# Chapter 3: Statistics

## Section 3.1 Overview of the Statistical Process

### Exercises 3.1

1. Describe the difference between a sample and a population. ADD
2. Describe the difference between a statistic and a parameter. ADD

A\

1. The ASPCC randomly selects 200 students from PCC Cascade campus to participate in a childcare survey in order to determine the demand for additional childcare options for PCC students.
   1. Who is the intended population?
   2. What is the sample?
   3. Is the collected data representative of the intended population? Why or why not? ADD
2. A local research firm randomly selects 1200 homes in Washington County to determine support for adding compost pick up to residents’ regular garbage service.
   1. Who is the intended population?
   2. What is the sample?
   3. Is the collected data representative of the intended population? Why or why not? ADD
3. A political scientist surveys 28 of the current 106 representatives in a state's congress. Of them, 14 said they were supporting a new education bill, 12 said there were not supporting the bill, and 2 were undecided.
   1. Who is the population of this study?
   2. What is the size of the population?
   3. What is the size of the sample?
   4. Give the statistic for the percentage of representatives surveyed who said they were supporting the education bill.
   5. If the margin of error was 5%, state the confidence interval for the percentage of representatives we might we expect to support the education bill, and explain what the confidence interval tells us.
4. The city of Raleigh has 9,500 registered voters. There are two candidates for city council in an upcoming election: Brown and Feliz. The day before the election, a telephone poll of 350 randomly selected registered voters was conducted. 112 said they'd vote for Brown, 207 said they'd vote for Feliz, and 31 were undecided.
   1. Who is the population of this survey?
   2. What is the size of the population?
   3. What is the size of the sample?
   4. Give the statistic for the percentage of voters surveyed who said they'd vote for Brown.
   5. If the margin of error was 3.5%, state the confidence interval for the percentage of voters surveyed that we might we expect to vote for Brown and explain what the confidence interval tells us.
5. To determine the average length of trout in a lake, researchers catch 20 fish and measure them. Describe the population and sample of this study.
6. To determine the average diameter of evergreen trees in a forested park, researchers randomly tag 45 specimens and measure their diameter. Describe the population and sample of this study. ADD
7. A college reports that the average age of their students is 28 years old. Is this a parameter or a statistic?
8. A local newspaper reports that among a sample of 250 subscribers, 45% are over the age of 50. Is this a parameter or a statistic? ADD
9. A recent survey reported that 64% of respondents were in favor of expanding the BIKETOWN bike share system to the greater Portland area. Is this a parameter or a statistic? ADD
10. Which sampling method is being described?
    1. In a study, the sample is chosen by separating all cars by size and selecting 10 of each size grouping.
    2. In a study, the sample is chosen by writing everyone’s name on a playing card, shuffling the deck, then choosing the top 20 cards.
    3. Every 4th person on the class roster was selected.
11. Which sampling method is being described?
    1. A sample was selected to contain 25 people aged 18-34 and 30 people aged 35-70.
    2. Viewers of a new show are asked to respond to a poll on the show’s website.
    3. To survey voters in a town, a polling company randomly selects 100 addresses from a database and interviews those residents.
12. Identify the most relevant source of bias in each situation.
    1. A survey asks the following: Should the mall prohibit loud and annoying rock music in clothing stores catering to teenagers?
    2. To determine opinions on voter support for a downtown renovation project, a surveyor randomly questions people working in downtown businesses.
    3. A survey asks people to report their actual income and the income they reported on their IRS tax form.
    4. A survey randomly calls people from the phone book and asks them to answer a long series of questions.
    5. The Beef Council releases a study stating that consuming red meat poses little cardiovascular risk.
    6. A poll asks, “Do you support a new transportation tax, or would you prefer to see our public transportation system fall apart?”
13. Identify the most relevant source of bias in each situation.
    1. A survey asks the following: Should the death penalty be permitted if innocent people might die?
    2. A study seeks to investigate whether a new pain medication is safe to market to the public. They test by randomly selecting 300 people who identify as men from a set of volunteers.
    3. A survey asks how many sexual partners a person has had in the last year.
    4. A radio station asks listeners to phone in their response to a daily poll.
    5. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score.
    6. High school students are asked if they have consumed alcohol in the last two weeks.
14. Identify whether each situation describes an observational study or an experiment.
    1. The temperature on randomly selected days throughout the year was measured.
    2. One group of students listened to music and another group did not while they took a test and their scores were recorded.
    3. The weights of 30 randomly selected people are measured.
15. Identify whether each situation describes an observational study or an experiment.
16. Subjects are asked to do 20 jumping jacks, and then their heart rates are measured.
17. Twenty coffee drinkers and twenty tea drinkers are given a concentration test.
18. The weights of potato chip bags are weighed on the production line before they are put into boxes.
19. A team of researchers is testing the effectiveness of a new vaccine for human papilloma virus (HPV). They randomly divide the subjects into two groups. Group 1 receives new HPV vaccine, and Group 2 receives the existing HPV vaccine. The patients in the study do not know which group they are in.
    1. Which is the treatment group?
    2. Which is the control group (if there is one)?
    3. Is this study blind, double-blind, or neither?
    4. Is this best described as an experiment, a controlled experiment, or a placebo-controlled experiment?
20. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment. Suppose that a new cancer treatment is under study. Of interest is the average length of time in months patients live once starting the treatment. Two researchers each follow a different set of 40 cancer patients throughout this new treatment.
21. What is the population of this study?
22. Would you expect the data from the two researchers to be identical? Why or why not?
23. If the first researcher collected their data by randomly selecting 10 nearby ZIP codes, then selecting 4 people from each, which sampling method did they use?
24. If the second researcher collected their data by choosing 40 patients they knew, what sampling method did they use? What concerns would you have about this data set, based upon the data collection method? MOVE LOCATION
25. For the clinical trials of a weight loss drug containing Garcinia Cambogia HIGHTLIG IN DOCUMENTthe subjects were randomly divided into two groups. The first received an inert pill along with an exercise and diet plan, while the second received the test medicine along with the same exercise and diet plan. The patients do not know which group they are in, nor do the fitness and nutrition advisors.
    1. Which is the treatment group?
    2. Which is the control group (if there is one)?
    3. Is this study blind, double-blind, or neither?
    4. Is this best described as an experiment, a controlled experiment, or a placebo-controlled experiment?
26. A study is conducted to determine whether people learn better with routine or crammed studying. Subjects volunteer from an introductory psychology class. At the beginning of the semester 12 subjects volunteer and are assigned to the routine studying group. At the end of the semester 12 subjects volunteer and are assigned to the crammed studying group.
    1. Identify the target population and the sample.
    2. Is this an observational study or an experiment?
    3. This study involves two kinds of non-random sampling: 1. Subjects are not randomly sampled from a specified population and 2. Subjects are not randomly assigned to groups. Which problem is more serious? What effect on the results does each have?
27. To test a new lie detector, two groups of subjects are given the new test. One group is asked to answer all the questions truthfully. The second group is asked to tell the truth on the first half of the questions and lie on the second half. The person administering the lie detector test does not know what group each subject is in. Does this experiment have a control group? Is it blind, double-blind, or neither? Explain.
28. A poll found that 30%, plus or minus 5%, of college freshmen prefer morning classes to afternoon classes.
    1. What is the margin of error?
    2. Write the survey results as a confidence interval.
    3. Explain what the confidence interval tells us about the percentage of college freshmen who prefer morning classes?
29. A poll found that 38% of U.S. employees are engaged at work, plus or minus 3.5%.
    1. What is the margin of error?
    2. Write the survey results as a confidence interval.
    3. Explain what the confidence interval tells us about the percentage of U.S. employees who are engaged at work.
30. A recent study reported a confidence interval of (24%, 36%) for the percentage of U.S. adults who plan to purchase an electric car in the next 5 years.
    1. What is the statistic from this study?
    2. What is the margin of error?

ADD

1. A recent study reported a confidence interval of (44%, 52%) for the percentage of two-year college students who are food insecure.
   1. What is the statistic from this study?
   2. What is the margin of error? ADD
2. A farmer believes that playing Barry Manilow songs to his peas will increase their yield. Describe a controlled experiment the farmer could use to test his theory. MOVE LOCATION
3. A sports psychologist believes that people are more likely to be extroverted as an adult if they played team sports as a child. Describe two possible studies to test this theory. Design one as an observational study and the other as an experiment. Which is more practical? MOVE LOCATION
4. Find a newspaper or magazine article, or the online equivalent, describing the results of a recent study (not a simple poll). Give a summary of the study’s findings, then analyze whether the article provided enough information to determine the validity of the conclusions. If not, produce a list of things that are missing from the article that would help you determine the validity of the study. Look for the things discussed in the text: population, sample, randomness, blind, control, margin of error, etc.

1. Use a polling website such as [www.pewresearch.org](http://www.pewresearch.org) or [www.gallup.com](http://www.gallup.com) and search for a poll that interests you. Find the result, the margin of error and confidence level for the poll and write the confidence interval.

### Exercises 3.1 Solutions

1. . A sample is a sub group of the population. A population is the entire group of subjects.
2. A Statistic is measurement obtained by taking a sample.

A parameter is a measurement obtained from the data of the entire population.

3. The intended population is all PCC students.

a. All PCC Students

b. 200 students form PCC Cascade campus

c. The collected data is not representative of all PCC students since it only includes responses from students at Cascade campus. This is an example of sampling bias.

4.

* 1. The intended population is all Washington County residences.
  2. 1200 homes in Washington County
  3. The collected data is likely representative since residences were selected at random from the entire county.

5.

a. The representatives in a state’s congress.

b.The population size is N=106

c.The sample size is n=28

* 1. The statistic is or 50%
  2. The confidence interval is (45%, 55%) and tells us that the true percentage of the state congress representatives in support of the new education (the parameter) likely lies between 45% and 55%.

6.

a. All registered voters in the city of Raleigh.

b. The population size is N=9500

c. The sample size is n=350.

d. The statistic is  or 32%

e. The confidence interval is (28.5%, 35.5%) and tells us that the true percentage of registered voters who will vote for Brown is likely to lie between 28.5% and 35.5%.

7. The population is all trout in the lake. The sample is the 20 that were caught.

8. The population is all trees in the park. The sample is the 45 that were tagged and measured.

9. Parameter

10. A sample was taken so it is a statistic

11. Statistics

12. a. Stratified

b. Simple Random Sample

c. Systematic

13.

a. Stratified

b. Volunteer

c. SRS

14.

a. Volunteer Bias

b. Sampling Bias

c. Response Bias

d. Non Response Bias

e. Response Bias

f. Loaded Questions

15.

a Loaded Questions

b. Volunteer Bias

c. Response Bias

d. Volunteer

e. response Bias

f .response bias or non-response bias

16.

a. Observational

b. Experimental

c. Observational

17.

a. Observational

b. Experimental

c. Observational

18.

a. Group 1

b. Group 2

c. blind because the patients in the study do not know.

19.

a. Cancer Patients

b. no because sampling has variability

c stratified

d. Convenient Sample ~ It does not represent the population.

20.

a. 2nd Group

b. Inert Pill Group

c. Double Blind because the patients and the advisors do not know who is in each group.

d. Controlled Experiment

21.

a. All Students

b. Experimental

c. It is only looking at one class and not all groups that are in the population so Subjects are not randomly sampled from a specified population.

22. The control group would be the group that were ask to tell the truth. It is a blind study because the person who is administering the lie detector test doesn’t know which group each person is in.

23.

a. .05 or 5%

b (25%, 35%)

c. I am confident that the percentage of college freshmen who prefer morning classes is between 25% to 35%

24.

a. 3.5 % or .035

b. (34.5, 41.5)

c. I am confident that the percentage of all U.S. Employees are engaged at work is between 34.5 % and 41.5%

25.

a. (24 + 36)/2 = 30%

b. 30 – 24 = 6. 6%

26.

a. 48%

b. 48 – 44 = 4%

27. Play Barry Manilow to half the crop and don’t play any music to the other half of the crop.