

22.1 Data-Gathering Techniques



Resource Locker

Essential Question: Under what circumstances should a sample statistic be used as an estimator of a population parameter?

Explore Finding the Mean of Samples Obtained from Various Sampling Methods

You collect data about a **population** by surveying or studying some or all of the individuals in the population. When *all* the individuals in a population are surveyed or studied, the data-gathering technique is called a **census**. A **parameter** is a number that summarizes a characteristic of the population. When only some of the individuals in a population are surveyed or studied, the data-gathering technique is called **sampling**. A **statistic** is a number that summarizes a characteristic of a sample. Statistics can be used to estimate parameters.



Consider the following table, which lists the salaries (in thousands of dollars) of all 30 employees at a small company. In this Explore, you will take samples from this population and compute the mean (the sum of the data divided by the sample size).

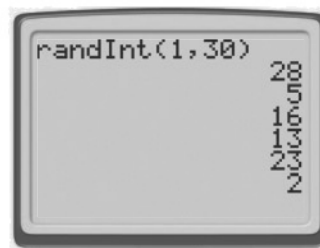
Salaries at a Small Company									
21	24	26	28	30	32	33	35	37	41
44	46	47	49	50	51	52	54	55	57
58	62	62	64	64	65	70	71	73	80

A Suppose the employees whose salaries are 51, 57, 58, 65, 70, and 73 volunteer to be in the sample. This is called a *self-selected sample*. Compute the sample's mean, rounding to the nearest whole number.

B Suppose the six salaries in the first two columns of the table are chosen. This is called a *convenience sample* because the data are easy to obtain. Record the salaries, and then compute the sample's mean, rounding to the nearest whole number.

C Suppose every fifth salary in the list, reading from left to right in each row, is chosen. This is called a *systematic sample*. Record the salaries, and then compute the sample's mean, rounding to the nearest whole number.

- D Label the data in the table with the identifiers 1–10 for the first row, 11–20 for the second row, and 21–30 for the third row. Then use a graphing calculator's random integer generator to generate six identifiers between 1 and 30, as shown. (If any identifiers are repeated, simply generate replacements for them until you have six unique identifiers.) This is called a *simple random sample*. Record the corresponding salaries, and then compute the sample's mean, rounding to the nearest whole number.



Reflect

1. Compute the mean of the population. Then list the four samples from best to worst in terms of how well each sample mean estimates the population mean.
2. With the way the table is organized, both the convenience sample and the systematic sample have means that are not too far from the population mean. Why?

Explain 1 Distinguishing Among Sampling Methods

The goal of sampling is to obtain a **representative sample**, because the statistic obtained from the sample is a good estimator of the corresponding population parameter. Some sampling methods can result in **biased samples** that may not be representative of the population and can produce statistics that lead to inaccurate conclusions about the corresponding population parameters.

Sampling Method	Description
Simple random sample	Each individual in the population has an equal chance of being selected.
Self-selected sample	Individuals volunteer to be part of the sample.
Convenience sample	Individuals are selected based on how accessible they are.
Systematic sample	Members of the sample are chosen according to a rule, such as every n th individual in the population.
Stratified sample	The population is divided into groups, and individuals from each group are selected (typically through a random sample within each group).
Cluster sample	The population is divided into groups, some of the groups are randomly selected, and either all the individuals in the selected groups are selected or just some of the individuals from the selected groups are selected (typically through a random sample within each selected group).

The Explore showed that simple random samples are likely to be representative of a population (as are other sampling methods that involve randomness) and are therefore preferred over sampling methods that don't involve randomness.

Example 1 Identify the population, classify the sampling methods, and decide whether the sampling methods could result in a biased sample. Explain your reasoning.

- A** The officials of the National Football League (NFL) want to know how the players feel about some proposed changes to the NFL rules. They decide to ask a sample of 100 players.
- The officials choose the first 100 players who volunteer their opinions.
 - The officials randomly choose 3 or 4 players from each of the 32 teams in the NFL.
 - The officials have a computer randomly generate a list of 100 players from a database of all NFL players.



The population consists of the players in the NFL.

- This is a self-selected sample because the players volunteer their opinions. This could result in a biased sample because the players who feel strongly about the rules would be the first ones to volunteer and get their opinions counted.
- This is a stratified sample because the players are separated by team and randomly chosen from each team. This is not likely to be biased since the players are chosen randomly and are taken from each team.
- This is a simple random sample because each player has an equally likely chance of being chosen. This is not likely to be biased since the players are chosen randomly.

- B** Administrators at your school want to know if students think that more vegetarian items should be added to the lunch menu.
- The administrators survey every 25th student who enters the cafeteria during the lunch period.
 - The administrators survey the first 50 students who get in the lunch line to buy lunch.
 - The administrators use a randomly generated list of 50 students from a master list of all students.

The population consists of _____.

- This is a _____ sample because _____. This method [is/isn't] likely to result in a biased sample because _____.
- This is a _____ sample because _____. This method [is/isn't] likely to result in a biased sample because _____.
- This is a _____ sample because _____. This method [is/isn't] likely to result in a biased sample because _____.

Your Turn

Identify the population, classify the sampling methods, and decide whether the sampling methods could result in a biased sample. Explain your reasoning.

3. A local newspaper conducts a survey to find out if adult residents of the city think the use of hand-held cell phones while driving in the city should be banned.
 - a. The newspaper sends a text message to a random selection of 1000 subscribers whose cell phones are listed in the paper's subscription database.
 - b. Using the 10 neighborhoods into which the city is divided, the newspaper randomly contacts 100 adults living in each of the neighborhoods.

Explain 2 Making Predictions from a Random Sample

In statistics, you work with data. Data can be **numerical**, such as heights or salaries, or **categorical**, such as eye color or political affiliation. While a statistic like the mean is appropriate for numerical data, an appropriate statistic for categorical data is a **proportion**, which is the relative frequency of a category.

Example 2 A community health center surveyed a small random sample of adults in the community about their exercise habits. The survey asked whether the person engages in regular cardio exercise (running, walking, swimming, or other) and, if so, what the duration and frequency of exercise are. Of the 25 people surveyed, 10 said that they do engage in regular cardio exercise. The table lists the data for those 10 people. Calculate statistics from the sample, and use the statistics to make predictions about the exercise habits of the approximately 5000 adults living in the community.

Type of exercise	Duration (minutes spent exercising)	Frequency (times per week)
Running	30	4
Walking	20	5
Running	40	3
Running	60	6
Swimming	40	4
Other	90	2
Running	30	3
Walking	20	5
Running	30	4
Other	120	1

- A** Calculate the proportion of adults who get regular cardio exercise and the proportion of runners among those who get regular cardio exercise. Use the proportions to predict the number of runners among all adults living in the community.

Proportion of adults who get regular cardio exercise: $\frac{10}{25} = 0.4$ or 40%

Proportion of runners among those who get regular cardio exercise: $\frac{5}{10} = 0.5$ or 50%

To predict the number of runners in the community, multiply the number of adults in the community by the proportion of adults who get regular cardio exercise and then by the proportion of runners among those who get regular cardio exercise.

Predicted number of runners in the community: $5000 \cdot 0.4 \cdot 0.5 = 1000$

- B** Calculate the mean duration of exercise for those who get regular cardio exercise and the mean frequency of exercise for those who get regular cardio exercise. Use the means to predict, for those who get regular cardio exercise, the number of hours spent exercising each week. Show your calculations and include units.

Mean duration of exercise for those who get regular cardio exercise: _____ minutes

Mean frequency of exercise for those who get regular cardio exercise: _____ times per week

To predict the number of hours spent exercising, multiply the mean duration of exercise (in minutes) for those who get regular cardio exercise by the mean frequency of exercise (in times per week) for those who get regular cardio exercise. This product will be in minutes per week. To convert to hours per week, also multiply by the conversion factor $\frac{1 \text{ hour}}{60 \text{ minutes}}$.

Predicted time spent exercising: minutes \cdot /week $\cdot \frac{1 \text{ hour}}{60 \text{ minutes}} \approx$ hours/week

Reflect

- 4. Discussion** How much confidence do you have in the predictions made from the results of the survey? Explain your reasoning.

Your Turn

- 5.** A ski resort uses the information gained from scanning season ski passes in lift lines to determine how many days out of each season the pass holders ski and how many lift rides they take each day. The table lists the data for a random sample of 16 pass holders. Calculate the mean number of days skied and the mean number of lift rides taken per day. Use the means to predict the number of lift rides taken per season by a pass holder.

Number of days	10	5	2	14	27	3	18	5
Number of lift rides	12	15	6	18	10	6	15	9
Number of days	4	16	7	12	19	14	25	13
Number of lift rides	11	13	14	10	8	6	15	18

Elaborate

6. Name a sampling method that is more likely to produce a representative sample and a sampling method that is more likely to produce a biased sample.

7. Why are there different statistics for numerical and categorical data?

8. **Essential Question Check-In** Explain the difference between a parameter and a statistic.

Evaluate: Homework and Practice



- Online Homework
- Hints and Help
- Extra Practice

1. A student council wants to know whether students would like the council to sponsor a mid-winter dance or a mid-winter carnival this year. Classify each sampling method.
- Survey every tenth student on the school's roster.
 - Survey all students in three randomly selected homerooms.
 - Survey 20 randomly selected freshmen, 20 randomly selected sophomores, 20 randomly selected juniors, and 20 randomly selected seniors.
 - Survey those who ask the council president for a questionnaire.
 - Survey a random selection of those who happen to be in the cafeteria at noon.
2. The officers of a neighborhood association want to know whether residents are interested in beautifying the neighborhood and, if so, how much money they are willing to contribute toward the costs involved. The officers are considering two methods for gathering data:
- Method A: Call and survey every tenth resident on the association's roster.
- Method B: Randomly select and survey 10 residents from among those who come to the neighborhood block party.
- Identify the population.

- b. Which sampling method is most likely to result in a representative sample of the population? Explain.
- c. Describe another sampling method that is likely to result in a representative sample of the population.
- d. Describe the categorical and numerical data that the officers of the neighborhood association want to gather.

Decide whether the sampling method could result in a biased sample. Explain your reasoning.

- 3. On the first day of school, all of the incoming freshmen attend an orientation program. Afterward, the principal wants to learn the opinions of the freshmen regarding the orientation. She decides to ask 25 freshmen as they leave the auditorium to complete a questionnaire.
- 4. The members of the school drama club want to know how much students are willing to pay for a ticket to one of their productions. They decide that each member of the drama club should ask 5 of his or her friends what they are each willing to pay.
- 5. A medical conference has 500 participating doctors. The table lists the doctors' specialties. A researcher wants to survey a sample of 25 of the doctors to get their opinions on proposed new rules for health care providers. Explain why it may be better for the researcher to use a stratified sample rather than a simple random sample.

Specialty	Number of Doctors
Dermatology	40
Geriatrics	120
Oncology	140
Pediatrics	100
Surgery	100

- 6. A researcher wants to conduct a face-to-face survey of 100 farmers in a large agricultural state to get their opinions about the risks and rewards of farming. The researcher has limited time and budget. Explain why it may be better for the researcher to use a cluster sample based on counties in the state rather than a simple random sample.

Identify the population and the sampling method.

- 7. A quality control inspector at a computer assembly plant needs to estimate the number of defective computers in a group of 250 computers. He tests 25 randomly chosen computers.
- 8. The manager of a movie theater wants to know how the movie viewers feel about the new stadium seating at the theater. She asks every 30th person who exits the theater each Saturday night for a month.

9. Eric is interested in purchasing a used sports car. He selects the make and model of the car at a website that locates all used cars of that make and model for sale within a certain distance of his home. The website delivers a list of 120 cars that meet his criteria. Eric randomly selects 10 of those cars and records what type of engine and transmission they have, as shown in the table.

Engine (L = liter)	Transmission
3.6 L V6	Manual
3.6 L V6	Automatic
6.2 L V8	Manual
3.6 L V6	Manual
6.2 L Supercharged V8	Manual
3.6 L V6	Automatic
6.2 L V8	Manual
3.6 L V6	Automatic
3.6 L V6	Manual
6.2 L V8	Automatic

a. If Eric can only drive cars with automatic transmissions, predict the number of such cars on the website's list. Show your calculations.

b. Eric knows that the 3.6 L V6 engine takes regular gasoline, but the 6.2 L V8 engine requires premium gasoline. To minimize his fuel costs, he wants a car with the 3.6 L V6 engine. Predict the number of such cars on the website's list. Show your calculations.

c. Predict the number of cars that have a 3.6 L V6 engine and an automatic transmission. Show your calculations.

- 10.** A community theater association plans to produce three plays for the upcoming season. The association surveys a random sample of the approximately 7000 households in the community to see if an adult member of the household is interested in attending plays and, if so, what type of plays the person prefers (comedy, drama, or musical), how many members of the household (including the person surveyed) might attend plays, and how many of the three plays those household members might attend. Of the 50 adults surveyed, 12 indicated an interest in attending plays. The table lists the data for those 12 people.



Preferred Type of Play	Number of People Attending	Number of Plays Attending
Comedy	2	1
Musical	3	2
Musical	1	2
Drama	2	3
Comedy	3	2
Comedy	2	3
Musical	4	1
Drama	2	3
Comedy	2	2
Musical	2	3
Comedy	5	1
Drama	1	2

- Describe the categorical and numerical data gathered in the survey.
- Calculate the proportion of adults who indicated an interest in attending plays and calculate the proportion of adults who prefer dramas among those who are interested in attending plays. If approximately 15,000 adults live in the community, predict the number of adults who are interested in attending plays that are dramas. Show your calculations.

- c. For an adult with an interest in attending plays, calculate the mean number of household members who might attend plays and the mean number of plays that those household members might attend. Round each mean to the nearest tenth. If the theater association plans to sell tickets to the plays for \$40 each, predict the amount of revenue from ticket sales. Show your calculations and include units.

11. Match each description of a sample on the left with a sampling technique on the right.

- | | | |
|---|-------|----------------------|
| A. A television reporter asks people walking by on the street to answer a question about an upcoming election. | _____ | Simple random sample |
| B. A television reporter randomly selects voting precincts and then contacts voters randomly chosen from a list of registered voters residing in those precincts to ask about an upcoming election. | _____ | Self-selected sample |
| C. A television reporter contacts voters randomly chosen from a complete list of registered voters to ask about an upcoming election. | _____ | Convenience sample |
| D. A television reporter contacts every 100th voter from a complete list of registered voters to ask about an upcoming election. | _____ | Systematic sample |
| E. A television reporter contacts registered voters randomly chosen from every voting precinct to ask about an upcoming election. | _____ | Stratified sample |
| F. A television reporter asks viewers to call in their response to a question about an upcoming election. | _____ | Cluster sample |

H.O.T. Focus on Higher Order Thinking

12. Critical Thinking A reporter for a high school newspaper asked all members of the school's track team how many miles they run each week.

- a. What type of data did the reporter gather?
- b. Was the reporter's data-gathering technique a census or a sample? Explain.
- c. Are the data representative of the entire student body? Why or why not?

13. Communicate Mathematical Ideas Categorical data can be nominal or ordinal. *Nominal data* refer to categories that do not have any “natural” ordering, while *ordinal data* refer to categories that do have an order. Similarly, numerical data can be discrete or continuous. *Discrete data* are typically counts or scores (which cannot be made more precise), while *continuous data* are typically measurements (which can be made more precise). For each description of a set of data, identify whether the data are nominal, ordinal, discrete, or continuous. Explain your reasoning. Also give another example of the same type of data.

- a. A researcher records how many people live in a subject’s household.
- b. A researcher records the gender of each subject.
- c. A researcher records the amount of time each subject spends using an electronic device during a day.
- d. A researcher records whether each subject is a young adult, a middle-aged adult, or a senior.

14. Analyze Relationships The grid represents the entire population of 100 trees in an apple orchard. The values in the grid show the number of kilograms of apples produced by each tree during one year. Given the data, obtain three random samples of size 20 from the population and find the mean of each sample. Discuss how the means of those samples compare with the population mean, which is 68.9.

109	52	62	72	110	61	51	50	100	50
54	104	54	111	74	73	77	68	65	66
108	53	27	75	52	117	76	60	64	67
73	36	103	71	67	60	59	26	80	61
38	63	35	112	75	68	51	72	79	62
58	105	55	53	118	57	101	66	116	31
29	57	74	33	102	69	28	71	30	58
39	55	34	120	64	114	70	113	78	63
107	37	56	25	76	70	69	77	30	115
56	40	106	32	119	65	80	78	79	59

Lesson Performance Task

Think about your school's cafeteria and the food it serves. Suppose you are given the opportunity to conduct a survey about the cafeteria.

- a. Identify the population to be surveyed.
- b. Write one or more survey questions. For each question, state whether it will generate numerical data or categorical data.
- c. Assuming that you aren't able to conduct a census of the population, describe how you could obtain a representative sample of the population.

- d. Suppose you asked a random sample of 25 students in your school whether they were satisfied with cafeteria lunches and how often in a typical week they brought their own lunches. The tables give the results of the survey. If the school has 600 students, use the results to predict the number of students who are satisfied with cafeteria lunches and the number of lunches brought to school in a typical week.

Satisfied with cafeteria lunches?	
Response	Number
Yes	18
No	7

Bring own lunches how often in a week?	
Response	Number
5 times	2
4 times	1
3 times	4
2 times	2
1 time	4
0 times	12