1.1 Section Exercises

Population and Sample

For each scenario, identify the population being studied and the sample chosen.

1. A national magazine wishes to determine America's favorite celebrities. A ballot is included in the November issue of the magazine. Readers are encouraged to mail in their ballots.
2. An education professor wants to gather information about parental involvement in early education for students attending a particular Ivy League university. She obtains a list of registered students from the registrar's office and randomly chooses 300 students to study.
3. A large discount store wants to determine the average income of its shoppers. A researcher chooses 100 shoppers at random between the hours of 1 and 5 p.m. on Thursday.
4. A hotel chain wants to build a new facility. The board of directors uses a list of the top 100 vacation spots in America and visits 20 of the cities on the list to determine their feasibility as the new hotel's location.

Population, Sample, Parameters, and Statistics

For each scenario, determine the population and sample, if given. Also, determine whether the highlighted value is a parameter or a statistic.

5. The manufacturer of Energy Bolt soda is testing for quality control. The company determines that 99.97% of all soda cans produced meet their quality standards.
6. Ms. Weigang wishes to determine the average ACT score for all students in her 11:00 algebra class. To do so, she anonymously collects information from each student. She calculates that the average ACT score for the class is 19.2.
7. The Nielsen Company wishes to determine which genre of program is streamed most often among Americans aged 18–25. They survey a group of 1045 adults aged 18–25 about what programs they stream on a regular basis. It is determined that 53% of survey respondents say that their favorite programs are comedies.
8. A research group wishes to know the average salary of professors at public universities in the United States. It is determined from public records that professors at public universities earn an average salary of $63,461 per year.
9. A horticulturist is testing to see whether a new fertilizer produces plants that are significantly taller than those grown with traditional fertilizers. His greenhouse contains 520 plants treated with the new fertilizer. He measures 60 plants and determines that their average height is 22.9 inches.
10. A study is conducted in order to determine the percentage of high school seniors in the Atlanta area who plan to major in business upon entering college. For the study, 230 seniors from Atlanta area high schools are randomly chosen and surveyed. Of these, 42% say that they intend to major in business upon entering college.
11. For her dissertation, Maria needs to estimate the average number of hours per week that children aged 10–12 spend in front of a screen. She randomly surveys 250 schoolchildren in her area who are between 10 and 12 years old and calculates an average of 17.4 hours of screen time per week.
12. For a news special, a reporter wishes to determine what percentage of adults in her viewing area are overweight. A random sample of 1067 adults from the area is chosen for the study. 40% of adults sampled are found to be overweight.

13. The governor of a Midwestern state wants to know his approval rating following a recent scandal. Using a telephone poll of registered voters in his state, 565 constituents are surveyed. The researchers determine that 37% of the governor's constituents that were surveyed currently approve of the job he is doing in office.

14. A large real estate firm wants to know the average price per square foot for condominiums in Okaloosa County, Florida. Based on real estate records for all condominiums in the county, the firm determines that condominiums in Okaloosa County sell for approximately $300 per square foot.

15. A consumer advocacy group wants to survey residents in the Northeast regarding hospital care. They mail out 10,500 surveys to randomly selected households in the Northeast. A total of 984 surveys are completed and returned. Out of this group, 64% say that their hospital care was above average. After analyzing all of the surveys, the consumer advocacy group determines that approximately 90% of people in the Northeast are satisfied with hospital care in their region.

16. The PTA of Brownsville is concerned about the number of hours that high school students spend each week using social networking websites. A group of 40 high school students is chosen at random and asked to take note of the amount of time spent using social networking websites during the following week. Out of these 40 students, 14 logged over 20 hours and another 9 logged between 15 and 20 hours on social networking sites that week. However, 18 said they were online on a near constant basis. That's a staggering 45% of teens surveyed.

Read the following fictitious studies. For each, determine the population, sample, parameter(s), and statistic(s).

17. Java Express is a company that sells high-end coffee appliances and gourmet coffee. Due to the current economic conditions, the company is looking into ways to change its strategy in order to appeal to more frugal consumers.

Before investing in costly setup expenses, Java Express needs to know the percentage of coffee consumers who would buy from a new line of less expensive products. To estimate this percentage, the company chooses 5 of its largest markets and surveys every registered customer from each of these chosen markets. A total of 6195 customers complete the survey.

The results show that 45% of those surveyed are very interested in the new line of products, 32% are somewhat interested, and the remaining 23% are not interested in the new products at all. Based on these results, the researchers estimate that approximately 77% of coffee consumers will be receptive to the new line of products. Java Express decides that it is worthwhile to introduce a new line of less expensive products to the market.

18. A study was conducted to investigate whether or not exercise is a factor in the overall outlook and positive spirit of Americans.

For the study, 400 Americans were selected at random from each of the following two categories: those who exercise on a regular basis and those who do not. For the purpose of the study, regular exercise was defined to be light to moderate activity at least three times per week. A total of 800 Americans completed the brief telephone survey. Of those who exercise regularly, 74% said that their overall outlook on life is positive. Of those who do not exercise regularly, 68% said that their overall outlook on life is positive.

Though not a definitive study, the researchers concluded that exercise does appear to influence the attitudes of Americans. In addition, the researchers estimated that approximately 71% of all Americans have a positive outlook. Further research would likely be conducted on the subject at a later time.

Descriptive vs. Inferential Statistics

Decide whether the following statements are examples of descriptive or inferential statistics.

19. Eighty-two percent of the employees from a small local company attended the annual company picnic.

20. Based on information from a recent survey, researchers estimate that the average American will spend $751 on gifts this Christmas.

21. The average age of entering freshmen at the University of Senatobia is 20.8 years old, based on information from the registrar's office.

22. Sixty-five percent of seniors at a local high school who are applying to college apply to at least one college out of state.

23. The average number of hours vacationers spend in national parks during the summer months is 4.5 hours, based on a survey of 1000 visitors in various national parks.

24. An outdoor sporting retailer predicts that the percentage of customers who sign up for guided adventures they offer will increase by 9% next year based on new predictive analytics.

For each scenario, answer the questions that follow.

25 Since 1878, the Bureau of Labor Statistics has calculated the unemployment rate by surveying a group of randomly selected Americans (currently around 60,000) in regard to their employment status. A person is considered unemployed if he or she does not have a job and has looked for a job in the previous four weeks. Recently the method for compiling the unemployment rate has been criticized for not giving an accurate estimate of the true percentage of jobless Americans.

a. What is the population of the study?

b. What groups of jobless Americans might be left out based on the way "unemployed" is defined?

c. The unemployment rate is often used for comparison. For example, in recent times economists compared the unemployment rates computed each month during the recession of 2009 and 2010 to those of the Great Depression, when unemployment peaked at around 25%. Could comparisons like this be made if the method used to collect data for the unemployment rate was changed?

26. Suppose you are given the task of determining whether it would be profitable to open a new ladies' health club in the Atlanta suburb of Marietta, Georgia. Which of the following options would be the best choice for the population of your study? Explain your answer.

a. All women in metro Atlanta who exercise regularly

b. All women in Marietta, Georgia

c. All women in Marietta, Georgia who exercise regularly

d. All residents of the metro Atlanta area

27. A car manufacturer is in the design process for a new sedan that it plans to launch. Before finalizing the body type and other details about the car, the company wants to know the preferences of their target market. The manufacturer would like its new car to be appealing to all age groups, but its particular focus is on drivers in their 20s and 30s.

a. The company wants to know the preferences of drivers in their "20s and 30s." How might the company better define the population of the study?

b. Imagine that you are the researcher. Brainstorm different ways that you might collect data and obtain a sample that represents the population well.

28. Suppose you are assigned the task of identifying the "average resident of your neighborhood".

a. What type of data should you collect from your neighbors?

b. Should you use a population or sample? Explain your answer.

c. Will your conclusions be descriptive or inferential?

c. Boiling points (on the Celsius scale) for various caramel candies

d. The top ten Spring Break destinations as ranked by USA Today

Solution

a. The amount of time it takes for each runner to run the race is *quantitative* since calculations performed on these data are meaningful. A finishing time is a measurement, therefore the data are *continuous*. Differences between finishing times are meaningful, and a time of zero represents the absence of racing. We could also say that Andrew finished the race in half of Peyton's time; thus, the data are at the *ratio* level of measurement.

b. Colors are labels, so these data are *qualitative*. Qualitative data are *neither* discrete nor continuous. There are many ways to order colors, such as alphabetically or based on the color spectrum. However, when discussing colors of crayons, order is not the primary factor, as opposed to data such as rankings, in which order is important. Therefore, the data are at the *nominal* level of measurement.

c. Calculations can be performed on boiling points because they are measurements, making these data *quantitative*. Temperatures are measurements, so the data are *continuous*. For the Celsius scale, a temperature of zero degrees is simply a placeholder and does not indicate the absence of heat. Therefore, data from the Celsius scale are always at the *interval* level of measurement.

d. Since the rankings cannot be meaningfully added or subtracted, the data must be *qualitative*. Qualitative data are *neither* discrete nor continuous. The rankings are in a specific order, so the data are at the *ordinal* level of measurement.

1.2 Section Exercises

Data Classification

Determine the following classifications for the given data sets.

a. Qualitative or quantitative
b. Discrete, continuous, or neither
c. Level of measurement

1. Prices of a particular pair of jeans at various department stores

2. Widths of the doors in a home

3. Low temperatures in degrees Fahrenheit across the state for one evening in March

4. Total dollar value of all items placed in each safe deposit box at a local bank

5. Yearly amounts of snowfall in Cleveland over 10 years

6. Heights of orchids on a windowsill

7. Amount of weight gained by each person in a group of college freshmen

8. Number of antique cars collected by each member of a car club

9. Number of six-foot wooden boards it takes to build any given desk

10. Bank account PIN numbers

11. Heights of men entering the armed forces

12. Sizes of T-shirts on sale

13. Temperatures in Kelvin of various sites on the planet Mars

14. Numbers of siblings that students in Ms. Pitcock's third grade class have

15. Jersey numbers of players on a lacrosse team

16. Types of pets reported in a recent survey

17. Positions in line at the checkout counter of a grocery store

18. Letter grades on students' English essays

19. Birth order of children in a family

20. Titles that precede people's names (Dr., Mr., Ms., and so forth.)

21. Number of students enrolled in each section of College Algebra at The Ohio State University

22. Average temperatures in degrees Celsius of the water in the Bahamas for each month in 2020

23. Birth years of members of your immediate family

24. Numbers of people per household reported on census forms

Respond thoughtfully to the following exercises.

25. Instead of classifying data into the groups qualitative and quantitative, some statisticians classify data as categorical or measurement data. **Categorical data** can be placed into categories, but no meaningful order can be assigned to the categories. **Measurement data** have numerical values assigned to them, and the data can be ordered according to those values.

a. Classify "genders of puppies adopted from an animal shelter" as categorical or measurement data.

b. Classify "high temperatures of Juneau, AK measured in degrees Fahrenheit" as categorical or measurement data.

c. What would be the problem with using this classification system for the data "T-shirt sizes listed as S, M, L, XL"?

26. Often, continuous data are measured in "discrete" units. In other words, the way that we express the units makes us think that we have discrete data because we are rounding the continuous values to whole number units. Give an example of a type of continuous data that might cause this confusion.

27. Explain why qualitative data are not further classified as discrete or continuous.

28. Discuss why it is important for researchers to know the classification of data they are interested in before they design a study.

protected file and only allowing authorized people to see the data; however, it is likely that the educator will know which student is associated with what data. Therefore, although the data are kept confidential, they are not collected anonymously.

Definition

An **Institutional Review Board (IRB)** is a group of people who review the design of a study to make sure that it is appropriate and that no unnecessary harm will come to the subjects involved.

Informed consent involves completely disclosing to participants the goals and procedures involved in a study and obtaining their agreement to participate.

Summary

In summary, a statistical study begins with a question to be answered. This question determines the population of the study and the variables of interest. Data are then collected by means of an observational study or an experiment, depending on whether the study hopes to determine a cause-and-effect relationship as established in Step 1. Once data are collected, they are organized through the use of tables, graphs, or numerical summaries. Lastly, the results are analyzed to answer the original question. The remainder of the book, beginning with Chapter 2, will be devoted to these last two steps of a study: organizing and analyzing data.

1.3 Section Exercises

Vocabulary

Determine if each statement is true or false. Explain why.

1. The first step in any statistical study is to state the question to be answered.
2. Data collection must be complete before variables are chosen so that the researcher can be sure he has the data needed to answer the question.
3. If a researcher wishes to determine a cause-and-effect relationship, she should use an observational study.
4. A random sample is the same thing as a simple random sample.
5. An Institutional Review Board will require that a researcher disclose to participants the goals and procedures involved in her study and obtain their agreement to participate.
6. The question in a statistical study dictates the population and variables.
7. Participants in an experiment should always be allowed to choose which group they are placed in so that they feel as comfortable as possible for the duration of the experiment.

Complete each statement.

8. An experiment in which both the participant and the person administering the treatment are unaware of whether the participant is in the treatment group or the control group is referred to as a __________ experiment.
9. A member of the population that is being studied in an experiment is called a __________.
10. The group in an experiment that receives a placebo is called the __________ group.
11. An experiment is a type of statistical study in which a __________ is applied to a group of the population.
12. A pill that looks like the treatment pill but has no active ingredient is called a __________.

13. An experiment in which only the participant is unaware of whether they are in the treatment group or the control group is referred to as a __________ experiment.

14. When a subject believes that he has recovered from an illness because he is taking a treatment drug, when in reality he is in the control group of an experiment, it is referred to as the __________.

15. A human subject in an experiment is referred to as a __________.

16. A researcher gives an active drug to the __________ group in an experiment.

Observational Studies vs. Experiments

Determine which type of study should be conducted: an observational study or an experiment.

17. A football coach wants to know the average weight of his offensive linemen.

18. A doctor wants to study the effect of ginseng on patients' memories.

19. A city planner wants to know the average number of vehicles parked in downtown parking lots on any given business day.

20. A cell phone company wants to know the average total length of time teenagers spend on the phone each day.

21. A dentist wants to look at the effects of a new dental material used for fillings.

22. A financial institution wants to know if customer behavior will change significantly when interacting with a new mobile app feature.

Sampling Methods

Identify the sampling method used in each observational study.

23. The FDA chooses 15 hospitals around the country at random. Every doctor in the chosen hospitals is asked to participate in the study.

24. Every 4th dorm room is selected for a survey regarding study hours and campus security.

25. A state politician wants to gauge public opinion in his area before deciding to run for reelection. For the study, 200 registered voters are chosen at random from each county in his district.

26. A computer program is used to randomly generate a list of student ID numbers in order to gather a group to give feedback about the Greek system on campus.

27. In order to complete a psychology project, you pass out surveys to the first 25 people you find in the student union.

28. A student asks all the people living on the 1st, 5th, and 8th floors of his dorm to answer a survey about dorm life on your campus.

29. A local church wanted to know the average age of its morning congregation, so they asked every 10th person leaving the service to put their age in a box.

30. Ten students from each of the 15 sections of College Algebra were asked about the quality of the textbook used in the course.

31. One thousand phone numbers were selected by a computer to be called for a telephone survey.

32. A local politician asks 20 people in his neighborhood what they think about the new school board proposal.

Cross-Sectional vs. Longitudinal Studies

Classify each scenario as either a cross-sectional study or a longitudinal study.

33. A social worker wants to determine the number of current foster children in her district who were placed in foster care due to neglect.

34. A budget-conscious person wishes to find which gas station in his area has the cheapest gas on his way to work one morning.

35. A local teachers' group creates charts to demonstrate that pay raises over the last five years have not kept up with inflation.
36. The child welfare office keeps track of how many reports of child abuse are received each month over a two-year period to determine if there are certain times of the year that generally have a higher report rate.
37. A human rights group gathers data from each state regarding the number of reported cases of human trafficking.
38. A patient with HIV gets her blood tested every three months to check her viral load to make sure that it is not increasing.

Meta-Analysis vs. Case Studies

Classify each scenario as either a meta-analysis or a case study.

39. A child prodigy's home life is examined in order to determine environmental factors that may have shaped him intellectually.
40. Studies performed on four different airlines are compared in order to determine which provides the best customer care.
41. For the purpose of studying sibling rivalry as affected by birth order, a typical American family is selected.
42. A medical researcher looks at multiple studies performed on a new drug in order to determine whether or not the drug is safe to put on the market.

Answer each question thoughtfully.

43. In the text we considered the research question: "Does taking 80 mg of aspirin each morning reduce the risk of heart attacks?" If we wanted to narrow our question to "Does taking 80 mg of aspirin each morning reduce the risk of heart attacks in African-American women over the age of 50?" how would this change the population of the study? If the results of the study showed that aspirin did indeed reduce the risk of heart attacks in this new population, would you be justified in recommending that your 52-year-old uncle begin taking aspirin daily? Does your answer change based on the ethnicity of your uncle? Explain.
44. Why not let a human choose the random sample? In reality, it is against our human nature to choose members of the population truly at random. To understand this phenomenon, take a moment to choose five random numbers between 1 and 100. Try to make sure they are truly random. Now, consider these questions. Are the numbers spaced out or grouped closely together? Are they all even, odd, or some of both? Did you have a reason for choosing any of the numbers? Did you alter any of your original responses and if so why? Did you repeat any of the numbers? In summary, would you say you were able to truly generate a set of random numbers?
45. In the text we considered an assembly line that has a mechanism with a defect so that it causes an error in every 4th label. If we sample every 4th label, we will get either a sample of labels that have no errors or a sample of labels that all have errors. Explain how to get a sample with no errors. Is it possible to get an unbiased sample by choosing every 5^{th} label from the assembly line? What about every 16^{th} part? Why or label not?
46. One type of convenience sampling is a **self-selected sample**. A self-selected sample is one in which the survey participants volunteer to be a part of the study rather than having been chosen by the researcher. The problem with a self-selected sample is that usually, only people with strong opinions will take the time to volunteer their time or information for the study. For instance, a popular American women's magazine wants to do research for a story regarding hospital care. The magazine lists a website in its June issue, inviting readers to log on and share stories about the care they received in the hospital. Describe the types of responses you could expect from people willing to log on and answer this survey. What would be the true population for the study described in this scenario? Can the results of this self-selected sample be generalized to describe hospital care for all patients in American hospitals?

Definition

Researcher bias occurs when a researcher influences the results of a study.

Response bias occurs when a researcher's behavior causes a participant to alter his or her response or when a participant gives an inaccurate response.

Participation bias occurs when there is a problem with the participation—or lack thereof—of those chosen for the study.

Nonresponse bias occurs when there is a lack of participation in a self-selected sample from certain segments of a population, when a person refuses to participate in a survey, or when a respondent omits questions when answering a survey.

Consider the *Conclusions*

In the end, it's the conclusions of the study that we should be most wise about. If we have a perfectly constructed, well-thought-out, unbiased study, we then want to make sure that our conclusions are actually correct. Do the data support the conclusion? Most often, the answer is yes, but why not make sure yourself? What if you read the following headline: "Most Americans Don't Like to See Pajamas Worn in the Grocery Store"? The results given to support the headline were: Of those who responded, 60% do not like to see it and 30% don't mind. What isn't stated is that 95% of the people surveyed couldn't be bothered to answer the question! (Not hard to imagine, is it?)

Here are some other things to consider: Do the results present the whole picture or just a part? Could there be other conclusions drawn? Could there be other reasons for the same conclusion drawn? Does the study have any practical applications? It's not that you should distrust all studies and results that you run across; it's that you should be an informed consumer of information and process critically what you read or hear through the news, journals, or any other outlet of information.

1.4 Section Exercises

Vocabulary

Complete each statement.

1. A __________ occurs when results from a study are tabulated incorrectly.
2. __________ occurs when the person administering the study influences the participants' responses.
3. __________ occurs when the results of a study tend to favor one outcome over another.
4. A participant who begins a study but fails to complete it is referred to as a __________.
5. A participant who does not fully comply with the instructions for a study is called a __________.
6. If participants are asked to answer survey questions in an uncomfortable setting, there is potential for __________ to occur.

Considering the Variables

Respond thoughtfully to the following exercises.

7. Consider the following science fair question: Does the quality of air get better with more rain? Name the variables indicated by this question, some different ways one might measure these variables, and some of the terms that need more precise definitions.
8. How would you measure the "Ten Best Colleges of the Midwest"? Name at least five distinct measurements.

9. How would you measure "quality of life"? What are some difficulties you might face?

10. Consider the magazine article that claims to know the "Best Vacation Spots in the World". Name at least 5 variables that could be possible measurements. What issues are there in this claim?

Considering the Setup

Describe as many potential sources of bias as you can for the following studies.

11. Star Crazy magazine wants to determine America's favorite celebrities. A survey is placed inside each subscription for readers to fill out and mail in to the company.

12. A struggling retailer wishes to improve sales. Store employees are given the task of polling local residents regarding their opinions about the store as well as asking them to make suggestions for improvement.

13. A major television network wants to know what TV shows people in one state are watching most often. For the study, televisions in 350 households in the four largest cities are monitored for one week.

14. American Star is a hit TV show in which Americans are asked to decide who has enough talent to be a star. Each week viewers call in and vote for their favorite performers. The contestant with the fewest votes is eliminated from the show.

Respond thoughtfully to the following exercises.

15. The governor of Tennessee would like to know if Tennesseans want to move to a state income tax. Using Sections 1.3 and 1.4, construct a plan to gather the information you would need to answer the governor and explain how you would carry out the plan.

16. A study regarding physical fitness is being conducted on campus. The title of the study is "Are We Fit?" All exercise science majors are required to be a part of the study. Using Sections 1.3 and 1.4, construct a plan to gather the information you would need to answer the question of "Are We Fit" and explain how you would carry out the plan.

$$\text{Upper Boundary of Class 1: } \frac{2.9+3.0}{2} = 2.95$$

$$\text{Midpoint of Class 1: } \frac{1.0+2.9}{2} = 1.95$$

Use the class width to find the other class boundaries and midpoints. To find the relative frequency, divide the frequency of each class by the number of data values, 18. The cumulative frequency is the rolling total of the class frequencies. The final frequency table with all of the values listed is shown below.

Numbers of Miles Professors Drive to Work Each Day

Class	Frequency	Class Boundaries	Midpoint	Relative Frequency	Cumulative Frequency
1.0–2.9	3	0.95–2.95	1.95	$\frac{3}{18}=\frac{1}{6}=0.1\bar{6}\approx 17\%$	3
3.0–4.9	3	2.95–4.95	3.95	$\frac{3}{18}=\frac{1}{6}=0.1\bar{6}\approx 17\%$	6
5.0–6.9	4	4.95–6.95	5.95	$\frac{4}{18}=\frac{2}{9}=0.\bar{2}\approx 22\%$	10
7.0–8.9	2	6.95–8.95	7.95	$\frac{2}{18}=\frac{1}{9}=0.\bar{1}\approx 11\%$	12
9.0–10.9	4	8.95–10.95	9.95	$\frac{4}{18}=\frac{2}{9}=0.\bar{2}\approx 22\%$	16
11.0–12.9	2	10.95–12.95	11.95	$\frac{2}{18}=\frac{1}{9}=0.\bar{1}\approx 11\%$	18

Using the relative frequencies of each class, we can see that the data is somewhat evenly spread through the classes since no single class has more than 22% of the data in it. In other words, the driving distance for professors is not a consistent distance. The cumulative frequencies show that more than half of the data is in the first 3 classes. In fact, 10/18 of the professors surveyed travel less than 7 miles each day.

Memory Booster

Relative frequency is a ratio that relates a class to the whole.

Cumulative frequency is a running total of the number of data values.

2.1 Section Exercises

Characteristics of Frequency Distributions

For Exercises 1–10, find the following for each frequency distribution.

a. Class width
b. Class boundaries for each class
c. Midpoint of each class
d. Relative frequency for each class
e. Cumulative frequency for each class

1.

Braking Times for Vehicles at 60 mph (in Minutes)

Class	Frequency
0.05–0.07	12
0.08–0.10	15
0.11–0.13	14
0.14–0.16	15
0.17–0.19	14

2.

Ages of Taste-Test Participants (in Years)	
Class	Frequency
15–19	7
20–24	8
25–29	10
30–34	2
35–39	3

3.

Age at Time of First Marriage (in Years)	
Class	Frequency
15–18	2
19–22	5
23–26	4
27–30	5
31–34	4

4.

Hourly Wage at First Job (in Dollars)	
Class	Frequency
7.50–8.49	12
8.50–9.49	50
9.50–10.49	48
10.50–11.49	45
11.50–12.49	34

5.

Cost of a 12 oz Soda (in Dollars)	
Class	Frequency
0.25–0.49	2
0.50–0.74	15
0.75–0.99	12
1.00–1.24	5
1.25–1.49	9

6.

Ages of Survey Participants (in Years)	
Class	Frequency
15–24	9
25–34	8
35–44	12
45–54	1
55–64	3

7.

Ages of First-Time Home Buyers (in Years)	
Class	**Frequency**
18–24	2
25–31	7
32–38	4
39–45	15
46–52	3

8.

Hourly Wages of Surveillance Operators (in Dollars)	
Class	**Frequency**
10.50–11.49	92
11.50–12.49	78
12.50–13.49	68
13.50–14.49	45
14.50–15.49	34

9.

Age at Time of First Car Purchase (in Years)	
Class	**Frequency**
16–19	12
20–23	8
24–27	15
28–31	12
32–35	9

10.

Grades on a Difficult Test	
Class	**Frequency**
A	2
B	5
C	7
D	13
F	10

Complete the frequency distribution that has been started for each set of data.

11. The following data describe the heights, in inches, of 30 volunteers for a bone density study.

72.8	71.2	70.3	73.4	72.6	74.1
70.9	71.6	72.1	74.6	75.0	72.0
69.1	69.5	72.6	72.4	73.6	75.1
71.8	71.6	71.9	70.9	70.2	69.3
72.1	72.3	72.5	73.4	74.0	75.0

Heights of Volunteers (in Inches)

Class	Frequency
69.0–69.9	3
70.0–70.9	
71.0–71.9	
72.0–72.9	
73.0–73.9	
74.0–74.9	
75.0–75.9	

12. The following data represent the number of exercises at the end of various sections in a traditional college algebra textbook.

145	137	138	112	137	100	78	127	97
70	143	133	150	124	115	110	45	141
119	92	84	94	105	71	95	117	104

Number of Section Exercises

Class	Frequency
40–59	
60–79	
80–99	
100–119	
120–139	
140–159	

13. At a state fair, one game involves guessing the number of marbles in a glass jar. The following data represent the guesses that people made during one hour at the state fair.

1234	1645	1469	1467	1549	1348	1671	1300	1200	1199
1621	1547	1501	1410	1487	1299	1500	1688	1301	1399

Number of Marbles in the Jar

Class	Frequency
1100–1199	
1200–	
1300–1399	4
	4
1500–1599	
1600–	4

14. The following data represent the amounts of pocket change, in dollars, in the pockets of a sample of business professionals in an office building.

0.23	0.52	0.76	0.79	0.8	0.21	0.13
1.05	1.24	1.15	1.10	0.98	0.28	0.64
1.34	0.38	0.31	0.42	0.41	0.24	1.42

Amount of Pocket Change (in Dollars)

Class	Frequency
–0.24	4
0.25–0.49	
0.50–	2
0.75–0.99	
–1.24	4
1.25–1.49	

Create a frequency distribution with the indicated number of classes for each set of data. Include the frequency, class boundaries, midpoint, relative frequency, and cumulative frequency of each class.

15. The following data represent the numbers of curl-ups completed in 60 seconds for a group of 16 eight-year-old children. Use six classes that have a class width of 5. Begin with a lower class limit of 15.

31	34	41	36	27	29	18	33
31	28	34	22	26	28	36	42

16. The following data represent times in minutes for completing a one mile run/walk from a group of 24 teenagers. Use six classes that have a class width of 2.00. Begin with a lower class limit of 6.00.

15.23	13.52	11.35	11.15	12.20	9.90	10.37	14.05
10.02	17.35	8.33	8.05	9.87	9.28	10.62	6.65
9.55	10.23	13.93	10.97	9.75	12.85	12.82	10.93

17. The following data represent the caloric intakes in one day for a group of 15 people between the ages of 20 and 39. Use five classes that have a class width of 400. Begin with a lower class limit of 1800.

2700	2200	2500	2800	2600
3000	2600	2200	3100	2800
1800	3500	2500	3000	2900

18. The following data represent the precipitation totals in inches for the month of September in 21 different towns in Alaska. Use six classes that have a class width of 3.50. Begin with a lower class limit of 0.00.

2.7	1.72	1.39	6.88	2.59	2.04	2.43
9.28	1.06	3.29	1.57	4.23	0.95	8.37
0.6	4.41	6.73	1.92	2.28	2.74	18.65

19. The following data represent the numbers of days absent from school in one school year for each of the 24 students in Ms. Jinn's fourth grade class. Use six classes that have a class width of 5 days. Begin with a lower class limit of 0.

17	8	12	3	0	5	13	12
25	10	6	8	11	0	1	4
19	21	22	9	16	9	3	2

20. The following data represent the weights in pounds of 24 collegiate offensive linemen in a particular state. Use six classes that have a class width of 25 pounds. Begin with a lower class limit of 175.

195	210	255	267	231	229	301	199
178	281	245	256	278	205	217	223
279	196	235	248	262	291	302	189

Use the given frequency distribution to answer the questions.

21.

Amounts of Weight (in Pounds) Lost Following a Low-Carbohydrate Diet

Weight Lost	Number of Women
1–5	4
6–10	8
11–15	11
16–20	3
21–25	0
26–30	1

a. How many participants were involved in the study?

b. How much weight did most participants lose?

c. What percentage of participants lost more than 15 pounds?

d. How long did the study last?

e. What is the lower class limit of the 4th class?

f. What is the midpoint for the 1st class?

g. If you were writing an article about this study, what would be your headline? Justify your response.

22.

Grams of Sugar per Serving in Children's Breakfast Cereal

Class	Frequency
10.00–11.99	21
12.00–13.99	19
14.00–15.99	11
16.00–17.99	18
18.00–19.99	18

a. How big was the original data set?

b. What percentage of the data fell in the 3rd class?

c. Estimate the average value of the data in the highest class.

d. Based on this distribution, do you believe that children's breakfast cereal contains too much sugar? Justify your response.

23. The frequency distribution shows the ages of students volunteering at a local animal shelter.

Ages of Student Volunteers (in Years)	
Class	**Frequency**
10–12	5
13–15	9
16–18	2
19–21	11
22–24	6

a. What is the relative frequency for the 4^{th} class?

b. What is the cumulative frequency for ages 18 and under?

c. What is the upper class boundary for the 2^{nd} class?

d. What is the class width?

e. What percentage of the student volunteers were between the ages 19 and 24?

f. How many students were surveyed?

g. What was the youngest age of the students surveyed?

24.

Distribution of Attitudes Toward College Success

Scale	Freshmen	Sophomores	Juniors	Seniors	Graduate Students
1.0–1.9	11	10	2	9	19
2.0–2.9	3	2	5	5	1
3.0–3.9	15	2	3	11	1
4.0–4.9	11	15	1	1	10
5.0–5.9	6	7	8	11	12
6.0–6.9	16	9	17	16	16
7.0–7.9	5	19	1	17	5
8.0–8.9	3	5	19	14	17
9.0–9.9	5	11	1	14	10
10.0–10.9	9	3	4	15	7

a. How many students were surveyed for the study?

b. What percentage of sophomores gave a response smaller than 5?

c. How many people gave a response in the 9^{th} class?

d. What percentage of respondents were graduate students?

e. Does this table easily display patterns in the data? In what ways might this data be better displayed?

25. Below is an example of a frequency distribution produced by a statistical software package called SPSS. Use it to answer the following questions.

Statistics

Respondent's Ethnicity

N	Valid	936
	Missing	4

Respondent's Ethnicity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	White	476	50.6	50.9	50.9
	Black or African American	74	7.9	7.9	58.8
	Asian	123	13.1	13.1	71.9
	Hispanic	198	21.1	21.2	93.1
	American Indian or Alaska Native	4	.4	.4	93.5
	Race and Ethnicity Unknown	61	6.5	6.5	100.0
	Total	936	99.6	100.0	
Missing	System	4	.4		
Total		940	100.0		

a. How many respondents gave their ethnicities?

b. What percentage of respondents were American Indian or Alaska Native?

c. What percentage of respondents who gave their ethnicities were White?

d. If you were to describe the racial diversity, or lack thereof, for this population, what would you say?

2.2 Section Exercises

Bar Graphs

Construct a bar graph for the given data. If it is appropriate, make the bar graph a Pareto chart.

1.

Math Grades on Test 1

Grade	Number of Students
A	30
B	56
C	47
D or F	12

2.

First United States Presidential Election, 1789

	Number of Electoral Votes
George Washington	69
John Adams	34
Another Candidate	26
Not Voted	44

Source: Encyclopedia Britannica. "United States Presidential Election Results." https://www.britannica.com/topic/United-States-Presidential-Election-Results-1788863 (30 Jan. 2019).

3.

Value-Added Tax

Country	Tax Percentage
Spain	16%
Canada	7%
Norway	25%
Japan	5%
United Kingdom	17.5%

Source: Wikipedia contributors. "Value added tax." *Wikipedia, The Free Encyclopedia.* 24 Jan. 2012. http://en.wikipedia.org/wiki/Value_added_tax (24 Jan. 2012).

4.

Cost of Attendance

	College Pricing
Public 2-year	$17,930
Public 4-year in state	$25,890
Public 4-year Out of state	$41,950
Private 4-year Nonprofit On-campus	$52,500

Source: Trends in Higher Education. College Board. "Average Estimated Undergraduate Budgets, 2018-19." https://trends.collegeboard.org/college-pricing/figures-tables/average-estimated undergraduate-budgets-2018-19 (18 July 2019).

Construct a side-by-side bar graph and a stacked bar graph for the given data. Then answer the questions.

5.

Math Enrollment

Math Class	Freshmen	Sophomores
Statistics	147	45
Algebra	160	73
Calculus	23	92
Quantitative Reasoning	12	120

a. Which course has the most students enrolled? Which graph did you use to answer this question?

b. Which course has the most sophomores enrolled? Which graph did you use to answer this question?

c. Which course has the most freshmen enrolled? Which graph did you use to answer this question?

d. Which course has the least students enrolled? Which graph did you use to answer this question?

6.

Pets Seen by the Vet in One Month

Type of Animal	Number Seen by Dr. Warren	Number Seen by Dr. Campbell
Cats	47	59
Dogs	56	37
Reptiles	13	6
Birds	28	30

a. Which veterinarian saw more dogs this month? Which graph did you use to answer this question?

b. Which type of animal was seen the least this month? Which graph did you use to answer this question?

c. Which veterinarian saw more cats this month? Which graph did you use to answer this question?

d. Which animal did Dr. Warren see the least this month? Which graph did you use to answer this question?

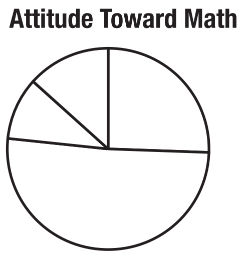
Pie Charts

Based on the data provided, label the pie charts with which section of the pie corresponds to which category. Provide the category label and the percentage of the pie associated with that category.

7.

Attitude Toward Math

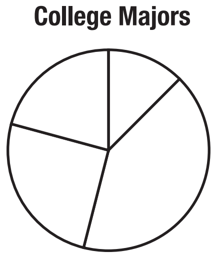
Attitude	Number of Students
Love	23
Like	46
Indifferent	9
Hate	12



8.

College Majors

Major	Number of Students
English	171
Business	569
Education	346
Science	285



Histograms

Do the following for each frequency distribution.

a. Construct a histogram.
b. Calculate the relative frequency for each class.
c. Construct a relative frequency histogram.

9.

Ages of Taste-Test Participants (in Years)

Class	Frequency
15–19	7
20–24	8
25–29	10
30–34	2
35–39	3

10.

Braking Times for Vehicles at 60 mph (in Minutes)	
Class	**Frequency**
0.05–0.07	12
0.08–0.10	15
0.11–0.13	14
0.14–0.16	15
0.17–0.19	14

11.

Age at Time of First Marriage (in Years)	
Class	**Frequency**
15–18	2
19–22	5
23–26	4
27–30	5
31–34	4

12.

Hourly Wage at First Job (in Dollars)	
Class	**Frequency**
7.50–8.49	12
8.50–9.49	50
9.50–10.49	48
10.50–11.49	45
11.50–12.49	34

Stem-and-Leaf Plots

Create a stem-and-leaf plot for the given data.

13. The following data represent the numbers of sit-ups completed in 60 seconds for a group of sixteen children What is the range of values from the stem-and-leaf plot that has the largest number of children completing sit-ups in 60 seconds?

31	34	41	36	27	29	18	33
31	28	34	22	26	28	36	42

14. The following data represent the caloric intakes in one day for a group of fifteen men between the ages of 20 and 39. Estimated Energy Requirements (EER) from the Institute of Medicine recommend a caloric intake for men between 2000 and 3000 calories per day. The following data represent the caloric intakes in one day for a group of fifteen men between the ages of 20 and 39. Based on the stem and leaf plot you created, are the majority of men in this group following the Institute of Medicine recommendation?

2700	2200	2500	2800	2600
3000	2600	2200	3100	2800
1800	3500	2500	3000	2900

Create an ordered stem-and-leaf plot for the given data.

15. The following data represent times in minutes for completing a one mile run from a group of twenty-four teenagers.

12.4	12.3	11.1	11.9	12.1	9.5	11.6	10.8
10.2	9.3	10.1	11.2	8.2	9.3	9.5	12.5
9.4	9.7	10.7	10.9	9.3	10.4	12.9	10.6

16. The following data represent the precipitation totals in inches for the month of September in 21 different towns in Alaska.

2.73	2.81	2.54	2.59	2.70	2.88	2.64
2.55	2.86	2.68	2.77	2.61	2.56	2.62
2.78	2.64	2.50	2.67	2.89	2.74	2.81

17. The following data represent the saline concentrations (in terms of specific gravity) in a saltwater aquarium on various days.

1.022	1.021	1.022	1.023	1.019	1.021	1.022	1.017	1.024
1.021	1.022	1.022	1.023	1.020	1.019	1.023	1.025	1.018

18. The following data represent the numbers of grams of fat per serving for a representative sample of various foods found in someone's kitchen pantry. Do you believe that this person is on a low-fat diet? Explain your reasoning using the data as your evidence.

0	2	0	6	8	1	10	2
3	14	4	21	13	7	9	17

Dot Plots

Create a dot plot for the given data.

19. The following data represent the numbers of visitors, in thousands, which various online clothing retailers had on their websites during the month of February.

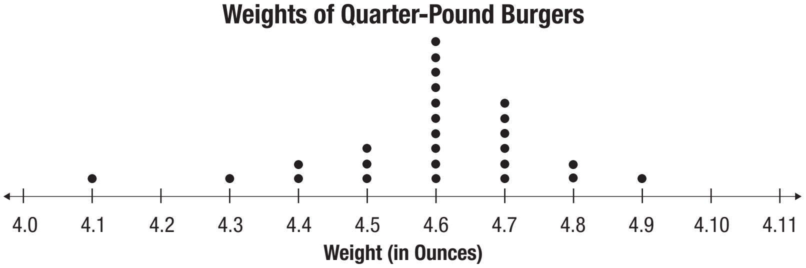
13	11	19	15	11	11	17
3	10	14	14	9	12	15
12	16	12	13	15	12	10

20. The following data represent the numbers of plastic shopping bags that customers used in a single shopping trip at a local grocery store one afternoon.

3	2	5	5	7	7	3	10	2	3	9	2	6	3	4
2	1	3	5	1	5	0	4	2	8	2	6	2	1	6
2	8	10	4	5	3	8	0	3	2	2	7	11	1	3
4	4	3	10	4	2	6	14	5	3	4	4	10	3	1
3	19	1	15	2	4	1	1	4	4	7	3	2	1	1
4	8	8	6	8	3	1	0	30	1	4	1	11	5	6
19	1	9	0	5	1	5	7	5	1	9	2	10	5	8
3	3	4	4	5	2	2	4	9	1	2	2	0	2	7
7	13	3	1	7										

Use the given dot plot to answer the questions.

21. A local fast food restaurant collected the following data while checking the weights of their quarter-pound burgers, which should be approximately 4 ounces.



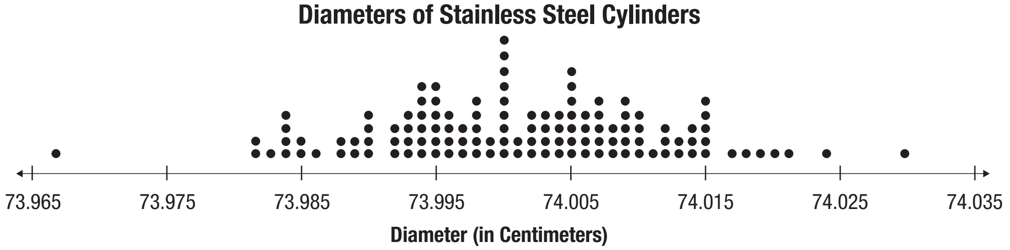
a. How many burgers were checked for weight?

b. What was the most common weight?

c. If you were describing the average weight for this sample rounded to the nearest whole ounce, would you say 4 ounces or 5 ounces? Explain your reasoning.

d. If you were a customer at this restaurant, would you feel satisfied that you were always getting your money's worth for a quarter-pound burger? Why or why not?

22. A quality control manager for a manufacturer of household products collected the following data while measuring the diameters of stainless steel cylinders that the company uses to make large trash cans.



a. Which diameter is the most common?

b. What is the largest diameter in the data?

c. Discuss the pros and cons of displaying this data set in this type of dot plot, keeping in mind any difficulties you had in answering part b.

Interpreting Graphs

Use the given graphs to answer the questions.

23. The following heat map shows the percentage of students enrolled in the Fall of each year who received a grade of D, F, W (withdraw), or I (incomplete). Use this map to answer the following questions:

DFWI Rates	2012	2013	2014	2015	2016	2017	2018
American History I	25%	23%	22%	23%	20%	19%	19%
College Algebra	44%	42%	40%	35%	37%	36%	34%
English Composition I	35%	35%	36%	38%	37%	39%	38%
Psychology I	28%	30%	35%	36%	30%	27%	24%
Principles of Biology	31%	29%	30%	29%	32%	33%	31%
Music Appreciation	10%	12%	14%	13%	17%	20%	22%

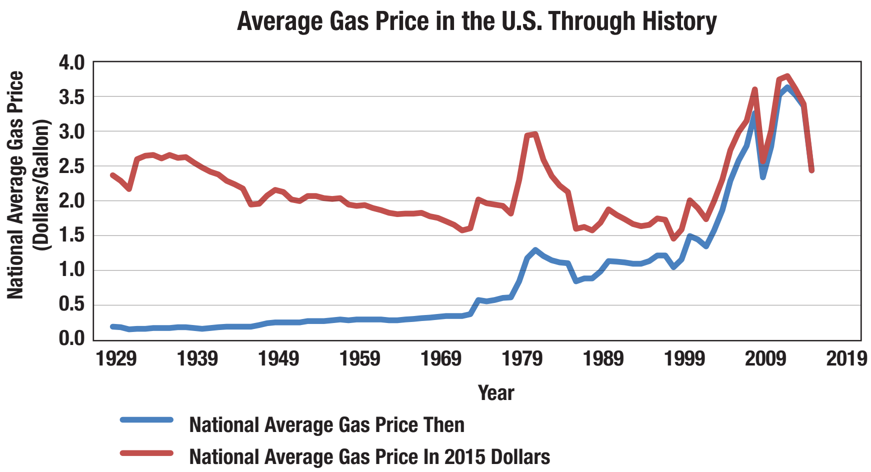
a. Which course had the worst DFWI rates?

b. Which course had the best DFWI rates?

c. Which course showed the most inconsistency in its DFWI rate?

d. Which year(s) had the most classes with poor (dark red or orange) DFWI rates?

24. Below is a line graph depicting the average national gas prices in the United States in that year's dollars ("then") and in 2015 dollars.



Source: Gringer, Bonnie. "Gas Prices Through History." https://www.titlemax.com/discovery-center/planes-trains-and-automobiles/average-gas-prices-through-history/ (30 Jan. 2019).

a. Approximately, what was the highest average price for gasoline? When did it occur? (Use the "then" values.)

b. Approximately, what was the lowest average price for gasoline? When did it occur? (Use the "then" values.)

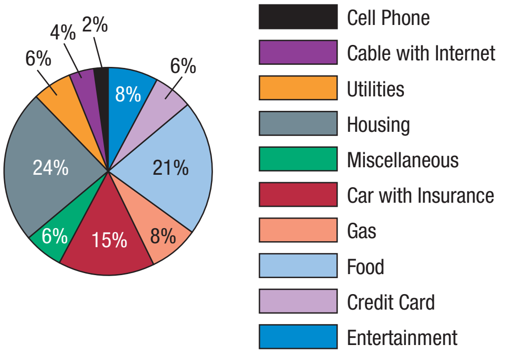
c. Approximately, what was the highest average price for gasoline? When did it occur? (Use the 2015 dollar values.)

d. Approximately, what was the lowest average price for gasoline? When did it occur? (Use the 2015 dollar values.)

e. How does the trend in the price of gas in that year's dollars compare to the trend in the price of gas adjusted to 2015 dollar values?

25. The following pie chart depicts Harrison's monthly budget.

Harrison's Monthly Budget



a. What does Harrison spend the most money on each month?

b. What does Harrison spend the least on each month?

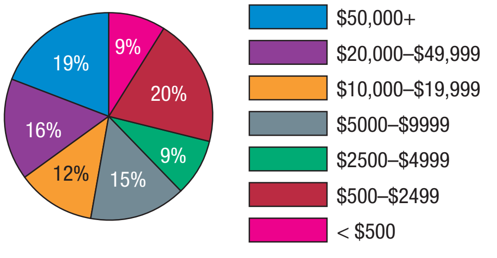
c. What percentage of Harrison's budget is spent on household bills? (Housing, Utilities, Cable with Internet, and Cell Phone)

d. What percentage of his monthly budget does Harrison spend on his car? (Car with Insurance and Gas)

e. Disposable income is the money remaining after all essentials have been paid. What percentage of Harrison's budget is disposable income? (Entertainment, Cable with Internet, Miscellaneous, Cell Phone)

26. Consider the pie chart below.

Initial Investments: Amount by Percentage of Respondents



a. What percentage of respondents initially invested $50,000 or more?

b. What percentage of respondents initially invested less than $10,000?

c. What percentage of respondents initially invested between $5000 and $19,999?

d. What percentage of respondents initially invested $10,000 or more?

27. Below is a stem-and-leaf plot of ACT math scores for a group of college music students.

ACT Math Scores of Students in a College Music Class

Stem	Leaves
1	2 3 3 4
1	5 7 8 8 8 8 9 9
2	0 0 0 1 1 2 2 2 3
2	5 5 5 6 7
3	2 3 4
3	

Key: 1 | 2 = 12

a. What was the lowest math score a student in this class received on the ACT?

b. What was the highest math score a student in this class received on the ACT?

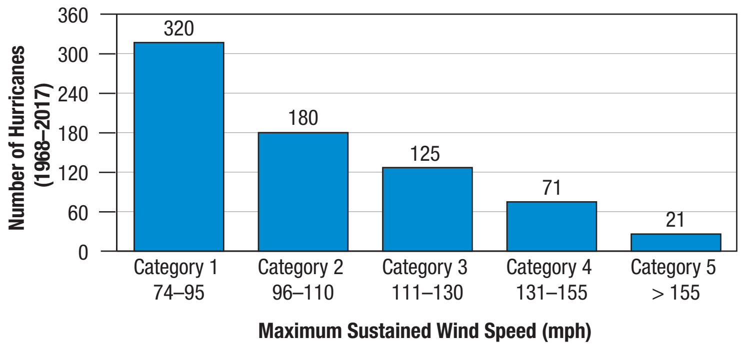
c. Which math score occurred most often?

d. How many students are represented by this information?

28. Meteorologists categorize hurricanes according to their maximum sustained wind speed using the Saffir-Simpson scale. This scale divides hurricanes into five categories, with Category 1 hurricanes having maximum sustained winds between 74 and 95 miles per hour (mph). The bar graph below depicts the number of hurricanes along the Atlantic coast of the United States from 1968 to 2017. Each bar in the graph represents one of the categories used to classify hurricanes.

Source: Wikipedia. Atlantic hurricane season. https://en.wikipedia.org/wiki/Atlantic_hurricane_season#Number_of_storms_of_each_strength_since_the_satellite_era (18 July 2019).

Hurricanes along the US Atlantic Coast



a. How many Category 3 hurricanes, with wind speeds of 111 to 130 mph, have hit the US Atlantic coast between 1968 and 2017?

b. A major hurricane is considered a hurricane in categories 3, 4, or 5 (that is, wind speeds greater than 111 mph). How many major hurricanes have hit the US Atlantic coast between 1968 and 2017?

c. What is the total number of hurricanes to hit the US Atlantic coast between 1968 and 2017?

29. Consider the following stem-and-leaf plot from a study published in an academic journal, which displays the measurements of the anterior radius of the left otolith for female stone flounder obtained from commercial capture in Tokyo Bay. The numbers to the left of the stems in this graph represent the cumulative frequencies of the data to the closest end. The middle stem is denoted by parentheses around the frequency of data in that category.

Source: Salgado Ugarte, Isaías H. "Exploratory Analysis of Some Measures of the Asymmetric Otoliths of Stone Flounder *Kareius bicoloratus* (Pisces: Pleuronectidae) in Tokyo Bay." *Anales del Instituto de Ciencias del Mar y Limnología*, Vol. 18, 1991 No. 2.

Anterior Radius of the Left Otolith for Female Stone Flounder

1	18	2
	19	
5	20	0 5 6 8
	21	
8	22	1 3 8
11	23	2 6 7
15	24	2 3 4 6
21	25	0 4 4 5 6 7
27	26	0 2 2 7 8 8
32	27	1 2 4 8 8
36	28	3 6 8 9
(7)	29	0 4 7 7 9 9 9
42	30	0 1 2 2 3 4 5 6 6 8 8 8 9
29	31	0 0 2 4 5 5 7 8 9
20	32	0 0 2 4 5 5 7
13	33	3 3 5 6 6 9
7	34	0 3
5	35	3 6
3	36	1 9
1	37	6

Unit = 0.01, N = 85 10 | 2 represents 1.02 mm

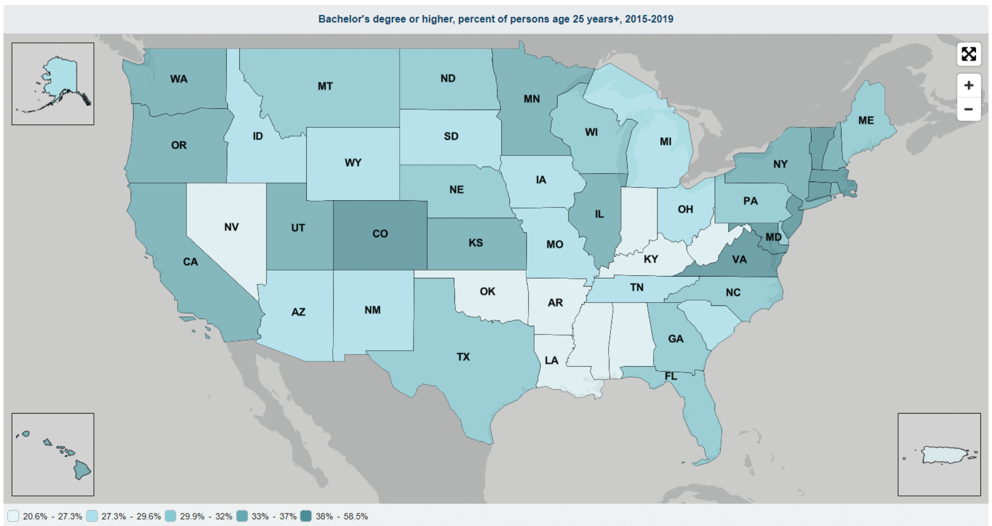
a. How many fish were in the study?

b. What were the smallest and largest measurements recorded?

c. Which length(s) appeared most often?

d. Did more fish have an otolith radius shorter or longer than the middle length category?

30. The following map depicts the percentage of each state population with a bachelor's degree or higher.



Source: US Census Bureau. "QuickFacts United States." 01 July 2018. https://www.census.gov/quickfacts/fact/map/US/EDU685217 (27 November 2019).

a. What area of the country has the largest percentage of bachelor's degrees or higher?

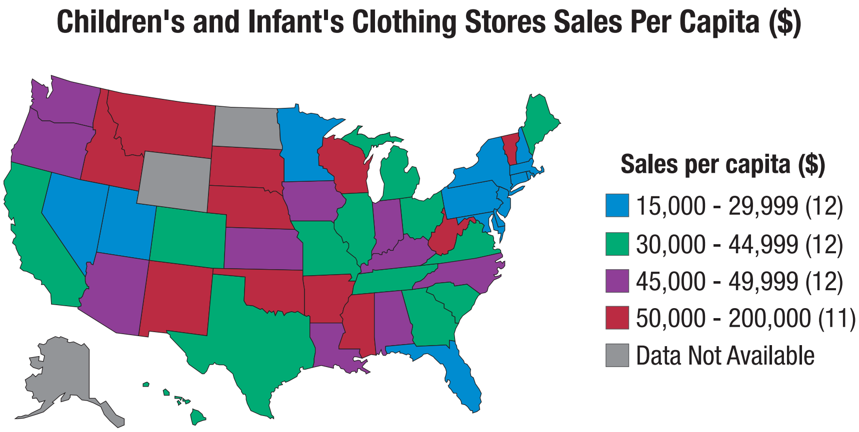
b. Look at the state you live in, what can you say about the percentage of bachelor's degrees or higher in your state compared to the rest of the United States?

31. For a school project, you need to graphically display data that you have collected. Determine the best way to display the following data. Explain your choice.

Number of Social Network Friends for Adults Over the Age of 40

213	583	114	317	80	497	524	352	126	627	250	710
528	438	347	944	721	753	302	349	101	812	74	849
314	841	411	160	174	439	435	569	323	444	529	556

32. The following map depicts the sales per capita for the children's clothing industry.



a. Of the states Texas, Oklahoma, and Louisiana, which state has the largest number of sales per capita in the children's clothing industry?

b. What is the range of sales per capita for California?

c. If you want to open a new children's clothing store either in the state you live or a state that shares a border with your state, where would you open? Explain your choice.

d. Notice that some of the wealthiest states (New York, Texas, California) have lower sales per capita. Conversely, some of the less wealthy states (Mississippi, West Virginia, Idaho) are among the highest sales per capita. Is this surprising to you? To what could you attribute these results?

Determine which type of graph would most clearly depict the data described.

33. The number of tickets sold at one theater over the course of a year

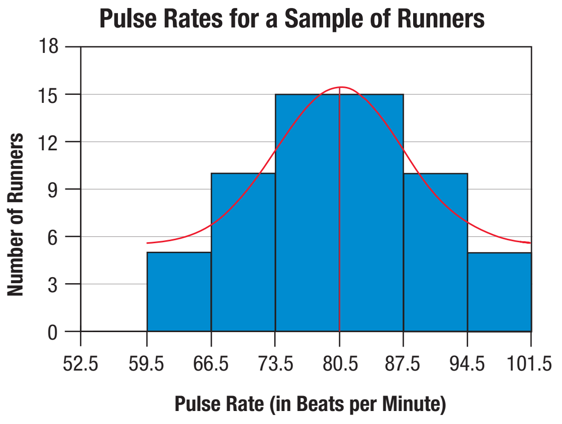
34. The number of tickets sold at one theater for each movie showing during one week

35. The number of tickets sold at one theater for each movie this week, specifically comparing the movie choices of patrons aged 18–35 to the choices of patrons aged 36 and older

36. Ticket prices for all theaters across the country

Solution

Notice that if we draw a smooth curve skimming the top of the histogram, we begin to see a curve similar to the shape of the symmetric curve. To be symmetric, the left and right sides of the graph should be close to mirror images. Drawing a line down the center of the graph, we can see that both sides of the graph are indeed mirror images of each other.



Thus, this histogram has a symmetric shape.

Summary

When analyzing a graph, there are several things you should consider. First, look at the title, the labels on the axes, the source, and how recently the data was published. Next, consider the overall picture that the graph is depicting. What is the graph trying to show? Is the type of graph that is used appropriate for the data being displayed? Lastly, consider the shape of the graph. Are there any outliers? These considerations should help you come to a clear conclusion regarding the meaning of the graph.

2.3 Section Exercises

Time-Series vs. Cross-Sectional Graph

For each graph described below, decide whether it is a time-series or a cross-sectional graph.

1. A map of Louisiana has the individual parishes shaded to represent the average income level for each parish.

2. A line graph depicts the change in the minimum wage since 1965.

3. A bar graph displays the average SAT scores each year for the last five years.

4. A stem-and-leaf plot displays the number of home runs hit by each player on the Yankees ball club.

Analyzing Graphs

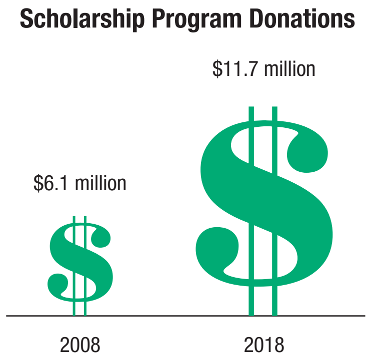
Use the given graphs to answer the questions.

5. Consider the following pictograph used to display the increase in funds donated to one university's scholarship program.

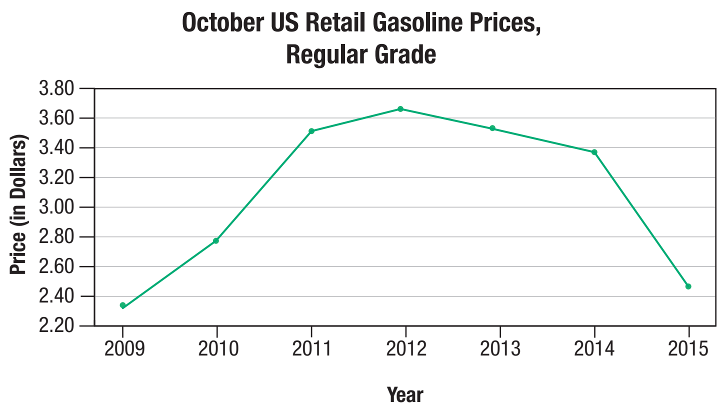
a. By what percentage did funds to the university scholarship program increase between 2008 and 2018?

b. Does the graph shown accurately depict the change in scholarship funds? Explain your answer.

c. What changes could be made to better display the given information?

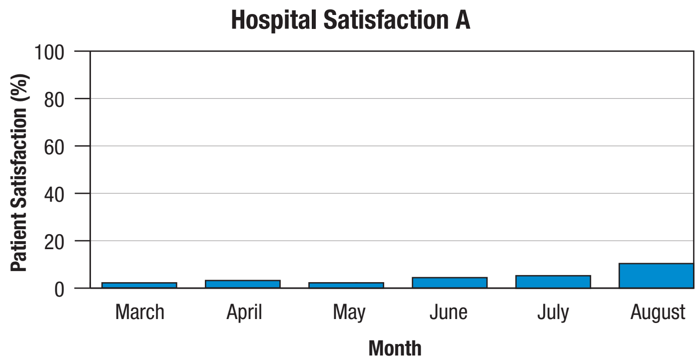


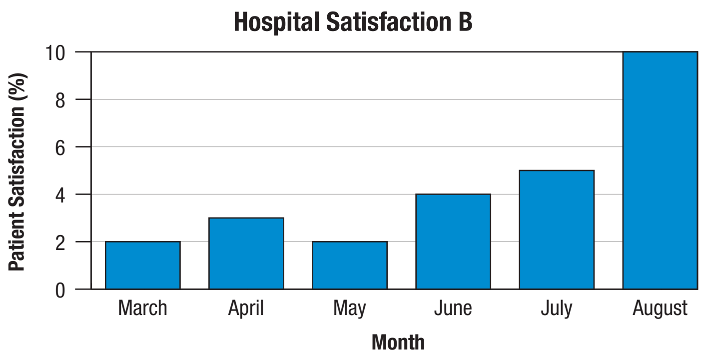
6. Does the scale on the following graph depict the situation accurately? Why or why not?



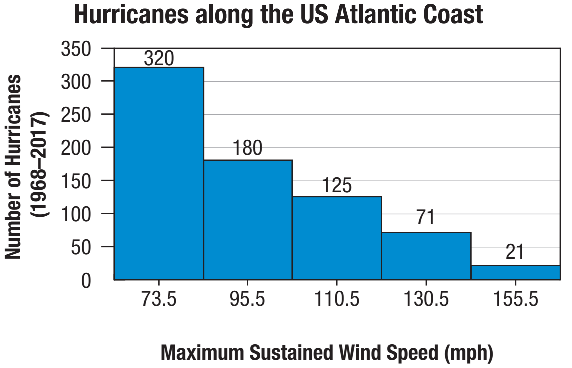
Source: Gringer, Bonnie. "Gas Prices Through History." https://www.titlemax.com/discovery-center/planes-trains-and-automobiles/average-gas-prices-through-history/ (24 July 2019).

7. Look at the two graphs shown below depicting the *same* data on people's overall satisfaction level with the care they received at their local hospital. Which of these two graphs gives a more accurate picture of hospital satisfaction? Has hospital satisfaction increased? Are people satisfied with the care at their local hospital according to these graphs? How do you know?



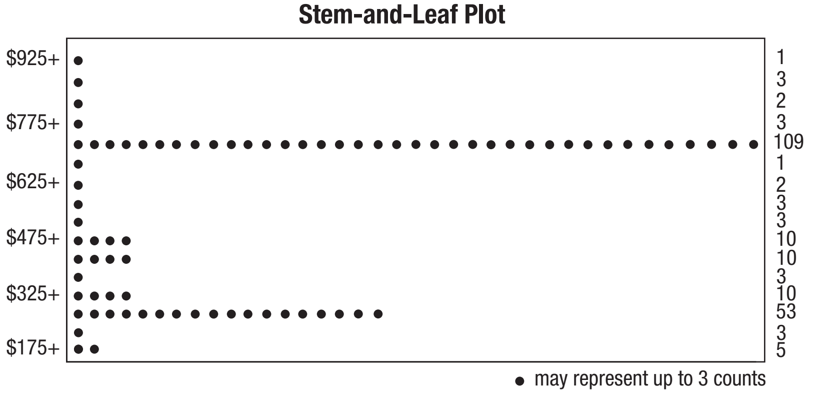


8. What errors occur in the following histogram?



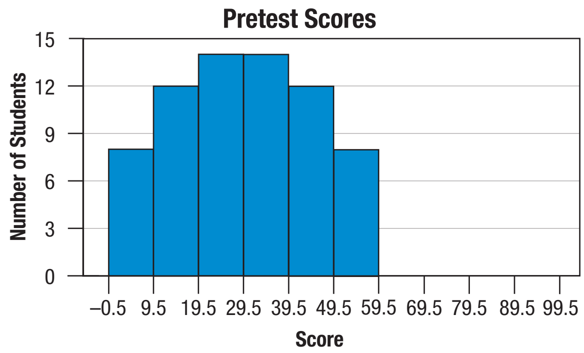
Source: Wikipedia. Atlantic hurricane season. https://en.wikipedia.org/wiki/Atlantic_hurricane_season#Number_of_storms_of_each_strength_since_the_satellite_era (18 July 2019).

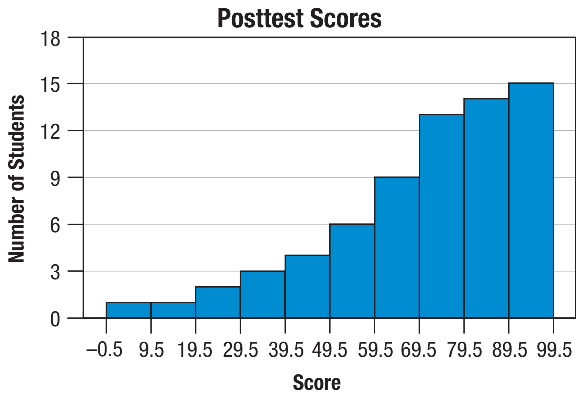
9. Consider the following excerpt from an online publication. Is the graph correctly labeled? If not, identify the corrections needed.



Source: NHHealthCost.org. "Health Costs for Consumers - Methodology." 16 Feb. 2007. http://www.nhhealthcost.org/method.aspx (24 Jan. 2012).

10. A pretest and a posttest were administered to one class at the beginning and end of two weeks of classes. The scores for each of the tests are shown graphically below.





a. Identify the general shape of the distribution of pretest scores.

b. Identify the general shape of the distribution of posttest scores.

c. Why do you think that the pretest scores are distributed the way they are?

d. Why do you think that the posttest scores are distributed the way they are?

Shapes of Graphs

For each set of data described below, discuss the most likely shape of its distribution.

11. The weights of the defensive linemen on football teams in the Big Ten Conference.

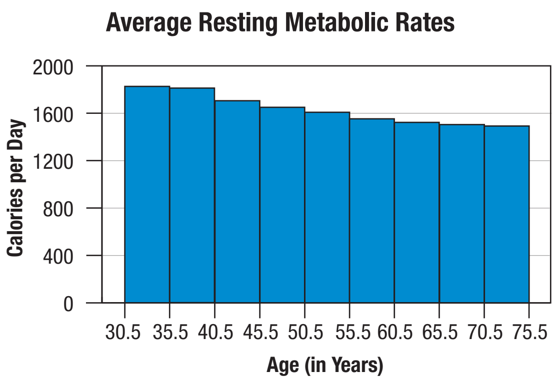
12. The lengths of the pregnancies of a group of gorillas being studied in the wild.

13. The last four digits used to generate telephone numbers.

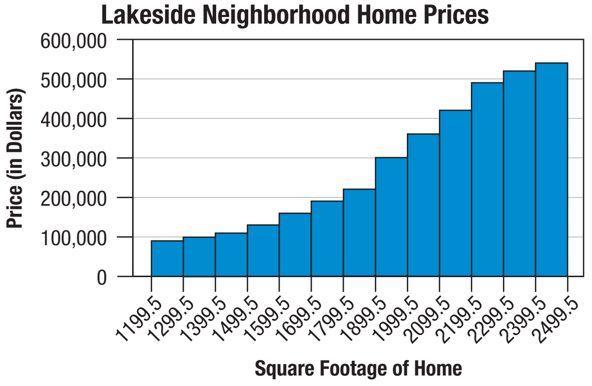
14. The income levels of a group of professional baseball players.

For each set of data displayed below, choose which of the four shapes defined in this section best describes the distribution.

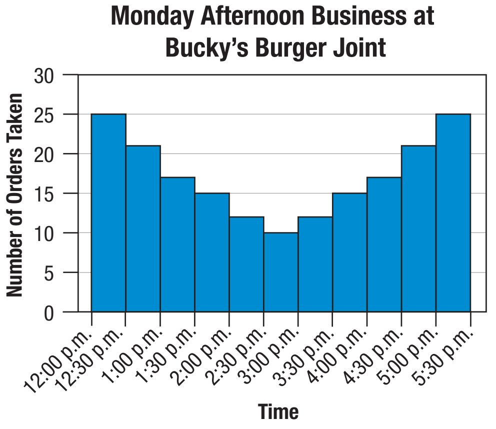
15.



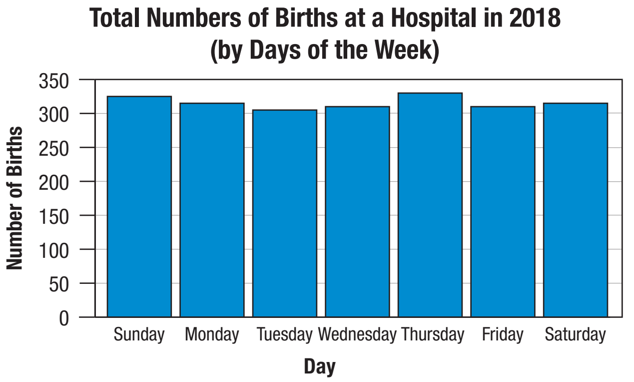
16.


Lakeside Neighborhood Home Prices
Price (in Dollars)
600,000
500,000
400,000
300,000
200,000
100,000
0
1199.5
1299.5
1399.5
1499.5
1599.5
1699.5
1799.5
1899.5
1999.5
2099.5
2199.5
2299.5
2399.5
2499.5
Square Footage of Home

17.


Monday Afternoon Business at Bucky's Burger Joint
Number of Orders Taken
30
25
20
15
10
5
0
12:00 p.m.
12:30 p.m.
1:00 p.m.
1:30 p.m.
2:00 p.m.
2:30 p.m.
3:00 p.m.
3:30 p.m.
4:00 p.m.
4:30 p.m.
5:00 p.m.
5:30 p.m.
Time

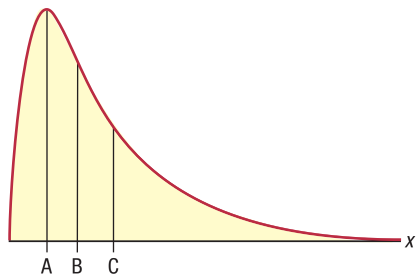
18.


Total Numbers of Births at a Hospital in 2018 (by Days of the Week)
Number of Births
350
300
250
200
150
100
50
0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day

Example 3.1.9

Determining Mean, Median, and Mode from a Graph

Determine which letter represents the mean, the median, and the mode in the graph to the right.



Solution

Mode: The mode is always located at the peak of a distribution, so it is line A.

Median: The median is the value that divides the area of the distribution in half. Here it is represented by line B.

Mean: This distribution is skewed to the right, so the mean will be the measure of center farthest to the right. Here it is represented by line C.

Table 3.1.1 Properties of Mean, Median, and Mode

Mean	Median	Mode
"Average"	"Middle value"	"Most frequent value"
Isn't necessarily a data value	Isn't necessarily a data value	Must be a data value
Single value	Single value	Could be one value, multiple values, or not exist
Affected by outliers	Not affected by outliers	Not affected by outliers
Use for quantitative data with *no* outliers	Use for quantitative data with outliers	Use for qualitative data

3.1 Section Exercises

Mean, Median, and Mode

Find the mean, median, and mode for each of the given data sets, and state whether the data set is unimodal, bimodal, multimodal or has no mode.

1. 10, 1, 5, 9, 4, 1
2. 190, 219, 160, 250, 175, 180, 240, 290
3. 5, 5, 5, 5, 5, 5, 5
4. 8, 9, 10, 12, 4, 3, 5, 7, 10
5. 4.25, 8.3, 6.0, 3.2, 1.9, 4.4, 4.9, 5.9, 5.1
6. 36, –76, 36, –20, 100, 98, 18, 100
7. –22, –83, 77, –79, –2, –42, 98, –66, –42
8. 0.5, 0.6, 1.7, 0.5, 1.3, 1.0, 1.3, 0.5
9. The following data set gives the cellular data overage fees for a group of 9 smart phone customers one month: \$25.00, \$42.00, \$15.00, \$6.16, \$194.73, \$60.00, \$16.00, \$6.40, \$1.10

10. The following data represent batting averages for a sample of professional baseball players.

.275	.265	.333	.244	.279
.250	.288	.292	.370	.243
.236	.321	.305	.292	.250

11. The following data represent average numbers of Tweets per day posted on Twitter for 16 high school students.

0.8	42.2	20.6	2.8
36.7	18.6	23.3	11.5
3.7	14.9	9.4	1.5
14.9	31.1	23.5	9.5

12. The following data represent the ages of 20 American entrepreneurs (in years).

28	39	43	53
35	32	34	29
33	31	32	31
25	22	30	29
41	36	23	47

13. The following data represent high temperatures for cities in the Southeast (in degrees Fahrenheit).

High Temperatures for Cities in the Southeast

Stem	Leaves
7	
7	7 9
8	2 3 4
8	5 5 7 8 8 9 9
9	0 0 1 1 2 2 3
9	5

Key: 7 | 7 = 77 °F

14. The following data represent weights of newborn babies (in pounds).

Weights of Newborn Babies

Stem	Leaves
5	4 9
6	5 6 8
7	3 3 5 5 8
8	2 5 7
9	1 3

Key: 5 | 4 = 5.4 pounds

Using the Mean to Find a Data Value

Use the given information to determine the unknown value.

15. The mean cost for items in a bag of groceries is \$1.96. There are 12 items in the bag, and the following are the prices for 11 of those items. Determine the price of the 12th item in the bag.

\$2.69, \$1.88, \$2.18, \$1.99, \$0.99, \$1.99, \$0.97, \$3.49, \$1.97, \$2.48, \$0.52

16. A plane that ferries visitors to a small resort island has strict guidelines on the weight allowed for passenger luggage. Consequently, the six passengers are limited to a maximum average luggage weight of 35 pounds (lb). The following are the weights of four out of six pieces of luggage: 39 lb, 22 lb, 35 lb, and 37 lb. The two pieces of luggage that haven't been weighed will have to split the remaining weight allowance. If the remaining weight allowance is split evenly between the bags, determine the maximum possible weight allowance for each remaining bag.

Weighted Mean

Calculate the weighted mean as described in each exercise.

17. Marquis is calculating his cumulative GPA. His grades are as follows: A (15 hours), B (18 hours), C (8 hours), and D (3 hours). Note that an A is equivalent to a 4.0, a B is equivalent to a 3.0, a C is equivalent to a 2.0, and a D is equivalent to a 1.0. What is Marquis' GPA?

18. Beth is calculating her cumulative GPA. Her grades are as follows: A (12 hours), B (22 hours), C (14 hours), and F (3 hours). Note that an A is equivalent to a 4.0, a B is equivalent to a 3.0, a C is equivalent to a 2.0, a D is equivalent to a 1.0, and an F is equivalent to 0. What is Beth's GPA?

19. The following table gives the average balances for one bank customer for the months of October through December.

Average Monthly Balances for a Bank Customer (October through December)

Month	Average Balance
October	\$2251.33
November	\$2490.51
December	\$1478.27

Calculate the average monthly balance for the three-month period of October through December. Note that, since each month contains a different number of days, the average balance for each month must be weighted by the number of days in that month.

20. The following table gives the average balances for one bank customer for the months of July through September.

Average Monthly Balances for a Bank Customer (July through September)

Month	Average Balance
July	\$402.45
August	\$322.97
September	\$298.64

Calculate the average monthly balance for the three-month period of July through September. Note that, since each month contains a different number of days, the average balance for each month must be weighted by the number of days in that month.

21. Susan is calculating her final grade in biology. The grade is broken down as follows: tests (50%), lab (30%), and final exam (20%). Susan's category averages are the following.

Tests:	82
Lab:	78
Final Exam:	86

What is her final grade?

22. Demarcus is calculating his final grade in biochemistry. The grade is broken down as follows: tests (40%), homework (10%), labs (20%), and final exam (30%). Demarcus's category averages are listed below.

Tests:	93
Homework:	98
Labs:	85
Final Exam:	89

What is Demarcus's final grade?

23. At the end of the semester, Jeffrey knows all of his grades in his psychology class except for the final exam. The breakdown of his grades and how much each category counts toward the final grade is shown below.

Tests (30%):	84
Homework (20%):	75
Semester Project (25%):	80
Final Exam (25%):	?

What grade must Jeffrey make on his final exam to finish the course with a B (80%)?

24. At the end of the semester, Brietta knows all of her grades in her anatomy and physiology class except for the final exam. The breakdown of her grades and how much each category counts toward the final grade is shown below.

Tests (40%):	88
Homework (15%):	85
Semester Project (25%):	94
Final Exam (20%):	?

What grade must Brietta make on her final exam to finish the course with an A (90%)?

25. Let's extend the concept of weighted mean to include the mean for a frequency distribution. The following frequency distribution gives the hourly wage for a person's first job.

Hourly Wage at First Job (in Dollars)

Class	Frequency
7.50 – 8.49	12
8.50 – 9.49	50
9.50 – 10.49	48
10.50 – 11.49	45
11.50 – 12.49	34

Since we do not know the exact value of each person's hourly wage, we will estimate that each value in a class is equal to the midpoint of that class. To calculate the mean of the frequency distribution by hand, we can use the formula for a weighted mean, with the data values equal to the class midpoints and the weights equal to the class frequencies. To use a TI-83/84 Plus calculator to find the mean of the frequency distribution, we can use the same procedure outlined in Example 3.3, again with the data values equal to the class midpoints and the weights equal to the class frequencies. Estimate the mean hourly wage for the given data.

26. The following frequency distribution shows water level readings for the Mississippi River at New Orleans over a period of 90 days.

Water Level Readings of the MS River at NOLA (in Feet)

Water Level	Frequency
14.5 - 15.49	19
15.5 - 16.49	19
16.5 - 17.49	18
17.5 - 18.49	20
18.5 - 19.49	14

If flood stage is 17 feet, is the average river level over this period of time above flood stage? Explain your answer. Use the instructions given in exercise 25 to help find the solution.

Determining the Most Appropriate Measure of Center

Determine the most appropriate measure of center for the described data set.

27. Hairstyles of students on a given college campus (length, color, straight/curly, and so forth)

28. Prices of used cars on a lot including the following models: Buick LaCrosse, Toyota 4Runner, Honda Prelude, Ford Escape, Dodge Grand Caravan, Ford Explorer, and Ferrari F430

29. Number of minutes students spend completing a homework assignment for an honors class

30. Prices of similar sofas at different furniture stores

31. B, C, D, B, D, B, D, A, B, A, D, C, A, B

32. A popular website rates vacation resorts using the following scale.

★★★★★ = Excellent

★★★★ = Above average

★★★ = Average

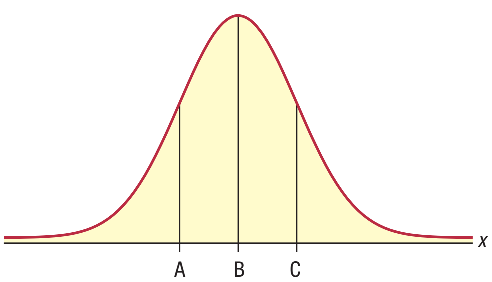
★★ = Below average

★ = Terrible

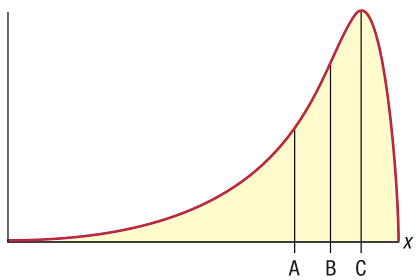
Graphs and Measures of Center

For each graph, determine which letter represents the mean, the median, and the mode.

33.



34.



35. Given the following measures of center for a data set, describe the most likely shape of the distribution.

Mean = 121.89, Median = 163.5, Mode = 165.2

36. Given the following measures of center for a data set, describe the most likely shape of the distribution.

Mean = 31.5, Median = 31.5, Mode = 31.5

Solution

Since we are not told in the problem whether the distribution of the data is bell-shaped, we cannot apply the Empirical Rule here. However, we can apply Chebyshev's Theorem to find a minimum estimate. To do so, we need to know how many standard deviations $1989 and $5229 are from the mean. By subtracting, we can find how far each of these figures lie from the mean. Then, dividing by the standard deviation, we can convert these differences into numbers of standard deviations. Here are the calculations.

Rent = $1989

$$\text{Distance from Mean} = x - \mu = \$1989 - 3609 = -\$1620$$

$$\text{Dev. from Mean} = \frac{\text{Distance}}{\text{Standard Deviation}} = \frac{-\$1620}{\$540} = -3$$

Rent = $5229

$$\text{Distance from Mean} = x - \mu = \$5229 - \$3609 = \$1620$$

$$\text{Dev. from Mean} = \frac{\text{Distance}}{\text{Standard Deviation}} = \frac{\$1620}{\$540} = 3$$

Thus these rent prices lie three standard deviations above and below the mean. Chebyshev's Theorem can then be applied for $K = 3$. Using the calculation previously shown in the box with the theorem, we can say that at least 88.9% of the rent prices lie within this range.

3.2 Section Exercises

Range, Population Standard Deviation, and Population Variance

Calculate the range, population standard deviation, and population variance of the given data set.

1. 3, 5, 7, 8

2. 8, 12, 10, 12, 4, 13, 5, 17, 10

3. 5, 5, 5, 5, 5, 5, 5

4. –5, 6, 3, –5, –7, –7, 8, –25

5. 17, 19, 21, 16, 20, 19, 17, 13

6. 1.6, 2.2, 1.1, 3.0, 4.9, 2.8, 5.7, 4.5, 6.0

Range, Sample Standard Deviation, and Sample Variance

Calculate the range, sample standard deviation, and sample variance of the given data set.

7. 2, 2, 3, 4, 7

8. 15, 12, 13, 14, 15, 17, 18, 11, 12, 15

9. 4, 4, 4, 4, 4, 4, 4

10. –1, 2, 3, –2, –4, 4, 5, –35, 2

11. 2.0, 1.7, 1.9, 1.5, 1.6, 1.9, 1.7, 3.0

12. 8.1, 9.5, 10.7, 12.3, 4.4, 3.9, 5.1, 7.3, 10.1

Sample Standard Deviation

Calculate the sample standard deviation for the given data set.

13. The following data represent weights of yorkshire terriers (in pounds).

6.8	9.1	8.7	7.5	8.2
5.4	6.5	8.5	7.3	6.6
5.9	7.3	9.3	7.5	7.8

14. The following data represent average numbers of Tweets per day posted on Twitter for 16 high school students.

0.8	42.2	20.6	2.8
36.7	18.6	23.3	11.5
3.7	14.9	9.4	1.5
14.9	31.1	23.5	9.5

15. The following data represent high temperatures for cities in the Southeast (in degrees Fahrenheit).

High Temperatures for Cities in the Southeast

Stem	Leaves
7	
7	7 9
8	2 3 4
8	5 5 7 8 8 9 9
9	0 0 1 1 2 2 3
9	5

Key: 7 | 7 = 77 °F

16. The following data represent ages of 20 American entrepreneurs (in years).

Ages of American Entrepreneurs

Stem	Leaves
2	2 3
2	5 8 9 9
3	0 1 1 2 2 3 4
3	5 6 9
4	1 3
4	7
5	3
5	

Key: 2 | 2 = 22 years old

Standard Deviation and Variance

Decide if each statement is true or false. Explain why.

17. If the standard deviation of a data set is zero, then all entries in the data must equal zero.

18. The population variance and sample variance are the same value for the same set of data.

19. It is possible to have a standard deviation of –3 for some data set.

20. It is possible to have a standard deviation of 435,000 for some data set.

Coefficient of Variation

Calculate the coefficient of variation, CV, for each data set and then answer the question.

21. The data in set A represent numbers of hours worked in one week for a sample of employees of a fast food restaurant. The data in set B represent numbers of minutes spent waiting for food for a sample of customers at the same restaurant. Which of the two data sets has the *larger* spread relative to its own mean?

A: 31, 33, 35, 39, 31, 32, 30, 40, 13, 41, 38, 32, 37, 33

B: 2.3, 3.5, 4.3, 2.1, 4.8, 3.9, 2.0, 3.3, 4.0, 1.6, 2.2

22. The data in set A represent prices (with tax included) of cookies sold at a sample of bakeries. The data in set B represent number of cookies sold in one weekend for the same sample of bakeries. Which of the two data sets has the *larger* spread relative to its own mean?

A: $1.23, $1.55, $1.01, $1.89, $2.35, $2.56, $2.71, $1.75, $2.01, $1.59

B: 119, 145, 97, 121, 118, 98, 102, 114, 118, 99

23. The data in set A represent numbers of orders received by an online retailer for a random sample of months from the past two years. The data in set B represent package weights (in pounds) for a random sample of customer orders from the same online retailer in the past two years. Which of the two data sets has the *smaller* spread relative to its own mean?

A: 21,568 20,888 20,037 21,932 22,000 21,123 21,567 22,298

B: 4.5, 4.3, 6.5, 3.3, 4.7, 3.67, 4.01, 3.89, 4.4, 2.99, 4.88, 3.77

24. The data in set A represent numbers of graduates from a high school for a random sample of years since the school opened. The data in set B represent the numbers of graduates who started college during the year after graduating from the same high school for the same sample of years. Which of the two data sets has the *smaller* spread relative to its own mean?

A: 328, 444, 283, 289, 345, 327, 298, 277, 419, 402, 399, 418, 401

B: 78, 73, 92, 89, 74, 88, 91, 71, 70, 89, 81, 83, 84

Empirical Rule

Use the Empirical Rule to answer the questions.

25. Suppose that private school tuitions in one region of the country have a bell-shaped distribution with a mean of $25,400 and a standard deviation of $1300. Approximately what percentage of tuitions are between $24,100 and $26,700?

26. Suppose that private school tuitions in one region of the country have a bell-shaped distribution with a mean of $25,400 and a standard deviation of $1300. Approximately what percentage of tuitions are between $22,800 and $28,000?

27. Suppose that electric bills for the month of May in one city have a bell-shaped distribution with a mean of $119 and a standard deviation of $22. Approximately what percentage of electric bills are greater than $97?

28. Suppose that electric bills for the month of May in one city have a bell-shaped distribution with a mean of $119 and a standard deviation of $22. Approximately what percentage of electric bills are less than $163?

29. Suppose it is known that verbal SAT scores have a bell-shaped distribution with a mean of 500 and a standard deviation of 100. Approximately what percentage of verbal SAT scores are no more than 600?

30. Suppose it is known that verbal SAT scores have a bell-shaped distribution with a mean of 500 and a standard deviation of 100. Approximately what percentage of verbal SAT scores are at least 300?

Chebyshev's Theorem

Use Chebyshev's Theorem to answer the questions.

31. Suppose that household electric bills for the months of May through August in a city in Florida have a mean of $230 and a standard deviation of $58. What is the minimum percentage of electric bills between $56 and $404?

32. Suppose that salaries for associate mathematics professors at one university have a mean of $64,900 and a standard deviation of $9400. What is the minimum percentage of associate professors with salaries between $46,100 and $83,700?

33. Car insurance premiums in one region have a quarterly mean of $246 and a standard deviation of $31. What is the minimum percentage of car insurance premiums between $184 and $308?

34. The average weight of cows auctioned at a large livestock event is 1614 with a standard deviation of 59 pounds. What is the minimum percentage of cows that weigh between 1437 and 1791 pounds?

Standard Deviation and Variance of Grouped Data

Estimate the sample standard deviation or variance of the data in each frequency distribution using the given formula.

35. Let's extend the concept of standard deviation to include the standard deviation for a frequency distribution. The following frequency distribution gives the final grades for students in a statistics class.

Final Grades

Grade	Frequency
66–72	4
73–79	7
80–86	12
87–93	8
94–100	5

Since we do not know the exact value of each final grade, we will estimate that each value in a class is equal to the midpoint of that class. Use the following formula to estimate the sample standard deviation of the data in the frequency distribution, if you calculate this value by hand. To calculate this estimate using a TI-83/84 Plus calculator, use the same directions given in Section 3.1 for calculating a weighted mean, entering the midpoints in L1 and the frequencies in L2.

$$s = \sqrt{\frac{n\left[\sum\left(f_i \cdot x_i^2\right)\right] - \left[\sum\left(f_i \cdot x_i\right)\right]^2}{n(n-1)}},$$

where n = sample size,

f_i = frequency of class i, and

x_i = midpoint of class i.

36. Use the formula given in Exercise 35 to estimate the sample standard deviation of the gas prices in the following frequency distribution.

Gas Prices

Price in Dollars per Gallon	Frequency
3.55–3.59	1
3.60–3.64	3
3.65–3.69	5
3.70–3.74	6
3.75–3.79	2
3.80–3.84	1

37. What is the approximate sample variance of the final grades in the frequency distribution given in Exercise 35?

38. What is the approximate sample variance of the gas prices in the frequency distribution given in Exercise 36?

3.3 Section Exercises

Percentiles

Answer the questions that follow each set of data.

1. The following data represent weights of Yorkshire Terriers (in pounds).

6.8	9.1	8.7	7.5	8.2
5.4	6.5	8.5	7.3	6.6
5.9	7.3	9.3	7.4	7.8

a. Which weight represents the 50^{th} percentile?

b. What is the percentile of a weight of 8.2 pounds?

2. The following data represent high temperatures for cities in the Southeast (in degrees Fahrenheit).

85	82	93	88
92	79	84	90
77	83	91	89
90	85	87	91
89	92	95	88

a. Which temperature represents the 75^{th} percentile?

b. What is the percentile of a temperature of 93 °F?

3. The following data represent average numbers of Tweets per day posted on Twitter for 16 high school students.

0.8	42.2	20.6	2.8
36.7	18.6	23.3	11.5
3.7	14.9	9.4	1.5
14.9	31.1	23.5	9.5

a. Which number represents the 25^{th} percentile?

b. What is the percentile of an average of 11.5 Tweets per day?

4. The following data represent scores on a test given in Mr. Jones's second period algebra class.

51	53	60	62	68	69
70	73	77	77	79	80
82	84	84	89	90	92
92	92	96	97	99	99

a. What test score represents the 40^{th} percentile?

b. What is the percentile of a test score of 92?

Exercises 5–10 reference data sets that are available for download from stat.hawkeslearning.com in Microsoft Excel or Minitab format. The Patient Data are medical statistics for a sample of patients from a hospital. The State Data are measurements collected by the US Census Bureau for each of the 50 states. Use the data sets to answer the following questions.

5. Refer to the Patient Data.

a. Which patient weight represents the 35^{th} percentile?

b. What is the percentile of a weight of 189 pounds?

6. Refer to the State Data.

a. Which population represents the 82^{nd} percentile?

b. What is the percentile of a mean travel time of 27 minutes?

7. What is the 85^{th} percentile for the patients' systolic blood pressure readings from the Patient Data?

8. In the State Data, what is the percentile of Oregon's mean travel time to work, 23.5 minutes?

9. What is the 50^{th} percentile of the patients' ages in the Patient Data?

10. What is the 30th percentile for state populations in the State Data?

Five-Number Summary

Find the five-number summary for each data set. Use the approximation method to calculate quartiles. These values will match those produced by a TI-83/84 Plus calculator.

11. The following data represent prices (in dollars) of used cars listed on www.autotrader.com for one zip code.

18,865	11,442	15,750	10,960	15,635
15,963	13,702	14,788	15,495	8250
	12,900	14,850	6450	

12. The following data represent weights of dimes, measured in grams.

2.268	2.267	2.269	2.268	2.271
2.266	2.267	2.268	2.270	2.272
	2.265	2.269	2.268	

13. The following data represent INR (International Normalized Ratio) readings of patients with blood clotting disorders.

1.5	2.1	1.7	3.5
4.1	1.2	1.7	1.8

14. The following data represent SAT Critical Reading scores for a randomly selected group of high school seniors.

520	750	620	470
520	660	780	580
390	460	660	570
290	500	690	540

15. The following data represent birth weights (in pounds) of eight newborn babies born on the same day at a local hospital.

5.4	6.5	7.8	9.1
9.3	10.1	7.8	9.0

16. The following data represent times, in minutes, taken by students in a physical education class to run/walk two miles.

12.22	12.35	13.45	16.78	19.01
21.34	24.87	25.10	26.93	29.81

17. The following data represent changes in weight, measured in pounds, from the beginning of a new diet to one month later.

–12	–11	–10
–8	–7	–5
–2	0	1
3	4	

18. The following data represent differences in high temperatures, in degrees Fahrenheit, for the same day from one year to the next in nine metropolitan areas.

9	–2	–8
12	15	19
21	–14	–11

Box Plots

Draw a box plot for each set of data on the same graph. Use the approximation method to calculate quartiles. These values will match those produced by a TI-83/84 Plus calculator. Use your box plots to answer the following questions.

a. Which data set has the smallest value?
b. Which data set has the larger median?
c. Which data set has the larger interquartile range?

19. Weight loss (in pounds) from diet A: 2, 3, 5, 5, 5, 6, 6, 6, 7, 7, 8

Weight loss (in pounds) from diet B: 3, 3, 4, 4, 4, 5, 6, 6, 9, 12

20. Respiratory rates at rest (in breaths per minute) of adults in group A:

10, 12, 13, 14, 15, 16, 17, 18

Respiratory rates at rest (in breaths per minute) of adults in group B:

11, 15, 17, 18, 19, 19, 20

21. Test scores for class A: 45, 60, 57, 83, 72, 93, 87, 73, 92

Test scores for class B: 23, 88, 67, 89, 91, 76, 72, 100, 95, 35

22. Diameters of cans (in cm) from assembly line A:

5.6, 5.7, 5.1, 5.7, 5.5, 5.9, 5.7, 5.5, 5.6, 5.6

Diameters of cans (in cm) from assembly line B:

5.4, 5.7, 5.6, 5.5, 5.6, 5.7, 5.7, 5.8, 5.6, 5.5

IQR and Outliers

Answer the following questions regarding interquartile range and outliers.

a. Calculate the IQR.
b. Use the IQR to identify any outliers.

23. The following data set represents the average sailing speed (in knots) for a sample of 15 racing yachts.

9.0	11.6	12.5	12.9	13.1
13.7	14.0	14.3	14.5	15.1
15.3	15.4	15.6	16.1	17.2

24. The following data gives the amount of student loans (in thousands of dollars) held by a group of 20 randomly selected recent college graduates.

0	0	6.7	7.6	8.1
9.9	10.8	17.0	20.5	24.6
25.5	29.3	30.0	34.8	44.5
50.5	65.2	85.2	95.8	102.5

Standard Scores

Calculate the standard score using the given values. Round your answer to two decimal places.

25. $\mu = 25, \quad \sigma = 3, \quad x = 27$

26. $\bar{x} = 37, \quad s = 8, \quad x = 34$

27. $\mu = 0.32, \quad \sigma = 0.01, \quad x = 0.29$

28. $\mu = 2, \quad \sigma = 0.5, \quad x = 1.8$

29. Carlita scored 32 on the ACT Mathematics Test and 730 on the mathematics section of the SAT. If the ACT Mathematics Test had a mean score of 21.0 with a standard deviation of 5.3, and the mathematics section of the SAT had a mean score of 516 with a standard deviation of 116, on which exam did Carlita earn a better math score with respect to her peers?

30. A manufacturer makes aluminum cans and longneck bottles. The average diameter of an aluminum can is supposed to be 4.2 inches, with an allowable standard deviation of 0.01 inches. The average diameter on a longneck bottle is supposed to be 3.8 inches, with an allowable standard deviation of 0.02 inches. A factory worker randomly selects a can from the assembly line and it has a diameter of 4.3 inches. The worker then selects a bottle from the assembly line and it has a diameter of 3.75 inches. Which assembly line is closest to specifications?

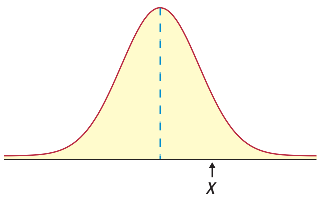
31. Don played in a local golf tournament for charity and scored a round of 63 while the average round for the day was a 74 with a standard deviation of 3 strokes. Later that week, Don played in a Pro-Am tournament and scored a 65 while the average score for the day was a 79 with a standard deviation of 4 strokes. Which was Don's better round of golf in comparison to the competition? (Remember, in golf, lower scores are better!)

32. A large sample of English labrador retrievers has a mean weight of 70.3 pounds with a standard deviation of 4.9 pounds. A similarly large sample of Siberian huskies has a mean weight of 48.8 pounds with a standard deviation of 3.5 pounds. Maggie, an English lab, is 74 pounds and Scout, a Siberian husky, is 52 pounds. Which dog is larger with respect to the average weight of its breed?

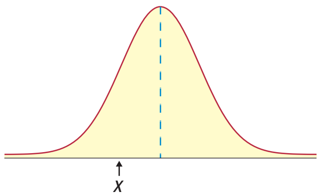
For each graph, where the mean is marked by the dotted line, which is a likely z-score for the indicated value of x? Choose from the following z-scores:

a. –1.3 **b.** 0 **c.** 1.7 **d.** 2.8

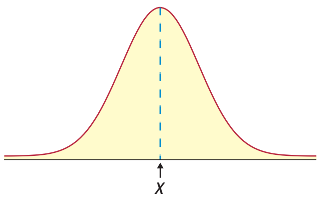
33.



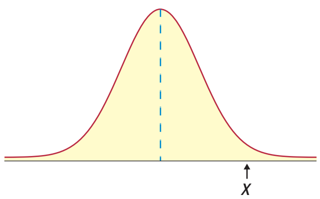
34.



35.



36.



4.1 Section Exercises

Sample Spaces

Find the sample space for the given experiment.

1. Two coins are tossed.

2. A child's board game contains a spinner with three colors: orange, yellow, and green. Give the sample space for two consecutive spins.

3. Choose an outfit consisting of one top and one bottom from three shirts (T-shirt, button-down, and sweater) and two pants (one pair of jeans and one pair of slacks).

4. A bag contains four marbles: one each of green, red, blue, and violet. Two marbles are drawn from the bag. Assume that the first marble is not put back in the bag before drawing the second marble.

5. When buying a new car, you've narrowed your choices to three colors: red, black, or silver. You also need to decide whether to have a sunroof or not, and whether you want leather or cloth interior.

6. When ordering a pizza with a coupon, you have a choice of crusts: thin, hand-tossed, or stuffed. You can also choose one topping from the following: pepperoni, ham, sausage, onion, bell pepper, or olives.

7. When choosing your seat at the opera, you can choose from three levels and then whether you want an aisle seat or not.

8. When building your new house, you have a choice of flooring for the kitchen: tile, concrete, or wood. You must also choose the counter tops from the following: granite, concrete, wood, or stainless steel.

Types of Probability

Determine whether each probability is subjective, experimental, or classical.

9. Jeff wants to know whether a certain coin is fair or not. He flips the coin 100 times and obtains tails 61 times. He calculates that the probability of obtaining a tail with his coin is 61%.

10. Caroline estimates that there is only a 10% chance that they will have a quiz in biology.

11. Mr. Dorrough's 18 students have dropped their names in the hat for a prize drawing. Stephanie calculates that she has a 1/18 chance of winning.

12. On a game show, the contestant must choose one of three doors, behind one of which is a new car. He has a one-in-three chance of winning the car.

13. A computer manufacturer estimates that there is less than a 5% chance that its customers will want to switch brands within the first year.

14. A student is randomly assigned to one of 5 College Algebra sections. The student's probability of getting assigned to one of two sections taught by the instructor the student prefers is $\frac{2}{5}$.

Experimental Probability

Calculate each experimental probability.

15. A very large bag contains more coins than you are willing to count. Instead, you draw a random sample of coins from the bag and record the following numbers of each type of coin in the sample before returning the sampled coins to the bag.

Coins in a Bag

Quarters	Dimes	Nickels	Pennies
23	29	17	38

If you randomly draw a single coin out of the bag, what is the probability that you will obtain:

a. A nickel?

b. A penny?

c. Either a quarter or a dime?

16. A telemarketer's computer selects phone numbers at random. The telemarketer has recorded the number of respondents in each age bracket for one evening in the following table.

Number of Respondents by Age

18–25	26–35	36–45	Over 45
29	40	55	51

What is the probability that the next respondent will be:

a. Over 45?

b. Between 26 and 35?

c. At least 36?

Classical Probability

Calculate each classical probability. Assume that individual outcomes are equally likely.

17. Martha has a box full of 17 different vintage vinyl records: 5 rock, 3 blues, 6 pop, and 3 R&B. If she randomly pulls out a record, what is the probability that it is a blues record?

18. Chloë puts a coin into a gumball machine that contains 12 blue, 15 pink, 9 orange, 16 yellow, and 14 white gumballs. What is the probability that Chloë gets a yellow gumball?

19. New players at a certain casino are issued a "players card" at random when they first enter the casino. The cards are preloaded with either \$25 or \$50 in free play amounts. There are 5 different designs of cards for each amount. What is the probability that the card issued to a new player is preloaded with \$50 in free play?

20. What is the probability that a person selected at random will have a March birthday? (Assume that every day of the year contains an equal number of birthdays, and the person was not born in a leap year.)

21. If Mark grabs a utensil out of a drawer without looking, what is the probability that he grabs a fork if there are 7 knives, 9 spoons, and 6 forks in the drawer?

22. If Timmy is fishing in his newly stocked pond and knows that there are 200 bream and 150 bass in it, what is the probability that the first fish he catches will be a bass?

23. John is at a cookout and wants to get a drink from the cooler. If there are 12 colas, 10 bottles of water, and 5 root beers in the cooler, what is the probability that he randomly grabs a root beer?

24. A college algebra class has 14 freshmen, 21 sophomores, 9 juniors, and a senior enrolled. What is the probability that the professor randomly selects the senior to answer a question?

25. Mary Ann is sewing and needs the spool of white thread. Her basket of sewing supplies is sitting next to her, and it contains 26 different colors of thread, including the white spool she needs. If she grabs one spool without looking, what is the probability that she has chosen the white spool of thread?

26. For a school fundraiser, 1000 raffle tickets are sold for $5 each. Each ticket is assigned a three-digit number using the digits 0–9. What is the probability that the winning ticket will be one with three repeating digits?

27. If you roll one six-sided die twice, what is the probability that the second number rolled is at least as large as the first?

28. Each night in a family of four siblings, the parents randomly assign the order for the siblings to take their nightly baths. What is the probability that the youngest sibling has to take her bath last?

4.2 Section Exercises

Complement

Describe the complement for the given event.

1. Out of 253 apple trees in the orchard, just 42 are ready for harvesting.
2. In the graduating class at one large high school, 39% of the graduates plan to attend college out-of-state.
3. Out of 21 players on the baseball team, 4 are left-handed.
4. The adult literacy rate for the United States is 97%.
5. A local news station claims that 70% of its viewers are over the age of 30.
6. Of the students not returning to Valley State Community College for the next semester, 75% of them either graduated or transferred to a university or both.

Complement Rule

Use the Complement Rule to find each probability.

7. A food distributor estimates that 2% of the eggs it delivers to grocery stores are cracked. What is the probability that an egg selected at random will not be cracked?
8. Every eleventh box of Blue Zinga cereal that is produced contains a special toy. If you buy a box of Blue Zinga, what is the probability that your box does not contain a toy?
9. Find the probability of rolling two dice and not getting doubles.
10. Given that every fifth person in line will get a coupon for a free box of popcorn at the movies, what is the probability that you don't get a coupon when you're in line?

Mutually Exclusive Events

Determine whether the events are mutually exclusive.

11. For a survey on campus, a sophomore or a history major is chosen.
12. When determining his schedule for the spring semester, Troy must decide whether to take a history class that meets at 9:00 a.m. on Mondays, Wednesdays, and Fridays or a science class that meets Mondays and Wednesdays at 9:30 a.m. (Assume that each class meets for 50 minutes.)
13. When assigning parts in a high school play, a senior or someone who's been in at least two other plays can play the lead role.
14. Choose a diet cola or a bottle of water out of the refrigerator.
15. Invest all disposable income in a startup tech company or real estate.
16. A student scores an A on an exam, or a student scores a passing grade.

Probability

Calculate each probability.

17. Suppose that 4 out of the 15 doctors in a small hospital are trained in special procedures. 11 out of the 15 are under the age of 45, and 2 are both trained in special procedures and under the age of 45. What is the probability that you are randomly assigned a doctor trained in special procedures or a doctor under the age of 45?
18. Nineteen college graduates are applying for internships at a Wall Street financial institution. Of these applicants, 12 graduated from Ivy League universities, 8 are business majors, 9 graduated summa cum laude, and 5 graduated summa cum laude from Ivy League universities. If one resume is selected at random for the first interview, what is the probability that the applicant is either an Ivy League graduate or a summa cum laude graduate?

19. The local arbor society is giving out free trees to the public. They have 200 pine trees, 100 dogwoods, 125 oaks, and 200 birch trees. If you are the first in line to receive your free tree and you cannot choose what type of tree you are given, what is the probability that you randomly get either an oak or a dogwood?

20. A flight attendant for USA Airlines is given an assignment that is randomly chosen from the following available flights: 15 European destinations, 6 Caribbean destinations, 8 South American destinations, 5 Asian destinations, and 11 destinations in the continental United States. What is the probability that his next assignment takes him either to Europe or Asia?

21. Consider the experiment in which two six-sided dice are tossed. What is the probability that the total is not four?

22. Consider the experiment in which two six-sided dice are tossed. What is the probability that either both dice are sixes or neither die is a six?

23. What's the probability of rolling at most one even number with two six-sided dice?

24. Jacki has a sizable collection of music downloaded to her phone. The relative frequencies for each genre of music she has stored are given in the table below.

Jacki's Music Collection

Genre	Relative Frequency
Rock	0.3256
Country	0.1181
R&B	0.2068
Oldies	0.0843
Show Tunes	0.1956
Bluegrass	0.0696

If Jacki chooses the shuffle option and one song is randomly chosen for play, what is the probability that it is either country or bluegrass?

25. Ana is looking forward to starting kindergarten in her Texas hometown. Of 14 kindergarten teachers, 3 are new this year, 11 are bilingual, and 1 is both new and bilingual. If class rosters are generated at random, what is the probability that Ana's teacher is either new or bilingual?

26. What is the probability of rolling two six-sided dice and obtaining either an odd total or a total less than six?

27. In the teacher's pencil box, 13 out of the 20 pencils have no eraser, and 12 out of the 20 are not mechanical pencils. There are 4 pencils that are both mechanical and have erasers on them. What is the probability that Ava chooses a pencil at random from the box and gets either a mechanical pencil or a pencil with an eraser?

28. For a story she is writing in her high school newspaper, Grace surveys moviegoers selected at random as they leave the new feature *Mystery on Juniper Island*. She simply asks each moviegoer to rate the show using a thumbs-up or thumbs-down. The results of her survey are given in the table below.

Survey Results

	Under 40 years old	40 years or older
Thumbs-Up	23	40
Thumbs-Down	19	11

What is the probability that one of Grace's survey respondents has either given a thumbs-up rating or is under 40 years old?

$$\begin{aligned}P(\text{choose same snack}) &= \frac{\text{number of ways to choose same snack}}{\text{total number of snacks}}\\ &= \frac{1}{49}\\ &\approx 0.0204\end{aligned}$$

Thus, there is about a 2% chance that each child will eat the same thing two days in a row.

4.3 Section Exercises

Independent Events

Determine whether the events are independent.

1. Finding the batteries in your calculator dead. Finding the battery in your car dead.
2. Getting a letter from your aunt. Getting a bank statement in the mail.
3. Claire's name being drawn in the school raffle, without replacement. Ben's name being drawn in the school raffle.
4. Buying a new shirt on sale. Having enough money for lunch that day.
5. Lois ordering a steak. Ann ordering a salad.
6. Winning the Mississippi state lottery. Winning the Florida state lottery.

Multiplication Rule for Independent Events

Calculate each probability.

7. Carriage rides in Charleston, South Carolina are unpredictable. You never know beforehand what part of the city you will see on your tour. Horse-drawn carriages pull up to a check-in gate at the start of the tour where they are randomly assigned one of the city's four routes. A monitor chooses a ball with a number on it assigning your carriage a route. What is the probability that three carriages in a row receive the same route assignment if balls are replaced after each assignment?
8. At the end of each month, two of the 221 employees at SpeedChart are randomly chosen to receive gift cards. Because of a computer glitch in April, receiving the first gift card did not exclude an employee from also receiving the second gift card that month. Find the probability that the same employee received both gift cards in the month of April.
9. There are 15 different colored crayons in a new box. What is the probability that the orange and then the green will be chosen at random, without replacement?
10. Insulin pens used to administer a patient's insulin at hospitals have a malfunction rate of 9%. This means that out of a box of 200 pens, 18 are defective in some way and must be thrown away. Find the probability of randomly selecting 3 defective insulin pens in a row from a brand new box of 200 pens, if a defective pen is immediately discarded.
11. Ashley's Internet service is terribly unreliable. In fact, on any given day, there is a 15% chance that her Internet connection will be lost at some point that day. What is the probability that her Internet service is not broken for five days in a row?
12. On a five-day vacation, the forecast is a 50% chance of rain every day. What's the probability that it rains every day?
13. On awards day at the end of the year, Jasmine has an 85% chance of winning the top award in English and a 4 out of 5 chance of winning an award for athletics. What's the probability that Jasmine wins both awards?

14. Suppose in any given year, the probability of a cactus blooming is 0.15 and the probability of an orchid blooming is 0.09. Determine the probability that both of these plants will bloom in the same year, assuming that they are independent events.

Conditional Probability

Calculate each conditional probability.

15. A high school swim team consists of 3 seniors, 2 juniors, 4 sophomores, and 2 freshmen. A relay team of four swimmers is chosen at random from the team members. What is the probability that the last swimmer chosen is a sophomore given that there is senior, a junior and a freshman already chosen?

16. A box of markers contains 10 black-inked (4 wide-tipped and 6 fine-tipped) and 15 red-inked (3 wide-tipped and 12 fine-tipped). What's the probability that a randomly chosen marker will be red, given that it is fine-tipped?

17. A six-sided die is rolled. What is the probability of rolling a two, assuming that you rolled an even number?

18. Josh is playing backgammon, a game played with two six-sided dice. What is the probability that the sum of the two dice he rolls is less than 4 given that he rolls an odd number?

19. Mrs. Harvey's algebra class has 42 students, classified by academic year and method of instruction as follows.

Mrs. Harvey's Algebra Class

	In-Class Instruction	Online Instruction
Freshman	9	13
Sophomore	4	5
Junior	4	2
Senior	2	3

Mrs. Harvey randomly chooses one student's homework to grade first.

a. What is the probability that she selects an in-class student, given that she chooses from only the sophomores?

b. What is the probability that she selects a junior, given that she chooses an online student?

c. What is the probability that she selects an online student, given that she chooses a junior?

20. The following table displays the breakdown of attendees at an International Biology conference by country and their role in the company they were representing.

International Biology Conference Attendees

	Canada	France	South Korea	United Kingdom	United States	Total
CEO	138	45	4	19	117	323
Director	8	4	25	6	63	106
Partner	23	7	3	20	103	156
Chairman	12	9	3	9	62	95
Other	112	146	154	143	2103	2658
Total	293	211	189	197	2448	3338

A random attendee is selected for an interview.

a. What is the probability that a Partner is selected, given that the attendee is from South Korea?

b. What is the probability that a Canadian is selected, given that the attendee is a director of the company?

c. What is the probability that a director is selected, given that the attendee is Canadian?

d. What is the probability that a CEO is selected, given that the attendee is from the continent of North America?

Multiplication Rule for Dependent Events

Calculate each probability.

21. David likes to keep a jar of change on his desk. Right now his jar contains 26 pennies, 19 nickels, 11 dimes, and 16 quarters. What is the probability that David reaches in and randomly grabs a quarter and then a nickel?

22. As part of an incentive, World Autos is offering any salesperson with at least 15 car sales in each month a chance to win one of three $50 gift cards. For every additional 5 cars, your name is entered again. The chart below shows the number of sales for each team member. Winning names are drawn without replacement from all entries.

Monthly Sales at World Autos					
Salesperson	**15 cars sold**	**20 cars sold**	**25 cars sold**	**30 cars sold**	**35 cars sold**
B. Carple	✓	✓			
R. Henry	✓				
M. Trent					
J. Owhan	✓	✓	✓	✓	
L. Prince	✓	✓	✓		
S. Smith	✓	✓	✓	✓	✓
H. Mowis	✓				
T. Hilson	✓				

a. Find the probability of S. Smith winning all three gift cards.

b. Find the probability of S. Smith winning the first gift card and M. Trent winning the second gift card.

c. Find the probability that all three winners are salespeople with only one entry in the drawing.

23. In a biology lab, vials considered defective must be thrown away. The following chart shows the breakdown of vials in a testing center by distributor.

Lab Vials by Distributor		
Lab Vials	**BioVial**	**I.B.C.**
Usable	1762	2001
Defective	3	2

a. Find the probability that two lab partners both randomly choose a defective vial, one after another, if all the vials are stored together.

b. Find the probability that a usable vial distributed by I.B.C. is selected followed by a usable vial distributed by BioVial.

24. Winning lottery numbers are randomly chosen from 101 balls. There are balls numbered 1- 99, along with 2 unnumbered blank balls.

a. What is the probability that the two blank balls are drawn first?

b. What is the probability that the ball numbered 99 is chosen first and then a blank ball is chosen?

c. What is the probability that a blank ball is not drawn if 5 balls are drawn in total?

Fundamental Counting Principle

Use the Fundamental Counting Principle to determine the total number of outcomes.

25. Determine the number of five-digit ZIP codes that can be made from the digits 0–9. (Assume that digits may repeat.)

26. Henry is setting a six-character password on his computer. He is told that the first two characters must be letters and the last four must be numbers. No character may be used twice. How many choices does Henry have for his password?

27. There are six children, three 1st graders and three 2nd graders, singing in the school program. As they line up to perform, the choral director insists that the first person be a 1st grader and the last a 2nd grader. How many ways can the children line up?

28. How many even four-digit PINs can be created from the digits 0–9?

29. How many odd six-digit PINs can be created from the digits 0–9?

30. As Candy is trying to decide what to wear, she can choose from eight blouses, three skirts and four pairs of shoes. Assuming everything coordinates, how many different possible outfits does she have?

Fundamental Counting Principle and Probability

Use the Fundamental Counting Principle to determine the total number of outcomes and then calculate the probability.

31. In the game of Clue, the guilty person can be chosen from 6 people, and there are 9 different possible weapons and 9 possible rooms. What is the probability of making a random guess of the guilty person, location, and murder weapon, and the guess being correct?

32. When ordering his new office equipment, Joe must choose one of 5 monitors, one of 4 printers, and one of 6 scanners. Joe likes 2 of the monitors, 1 of the printers, and 3 of the scanners. If his boss randomly chooses his office equipment for him, what is the probability that Joe will receive a system that makes him happy?

33. A couple having twins is deciding on names. They narrowed their choices to 5 family names and 7 non-family names. The new father's parents like only 1 of the family names and 2 of the non-family names. Assuming that the new parents chose one family name and one non-family name, what is the probability that the names they choose will make the new grandparents happy?

34. Stephanie has 11 different outfits in her closet that she wears on Sundays. Ann has 14 different outfits, of which 3 are the same as Stephanie's. What is the probability that they wear the same outfit on a particular Sunday?

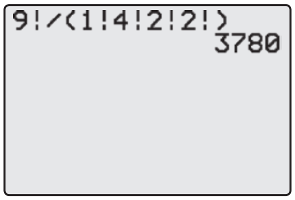
Since there are a total of 9 letters in TENNESSEE, substituting these values into the special permutation formula gives the following.

By Hand:

$$\frac{9!}{1!4!2!2!} = \frac{9 \cdot \overset{4}{\cancel{8}} \cdot 7 \cdot \overset{3}{\cancel{6}} \cdot 5 \cdot \cancel{4 \cdot 3 \cdot 2 \cdot 1}}{(1)\,\cancel{(4 \cdot 3 \cdot 2 \cdot 1)}\,\cancel{(2 \cdot 1)}\,\cancel{(2 \cdot 1)}}$$

$$= 9 \cdot 4 \cdot 7 \cdot 3 \cdot 5$$

$$= 3780$$



TI-83/84 Plus: Using the calculator we obtain the same result shown in the screenshot. Remember that you need parentheses around the calculation in the denominator.

Thus, there are 3780 ways to arrange the letters in the word TENNESSEE.

4.4 Section Exercises

Factorials

Evaluate each factorial expression.

1. $6!$ **2.** $8!$ **3.** $\frac{6!}{4!}$ **4.** $\frac{8!}{5!}$

5. $\frac{6!}{4!2!}$ **6.** $\frac{8!}{5!3!}$ **7.** $\frac{6!}{4!(6-4)!}$ **8.** $\frac{8!}{5!(8-5)!}$

9. $0!$ **10.** $\frac{7!}{0!}$

Combinations and Permutations

Evaluate each combination or permutation expression.

11. ${}_5C_2$ **12.** ${}_8C_5$ **13.** ${}_4C_4$ **14.** ${}_5C_1$

15. ${}_{12}P_1$ **16.** ${}_7\mathrm{P}_4$ **17.** ${}_5\mathrm{P}_3$ **18.** ${}_7\mathrm{P}_6$

19. ${}_8P_1$ **20.** ${}_3P_3$ **21.** $\frac{{}_3C_2}{{}_3P_1}$ **22.** $\frac{{}_{10}P_2}{{}_{10}C_2}$

23. ${}_6C_4 + {}_6C_3 + {}_6C_2 + {}_6C_1$ **24.** ${}_5P_4 + {}_5P_3 + {}_5P_2 + {}_5P_1$

Simplify the formula for each expression.

25. ${}_nC_n$ **26.** ${}_nC_1$ **27.** ${}_nP_1$ **28.** ${}_nP_n$

29. ${}_nP_{n-1}$ **30.** ${}_nC_{n-1}$

Use a combination or permutation expression to determine the total number of outcomes.

31. Farmer John has nine prize-winning cows. How many ways can he choose three of his cows to show at the state fair?

32. When a large group visits the Library of Congress, 3 members of the group are randomly chosen as "Library Ambassadors" for the day. How many ways can the ambassadors be chosen from a group of 38 visitors?

33. There are 12 members in the local garden club. In how many ways can a president and secretary be chosen? (Assume that no member can hold both positions at the same time.)

34. A teacher must choose parts for the upcoming Thanksgiving play from her class of 17 students. How many ways can she choose the parts of pilgrim, Native American, and turkey?

35. A teacher must choose parts for the upcoming Thanksgiving play from her class of 18 students. She needs a group of four students to serve as program attendants before the start of the play. In how many ways can this group be chosen?

36. There are eight people hosting a party. Three people are needed to stay and clean up after the party is over. How many ways can the clean-up crew be chosen?

37. There are eight people hosting a party. One person must set up the catering, another must bring flowers, and someone else needs to bring drinks. In how many ways can these tasks be assigned?

38. In assigning seats for a classroom, how many ways can a teacher place 8 students in the front row from her roll of 35?

39. In how many ways can a graduate student fulfill the master's degree requirements in mathematics if 10 classes are needed from a choice of 15 classes?

40. In how many ways can 1^{st}, 2^{nd}, and 3^{rd} place prizes be awarded in a local science fair if there are 25 participants?

41. In how many ways can a task force of 4 people be chosen from a group of 12 employees?

42. If 3 people need to serve as chaperones on a school trip, in how many ways can they be chosen from the parents of the 20 students? (Assume that each child has two parents available.)

43. The Seago family is planning their vacation. Each of the five family members is allowed to nominate three places they would like to visit. If they want to visit four different places during the trip, in how many ways can they plan their road trip, assuming that no family members choose the same place?

44. In how many ways can the letters in the word STATISTICS be arranged?

45. In how many ways can the letters in the word PROBABILITY be arranged?

46. Karran was born on 11/21/1992. He would like to make an eight-digit code from all of the digits in his birth date. How many different eight-digit codes could he create?

47. Employees at a local factory need a unique seven-digit code to access the building. The manager wants to make each person's code from the factory's phone number, 555-9313.

- **a.** If there are 509 employees who need codes, will the manager have enough unique codes using only the digits in the phone number?
- **b.** Would there be enough ten-digit codes if he used the area code, 514, as well?

48. Which of the following words would produce the greatest number of different five-letter arrangements?

- **a.** BEAST
- **b.** ORDER
- **c.** TESTS
- **d.** GOING

Counting Techniques and Probability

Use counting techniques to compute each classical probability.

49. At a carnival entrance, tickets are assigned five-digit numbers using the digits 0–9. If one ticket number is chosen randomly for a prize, what is the probability that every number on the ticket is even? (Assume that all possible ticket numbers are eligible to be chosen.)

50. Suppose that your boss must choose three employees in your office to attend a conference in Jamaica. Because all 20 of you want to go, he decides that the only fair way is to draw names out of a hat. What is the probability that you, Suzanne, and Alex are chosen?

51. Rhonda and Laura are planning to watch two movies over the weekend from Laura's collection of 24 DVDs. Rhonda has two favorites among the collection. What is the probability that the girls would randomly choose those two movies to watch?

52. Bill is planting tulip bulbs in the front yard. There are two white bulbs and two red bulbs mixed together in a bucket. What is the probability that Bill plants the four bulbs in a row so that they are alternating in color?

53. Nye is playing Scrabble. What is the probability that she chooses the tiles with the letters of her name, in order, when she draws three tiles from the bag? (Assume that when she begins there is one tile of each letter in the alphabet in the bag.)

54. Every 6 months, university email requires that a new 5-digit password be set up. No digits are allowed to be repeated and it must be different from your last two passwords. If you let your computer randomly choose a 5-digit code for you with no repeating digits, what is the probability that it will choose one of the last 2 passwords you've had?

4.5 Section Exercises

Combining Probability and Counting Techniques

Use the Fundamental Counting Principle, combination formula, permutation formula, or a combination of the methods to solve each problem.

1. Suppose that license tags in a particular area of the state must begin with one of the following letters: L, D, Y, or K. The rest of the tag must contain two letters followed by three digits (0–9).

a. If characters cannot repeat, how many unique tags can be made?

b. What's the probability that you are randomly assigned a license tag that begins with the letter K? (Assume that all possible license tags are available to be assigned.)

2. Suppose that license tags in another state are made up of three letters followed by three digits (0–9), none of which can repeat. Additionally, each tag number may be assigned to a regular tag, a wildlife-conservation tag, or a veterans tag.

a. How many unique tags can be made?

b. What's the probability that you are parked next to a car from this state that has a tag that ends in a 9?

3. **a.** How many four-digit numbers can be created from the digits 0–9 if the first and last digits must be odd and no digit can repeat?

b. What's the probability that a number that starts with 5 is randomly chosen from this group?

4. How many four-digit numbers can be made from the digits 0–9 if each number created must be greater than 5000?

5. A non-profit is required by statute to keep a 6 person board of advisors from the community. The board must be comprised of 2 student representatives and 4 local leaders.

a. How many advising boards can be formed if there are 13 applicants for the student positions and 19 applicants for the local leader positions?

b. If two brothers both applied for the student positions, what is the probability that they were both chosen?

6. Suppose an orchestra needs between 16 and 18 1st violins for an upcoming performance.

a. How many possible ways can the 1^{st} violin section be put together if there are 36 violinists able to play for the orchestra?

b. What is the probability that a 17-person 1^{st} violin section is chosen from the possible options found in part **a.**?

7. Virginia's Veggie Café offers 5 types of homemade bread, 10 toppings, and 6 different condiments. How many different super sandwiches can be made if a super sandwich consists of 6 different toppings and 2 different condiments?

8. Mysti is picking out material for her new quilt. There are 12 possible plaids, 8 different solids, and 4 floral prints that she can choose from. If she needs 3 plaids, 2 solids, and 2 floral prints for her quilt, how many different ways can she choose the materials?

9. Because Tristan has diabetes, he must make sure that his daily diet includes 2 vegetables, 3 fruits, and 2 breads. At the grocery store, he has a choice of 20 vegetables, 8 fruits, and 5 breads.

a. In how many ways can he make up his daily requirements if he doesn't like to eat 2 helpings of the same thing in one day?

b. What's the probability that a random choice from his possibilities would yield either carrots or spinach in its menu, given that carrots and spinach are both vegetable choices at the grocery store?

10. How many different teams of 4 can be chosen from a group of 20 adults and 15 children if each team must have at least 1 child on it?

11. A football coach needs to choose 11 players to start on offense. There are 6 freshmen, 6 sophomores, 8 juniors, and 7 seniors on the team. In how many ways can the starting 11 be chosen if the coach wants all 7 seniors to play?

12. Lindsay is checking out books at the library, and she is primarily interested in mysteries and nonfiction. She has narrowed her selections down to seven mysteries and eight nonfiction books.

a. How many different combinations of books can she check out if she is only allowed three books at a time?

b. How many different combinations of books can she check out if she is only allowed three books at a time, and she wants at least one mystery?

c. If she randomly chooses three books from her selections, what's the probability that they will all be mysteries?

13. In choosing what music to play at a charity fund-raising event, Marlow has 41 Mozart, 104 Haydn, and 8 Schubert symphonies from which to choose. He is setting up a schedule of the 12 songs to be played during the show.

a. How many different schedules are possible if he needs to have an equal number of symphonies from Mozart, Haydn, and Schubert?

b. If the songs are chosen randomly, what's the probability that all 12 symphonies are by Mozart?

14. A baseball coach needs to choose 9 players to be in the batting lineup for the first game of the season. There are 5 freshmen, 4 sophomores, 7 juniors, and 4 seniors on the team.

a. In how many ways can the batting order be chosen if the coach wants no more than 2 freshmen to play?

b. In how many ways can the batting order be chosen if the coach wants all 4 seniors to play?

What do these results tell us? Comparing the standard deviations, we see that not only does Plan B have a higher expected value, but its profits vary slightly less than those of Plan A. We may conclude that Plan B carries a slightly lower amount of risk than Plan A.

5.1 Section Exercises

Properties of a Probability Distribution

For each table, determine whether it could represent a valid discrete probability distribution. If not, explain why.

1.

x	$P(X = x)$
1	0.2
2	0.6
3	0.05
4	0.15
5	0.0

2.

x	$P(X = x)$
–22	0.4
53	1.05
–15	0.05

3.

x	$P(X = x)$
0.5	0.4
2.5	0.7
4.5	–0.3
6.5	0.2

4.

x	$P(X = x)$
–15	0.4
–20	0.3
–25	0.4

Discrete Probability Distributions

Create the probability distribution for each random variable described.

5. The number of tails showing when flipping four coins.

6. The number of even numbers showing when a pair of standard six-sided dice are rolled.

7. The difference between the two numbers showing when a pair of standard six-sided dice are rolled (largest value – smallest value).

8. The number of heads showing in five tosses of a coin.

Mean and Standard Deviation for Discrete Probability Distributions

For each discrete probability distribution, find the mean and the standard deviation.

9.

x	$P(X = x)$
15	0.6
22	0.4

10.

x	$P(X = x)$
–55	0.45
30	0.55

11.

x	$P(X = x)$
14	0.3
21.5	0.4
–2	0.3

12.

x	$P(X = x)$
–$1.50	0.3
$0.00	0.5
$2.75	0.1
$5.00	0.1

Expected Values

Determine the expected values for each scenario.

13. Scott likes to trade stocks online. On a good day, he averages a $2200.00 gain. On a bad day, he averages a $1600.00 loss. Suppose that he has good days 25% of the time, bad days 35% of the time, and the rest of the time he breaks even.

a. What is the expected value for one day of Scott's online trading?

b. If Scott trades online every weekday for three weeks, how much money should he expect to gain or lose?

c. What is the variance for one day of Scott's online trading?

14. Mike's older brother, Jack, bets him that he can't roll two dice and get doubles three times in a row. If Mike does it, Jack will give him $100.00. Otherwise, Mike has to give Jack $5.00.

a. What is the expected value of Mike's bet?

b. What is the expected value of Jack's bet?

c. If Mike and Jack make the same bet 30 times, how much can Mike expect to win or lose?

15. An insurance company offers Mississippi adults between the ages of 25 and 34 a $100,000 life insurance policy for $18 a month. They use the fact that Mississippi has a yearly death rate of 172.8 per 100,000 residents aged 25–34 years.

a. Find the expected value per customer for the insurance company at the end of one year for the policy described.

b. If the insurance company has 10,000 customers with these life insurance policies in Mississippi, what is its profit at the end of the year?

16. Suppose the same insurance company as in the previous question insures adults ages 25 to 34 in California for the same amount of money per month, but offers a $175,000 policy for that amount of money. The reason for the difference in the payout is that the death rate in California for that age group is 81.6 per 100,000 residents.

a. Find the expected value per customer for the insurance company at the end of one year for the policy described in California.

b. If the insurance company has 10,000 customers with these life insurance policies in California, what is its profit at the end of the year?

c. Which state is more profitable for the insurance company (as compared to Mississippi in the previous problem)?

17. A carnival game consists of choosing eight winning numbers, then randomly choosing one ball out of 50 balls, numbered 1 - 50. If the numbered ball you pick is one of your winning numbers, you win a $5 prize.

a. What is the expected value of the game if it costs $2 to play?

b. How much can you expect to win or lose if you play the game 10 times?

18. A church in town is raffling off $50.00. You can buy one ticket for $1.00, three tickets for $2.50, or five tickets for $4.00. Assume that the church sells 100 tickets.

a. Find the expected value for each of the three ticket options.

b. Should you buy one, three, or five tickets in order to maximize your expected winnings?

19. A department store is running a promotion one Saturday by giving out coupons for \$10 of free merchandise. Based on data collected in the past, only one-fourth of customers who shop on that Saturday use the coupon but do not purchase any other merchandise. However, one-third of customers purchase \$40 in merchandise and then use the coupon. Another one-third of customers use the coupon after ringing up a total of \$75 in merchandise. The remainder of customers who come in the store do not take advantage of the promotion at all.

a. Find the expected value of the promotion per customer for the department store.

b. If the store has 720 customers on the promotional Saturday, what is its expected revenue for the day?

20. A car dealership is offering an interesting incentive in order to get people to come and test drive its new sports cars. Everyone who agrees to a test drive gets to choose a key. There are 75 car keys in the bag, and 4 of them unlock a sports car. For the customers who choose a winning key, the dealership agrees to knock \$1000 off of the price if they buy a new car. (Assume that each key is returned after being drawn.)

a. From the perspective of the car dealership, what is the expected value of the incentive for one customer who chooses a key and buys a new car?

b. If 90 customers come in and choose a key and all of them buy new cars, how much can the dealership expect to give up in sales?

21. a. In the following probability distribution for the cost of textbooks in a fall semester for various liberal arts majors, each probability represents the chance that the total cost of a student's books will be the given amount. Find the expected value for the cost of books for the semester.

Cost of Textbooks for Liberal Arts Majors

x	$P(X = x)$
\$262	0.19
\$410	0.21
\$590	0.17
\$653	0.43

b. In the following probability distribution for the cost of textbooks in a fall semester for various business majors, each probability represents the chance that the total cost of a student's books will be the given amount. Find the expected value for the cost of books for the semester.

Cost of Textbooks for Business Majors

x	$P(X = x)$
\$378	0.35
\$389	0.14
\$392	0.25
\$401	0.26

c. Which of the groups of majors would you guess is more likely to feel that their textbooks are priced fairly? Explain your answer. (**Hint:** Use the standard deviations to help you make your *informed* decision.)

22. The managing director of a traveling carnival needs to add a new game to the carnival lineup. Given below are the probability distributions for his top two choices. The values of the random variable are the amounts the carnival would either gain (positive values) or have to pay out (negative values). Which game would you advise the director to choose? Why?

Carnival Game 1

x	$P(X = x)$
$1.00	0.8
–$5.00	0.135
–$10.00	0.06
–$100.00	0.005

Carnival Game 2

x	$P(X = x)$
$1.00	0.65
–$1.00	0.25
–$2.00	0.1

is that the precise probabilities would not actually change enough to affect the value of the answer. We are considering eight programs; thus, $n = 8$. If we define a success to be receiving a discount coupon, then the probability of obtaining a success is $p = 0.2$. Let X be the number of discount coupons received in the eight programs bought by your friends. We are interested in the probability that *at least half of the eight* friends get a discount coupon, so at least four out of the eight, or $P(X \geq 4)$. As in the previous example, in order to use the cumulative binomial probability function to solve this problem, we will need to use the Complement Rule.

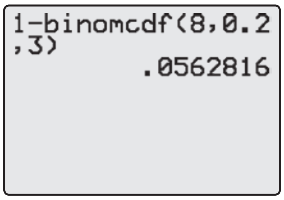
$$P(X \geq 4) = 1 - P(X < 4)$$

This is still not exactly what we need because the cumulative binomial distribution is only cumulative probabilities of the form $P(X \leq x)$. Fortunately, this situation is not too difficult to deal with due to one of the characteristics of the binomial distribution. The value for x must be a whole number; therefore, $P(X < 4) = P(X \leq 3)$.

Tables: Find the row for $n = 8$, $x = 3$, and the column for $p = 0.2$. The probability where they intersect is 0.9437. Thus,

$$\begin{aligned} P(X \geq 4) &= 1 - P(X \leq 3) \\ &= 1 - 0.9437 \\ &= 0.0563 \end{aligned}$$

TI-83/84 Plus: Using all of this information we calculate the probability as shown below and in the screenshot in the margin.



$$\begin{aligned} P(X \geq 4) &= 1 - P(X < 4) \\ &= 1 - P(X \leq 3) \\ &= \texttt{1-binomcdf(8,0.2,3)} \\ &\approx 0.0563 \end{aligned}$$

Therefore, the probability that at least half of the eight friends find discount coupons in their programs is approximately 0.0563, or 5.63%, which indicates that it is not very likely.

5.2 Section Exercises

Properties of a Binomial Distribution

Determine whether the given procedure meets the criteria of a binomial distribution. If not, identify at least one requirement that is not satisfied.

1. A survey of college students rating the food in the campus dining hall on a scale from 1–10.
2. Drawing a card from a standard deck of 52 cards and recording whether the card drawn is a face card, number card, or an ace.
3. The number of times a professional baseball player hits a home run each time at bat.
4. Surveying 124 people living in the United States who use Internet service and recording their "no" responses to the question: "Do you think that Internet sites should be federally regulated?"

Probability for Binomial Distributions

Assume that the random variable X has a binomial distribution with the given probability of obtaining a success. Find each specified probability, given the number of trials.

5. $P(X=3),\quad n=5,\quad p=0.4$

6. $P(X=4),\quad n=10,\quad p=0.3$

7. $P(X\le 8),\quad n=12,\quad p=0.1$

8. $P(X\le 2),\quad n=3,\quad p=0.9$

9. $P(X<7),\quad n=18,\quad p=0.4$

10. $P(X<9),\quad n=17,\quad p=0.5$

11. $P(X>3),\quad n=4,\quad p=0.8$

12. $P(X>5),\quad n=10,\quad p=0.7$

13. $P(X\ge 6),\quad n=7,\quad p=0.2$

14. $P(X\ge 8),\quad n=15,\quad p=0.6$

Find each specified probability for the given scenario.

15. Suppose that the probability of Thad making a free throw in the championship basketball game is 60%, and each throw is independent of his last throw. Assume that Thad attempts seven free throws during the game.

a. What is the probability that he will make more than four of his free throws?

b. What is the probability that he will make all of his free throws?

c. How many free throws should we expect Thad to make of the seven attempts during this game?

16. At one large university, freshmen account for 30% of the student body.

a. If a group of twelve students is randomly chosen by the school newspaper to comment on textbook prices, what is the probability that fewer than three of the students are freshmen? (Assume that this situation can be modeled using a binomial distribution.)

b. If a group of ten students is randomly chosen by the school newspaper to comment on textbook prices, what is the probability that more than three of the students are freshmen? (Assume that this situation can be modeled using a binomial distribution.)

c. If a group of 40 students is randomly chosen by the school newspaper to comment on textbook prices, how many of the students can you expect to be freshmen?

17. The SugarBear Candy Factory makes two types of chocolate candy bars—milk chocolate and milk chocolate with almonds. In a typical day, 60% of the candy bars being made are milk chocolate, and the rest are milk chocolate with almonds. At the end of the day, a quality control expert randomly chooses 14 chocolate bars for inspection.

a. What is the probability that half of them contain almonds?

b. What is the probability that at least half of them contain almonds?

c. Of the 14 chocolate bars chosen for inspection, what is the variation in the number of bars containing almonds that the inspector could get?

18. Suppose that the probability of your favorite baseball player getting a hit at each at-bat is 0.30. Assume that each at-bat is independent of any other at-bat.

a. What is the probability that he bats six times and gets fewer than two hits?

b. What is the probability that he bats nine times and gets at most four hits?

c. In a three-game series where the player bats a total of 10 times, what is the variation in the number of hits he might get?

19. Suppose that Carlos is taking a multiple-choice test where there are five answer choices for each question, and he randomly guesses on four questions.

a. What is the probability that he gets exactly three of those questions correct?

b. What is the probability that he gets at least one out of the four questions correct?

c. What is the probability that he gets none of the four questions correct?

20. In a pediatrician's office, the probability of a "no show" for any checkup appointment on any given day is 1 out of 10. Suppose that there are 18 appointments scheduled for one day.

a. Find the probability that fewer than 4 don't show.

b. Find the probability that at least 2 don't show.

c. Find the probability that the doctor sees every patient scheduled.

21. The probability of any plant surviving in Kerry's garden is 0.8. Suppose she plants 19 new plants this year.

a. What is the probability that at least 5 of them survive?

b. What is the probability that more than $\frac{3}{4}$ of them survive?

c. What is the probability that she has a bad year and none of them survive?

22. In a national park in Alaska, there are 100 polar bears. As part of the monitoring of the park, the rangers caught 20 bears, tagged them and released them back into the park during the first year. A year later they caught 13 bears. (Assume that the bears are caught at different times, and the same bear could be caught more than once.)

a. What is the probability that 6 are tagged?

b. What is the probability that at least 10 are tagged?

c. What is the probability that none are tagged?

23. Underneath each cap of Brand X cola bottles is a chance to win a free cola. Suppose that the probability of winning is 1 out of 11 and you buy 16 colas.

a. What is the probability that you win at least once?

b. What is the probability that you don't win at all?

c. What is the probability that you win with half of the bottles?

24. Ronnie owns a fireworks stand and knows that in the fireworks business, 1 out of every 13 fireworks is a dud. Suppose that Juanita buys 10 firecrackers at Ronnie's stand.

a. What is the probability that no more than 3 are duds?

b. What is the probability that she has a perfect display with no duds?

c. What is the probability that more than half are duds?

5.3 Section Exercises

Determining the Value of λ for a Poisson Distribution

Determine the mean, which is the value of λ, for each scenario.

1. A fifth-grader averages three grammatical errors per paragraph. What is his average for four paragraphs?

2. A surveillance officer reports two incidents of shoplifting per month, on average. How many incidents of shoplifting on average are reported per year?

3. Baggage handlers at a certain airport move an average of 1500 bags on an eight-hour shift. On average, how many bags are moved per hour?

4. An assembly line, on average, produces 1 defective part for every 100 parts that roll off the line. What is the average number of defects for a group of 20 parts?

5. You average 70 heartbeats per minute. What is the average number of heartbeats you have in 10 seconds?

6. Someone with mild sleep apnea, a condition where the person stops breathing for 10 seconds or longer during sleep, has 10 such episodes where their breathing stops per hour. If on a typical night they sleep for 7 hours, what is the average number of episodes that they have per night?

Probability for Poisson Distributions

Find each specified probability. Assume that the random variable X has a Poisson distribution with the given value of λ.

7. $P(X=2), \quad \lambda=1.80$

8. $P(X=0), \quad \lambda=2.60$

9. $P(X=13), \quad \lambda=6.50$

10. $P(X=8), \quad \lambda=9.30$

11. $P(X\le 2), \quad \lambda=3.80$

12. $P(X\le 5), \quad \lambda=7.10$

13. $P(X<4), \quad \lambda=8.30$

14. $P(X<8), \quad \lambda=2.90$

15. $P(X\ge 3), \quad \lambda=4.90$

16. $P(X\ge 12), \quad \lambda=9.60$

Find each specified probability for the given scenario. Assume that each scenario can be modeled by a Poisson distribution.

17. The Oxford Gift Shop averages four sales each hour. Betty is scheduled to work the cash register from 1:00–3:00 on Saturday afternoon.

a. What is the probability that Betty rings up exactly ten customers?

b. What is the probability that Betty rings up more than ten customers?

18. The Pancake House is so popular that it boasts of selling a stack of pancakes every two minutes.

a. What is the probability that there is a ten-minute interval in which no pancakes are sold?

b. What is the probability that there is a five-minute interval in which fewer than three stacks of pancakes are sold?

19. The pizza place next to the local college receives an average of 20 pizza orders per hour during lunch.

a. In any given hour during lunch, what is the probability that the pizza place receives at least 22 pizza orders?

b. In any given hour during lunch, what is the probability that the pizza place receives more than 22 pizza orders?

20. Rob is a busy physician in the emergency room. He sees an average of four major trauma patients each night.

- **a.** What is the probability that fewer than three major trauma patients will be admitted on any given night?
- **b.** What is the probability that no more than five major trauma patients will be admitted on any given night?

21. Suppose that a bank drive-through serves customers at a rate of twelve cars every hour.

- **a.** What is the probability that the bank drive-through will serve fewer than five customers in 30 minutes?
- **b.** What is the probability that the bank drive-through will serve three customers in 15 minutes?

22. On average, Patrick sees a spider in his home once a month.

- **a.** What is the probability that Patrick sees two spiders in a month and a half?
- **b.** What is the probability that Patrick sees no more than one spider in a month and a half?

23. Jane cannot sell a finished piece of pottery if she discovers that the clay has a defect in it. Suppose that she has to discard 2 pieces for every 56 pieces she makes.

- **a.** What is the probability that in 14 pieces of pottery, just 1 piece is defective?
- **b.** What is the probability that in 28 pieces of pottery, at least 1 piece is defective?

24. An experienced transcriber misspells only 2 out of every 100 words on average. The transcriber is currently transcribing a 1000 word essay.

- **a.** What is the probability that the transcriber will misspell 15 words?
- **b.** What is the probability that the transcriber will misspell no more than 10 words?

25. Suppose that, on average, 45 books are checked out of the local public library per day.

- **a.** What is the probability that 100 books are checked out in two days?
- **b.** What is the probability that at most 200 books are checked out in four weeks?

26. A landscape architect knows that in the cable that he lays for landscape lighting he can expect one defect in 300 yards of cable.

- **a.** What is the probability that in 100 yards of cable, he would find two defects?
- **b.** What is the probability that in 200 yards of cable, he would find fewer than three defects?

b. In this scenario, there is a fixed number of trials, namely $n = 250$. These trials are independent since where one applicant lives does not affect where the next applicant lives. For each applicant, there are only two possible outcomes for where the applicant lives: in-state or not in-state. Thus, a binomial distribution should be used.

c. In the last scenario, there is a fixed number of trials, namely $n = 50$. However, these trials are dependent since the probability of getting a coupon changes with each program that is handed out. With a fixed number of dependent trials, use a hypergeometric distribution.

5.4 Section Exercises

Probability for Hypergeometric Distributions

Find each specified probability for the given scenario. Assume that every scenario follows a hypergeometric distribution.

1. In a standard deck of 52 cards, 13 are hearts. Assume that 5 cards are selected without replacement out of a well-shuffled deck.

a. What is the probability of getting exactly 2 hearts?

b. What is the probability that all 5 cards will be hearts?

c. What is the average number of hearts you would be dealt in five cards?

2. Suppose that one Christmas, Abby and Andrew's mother forgot to label their gifts. Out of ten wrapped presents, five are for Abby and five are for Andrew.

a. What is the probability that exactly one of the first four presents opened is for Abby?

b. What is the probability at most three of the first five gifts opened are for Andrew?

c. Of the first four presents opened, what is the average number that would be for Abby?

3. Suppose that 12 of the 20 azaleas for sale at a large nursery have pink flowers and the rest have red flowers. Because it is early in the season, they have not begun to bloom and you cannot yet tell what color each plant will be.

a. If eight azaleas are chosen at random without replacement, what is the probability that exactly 6 will be pink?

b. If five azaleas are chosen at random without replacement, what is the probability that none of the azaleas will be pink?

c. Calculate the variance for this hypergeometric distribution in part b.

4. Eloise loves jelly beans, and the yellow ones are her favorite. One afternoon she is snacking on a bag of 18 jelly beans, 5 of which are yellow. She grabs a handful of 6 jelly beans.

a. What is the probability that more than half of the jelly beans in her hand are yellow?

b. What is the probability that fewer than 2 of the jelly beans in her hand are yellow?

c. Calculate the variance for this hypergeometric distribution.

5. The manager of a furniture store has just received a shipment of sofas and recliners. He knows that the order contains five sofas and nine recliners.

a. What is the probability that the first three items brought into the store are recliners?

b. What is the probability that out of the first seven items brought into the store, no more than two are sofas?

6. Suppose Audrey received a box of chocolates for Valentine's Day. Just after opening the box, she lost the paper which had the description of each chocolate on it. However, she knows that there were six truffles and five caramel candies left.

a. What is the probability that the first two chocolates she eats are both truffles?

b. What is the probability that at least one of the first three chocolates she eats is caramel?

7. Karen has 20 squares of material to use for her quilt; 8 are polka-dotted and the rest are floral.

a. What's the probability that she randomly uses all floral squares for the first 5 pieces of the quilt?

b. What's the probability that 2 out of the first 6 pieces she randomly chooses are polka-dotted?

8. Jay has ten pieces of mail to open, four of which are junk mail.

a. What is the probability that he randomly opens two pieces of mail and they are both junk mail?

b. What is the probability that he randomly opens three pieces of mail and at least two of them are junk mail?

9. Grab bags at the town festival are filled with either a coupon for a free hamburger or a coupon for a free order of french fries. Suppose there are 22 hamburger coupons left and 18 french-fry coupons left.

a. What is the probability that you and your friend both get bags with hamburger coupons in them?

b. Given that you both got hamburger coupons, what is the probability that if you each choose again, you both get french-fry coupons?

10. Suppose there are eight green tags, twelve white tags, and four red tags left to use as name tags at a conference. Tags are given out randomly at the registration desk.

a. What is the probability that the first two tags given out are red?

b. What is the probability that if part a. did happen, then more than three out of the next five tags would be white?

11. An antiques dealer has fifteen antique cedar chests for sale. Unknown to the dealer, one of these cedar chests is actually a modern reproduction. If the dealer randomly chooses three of these cedar chests for display, what is the probability that one of the cedar chests on display will be the reproduction?

12. Of the 30 pairs of size 8 jeans on the display table at a local retail store, 2 of the pairs are incorrectly labeled with the wrong size. A customer comes into the store and randomly chooses 5 pairs of jeans from the table, believing them all to be the same size.

a. What is the probability that all 5 pairs of jeans that the customer chose are the correct size?

b. What is the probability that 1 of the 5 pairs of jeans that the customer chose is incorrectly labeled?

c. What is the probability that both of the incorrectly labeled pairs of jeans are in the 5 pairs that the customer chose?

Determining Which Discrete Probability Distribution to Use

Each of the following exercises can be solved using the binomial distribution, the Poisson distribution, or the hypergeometric distribution. Begin by stating which distribution to use, and then solve the problem.

13. A bank receives an average of 8 bad checks per day. What is the probability that during a five-day bank week, the bank will receive 50 bad checks?

14. The IRS is considering auditing 250 tax returns within a given region, 15 of which have serious errors (unknown to the IRS, of course). If an IRS agent randomly chooses 25 of these 250 tax returns on which to perform an audit, what is the probability that 3 of these tax returns will have serious errors?

15. It is commonly stated that one out of every two marriages will end in divorce. Assuming this is true, what is the probability that of ten randomly selected married couples, seven of the couples' marriages will end in divorce?

16. According to records from a large public university, 85% of students who graduate from the university successfully find employment in their chosen field within six months of graduation. What is the probability that of eight randomly selected students who have graduated from this university, at least five of them find employment in their chosen field within six months?

17. Suppose that one of a local breeder's ten newborn poodles has a genetic birth defect that will only appear later in life. Due to the high expense of the testing, the breeder randomly chooses only three puppies to test for genetic birth defects. What is the probability that one of the puppies will test positive for a genetic birth defect?

18. If a child welfare office receives an average of five reports of child abuse per day, what is the probability that the child welfare office will have a slow day and receive no more than two reports?

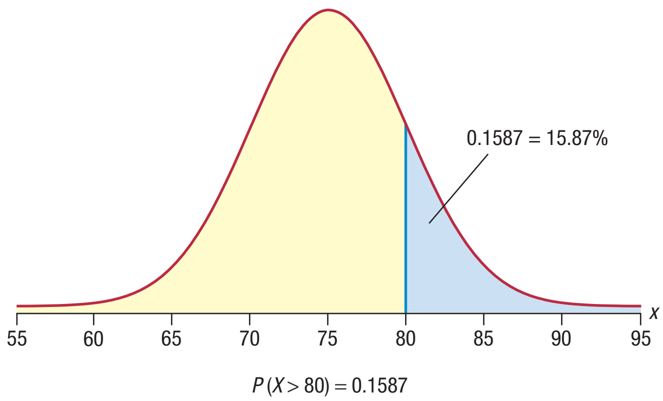


Figure 6.1.9 Normal Distribution of Labrador Retriever Weights (in Pounds)

In order to apply this powerful connection between the area under a normal curve and probability, we need to be able to calculate the area under a normal curve in any given region. The next two sections look at how to calculate the area under a given normal curve and hence use it to identify certain probabilities.

6.1 Section Exercises

Properties of Normal Distributions

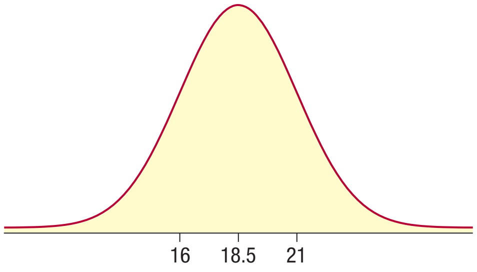
Decide if each statement is true or false. Explain why.

1. There are a limited number of normal distributions.
2. The total area under a normal curve changes depending on the height of the mean of the curve.
3. The mean of a normal distribution is always 0.
4. A normal distribution with a small standard deviation is "flatter" than one with a large standard deviation.
5. The area under any region of a normal curve is equal to the probability that the random variable will fall within that region.
6. For any normal distribution, the mean, median, and mode are equal.
7. The line of symmetry for a normal distribution is $x = \mu$.
8. The y-axis is a vertical asymptote for all normal distributions.
9. The inflection points for any normal distribution are one standard deviation on either side of the mean.
10. Normal distributions are symmetric, but they do not have to be bell-shaped.

Identifying Properties of Normal Curves

Use the normal distribution curves to answer the questions.

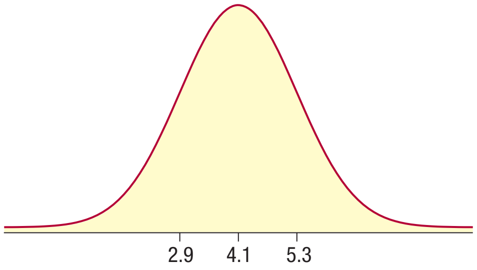
11. Use the normal distribution curve in the figure to answer the following questions.



a. Approximate the mean of the distribution.

b. Approximate the standard deviation of the distribution.

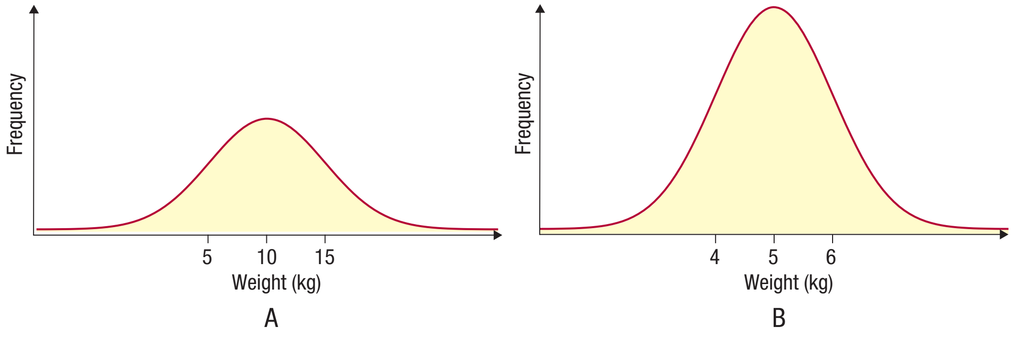
12. Use the normal distribution curve in the figure to answer the following questions.



a. Approximate the mean of the distribution.

b. Approximate the standard deviation of the distribution.

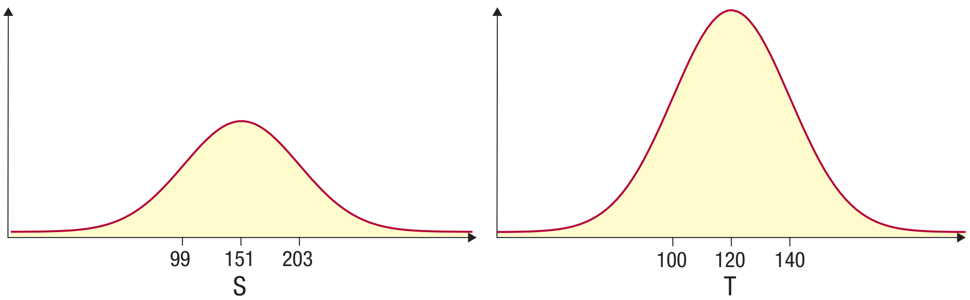
13. Use the normal distribution curves in the figure to answer the following questions.



a. Which distribution has the larger standard deviation?

b. Which distribution has the larger mean?

14. Use the normal distribution curves in the figure to answer the following questions.



a. Which distribution has the smaller standard deviation?

b. Which distribution has the smaller mean?

Drawing Normal Distributions

Draw a normal curve with the given characteristics. Indicate the value of the mean and the location of the given x-value on the x-axis, and indicate the inflection points using vertical lines extending from the appropriate values on the x-axis to the inflection points on the curve.

15. $\mu = 65$ and $\sigma = 20$; $x = 40$

16. $\mu = 5.00$ and $\sigma = 0.25$; $x = 4.80$

17. $\mu = 10.75$ and $\sigma = 0.35$; $x = 11.15$

18. $\mu = 15$ and $\sigma = 2$; $x = 19$

19. $\mu = 0.023$ and $\sigma = 0.001$; $x = 0.020$

20. $\mu = 12{,}000$ and $\sigma = 2000$; $x = 9{,}750$

Comparing Normal Distributions

Draw two normal distributions, on the same x-axis, that have the given characteristics.

21. Population means that differ by 5 units

22. The same mean but different standard deviations

Determining if a Distribution is Normal

Determine if the following distributions could be classified as normal distributions based on the information given . Explain your answer.

23. The distribution of long jump distances in the NCAA Outdoor Track and Field competition

24. The distribution of scores on a college entrance exam based on whole numbers 1-33

25. The distribution of the number of sick days employees took at Viant & Company over the past 10 years

26. The distribution of daily water temperatures for each of the Great Lakes

6.2 Section Exercises

Converting to the Standard Normal Distribution

Calculate the standard score of the given x-value. Indicate where the z-value would be on the standard normal distribution.

1. $\mu = 65$ and $\sigma = 20$; $x = 40$

2. $\mu = 5.00$ and $\sigma = 0.25$; $x = 4.80$

3. $\mu = 15$ and $\sigma = 2$; $x = 19$

4. $\mu = 10.75$ and $\sigma = 0.35$; $x = 11.15$

5. $\mu = 0.023$ and $\sigma = 0.001$; $x = 0.020$

6. $\mu = 12{,}000$ and $\sigma = 2000$; $x = 10{,}750$

Using the *z*-Score Formula

Use the z-score formula to complete each exercise.

7. Find the missing value in each row of the table.

	z	x	μ	σ
a.	?	35.0	37.0	1.6
b.	1.75	12.3	?	2.8
c.	–3.10	?	3.40	0.20
d.	2.15	479	436	?

8. Find the missing value in each row of the table.

	z	x	μ	σ
a.	?	2.87	2.45	0.21
b.	–2.80	579.0	?	8.5
c.	1.40	?	89.10	7.45
d.	–3.20	13.67	15.11	?

9. Write a formula for x in terms of the population mean, population standard deviation, and z-score.

Area under the Standard Normal Curve

Find the area under the standard normal curve to the left of the given z-value.

10. $z = 2.35$

11. $z = 1.78$

12. $z = -1.25$

13. $z = -0.19$

14. $z = 2$

15. $z = 1.3$

16. $z = 3.57$

17. $z = -4.12$

Find the area under the standard normal curve to the right of the given z-value.

18. $z = 1.35$

19. $z = 1.7$

20. $z = -2.51$

21. $z = -0.39$

22. $z = -1$

23. $z = 3.68$

Find the area under the standard normal curve between the given z-values.

24. $z_1 = 0.35, \quad z_2 = 1.85$

25. $z_1 = -1.25, \quad z_2 = 2.16$

26. $z_1 = -1.78, \quad z_2 = -0.95$

27. $z_1 = -0.19, \quad z_2 = 1$

28. $z_1 = -3.57, \quad z_2 = 1.85$

29. $z_1 = 1.51, \quad z_2 = 2.61$

30. Find the area under the standard normal curve between $z_1 = -1.00$ and $z_2 = 1.00$. What is the difference between the amount you found for this area and the amount the Empirical Rule approximates for this area? (See Section 3.2 to review the Empirical Rule.)

31. Find the difference between the more precise value found using the normal distribution tables or technology and the approximation given by the Empirical Rule for the area under the standard normal curve within two standard deviations of the mean. Do the same for the area within three standard deviations of the mean. (See Section 3.2 to review the Empirical Rule.)

Find the total of the areas under the standard normal curve to the left of z_1 and to the right of z_2.

32. $z_1 = -1.46, \quad z_2 = 1.46$

33. $z_1 = -2.11, \quad z_2 = 2.11$

34. $z_1 = -3.05, \quad z_2 = 3.05$

35. $z_1 = -2.31, \quad z_2 = 1.67$

36. $z_1 = -1.75, \quad z_2 = 1.89$

37. $z_1 = -2, \quad z_2 = 1$

38. $z_1 = -3.81, \quad z_2 = 2.37$

39. $z_1 = 1.31, \quad z_2 = 1.93$

40. $z_1 = 0.35, \quad z_2 = 1.75$

41. $z_1 = -2.31, \quad z_2 = -1.67$

42. $z_1 = -3, \quad z_2 = -2$

43. $z_1 = 1.51, \quad z_2 = 2.61$

Probability for the Standard Normal Distribution

Find each specified probability.

44. $P(z < -3.14)$

45. $P(z < 1.43)$

46. $P(z > 2.72)$

47. $P(z > -0.81)$

48. $P(-1.86 < z < 3.14)$

49. $P(0.78 < z < 2.64)$

50. $P(0 < z < 2.78)$

51. $P(-2.81 < z < -1.14)$

52. $P(z < -1.26 \text{ or } z > 1.26)$

53. $P(z < -2.39 \text{ or } z > 2.39)$

54. Find the probability that z differs from the mean by more than one standard deviation.

55. Find the probability that z differs from the mean by more than two standard deviations.

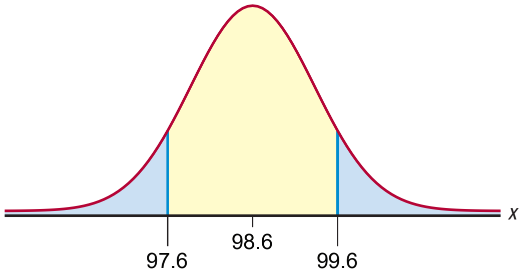
56. Find the probability that z differs from the mean by less than two standard deviations.

57. Find the probability that z differs from the mean by less than 1.5 standard deviations.

58. $P(z < 0.85 \text{ or } z > -2.34)$

59. $P(-4.92 < z < 3.68)$

60. $P(z > 4.08)$



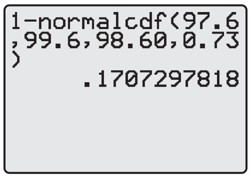
Now, using either tables or available technology find the value of the shaded area.

Tables: Convert both temperatures to standard scores and then use the cumulative probability tables.

For 97.6 °F: $z_1 = \frac{x-\mu}{\sigma} = \frac{97.6-98.60}{0.73} \approx -1.37$

For 99.6 °F: $z_2 = \frac{x-\mu}{\sigma} = \frac{99.6-98.60}{0.73} \approx 1.37$

Notice that these z-scores are equal distances from the mean. Because of the symmetry of the curve, the shaded areas must then be equal. The easiest way to calculate the total area is to find the area to the left of $z_1 \approx -1.37$ and simply double it. We then have $(0.0853)(2) = 0.1706$.



TI-83/84 Plus: The area in the tails can be found in one step on the calculator by subtracting the area between the two z-values from one. Therefore, the probability can be found by entering `1-normalcdf(lower bound, upper bound, μ,σ)`. For our question, these values are

lower bound: 97.6
upper bound: 99.6
$\mu = 98.6$
$\sigma = 0.73$

As shown in the margin screenshot, the calculator gives the more accurate value of 0.1707.

The probability would be reported as 0.1706 if using the the tables or 0.1707 if using the calculator. Thus, the probability of a healthy adult having a body temperature that differs from the population mean by more than 1°F is approximately 17%.

6.3 Section Exercises

Probability for Normal Distributions

Complete the exercises for the given scenario. Round your answer to two decimal places for a percentage and four decimal places for a probability.

1. Recently, birth weights of Norwegians were reported to be normally distributed with a mean of 3668 grams (g) and a standard deviation of 511 g. Suppose that a Norwegian baby was chosen at random.

- **a.** Find the probability that the baby's birth weight was less than 4000 g.
- **b.** Find the probability that the baby's birth weight was greater than 3750 g.
- **c.** Find the probability that the baby's birth weight was between 3000 g and 4000 g.
- **d.** Find the probability that the baby's birth weight was less than 2650 g or greater than 4650 g.

2. Systolic blood pressure is normally distributed with a mean of 113.8 and a standard deviation of 10.8. Suppose a person was chosen at random.

a. Find the probability that a randomly selected person will have a systolic blood pressure of less than 124.

b. Find the probability that a randomly selected person will have a systolic blood pressure of more than 126.5.

c. Find the probability that a randomly selected person will have a systolic blood pressure between 93 and 130.

d. Find the probability that a randomly selected person will have a systolic blood pressure of less than 95 or greater than 120.

3. Data collected on the total number of minutes people spent in the local emergency room revealed that patients spent on average 166.9 minutes in the ER, with a standard deviation of 55.4 minutes. Assume that the times collected follow a normal distribution.

a. What percentage of the ER patient population spent less than 3 hours there?

b. What percentage of the ER patient population spent more than 2 hours there?

c. What percentage of the ER patient population spent between 2.5 hours and 3.5 hours there?

d. What percentage of the ER patient population spent either less than 2.5 hours or more than 3.5 hours there?

4. Deer hunters score the antlers of their quarry in order to compare the most spectacular specimens. White-tailed bucks are given a score based on lengths and circumferences of their antlers. These scores are normally distributed with a mean of 133 and a standard deviation of 18.

a. What is the probability that a random white-tailed buck in a herd has an antler score of more than 150?

b. What is the probability that a random white-tailed buck in a herd has an antler score of less than 160?

c. What is the probability that a random white-tailed buck in a herd has an antler score of between 130 and 140?

d. What is the probability that a random white-tailed buck in a herd has an antler score of less than 100 or greater than 170?

5. Suppose that the mean calorie intake is 2050 calories per day, with a standard deviation of 175 calories. Assume that calorie intakes follow a normal distribution.

a. Find the probability that a peer in your class consumes more than 1800 calories per day.

b. Find the probability that a randomly selected stranger consumes fewer than 1500 calories per day.

c. Even though the "freshman fifteen" is a common occurrence, for a freshman to put on 15 pounds in one semester they would have to consume between 2100 and 2600 calories per day. Find the probability that a randomly chosen freshman would consume this many calories.

6. The total blood cholesterol levels in a certain Mediterranean population are found to be normally distributed with a mean of 160 milligrams/deciliter (mg/dL) and a standard deviation of 50 mg/dL. Researchers at the National Heart, Lung, and Blood Institute consider this pattern ideal for a minimal risk of heart attacks.

a. Find the percentage of this population who have blood cholesterol levels less than 150 mg/dL.

b. Find the percentage of this population who have blood cholesterol levels which exceed the ideal level by at least 10 mg/dL.

c. Find the percentage of this population who have blood cholesterol levels between 150 and 200 mg/dL.

d. Find the percentage of this population who have blood cholesterol levels less than 100 mg/dL or greater than 220 mg/dL.

7. Suppose that motorists in the southeastern United States use a mean of 8.20 gallons of gasoline per week with a standard deviation of 0.47 gallons. Assume that gasoline consumption levels are approximately normally distributed.

a. What percentage of southeastern drivers use more than 9.0 gallons of gasoline per week?

b. What percentage of southeastern drivers use no more than 7.0 gallons of gasoline per week?

c. What percentage of southeastern drivers use an amount of gasoline that differs from the mean by more than 0.5 gallons per week?

8. Abby is on very strict sugar diet and is trying to account for all of the grams of sugar she intakes in a week. For a snack, she loves to eat raw carrots and knows that a carrot of average length contains approximately 3.5g of sugar. Answer the following questions assuming that the lengths of carrots are normally distributed with a mean of 17.78 cm and a standard deviation of 3.81 cm.

a. Find the probability that a randomly chosen carrot is less than 10 cm long.

b. Find the probability that in a bag of raw carrots, Abby randomly chooses a carrot which is shorter than the average carrot.

c. Find the probability that in a bag of raw carrots, Abby randomly eats a carrot with more than 3.5 grams of sugar.

d. Suppose Abby randomly selects 2 carrots to eat. Find the probability that both will have at most 3.5 grams of sugar. (Hint: See Chapter 4 for probability of 2 events occuring)

9. Assume for the moment that the distribution of weights of adults is approximately normal with a mean of 179.8 lb and a standard deviation of 93.0 lb.

a. Find the probability that a randomly selected adult would have a weight at least two standard deviations above the mean.

b. Find the probability that a randomly selected adult would have a weight at least two standard deviations below the mean.

c. Given your answers in parts **a.** and **b.**, do you think it is reasonable to assume that weights of adults follow this normal distribution? Explain your answer.

10. Suppose that lifetimes for a particular car battery are normally distributed with a mean of 148 weeks and a standard deviation of 8 weeks.

a. If the company guarantees its battery for 3 years, what percentage of the batteries sold would you expect to be returned before the end of the warranty period? Assume that there are 52 weeks in a year.

b. Imagine you were the CEO of the battery company. Evaluate the warranty offer and list any changes you would make as the CEO.

11. Ella refuses to tell you her weight in her ninth month of pregnancy. However, she does tell you that her weight is above the mean. Which of the following *z*-scores is possible for her weight?

a. –0.08

b. 1.43

c. 0

d. Not enough information

12. The mean distance a new sales representative travels per week in the first year is 300 miles. At Olivia's job interview, she found out that she would travel at most the mean distance per week for first-year sales reps. Which of the following *z*-scores are possible for her mean weekly traveling distance during her first year?

a. 0

b. −1.42

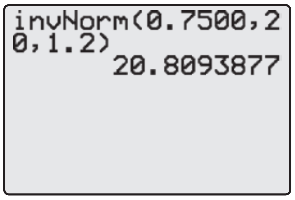
c. 0.78

d. Not enough information

baby. We can use the formula for x that we saw previously.

$$\begin{aligned} x &= z \cdot \sigma + \mu \\ &= (0.67)(1.2) + 20.0 \\ &\approx 20.8 \end{aligned}$$

TI-83/84 Plus: We can find the value of the normally distributed variable in one step using the function syntax `invNorm(area,μ,σ)`, where *area* is the area to the left.



Here we have

$$\begin{aligned} area &= 0.7500 \\ \mu &= 20.0 \\ \sigma &= 1.2 \end{aligned}$$

Enter `invNorm(Ø.75ØØ, 2Ø, 1.2)`, as shown in the screenshot in the margin. This gives a value of 20.8 rounded to one decimal place.

Therefore, the minimum length a baby can be and still be in the top 25% of lengths of full-term newborn babies is approximately 20.8 inches.

6.4 Section Exercises

Finding the *z*-Value That Corresponds to a Given Area

Find the indicated value of z.

1. What z-value has an area of 0.0038 to its left?
2. What z-value has an area of 0.9803 to its left?
3. What z-value has an area of 0.9738 to its left?
4. What z-value has an area of 0.0212 to its left?
5. What z-value represents the 95^{th} percentile?
6. What z-value represents the 30^{th} percentile?
7. What value of z is the 25th percentile?
8. What z-value represents the 75^{th} percentile?
9. What z-value has an area of 0.0838 to its right?
10. What z-value has an area of 0.0049 to its right?
11. What z-value has an area of 0.9706 to its right?
12. What z-value has an area of 0.5987 to its right?
13. Find the value of z such that the area between $-z$ and z is 0.99.
14. Find the value of z such that the area between $-z$ and z is 0.80.
15. Find the value of z such that the area between $-z$ and z is 0.98.
16. Find the value of z such that the area between $-z$ and z is 0.95.
17. Find the value of z such that the area to the left of $-z$ plus the area to the right of z is 0.5686.
18. Find the value of z such that the area to the left of $-z$ plus the area to the right of z is 0.7642.
19. Find the value of z such that the area to the left of $-z$ plus the area to the right of z is 0.0286.
20. Find the value of z such that the area to the left of $-z$ plus the area to the right of z is 0.0040.
21. What z-value represents the third quartile?
22. What z-value represents the first quartile?

Applications

Answer each question for the given scenario.

23. If a normal distribution has a mean of 95.0 and a standard deviation of 8.8, what is the value of the random variable X that has an area to its right equal to 0.0526?

24. If a normal distribution has a mean of 38.0 and a standard deviation of 1.25, what is the value of the random variable X that has an area to its right equal to 0.3121?

25. If a normal distribution has a mean of 33.7 and a standard deviation of 10.5, what is the value of the random variable X that has an area to its left equal to 0.9904?

26. If a normal distribution has a mean of 152.1 and a standard deviation of 22.0, what is the value of the random variable X that has an area to its left equal to 0.8238?

27. The body temperatures of adults are normally distributed with a mean of 98.60 °F and a standard deviation of 0.73 °F. What temperature represents the 85^{th} percentile?

28. Heights of river birch trees at a large nursery are approximately normally distributed with a mean of 92.3 inches and a standard deviation of 4.1 inches. What is the cutoff height for birch trees in the tallest 10%?

29. The weights of Jersey cows offered at auction in one region are normally distributed with a mean of 825.0 pounds and a standard deviation of 74.8 pounds. One rancher endeavors to only bid on cows that are in the top 5% of weight. What is the lowest weight cow that the rancher should bid on?

30. Suppose that the weights of college students are normally distributed with a mean of 150 pounds and a standard deviation of 20 pounds. What weight represents the first quartile for college students?

31. The cruising altitudes for the fleet of one commercial airliner are normally distributed with a mean of 34,950 feet and a standard deviation of 1830 feet. What cruising altitude represents the 80^{th} percentile?

32. In one region of the Caribbean Sea, daily water temperatures are normally distributed with a mean of 77.9 °F and a standard deviation of 2.4 °F. What is the third quartile for water temperatures in this region?

33. During one season of racing at the Talladega Superspeedway, the mean speed of the cars racing there was found to be 158.900 mph with a standard deviation of 6.700 mph. What speed represents the 30^{th} percentile for speeds of race cars at Talladega? Assume that the racing speeds are normally distributed.

34. Suppose that preschoolers spend a mean of 25 hours per week in day care with a standard deviation of 5 hours per week. A newspaper journalist wants to point out that preschoolers are staying in day cares too long, from his perspective. If he only looks at the extreme end of the distribution, that is, the top 3%, what's the minimum number of hours per week those preschoolers spend in day care? Assume that the number of hours per week that preschoolers spend in day care are normally distributed.

35. School-age children should drink approximately 40.5 oz of water per day according to a recent report. Suppose the amounts of water that schoolchildren actually consume in a day are approximately normally distributed with a mean of 32.0 oz and a standard deviation of 7.1 oz.

a. What is the probability that a randomly selected student will drink less than the suggested amount of water in a day?

b. If the standard deviation remained the same and the daily water intakes were still normally distributed, how much would the mean need to increase so that only 5% of students drink less than 36 oz per week?

36. Pop-It popcorn maker has a mean time before failure of 36 months with a standard deviation of 5 months, and the failure times are normally distributed. What should be the warranty period, in months, so that the manufacturer will not have more than 10% of the poppers returned?

6.5 Section Exercises

Continuity Correction

Describe the area under the normal curve that would be used to approximate the binomial probability.

1. Consider the probability that at least 40 out of 232 planes will malfunction on the runway.
2. Consider the probability that at least 100 out of 250 new mothers received prenatal care.
3. Consider the probability that more than 500 out of 12,000 tax returns were filed incorrectly.
4. Consider the probability that more than 180 out of a sample of 1200 elderly people in the United States will have the flu this winter.
5. Consider the probability that at most 30 out of 534 smartphones on the assembly line are defective.
6. Consider the probability that at most 15 out of 400 high school seniors will not graduate on time.
7. Consider the probability that fewer than 12 out of 367 fourth graders will not pass the state placement test.
8. Consider the probability that 70 out of 100 trees planted in an orchard will live to maturity.
9. Consider the probability that 35 out of 200 registered voters will not vote in the election.
10. Consider the probability that at least 225 out of a sample of 1200 people will use the emergency room at the hospital this year.

Conditions for Using the Normal Distribution Approximation

Verify that a normal distribution can be used to approximate the binomial probability, or show how the conditions have not been met.

11. Consider the probability that fewer than 15 out of the 123 people watching a movie have already read the book. Assume that the probability of a given person having read the book is 40%.
12. Consider the probability that more than 100 out of 238 fifth graders have seen all of the Marvel movies. Assume that the probability of a given fifth grader having seen all of the Marvel movies is 54%.
13. Consider the probability that at most 2 out of 30 television sets on an assembly line are defective. Assume that the probability of a given television set being defective is 5%.
14. Consider the probability that no more than 5 out of 120 teenage girls become pregnant before finishing high school. Assume that the teen pregnancy rate is 4%.

Normal Distribution Approximation of Binomial Probability

Approximate the binomial probability using the normal distribution. You may safely assume that the conditions for using the normal distribution approximation have been met for each scenario.

15. What is the probability that more than 150 out of 230 eighth-graders at a local middle school have been exposed to drugs? Assume that a previous study at this school reported that the probability of an individual eighth-grade student having been exposed to drugs is 63%.
16. What is the probability that more than 100 out of 300 elections end in a runoff situation? One report suggests that there is a 32% chance of an individual election ending in a runoff situation.
17. What is the probability that more than 20 out of a class of 347 high school seniors will drive under the influence of alcohol on prom night? The local chapter of MADD fears that the probability of a high school senior drinking and driving on prom night is 38%.
18. What is the probability that more than 200 out of the 248 hunters staying at a hunting club this season will obtain their game of choice? Club records indicate that hunters have an 83% chance of obtaining their desired game animal.

19. What is the probability that at least 67 out of 100 cars stopped at a roadblock will not be given a ticket? Local authorities report that tickets usually are given to 23% of cars stopped.

20. What is the probability that at least 25 out of a survey of 200 Cajuns do not actually like Cajun food? A regional food critic believes that the probability of a Cajun not liking Cajun food is 14%.

21. What is the probability that at least 140 out of 200 drivers surveyed speed on a regular basis? The state's highway patrol estimates that 78% of drivers typically exceed the speed limit.

22. What is the probability that at least 130 out of 145 preschoolers watch more than four hours of television per night? One previous study indicates that the probability that a preschooler watches more than four hours of television per night is 84%.

23. What is the probability that no more than 32 out of 150 vehicles inspected at a service station will fail their yearly inspection and not receive an inspection sticker? One service station's records indicate that the probability of a car failing its inspection is 18%.

24. What is the probability that no more than 130 out of the 2300 tax returns filed at a local CPA's office will be inaccurate? Previous records indicate only a 7% probability that a given tax return from this office is incorrect.

25. What is the probability that no more than 5 out of 250 puppies born to a well-respected dog breeder will have birth defects? This dog breeder usually averages only 2 birth defects in 50 births.

26. What is the probability that no more than 50 out of 150 former smokers will resume smoking within six months of quitting? Assume that the probability of a former smoker resuming smoking is 35%.

27. What is the probability that fewer than 100 homes out of a sample of 1200 homes will not have a TV? A study indicates that 9 out of 10 homes have televisions.

28. What is the probability that fewer than 100 out of 320 high school graduates will not attend college? The registrar's office at a local college estimates that the probability of a high school graduate going on to college is 68.1%.

29. What is the probability that 125 out of 200 people are overweight? Estimates show that the probability of someone being overweight is 65%.

30. What is the probability that out of 350 entering college freshmen, 100 will need to take a co-requisite mathematics class? Research shows that there is a 36% chance of a college freshman needing to take a co-requisite mathematics class.

7.1 Section Exercises

Properties of Sampling Distributions

Decide if each statement is true or false. Explain why.

1. A sampling distribution refers to individuals rather than groups.

2. The shape of a sampling distribution of sample means that follows the requirements of the Central Limit Theorem will be approximately bell-shaped.

3. A sampling distribution of sample means has a standard deviation equal to $\frac{\sigma}{\sqrt{n}}$.

4. A sampling distribution of sample means has a mean equal to $\frac{\mu}{\sqrt{n}}$.

Means and Standard Deviations of Sampling Distributions

Find the mean and standard deviation of the sampling distribution of sample means using the given information.

5. $\mu = 35$ and $\sigma = 9$; $n = 64$

6. $\mu = 28$ and $\sigma = 6$; $n = 81$

7. $\mu = 12.0$ and $\sigma = 2.3$; $n = 36$

8. $\mu = 52$ and $\sigma = 7$; $n = 100$

9. $\mu = 9.5$ and $\sigma = 10.0$; $n = 39$

10. $\mu = 582.0$ and $\sigma = 23.6$; $n = 1201$

Applications for Sampling Distributions of Sample Means

Answer each question for the given scenario.

11. According to the Bangor Daily News, heating fuel prices in Maine averaged \$2.88 per gallon in October 2014, an increase of 14 cents from the previous month. If samples of 50 heating oil prices are collected, what would be the mean of the sampling distribution of sample means?

Source: Bangor Daily News. November 2014. https://www.bangordailynews.com (12 Aug. 2019).

12. According to a local school district, middle school students are assigned a mean of 2.5 hours of homework per night. If 144 samples of 50 students from this school district are collected and the amount of time spent per night on homework is recorded for each student, what would be the mean of the sampling distribution of the sample means?

13. Some health reports claim that the mean duration of a cold is seven days. If 120 samples of 100 people with colds are taken from across the United States and the duration of each person's cold is recorded, what would be the mean of the sampling distribution of the sample means?

14. Suppose that an Internet source shows that the mean fare for one-way flights for business travel is \$217, the lowest in five years. If 215 samples of 45 one-way fares for business travel are collected from across the United States, what would be the mean of the sampling distribution of the sample means?

15. For a large internet retailer, their average customer spent \$51.28 during the Black Friday sale with a standard deviation of \$9.53. If a sampling distribution is created for samples of 75 customers, what would be the standard deviation of the sampling distribution of sample means?

16. Suppose that a study of elementary school students reports that the mean age at which children begin reading is 5.7 years with a standard deviation of 1.1 years. If a sampling distribution is created using samples of the ages at which 55 children began reading, what would be the standard deviation of the sampling distribution of the sample means?

17. A study on the latest fad diet claimed that the amounts of weight lost by all people on this diet had a standard deviation of 5.8 pounds. If a sampling distribution is created using samples of the amounts of weight lost by 100 people on this diet, what would be the standard deviation of the sampling distribution of the sample means?

18. According to aamc.org, the average tuition for students attending a public medical school without resident status for that state is $60,802 per year. If a sampling distribution that has a standard error of the mean equal to $100 is desired, how many medical school tuitions must be in each sample? Assume a population standard deviation of $3150.

Source: Tuition and Student Fees. Association of American Medical Colleges. 2019.https://www.aamc.org/data/tuitionandstudentfees/ (12 Aug. 2019).

19. Shipping costs for a large national distributor have a mean of $7.94 per item with a standard deviation of $2.29. The population distribution is bell-shaped. Consider the sampling distribution created for samples of size 25. Can a normal approximation be used for this sampling distribution? Explain your answer.

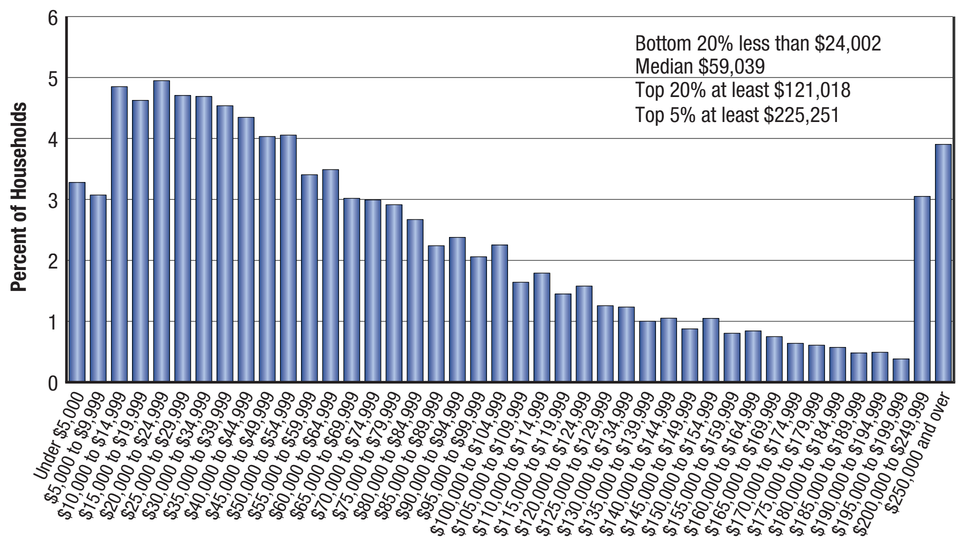
20. Shipping weights for a large national distributor have a mean of 12.2 pounds per package and a standard deviation of 3.8 pounds. The graph of the population is skewed right. If a sampling distribution using samples of 20 packages each is created, can a normal approximation be applied? Explain your answer.

21. A television streaming service has found that its customers have on average 65.5 hours of programming recorded to their DVR systems with a standard deviation of 20.4 hours. The population distribution is found to have a multimodal shape. Consider a sampling distribution created from this population using a sample size of 25 customers. Can a normal approximation be used for the sampling distribution? Explain your answer.

22. An airline company recorded the delay times for all 5592 flights completed one week. They found that the delay times had a mean of 16.5 minutes and standard deviation of 4.7 minutes. The frequency distribution created from the population of delay times was found to be skewed right. If a sampling distribution is created using samples of 50 flights each, can a normal approximation be applied? Explain your answer.

23. Consider the following frequency histogram depicting US household incomes for 2016.

Distribution of Annual Household Income in the United States (2016)



Source: U.S. Census Bureau, Current Population Survey, 2017 Annual Social and Economic Supplement https://commons.wikimedia.org/wiki/File:Distribution_of_ Annual_Household_Income_in_the_United_States_2016.svg (28 July 2019).

a. How would you describe the general shape of this distribution?

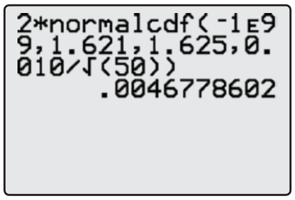
b. Consider the sampling distribution of sample means created for samples of 15 household incomes each. Could a normal approximation be used for this sampling distribution? Explain your answer.

c. Consider the sampling distribution of sample means created for samples of 80 household incomes each. Could a normal approximation be used for this sampling distribution? Explain your answer.

$$
\begin{aligned}
z &= \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \\
&= \frac{0.004}{\left(\frac{0.010}{\sqrt{50}}\right)} \\
&\approx 2.83
\end{aligned}
$$

Thus, the two z-scores are $z_1 = -2.83$ and $z_2 = 2.83$. Because we want to find the total area in the two tails, using the symmetry property of the normal distribution, we can find the area to the left of $z_1 = -2.83$ and double it. Using the tables we find that the area to the left of $z_1 = -2.83$ is 0.0023, so the total area in the two tails is $(0.0023)(2) = 0.0046$.

TI-83/84 Plus: In order to find the total area located in two symmetric tails of the distribution, we will use the calculator to double the area found in the left tail of the curve. To do this, go to the `DISTR` menu and compute `2*normalcdf(lower bound, upper bound, μ, σ)` using the following values.



lower bound: $-1\text{E}99$
upper bound: 1.621
$\mu = 1.625$
$\sigma = \dfrac{0.010}{\sqrt{50}}$

As shown in the margin screenshot, the calculator gives the more accurate value of approximately 0.0047 for the probability when this method is used.

Hence, the probability would be reported as 0.0046 if using tables or 0.0047 if using the calculator. This means that the probability of the sample batch not meeting the requirements is very small, approximately 0.5%.

7.2 Section Exercises

Calculating Standard Scores for Sample Means

Calculate the standard score (z-score) of the given sample means. Round your answer to two decimal places.

1. $\mu = 36$ and $\sigma = 39$; $n = 73$; $\bar{x} = 38$
2. $\mu = 11.0$ and $\sigma = 10.2$; $n = 100$; $\bar{x} = 10.0$
3. $\mu = 1.10$ and $\sigma = 1.10$; $n = 40$; $\bar{x} = 0.55$
4. $\mu = 0.49$ and $\sigma = 0.60$; $n = 1500$; $\bar{x} = 0.52$
5. $\mu = 114{,}023$ and $\sigma = 30{,}000$; $n = 100$; $\bar{x} = 110{,}045$
6. $\mu = 514.2$ and $\sigma = 30.2$; $n = 1000$; $\bar{x} = 513.0$

Probability

Find each specified probability.

7. Suppose that the walking step lengths of adults are normally distributed with a mean of 2.4 feet and a standard deviation of 0.3 feet. A sample of 34 step lengths is taken.

   a. Find the probability that an individual adult's step length is less than 2.1 feet.

b. Find the probability that the mean of the sample taken is less than 2.1 feet.

c. Find the probability that the mean of the sample taken is more than 2.5 feet.

d. Find the probability that the sample mean differs from the population mean by more than 0.06 feet.

8. The average number of miles between oil changes for vehicles serviced in one large facility is 6520 with a standard deviation of 1100 miles. A random sample of 50 vehicles serviced at this facility is taken.

a. Calculate the probability that the sample mean is more than 6800.

b. Calculate the probability that the sample mean is less than 6700.

c. Calculate the probability that the sample mean is between 6200 and 6400.

9. Intelligence quotient (IQ) scores are often reported to be normally distributed with $\mu = 100.0$ and $\sigma = 15.0$.

a. What is the probability of a person chosen at random having an IQ score of less than 95?

b. If a random sample of 50 people is taken, what is the probability that their mean IQ score will be less than 95?

c. If a random sample of 50 people is taken, what is the probability that their mean IQ score will be more than 95?

d. If a random sample of 50 people is taken, what is the probability that their mean IQ score will be more than 105?

e. If a random sample of 50 people is taken, what is the probability that their mean IQ score will differ from the population mean by more than 5?

10. Suppose that the diameters of oak trees are normally distributed with a mean of 4.000 feet and a standard deviation of 0.375 feet.

a. What is the probability of walking down the street and finding an oak tree with a diameter of more than 5 feet?

b. What is the probability of sampling a set of 87 oak trees and finding their mean to be more than 4.1 feet in diameter?

c. What is the probability of sampling a set of 87 oak trees and finding their mean to be less than 3.92 feet in diameter?

d. What is the probability of sampling a set of 87 oak trees and finding their mean to differ from the population mean by less than 0.1 feet in diameter?

11. A glass tube maker claims that his tubes have lengths that are normally distributed with a mean of 9.00 cm and a variance of 0.25.

a. What is the probability that a quality control regulator will pull a tube off the assembly line that has a length between 8.6 and 9 cm?

b. What is the probability that a random sample of 40 tubes will have a mean of less than 8.8 cm?

c. What is the probability that a random sample of 35 tubes will have a mean of more than 9.2 cm?

d. What is the probability that a random sample of 75 tubes will have a mean that differs from the population mean by more than 0.1 cm?

12. A pencil manufacturer claims that its pencils have lengths that are normally distributed with a mean of 6.0 inches and a variance of 0.2. What is the probability that a randomly chosen pencil will have a length of more than 6.4 inches?

13. A vineyard claims that the mean berry count per grape cluster is 43 with a standard deviation of 3.6 berries. What is the probability that, in next year's crop, the mean berry count for a sample of 50 clusters will be between 42 and 44?

14. A book publisher claims that its mean book length is 250 pages with a standard deviation of 70 pages. As a reviewer, you get paid per book, and not per page, to read over a manuscript. What is the probability that for a sample of 45 randomly selected books, the mean length of a book is less than 230 pages?

15. Airlines predict that the mean number of "no shows" per 100 seats on each flight is 10.0 with a standard deviation of 3.4. What is the probability that a random sample of 45 flights has a mean of more than 12 "no shows" per 100 seats?

16. In a Scrabble tournament, the scores were normally distributed and the mean score was 420.2 points with a standard deviation of 105.0 points. What is the probability that the score of a randomly selected competitor differs from the mean score by less than 50 points?

17. A tea bag manufacturer needs to place 2 g of tea in each bag. If the machinery places a mean of 2.6 g of tea in each bag with a standard deviation of 0.3 g, what is the probability that a randomly chosen bag will have between 2 and 2.8 g of tea, assuming that the amounts of tea per bag are normally distributed?

18. A pollster claims that the mean amount spent on Christmas gifts by an American family is \$927 with a standard deviation of \$200. What is the probability that the mean amount spent on Christmas gifts for a sample of 500 families differs from the population mean by less than \$15?

19. A medical journal reports the mean fetal heart rate to be 140 beats per minute (bpm) with a standard deviation of 12 bpm. What is the probability that a fetal heart rate differs from the mean by more than 25 bpm, assuming that fetal heart rates are normally distributed?

20. A local sports magazine reports the mean length of a baseball game to be 175.9 minutes with a standard deviation of 27.0 minutes. For a random sample of 30 games, what is the probability that the mean game length is at most 170 minutes?

21. Teenagers reportedly spend a mean of 17.5 hours per week playing computer and video games. For a random sample of 110 teenagers, what is the probability that the mean amount of time spent playing games is more than 18 hours per week? Let $\sigma = 3.0$.

22. The mean wait time for a drive-through chain is 193.2 seconds with a standard deviation of 29.5 seconds. What is the probability that for a random sample of 45 wait times, the mean is between 185.7 and 206.5 seconds?

23. The mean amount spent per order at one fast food restaurant is \$8.43 with a standard deviation of \$1.52. What is the probability that for a random sample of 75 orders from this week's customers, the mean amount spent will be between \$8 and \$9?

24. The mean hourly rate for babysitters in one town is \$7.05 with a standard deviation of \$0.55. What is the probability that a babysitter chosen at random will charge an hourly rate that differs from the mean by less than \$1.00, assuming that the hourly rates for babysitters in this town are normally distributed?

25. The mean sale price for a piece of art from members of a large artists' guild is \$545 with a standard deviation of \$76. If, at one show, 45 pieces are for sale, what is the probability that the mean sale price for the show will be higher than \$575?

26. The mean cost for plumbing repairs in one area is \$208 with a standard deviation of \$64. If 31 homeowners in that area are surveyed, what is the probability that the mean cost for their plumbing repairs will be over \$200?

27. One pediatric clinic sees a mean of 42.0 patients per day with a standard deviation of 3.4 patients per day. What is the probability that the mean number of patients per day for March (31 days) will be between 41 and 44?

28. At a large university, the mean amount spent by students for cellular phone service is \$38.90 per month with a standard deviation of \$3.64 per month. Consider a group of 44 randomly chosen university students. What is the probability that the mean amount of their monthly cell phone bills differs from the mean for the university by more than \$1?

29. A large bakery sells a mean of 11.0 dozen cookies per day with a standard deviation of 1.1 dozen cookies per day. Consider the mean number of cookies sold per day in July (31 days) of one year. What is the probability that the mean differs from the population mean by less than 0.5 dozen?

30. According to data released by the Jackson City Chamber of Commerce, the weekly wages of office workers have a mean of \$723 and a standard deviation of \$151. If 57 office workers are chosen at random from Jackson City, what is the probability that the mean weekly wage of these workers will be greater than \$750?

Hence, the probability would be reported as 0.4902 if using tables or 0.4892 if using the calculator. This means that the probability that the sample proportion differs from the population proportion by more than 2% is approximately 49%.

7.3 Section Exercises

Calculating Standard Scores for Sample Proportions

Given the following parameters for a sampling distribution of sample proportions, calculate the standard score of the sample proportion. Round the standard score to two decimal places.

1. $p = 0.34, \quad x = 35, \quad n = 100$
2. $p = 0.54, \quad x = 563, \quad n = 1000$
3. $p = 0.8, \quad x = 24, \quad n = 29$
4. $p = 0.61, \quad x = 66, \quad n = 98$
5. $p = 0.15, \quad x = 1152, \quad n = 12{,}500$
6. $p = 0.93, \quad x = 987, \quad n = 1012$

Probability

Find each specified probability.

7. A large car dealership claims that 47% of its customers are looking to buy a sport utility vehicle (SUV). A random sample of 61 customers is surveyed. What is the probability that less than 40% are looking to buy an SUV?
8. The local nursery is waiting for its spring annuals to be delivered, and 20% of the plants ordered are petunias. If the first truck contains 120 plants packed at random, what is the probability that no more than 30 of the plants are petunias?
9. A car lot sells pre-owned cars. Its sales manager reports that 78% of the cars are sold for less than $10,000. What is the probability that in a random sample of 45 cars, more than 75% are sold for less than $10,000?
10. A news report states that 65% of vehicles sold nationally are SUVs. A random sample of 100 vehicles sold in the last month is taken. What is the probability that at least 70 of the vehicles in the sample were SUVs?
11. A local florist says that 85% of all flowers sold for Valentine's Day are roses. Consider a random sample of 150 orders for Valentine's Day flowers. What is the probability that less than 80% of the orders in the sample are for roses?
12. At one private college, 34% of students are business majors. Suppose that 260 students are randomly selected from a list in the registrar's office. What is the probability that the proportion of business students in the sample differs from the population proportion by less than 2%?
13. At a large grocery store, 72% of shoppers are women. In order to obtain information about spending habits, 40 shoppers are randomly chosen for a survey. What is the probability that the proportion of women in the sample differs from the population proportion by more than 3%?
14. A major appliance retailer claims that 18% of all appliances sold are dishwashers. If 112 appliances sold are randomly selected, what is the probability that the proportion of dishwashers sold differs from the population proportion by over 2%?

15. A popular restaurant downtown says that 20% of diners order the daily special. Consider a random sample of 128 people who dined at the restaurant. What is the probability that the proportion of diners who ordered the special differs from the population proportion by less than 1%?

16. A study of college freshmen revealed that 29% had no work experience prior to entering college. Suppose a group of 75 college freshmen is chosen at random. What is the probability that less than 18 of the students selected had no work experience prior to entering school?

8.1 Section Exercises

Note: *For all exercises in this section, you may assume that the requirements mentioned in this section are met; namely, the population standard deviation is known, all samples are simple random samples, and either the sample size is at least 30 or the population distribution is approximately normal.*

Point Estimates and Confidence Intervals for Population Means

Find each specified point estimate or confidence interval.

1. A survey of 42 randomly selected teachers finds that they spend a mean of \$18 per week on lunch. What is the best point estimate for the mean amount of money spent per week on lunch for all teachers?

2. The mean number of pets per student for a random sample of 47 students at Brown Elementary is 2.5 pets. What is the best point estimate for the mean number of pets per student for all students at Brown Elementary?

3. A survey of teachers reports that a point estimate for the mean amount of money spent each week on lunch is \$18.00. If the margin of error for a 95% confidence interval for the mean amount of money spent each week on lunch by all teachers is \$1.70, construct a 95% confidence interval for the mean amount of money spent each week on lunch for all teachers.

4. The mean batting average for a random sample of 35 professional baseball players is .283. If the margin of error for the population mean with a 99% level of confidence is .051, construct a 99% confidence interval for the mean batting average for professional baseball players.

Margins of Error of Confidence Intervals for Population Means (σ Known)

Calculate the margin of error of a confidence interval for the population mean at the given level of confidence.

5. $n = 56, \quad \sigma = 3.14, \quad c = 0.90$

6. $n = 81, \quad \sigma = 2.45, \quad c = 0.95$

7. $n = 93, \quad \sigma = 1.25, \quad c = 0.95$

8. $n = 134, \quad \sigma = 0.27, \quad c = 0.99$

Confidence Intervals for Population Means (σ Known)

Construct a confidence interval for the population mean at the given level of confidence.

9. $n = 89, \quad \sigma = 2.01, \quad c = 0.95, \quad \bar{x} = 45.00$

10. $n = 64, \quad \sigma = 8.01, \quad c = 0.90, \quad \bar{x} = 90.40$

11. $n = 607, \quad \sigma = 1.92, \quad c = 0.99, \quad \bar{x} = 18.45$

12. $n = 1123, \quad \sigma = 7.31, \quad c = 0.95, \quad \bar{x} = 87.12$

Minimum Sample Sizes for Estimating Population Means

Calculate the minimum sample size needed to construct a confidence interval with the desired characteristics. The value of the population standard deviation is an estimate based on a previous reliable study.

13. $E = 0.5, \quad \sigma = 5.25, \quad c = 0.95$

14. $E = 2, \quad \sigma = 12.10, \quad c = 0.90$

15. $E = 1.5, \quad \sigma = 4.75, \quad c = 0.99$

16. $E = 3, \quad \sigma = 15.03, \quad c = 0.95$

Confidence Intervals for Population Means (σ Known)

Construct and interpret each specified confidence interval.

17. A professor wants to estimate how many hours per week her students study. A simple random sample of 78 students had a mean of 15.0 hours of studying per week. Construct and interpret a 90% confidence interval for the mean number of hours a student studies per week. Assume that the population standard deviation is known to be 2.3 hours per week.

18. A faculty advocacy group is concerned about the amount of time teachers spend each week doing schoolwork at home. A simple random sample of 56 teachers had a mean of 8.0 hours per week working at home after school. Construct and interpret a 95% confidence interval for the mean number of hours per week a teacher spends working at home. Assume that the population standard deviation is 1.5 hours per week.

19. A writer for a computer magazine is working on an article about computer usage in American households. A simple random sample of 120 American households has a mean computer usage time of 19.2 hours per week. Construct and interpret a 95% confidence interval for the mean computer usage time per week for all American households. Assume that the population standard deviation is 3.3 hours per week.

20. A survey of 85 randomly selected homeowners finds that they spend a mean of $67 per month on home maintenance. Construct and interpret a 99% confidence interval for the mean amount of money spent per month on home maintenance by all homeowners. Assume that the population standard deviation is $14 per month.

21. A survey of 97 randomly selected homeowners found that the mean amount spent on lawn service was $720 per year. Construct and interpret a 98% confidence interval for the mean amount of money spent on lawn service per household each year. Assume that the population standard deviation is $123 per year.

22. A survey of a simple random sample of 140 dieters revealed that the numbers of times they "cheated" on their diets had a mean of 7.0 times per week. Construct and interpret a 99% confidence interval for the mean number of times dieters "cheat" on their diets each week. Assume that the population standard deviation is 1.5 times per week.

23. A physical therapist is investigating the mean recovery time after ACL surgery for patients involved in a new therapy regimen. For the purpose of the study, a successful recovery was defined to be the ability to walk without crutches. For 38 randomly selected patients, the mean recovery time after ACL surgery was found to be 22.6 days. Assume that the population standard deviation is 3.7 days. Find and interpret a 99% confidence interval for the mean recovery time for all ACL surgery patients undergoing the same new therapy.

24. The manufacturers of Caudill automotive oil wish to estimate the mean number of miles that motorists drive between oil changes. A random sample of 54 motorists has a mean of 5900 miles driven between oil changes. Assume that the population standard deviation is 1350 miles. Construct and interpret a 95% confidence interval for the mean number of miles driven between oil changes for all motorists.

Minimum Sample Sizes for Estimating Population Means

Calculate the minimum sample size needed to construct a confidence interval with the desired characteristics. The value of the population standard deviation is an estimate based on a previous reliable study.

25. The upper management at a bank would like to estimate the mean number of credit cards held by Millennials. They would like to create a 98% confidence interval with a maximum error of 1 card. Assuming a standard deviation of 3.25 cards, what is the minimum number of Millennials they must include in their sample?

26. A social worker is concerned about the number of prescriptions her elderly clients have. She would like to create a 99% confidence interval for the mean number of prescriptions per client with a maximum error of 2 prescriptions. Assuming a standard deviation of 5.2 prescriptions, what is the minimum number of clients she must sample?

27. Suppose you are interested in determining the mean number of hours students spend working out each week. You want a 95% level of confidence and a maximum error of 0.5 hours. Assuming the standard deviation is 2.5 hours, what is the minimum number of students you must include in your sample?

28. For a psychology experiment, Emma is assigned the task of finding the mean number of hours students sleep per night. Her results must be at the 99% level of confidence with a maximum error of 0.25 hours. Assuming the standard deviation is 1.4 hours, how many students must Emma survey?

Respond thoughtfully to the following exercises.

29. Given a 99% confidence interval for a population mean of $(1.01, 1.97)$, is it possible to determine the original point estimate for the interval? Explain your answer.

30. Two individual researchers report a confidence interval for the mean number of unreported domestic violence incidents per month in one small county over the past year. One gives a 95% confidence interval of $(4.93, 6.14)$ while the other gives a 98% confidence interval of $(3.72, 5.01)$. Is it possible for both of the intervals to contain the true mean number of unreported domestic violence incidents? Explain your answer.

31. Last quarter, your sales team reported a mean amount of revenue of \$131,540 per salesperson. Based on a survey of a portion of the sales team and last quarter's data, you estimate that this quarter, the mean revenue will be between \$129,915 and \$154,798 per salesperson, with a 95% level of confidence. Would you report to your supervisor that this quarter's sales will be up from last quarter? Why or why not?

32. The formula for the margin of error for a confidence interval for a population mean (when the population standard deviation is known) is given as $E = \left(z_{\alpha/2}\right)\left(\sigma_{\bar{x}}\right)$ or $E = \left(z_{\alpha/2}\right)\left(\frac{\sigma}{\sqrt{n}}\right)$. Explain how the Central Limit Theorem can be used to rewrite the first formula as the second one.

8.2 Section Exercises

Finding *t*-Values

Find each specified t-value.

1. Find the value of $t_{0.05}$ for a t-distribution with 15 degrees of freedom.
2. Find the value of $t_{0.01}$ for a t-distribution with 8 degrees of freedom.
3. Find the value of $t_{0.005}$ for a t-distribution with 26 degrees of freedom.
4. Find the value of $t_{0.10}$ for a t-distribution with 20 degrees of freedom.
5. Find the value of $t_{0.01}$ for a t-distribution with 29 degrees of freedom.
6. Find the value of $t_{0.025}$ for a t-distribution with 6 degrees of freedom.
7. Find the value of t for a t-distribution with 9 degrees of freedom such that the area to the right of t equals 0.05.
8. Find the value of t for a t-distribution with 11 degrees of freedom such that the area to the right of t equals 0.025.
9. Find the value of t for a t-distribution with 5 degrees of freedom such that the area to the right of t equals 0.01.
10. Find the value of t for a t-distribution with 14 degrees of freedom such that the area to the right of t equals 0.005.
11. Find the value of t for a t-distribution with 10 degrees of freedom such that the area to the left of t equals 0.05.
12. Find the value of t for a t-distribution with 3 degrees of freedom such that the area to the left of t equals 0.10.
13. Find the value of t for a t-distribution with 12 degrees of freedom such that the area to the left of t equals 0.005.
14. Find the value of t for a t-distribution with 27 degrees of freedom such that the area to the left of t equals 0.025.
15. Find the value of t for a t-distribution with 13 degrees of freedom such that the area to the left of $-t$ plus the area to the right of t equals 0.01.
16. Find the value of t for a t-distribution with 23 degrees of freedom such that the area to the left of $-t$ plus the area to the right of t equals 0.20.
17. Find the value of t for a t-distribution with 12 degrees of freedom such that the area to the left of $-t$ plus the area to the right of t equals 0.02.
18. Find the value of t for a t-distribution with 18 degrees of freedom such that the area to the left of $-t$ plus the area to the right of t equals 0.05.
19. Find the value of t for a t-distribution with 18 degrees of freedom such that the area between $-t$ and t equals 95%.
20. Find the value of t for a t-distribution with 4 degrees of freedom such that the area between $-t$ and t equals 98%.
21. Find the value of t for a t-distribution with 25 degrees of freedom such that the area between $-t$ and t equals 90%.
22. Find the value of t for a t-distribution with 22 degrees of freedom such that the area between $-t$ and t equals 95%.
23. Find the critical t-value for a 90% confidence interval using a t-distribution with 25 degrees of freedom.

24. Find the critical t-value for a 99% confidence interval using a t-distribution with 18 degrees of freedom.

25. Compare and contrast Exercises 21 and 23.

26. Compare and contrast Exercise 15 and Example 8.2.5.

27. In the paragraph after Figure 8.2.2 in this section, we mention that the last row of the t-distribution table has some familiar numbers in it. Why should these numbers be familiar?

$$0.5240 < \mu < 0.7624$$
or
$$(0.5240, 0.7624)$$

We are 99% confident that the mean amount of water used per household for brushing teeth is between 0.5240 and 0.7624 gallons per day.

8.3 Section Exercises

Note: *For all exercises in this section, you may assume that the requirements mentioned in this section are met; namely, the population standard deviation is unknown, all samples are simple random samples, and either the sample size is at least 30 or the population distribution is approximately normal.*

Confidence Intervals for Population Means (σ Unknown)

Construct a confidence interval for the population mean at the given level of confidence using the information provided. Assume sample data are simple random samples.

1. $n = 14$, $\bar{x} = 95.0$, $s = 4.8$, level of confidence is 95%
2. $n = 25$, $\bar{x} = 56$, $s = 8$, level of confidence is 90%
3. $n = 8$, $\bar{x} = 7.0$, $s = 1.2$, level of confidence is 99%
4. $n = 13$, $\bar{x} = 1.97$, $s = 0.03$, level of confidence is 98%
5. $c = 0.98$

192	465	321	299	516	256	339	311	407

6. $c = 0.95$

Stem-and-Leaf Plot

Stem	Leaves
2	2 3 3 8 9 9
3	0 1 1 4
4	5 6 7 7 8 8

Key: 2 | 2 = 22

7. Level of confidence is 90%

11	47	95	54	33	64	4	8	57	9	80	32	19
8	90	3	49	4	44	79	80	48	16	64	55	68
31	7	15	21	52	6	78	109	40	50	12	29	22

8. Level of confidence is 80%

Ages of Participants in a Study

Age	Frequency
16	9
17	15
18	16
19	5
20	3

Margins of Error of Confidence Intervals for Population Means (σ Unknown)

Calculate each specified margin of error.

9. The mean distance commuters drove to work each day was estimated to be 40.8 miles from a sample of 45 commuters. The sample standard deviation was 5.8 miles. Calculate the margin of error for a 95% confidence interval.

10. The mean amount of money spent per week on gas by a sample of 25 drivers was found to be $57.00 with a standard deviation of $2.36. Calculate the margin of error for a 90% confidence interval. Assume that the population distribution is approximately normal.

Confidence Intervals for Population Means (σ Unknown)

Construct and interpret each specified confidence interval.

11. Wildlife conservationists studying grizzly bears in the United States found that the mean weight of 25 adult male grizzly bears was 600 pounds with a standard deviation of 90 pounds. Construct and interpret a 98% confidence interval for the mean weight of all adult male grizzly bears in the United States. Assume that the weights of all adult male grizzly bears in the United States are normally distributed.

12. The mean length of 12 newly hatched iguanas is 7.00 inches with a standard deviation of 0.75 inches. Construct and interpret a 90% confidence interval for the mean length of all newly hatched iguanas. Assume that the lengths of all newly hatched iguanas are normally distributed.

13. Suppose that you sample 59 high school baseball pitchers in one county and find that they have a mean fastball pitching speed of 80.00 miles per hour (mph) with a standard deviation of 4.98 mph. Find a 95% confidence interval for the mean fastball pitching speed of all high school baseball pitchers in the county. Interpret the interval.

14. Given the following data, construct and interpret a 99% confidence interval for the mean face value of an individual life insurance policy.

Face Values of Individual Life Insurance Policies

Stem	Leaves
15	0 0 0 0 0 0 1 2 2 3 3 4 5 8 9 9
16	0 0 0 2 3 3 3 3 5 5 8 8 8
17	1 1 2 2 2 3 4 5 5 5 6
18	2 2 3 5 5
19	0 0 5 5 5 6
20	0 0 0 0 0 2 2

Key: 15 | 0 = $150,000

15. The attendance records for a random sample of 28 men's basketball games at one university revealed that the mean number of fans at each game was 4125.0 with a standard deviation of 741.0. To predict ticket sales for next year, the athletic office needs a confidence interval for attendance at these games. Using a confidence level of 95%, construct and interpret a confidence interval for the mean number of fans at men's basketball games at this university. Assume that the population distribution is approximately normal.

16. Suppose you are thinking about getting a puppy and want to know the amount of time people spend caring for puppies. You survey 31 puppy owners and find that the mean amount of time they spend caring for their puppies is 108.0 minutes per day. If the standard deviation is 17.0 minutes, construct a 98% confidence interval for the mean amount of time puppy owners spend on their puppies per day. Interpret the interval.

17. A company that manufactures gas fireplaces wishes to estimate the average amount of heat produced by its newest model. A sample of 35 fireplaces produced an average of 22,770 BTUs of heat with a standard deviation of 320 BTUs. Create a 90% confidence interval for the average amount of heat produced by all of the newest model fireplaces manufactured by this company.

18. A fitness publication wishes to determine the average number of calories contained in a fast food combo meal. A sample of 46 combo meals is selected at random from a variety of fast food restaurants. The average number of calories contained in the sample meals was 860 with a standard deviation of 110 calories. Construct a 99% confidence interval for the average number of calories in all fast food combo meals.

19. As part of a fuel efficiency study, a large school district wishes to estimate the average weekly mileage for its school buses. A sample of 18 buses is selected at random and the mileages are recorded for that week. The same mean was found to be 215 with a standard deviation of 31 miles. Construct a 99% confidence interval for the average weekly mileage for all school buses in the school district. Assume that the population distribution is approximately normal.

20. The following is a random sample of the annual salaries of high school counselors in the United States. Assuming that the distribution of salaries is approximately normal, construct a 90% confidence interval for the mean salary of high school counselors across the United States. Interpret the interval.

\$51,050	\$38,740	\$65,360	\$42,640	\$55,340	\$32,980	\$49,540

Respond thoughtfully to the following exercises.

21. Suppose you were told that the confidence interval for a population mean is $(16.30,\ 19.70)$. Is it possible for you to determine what the margin of error, E, is? If so, what is it?

22. Suppose you are told that a 95% confidence interval for the mean monthly household electric bill in the county where you live is $(205.56,\ 253.90)$. Is it possible for you to determine how many electric bills were sampled to construct the interval?

23. You are presented with the following reports estimating the mean increase in monthly household spending on gasoline over the past six months. Determine which report you find most convincing and explain your reasoning.

i. Based on a simple random sample of 15 households, we are 98% confident that the mean increase in monthly household spending on gasoline over the past six months is between \$43.65 and \$58.93.

ii. Based on a simple random sample of 52 households, we are 95% confident that the mean increase in monthly household spending on gasoline over the past six months is between \$48.82 and \$52.10.

24. If you were presented with a margin of error of 15,642, would you believe that the margin of error was reported correctly?

Discussion Questions

Discuss each question with your classmates. Focus on the relationships between the parameters in each question.

25. Lisa sets out to survey 500 people; however, only 387 responses were received.

a. How will this decrease in her sample size affect the margin of error for her confidence interval for a population mean?

b. How will this decrease in her sample size affect the width of her confidence interval for a population mean?

26. How will increasing the level of confidence without changing the sample size affect the width of a confidence interval for a population mean?

27. How will increasing the level of confidence without changing the sample size affect the margin of error for a confidence interval for a population mean?

28. Which level of confidence will produce a wider confidence interval for a population mean: a 95% level of confidence or a 99% level of confidence?

29. If you decrease the sample size while keeping the margin of error for a confidence interval for a population mean constant, what effect will this have on the level of confidence?

30. How will the width of a confidence interval for a population mean change if you increase the sample size and keep the same level of confidence?

$$\hat{p} = \frac{x}{n} = \frac{84}{383} \approx 0.219321$$

c. To construct the confidence interval, we need to calculate the margin of error. Substituting the appropriate values into the formula, we have the following.

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.44\sqrt{\frac{0.219321(1-0.219321)}{383}} \approx 0.030447$$

Notice that the margin of error we calculated is very close to the value of E that we were willing to accept in part a. Since we used the minimum sample size calculated in part a., this is to be expected.

Therefore, subtracting the margin of error from the sample proportion and then adding the margin of error to the sample proportion gives us the following endpoints for the confidence interval.

$$\text{Lower endpoint: } \hat{p} - E = 0.219321 - 0.030447 \approx 0.189$$

$$\text{Upper endpoint: } \hat{p} + E = 0.219321 + 0.030447 \approx 0.250$$

Thus, the 85% confidence interval for the population proportion ranges from 0.189 to 0.250. The confidence interval can be written mathematically using either inequality symbols or interval notation, as follows.

$$0.189 < p < 0.250$$

or

$$(0.189, 0.250)$$

In other words, the state education commission can be 85% confident that the proportion of tenth-grade students in the state who read at or below the eighth-grade level is between 18.9% and 25.0%.

8.4 Section Exercises

Determine if the conditions $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$ are met for the following scenarios.

1. $n = 80, \hat{p} = 0.43$
2. $n = 110, \hat{p} = 0.95$
3. A hospital administrator wishes to estimate the proportion of his medical staff that follows the appropriate handwashing protocol. Out of 35 staff members randomly chosen, 29 were found to follow the appropriate handwashing protocol.

4. A market analyst wants to predict the proportion of mortgages in the San Francisco Bay area that are distressed, meaning that borrowers are more than 3 months behind on payments. Out of 350 mortgages selected at random, 17 are found to be distressed.

Note: *For the following exercises in this section, you may assume that the necessary requirements are met; namely, all samples are simple random samples, the conditions for a binomial distribution are met, and the sample size is large enough to ensure that* $n\hat{p} \geq 10$ *and* $n(1-\hat{p}) \geq 10$.

Point Estimates and Confidence Intervals for Population Proportions

Find each specified point estimate or confidence interval.

5. Out of 50 randomly selected students who were surveyed, 41 felt that dorm renovation should be the administration's top priority. What is the best point estimate for the proportion of all students who feel that dorm renovation should be the administration's top priority?

6. In a random sample of 130 students, only 7 had been placed in the wrong math class. What is the best point estimate for the population proportion of all students who have been placed in the wrong math class?

7. A survey revealed that 48 out of 112 randomly selected homeowners in one neighborhood pay for lawn service every month. What is the best point estimate for the population proportion of all homeowners in that neighborhood who pay for lawn service every month?

8. Out of 210 randomly selected registered voters who were polled, 107 declared that they intended to vote for the incumbent in the upcoming election. What is the best point estimate for the proportion of all registered voters who intend to vote for the incumbent?

9. Out of 50 randomly selected students who were surveyed, 41 felt that dorm renovation should be the administration's top priority. Using a margin of error of 8.9%, give the interval estimate for the proportion of all students who feel that dorm renovation should be the administration's top priority.

10. Out of 210 randomly selected registered voters who were polled, 107 declared that they intended to vote for the incumbent in the upcoming election. Using a margin of error of 4.4%, give the interval estimate for the proportion of all registered voters who intend to vote for the incumbent.

Confidence Intervals for Population Proportions

Construct and interpret each specified confidence interval.

11. Thirteen out of 147 randomly selected faculty members who were surveyed at a community college know sign language. Construct and interpret a 90% confidence interval for the proportion of all faculty members at the community college who know sign language.

12. A random sample of 200 computer chips is obtained from one factory and 4% are found to be defective. Construct and interpret a 95% confidence interval for the proportion of all computer chips from that factory that are defective.

13. Out of 140 randomly selected kindergartners who were surveyed, 32 said that pancakes are their favorite breakfast food. Construct and interpret a 90% confidence interval for the percentage of all kindergartners who say pancakes are their favorite breakfast food.

14. Out of 54 randomly selected patients of a local hospital who were surveyed, 49 reported that they were satisfied with the care they received. Construct and interpret a 95% confidence interval for the percentage of all patients satisfied with their care at that hospital.

15. A survey of 145 randomly selected students at one college showed that only 87 checked their campus e-mail account on a regular basis. Construct and interpret a 90% confidence interval for the percentage of students at that college who do not check their e-mail account on a regular basis.

16. A random sample of 200 computer chips is obtained from one factory and 4% are found to be defective. Construct and interpret a 99% confidence interval for the percentage of all computer chips from that factory that are *not* defective.

17. Out of 190 randomly selected adults in the United States who were surveyed, 71 exercise on a regular basis. Construct and interpret a 95% confidence interval for the proportion of all adults in the United States who exercise on a regular basis.

18. A survey of 47 randomly selected members of a health club reported that 68.1% prefer walking or running on a treadmill to any other type of aerobic exercise. Construct and interpret a 95% confidence interval for the proportion of health club members who prefer walking or running on a treadmill.

19. Suppose that a national wireless phone company is interested in determining how many of its customers prefer paperless billing. In a survey of 730 customers, 323 said they prefer paperless billing. Construct and interpret a 95% confidence interval for the percentage of all customers who prefer paperless billing.

20. Of 542 mortgages made to borrowers with credit scores between 620 and 679, 41 defaulted on their loans within 5 years. Create a 90% confidence interval for the percentage of all borrowers with credit scores between 620 and 679 who default on their mortgages.

Minimum Sample Sizes for Estimating Population Proportions

Calculate the minimum sample size needed to construct a confidence interval with the desired characteristics.

21. Suppose you wish to determine the proportion of college students in your state who receive some form of financial aid. You want to be 98% confident of your results and have a maximum error of 5%. Calculate the minimum sample size that you must have to meet these requirements, given that the financial aid office at a local institution estimates the percentage to be 78%.

22. Suppose that you wish to determine the proportion of college students in the United States who work at least 20 hours per week while enrolled as full-time students. You want to be 95% confident in your results and have a maximum error of 3%. Calculate the minimum number of college students that you must survey to meet these requirements, given that a previous study estimated that 68% of all college students in the United States work at least 20 hours per week while enrolled as a full-time student.

23. A market researcher wishes to determine the proportion of American consumers who shop online. The results must be accurate at the 90% level of confidence with a maximum error of 2%. Calculate the minimum sample size needed to meet these requirements, given that a previous study estimated that 59% of American consumers shop online.

24. Suppose a researcher needs to verify the proportion of patients who will have mild side effects when taking a recently approved drug. The researcher must be 99% sure of her results with a maximum error of 1%. Calculate the minimum sample size the researcher needs to meet her requirements, given that the studies used to get the drug approved reported that only 12% of patients experienced mild side effects.

Respond thoughtfully to the following exercises.

25. The Deseret News in Salt Lake City, Utah reported in April 2012 that the Zions Bank Consumer Attitude Index for Utah increased to 85.3, which was 5.4 points higher than the month prior. It stated that the index was "based on a representative sample of 500 Utah households … and has a confidence interval of plus or minus 4.38 percent at a 95 percent confidence level." Based on the information given in the article, is it possible that the index reported in April actually did not increase over the previous month's index? Why or why not?

Source: "Utah Consumers Remain Confident, New Report States." *Deseret News*. 24 Apr. 2012. http://www.deseretnews.com/article/765571132/Utah-consumers-remain-confident-new-report-states.html (20 May 2012).

26. A Bloomberg Politics National Poll surveyed 982 Americans in November 2016 regarding the upcoming Presidential election. Of this group, 46% favored Hillary Clinton and 43% favored Donald Trump, with a margin of error of 3.5%. Donald Trump went on to win the election a few weeks later. Based on this information, is it true to say that the Bloomberg poll predicted the election wrong? Explain your answer.

Source: Bloomberg Politics Poll. Selzer & Company. 7 Nov. 2016. https://assets.bwbx.io/documents/users/iqjWHBFdfxlU/rklC

27. Solve the formula for the margin of error of a confidence interval for a proportion, $E = z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$, for n, to show that it is equivalent to the formula for the minimum sample size for estimating a population proportion, $n = \hat{p}\left(1-\hat{p}\right)\left(\frac{z_{\alpha/2}}{E}\right)^2$. (Recall that a value of $\hat{p}$ from a previous study is typically used as an estimate for p in the formula for the minimum sample size.)

Solution

According to the table, we see that to be 99% confident that the sample standard deviation will be within 5% of the true population standard deviation, the minimum sample size required is 1337.

Minimum Sample Sizes for Estimating Standard Deviation

s is Within this Percentage of the Value of σ	Minimum Sample Size Needed for 95% Level of Confidence	Minimum Sample Size Needed for 99% Level of Confidence
1%	19,206	33,220
5%	769	1337
10%	193	337
20%	49	86
30%	22	39
40%	13	23
50%	9	15

Therefore, the market researcher must include at least 1337 home prices for her study to have the level of precision that she needs.

8.5 Section Exercises

Note: *For all exercises in this section, you may assume that the requirements mentioned in this section are met; namely, all samples of a given size have an equal probability of being chosen and the population distribution is approximately normal.*

Point Estimates for Population Variances and Population Standard Deviations

Find each specified point estimate.

1. What is the best point estimate for the population standard deviation if the sample standard deviation is 3.5?

2. What is the best point estimate for the population standard deviation if the sample variance is 25?

3. Consider a sample of caps for three-inch pipes. Their diameters are measured and found to have a variance of 0.12. Give a point estimate for the population variance in diameter lengths of the caps.

4. For a random sample of 112 largemouth bass, the mean weight was found to be 2.7 pounds with a standard deviation of 0.8 pounds. Give a point estimate for the population variance of weights of largemouth bass.

Critical Values for Confidence Intervals for Population Variances

Determine the critical values for the left and right endpoints of a confidence interval for the population variance using the given information.

5. $n = 25, \quad \alpha = 0.05$

6. $n = 17, \quad \alpha = 0.10$

7. $n = 22, \quad c = 0.90$

8. $n = 10, \quad c = 0.95$

Confidence Intervals for Population Variances and Population Standard Deviations

Construct a confidence interval for the population variance at the given level of confidence.

9. $n = 12, \quad s^2 = 19.2, \quad c = 0.95$

10. $n = 20, \quad s^2 = 14.2, \quad c = 0.99$

11. $n = 23, \quad s^2 = 11.9, \quad c = 0.99$

12. $n = 29, \quad s^2 = 26.5, \quad c = 0.98$

Construct a confidence interval for the population standard deviation at the given level of confidence.

13. $n = 63, \quad s = 2.4, \quad c = 0.99$

14. $n = 41, \quad s = 1.2, \quad c = 0.95$

15. $n = 50, \quad s = 6.8, \quad c = 0.98$

16. $n = 56, \quad s = 3.5, \quad c = 0.99$

Construct and interpret each specified confidence interval.

17. A commercial grocer is testing the variance in the weights of packages of strawberries. The weights of the packages are measured in ounces, and a random sample of 15 packages of strawberries has a variance of 3.80. Construct and interpret a 90% confidence interval for the variance in weights of all packages of strawberries.

18. A tire manufacturer is testing the tire pressure in its new line of SUV tires. A random sample of 10 tire pressure readings, measured in pounds per square inch (psi), yields a variance of 31.8. Construct and interpret a 95% confidence interval for the variance in tire pressures for all new SUV tires produced by the manufacturer.

19. Speeds were measured in miles per hour (mph) for a random sample of 26 fastballs thrown by major league pitchers, and the variance in speed was 21.5. Construct and interpret a 98% confidence interval for the variance in speeds of all fastballs thrown by major league pitchers.

20. A company is testing 63 randomly selected compact fluorescent light bulbs to see how long they last. The study results show a standard deviation of 121.4 hours. Construct and interpret a 99% confidence interval for the true standard deviation for the lifetimes of the compact fluorescent light bulbs.

21. A transit system at a large theme park is testing how long it takes for a driver to complete one circuit. A sample of 40 completed circuits shows a standard deviation of 7.7 minutes. Construct and interpret a 98% confidence interval for the standard deviation in completion times for all circuits driven.

22. After testing 31 pairs of noise-reducing earmuffs, the standard deviation of the reductions in noise levels is calculated to be 0.9 decibels. Construct and interpret a 95% confidence interval for the standard deviation of noise level reductions produced by all pairs of this type of noise-reducing earmuffs.

23. Consider the following sample data.

3.2	3.6	2.9	3.0	3.0
3.1	3.2	3.3	2.9	3.3
2.9	3.1	3.4	3.3	3.0

Build and interpret a 99% confidence interval for the population variance.

24. The following sample of weights (in ounces) was taken from 14 boxes of crackers randomly selected from the assembly line.

16.87	16.92	17.01	16.98	16.99	16.92	16.91
17.00	17.01	16.96	16.95	16.94	17.00	16.92

Build and interpret a 98% confidence interval for the population variance for the weights of all boxes of crackers that come off the assembly line.

25. The weights of 89 randomly selected new truck engines from one factory were found to have a standard deviation of 1.59 pounds. Construct and interpret a 95% confidence interval for the population standard deviation of the weights of all new truck engines in this particular factory.

26. A butcher uses a machine that packages ground beef in one-pound portions. A sample of 52 packages of ground beef has a standard deviation of 0.2 pounds. Construct and interpret a 99% confidence interval to estimate the standard deviation of the weights of all packages prepared by the machine.

Minimum Sample Sizes for Estimating Population Variances and Population Standard Deviations

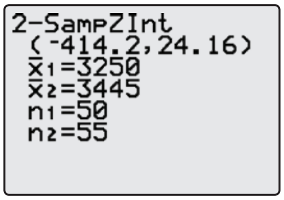
Find the minimum sample size needed to construct a confidence interval with the desired characteristics.

27. Find the minimum sample size needed to be 95% confident that the sample variance is within 20% of the population variance.

28. Find the minimum sample size needed to be 99% confident that the sample variance is within 5% of the population variance.

29. Find the minimum sample size needed to be 99% confident that the sample standard deviation is within 10% of the population standard deviation.

30. Find the minimum sample size needed to be 95% confident that the sample standard deviation is within 1% of the population standard deviation.



The 85% confidence interval for the difference between the two means ranges from approximately –\$414 to \$24. The confidence interval can be written mathematically using either inequality symbols or interval notation, as shown below.

$$-414 < \mu_1 - \mu_2 < 24$$

or

$$(-414, 24)$$

Because the interval ranges from a negative number to a positive one, the interval contains the number 0. In other words, the data do not provide evidence that the two population means are unequal at this level of confidence. Therefore, with 85% confidence, we can say that there is not sufficient evidence to support the claim that using the fuel additive results in a decrease in repair costs.

9.1 Section Exercises

Note: *For all exercises in this section, you may assume that the requirements mentioned in this section are met; namely, all samples are independent, simple random samples, both population standard deviations are known, and either both sample sizes are at least 30 or both population distributions are approximately normal.*

Point Estimates for Differences between Two Population Means

Find the point estimate for the true difference between the population means.

1. $\bar{x}_1 = 14$ and $\bar{x}_2 = 11$

2. Sample 1's mean is 45 and Sample 2's mean is 56.

3. One-mile running times for a sample of 65 participants had a mean of 8.45 minutes. The mean time for a sample of 77 participants is 8.25 minutes.

4. The mean temperature for Group A is 72.4 °F and the mean temperature for Group B is 77.3 °F.

5. The following data represent temperatures in degrees Celsius for random samples of water taken from two different geysers at a National Park.

Water Temperatures of Geyser A

Stem	Leaves
34	6
35	2 4 4 5 5 6 7 7 7 7
36	1 2 3 4 4 5 5 5 6 7 8 8 9 9
37	0 0 0 0 1 1 2 2 2 2 5 6 6 9
38	0 0 1 1 2

Key: 34 | 6 = 34.6 °C

Water Temperatures of Geyser B

Stem	Leaves
33	3 3 4 4 5 5 6 7 8 8 9 9
34	0 0 1 1 2 2 2 3 4 5 5 7 9 9
35	2 3 3 3 6 6 8 8 8
36	5
37	2 2 4 4 5 8

Key: 33 | 3 = 33.3 °C

6. Weights (in Grams) of Candy Bar A: 93, 92, 93, 94, 92, 93, 95, 94, 93, 93, 92

   Weights (in Grams) of Candy Bar B: 93, 94, 95, 92, 93, 91, 96, 92, 93, 93, 94, 95, 92

Margins of Error of Confidence Intervals for Differences between Two Population Means (σ Known, Independent Samples)

Calculate the margin of error of a confidence interval for the difference between the two population means using the given information.

7. $\sigma_1 = 2.88$, $n_1 = 73$, $\sigma_2 = 3.01$, $n_2 = 99$, $c = 0.99$
8. $\sigma_1 = 5.3$, $n_1 = 31$, $\sigma_2 = 4.9$, $n_2 = 40$, $\alpha = 0.02$
9. $\sigma_1 = 0.36$, $n_1 = 88$, $\sigma_2 = 1.09$, $n_2 = 83$, $c = 0.95$
10. $\sigma_1 = 11.54$, $n_1 = 115$, $\sigma_2 = 10.65$, $n_2 = 122$, $\alpha = 0.01$

Confidence Intervals for Differences between Two Population Means (σ Known, Independent Samples)

Construct and interpret each specified confidence interval.

11. A local pizza shop claims to have a shorter delivery time than its competitor, a national chain. A local reporter collects data on random samples of deliveries from the local shop and from the competitor. Based on 48 deliveries, the local store has a mean delivery time of 29 minutes. Assume that the population standard deviation of the local shop's delivery times is 8 minutes. Based on 52 deliveries, the national chain store has an average delivery time of 31 minutes. Assume that the population standard deviation of the chain store's delivery times is 9 minutes. Construct and interpret a 90% confidence interval for the true difference between the mean delivery times for the two restaurants.

12. Dogsled drivers, known as mushers, use several different breeds of dogs to pull their sleds. One proponent of Siberian Huskies believes that sleds pulled by Siberian Huskies are faster than sleds pulled by other breeds. He times 35 teams of Siberian Huskies on a particular short course, and they have a mean time of 5.3 minutes. The mean time on the same course for 32 teams of other breeds of sled dogs is 6.1 minutes. Assume that the times on this course have a population standard deviation of 1.1 minutes for teams of Siberian Huskies and 1.9 minutes for teams of other breeds of sled dogs. Construct and interpret a 99% confidence interval for the true difference between the mean times on this course for teams of Siberian Huskies and teams of other breeds of sled dogs.

13. There are two elementary schools in Caleb's community. Third graders in both schools recently took the same standardized exam. A random sample of 45 third graders from the first school has a mean exam score of 75. A random sample of 39 third graders from the second school has a mean score of 78. Assume that the population standard deviations of the third graders' scores are known to be 10 for the first school and 5 for the second school. Construct and interpret a 95% confidence interval for the true difference between the mean exam scores for the two schools.

14. A contributor for the local newspaper is writing an article for the weekly fitness section. To prepare for the story, she conducts a study to compare the exercise habits of people who exercise in the morning to the exercise habits of people who work out in the afternoon or evening. She selects three different health centers from which to draw her samples. The 49 people she sampled who work out in the morning have a mean of 4.1 hours of exercise each week. The 54 people surveyed who exercise in the afternoon or evening have a mean of 3.7 hours of exercise each week. Construct and interpret a 95% confidence interval for the true difference between the mean amounts of time spent exercising each week by people who work out in the morning and those who work out in the afternoon or evening at the three health centers. Assume that the weekly exercise times have a population standard deviation of 0.7 hours for people who exercise in the morning and 0.5 hours for people who exercise in the afternoon or evening.

15. Russell is doing some research before buying his first house. He is looking at two different areas of the city, and he wants to know if there is a significant difference between the mean prices of homes in the two areas. For the 34 homes he samples in the first area, the mean home price is $168,500. Public records indicate that home prices in the first area have a population standard deviation of $22,950. For the 39 homes he samples in the second area, the mean home price is $171,800. Again, public records show that home prices in the second area have a population standard deviation of $30,995. Construct and interpret a 90% confidence interval for the true difference between the mean home prices in the two areas.

16. A sports news station wanted to know whether people who live in the North or the South are bigger sports fans. For its study, 121 randomly selected Southerners were surveyed and found to watch a mean of 4.1 hours of sports per week. In the North, 137 randomly selected people were surveyed and found to watch a mean of 3.9 hours of sports per week. Find and interpret a 99% confidence interval for the true difference between the mean numbers of hours of sports watched per week for the two regions if the South has a population standard deviation of 1.8 hours per week and the North has a population standard deviation of 1.9 hours per week.

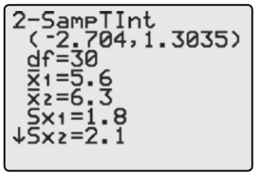
Step 3: Subtract the margin of error from and add the margin of error to the point estimate.

$$\text{Lower endpoint: } (\bar{x}_1 - \bar{x}_2) - E = -0.7 - 2.003538 \approx -2.7$$

$$\text{Upper endpoint: } (\bar{x}_1 - \bar{x}_2) + E = -0.7 + 2.003538 \approx 1.3$$

This confidence interval can be expressed mathematically as (–2.7, 1.3).

TI-83/84 Plus: We can compute the confidence interval in one step by using the calculator. Since we already have the sample statistics, go to STAT > TESTS and choose 2-SampTInt. The general formula for this computation is 2-SampTInt($\bar{x}_1$, s_1, n_1, $\bar{x}_2$, s_2, n_2, C-Level, Pooled). If the data is pooled, select Yes. If the data is not pooled, select No. All of the other statistics are given to us in the problem.



$\bar{x}_1 = 5.6$; $s_1 = 1.8$; $n_1 = 12$

$\bar{x}_2 = 6.3$; $s_2 = 2.1$; $n_2 = 20$

C-Level = 0.99

Pooled = Yes

The confidence interval returned by the calculator is (–2.7, 1.3) as shown in the margin.

Therefore, we are 99% confident that the population mean birth weight for babies whose mothers took the pain reliever while pregnant is between 2.7 pounds less than and 1.3 pounds more than the population mean birth weight for babies whose mothers didn't take the pain reliever. Since the confidence interval contains 0, there is not sufficient statistical evidence to indicate that the mean birth weights are different for the two populations. Thus, we cannot conclude that the pain reliever causes lower birth weights.

9.2 Section Exercises

Note: *For all exercises in this section, you may assume that the requirements mentioned in this section are met; namely, all samples are independent, simple random samples, both population standard deviations are unknown, and either both sample sizes are at least 30 or both population distributions are approximately normal.*

Critical Values for Confidence Intervals for Differences between Two Population Means (σ Unknown, Independent Samples)

Determine the critical value for a confidence interval for the difference between the two population means using the given information.

1. $c = 0.95$, population variances assumed to be equal, $n_1 = 11$, $n_2 = 20$

2. $c = 0.99$, population variances assumed to be equal, $n_1 = 9$, $n_2 = 6$

3. $c = 0.90$, population variances assumed to be unequal, $n_1 = 14$, $n_2 = 12$

4. $c = 0.95$, population variances assumed to be unequal, $n_1 = 25$, $n_2 = 26$

5. $c = 0.95$, population variances assumed to be unequal, $n_1 = 4$, $n_2 = 7$

6. $c = 0.99$, population variances assumed to be equal, $n_1 = 22$, $n_2 = 18$

7. $\alpha = 0.02$, population variances assumed to be equal, $n_1 = 13$, $n_2 = 8$
8. $\alpha = 0.05$, population variances assumed to be equal, $n_1 = 11$, $n_2 = 16$

Margins of Error of Confidence Intervals for Differences between Two Population Means (σ Unknown, Independent Samples)

Calculate the margin of error of a confidence interval for the difference between the two population means using the given information.

9. Population variances assumed to be unequal, $t_{0.01} = 2.764$, $s_1 = 6$, $n_1 = 16$, $s_2 = 5$, $n_2 = 11$
10. Population variances assumed to be equal, $t_{0.025} = 2.021$, $s_1 = 2$, $n_1 = 22$, $s_2 = 3$, $n_2 = 20$
11. Population variances assumed to be equal, $\alpha = 0.01$, $s_1 = 10$, $n_1 = 15$, $s_2 = 13$, $n_2 = 10$
12. Population variances assumed to be unequal, $c = 0.90$, $s_1 = 2$, $n_1 = 6$, $s_2 = 5$, $n_2 = 5$

Confidence Intervals for Differences between Two Population Means (σ Unknown, Independent Samples)

Construct and interpret each specified confidence interval.

13. Different breeds of dogs produce different-sized litters. Wesley believes that his Coonhounds will have litters that are twice as big as his Schnauzers. He has ten female Coonhounds and the mean size of their most recent litters was 11.0 puppies with a standard deviation of 3.0 puppies. Wesley also has nine female Schnauzers, whose most recent litters had a mean litter size of 5.0 puppies with a standard deviation of 2.0 puppies. Assume that the population variances are not the same. Create and interpret a 90% confidence interval to estimate the true difference between the mean sizes of the litters of Wesley's Coonhounds and Schnauzers.
14. Midwives claim that a water birth reduces the amount of pain the mother perceives. To test this theory, one midwife asks six of her clients who had a water birth and eight of her clients who did not have a water birth to rate their pain on a scale of 1 to 10. The mean pain score of mothers who had a water birth was a 7.0 with a standard deviation of 1.2. The mean pain score of mothers who did not have a water birth was 9.0 with a standard deviation of 0.9. Construct and interpret an 80% confidence interval for the true difference between the pain scores of mothers who had a water birth and mothers who did not. Assume that the variances of the two populations are not equal.
15. Steve believes that his wife's cell phone battery does not last as long as his cell phone battery. On eight different occasions, he measured the length of time his cell phone battery lasted and calculated that the mean was 24.3 hours with a standard deviation of 6.1 hours. He measured the length of time his wife's cell phone battery lasted on nine different occasions and calculated a mean of 22.8 hours with a standard deviation of 8.3 hours. Construct and interpret a 95% confidence interval for the true difference in battery life between Steve's cell phone and his wife's. Assume that the population variances are the same.
16. Dentists believe that a diet low in sugary foods can reduce the number of cavities in children. Ten children whose diets are believed to be high in sugar are examined and the mean number of cavities is 3.0 with a standard deviation of 0.7. Twelve children whose diets are believed to be low in sugar are examined and the mean number of cavities is 1.8 with a standard deviation of 0.6. Construct and interpret a 99% confidence interval for the true difference between the mean numbers of cavities for children whose diets are high in sugar and those whose diets are low in sugar. Assume that the variances of the two populations are the same.

17. A company that manufactures baseball bats believes that its new bat will allow players to hit the ball 30 feet farther than its current model. The owner hires a professional baseball player known for hitting home runs to hit ten balls with each bat and he measures the distance each ball is hit to test the company's claim. The results of the batting experiment are shown in the following table. Construct and interpret a 90% confidence interval for the true difference between the mean distance hit with the new model and the mean distance hit with the older model. Assume that the variances of the two populations are the same.

Hitting Distance (in Feet)

New Model	Old Model
235	200
240	210
253	231
267	218
243	210
237	209
250	210
241	229
251	234
248	231

18. A marketing firm is doing research for an Internet-based company. It wants to appeal to the age group of people who spend the most money online. The company wants to know if there is a difference in the mean amount of money people spend per month on Internet purchases depending on their age bracket. The marketing firm looked at two age groups, 18–24 years and 25–30 years, and collected the data shown in the following table. Assume that the population variances are not the same. Construct and interpret an 80% confidence interval to estimate the true difference between the mean amounts of money per month that people in these two age groups spend on Internet purchases.

Internet Spending per Month

	18–24 Years	25–30 Years
Mean Amount Spent	\$52.00	\$45.49
Standard Deviation	\$21.30	\$13.19
Sample Size	18	25

19. Is it worth pursuing a doctoral degree in education if you already have an undergraduate degree? One way to help make this decision is to look at the mean incomes of these two groups. Suppose that 19 people with bachelor's degrees in education were surveyed. Their mean annual salary was \$28,500 with a standard deviation of \$6800. Eleven people with doctoral degrees in education were found to have a mean annual salary of \$55,500 with a standard deviation of \$10,200. Assume that the population variances are not the same. Construct and interpret a 90% confidence interval to estimate the true difference between the mean salaries for people with doctoral degrees and undergraduate degrees in education.

20. In evaluating his teaching, a literature professor decides to have a random sample of 15 students rate him on a scale of 1–10 with 10 being excellent. After getting the results, a mean of 6.2 with a standard deviation of 1.5, he decides to make an effort to improve his teaching. Next semester, for a random sample of 13 students, the results are a mean of 7.5 with a standard deviation of 1.6. Construct and interpret a 95% confidence interval to estimate the true difference between the literature professor's mean ratings from semester to semester. You can assume that the population variances are the same.

21. To conform to market trends, a prominent syrup manufacturer is changing the design of its syrup bottle from a tall slender bottle to a shorter rounder bottle that fits more easily in a microwave oven. The shareholders are concerned that the mean amount of syrup in each bottle remains the same. A sample of 16 bottles of the older design has a mean capacity of 36.20 fluid ounces (fl oz) with a standard deviation of 0.90 fl oz, and a sample of 20 bottles from the new design has a mean capacity of 35.90 fl oz with a standard deviation of 0.73 fl oz. Because the same machine is being used to fill the bottles, you may assume that the population variances are the same. Construct and interpret a 90% confidence interval to estimate the difference between the mean volumes for the bottles.

22. A researcher is interested in exploring the relationship between calcium intake and weight loss. Two different groups, each with 25 dieters, are chosen for the study. Group A is required to follow a specific diet and exercise regimen, and also take a 500-mg supplement of calcium each day. Group B is required to follow the same diet and exercise regimen, but with no supplemental calcium. After six months on the program, the members of Group A had lost a mean of 12.7 pounds with a standard deviation of 2.2 pounds. The members of Group B had lost a mean of 10.8 pounds with a standard deviation of 2.0 pounds during the same time period. Assume that the population variances are not the same. Create and interpret a 95% confidence interval to estimate the true difference between the mean amounts of weight lost by dieters who supplement with calcium and those who do not.

23. Gavin likes to grow tomatoes in the summer, and each year he experiments with ways to improve his crop. This summer, he wants to determine whether the new fertilizer he has seen advertised will increase the mean number of tomatoes produced per plant. Being careful to control the conditions in his garden, he uses the old fertilizer on 12 tomato plants and the new fertilizer on the other 8 plants. Over the course of the growing season, he calculates a mean of 18.2 tomatoes per plant for the old fertilizer, with a standard deviation of 1.9 tomatoes. He calculates a mean of 21.4 tomatoes per plant for the new fertilizer, with a standard deviation of 2.6 tomatoes. Create and interpret a 90% confidence interval for the true difference in tomato production. Assume that the population variances are the same.

24. A retailer is interested in comparing the shopping habits of customers as different types of music play throughout the store. One day when the store played slow instrumental music, the 17 customers who made purchases spent a mean of \$84 with a standard deviation of \$20. The next day, the store played upbeat instrumental music, and the 13 customers who made purchases spent a mean of \$93 with a standard deviation of \$22. Construct and interpret a 99% confidence interval for the true difference between the mean amounts spent by customers while listening to different types of music in the store. Assume that the population standard deviations are the same.

25. A researcher wants to know who reads more, Millenials or GenXers. She surveys 28 Millenials and finds that they read a mean of 2.5 books a month, with a standard deviation of 0.9 books. For the 25 members of Generation X that she surveys, she finds that they read a mean of 2.1 books per month with a standard deviation of 0.8 books. Assuming that the population standard deviations are different, construct and interpret a 90% confidence interval for the true difference between the mean numbers of books read each month by these two generations.

26. A manufacturer of bicycle tires wishes to perform a test of its new tire prototype before putting it into consideration for mass production. They wish to know if there is a significant difference in the lifetime of the new tires versus their standard model. The new tires are given to a group of 15 cyclists and they are able to ride for an average of 2040 miles before needing replacement, with a standard deviation of 22 miles. Another group of 14 cyclists is given standard tires from the company, and they are able to ride on average 1895 miles before needing replacement, with a standard deviation of 20 miles. Construct and interpret the 99% confidence interval for the true difference in mileage between the two tires. Assume that the variances of the tires produced from the same company have the same population standard deviations.

Step 3: Subtract the margin of error from and add the margin of error to the point estimate.

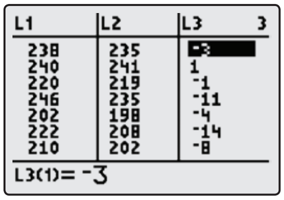
Finally, add and subtract E from $\bar{d}$ to obtain the endpoints of the interval.

$$\text{Lower endpoint: } -10.461538 - 6.983759 \approx -17.4$$

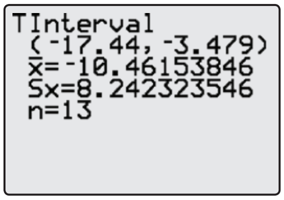
$$\text{Upper endpoint: } -10.461538 + 6.983759 \approx -3.5$$

Hence we find the 99% confidence interval to be (–17.4, –3.5).

TI-83/84 Plus: Begin by entering the data in the lists L1 and L2. Let L1 be the "before" levels and L2 be the "after" levels. Since we need the differences between the pairs of data, we use L3 to calculate those differences for us. To do so, highlight L3 and enter the formula to subtract the "before" levels from the "after" levels (L2−L1) by using the 2ND button to select the lists. The screenshot in the margin shows how the data and the paired differences will appear in the calculator.



Now that we have the paired differences in L3, we can create a one-sample t-interval using those paired differences as our raw data. Press STAT, scroll to TESTS, and choose option TInterval. We want to calculate the confidence interval from the data, so choose the Data option. Our data are in List 3, so enter L3 by pressing 2ND and then 3. The frequency of the data (Freq) is the default value, which is 1. Also, we want a 99% confidence interval, so enter 0.99 for C-Level. Highlight Calculate and press ENTER. The calculator returns a confidence interval of (–17.4, –3.5) as shown in the margin.



Note that both endpoints of the confidence interval are negative, which indicates that the mean cholesterol level decreased significantly. We are 99% confident that, after taking the new cholesterol-lowering drug for four weeks, the mean decrease in the total cholesterol levels for the population from which the participants were sampled is between 3.5 and 17.4 mg/dL.

9.3 Section Exercises

Note: *For all exercises in this section, you may assume that the requirements mentioned in this section are met; namely, the samples are dependent samples of paired data, and the population distributions of the paired differences are approximately normal.*

Means and Standard Deviations of Paired Differences

Calculate $\bar{d}$ and s_d for each set of paired data.

1.

Sample A	22	21	19	20	17	18	20	19	22
Sample B	24	23	23	19	20	21	23	19	23

2.

Sample 1	2	1	1	2	1	1	2	1	2
Sample 2	4	3	2	1	2	2	2	1	3

3.

Diastolic Blood Pressure Readings for Patients at Consecutive Check-ups								
Check-up 1	78	79	91	83	79	92	45	68
Check-up 2	67	85	90	84	86	96	51	66

4.

Number of Minutes Taken to Complete a 5K Race for 9 Runners									
Fall 5K	18	23	24	19	17	22	18	16	20
Spring 5K	24	23	23	19	20	21	23	19	23

Confidence Intervals for Means of Paired Differences for Two Populations (σ Unknown, Dependent Samples)

Formula practice; construct a confidence interval for the mean of the paired differences for the two populations using the given information.

5. $n = 41, \quad \bar{d} = 2.230, \quad \alpha = 0.01, \quad s_d = 0.567$

6. $n = 30, \quad \bar{d} = 12.0, \quad \alpha = 0.05, \quad s_d = 2.7$

7. $n = 5, \quad \bar{d} = 1.14, \quad \alpha = 0.10, \quad s_d = 1.30$

8. $n = 25, \quad \bar{d} = 3.40, \quad \alpha = 0.05, \quad s_d = 1.08$

Construct and interpret each specified confidence interval.

9. To determine if a new cold medicine works better than the traditional cold medicine, 16 people who had a cold volunteered for a study. The volunteers were matched based on age to create eight pairs. A double-blind study was constructed where one of the volunteers in the pair was given the new cold medicine and the other member of the pair was given the traditional medicine. The duration of the cold (in days) was measured for each person and the results are shown in the following table. Construct and interpret a 99% confidence interval for the true mean difference between the durations of a cold for those taking the traditional medication and those taking the new medication.

Duration of Cold (in Days)								
New Medicine	4	5	3	6	3	4	5	7
Traditional Medicine	6	5	4	8	4	5	7	7

10. To determine if his teaching method increases students' learning, a professor administers a pretest to his class at the beginning of the semester and then a posttest at the end of the semester. The results from 15 randomly chosen students are given below. Construct and interpret a 95% confidence interval for the true mean difference between the scores to determine if the teaching method increases students' knowledge of the course material.

Test Scores

Pretest	Posttest
60	80
61	87
65	91
71	97
68	89
67	86
65	85
62	83
63	89
68	93
69	94
70	99
65	92
62	91
57	83

11. A personal trainer believes that walking every day can produce the same health benefits as jogging. Ten volunteers are paired based on significant characteristics; half of the group is asked to walk every day and half of the group is asked to jog every day. The amounts of weight lost (in pounds) over a 30-day period are recorded in the following table. Construct and interpret a 95% confidence interval for the true mean difference between the amount of weight lost by jogging and the amount of weight lost by walking.

Weight Loss (in Pounds)					
Walking	8	9	10	7	9
Jogging	10	12	14	9	12

12. Researchers have developed a method to improve memory. To test their method, 12 participants are asked to memorize a list of words and the number of words remembered correctly is recorded. The participants are then taught the method to improve their memory and are asked to memorize another list of words and the number of words remembered correctly is again recorded. The results are shown in the following table. Construct and interpret a 98% confidence interval to estimate the true mean increase in the number of words that people can memorize after learning the memory method.

Number of Words Memorized	
Before	After
8	13
5	16
6	12
7	16
6	15
3	13
8	17
10	20
12	19
6	17
8	18
5	13

13. A pharmaceutical company is running tests to see how well its new drug lowers cholesterol. Ten adults volunteer to participate in the study. The total cholesterol level of each participant (in mg/dL) is recorded once at the start of the study and then again after three months of taking the drug. The results are given in the following table. Construct and interpret a 99% confidence interval for the true mean difference between the cholesterol levels for people who take the new drug.

Total Cholesterol Levels (in mg/dL)	
Initial Level	Level after Three Months
210	201
200	195
215	208
194	197
206	200
221	203
203	190
189	188
208	210
211	210

14. An infomercial claims that its new cooking device will dramatically reduce the time you spend preparing meals. Wondering if the claim is true, you set out to determine how much time the new cooking device will really save someone on average. Eight people who have purchased the item agree to participate in your study, and they each estimate the time they spent cooking dinner before they bought the new device and then after they began using it. Use the table of results to create and interpret a 90% confidence interval for the true mean change in the amount of time spent cooking dinner by using the infomercial's item.

Time Spent Preparing Dinner (in Minutes)

Without Device	With Device
50	45
60	50
45	30
30	30
45	60
50	40
20	15
25	30

15. Philip wants to take a speed-reading course, but his wife thinks that it is a waste of time. To convince her that the course will really change the way that he reads, Philip decides to conduct an informal study. He polls seven people, asking them to tell the number of pages they were able to read in an hour before and then after they took the course. The results he obtained are found in the following table. Construct and interpret a 95% confidence interval for the true mean increase in reading speeds for people who have taken the speed-reading course.

Number of Pages Read in One Hour

Before Course	35	50	45	50	60	70	65
After Course	50	60	80	70	85	100	90

16. A home improvement show gives tips on ways to improve your house before it is listed for sale. The show claims that their tips will help your house sell faster. To test the show's claim, a researcher found five pairs of houses that were similar in condition, area, and asking price. The homeowners from one house in each pair were asked to follow the tips given in the show before putting their houses on the market. (No major renovations were allowed.) The researchers then kept tabs on the subsequent length of time that it took for each house to sell. Use the following results to find and interpret a 90% confidence interval for the true mean difference between the numbers of weeks required to sell a house for homeowners who follow the show's tips and those who do not follow the tips.

Number of Weeks to Sell a House

Without Tips	9	10.5	3	5.5	14.5
With Tips	7	6	5	1	4

17. During the fall semester of biology, the instructor teaching the class became ill and had to have another instructor stand in for him. After the students had taken a test under both instructors, the chairman of the biology department wanted to know if there was a significant difference in the performance of students under the different instructors. The results from each test are given in the table below. Construct and interpret a 90% confidence interval for the true mean difference between students' test scores under the two instructors.

Students' Test Scores

Instructor A	58	82	78	91	83	77	45	87	94	92	68	77
Instructor B	52	60	83	90	87	70	48	81	90	90	71	70

18. At the beginning of an elementary physical education class, students are asked to do as many sit-ups as possible in a one-minute period. After practicing the proper technique for doing sit-ups, the students are once again timed to see if they increased the number of sit-ups they can do in one minute. Use the table of results to create and interpret a 95% confidence interval for the true mean change in the number of sit-ups that students were able to complete in one minute after the training.

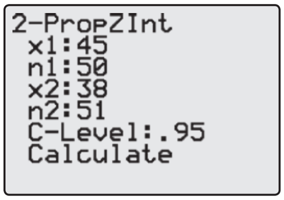
Number of Sit-Ups								
Before	10	20	15	30	31	40	20	25
After	28	30	27	42	50	45	19	36

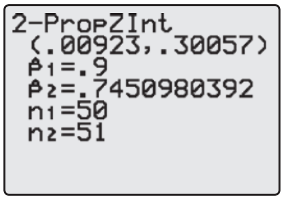
19. A learn-to-type software program claims that it can improve your typing skills. To test the claim and possibly help yourself out, you and three of your friends decide to try the program and see what happens. Use the table below to construct and interpret an 80% confidence interval for the true mean change in the typing speeds for people who have completed the typing program.

Typing Speeds (in Words per Minute)	
Before	**After**
42	54
50	52
37	41
22	30

20. After students were not doing so well in her math class, Ms. Comeaux decided to try a different approach and use verbal positive reinforcement at least once every hour. Use the following results to find and interpret a 99% confidence interval for the true mean change in the students' test scores after Ms. Comeaux started using positive reinforcement.

Students' Test Scores	
Without Reinforcement	**With Reinforcement**
66	72
57	55
61	80
72	79
38	49
47	56





Notice that the calculator gives the same interval as the interval that is calculated by hand but with more decimal places.

The confidence interval can be written mathematically using either inequality symbols or interval notation, as shown below.

$$0.009 < p_1 - p_2 < 0.301$$

or

$$(0.009,\ 0.301)$$

Therefore, we are 95% confident that the percentage of students who passed the class is between 0.9% and 30.1% higher for the population of students who used the new instructional technology (Population 1) than for the population of students who did not use the technology (Population 2). Thus, with 95% confidence, the professor can conclude that the new instructional technology does improve students' scores.

9.4 Section Exercises

Necessary Conditions for Using the Normal Distribution to Compare Two Population Proportions

Verify that the normal distribution can be used to compare the population proportions using the given information, or show how the conditions have not been met.

1. $n_1 = 130,\quad n_2 = 200,\quad \hat{p}_1 \approx 0.869,\quad \hat{p}_2 = 0.72$
2. $n_1 = 1100,\quad n_2 = 1200,\quad \hat{p}_1 = 0.05,\quad \hat{p}_2 = 0.01$
3. $n_1 = 23,\quad n_2 = 14,\quad \hat{p}_1 \approx 0.435,\quad \hat{p}_2 \approx 0.714$
4. $n_1 = 19,\quad n_2 = 12,\quad \hat{p}_1 \approx 0.158,\quad \hat{p}_2 \approx 0.333$

Point Estimates for Differences between Two Population Proportions

Find each specified point estimate.

5. Sample 1 has 17 "yes" responses out of 97 responses in the sample, and Sample 2 has 46 "yes" responses out of 131 responses in the sample. Calculate the point estimate for the difference between the population proportions of "yes" responses.
6. A random sample of records from Public School A shows that 61 out of 156 students started first grade having already lost their first tooth, while at Public School B, 46 out of a random sample of 121 students had lost their first tooth before entering first grade. Find the point estimate for the difference between the population proportions of students who have lost their first tooth before entering first grade at the two schools.
7. Given that the first sample had 15 broken eggs out of 360 eggs in the sample and that the second sample had 12 broken eggs out of 540 eggs, find the point estimate for the difference between the population proportions of broken eggs.

8. Find the point estimate for the difference between the population proportions of "no" votes given the following information. Sample A: 16 no, 32 yes. Sample B: 43 no, 55 yes.

Margins of Error of Confidence Intervals for Differences between Two Population Proportions

Calculate the margin of error of a confidence interval for the difference between the two population proportions using the given information.

9. $n_1 = 130$, $n_2 = 200$, $\hat{p}_1 \approx 0.869231$, $\hat{p}_2 = 0.72$, 95% level of confidence

10. $n_1 = 1100$, $n_2 = 1200$, $\hat{p}_1 = 0.05$, $\hat{p}_2 = 0.01$, $\alpha = 0.05$

11. $n_1 = 88$, $n_2 = 74$, $\hat{p}_1 \approx 0.431818$, $\hat{p}_2 \approx 0.608108$, $c = 0.99$

12. $n_1 = 24$, $n_2 = 39$, $\hat{p}_1 \approx 0.542$, $\hat{p}_2 \approx 0.641$, $c = 0.90$

13. The following table shows a random sample of data collected by a state highway patrolman on whether the driver in the vehicle in a traffic stop was wearing a seat belt. Group A consists of vehicles that were stopped for what the officer considered serious offenses, and Group B consists of vehicles that were stopped for minor offenses. Using a 99% level of confidence, calculate the margin of error for the true difference between the proportions of drivers who do not wear their seat belts for the populations of drivers stopped for major offenses and those stopped for minor offenses.

Seat Belt Use

	No	Yes
Group A	73	486
Group B	159	453

14. The data below represent a random sample of students from a community college (which would only have freshmen and sophomores), showing how many students in the sample are attending school without financial aid and with financial aid. Calculate the margin of error of the 90% confidence interval for the true difference between the proportions of freshmen and sophomores who are attending school without financial aid.

Financial Aid

	Freshmen	Sophomores
Without Aid	36	68
With Aid	72	36

15. The data below are the proportions of respondents from surveys conducted in January and May who said they feel strongly that the mayor is doing a good job.

January: 365 out of 500

May: 402 out of 600

Using a 90% level of confidence, calculate the margin of error for the true difference between the population proportions of people who feel strongly that the mayor is doing a good job.

Confidence Intervals for Differences between Two Population Proportions

Construct and interpret each specified confidence interval.

Note: *You may assume the requirements mentioned in this section are met; namely, all samples are independent, simple random samples, the conditions for a binomial distribution are met, and the sample sizes are large enough to ensure that* $n_1\hat{p}_1 \geq 10$, $n_1\left(1-\hat{p}_1\right) \geq 10$, $n_2\hat{p}_2 \geq 10$, and $n_2\left(1-\hat{p}_2\right) \geq 10$.

16. Doctors at a fertility clinic wish to determine if taking fertility drugs increases the chances of a multiple birth (having two or more babies at once). Doctors record the number of multiple births and the number of single births for a sample of patients taking fertility drugs and a sample of patients not taking fertility drugs. The data are shown below. Construct and interpret a 99% confidence interval for the true difference between the proportions of multiple births for women taking fertility drugs and those who are not.

Fertility Drugs

	With Drugs	Without Drugs
Single Birth	32	43
Multiple Birth	12	13

17. Psychiatrists wish to determine if there is a higher incidence of divorce among couples in which one of the spouses has suffered a serious head injury than among the general population. Of 51 couples who were married at the time one of the spouses had a serious head injury, 20 are still married. Of 50 randomly selected couples in which no head injury occurred, 24 have remained married during this same time period. Construct and interpret a 95% confidence interval for the true difference between the divorce rates of couples in which a head injury occurred and the general population.

18. Do a larger percentage of Southerners than Northerners attend church on a weekly basis? Random samples of Northerners and Southerners are interviewed about their church attendance and the results of the survey are shown below. Construct and interpret a 90% confidence interval for the true difference between the percentages of Northerners and Southerners who attend church on a weekly basis.

Church Attendance

	Northerners	Southerners
Attend Church Weekly	12	35
Do Not Attend Church Weekly	34	16

19. For his senior research project, Ryan is investigating factors that contribute to high school dropout rates. One factor of interest to him is whether kids who are involved in team sports as children are less likely to drop out of high school. He surveys 41 high school graduates and finds that 30 were involved in team sports as children. He surveys 38 high school dropouts and finds that 23 were involved in team sports as children. Construct and interpret a 90% confidence interval for the true difference between the proportions of high school graduates and dropouts who participated in team sports as children.

20. A hedge fund analyst wishes to know which is more open to risk in their investment portfolios, traditional banks or credit unions. He sends out a survey to 100 traditional banks and 100 credit unions, asking them whether they are open to a specified level of risk in their investment portfolios. In response to the survey, 49 of the 100 traditional banks in the sample said yes, and 52 of the 100 credit unions in the sample said yes. Construct and interpret a 95% confidence interval for the true difference between the proportions of traditional banks and credit unions who are open to risk in their investment portfolios.

21. For his senior psychology project, Robert wants to explore which generation is more honest, Baby Boomers or Millennials. To do this, he leaves a wallet filled with money and identification cards in a local store and hides a video camera to record the results. He repeats the experiment 80 times. Of the 37 Millennials who picked up the wallet, 30 returned it to Robert. Of the 43 Baby Boomers who found the wallet, 29 returned it to him. Construct and interpret a 90% confidence interval for the true difference between the proportions of Baby Boomers and Millennials who will return a lost wallet.

22. Ann, a teacher, is concerned with the obsession many of her students have with video games. She is afraid that the video games have a negative impact on her students' performance in the classroom. To test her theory, Ann sends home a letter explaining the study, and 82 parents agree to let their children participate. Ann then randomly divides the students into two equal groups. Group A is required to play video games for two hours one evening, while Group B is not allowed any time to play video games. The following day, the students are given a review test over previously learned material. From Group A, 29 students pass the test. From Group B, 34 students pass the test. Construct and interpret a 95% confidence interval for the true difference between the proportions of students who pass the test and play video games for two hours on the night before the test and those who pass the test but do not play games. Use the confidence interval to determine whether the evidence suggests that playing video games negatively impacts students' performance.

23. Frank owns a nursery, and he has had trouble in the past getting his gerbera daisies to bloom. This spring, he has decided to experiment with a new, more expensive fertilizer called FertiGro. Of the 40 plants that he treats with his regular fertilizer, 27 bloom. Of the 35 that he treats with FertiGro, 29 bloom. Construct and interpret a 99% confidence interval for the true difference between the proportions of gerbera daisy plants that bloom for the populations of plants treated with FertiGro and those treated with regular fertilizer.

24. A state politician is interested in knowing how voters in rural areas and cities differ in their opinions about gun control. For his study, 75 rural voters were surveyed, and 41 were found to support gun control. Also included in the study were 75 voters from cities, and 53 of these voters were found to support gun control. Construct and interpret a 90% confidence interval for the true difference between the proportions of rural and city voters who favor gun control.

25. Aaliyah wants to know if there is a difference between the proportions of customers who just order water to drink at two popular restaurants in town. She collected the data in the following table. Construct and interpret a 95% confidence interval for the true difference between the proportions of customers who just order water to drink at Restaurants A and B.

Drink Orders

	Water	Other Beverage
Restaurant A	61	52
Restaurant B	68	72

26. A local city government is trying to promote road safety by encouraging drivers to buckle up. Its campaign director is trying to decide to which age group she should direct most of the promotions. She believes that fewer older adults buckle up, as they came from a generation who grew up without seat belts. She surveyed 49 senior adults and found that 30 of them buckle up on a regular basis. She then surveyed 52 middle-aged adults and found that 42 of them buckle up on a regular basis. Construct and interpret a 90% confidence interval to estimate the true difference between the proportions of senior adults and middle-aged adults who wear seat belts on a regular basis.

27. Early-childhood-development studies indicate that the more often a child is read to from birth, the earlier the child begins to read. A local parents' group wants to test this theory and samples families with young children. They find the following results. Construct and interpret a 98% confidence interval to estimate the true difference between the proportions of children who read at an early age when they are read to frequently compared to those who were read to less often, as described in the table of results.

Ages When Children Begin to Read

	Read to at Least Three Times per Week	Read to Fewer than Three Times per Week
Started Reading by Age 4	46	31
Started Reading after Age 4	40	57

9.5 Section Exercises

Note: *For all exercises in this section, you may assume that the requirements mentioned in this section are met; namely, the samples are independent, simple random samples, and both population distributions are approximately normal.*

Point Estimates for Ratios of Two Population Variances

Calculate the point estimate for the ratio of the population variances.

1. $n_1 = 13, \quad n_2 = 17, \quad s_1^2 = 3.467, \quad s_2^2 = 2.903$

2. $n_1 = 23, \quad n_2 = 20, \quad s_1^2 = 0.689, \quad s_2^2 = 0.542$

3. $n_1 = 8, \quad n_2 = 10, \quad s_1^2 = 12.103, \quad s_2^2 = 10.874$

4. $n_1 = 26, \quad n_2 = 28, \quad s_1^2 = 1.472, \quad s_2^2 = 1.327$

Critical Values for Confidence Intervals for Ratios of Two Population Variances

Determine the critical values for the left and right endpoints of a confidence interval for the ratio of the two population variances using the given information.

5. $n_1 = 18, \quad n_2 = 23, \quad s_1^2 = 12.470, \quad s_2^2 = 12.205,$ 95% level of confidence

6. $n_1 = 21, \quad n_2 = 20, \quad s_1^2 = 134.943, \quad s_2^2 = 125.908, \quad \alpha = 0.10$

7. $n_1 = 9, \quad n_2 = 9, \quad s_1^2 = 0.974, \quad s_2^2 = 0.903, \quad c = 0.95$

8. $n_1 = 18, \quad n_2 = 19, \quad s_1^2 = 20.461, \quad s_2^2 = 18.321, \quad c = 0.90$

9. $n_1 = 12, \quad n_2 = 13, \quad s_1^2 = 5.100, \quad s_2^2 = 5.098, \quad \alpha = 0.01$

10. $n_1 = 7, \quad n_2 = 6, \quad s_1^2 = 4.110, \quad s_2^2 = 3.873,$ 99% level of confidence

Confidence Intervals for Ratios of Two Population Variances

Construct a confidence interval for the ratio of the two population variances using the given information.

11. $\frac{s_1^2}{s_2^2} = 1.23, \quad F_{\alpha/2} = 2.4499, \quad F_{(1-\alpha/2)} = 0.3556$

12. $\frac{s_1^2}{s_2^2} = 6.31, \quad F_{\alpha/2} = 5.4160, \quad F_{(1-\alpha/2)} = 0.0688$

13. $n_1 = 13, \quad n_2 = 12, \quad \frac{s_1^2}{s_2^2} = 1.14, \quad \alpha = 0.05$

14. $n_1 = 14, \quad n_2 = 18, \quad \frac{s_1^2}{s_2^2} = 2.84, \quad \alpha = 0.05$

15. $n_1 = 9, \quad n_2 = 12, \quad \frac{s_1^2}{s_2^2} = 1.23, \quad \alpha = 0.01$

16. $n_1 = 20, \quad n_2 = 20, \quad \frac{s_1^2}{s_2^2} = 1.87, \quad \alpha = 0.10$

17. $n_1 = 15, \quad n_2 = 16, \quad s_1^2 = 8.455, \quad s_2^2 = 2.897,$ 95% level of confidence

18. $n_1 = 18, \quad n_2 = 17, \quad s_1^2 = 4.067, \quad s_2^2 = 3.903,$ 95% level of confidence

Construct each specified confidence interval and interpret the interval.

19. A medical researcher is trying to determine whether the population variances of systolic blood pressure levels are the same for patients who take a new medication for high blood pressure and patients who do not take the new medication. The control group consists of 20 patients with a sample variance in systolic blood pressure of 124.940. The treatment group, who is taking the new medication, consists of 23 patients with a sample variance in systolic blood pressure of 123.980. Construct and interpret a 90% confidence interval for the ratio of the population variances of systolic blood pressure levels for the two groups.

20. A track coach wants to make sure that two of her star runners are equally consistent in their times for the race. She times 16 of Amy's practice runs and calculates that the sample variance in times (measured in seconds) is 2.560. She then times 15 of Veronika's practice runs and calculates that the sample variance in times is 2.519. Construct and interpret a 90% confidence interval for the ratio of the population variances of Amy's and Veronika's run times.

21. A quality control inspector is testing two machines that are used to fill bags of flour to determine if they are running consistently. In particular, he wants to know whether the population variances of the amounts of flour per bag are equal for the two machines. He measures the weights, in pounds, of 11 bags of flour filled by Machine A and calculates a sample variance of 0.584. He then measures the weights of 10 bags of flour filled by Machine B and calculates a sample variance of 0.499. Construct and interpret a 99% confidence interval for the ratio of the population variances of the weights of the bags of flour for the two machines. If it is important that the two machines have the same variance, should the inspector adjust one of the machines, and if so, which machine needs adjusting?

22. A nutritionist is comparing two new diets, and after running previous tests to determine if the same amount of weight can be lost using both diets, he now wants to know if the amounts of weight lost are consistent for the two diets. In other words, he wants to estimate the ratio of the variances of the amounts of weight lost by people on the two diets. Group A consists of six people on Diet A, and after two months on the diet, the sample variance of the amounts of weight lost (measured in pounds) is 26.041. Group B consists of seven people on Diet B, and after the same two months on their diet, the sample variance of the amounts of weight lost by this group is 25.084. Construct and interpret a 99% confidence interval for the ratio of the population variances of the amounts of weight lost by people on these two diets.

23. A new brand of golf balls claims that using these golf balls can increase your precision. A golf pro tests this claim by hitting 45 golf balls of the old brand and 30 golf balls of the new brand from the same tee on the driving range and measuring the distance, in yards, that each ball travels. The sample variance of the driving distances for the old brand was 29.752 and the sample variance for the new brand was 15.910. Construct and interpret a 90% confidence interval for the ratio of the population variances.

24. A professor wants to make sure that two different versions of a test are equivalent. He decides to compare the variances of the test scores from each version. A sample of 19 scores on Version A has a sample variance of 2.450, while a sample of 20 scores from Version B has a sample variance of 2.391. Construct and interpret a 95% confidence interval for the ratio of the population variances of the scores on the two versions of the test.

25. A graduate student walking around a large college campus notices that students, not faculty, often own the more expensive cars. He collects data from 25 professors and 28 students on the prices of their cars and calculates the variance of the car prices from each sample. The sample variance for the sample of professors' cars is 7849.38 and the sample variance for the sample of students' cars is 3567.90. Construct and interpret a 95% confidence interval for the ratio of the population variances of prices of professors' and students' cars for this college.

26. A researcher wants to compare the consistency with which two marksmen hit the bull's-eye of a target. The first marksman hits an average of 4.5 bull's-eyes per session with a variance of 2.78. The second marksman hits an average of 5.3 bull's-eyes per session with a variance of 2.34. If 20 sessions were recorded for each marksman, construct and interpret a 95% confidence interval for the ratio of the population variances of the numbers of times per session that these two marksmen hit the bull's-eye.

27. A paint technician for an oil company has to make sure that the painted coatings on the insides of the oil tankers have a consistent thickness. He measures the thickness of the coating (in millimeters) at 15 spots inside the first tanker and calculates a sample variance of 0.3918. He then measures the thickness of the coating at 15 spots inside a second tanker sprayed using the same equipment and calculates a sample variance of 0.4231. Construct and interpret a 99% confidence interval for the ratio of the population variances to help determine if there was a significant difference in how consistently the equipment applied the coating in one tanker versus the other.

28. When shopping for a new car, Emily became interested in how consistently cars are priced in her area; she was planning to buy either a Honda Accord or a Toyota Camry. She priced ten different Accords with similar features at several dealerships in her area. The sample variance of the prices for the Accords was 273,529. Emily then priced nine different Camrys with similar features at several different dealerships in her area. The sample variance of the prices for the Camrys was 231,361. Construct and interpret a 99% confidence interval for the ratio of the population variances of the prices for the two different car models.

Example 10.1.8

Determining the Type of Error

A study regarding the effects of television viewing on children reports that children watch a mean of 4.0 hours of television per night. Kiko believes the mean number of hours that children in her neighborhood watch television per night is not 4.0. She performs a hypothesis test and rejects the null hypothesis. Assume that in reality, children in her neighborhood do watch a mean of 4.0 hours of television per night. Did she make an error? If so, what type?

Solution

Begin by writing the null and alternative hypotheses. Kiko wishes to gather data in support of her belief that the mean is not 4.0 hours per night. Therefore, her research hypothesis is written mathematically as $H_a\colon \mu \neq 4.0$. Thus, the null and alternative hypotheses are written as follows.

$$H_0\colon \mu = 4.0$$
$$H_a\colon \mu \neq 4.0$$

The decision was to reject the null hypothesis, when in reality, $\mu = 4.0$, so Kiko rejected a true null hypothesis. This is a Type I error.

Memory Booster

We would usually have no basis for determining if a Type I error or a Type II error has been made because population parameters are usually unknown.

Now that you have had practice distinguishing between Type I and Type II errors, it is important to note that in almost every case, we would not know the "reality" of a situation because population parameters are usually unknown. Thus, we would usually have no basis for determining if a Type I error or a Type II error has been made. Knowing what types of errors exist and how they are related to each other, as well as being able to distinguish between them, allows us to interpret decisions in the context of possible realities.

In conclusion, in this section we discussed the nuances of each step of a hypothesis test as well as the types of errors that might be made when stating the conclusion. In the next several sections, we will discuss the details of determining the appropriate probability distribution to use as well as the test statistic to use for a given population parameter and the various methods for determining the values of that test statistic. In these sections, we will look at complete examples of hypothesis tests from start to finish for several different population parameters. These complete examples will put all the general discussion from this section into specific focus.

10.1 Section Exercises

Fundamentals of Hypothesis Testing

Decide if each statement is true or false. Explain why.

1. When we reject a null hypothesis, we have proven the alternative hypothesis to be true.
2. There are only two possible conclusions in a hypothesis test: reject or fail to reject the null hypothesis.
3. A Type I error is made when a true null hypothesis is rejected.
4. A Type II error is made when we fail to reject a true null hypothesis.
5. The level of significance is the probability of making a Type II error.
6. The probability of making a Type I error is inversely related to the probability of making a Type II error.

Null and Alternative Hypotheses

State the null and alternative hypotheses for each scenario.

7. Based on past sales, a shoe manufacturer considers the mean size of women's shoes to be 7.5. The manufacturer would like to test if this is still the case.

8. Austin's company manufactures test tubes. He needs his tubes to be exactly 4.0 mm in diameter. If they are too narrow or too wide, he must recalibrate his machine. Austin randomly measured 150 tubes off the production line to perform a hypothesis test.

9. A sports analyst is testing his claim that the mean weight of this year's NFL linemen is heavier than in past years. Suppose that over the past five years, the mean weight of NFL linemen was 320.0 pounds.

10. A nationwide study shows that children watch a mean of 3.0 hours of television per day. A mother in Louisiana believes that this is an underestimate for children in her area. She conducts a local survey to test her belief.

11. After reading a headline claiming that half of Americans think space travel will become routine during the next 50 years of space exploration, a science teacher had students survey friends and relatives nationwide to research whether the true proportion is less than half.

12. Madison, an ambulance driver in a small town, believes a typical 24-hour shift receives fewer than 5 emergency calls on average. To test this belief, the number of calls each shift receives is recorded for 10 randomly selected shifts.

13. While continuing to keep abreast of local trends in education, a school administrator read a journal article that reported only 42% of high school students study on a regular basis. The administrator claims that this percentage is too low for the district.

14. The city council of Oxford is thinking of building a road around the city. The council wants to know if the majority of the residents are in favor of the new road before pursuing the issue further.

15. As a general guideline, tap water should contain 3 ppm (parts per million) chlorine. A local council wants to test the chlorine levels on 15 randomly selected days over the next two months to see if the guidelines are being exceeded.

16. Seawater is believed to have a mean fluoride concentration of 1.3 mg/L (milligrams per liter). A marine biologist is concerned that the level of fluoride is too high in a particular area and is killing the ocean life.

Source: "Fluoride in Drinking-Water: Background Document for Development of WHO *Guidelines for Drinking-Water Quality*." World Health Organization. 2004. http://www.who.int/water_sanitation_health/dwq/chemicals/fluoride.pdf (9 Nov. 2011).

Interpreting Conclusions of Hypothesis Tests

Answer each question.

17. A television network has believed for many years that 40% of its viewers are below the age of 22. For marketing purposes, a potential advertiser wishes to test the claim that the percentage is actually less than 40%. After performing the test at the 0.05 level of significance, the advertiser decides to reject the null hypothesis. What does this conclusion lead us to believe about the potential advertiser's claim?

18. A radio station has always believed that the mean age of its listeners is 26. Because their ratings are slipping, the executives need to know if the mean age has changed, so that they can alter their programs accordingly. After information is collected from 329 listeners and a hypothesis test is completed, the radio station executives decide that they should reject the null hypothesis with a 95% level of confidence. Based on this conclusion, should the radio station look into changing its programming?

19. A pharmaceutical company has publicized that approximately 4% of people who take a particular drug experience significant side effects. A researcher is concerned that the percentage is more than 4%, and she decides to test her claim with a hypothesis test. Based on the sample she collects, she decides to fail to reject the null hypothesis at the 0.01 level of significance. What does this conclusion tell us about the researcher's claim?

20. A cellular phone company promotes the claim that its customers spend a mean of \$60 per month on cell phone service. One skeptical college student is convinced that the mean is higher than \$60 per month. He surveys a random sample of the company's customers and performs a hypothesis test to test his claim. In the end, he decides to fail to reject the null hypothesis with $\alpha = 0.01$. What does this conclusion tell us about the college student's claim?

Types of Errors in Hypothesis Testing

For each scenario, determine the type of error that was made, if any. (Hint: Begin by determining the null and alternative hypotheses.)

21. A software company advertises that its software can improve students' grades by 15%. One student conducts a hypothesis test to see if the percentage increase is less than 15%. The conclusion of the hypothesis test is to reject the null hypothesis. If the true percentage increase in grades for all students using the software is 13%, was an error committed? If so, what type?

22. A prominent travel agency claims that the mean cost of a four-day theme park vacation is \$1500 per person. A researcher believes that the mean cost is much higher, and decided to test the theory with a hypothesis test. According to the sample obtained, the decision is to fail to reject the null hypothesis. If, in reality, the mean cost of a four-day theme park vacation is \$1500 was an error made with the hypothesis test? If so, what type?

23. The mathematics department at one university reports that the failure rate for College Algebra is 35% in any given semester. A group of students believes the rate is higher, and they decided to use a hypothesis test to see if they are correct. Based on the sample they obtained, they decided to reject the null hypothesis. If, in reality, the failure rate is 35%, did the students make an error? If so, what type?

24. Dreamfilms Studios boasts that its summer features are so good that half of all tickets sold one summer are for its films. A competing studio claims that less than half of all tickets sold are for Dreamfilms' movies. After the hypothesis test is completed, the conclusion is to fail to reject the null hypothesis. If the true percentage of tickets sold for Dreamfilms Studios' movies that summer was only 48%, was an error committed? If so, what type?

25. An ice cream company must keep the temperature inside its delivery trucks at 30 degrees Fahrenheit. If the temperature is too warm, the ice cream will melt. The company also does not want to allow the temperature to be too low, because it does not want to waste the cost and energy it takes to keep the trucks cool. Periodically the company runs a hypothesis test to ensure that the temperature is as claimed. Suppose that after one hypothesis test, the company fails to reject the null hypothesis. If, in reality, the mean temperature of its trucks is 29 degrees Fahrenheit, was an error committed? If so, what type?

26. A survey of undergraduate college students found that 52% of students think the FAFSA is only for federal financial aid. A local college wants to make sure that their students understand that this is not true. After presenting numerous programs aimed at helping students understand the reasons for filling out the FAFSA forms for financial aid, the college believes that the percentage of students who believe this myth is now lower at their college. A hypothesis test of their claim results in failing to reject the null hypothesis. If, in reality, 52% of students at this college believe that the FAFSA is only for federal financial aid, was an error made? If so, what type?

10.2 Section Exercises

Rejection Regions for Hypothesis Tests for Population Means (σ Known)

Draw the rejection region for the hypothesis test for the population mean with the given hypotheses. Assume that the population standard deviation is known and the sample size is at least 30.

1. H_0: $\mu = 65$ and H_a: $\mu > 65$, $\alpha = 0.05$
2. H_0: $\mu = 5$ and H_a: $\mu \neq 5$, $c = 0.90$
3. H_0: $\mu = 3.2$ and H_a: $\mu < 3.2$, $\alpha = 0.01$
4. H_0: $\mu = 700$ and H_a: $\mu > 700$, $c = 0.98$
5. H_0: $\mu = 0.19$ and H_a: $\mu \neq 0.19$, $c = 0.95$
6. H_0: $\mu = 0.166$ and H_a: $\mu < 0.166$, $\alpha = 0.10$

p-Values for Hypothesis Tests for Population Means (σ Known)

Calculate the p-value for the hypothesis test for the population mean with the given hypotheses and test statistic.

7. H_0: $\mu = 60$ and H_a: $\mu < 60$, $z = -1.48$
8. H_0: $\mu = 1.07$ and H_a: $\mu > 1.07$, $z = 2.27$
9. H_0: $\mu = 2.89$ and H_a: $\mu \neq 2.89$, $z = -2.21$
10. H_0: $\mu = 144$ and H_a: $\mu \neq 144$, $z = 1.65$
11. H_0: $\mu = 0.56$ and H_a: $\mu > 0.56$, $z = 1.61$
12. H_0: $\mu = 760$ and H_a: $\mu < 760$, $z = -2$

Conclusions of Hypothesis Tests for Population Means (σ Known)

Determine the appropriate conclusion for a hypothesis test with the given level of significance.

13. p-value $= 0.0166$, $\alpha = 0.01$
14. p-value $= 0.0197$, $\alpha = 0.02$
15. p-value $= 0.0465$, $\alpha = 0.05$
16. p-value $= 0.0485$, $\alpha = 0.02$
17. p-value $= 0.1190$, $\alpha = 0.10$
18. p-value $= 0.0094$, $\alpha = 0.01$
19. H_0: $\mu = 409$ and H_a: $\mu < 409$, $\alpha = 0.05$, $z = -1.87$
20. H_0: $\mu = 0.65$ and H_a: $\mu > 0.65$, $\alpha = 0.10$, $z = 1.22$
21. H_0: $\mu = 2$ and H_a: $\mu \neq 2$, $\alpha = 0.02$, $z = -2.28$
22. H_0: $\mu = 3010$ and H_a: $\mu \neq 3010$, $\alpha = 0.01$, $z = 2.69$
23. H_0: $\mu = 10.3$ and H_0: $\mu > 10.3$; $\alpha = 0.10$; $z = 1.97$
24. H_0: $\mu = 408$ and H_a: $\mu < 408$; $\alpha = 0.05$; $z = -1.76$

Hypothesis Tests for Population Means (σ Known)

Perform each hypothesis test. For each exercise, complete the following steps.

a. State the null and alternative hypotheses.

b. Determine which distribution to use for the test statistic, and state the level of significance.

c. Calculate the test statistic.

d. Draw a conclusion using the given level of significance and interpret the decision.

25. A manufacturer must test that his bolts are 2.00 cm long when they come off the assembly line. He must recalibrate his machines if the bolts are too long or too short. After sampling 100 randomly selected bolts off the assembly line, he calculates the sample mean to be 1.90 cm. He knows that the population standard deviation is 0.50 cm. Assuming a level of significance of 0.05, is there sufficient evidence to show that the manufacturer needs to recalibrate the machines?

26. The U.S. Energy Information Administration claimed that in 2017, U.S. residential customers used an average of 10,399 kilowatt hours (kWh) of electricity. A local power company believes that residents in their area use more electricity on average than EIA's reported average. To test their claim, the company chooses a random sample of 125 of their customers and calculates that these customers used an average of 10,678 kWh of electricity in the prior year. Assuming that the population standard deviation is 1361 kWh, is there sufficient evidence to support the power company's claim at the 0.05 level of significance?

Source: U.S. Energy Information Administration. "How much electricity does an American home use?" 26 Oct. 2018. https://www.eia.gov/tools/faqs/faq.php?id=97&t=3 (20 Mar. 2019).

27. A wedding website states that the average cost of a wedding in 2017 was $25,764. One concerned bride hopes that the average is less than reported. To see if her hope is correct, she surveys 55 recently married couples and finds that the average cost of weddings in the sample was $23,015. Assuming that the population standard deviation is $7235, is there sufficient evidence to support the bride's hope at the 0.10 level of significance?

Source: The Wedding Report, Inc. "Cost of Wedding." 2019. https://www.costofwedding.com (20 Mar. 2019).

28. A national business magazine reports that the mean age of retirement for women executives is 61.0. A women's rights organization believes that this value does not accurately depict the current trend in retirement. To test this, the group polled a simple random sample of 95 recently retired women executives and found that they had a mean age of retirement of 61.5. Assuming the population standard deviation is 2.5 years, is there sufficient evidence to support the organization's belief at the 0.05 level of significance?

29. The board of a major credit card company requires that the mean wait time for customers when they call customer service is 3.00 minutes. To make sure that the mean wait time is not exceeding the requirement, an assistant manager tracks the wait times of 45 randomly selected calls. The mean wait time was calculated to be 3.40 minutes. Assuming the population standard deviation is 1.45 minutes, is there sufficient evidence to say that the mean wait time for customers is longer than 3.00 minutes with a 98% level of confidence?

30. A survey by Constant Contact reported that small businesses spend 20 hours a week marketing their business. A local chamber of commerce claims that small businesses in their area are not growing because these businesses are spending less than 20 hours a week on marketing. The chamber conducts a survey of 80 small businesses within their state and finds that the average amount of time spent on marketing is 19.3 hours a week. Assuming that the population standard deviation is 4.3 hours, is there sufficient evidence to support the chamber of commerce's claim at the 0.01 level of significance?

Source: Small Business Trends. "Small Businesses Spend 20 Hours Per Week on Marketing." 1 Nov. 2017. https://smallbiztrends.com/2016/09/how-much-time-do-you-spend-marketing-your-business.html (20 Sept. 2019).

Method 2: *p*-Values

The p-value given by the calculator is approximately 0.1912 as shown in the previous screenshot. Since $0.1912 > 0.01$, we have p-value $> \alpha$. Thus, the conclusion is to fail to reject the null hypothesis.

Interpretation: We interpret failing to reject the null hypothesis to mean that the evidence collected is not strong enough at the 0.01 level of significance to say that the machine is working improperly.

10.3 Section Exercises

Rejection Regions for Hypothesis Tests for Population Means (σ Unknown)

Determine the rejection region for a hypothesis test for the population mean using the given information. Assume that the population standard deviation is unknown and the population distribution is approximately normal.

1. $\alpha = 0.05$, $df = 14$, left-tailed test

2. $\alpha = 0.01$, $df = 23$, right-tailed test

3. $c = 0.90$, $df = 18$, H_a: $\mu \neq 18$

4. $c = 0.95$, $df = 26$, H_a: $\mu < 0.18$

5. $\alpha = 0.10$, $n = 10$, H_a: $\mu \neq 102$

6. $\alpha = 0.025$, $n = 6$, H_a: $\mu > 203$

Draw the rejection region for the hypothesis test for the population mean and give the appropriate conclusion for the hypothesis test using the given information. Assume that the population distribution is approximately normal.

7. H_0: $\mu = 195$ and H_a: $\mu \neq 195$, $t = -3.11$, $n = 20$, $\alpha = 0.05$

8. H_0: $\mu = 17$ and H_a: $\mu < 17$, $t = -2.05$, $n = 6$, $\alpha = 0.05$

9. H_0: $\mu = 28$ and H_a: $\mu > 28$, $t = 1.99$, $n = 29$, $\alpha = 0.01$

10. H_0: $\mu = 98$ and H_a: $\mu < 98$, $t = -2.73$, $n = 10$, $\alpha = 0.01$

11. H_0: $\mu = 7$ and H_a: $\mu \neq 7$, $t = -3$, $n = 15$, $\alpha = 0.01$

12. H_0: $\mu = 39$ and H_a: $\mu > 39$, $t = 2.01$, $n = 8$, $\alpha = 0.05$

Hypothesis Tests for Population Means (σ Unknown)

Perform each hypothesis test. For each exercise, complete the following steps. Assume that each population distribution is approximately normal.

a. State the null and alternative hypotheses.

b. Determine which distribution to use for the test statistic, and state the level of significance.

c. Calculate the test statistic.

d. Draw a conclusion and interpret the decision.

13. One cable company claims that it has excellent customer service. In fact, the company advertises that a technician will arrive in approximately 30 minutes after a service call is placed. One frustrated customer believes this is not accurate, claiming that it takes over 30 minutes for the cable technician to arrive. The customer asks a simple random sample of 9 other cable customers how long it has taken for the cable technician to arrive when they have called for one. The sample mean for this group is 33.2 minutes with a standard deviation of 3.4 minutes. Test the customer's claim at the 0.025 level of significance.

14. A parenting magazine reports that the average amount of wireless data used by teenagers each month is 10 Gb. For her science fair project, Ella sets out to prove the magazine wrong. She claims that the mean among teenagers in her area is less than reported. Ella collects information from a simple random sample of 25 teenagers at her high school and calculates a mean of 9.8 Gb per month with a standard deviation of 2.7 Gb per month. Test Ella's claim at the 0.01 level of significance.

15. A children's clothing company manufactures hand-smocked dresses for girls. The length of one particular size of dress is designed to be 26 inches. The company regularly tests the lengths of the garments to ensure quality control, and if the mean length is found to be significantly longer or shorter than 26 inches, the machines must be adjusted. The most recent simple random sample of 28 dresses had a mean length of 26.30 inches with a standard deviation of 0.77 inches. Perform a hypothesis test on the accuracy of the machines at the 0.01 level of significance.

16. A pizza delivery chain advertises that it will deliver your pizza in 20 minutes from when the order is placed. Being a skeptic, you decide to test and see if the mean delivery time is actually more than 20 minutes. For the simple random sample of 7 customers who record the amount of time it takes for each of their pizzas to be delivered, the mean is 22.7 minutes with a standard deviation of 4.3 minutes. Perform a hypothesis test using a 0.05 level of significance.

17. Community college instructors' salaries in one state are very low, so low that educators in that state regularly complain about their compensation. The national mean is $51,878, but instructors from Mississippi claim that the mean in their state is significantly lower. They survey a simple random sample of 23 colleges in the state and calculate a mean salary of $46,005 with a standard deviation of $8833. Test the instructors' claim at the 0.01 level of significance.

Source: Based on 2017-18 NCES data for average 9-month salary for full-time instructional staff nationwide and in Mississippi for colleges where the highest degree is an Associate's degree.

18. It currently takes users a mean of 5 minutes to install the most popular computer program made by RodeTech, a software design company. After changes have been made to the program, the company executives want to know if the new mean is now different from 5 minutes so that they can change their advertising accordingly. A simple random sample of 20 new customers are asked to time how long it takes for them to install the software. The sample mean is 4.1 minutes with a standard deviation of 1.9 minutes. Perform a hypothesis test at the 0.05 level of significance to see if the mean installation time has changed.

10.4 Section Exercises

Necessary Conditions for Using the Normal Distribution in Hypothesis Tests for Population Proportions

Determine whether the normal distribution can be used to perform a hypothesis test for the population proportion in each scenario.

1. After collecting data from a simple random sample of 43 townspeople on whether they approve of the local mayor, a statistics student wants to run a hypothesis test for the population proportion using a 95% level of confidence. Has he collected enough data to test the claim using a normal distribution that the currently accepted approval rating is incorrect if the current belief is that 35% of residents approve of the mayor?

2. An environmentalist wishes to conduct a hypothesis test on the percentage of cars driven in the city that are hybrids. Is it sufficient for him to use a simple random sample of 54 cars to perform a hypothesis test using a normal distribution if hybrids currently account for 4% of the car sales in the country and he claims that the percentage of hybrids in the city is higher than that?

Source: Hirsch, Jerry. Los Angeles Times. "Hybrid vehicle sales speed up." 29 Mar. 2013. https://www.latimes.com/business/la-xpm-2013-mar-29-la-fi-hy-autos-hybrid-20130330-story.html (25 Mar. 2019).

3. A scientist believes that 12 out of every 100 office workers suffer from extreme sensitivity to dust in the workplace. Is it appropriate to use a normal distribution to conduct a hypothesis test at the 99% level of confidence if only 57 office workers are surveyed?

4. In a recent study at a local college, 48 students admitted that they had tried alcohol at least once while under the legal drinking age, while 34 students said they had not. Is there a large enough sample to use a normal distribution to conduct a hypothesis test to see if the percentage of students at the college who admit to drinking underage is the same as the national percentage of 65%?

Source: National Institute on Alcohol Abuse and Alcoholism. "Fall Semester—A Time for Parents To Discuss the Risks of College Drinking." Aug. 2019. https://pubs.niaaa.nih.gov/publications/CollegeFactSheet/CollegeFact.htm (20 Sept. 2019).

Hypothesis Tests for Population Proportions

Perform each hypothesis test. For each exercise, complete the following steps.

a. State the null and alternative hypotheses.
b. Determine which distribution to use for the test statistic, and state the level of significance.
c. Calculate the test statistic.
d. Draw a conclusion using the given level of significance and interpret the decision.

5. One study claimed that a whopping 95% of college students identify themselves as procrastinators. Since the report also claims that only 20% of the general population claim to be procrastinators, one professor believes that the claim regarding college students is too high. The professor conducts a simple random sample of 275 college students and finds that 251 of them identify themselves as procrastinators. Does this evidence support the professor's claim that fewer than 95% of college students are procrastinators? Use a 0.10 level of significance.

Source: Gaille, Brandon. "19 Lazy Procrastination Statistics." 19 May 2017. https://brandongaille.com/17-lazy-procrastination-statistics/ (25 Mar. 2019).

6. The National Academy of Science reported in a 1997 study that 40% of research in mathematics is published by US authors. The mathematics chairperson of a prestigious university wishes to test the claim that this percentage is no longer 40%. He has no indication of whether the percentage has increased or decreased since that time. He surveys a simple random sample of 130 recent articles published by reputable mathematics research journals and finds that 62 of these articles have US authors. Does this evidence support the mathematics chairperson's claim that the percentage is no longer 40%? Use a 0.10 level of significance.

Source: Panel on International Benchmarking of US Mathematics Research and Committee on Science, Engineering, and Public Policy. *International Benchmarking of US Mathematics Research.* National Academy of Sciences, National Academy of Engineering, Institute of Medicine (SEM). 1997. http://www.nap.edu/openbook.php?record_id=9089&page=22 (2 Dec. 2011).

7. Sleep apnea is a condition in which the sufferers stop breathing momentarily while they are asleep. This condition results in lack of sleep and extreme fatigue during waking hours. A current estimate is that 18 million out of the 312.7 million Americans suffer from sleep apnea, or approximately 5.8%. A safety commission is concerned about the percentage of commercial truck drivers who suffer from sleep apnea. They do not have any reason to believe that it would be higher or lower than the population's percentage. To test the claim that the percentage of commercial truck drivers who suffer from sleep apnea is not 5.8%, a simple random sample of 350 commercial truck drivers is examined by a medical expert, who concludes that 30 suffer from sleep apnea. Does this evidence support the claim that the percentage of commercial truck drivers who suffer from sleep apnea is not 5.8%? Use a 0.01 level of significance.

Source: American Sleep Apnea Association. "Sleep Apnea." 2011. http://www.sleepapnea.org/learn/sleepapnea.html (2 Dec. 2011).

Source: US Census Bureau. "American FactFinder." http://factfinder2.census.gov (2 Dec. 2011).

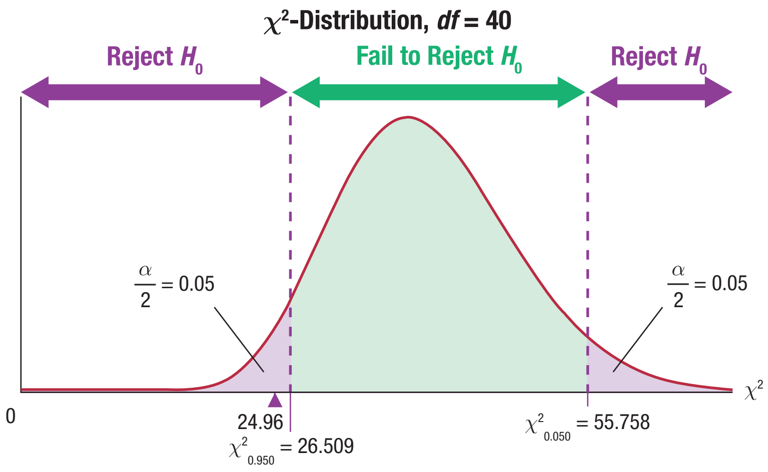
8. In the romantic comedy *Sleepless in Seattle*, Meg Ryan's character, Annie, begins the movie concerned over a report that "a woman over age 40 has a better chance of being killed by a terrorist than of getting married." While this claim is clearly exaggerated, according to snopes.com it is actually based on a true, though flawed, study that "did conclude that the likelihood of marriage for a never-previously-wed, 40-year-old university-educated American woman was 2.6%." To demonstrate that this percentage is too small, Annie uses her resources at the Baltimore Sun to conduct a simple random sample of 450 never-previously-wed, university-educated, American women who were single at the beginning of their 40s and who are now 45. Of these women, 15 report now being married. Does this evidence support Annie's claim, at the 0.10 level of significance, that the chances of getting married for this group are greater than 2.6%?

Source: Mikkelson, David. "Woman's Chance of Marriage Over 40." 11 Apr. 2008. https://www.snopes.com/fact-check/marry-go-round/ (25 Mar. 2019).

9. Postpartum depression and anxiety (PPD), is a common medical condition affecting mothers and their families after the birth of a baby. In 2008, the CDC estimated that 15% of women who have recently given birth suffer from PPD. However, this research only reflected self-reported cases. Therefore, one group dedicated to helping women and their families with PPD believes that the true percentage of women who suffer from PPD is much higher. The group conducts a simple random sample of 85 women who had given birth in the last year and discovers that 19 of them report having PPD. Based on this evidence, can the group claim that the true percentage of women who have PPD is greater than 15%? Use a 0.05 level of significance.

Source: Stone, Katherine. Postpartum Progress. "How Many Women Get Postpartum Depression? The Statistics on PPD." https://postpartumprogress.com/how-many-women-get-postpartum-depression-the-statistics-on-ppd (25 Mar. 2019).

10. A direct mail appeal for contributions from a university's alumni and supporters is considered to be too costly if less than 15% of the alumni and supporters provide monetary contributions. To determine if a direct mail appeal is cost effective, the fundraising director sends the direct mail brochures to a simple random sample of 250 people on the alumni and supporters mailing lists. They receive monetary contributions from 36 people. Does this evidence demonstrate that the direct mail campaign is not cost effective? Use a 0.05 level of significance.



Since $\chi^2 = 24.96$, which is in the rejection region, we reject the null hypothesis.

Interpretation: Thus, there is sufficient evidence at the 0.10 level of significance to support the farmer's claim that the variance of the heights of the two-year-old trees is not 16.

10.5 Section Exercises

Rejection Regions for Hypothesis Tests for Population Variances and Population Standard Deviations

State the critical value(s) of the test statistic, and determine the rejection region for a hypothesis test for the population variance or population standard deviation using the given information. Assume that the population distribution is approximately normal.

1. $\alpha = 0.05$, $df = 26$, left-tailed test

2. $\alpha = 0.10$, $df = 15$, left-tailed test

3. $\alpha = 0.05$, $n = 18$, $H_a: \sigma > 0.04$

4. $c = 0.90$, $n = 31$, $H_a: \sigma^2 > 0.65$

5. $c = 0.95$, $n = 25$, $H_a: \sigma^2 \neq 0.80$

6. $c = 0.99$, $n = 30$, $H_0: \sigma \neq 0.009$

Hypothesis Tests for Population Variances and Population Standard Deviations

Perform each hypothesis test. For each exercise, complete the following steps. Assume that each population is normally distributed.

a. State the null and alternative hypotheses.
b. Determine which distribution to use for the test statistic, and state the level of significance.
c. Calculate the test statistic.
d. Draw a conclusion and interpret the decision.

7. A dairy supplier fills hundreds of cartons with milk each day. They are contracted to fill each carton with exactly one gallon of milk. Because of the moving parts on the machine that fills the cartons, the amount of milk dispensed begins to vary slightly over time. When this happens, the machine must be serviced to realign it correctly. When testing the accuracy of the machine, the amount of milk dispensed into each carton sampled is measured in milliliters (mL). A variance of 4 is acceptable. Servicing must occur when the variance in the amounts of milk in the cartons is more than 4 with a level of significance of 0.01. Twenty-five cartons are randomly chosen to be tested and the amounts of milk in the cartons are found to have a standard deviation of 2.5 mL. Perform a hypothesis test to determine if the machine needs servicing.

8. A potato chip manufacturer produces bags of potato chips that are supposed to have a net weight of 326 grams. Because the chips vary in size, it is difficult to fill the bags to the exact weight desired. However, the bags pass inspection so long as the standard deviation of their weights is 3 grams. A quality control inspector wished to test the claim that one batch of bags has a standard deviation of more than 3 grams, and thus does not pass inspection. If a sample of 25 bags of potato chips is taken and the standard deviation is found to be 3.4 grams, does this evidence, at the 0.05 level of significance, support the claim that the bags should fail inspection?

9. At an archery range, instructors can determine if an archer is consistently missing the target because of the sight or because of the archer's ability. If a sight is off, the variance of the distances between the shots and the center of the shot pattern will be small (even if the shots are not in the center of the target). A student claims that it is the sight that is off, not his aim, and wants the instructor to confirm his claim. If a skilled archer shoots an arrow at a target multiple times, the distances between the shots and the center of the shot pattern, measured in centimeters (cm), will have a variance of 0.33. After the student shoots 20 shots at the target, the instructor calculates that the distances between his shots and the center of the shot pattern, measured in cm, have a variance of 0.17. Does this evidence support the student's claim that the sight is off? Use a 0.10 level of significance.

10. To ensure that there is not a large disparity in the quality of education at various schools in a school district, the school board wants to make sure that the variance of the mean standardized-test scores for all students at each school in the district is less than 0.05. To test this claim, they looked at the mean student scores for the standardized test from a random sample of 18 schools in the district. The results from the survey found that the overall mean was a score of 192.560 with a standard deviation of 0.162. With $\alpha = 0.10$, perform a hypothesis test to determine if the variance is less than 0.05.

11. The manufacturer of a popular antibiotic must ensure that each 5-mL dose contains 250 milligrams (mg) of the active ingredient. It is also essential that the variance of the amounts of active ingredient per dose be less than 0.1. For testing purposes, a random sample of 100 doses is taken, and the standard deviation of the amounts of active ingredient per dose in the sample is found to be 0.3 mg. Does the evidence support the claim that the variance is within the necessary bounds, at the 0.01 level of significance?

12. Claudia's bakery business has taken off. Her goods are so popular that she has decided to invest in kitchen equipment that will do some of the work for her. One machine, in particular, prepares bread dough, and it is set to measure 8 grams of yeast per loaf. Because baking requires precise measurements, she wants to test the new machine to make sure that the variance in the amounts of yeast per loaf is less than 0.30, at the 0.05 level of significance. A sample of the amounts of yeast added to the dough for 8 loaves has a variance of 0.25. Does this evidence support the claim that the new machine produces dough with a variance in the amounts of yeast per loaf of less than 0.30?

13. The temperatures in chicken incubators on a chicken farm, measured in degrees Fahrenheit (°F), are generally believed to have a variance of 0.50. The manager of the chicken farm claims that the variance has changed. A random test of 25 incubators finds that their temperatures have a standard deviation of 0.55 °F. At the 0.05 level of significance, does this evidence support the manager's claim that the variance is not 0.50?

14. A grocery store needs the refrigeration section to have its coolers stay at the same temperature on a daily basis with little variance to help ensure quality. Daily temperatures are measured in degrees Fahrenheit (°F), and the manager of the store assumes that the variance in the daily temperatures is 3.8. The assistant manager claims that the variance is not 3.8 and decides to test the claim using a hypothesis test. For a random sample of 30 days, the assistant manager finds that the standard deviation in the daily temperatures for one cooler is 2.9 °F. At the 0.01 level of significance, does this evidence support the claim that the variance in the daily temperatures for that cooler is not 3.8?

15. A health club needs to ensure that the temperature in its heated pool stays constant throughout the winter months. Otherwise, it needs to invest in a new heater for the pool. It is assumed that the daily water temperatures, measured in degrees Fahrenheit (°F), have a variance of 2.25, which is considered to be within normal limits for a properly operating heater. The pool manager needs to determine if the variance is still 2.25, so a hypothesis test is performed to test the claim that the variance in the temperature is not 2.25. After testing the pool water for a random sample of 15 winter days, the pool manager finds a mean daily temperature of 78.60 °F with a variance of 3.81. At the 0.10 level of significance, does this evidence support the manager's claim that the variance in the water temperature is no longer 2.25?

16. Many properties of the fabric used in different textiles, such as sheets and clothing, are influenced by whether the diameter of yarn used in the creation of the textile is consistent. Therefore, various methods can be used to measure the diameter of yarn at specified intervals, such as every 2 mm, to determine the consistency of the diameter. These measurements are normally distributed. Suppose that one textile manufacturer will not use any yarn in which the standard deviation of the diameters is greater than 0.02 mm. To ensure that the yarn is usable, the diameter of a length of yarn is measured at 100 random intervals and the standard deviation of those measurements is found to be 0.021402 mm. Does this evidence provide support for the manufacturer to use this batch of yarn? Use a 0.05 level of significance.

Source: Ibrahim, S. et.al. "Characterization of Yarn Diameter Measured on Different." 4 July 2012. http://textileconference.rmutp.ac.th/wp-content/uploads/2012/10/005-Characterization-of-Yarn-Diameter-Measured-on-different.pdf (27 Mar. 2019).

Step 4: Draw a conclusion and interpret the decision.

Method 1: Rejection Regions

The number of degrees of freedom for the chi-square distribution for this test is $df = 5 - 1 = 4$, and $\alpha = 0.05$. Using the table, we find that the critical value is $\chi^2_{0.05} = 9.488$. Comparing the test statistic to the critical value, we have $2.950 < 9.488$, so $\chi^2 < \chi^2_{0.050}$, and thus we must fail to reject the null hypothesis.

Method 2: *p*-Values

We see that $\chi^2 \approx 2.950$ and the p-value ≈ 0.5662 from the output screen for the TI-83/84 Plus calculator. This p-value can be compared to the level of significance, $\alpha = 0.05$, to draw a conclusion. Remember that if p-value $\leq \alpha$, then the conclusion is to reject the null hypothesis. In this case, p-value $> \alpha$, so the conclusion is to fail to reject the null hypothesis.

Interpretation: The evidence does not support the claim that the proportions of kids playing youth sports have changed. The educational campaign does not appear to have made any difference based on this evidence.

10.6 Section Exercises

Null and Alternative Hypotheses for Chi-Square Tests for Goodness of Fit

State the null and alternative hypotheses in words for each scenario.

1. A pharmacist believes that more people get their prescriptions filled on Fridays than on any other day of the week.
2. A game warden believes that the numbers of ducks, geese, and swans in the national park are not equal.
3. The hairstylists at a salon think that they are twice as busy on Thursdays and Fridays than the rest of the week.
4. A music producer believes that Southerners prefer country music over either pop music or R&B.

Test Statistics for Chi-Square Tests for Goodness of Fit

Calculate the test statistic, χ^2, for a chi-square test for goodness of fit using the given information.

5. The following data represent numbers of drive-through customers at a fast-food restaurant on a weekday.

Drive-Through Customers

	8:00–9:00 a.m.	12:00–1:00 p.m.	3:00–4:00 p.m.	5:00–6:00 p.m.
Observed Values, O_i	35	54	11	21
Expected Values, E_i	$\frac{121}{4}$	$\frac{121}{4}$	$\frac{121}{4}$	$\frac{121}{4}$

6. The following data represent numbers of four different candy bars sold at a concession stand in one day.

Candy Bar Sales

	A	B	C	D
Observed Values, O_i	11	12	5	15
Expected Values, E_i	$\frac{43}{4}$	$\frac{43}{4}$	$\frac{43}{4}$	$\frac{43}{4}$

7. The following data represent job applicants' scoring levels on a company's entrance tests.

Applicants' Scoring Levels			
	Level 1	Level 2	Level 3
Observed Values, O_i	29	201	169
Expected Values, E_i	20.5	218.8	159.7

8. The following data represent numbers of hamburgers sold at three competing restaurants between 12 p.m. and 2 p.m. one weekday.

Hamburger Sales			
	Handy Andy	Philip's Grocery	Proud Larry
Observed Values, O_i	37	26	24
Expected Values, E_i	29	29	29

9. The following data represent numbers of kids' jeans sold in various sizes at a department store in one day.

Kids' Jeans Sales				
	Size 6	Size 8	Size 10	Size 12
Observed Values, O_i	4	7	5	5
Expected Values, E_i	5.25	5.25	5.25	5.25

10. The following data represent numbers of children at the local library on a Saturday.

Children at the Library				
	9:00–10:00 a.m.	10:00–11:00 a.m.	1:00–2:00 p.m.	3:00–4:00 p.m.
Observed Values, O_i	6	10	9	8
Expected Values, E_i	8.25	8.25	8.25	8.25

Critical Values for Chi-Square Tests for Goodness of Fit

Find the critical value of χ^2 for a chi-square test for goodness of fit using the given information.

11. $\alpha = 0.10, \quad k = 47$

12. $\alpha = 0.05, \quad k = 22$

13. $\alpha = 0.005, \quad k = 13$

14. $\alpha = 0.01, \quad k = 9$

Conclusions of Chi-Square Tests for Goodness of Fit

State the critical value of χ^2 and determine the appropriate conclusion for a chi-square test for goodness of fit using the given information.

15. $\alpha = 0.005, \quad k = 15, \quad \chi^2 = 29.674$

16. $\alpha = 0.10, \quad k = 8, \quad \chi^2 = 13.219$

17. $\alpha = 0.05, \quad k = 29, \quad \chi^2 = 43.222$

18. $\alpha = 0.025, \quad k = 17, \quad \chi^2 = 27.008$

Hypothesis Tests for Goodness of Fit

Perform each test for goodness of fit. For each exercise, complete the following steps.

a. State the null and alternative hypotheses.
b. Determine which distribution to use for the test statistic, and state the level of significance.
c. Find the expected value for each possible outcome, and calculate the test statistic.
d. Draw a conclusion and interpret the decision.

19. A school principal claims that the number of students who are tardy to school does not vary from month to month. A survey over the school year produced the following results. Using a 0.05 level of significance, test a teacher's claim that the number of tardy students does vary by the month.

Tardy Students

	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May
Number	7	18	16	5	8	12	15	18	11	15

20. A service station owner believes that equal numbers of customers prefer to buy gasoline on every day of the week. A manager at the service station disagrees with the owner and claims that the number of customers who prefer to buy gasoline on each day of the week varies. Test the manager's claim using $\alpha = 0.10$. The owner surveyed 739 customers over a period of time to record each customer's preferred day of the week. Here's what he found.

Preferred Day to Buy Gasoline

	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
Number	103	103	126	103	111	96	97

21. At the emergency room at one hospital, the nurses are convinced that the number of patients that they see during the midnight shift is affected by the phases of the moon. The doctors think this is an old wives' tale. The nurses decide to test their theory at the 0.10 level of significance by recording the number of patients they see during the midnight shift for each moon phase over the course of one lunar cycle. The results are summarized in the table below.

ER Patients During the Midnight Shift

	New Moon	1st Quarter	Full Moon	3rd Quarter
Number of Patients	85	66	97	68

22. The management of the local zoo wants to know if all of their animal exhibits are equally popular. If there is significant evidence that some of the exhibits are not being visited frequently enough, then changes may need to take place within the zoo. A tally of visitors is taken for each of the following animals throughout the course of a week, and the results are contained in the following table. At $\alpha = 0.05$, determine whether there is sufficient evidence to conclude that some exhibits are less popular than others.

Animal Exhibits at the Zoo

	Elephants	Lions/Tigers	Giraffes	Zebras	Monkeys	Birds	Reptiles
Number of Visitors	157	154	168	162	185	129	133

23. The manager of the city pool has scheduled extra lifeguards to be on staff for Saturdays. However, he suspects that Fridays may be more popular than the other weekdays as well. If so, he will hire extra lifeguards for Fridays, too. To test his theory that the daily number of swimmers varies on weekdays, he records the number of swimmers each day for the first week of summer. Test the manager's theory at the 0.01 level of significance.

Swimmers at the City Pool

	Monday	Tuesday	Wednesday	Thursday	Friday
Number	46	47	43	53	54

24. A manufacturer of children's vitamins claims that its vitamins are mixed so that each batch has exactly the following percentages of each color: 20% green, 40% yellow, 10% red, and 30% orange. To test the claim that these percentages are incorrect, 100 bottles of vitamins were pulled and the colors of the vitamins were tallied. The results are listed in the following table. At $\alpha = 0.05$, determine whether there is sufficient evidence to conclude that the percentages stated by the vitamin manufacturer are incorrect.

Children's Vitamins

	Green	Yellow	Red	Orange
Number	1149	1948	552	1401

Method 2: ***p*****-Values**

The calculator reported a p-value of approximately 0.5831, so we can compare that to the level of significance, $\alpha = 0.10$. Since p-value $> \alpha$, we fail to reject the null hypothesis.

Interpretation: Thus, there is not enough evidence at this level of significance to conclude that there is an association between hair color and the combination of gender and marital status for this hairdresser's clients.

10.7 Section Exercises

Contingency Tables of Expected Values

Use each contingency table of observed values to find the contingency table of expected values.

1. The following data represent preferred cereal brands for a random sample of adults.

Observed Sample of 146 Adults

	Brand A	Brand B	Brand C	Brand D	Total
18-22 year olds	28	14	11	27	80
23-27 year olds	9	16	8	33	66
Total	37	30	19	60	146

2. The following data represent preferred writing hands for a random sample of adults.

Observed Sample of 237 Adults

	Ages 20–29	Ages 30–39	Ages 40–49	Total
Left Hand	15	42	31	88
Right Hand	35	41	50	126
Ambidextrous	10	5	8	23
Total	60	88	89	237

3. The following data represent preferences for user interfaces of proposed tablet models, as chosen by participants of four different consumer focus groups.

Observed Sample of 119 Consumers

	Group A	Group B	Group C	Group D	Total
Model 1	8	12	4	9	33
Model 2	11	15	6	3	35
Model 3	4	10	3	11	28
Model 4	6	11	4	2	23
Total	29	48	17	25	119

4. The following data represent the favorite sports of adults living in various regions of the United States for a random sample of adults.

Observed Sample of 336 Adults					
	Football	Basketball	Baseball	Soccer	Total
Northeast	27	25	23	20	95
Southeast	26	21	19	17	83
Midwest	24	25	18	11	78
West	20	22	17	21	80
Total	97	93	77	69	336

Test Statistics for Chi-Square Tests for Association

Calculate the test statistic, χ^2, for a chi-square test for association using the given contingency tables.

5. The following data represent the observed values and expected values of the eating habits of runners and swimmers for a random sample of adult athletes.

Observed Sample of 390 Adult Athletes			
	Prefer to Eat Before a Workout	Prefer to Eat After a Workout	Total
Runner	68	121	189
Swimmer	73	128	201
Total	141	249	390

Expected Values			
	Prefer to Eat Before a Workout	Prefer to Eat After a Workout	Total
Runner	68.330769	120.669231	189
Swimmer	72.669231	128.330769	201
Total	141	249	390

6. The following data represent the observed values and expected values of the song preferences for a random sample of adults.

Observed Sample of 343 Adults					
	Song 1	Song 2	Song 3	Song 4	Total
18–22 year-olds	34	3	46	20	103
23–27 year-olds	33	25	48	17	123
28–32 year-olds	28	18	17	9	72
33–37 year-olds	3	15	20	7	45
Total	98	61	131	53	343

Expected Values					
	Song 1	Song 2	Song 3	Song 4	Total
18–22 year-olds	29.428571	18.317784	39.338192	15.915452	103
23–27 year-olds	35.142857	21.874636	46.976676	19.005831	123
28–32 year-olds	20.571429	12.804665	27.498542	11.125364	72
33–37 year-olds	12.857143	8.002915	17.186589	6.953353	45
Total	98	61	131	53	343

7. The following data represent the observed values and expected values of the fat contents and brands of microwaveable meals sold at a local grocery store for a random sample of meals sold.

Observed Sample of 3283 Microwaveable Meals Sold				
	Less Than 5 g Fat	**5–10 g Fat**	**More Than 10 g Fat**	**Total**
Brand A	160	27	1380	1567
Brand B	86	9	789	884
Brand C	55	7	770	832
Total	301	43	2939	3283

Expected Values				
	Less Than 5 g Fat	**5–10 g Fat**	**More Than 10 g Fat**	**Total**
Brand A	143.669510	20.524216	1402.806275	1567
Brand B	81.049041	11.578434	791.372525	884
Brand C	76.281450	10.897350	744.821200	832
Total	301	43	2939	3283

8. The following data represent the observed values and expected values of the movie preferences for a random sample of college students.

Observed Sample of 409 College Students					
	Suspense	**Drama**	**Comedy**	**Horror**	**Total**
Freshman	24	28	37	18	107
Sophomore	19	25	35	15	94
Junior	31	33	30	12	106
Senior	26	29	34	13	102
Total	100	115	136	58	409

Expected Values					
	Suspense	**Drama**	**Comedy**	**Horror**	**Total**
Freshman	26.161369	30.085575	35.579462	15.173594	107
Sophomore	22.982885	26.430318	31.256724	13.330073	94
Junior	25.916870	29.804401	35.246944	15.031785	106
Senior	24.938875	28.679707	33.916870	14.464548	102
Total	100	115	136	58	409

Conclusions of Chi-Square Tests for Association

State the critical value of χ^2 and determine the appropriate conclusion for a chi-square test for association using the given information.

9. $\alpha = 0.10$, Number of rows = 2, Number of columns = 3, $\chi^2 = 5.13$

10. $\alpha = 0.025$, Number of rows = 5, Number of columns = 5, $\chi^2 = 31.1$

11. $\alpha = 0.005$, Number of rows = 5, Number of columns = 7, $\chi^2 = 40.8$

12. $\alpha = 0.10$, Number of rows = 7, Number of columns = 6, $\chi^2 = 40.3$

Hypothesis Tests for Association

Perform each test for association using the method of your choice or the one assigned by your instructor. For each exercise, complete the following steps.

a. State the null and alternative hypotheses.

b. Determine which distribution to use for the test statistic, and state the level of significance.

c. Find the expected value for each possible outcome, and calculate the test statistic.

d. Draw a conclusion and interpret the decision.

13. One state's Department of Education wants to know if there is a relationship between grades and the particular sport student athletes play in the state. A random sample of student athletes in the state produces the following results. Is there sufficient evidence at the 0.005 level of significance to show that there is a relationship between grades and the sport student athletes play?

Observed Sample of 440 Students

	A	B	C	D	F	Total
Swimming	12	30	44	15	9	110
Tennis	5	24	37	36	8	110
Cross Country	16	41	39	10	4	110
Basketball	10	40	35	19	6	110
Total	43	135	155	80	27	440

14. An insurance company wants to know if the color of an automobile has a relationship with the number of moving violations. The following contingency table gives the results of data collected from police reports across the nation. The columns list the numbers of reported moving violations in a year. Use a level of significance of $\alpha = 0.01$ to conduct this test.

Observed Sample of 443 Cars

	0–1	2–3	More Than 3	Total
White	76	33	13	122
Black	44	21	8	73
Red	50	46	12	108
Silver	33	27	7	67
Other	28	34	11	73
Total	231	161	51	443

15. A soft drink company is interested in knowing whether there is a relationship between cola preference and age. A random sample of 800 people is chosen for a taste test. The results of the study are found in the following table. Is there sufficient evidence at the 0.005 level of significance to lead you to believe that there is an association between cola preference and age?

Observed Sample of 800 People				
	Cola A	**Cola B**	**Cola C**	**Total**
15–29	94	102	105	301
30–44	99	97	86	282
45–59	68	73	76	217
Total	261	272	267	800

16. A marketing firm wants to know if there is a difference in the best marketing strategy for new customers compared to returning customers. To determine the relationship between marketing strategy and customer type, the following information on number of sales generated is obtained for each group. Does the evidence gathered support the claim at $\alpha = 0.01$?

Observed Sample of 926 Customers				
	Paid search	**Social Media**	**Web retargeting**	**Total**
New customer	158	165	153	476
Returning customer	149	147	154	450
Total	307	312	307	926

17. Suppose that a bookseller wants to study the relationship between book preference and residential area. A random sample of readers is chosen for the study, and each participant is asked to choose their favorite genre out of the following choices: mystery, fiction, nonfiction, and self-help. The results are detailed below. Does the evidence gathered show a relationship between book preference and residential area at $\alpha = 0.005$?

Observed Sample of 266 Readers					
	Mystery	**Fiction**	**Nonfiction**	**Self-Help**	**Total**
Rural	28	30	41	22	121
Urban	32	59	28	26	145
Total	60	89	69	48	266

18. A travel agency is interested in finding out if different age groups frequent different Spring Break destinations, to better target the appropriate audiences. A random sample of college Spring Break vacationers produces the results given in the table below. Is there enough evidence at the 0.05 level of significance to show that there is a relationship between age (by college classification) and destination?

Observed Sample of 192 College Students

	Beach	Mountains	City	Home	Total
Freshman	19	2	7	24	52
Sophomore	15	4	3	20	42
Junior	18	1	9	19	47
Senior	21	6	4	20	51
Total	73	13	23	83	192

19. A marketing firm wants to know if there is an association between a person's age and level of educational attainment and his or her favorite Super Bowl commercial. A random sample of people is asked to choose between two commercials, and the results are in the following table. Given these results, is there enough evidence at the 0.10 level of significance to conclude that an association exists?

Observed Sample of 80 Adults

	Commercial 1	Commercial 2	Total
College Degree, 18–30	14	6	20
No College Degree, 18–30	9	11	20
College Degree, 31–45	12	8	20
No College Degree, 31–45	6	14	20
Total	41	39	80

20. Pollsters want to test if an association exists between a person's profession and their political party. A random sample of 236 voters is polled, resulting in the data in the following table. Based on these results, is there enough evidence at the 0.025 level of significance to say that an association exists?

Observed Sample of 236 Voters

	Democrat	Republican	Total
Doctor	13	35	48
Lawyer	33	19	52
Teacher	23	25	48
Farmer	39	11	50
Laborer	28	10	38
Total	136	100	236

11.1 Section Exercises

Test Statistics for Hypothesis Tests for Two Population Means

Calculate the test statistic for a hypothesis test for two population means using the given information.

1. $\bar{x}_1 = 9.21,\quad \sigma_1 = 2.01,\quad n_1 = 45,\quad \bar{x}_2 = 8.76,\quad \sigma_2 = 1.77,\quad n_2 = 51,\quad H_0\colon \mu_1 - \mu_2 \geq 0$

2. $\bar{x}_1 = 72.82,\quad \sigma_1 = 7.90,\quad n_1 = 31,\quad \bar{x}_2 = 75.11,\quad \sigma_2 = 6.54,\quad n_2 = 39,\quad H_0\colon \mu_1 - \mu_2 \leq 0$

3. $\bar{x}_1 = 118.4,\quad \sigma_1 = 5.93,\quad n_1 = 64,\quad \bar{x}_2 = 104.3,\quad \sigma_2 = 5.74,\quad n_2 = 65,\quad H_0\colon \mu_1 - \mu_2 \geq 15$

4. $\bar{x}_1 = 43.1,\quad \sigma_1 = 2.33,\quad n_1 = 71,\quad \bar{x}_2 = 34.3,\quad \sigma_2 = 2.96,\quad n_2 = 70,\quad H_0\colon \mu_1 - \mu_2 = 8$

Null and Alternative Hypotheses for Hypothesis Tests for Two Population Means

State the null and alternative hypotheses for each scenario.

5. The claim is that the mean of Population 1 is more than 30 units less than the mean of Population 2.

6. Ann claims that the mean drive time for her to get home from work on a Friday (Population 1) is less than her mean drive time from work to home on a Thursday (Population 2).

7. Biontrix (Population 1) claims that their customer rebates are higher on average than that of their competitor, E.D.G. Inc. (Population 2).

8. Carly claims that 6 months ago her mean time to walk a mile (Population 1) was more than 4 minutes longer than it is currently (Population 2).

9. A newspaper claims that the mean age of its current readers (Population 1) has dropped 3.6 years over the last 10 years (Population 2). A market researcher believes that the newspaper's claim is incorrect and plans to conduct a hypothesis test to provide evidence against the newspaper's claim.

10. East Wind Hospital (Population 1) claims that their patients spend less time in the waiting room on average than Brown County Hospital (Population 2).

Hypothesis Tests for Two Population Means (σ Known)

Perform each hypothesis test using the method of your choice or the one assigned by your instructor. For each exercise, complete the following steps.

a. State the null and alternative hypotheses.
b. Determine which distribution to use for the test statistic, and state the level of significance.
c. Calculate the test statistic.
d. Draw a conclusion by comparing the *p*-value to the level of significance and interpret the decision.

11. A car company claims that its new SUV gets better gas mileage than its competitor's SUV. A random sample of 35 of its SUVs has a mean gas mileage of 12.6 miles per gallon (mpg). The population standard deviation is known to be 0.4 mpg. A random sample of 31 competitor's SUVs has a mean gas mileage of 12.4 mpg. The population standard deviation for the competitor is known to be 0.3 mpg. Test the company's claim at the 0.05 level of significance.

12. A college student is interested in investigating the claim that students who graduate with a master's degree earn higher salaries, on average, than those who finish with a bachelor's degree. She surveys, at random, 42 recent graduates who completed their master's degrees, and finds that their mean salary is $38,400 per year. The standard deviation of annual salaries for the population of recent graduates who have master's degrees is known to be $3100. She also surveys, at random, 45 recent graduates who completed their bachelor's degrees, and finds that their mean salary is $36,750 per year. The standard deviation of annual salaries for the population of recent graduates with only bachelor's degrees is known to be $3700. Test the claim at the 0.05 level of significance.

13. Fran is training for her first marathon, and she wants to know if there is a significant difference between the mean number of miles run each week by group runners and individual runners who are training for marathons. She interviews 32 randomly selected people who train in groups, and finds that they run a mean of 49.0 miles per week. Assume that the population standard deviation for group runners is known to be 4.2 miles per week. She also interviews a random sample of 30 people who train on their own and finds that they run a mean of 47.2 miles per week. Assume that the population standard deviation for people who run by themselves is 4.8 miles per week. Test the claim at the 0.05 level of significance.

14. Rob and Phil are both internal medicine residents in the Southeast, but they work at different hospitals. When they compare their schedules, Rob is convinced that, on average, residents at his hospital work for over 3 hours more per week than those at Phil's hospital. Rob asks a random sample of 30 residents at his hospital to record their hours for one week. He finds that they worked for a mean of 74.3 hours. Phil also asks 30 randomly selected residents at his hospital to record their hours for the same week. He calculates that they worked for a mean of 70.1 hours. Assume that the population standard deviation for the hospital where Rob works is known to be 2.6 hours and the population standard deviation for the hospital where Phil works is known to be 2.9 hours. Test Rob's claim at the 0.05 level of significance.

15. A weight-loss company wants to make sure that its clients lose more weight, on average, than they would without the company's help. An independent researcher collects data on the amount of weight lost in one month from 45 of the company's clients and finds a mean weight loss of 12 pounds. The population standard deviation for the company's clients is known to be 7 pounds per month. Data from 50 dieters not using the company's services reported a mean weight loss of 10 pounds in one month. The population standard deviation for dieters not using the company's services is known to be 6 pounds per month. Test the company's claim that using its services results in a greater mean weight loss at the 0.05 level of significance.

16. Two friends, Karen and Jodi, work different shifts for the same ambulance service. They wonder if the different shifts average different numbers of calls. Looking at past records, Karen determines from a random sample of 35 shifts that she had a mean of 5.2 calls per shift. She knows that the population standard deviation for her shift is 1.3 calls. Jodi calculates from a random sample of 34 shifts that her mean was 4.8 calls per shift. She knows that the population standard deviation for her shift is 1.2 calls. Test the claim that there is a difference between the mean numbers of calls for the two shifts at the 0.05 level of significance.

17. A professor believes that, for the introductory art history classes at his university, the mean test score of students in the evening classes is more than 5 points lower than the mean test score of students in the morning classes. He collects data from a random sample of 250 students in evening classes and finds that they have a mean test score of 80.2. He knows the population standard deviation for the evening classes to be 11.9 points. A random sample of 300 students from morning classes results in a mean test score of 86.8. He knows the population standard deviation for the morning classes to be 10.2 points. Test his claim with a 95% level of confidence.

18. A state board of directors for higher education is comparing the mean salaries of entry-level Ph.D. positions at the state's two major universities to make sure there is not a difference in entry-level pay. Use the information given in the following table, which was collected by the board of directors for higher education, to perform the hypothesis test at the 0.05 level of significance.

Salaries of Entry-Level Ph.D. Positions

	University A	University B
Sample Size	42	51
Mean Entry-Level Salary	\$58,500	\$60,200
Population Standard Deviation	\$3200	\$11,700

19. A parent interest group is looking at whether birth order affects scores on the ACT test. It was suggested that, on average, first-born children earn lower ACT scores than second-born children. After surveying a random sample of 100 first-born children, the parents' group found that they had a mean score of 20.9 on the ACT. A survey of 175 second-born children resulted in a mean ACT score of 21.1. Assume that the population standard deviation for first-born children is known to be 1.8 points and the population standard deviation for second-born children is known to be 2.3 points. Is there sufficient evidence at the 10% level of significance to say that the mean ACT score of first-born children is lower than the mean ACT score of second-born children?

20. Lauren and Keri live in different states and disagree about who has higher electric bills. To settle their disagreement, the girls decide to sample electric bills in their area for the month of June and perform a hypothesis test. The electric company in Lauren's state reports that a random sample of 35 monthly residential electric bills has a mean of \$104.53. For a random sample of 51 monthly residential electric bills in Keri's state, the mean is \$101.48. Assume that the population standard deviation in Lauren's state is known to be \$17.81, and the population standard deviation in Keri's state is known to be \$25.30. Is there evidence at the 0.01 level to say that the mean monthly residential electric bill is higher for Lauren's state than for Keri's state?

21. A car servicing shop wants to use the best windshield wiper blades for its customers. It has kept track of the mean numbers of sets of blades needed per year for two different brands of blades. A random sample of 35 customers using Brand A needed a mean of 1.2 sets of blades per year, and 30 randomly selected customers using Brand B needed a mean of 1.3 sets of blades per year. Assume that the population standard deviations for Brand A and Brand B are 0.3 and 0.7 sets per year, respectively. Is there sufficient evidence at the 0.15 level to say that the mean number of sets of wiper blades needed per year is lower for Brand A than for Brand B?

22. Two college friends are big sports fans. While they are watching baseball one season, they think that there is probably a difference between the mean batting averages of players in the SEC East and SEC West divisions. To test their theory, they find the mean batting average of 40 randomly selected SEC West players to be .260. They also find the mean batting average of a random sample of 50 SEC East players to be .249. Assume that the population standard deviation for the SEC West division is known to be .026 and the population standard deviation for the SEC East division is known to be .051. Is there sufficient evidence at the 0.10 level of significance to say that there is a difference between the mean batting averages for the two SEC divisions?

11.2 Section Exercises

Test Statistics and Degrees of Freedom for Hypothesis Tests for Two Population Means (σ Unknown)

Calculate the test statistic and determine the number of degrees of freedom for a hypothesis test for two population means using the given information. Assume that both population distributions are approximately normal.

1. $\bar{x}_1 = 93.0$, $s_1 = 10.4$, $n_1 = 21$, $\bar{x}_2 = 89.2$, $s_2 = 9.5$, $n_2 = 18$, $H_0: \mu_1 - \mu_2 = 0$
Assume that the population variances are not the same.

2. $\bar{x}_1 = 3.4$, $s_1 = 0.3$, $n_1 = 5$, $\bar{x}_2 = 3.7$, $s_2 = 0.5$, $n_2 = 7$, $H_0: \mu_1 - \mu_2 \le 0$
Assume that the population variances are not the same.

3. $\bar{x}_1 = 33.5$, $s_1 = 2.1$, $n_1 = 14$, $\bar{x}_2 = 31.1$, $s_2 = 2.8$, $n_2 = 11$, $H_a: \mu_1 - \mu_2 > 0$
Assume that the population variances are equal.

4. $\bar{x}_1 = 24.1$, $s_1 = 1.3$, $n_1 = 19$, $\bar{x}_2 = 23.0$, $s_2 = 1.1$, $n_2 = 22$, $H_0: \mu_1 - \mu_2 = 0$
Assume that the population variances are equal.

Note: *The* `2-SampTTest` *was used to obtain the p-value answers for all exercises.*

Hypothesis Tests for Two Population Means (σ Unknown)

Perform each hypothesis test using the method of your choice or the one assigned by your instructor. For each exercise, complete the following steps. Assume that both population distributions are approximately normal in each scenario.

a. State the null and alternative hypotheses.
b. Determine which distribution to use for the test statistic, and state the level of significance.
c. Calculate the test statistic.
d. Draw a conclusion and interpret the decision.

5. A manufacturer fills soda bottles. Periodically the company tests to see if there is a difference between the mean amounts of soda put in bottles of regular cola and diet cola. A random sample of 14 bottles of regular cola has a mean of 501.6 mL of soda with a standard deviation of 3.9 mL. A random sample of 16 bottles of diet cola has a mean of 498.9 mL of soda with a standard deviation of 5.3 mL. Test the claim that there is a difference between the mean fill levels for the two types of soda using a 0.01 level of significance. Assume that the population variances are not equal since different machines are used to fill bottles of regular cola and diet cola.

6. A professor is concerned that the two sections of college algebra that he teaches are not performing at the same level. To test his claim, he looks at the mean exam score for a random sample of students from each of his classes. In Class 1, the mean exam score for 12 students is 78.7 with a standard deviation of 6.5. In Class 2, the mean exam score for 15 students is 81.1 with a standard deviation of 7.4. Test the professor's claim at the 0.05 level of significance. Assume that the population variances are equal.

7. While shopping for a cookout, Ian notices that a particular brand of charcoal briquettes claims to burn longer because the briquettes are 60% thicker than the competitor's briquettes. Feeling tired of being taken for a shopper who just believes what the manufacturer wants him to believe, Ian decides to test the manufacturer's claim that the charcoal briquettes from Brand A are thicker than those from Brand B. He buys a bag of each kind of charcoal, randomly selects a few briquettes from each brand, and measures the thicknesses of the briquettes in his samples. His findings are given in the following table. Test the manufacturer's claim that the charcoal briquettes from Brand A are thicker than those from Brand B at the 0.01 level of significance. Assume that the population variances are different.

Thickness of Charcoal Briquettes (in cm)

	Brand A	Brand B
Sample Size	8	6
Mean Thickness	3.21	2.13
Standard Deviation	0.50	0.85

8. A new small business wants to know if its current radio advertising is effective. The owners decide to look at the mean number of customers who make a purchase in the store on days immediately following days when the radio ads are played as compared to the mean for those days following days when no radio advertisements are played. They found that for 11 days following no advertisements, the mean was 17.8 purchasing customers with a standard deviation of 3.5 customers. On 6 days following advertising, the mean was 22.8 purchasing customers with a standard deviation of 2.8 customers. Test the claim, at the 0.01 level, that the mean number of customers who make a purchase in the store is lower for days following no advertising compared to days following advertising. Assume that the population variances are equal.

9. Gary has discovered a new painting tool to help him in his work. If he can prove to himself that the painting tool reduces the amount of time it takes to paint a room, he has decided to invest in a tool for each of his helpers as well. From records of recent painting jobs that he completed before he got the new tool, Gary collected data for a random sample of 6 medium-sized rooms. He determined that the mean amount of time that it took him to paint each room was 4.2 hours with a standard deviation of 0.5 hours. For a random sample of 4 medium-sized rooms that he painted using the new tool, he found that it took him a mean of 3.9 hours to paint each room with a standard deviation of 0.7 hours. At the 0.10 level, can Gary conclude that his mean time for painting a medium-sized room without using the tool was greater than his mean time when using the tool? Assume that the population variances are equal.

10. A supermarket chain is convinced that customers spend more money at the grocery store when the store plays music with a slow tempo over the loud speaker rather than music with a faster tempo. They select 2 stores to test their theory. The first store plays music with a slow tempo for the noon hour and randomly chooses 4 of the receipts with a mean of \$122.56 and a standard deviation of \$13.12. The second store plays music with a fast tempo for the noon hour. Five randomly chosen receipts from that time have a mean of \$108.31 with a standard deviation of \$17.06. Assume that the population variances are different. At the 0.10 level, can the supermarket chain say that music with a slower tempo makes customers buy more groceries?

11. While running some quality control tests, a manager at a factory that makes potato chips noticed a difference in the mean bag weights for chips coming from two different production lines. To see if the bags in Line A did actually have a mean weight lower than those in Line B, he randomly selected some of the bags from each line. His results are given in the following table. If he assumes that the population variances for the two lines are different, can he conclude at the 0.10 level that the mean weight of bags from Line A is lower than the mean weight of bags from Line B?

Weights of Bags of Chips (in g)

	Line A	Line B
Sample Size	20	15
Mean Weight	309.63	311.87
Standard Deviation	15.91	13.21

12. A physician wants to test the claim that the average adult height of premature baby boys is different from that of full-term baby boys. To do this, he finds a random sample of 18 men who were born prematurely and calculates that their mean height is 68.1 inches with a standard deviation of 2.3 inches. He also finds a random sample of 20 men who were carried full term and calculates their mean height to be 68.9 inches with a standard deviation of 2.0 inches. Assume that the population standard deviations are unequal, and test the claim at the 0.05 level of significance.

13. A pharmaceutical company needs to know if its new cholesterol drug, Praxor, is effective at lowering cholesterol levels. It believes that people who take Praxor will average a greater decrease in cholesterol level than people taking a placebo. After the experiment is complete, the researchers find that the 25 participants in the treatment group lowered their cholesterol levels by a mean of 23.5 points with a standard deviation of 5.8 points. The 25 participants in the control group lowered their cholesterol levels by a mean of 18.9 points with a standard deviation of 4.1 points. Assume that the population variances are not equal, and test the company's claim at the 0.01 level.

14. A speech pathology professor believes from experience that, on average, boys begin talking at a later age than girls. To test her theory, she gathers information from the parents of random samples of 12 boys and 14 girls. The boys began talking at a mean of 1.33 years of age with a standard deviation of 0.15 years. The girls began talking at a mean of 1.23 years of age with a standard deviation of 0.12 years. Assume that the population standard deviations are equal, and test the professor's claim at the 0.05 level of significance.

15. Insurance Company A claims that its customers pay less for car insurance, on average, than customers of its competitor, Company B. You wonder if this is true, so you decide to compare the average monthly costs of similar insurance policies from the two companies. For a random sample of 9 people who buy insurance from Company A, the mean cost is \$152 per month with a standard deviation of \$17. For 11 randomly selected customers of Company B, you find that they pay a mean of \$155 per month with a standard deviation of \$14. Assume that the population variances are equal, and test Company A's claim at the 0.01 level of significance.

16. A women's group believes that women are offered lower starting salaries than men applying for similar jobs. To test this claim, the group sends 10 women and 10 men to interviews for similar positions at various companies. The women were offered a mean starting salary of \$29,500 with a standard deviation of \$1100. The men were offered a mean starting salary of \$30,500 with a standard deviation of \$950. Assume that the population standard deviations are different, and test the group's claim at the 0.05 level of significance.

17. A psychologist wants to test the claim that the mean lengths of time spent to complete a piece of art are different for trauma patients and non-trauma patients. He randomly chooses 12 trauma patients and 11 non-trauma patients and asks them to estimate the amount of time it takes for them to complete a piece of artwork. According to the survey, the trauma patients spent a mean of 4.6 hours to complete the artwork with a standard deviation of 1.3 hours. The non-trauma patients spent a mean of 5.4 hours to complete a piece of artwork with a standard deviation of 1.1 hours. Assume that the population variances are equal, and test the claim at the 0.05 level of significance.

18. The results of a state-wide 9^{th} grade literature state assessment for random samples of students in two neighboring counties are shown in the table.

Grappling County	Hammond County
$\bar{x}_1 = 517$	$\bar{x}_2 = 495$
$s_1 = 39.7$	$s_2 = 24.8$
$n_1 = 134$	$n_2 = 40$

Can you conclude that there is a difference in the mean literature test scores for the students of the two counties? Use $\alpha = 0.01$. Assume the populations are normally distributed and the population variances are not equal.

11.3 Section Exercises

Hypothesis Tests for the Mean of the Paired Differences for Two Populations (σ Unknown, Dependent Samples)

Perform each hypothesis test using the method of your choice or the one assigned by your instructor. Assume that the population of paired differences is normally distributed. For each exercise, complete the following steps.

a. State the null and alternative hypotheses.
b. Determine which distribution to use for the test statistic, and state the level of significance.
c. Calculate the necessary sample statistics then compute the test statistic.
d. Draw a conclusion and interpret the decision.

1. An anger-management course claims that, after completing its seminar, participants will lose their tempers less often. Always a skeptic, you decide to test this claim. A random sample of 12 seminar participants is chosen, and these participants are asked to record the number of times that they lost their tempers in the two weeks prior to the course. After the course is over, the same participants are asked to record the number of times that they lost their tempers in the next two weeks. The following table lists the results of the survey. Using these data, test the claim at the 0.05 level of significance.

Number of Times Temper Was Lost during a Two-Week Period

Before	8	10	6	7	4	11	12	5	6	3	6	4
After	6	5	6	6	5	9	4	5	4	4	5	4

2. The manufacturer of a new eye cream claims that the cream reduces the appearance of fine lines and wrinkles after just 14 days of application. To test the claim, 10 women are randomly selected to participate in a study. The number of fine lines and wrinkles that are visible around each participant's eyes is recorded before and after the 14 days of treatment. The following table displays the results. Test the claim at the 0.01 level of significance.

Numbers of Fine Lines and Wrinkles

Before	8	14	13	15	10	16	9	10	11	10
After	6	14	11	14	10	14	9	9	11	8

3. An SAT prep course claims to increase student scores by more than 60 points, on average. To test this claim, 9 students who have previously taken the SAT are randomly chosen to take the prep course. Their SAT scores before and after completing the prep course are listed in the following table. Test the claim at the 0.01 level of significance.

SAT Scores									
Before Prep Course	1010	980	1170	1200	1040	1280	1450	1470	1500
After Prep Course	1100	1260	1190	1280	1170	1370	1440	1500	1520

4. Sarah believes that completely cutting caffeine out of a person's diet will allow him or her more restful sleep at night. In fact, she believes that, on average, adults will have more than two additional nights of restful sleep in a four-week period after removing caffeine from their diets. She randomly selects 8 adults to help her test this theory. Each person is asked to consume two caffeinated beverages per day for 28 days and then cut back to no caffeinated beverages for the following 28 days. During each period, the participants record the numbers of nights of restful sleep that they had. The following table gives the results of the study. Test Sarah's claim at the 0.05 level of significance.

Numbers of Nights of Restful Sleep in a Four-Week Period								
With Caffeine	21	20	26	20	24	21	18	15
Without Caffeine	22	24	27	23	21	26	22	23

5. To test the claim that children with the same parents do not have the same weights, some students in an upper-level statistics class surveyed nine families with at least two boys in the family having the same parents. The boys' weights, taken at the same age, are listed in the following table. Is there enough evidence, at the 0.10 level of significance, to support the claim that, on average, two boys with the same parents do not have equal weights?

Boys' Weights (in Pounds)									
Boy A	94	138	171	131	159	110	148	90	170
Boy B	93	140	176	130	173	112	145	90	178

6. One of the top golf camps in the country advertises that a week with its coaches will lower your average golf score by two points. One disgruntled customer claims that the camp does not live up to its advertised claim. To test the customer's claim, eight randomly selected people attending the camp agreed to participate in a study, and their pre-camp and post-camp average scores are listed in the following table. (These are their average scores on a par-72 golf course.) Test the customer's claim at the 0.10 level of significance.

Average Golf Scores								
Before Camp	75	74	76	75	76	76	74	77
After Camp	73	72	73	74	74	72	74	75

7. A violin teacher wants to convince parents of 5-year-olds that, after a year of lessons, the students will increase their stamina for standing in the correct position by more than 30 minutes, on average. She recorded the length of time that each student could hold the correct position while practicing at the beginning of the year, and then again at the end of the year. Her results are listed in the following table. Is there evidence, at the 0.10 level of significance, that the 5-year-olds' stamina times increase by a mean of more than 30 minutes after one year of violin lessons?

Stamina for Correct Position (in Minutes)

Start of Lessons	5	7	5	4	10	6	8	5
After a Year of Lessons	36	39	40	35	41	37	38	37

8. A psychology graduate student wants to test the claim that there is a significant difference between the IQs of spouses. To test this claim, she measures the IQs of 9 married couples using a standard IQ test. The results of the IQ tests are listed in the following table. Using a 0.05 level of significance, test the claim that there is a significant difference between the IQs.

IQs of Married Couples

Spouse 1	100	110	132	120	90	115	124	121	107
Spouse 2	98	111	134	119	95	116	122	118	110

9. An economist studying inflation in electricity prices in 2018 and 2019 believes that the average price of electricity, even after adjusting for inflation, changed between these two years. To test his claim, he samples 9 different counties and records the average price of electricity in each county from each year. He then adjusts the prices for inflation. His results are given in the following table. Test the economist's claim at the 0.01 level of significance.

Average Residential Retail Prices of Electricity ($/kWh)

2018	2019
20.37	19.30
15.44	15.48
17.45	15.29
16.47	16.06
15.73	15.99
14.82	15.36
16.63	16.55
17.66	18.63
11.67	12.79

10. A local school district is looking at adopting a new textbook that, according to the publishers, will increase standardized test scores of second graders by more than 10 points, on average. Never willing to believe a publisher's claim without evidence to support it, the school board decides to test the claim. The school board chooses two second-grade classes for the study. One class was assigned the new textbook and the other class used the traditional textbook. Eight children from each class were then paired based on demographics and ability levels. The following table lists the standardized test scores for the pairs. Do the data support the company's claim at the 0.05 level of significance?

Standardized Test Scores of Second Graders

New Book	78	82	90	67	79	83	89	93
Old Book	67	70	79	54	68	71	78	82

11.4 Section Exercises

Test Statistics for Hypothesis Tests for Two Population Proportions

Calculate the test statistic for a hypothesis test for two population proportions using the given information. Assume that the presumed difference between the population proportions is 0.

1. $x_1 = 15, \quad n_1 = 28, \quad x_2 = 21, \quad n_2 = 33,$
2. $x_1 = 139, \quad n_1 = 180, \quad x_2 = 178, \quad n_2 = 208,$
3. $x_1 = 18, \quad n_1 = 67, \quad x_2 = 20, \quad n_2 = 59,$
4. $x_1 = 62, \quad n_1 = 101, \quad x_2 = 71, \quad n_2 = 130,$

Hypothesis Tests for Two Population Proportions

Perform each hypothesis test using the method of your choice or the one assigned by your instructor. For each exercise, complete the following steps.

a. State the null and alternative hypotheses.
b. Determine which distribution to use for the test statistic, and state the level of significance.
c. Calculate the test statistic.
d. Draw a conclusion by comparing the *p*-value to the level of significance and interpret the decision.

5. A newspaper story claims that more houses are purchased by singles now than singles 5 years ago. To test this claim, two studies were conducted on the buying habits of singles over the past 5 years. In the first study, 500 house purchases in the current year were randomly selected and 100 of those were made by singles. In the second study, again 500 house purchases were randomly selected from 5 years ago and 72 of those were made by single people. Test the newspaper's claim using a 0.01 level of significance. Is there sufficient evidence to support the newspaper's claim?
6. Researchers claim that the birth rate in Bonn, Germany is higher than the national average. A random sample of 1200 Bonn residents produced 12 births, whereas a random sample of 1000 people from all over Germany had 8 births during the same year. Test the researchers' claim using a 0.05 level of significance.
7. University officials hope that changes they have made have improved the retention rate. Last year, a sample of 1926 freshmen showed that 1400 returned as sophomores. This year, 1508 of 2011 freshmen sampled returned as sophomores. Determine if there is sufficient evidence at the 0.05 level to say that the retention rate has improved.
8. Adrian hopes that his new training methods have improved his batting average. Before starting his new regimen, he was batting .220 in a random sample of 50 at-bats. For a random sample of 24 at-bats since changing his training techniques, his batting average is .375. Determine if there is sufficient evidence to say that his batting average has improved at the 0.10 level of significance.

9. There is an old wives' tale that women who eat chocolate during pregnancy are more likely to have happy babies. A pregnancy magazine wants to test this claim, and it gathers 100 randomly selected pregnant women for its study. Half of the women sampled agree to eat chocolate at least once a day, while the other half agree to forego chocolate for the duration of their pregnancies. A year later, the ladies complete a survey regarding the overall happiness of their babies. The results are given in the following table. At the 0.01 level of significance, test the claim of the old wives' tale.

Numbers of Babies

	Happy Babies	Unhappy Babies
With Chocolate	24	26
Without Chocolate	22	28

10. A new government program claims to lower high school dropout rates. In one school district, from a sample of 3400 students, the previous dropout rate was 4.5%. Two years after the start of the new program, the dropout rate has been lowered to 3.8% out of 1450 students sampled. Is there enough evidence to say that the government program is effective at lowering the high school dropout rate? Test the government's claim at the 0.10 level of significance.

11. To test the fairness of law enforcement in its area, a local citizens' group wants to know whether women and men are unequally likely to get speeding tickets. Two hundred randomly selected adults were phoned and asked whether or not they had been cited for speeding in the last year. Using the results in the following table and a 0.05 level of significance, test the claim of the citizens' group.

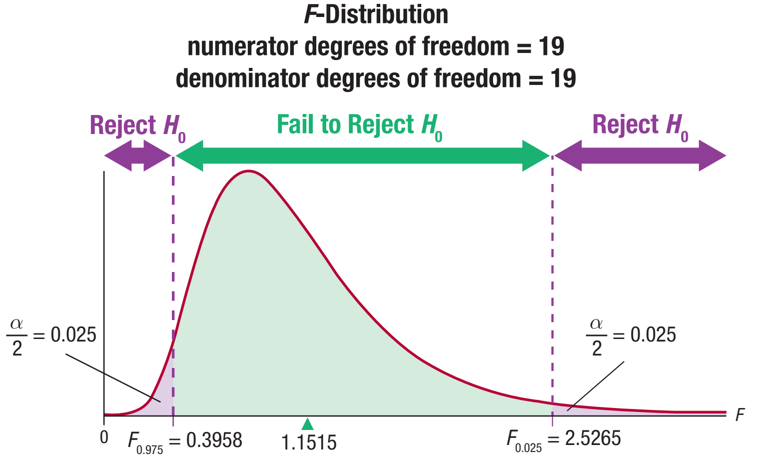
Speeding Tickets

	Ticketed	Not Ticketed
Men	11	75
Women	12	102

12. A study was performed to determine the percentage of people who wear life vests while out on the water. A researcher believed that the percentage was different for those who rode jet skis compared to those who were in boats. Out of 200 randomly selected people who rode a jet ski, 91% wore life vests. Out of 250 randomly selected boaters, 83.2% wore life vests. Using a 0.10 level of significance, test the claim that the proportion of people who wear life vests while riding a jet ski is not the same as the proportion of people who wear life vests while riding in a boat.

$F_{(1-\alpha/2)} = F_{0.975} = 0.3958$. Using the section of Table H for an area of $\alpha/2 = 0.025$ gives us $F_{\alpha/2} = F_{0.025} = 2.5265$. Therefore, we will reject the null hypothesis if either $F < 0.3958$ or $F > 2.5265$.

Because $F \approx 1.1515$ is not in the rejection region, we must fail to reject the null hypothesis.



Method 2: *p*-Values

P-values for the *F*-test must be found using technology. We will show the calculator method here. Other methods of technology can be found at stat.hawkeslearning.com.

TI-83/84 Plus: The *p*-value shown in the output screenshot in Step 3 is approximately 0.7616. Because this is greater than $\alpha = 0.05$, we fail to reject the null hypothesis.

Interpretation: Failing to reject the null hypothesis means that, at the 0.05 level of significance, there is not sufficient evidence to say that the scores on the two versions of the test have different variances.

11.5 Section Exercises

Null and Alternative Hypotheses for Hypothesis Tests for Two Population Variances

State the null and alternative hypotheses for each scenario.

1. A medical researcher believes that the variance in the lung capacities of smokers is less than that of nonsmokers. Let σ_1^2 represent the population variance for smokers.
2. A professor believes that the variance of SAT scores of honor students is less than that of all students who take the SAT. Let σ_1^2 represent the population variance for honor students.
3. A quality control inspector believes that the variance in the diameters of soda cans produced by Machine 1 is greater than the variance in the diameters of soda cans produced by Machine 2. Let σ_1^2 represent the population variance for Machine 1.
4. A psychologist is interested if the variance of the fastest speed driven by current college students, σ_1^2, is greater than the variance of the fastest speed driven by those with a college degree.
5. A golf pro believes that the variances of his driving distances are not the same when he uses different brands of golf balls. He is especially interested in comparing Titleist golf balls to a generic store brand. Let σ_1^2 represent the population variance for Titleist golf balls.

6. A paint technician believes that the variance of the thickness of the special coating in one tank is not the same as the variance of the thickness of the coating in another tank. Let σ_1^2 represent the population variance for Tank 1.

Test Statistics for Hypothesis Tests for Two Population Variances

Calculate the test statistic for a hypothesis test for two population variances using the given information. Assume that both population distributions are approximately normal.

7. $n_1 = 15, \quad s_1^2 = 3.007, \quad n_2 = 16, \quad s_2^2 = 2.897$
8. $n_1 = 4, \quad s_1^2 = 0.961, \quad n_2 = 6, \quad s_2^2 = 0.899$
9. $n_1 = 31, \quad s_1^2 = 46.821, \quad n_2 = 28, \quad s_2^2 = 57.024$
10. $n_1 = 23, \quad s_1^2 = 35{,}679, \quad n_2 = 24, \quad s_2^2 = 39{,}018$

Rejection Regions for Hypothesis Tests for Two Population Variances

State the critical value(s) of the test statistic, and determine the rejection region for the hypothesis test for the two population variances using the given information. Then give the appropriate conclusion for the hypothesis test. Assume that both population distributions are approximately normal.

11. $n_1 = 19, \quad s_1^2 = 0.3891, \quad n_2 = 24, \quad s_2^2 = 0.9579, \quad H_a\colon \sigma_1^2 < \sigma_2^2, \quad \alpha = 0.05$
12. $n_1 = 14, \quad s_1^2 = 3.152, \quad n_2 = 11, \quad s_2^2 = 9.300, \quad H_a\colon \sigma_1^2 < \sigma_2^2, \quad \alpha = 0.05$
13. $n_1 = 11, \quad s_1^2 = 31{,}207, \quad n_2 = 11, \quad s_2^2 = 38{,}916, \quad H_a\colon \sigma_1^2 < \sigma_2^2, \quad \alpha = 0.01$
14. $n_1 = 11, \quad s_1^2 = 3.007, \quad n_2 = 25, \quad s_2^2 = 2.897, \quad H_a\colon \sigma_1^2 > \sigma_2^2, \quad \alpha = 0.05$
15. $n_1 = 20, \quad s_1^2 = 10.453, \quad n_2 = 23, \quad s_2^2 = 3.199, \quad H_a\colon \sigma_1^2 > \sigma_2^2, \quad \alpha = 0.10$
16. $n_1 = 12, \quad s_1^2 = 1893, \quad n_2 = 26, \quad s_2^2 = 1066, \quad H_a\colon \sigma_1^2 > \sigma_2^2, \quad \alpha = 0.01$
17. $n_1 = 16, \quad s_1^2 = 18.01, \quad n_2 = 21, \quad s_2^2 = 17.07, \quad H_a\colon \sigma_1^2 \neq \sigma_2^2, \quad \alpha = 0.05$
18. $n_1 = 20, \quad s_1^2 = 27.08, \quad n_2 = 29, \quad s_2^2 = 11.77, \quad H_a\colon \sigma_1^2 \neq \sigma_2^2, \quad \alpha = 0.05$
19. $n_1 = 20, \quad s_1^2 = 8.12, \quad n_2 = 18, \quad s_2^2 = 16.78, \quad H_a\colon \sigma_1^2 \neq \sigma_2^2, \quad \alpha = 0.10$
20. $n_1 = 11, \quad s_1^2 = 12{,}047, \quad n_2 = 12, \quad s_2^2 = 18{,}019, \quad H_a\colon \sigma_1^2 \neq \sigma_2^2, \quad \alpha = 0.01$

Hypothesis Tests for Two Population Variances

Perform each hypothesis test using the method of your choice or the one assigned by your instructor. For each exercise, complete the following steps. Assume that both population distributions are approximately normal in each scenario.

a. State the null and alternative hypotheses.
b. Determine which distribution to use for the test statistic, and state the level of significance.
c. Calculate the test statistic.
d. Draw a conclusion and interpret the decision.

21. A golf pro believes that the variances of his driving distances are different for different brands of golf balls. In particular, he believes that his driving distances, measured in yards, have a smaller variance when he uses Titleist golf balls than when he uses a generic store brand. He hits 10 Titleist golf balls and records a sample variance of 201.65. He hits 10 generic golf balls and records a sample variance of 364.57. Test the golf pro's claim using a 0.05 level of significance. Does the evidence support the golf pro's claim?

22. A quality control inspector believes that the variance in the diameters of soda cans, measured in millimeters, is greater for soda cans produced by Machine A than for soda cans produced by Machine B. The sample variance of a random sample of 15 soda cans from Machine A is 2.788. The sample variance for a random sample of 17 soda cans from Machine B is 1.982. Test the inspector's claim using a 0.10 level of significance. Does the evidence support the inspector's claim?

23. A medical researcher believes that the variance of total cholesterol levels in men is greater than the variance of total cholesterol levels in women. The sample variance for a random sample of 8 men's cholesterol levels, measured in mg/dL, is 277. The sample variance for a random sample of 7 women is 89. Test the researcher's claim using a 0.10 level of significance. Does the evidence support the researcher's belief?

24. A basketball coach believes that the variance of the heights of adult male basketball players is different from the variance of heights for the general population of men. The sample variance of heights, measured in inches, for a random sample of 12 basketball players is 24.76. The sample variance for a random sample of 13 other men is 25.87. Test the coach's claim using a 0.01 level of significance. Does the evidence support the coach's claim?

25. One study claims that the variance in the resting heart rates of smokers is different than the variance in the resting heart rates of nonsmokers. A medical student decides to test this claim. The sample variance of resting heart rates, measured in beats per minute, for a random sample of 5 smokers is 545.1. The sample variance for a random sample of 5 nonsmokers is 103.7. Test the study's claim using a 0.01 level of significance. Does the evidence support the study's claim?

26. A professor believes that the variance of ACT composite scores of honor students is less than that of all students who take the ACT. The sample variance of the ACT composite scores for a random sample of 18 honors students is 12.1. The sample variance of the ACT composite scores for a random sample of 20 other students is 28.9. Test the professor's claim using a 0.05 level of significance. Does the evidence support the professor's claim?

Method 2: *p*-Values

The output screen in Step 3 displays the p-value = 0.074407. Since the p-value is less than our level of significance, $\alpha = 0.10$, we reject the null hypothesis.

Interpretation: Thus, the researchers can conclude that there is sufficient evidence, at the 0.10 level of significance, to support the claim that at least one of the population means is different. That is, the evidence suggests that the mean decrease in a woman's cholesterol level for at least one of the drugs is different from the others. However, we cannot conclude, from the one-way ANOVA test alone, how the population means differ. Looking at the descriptive statistics for the sample data, we see that Drugs 1 and 3 lowered the women's cholesterol levels by a mean of 20.1 points and Drug 2 lowered the women's cholesterol levels by a mean of 18.2 points, which suggests that the population mean for Drug 2 differs from the others, but we cannot make this conclusion from the one-way ANOVA test alone.

Note that the one-way ANOVA test only tells us whether there exists a difference between the population means. It does not tell you which mean, or means, are different from the others. If you want to determine which means are different, you would use a multiple comparisons test such a Tukey, Dunnet, or an MCB test, which are beyond the scope of this book.

11.6 Section Exercises

Critical Values for One-Way ANOVA Tests

Find the critical value for a one-way ANOVA test using the given information.

1. $\alpha = 0.05,\quad df_1 = 10,\quad df_2 = 12$

2. $\alpha = 0.01,\quad df_1 = 9,\quad df_2 = 15$

3. $\alpha = 0.10,\quad df_1 = 7,\quad df_2 = 8$

4. $\alpha = 0.05,\quad df_1 = 3,\quad df_2 = 9$

5. $\alpha = 0.10,\quad df_1 = 1,\quad df_2 = 5$

6. $\alpha = 0.01,\quad df_1 = 6,\quad df_2 = 4$

ANOVA Tables

Complete each ANOVA table.

7.	*SS*	*df*	*MS*	*F*
Treatments (T)	4	5		
Error (E)	11	6		
Total				

8.	*SS*	*df*	*MS*	*F*
Treatments (T)	49.3	4		
Error (E)		3		
Total	142.5			

9.	*SS*	*df*	*MS*	*F*
Treatments (T)	50		12.5	
Error (E)		3		
Total	74			

10.	*SS*	*df*	*MS*	*F*
Treatments (T)		9		
Error (E)	21.8		5.45	
Total	101.9			

One-Way ANOVA Tests

Perform each one-way ANOVA test. Assume that the population distributions for each test are all approximately normal with equal population variances. For each exercise, complete the following steps.

a. Find all the values for the ANOVA table.b. Draw a conclusion and interpret the decision.

c. If there is enough evidence to support the claim that at least one of the population means differs from the others, determine, if possible, which population mean(s) is (are) different.

11. A regional manager wants to know if there is a difference between the mean amounts of time that customers wait in line at the drive-through window for the three stores in her region. She randomly samples the wait times at each store. Her data are given in the following table. Use a one-way ANOVA test to determine if there is a difference between the mean wait times for the three stores, at the 0.05 level of significance.

Drive-Through Wait Times (in Minutes)

Store 1	Store 2	Store 3
2.34	2.87	1.32
1.23	1.94	1.45
1.89	2.36	1.78
2.31	1.85	2.01
3.02	1.75	2.45
1.95	2.82	1.92
2.45	3.32	1.83

12. A manager is concerned that one of his workers is producing more defective parts, on average, than the other similarly skilled workers. He records the number of defective parts made by this worker and two others with similar experience during their shifts each day for one week. The results are provided in the following table. At the 0.10 level of significance, can you conclude that there is a difference between the mean numbers of defective parts produced each day by these three workers?

Number of Defective Parts per Day

Worker A	Worker B	Worker C
3	2	1
2	2	2
2	1	1
4	0	1
2	1	2

13. The Panhellenic Council is comparing the mean GPAs of four sororities on campus. The GPAs for 10 girls in each sorority are given in the following table. Based on these data, can you conclude that there is a difference between the mean GPAs for these four sororities? Use a 0.05 level of significance.

Sority GPAs			
ΛΧΒ	ΩΣ	ΘΠ	ΔΜΟ
2.5	3.1	2.7	3.1
3.4	3.4	2.9	3.2
4.0	3.6	2.8	3.1
3.8	3.6	2.8	2.8
2.7	3.7	2.9	2.9
2.8	3.0	3.1	3.1
2.8	4.0	2.9	3.0
3.2	3.9	2.7	3.3
3.1	4.0	2.6	2.7
3.9	3.4	3.3	2.6

14. A group of paramedics does not believe that the mean numbers of calls received in one shift are the same for the morning, afternoon, and night shifts. To test this claim, they record the number of calls received during each shift for seven days. Based on this evidence, can the paramedics conclude that the mean numbers of calls are different for the three shifts? Use a 0.01 level of significance.

Number of Calls per Shift		
Morning	Afternoon	Night
2	3	5
3	4	4
2	4	5
4	5	3
3	3	2
1	2	5
3	5	6

15. A researcher believes there is a difference in the mean number of days before visible results begin to show among three types of facial creams that reduce wrinkle lines. Several consumers are randomly selected and given one of the three creams. Each participant then recorded the number of days it took to see results. The results are shown in the table. Based on these data, can you conclude that there is a difference between the mean number of days for these three creams? Use a 0.01 level of significance.

Cream #1	Cream #2	Cream #3
16	14	12
14	17	17
21	15	12
15	20	15
19		

16. A local weather team is comparing the mean amount of snowfall reported by viewers in four different regions of the city. Based on the data, can you conclude that there is a difference between the mean amount of snowfall for these four regions? Use a 0.10 level of significance.

Region 1	Region 2	Region 3	Region 4
3.26	2.34	2.80	2.34
2.22	2.38	2.87	2.43
3.26	3.31	3.03	2.38
3.39	2.39	2.49	3.31
2.84	3.40	2.84	2.39
3.01	2.70	3.23	3.40
3.13	2.34	3.49	2.71
2.84	2.11	3.59	2.34
3.00		3.81	
2.74			

Solution

Recall that the coefficient of determination tells us the amount of variation in the response variable (number of cups of hot chocolate) that is associated with the variation in the explanatory variable (temperature). Thus, the coefficient of determination for the relationship between the number of cups of hot chocolate and the temperature will tell us the proportion or percentage of the variation in cups of hot chocolate that can be associated with the variation in temperature. Also, recall that the coefficient of determination is equal to the square of the correlation coefficient. Since we know that the correlation coefficient for these data is $r = -0.65$, we can calculate the coefficient of determination as $r^2 = (-0.65)^2 = 0.4225$. Thus, approximately 42.3% of the variation in the number of cups of hot chocolate sold can be associated with the variation in the average temperature that evening.

12.1 Section Exercises

Type and Strength of a Linear Relationship

Predict the type and strength of the linear relationship between each pair of variables: weak negative, strong negative, weak positive, strong positive, or no linear relationship at all.

1. Height and shoe size
2. Cholesterol level and IQ
3. Hours of physical activity per week and body fat percentage
4. Number of fruits and vegetables eaten per day and risk of heart disease
5. Number of pets owned and household income
6. Employee satisfaction and their salaries.

Statistical Significance of Correlation Coefficients

Determine whether the correlation coefficient is statistically significant at the specified level of significance for the given sample size. Assume that a scatter plot of the data shows a linear pattern.

7. $r = 0.731, \quad \alpha = 0.01, \quad n = 11$
8. $r = 0.638, \quad \alpha = 0.05, \quad n = 11$
9. $r = -0.499, \quad \alpha = 0.01, \quad n = 26$
10. $r = -0.443, \quad \alpha = 0.01, \quad n = 30$
11. $r = -0.462, \quad \alpha = 0.05, \quad n = 18$
12. $r = 0.305, \quad \alpha = 0.05, \quad n = 45$

Scatter Plots and Correlation

Complete the following objectives for each data set.

a. Draw a scatter plot.

b. Estimate the correlation in words (that is, positive, negative, or no correlation).

c. Calculate the correlation coefficient, *r*.

d. Determine whether *r* is statistically significant at the 0.01 level of significance.

13. Consider the relationship between the number of bids an item on eBay receives and the item's selling price. The following is a sample of 10 items sold through auctions on eBay.

Numbers of Bids and Selling Prices

Number of Bids, *x*	7	18	32	11	20	1	5	4	2	1
Price (in Dollars), *y*	106.00	100.00	129.99	176.00	200.00	25.00	20.50	36.00	36.50	49.99

14. The following data represent the number of hours 10 students spent studying and their corresponding grades on their midterm exams.

Hours Spent Studying and Midterm Grades

Hours Studying, *x*	0	0	0.5	1.5	2	3	3.5	3.5	4.5	6
Midterm Grade, *y*	64	83	72	74	85	89	79	93	98	95

15. The following table gives the average number of hours 12 junior high students were left unsupervised each day and their corresponding overall grade averages.

Hours Unsupervised and Overall Grade Averages

Hours Unsupervised, *x*	0	0	0.5	1.0	1.0	1.5	2.0	3.0	3.0	4.0	4.5	5.5
Overall Grade Average, *y*	96	91	88	92	94	91	87	85	81	80	77	72

16. The following data represent the completion percentages and interception percentages of eight NFL quarterbacks.

Completion Percentages and Interception Percentages

Completion %, *x*	62.5	67.0	60.2	65.4	57.8	57.5	59.1	59.2
Interception %, *y*	1.8	1.8	2.3	4.5	4.2	2.3	1.5	3.6

17. The following table gives the weeks of gestation and corresponding birth weights (in pounds) for a sample of 14 babies.

Weeks of Gestation and Birth Weights

Weeks of Gestation, *x*	34	35	37	37	38	38	39	39	39	40	40	40	41	41
Weight (in Pounds), *y*	4.8	5.9	5.3	6.7	5.6	7.3	7.7	8.4	8.2	9.5	7.6	8.6	8.9	9.4

18. A study on bone density at a local hospital produced the following results.

Ages and Bone Density Levels

Age (in Years), *x*	34	43	49	51	65
Bone Density (in mg/cm^2), *y*	946	875	804	723	691

Correlation Coefficients and Coefficients of Determination

Complete the following objectives for each data set.

a. Calculate the correlation coefficient, r.

b. Determine if r is statistically significant at the 0.05 level of significance.

c. Calculate the coefficient of determination, r^2.

d. Interpret the meaning of r^2 for the given set of data.

19. The following table contains data from 16 former elementary statistics students, where x represents the number of absences a student had for the semester and y represents the student's final class average.

Number of Absences and Class Average

Absences, x	2	2	3	10	3	7	9	1	12	9	1	1	13	1	10	3
Class Average, y	86	83	81	53	92	71	68	79	53	78	77	85	62	97	54	79

20. The following table gives data collected on the birth weights of mothers and the birth weights of their babies. (**Hint:** To do calculations on these data, first convert the data to a single unit of measurement.)

Mothers' Birth Weights and Babies' Birth Weights

Mother, x	**lb**	6	6	7	8	7	7	7	6	9	6	9	8	6	5	6	5
	oz	6	2	9	10	6	15	7	5	3	10	3	9	1	0	1	0
Baby, y	**lb**	5	6	7	7	10	5	9	5	8	7	9	8	5	7	7	8
	oz	11	14	10	2	0	13	0	9	13	11	10	6	8	15	2	5

21. Many universities use a student's ACT score to help with admission decisions as well as placement decisions. From a sample of 20 students, the following data are obtained.

ACT Scores and College GPAs

ACT Score, x	College GPA, y
16	1.85
18	2.20
24	2.80
25	3.50
34	4.00
27	3.18
29	3.90
25	2.90
30	4.00
21	2.60
17	2.50
21	3.65
28	3.10
31	3.72
35	3.24
18	2.30
17	1.70
26	3.10
28	3.50
23	2.76

22. The following table contains information on the number of hours spent marketing per week and the number of new customers brought in that week for several multi-level marketing professionals.

Time Spent on Marketing and Number of New Customers

Time Spent on Marketing (in Hours per Week), x	Number of New Customers, y
10	3
11	3
11	4
12	5
5	3
13	4
15	6
17	5
7	3
10	5

g. The data were collected in Key West—not Destin, Florida. Therefore, it is not appropriate to use the linear regression equation to make predictions regarding the precipitation in Destin.

h. The slope of the regression line indicates the amount of increase in the y-variable for each increase of one unit in the x-variable. Thus, for each degree that the temperature rises, there would be a corresponding increase of 0.225 inches of precipitation based on the model.

12.2 Section Exercises

Finding Least-Squares Regression Lines and Making Predictions

Complete the following objectives for each data set.

a. Find the equation of the regression line if a statistically significant linear relationship exists at the 0.05 level of significance.

b. If a linear regression model is appropriate, use that model to make the indicated prediction.

1. a. The following table shows students' test scores on the first two tests in an introductory biology class.

Biology Test Scores

First Test, *x*	55	40	71	82	90	50	83	75	65	52	77	93
Second Test, *y*	58	43	68	86	87	51	87	70	67	55	77	90

b. If a student scored a 70 on his first test, make a prediction for his score on the second test.

2. a. The following table lists the birth weights (in pounds), x, and the lengths (in inches), y, for a set of newborn babies at a local hospital.

Birth Weights and Lengths

Birth Weight (in Pounds), *x*	5	3	8	6	9	7	7	6	6	10
Length (in Inches), *y*	16	15	18	19	19	18	20	17	15	21

b. Predict the length of an 8-pound baby.

3. a. The following table contains the number of layoffs at eight different companies over the last year and the employee satisfaction rating for each company, on a scale of 1 to 5 with 5 being the most satisfied.

Employee Satisfaction

Number of Layoffs, *x*	23	22	28	35	23	25	20	25
Employee Satisfaction Rating, *y*	4.3	4.2	3.4	2.8	3.7	3.8	4.1	3.8

b. For a company that laid off 21 employees, what would be the predicted employee satisfaction rating?

4. a. Students studying for a state achievement test were asked to keep track of the total number of hours that they spent playing video games during the two weeks before the test. In the following table, x is the number of hours spent playing video games (on any type of device) and y is the student's score on the achievement test (out of 130).

Hours Playing Video Games and Achievement Test Scores								
Hours Playing Video Games, *x*	12	15	14	10	11	13	12	14
Test Score, *y*	99	63	79	115	108	82	98	73

b. If a student spent 11.5 hours playing video games, predict his achievement test score.

5. a. The table shows the average per-paper circulation of the top 20 U.S. alternative weekly newspapers by circulation and the total number of newsroom employees in the newspaper sector.

Number of Newsroom Employees and Average Per-Paper Circulation						
Average Per-Paper Circulation, *x*	87,186	79,942	72,910	65,936	61,654	55,347
Total Number of Newsroom Employees in the Newspaper Sector, *y*	51,430	48,920	46,310	44,120	42,450	39,210

Source: Pew Research Center. Journalism and Media. 13 June 2018. http://www.journalism.org/fact-sheet/newspapers/ (15 Feb. 2019).

b. How many employees might we expect in the newspaper sector if average circulation for newspapers was 75,000?

6. a. The table below is a record of the number of miles driven between stops for gas and the amount of money spent to fill up the gas tank in Evelyn's car over several months.

Miles Driven and Cost of Gasoline							
Miles Driven, *x*	350	275	310	289	341	237	336
Cost to Fill up, *y*	27.30	20.79	25.74	21.11	28.66	21.33	27.58

b. If Evelyn needs to drive 325 miles home from college, how much should she budget for gas for the trip?

7. a. The following table shows the average mathematics and science achievement test scores of eighth-grade students from 12 countries around the world.

Mathematics and Science Achievement Test Scores												
Math, *x*	525	558	511	531	392	585	476	496	520	582	532	502
Science, *y*	540	535	518	533	420	569	450	538	535	530	552	515

b. Predict the average science score for a country with an average eighth-grade math score of 570.

8. a. The following table shows the average amount of time wasted at work per day by year of birth.

Birth Years and Time Wasted per Day					
Birth Year, *x*	1940	1955	1965	1975	1985
Time Wasted (in Hours), *y*	0.5	0.68	1.19	1.61	1.95

b. Predict the amount of time that would be wasted by someone born in 1980.

Use the regression equations you calculated in the previous examples to answer the following questions.

9. Refer to exercise 3 regarding the number of layoffs and employee satisfaction rating. How does each additional layoff affect employee satisfaction?

10. Refer to exercise 4 regarding students' test scores and the number of hours spent playing video games. What is the effect on a student's test score for each additional hour playing video games?

11. Refer to exercise 5 regarding the circulation of newspapers and the number of newsroom employees. What is the effect on the number of employees for each additional 10,000 papers in circulation?

12. Refer to exercise 6 regarding Evelyn's spending on gas for her car. We are used to seeing the price of gas per gallon, and even calculating miles per gallon. However, using the regression line you calculated in exercise 6 we can look at the cost in a different way. What is the cost of gas for each additional mile driven?

12.3 Section Exercises

Regression Analysis

Complete the following objectives for each data set.

a. Calculate the sum of squared errors, SSE.

b. Calculate the standard error of estimate, S_e.

c. Construct a 95% prediction interval for the given value of the explanatory variable.

1. The following table contains data from a sample of ten people regarding their weights and the times it took them to run/walk one mile. The times are in minutes and the weights are in pounds.

Construct a 95% prediction interval for time given $x = 175$.

Weights and Times to Run/Walk One Mile

Weight (in Pounds), *x*	178	182	180	165	159	170	189	193	195	203
Time (in Minutes), *y*	13.24	15.32	15.21	12.04	12.21	13.10	16.75	16.98	17.02	17.19

2. The following table contains data from a sample of ten students regarding their final exam scores in the first and second semesters of English Composition.

Construct a 95% prediction interval for the final exam score in the second semester given $x = 70$.

Exam Scores

Final Exam 1, *x*	87	65	99	78	81	86	61	63	90	100
Final Exam 2, *y*	89	72	93	81	75	87	72	69	88	89

3. The following table contains the scores from a group of 15 high school seniors on a psychological assessment of positive affect for the subject of math and their scores on the same assessment for the subject of chemistry. The assessment measures the strength of the student's positive feelings towards the subject on a 40 point scale, with 40 being the most positive.

Construct a 95% prediction interval for score on the chemistry assessment given $x = 20$.

Positive Affect Assessment

Score for Math Assessment, *x*	15	5	32	24	33	20	19	8	12	40	21	15	24	3	10
Score for Chemistry Assessment, *y*	20	9	12	25	38	12	23	10	9	33	22	17	18	5	11

4. The following table contains data regarding golf scores and the average lengths of drives (in yards) for a sample of five golfers.

Construct a 95% prediction interval for the length of the golfer's drive given $x = 72$.

Golf Scores and Average Drive Lengths

Golf Score, *x*	68	72	69	70	75
Length of Drive (in Yards), *y*	275	236	289	245	225

5. The following table contains data from a sample of eight people regarding the average number of fat grams consumed per day and the amount of weight lost (in pounds) over a period of six months.

Construct a 95% prediction interval for the amount of weight lost given $x = 55$.

Fat Consumption and Weight Lost

Fat Consumed (in Grams), *x*	40	20	25	50	65	75	78	80
Weight Loss (in Pounds), *y*	23	30	38	15	10	8	7	5

6. The following table contains data from a sample of eight people regarding a person's physical fitness (measured on a scale of 1 to 10, with 10 being perfectly fit) and the number of days it took for the person to recover from gall bladder surgery.

Construct a 95% prediction interval for the length of the recovery time given $x = 5$.

Physical Fitness Levels and Recovery Times

Physical Fitness, *x*	8	7	6	5	4	3	2	1
Recovery Time (in Days), *y*	5	5	6	7	7	9	10	12

7. The following table contains data from a sample of eight people regarding the number of bowls of chicken soup consumed while having cold symptoms and the number of days the symptoms persisted.

Construct a 95% prediction interval for the duration of the cold given $x = 3$.

Bowls of Soup and Duration of Cold Symptoms

Bowls of Soup, *x*	5	5	4	4	3	2	1	0
Duration of Cold Symptoms, *y*	3	4	3	5	4	5	6	7

8. The following table contains data from a sample of five people and their cholesterol scores before and after taking a cholesterol medication.

Construct a 95% prediction interval for the cholesterol level after medication given $x = 225$.

Cholesterol Level Before and After Medication

Cholesterol Level Before Medication, *x*	220	230	267	200	195
Cholesterol Level After Medication, *y*	154	167	186	132	125

Constructing Confidence Intervals for the *y*-intercept and the Slope of the Regression Line

Complete the following objectives for each data set.

a. Calculate the sum of squared errors, SSE.
b. Calculate the standard error of estimate, S_e.
c. Construct a 95% confidence interval for the *y*-intercept of the regression line.
d. Construct a 95% confidence interval for the slope of the regression line.

Note: *Microsoft Excel can be used to calculate the answers for all parts simultaneously.*

9. The following table contains data from a sample of ten students regarding the numbers of parking tickets they received during one semester and their monthly incomes (including allowances from parents as well as paychecks from employment as income).

Parking Tickets and Monthly Income

Number of Tickets, *x*	10	8	3	2	0	5	4	2	1	0
Monthly Income (in Dollars), *y*	4000	3800	1500	2000	870	2500	1800	1000	1200	1400

10. The following data were collected from a sample of fathers and sons. The heights are given in inches.

Heights of Fathers and Sons (in Inches)

Height of Father, *x*	72	70	73	68	74	73	69	70	75
Height of Son, *y*	73	69	73	70	73	70	68	71	76

11. The following data were collected from a sample of students on the numbers of times they were tardy for their world history class and their final exam grades in the course.

Numbers of Tardies and Final Exam Grades

Number of Tardies, *x*	9	8	7	6	5	2	0
Final Exam Grade, *y*	57	64	69	70	72	84	91

12. The table below shows the heights of clients, measured in inches, and their score, out of 20 points, on a self-esteem test administered by a psychologist.

Heights of Clients

Height (in Inches), *x*	69	68	71	72	68	67	70	75
Self-Esteem, *y*	15	13	17	18	15	14	17	19

the old equation to determine which equation is the best model for the given data. Most importantly, however, we should remember that a multiple regression model does not imply causation any more than a regression model with one explanatory variable does. Just as we know that an increase in a child's age does not *cause* the child's reading level to increase, we should not presume that adding more variables to a regression equation accounts for all the possible influences on the response variable, or that the variables included are even *causing* the changes in the response variable at all.

12.4 Section Exercises

Interpreting ANOVA Tables

The following ANOVA tables were generated using Microsoft Excel. Based on the ANOVA table given, is there sufficient evidence at the 0.05 level of significance to conclude that the linear relationship between the explanatory variables and the response variable is statistically significant?

1.

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2677.934566	1338.967283	167.6684672	5.21957E-10
Residual	13	103.8154339	7.985802609		
Total	15	2781.75			

2.

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	4.272742323	2.136371162	8.88737913	0.004285466
Residual	12	2.88459101	0.240382584		
Total	14	7.157333333			

3.

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1691672.173	563890.7244	11.50954064	0.080992754
Residual	2	97986.66027	48993.33013		
Total	5	1789658.833			

4.

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	8.553919903	4.276959952	3.555389644	0.061293877
Residual	12	14.43541343	1.202951119		
Total	14	22.98933333			

For each of the following computer outputs, write the multiple regression equation. Define each of the variables used in the multiple regression equation. Could one of the explanatory variables be eliminated from the model? If so, which one?

5. The following table was generated from a sample of 16 statistics students regarding each student's number of absences in the class, average number of hours spent studying for the class per week, and final grade in the class.

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	63.88047809	4.914092338	12.99944602	7.95988E-09
Absences	−1.416204746	0.422223932	−3.354155552	0.005179159
Hours Studied per Week	4.168586523	0.73626183	5.661826206	7.77404E-05

6. The following table was generated from a sample of 16 babies regarding the weight of the mother at birth, the weight of the father at birth, and the weight of the baby at birth. All weights are measured in pounds.

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	−2.882473037	0.9737576	−2.960154597	0.01105403
Mother's Weight	0.356055126	0.090343376	3.941131517	0.001689591
Father's Weight	1.02230869	0.099308161	10.29430688	1.28376E-07

7. The following table was generated from a sample of 15 babies regarding the number of weeks of gestation, the number of prenatal doctor visits the mother had, and the birth weight of the baby in pounds.

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	−19.56998222	5.716106062	−3.423656245	0.005044399
Weeks of Gestation	0.684805526	0.12588231	5.44004576	0.000150053
Number of Prenatal Visits	0.063077042	0.126763644	0.497595684	0.627761713

8. The following table was generated from a sample of 20 college freshmen regarding their high school GPAs, their ACT scores, and their GPAs after their first year at college.

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	−0.616112354	0.345507101	−1.783211841	0.092415337
High School GPA	0.975068267	0.186279465	5.234437779	6.7297E-05
ACT Score	0.015000337	0.019048101	0.787497806	0.441832554

Constructing and Analyzing a Multiple Regression Model

For each data set, use available technology, such as Microsoft Excel, to determine if a statistically significant linear relationship exists between the explanatory variables and the response variable at the 0.05 level of significance. If the linear relationship is significant, identify the multiple regression equation that best fits the data.

9. The following data were collected to explore how tuition costs, student-to-faculty ratios, and extra time to graduation index affect freshman enrollment. The student-to-faculty ratio is expressed as the number of students per faculty member. The extra time to graduation index is expressed as the average number of semesters students take above the normal time to completion.

Effects on Freshman Enrollment

Tuition Cost (in Dollars), x_1	Student-to-Faculty Ratio, x_2	Extra Time to Grad Index, x_3	Freshman Enrollment, y
5495	18	1.5	3219
7912	18	2	2781
5812	6	1.7	3556
7180	4	2.1	2987
8883	15	3	1794
6495	19	1.8	3014

10. The following data were collected to explore how the average number of hours a junior high student is unsupervised per night and the average number of hours per night the student watches television affect his class average.

Effects on Junior High Students' Class Averages

Hours Unsupervised, x_1	Hours Watching TV, x_2	Class Average, y
0	2	96
0	3	91
0.5	3	88
1	2	92
1	1	94
1.5	2	91
2	4	87
3	3	85
3	4	81
4	5	80
4.5	4	77
5.5	5	72

11. The following data were collected to explore how people's ages and their average daily calcium intakes (measured in mg per day) affect their bone mineral density.

Effects on Bone Density

Age (in Years), x_1	Daily Calcium Intake (in mg per Day), x_2	Bone Density (in mg/cm^2), y
34	791	950
43	832	875
47	865	804
52	901	723
58	923	691

12. Common predictors for freshman grade point average are ACT scores and high school GPA. Let's explore how well two unusual factors, a student's age when entering college and the number of parking tickets the student receives during the student's first semester at school, predict a student's GPA after one semester in college.

Effects on GPA

Student's Age, x_1	Number of Parking Tickets, x_2	GPA, y
20	0	3.6
21	0	3.9
17	0	2.4
18	0	3.1
19	1	3.5
35	1	4.0
23	1	3.6
19	2	2.8
18	2	3.0
18	2	2.2
24	3	3.9
19	3	3.1
18	5	2.1
17	7	2.8
18	8	1.7

Challenge Questions

Answer each of the following questions.

13. Which value in the ANOVA table tells you the number of explanatory variables in the multiple regression model?

14. Should you always use as many explanatory variables as possible?

15. What is the disadvantage of the multiple coefficient of determination?