number of crayfish and then randomly select 30 crayfish from the net to take blood samples and measure their length and weight.

For problems 7–8, explain your thinking.

7) Why is it important that being selected for a sample is randomized?

8) What is the reason for having many sampling methods in statistics?

For problems 9–10, identify the bias in the question. Then rewrite the question to remove the bias.

9) Do you prefer the delicious taste and superior quality of Brand X coffee over other bland coffee brands?

10) Given the obvious advantages of playing video games for improving problem-solving skills, what is the minimum number of hours a day students should be allowed to play?

Identify and explain the misuse of statistics.

11) A company wanted to know if its employees' eyes were less strained if a screen filter was used during computer work. They selected 100 employees to participate. The list of participants was alphabetized, and the first 50 employees had a screen filter added to their computers without them knowing (treatment group). The remaining 50 employees did not have a screen filter (control).

12) A call center tracked one employee’s first ten calls of the day. The employee was able to resolve seven out of ten calls in ten minutes or less each. On any given day, the call center receives an average of 375 calls. The call center reported: “Our representatives resolve 70% of all issues in under ten minutes per call.”

To continue, return to the Online Lesson.

Mastery Check

Show What You Know

Following six consecutive weeks of daily accidents on a major highway near Chicago, an Illinois civil engineer is investigating traffic congestion. The engineer's first step toward addressing the issue is to determine the average speed and type of vehicles on the highway.

To collect data, one clearly marked, fully visible police car is parked along the highway. Over one single morning rush hour, the police officer used radar to record the speed and vehicle type of every tenth vehicle that passed a designated checkpoint. The officer reported speeds were 3-5 mph under the posted speed limit.

A) Identify the population and the sampling method used.

B) Identify the flaws and/or bias in the sampling. Make suggestions for improvement.

Unsatisfied with the initial sampling method, the civil engineer seeks additional feedback from highway drivers. A billboard on the highway encourages drivers to complete a survey after their journeys. One survey question asks, “Given the unacceptable levels of traffic congestion, how strongly do you agree that the state must immediately dedicate funding to improve the highway?”

C) Identify the biased wording, then rewrite the question to remove bias.

Say What You Know

In your own words, talk about what you have learned using the objectives for this lesson and your work on this page.

To continue, return to the Online Lesson.

Practice 2

Complete problems on a separate sheet of paper.

For problems 1–6, select a sampling method, then name the population.

A) Simple random **B)** Systematic **C)** Stratified **D)** Cluster

1) A quality control manager at a light bulb factory wants to check the lifespan of a batch of 10,000 light bulbs. He uses a random number generator that selects 100 light bulbs from the production line for testing.

2) A city planner is studying traffic patterns on Champion Avenue. They set up a camera that records cars passing a specific point. Starting at a random minute in the hour, they record the average traffic speed every ten minutes for a full day.

3) A factory manager needs to inspect a batch of 500 newly manufactured car parts for defects. They line up all the parts, pick a random starting part, and then inspect every 20th part on the line after that.

4) A national research organization wants to survey adults in a specific state about their exercise habits. They divide the state into counties (geographic regions), randomly select 15 counties, and then survey every adult residing in those selected counties.

5) A university wants to understand student opinions on a new campus policy. They divide the student body into undergraduate, graduate, and doctoral students. From each of these groups, a proportional number of students is randomly selected to participate in a survey.

6) A company is conducting a prize drawing for its employees. They assign a unique number to each of the 250 employees and then use a random number generator to pick ten winning numbers.

For problems 7–8, identify the bias in the question. Then rewrite the question to remove the bias.

7) Because everyone wants to be stylish, what brand name clothing is the best?

8) As a person who cares about the environment, don't you agree that nothing else should ever be printed on paper in order to save the trees?

For problems 9–10, identify and explain the misuse of statistics.

9) An advertisement for a new job-search website claims that people who use the site are highly successful. The ad states that users have a 200% higher chance of finding a job. A footnote clarifies that this is based on a comparison to a baseline control group of non-users who had a 1% chance of finding a job. The claim: "Use our site for a 200% higher chance of finding your dream job!"

10) A university randomly selected 500 college seniors to ask them how many hours per week they spent studying and what their final GPA was for that semester. The report: The researchers concluded that there is a strong positive relationship between the two variables.

To continue, return to the Online Lesson.