

Additional Exercises

1. In a study, seat belt users were found to have 20% fewer fatalities than those who do not wear seat belts. Do these results prove that seat belts reduce the chances of a fatality?
2. States having an abundance of coastline have an obvious advantage over landlocked states, or states with little coastline, in that their economies may profit from an extensive fishing industry, tourism, shipping or other water related activities. Alaska, the leader by far in miles of coastline, has a total of 6640 miles, of which 5580 miles border the Pacific Ocean and 1060 miles border the Arctic. Florida, the leader in the contiguous United States, has a total of 1350 miles with 580 miles on the Atlantic Ocean and 770 miles on the Gulf of Mexico. Of all states with some coastline, New Hampshire, with 13 miles of coastline, is in last place.
 - a. Identify the population.
 - b. What characteristic is being measured?
3. A young actuary (statistician usually working in the insurance industry) has been asked to summarize the number of automobile accident claims by region for his company. He randomly selects 50 automobile accident claims which his company has settled in the last year and counts the number of accidents in each region: North, South, East and West. He summarizes the counts by region in a chart and gives the results to his supervisor.
 - a. Identify the population.
 - b. What characteristic of the population is being measured?
 - c. Identify the sample.
 - d. Is the purpose of the data collection to perform descriptive or inferential statistics?
4. In a study in *Psychological Reports*, a sample of 547 nonsmokers was selected from a large Midwestern university. The sample contained 330 women and 217 men. These students were asked to complete a questionnaire concerning students who smoked. Seventy-six percent of those surveyed strongly agreed with the statement, *I am less likely to want a smoker for a roommate*. Sixty-three percent of those surveyed strongly agreed with the statement, *I am less likely to want a smoker for a date*. Sixty-seven percent of those surveyed strongly agreed with the statement, *I am less likely to want a smoker for a husband or wife*.
 - a. Identify the population.
 - b. What characteristics of the population are being measured?
 - c. Identify the sample.
 - d. Give one statistical inference that may be made from the data.
 - e. Why is this called an inference?

Additional Exercises

- Suppose you are the administrator of a public school system. What kinds of variables would you measure and how would you collect the measurements on the following subjects:
 - Student learning
 - School discipline
 - Teacher preparation
 - Absenteeism (pupil and teacher)
 - Cafeteria food quality
- The head of the Veterans Administration has been receiving complaints from a Vietnam Veterans organization concerning disability checks. The organization claims that checks are continually late. The checks are to arrive no later than the tenth of each month.
 - What variables would you measure to explore this problem?
 - How would you collect measurements on these variables?
- How would you make informed investment decisions with a sizable sum of money that has been unexpectedly bequeathed to you by a family member?
 - What specific measurements would you use to evaluate potential investments?
 - What types of data and information would you need to consider to make informed decisions?
- Flying Eagle Airlines advertises that it surpasses all other airlines in flights that arrive on time. A competitor states that it has a better on-time record than any other airline. Can both claims be correct? Explain.
- Two local grocery stores both claim to have the lowest prices in town. Develop a measurement that you believe could be used as a criterion to determine which store actually has the lowest prices.
- At the end of 2001, the United States had 32.9 million people living in poverty according to the Census Bureau.⁸⁹ This was an increase of 1.3 million from the previous year. Poverty was defined by the Census Bureau as having a cash income less than \$14,255 a year. The Census Bureau does not include in their income measurement any part of \$167 billion spent on Medicaid, a federal program by which medical care is provided to the poor. The Census Bureau only includes \$34.9 billion out of the \$205 billion spent annually on public welfare. Forty percent of those classified as impoverished own their own homes. How do you think poverty should be defined?
- The quality movement has compelled American businesses to address the problem of measuring customer satisfaction. How would you measure customer satisfaction if you owned a car dealership?
- Identify the following variables as discrete or continuous.
 - Average test score on a test ranging from 0 to 100
 - Number of boot errors on a computer

- c. A stock's earnings per share
 - d. Energy usage in a production process
9. Determine the level of measurement for each of the following variables.
- a. Golf score in relation to par
 - b. SAT score
 - c. Rating from 1 to 5 of quality of service in a restaurant
 - d. Make and model of a vehicle
 - e. The number of students with a business major
10. According to a Danish researcher, if you drop your average daily activity level by taking elevators instead of stairs, by parking your car in the closest space, or by never walking to run errands, you increase your risk of diabetes, heart disease, and premature death. The researcher studied two groups of healthy men (eight in the first group with an average age of 27 and an average body mass index (BMI) of 22.9, which is well within the normal range; and ten in the second group with an average age of 23.8 years and a BMI of 22.1). In addition to age and BMI, researchers also collected information such as number of steps per day (each group of men was fitted with pedometers), height, weight, and race. With the first group of men, the researchers asked that they reduce their daily activity (steps) by taking cars on short trips and elevators instead of stairs. The insulin levels were also measured for each group and the researchers found that with the reduced activity, insulin levels rose by nearly 60 percent after two weeks of inactivity, thus increasing the risk of diabetes and heart disease.⁹⁰ However, the good news is that by increasing activity over a two-week period of time, one can begin to reduce his or her risk of diabetes and heart disease.
- a. List the different variables measured in this study.
 - b. Which variables are quantitative and which are qualitative?
 - c. Of the variables that are quantitative, are they discrete or continuous?
 - d. Give the levels of measurement for these variables.
11. Consider the world production of crude oil given in millions of barrels per day.

Data

The data set is available on stat.hawkeslearning.com under **Discovering Statistics and Data, Fourth Edition > Data Sets > World Crude Oil Production**.

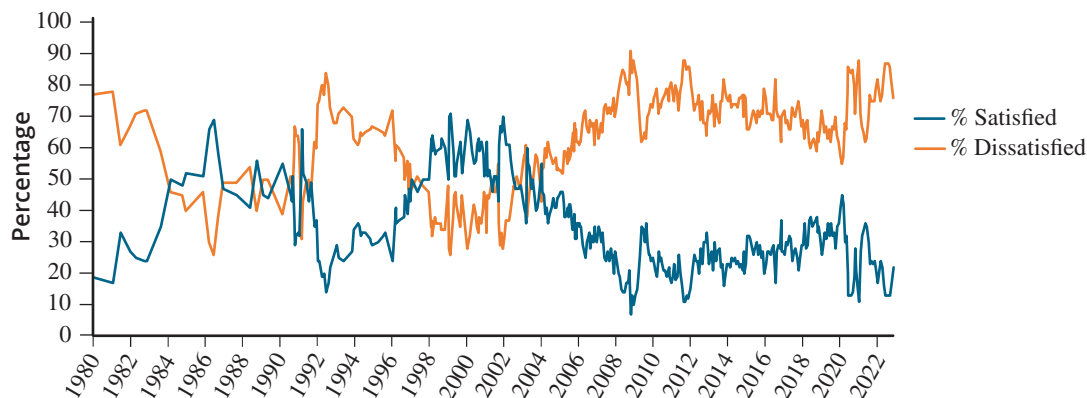
| World Production of Crude Oil | | | |
|-------------------------------|--|------|--|
| Year | World Crude Oil Production (Million barrels per day) | Year | World Crude Oil Production (Million barrels per day) |
| 1972 | 53.7 | 1997 | 70.4 |
| 1973 | 55.9 | 1998 | 71 |
| 1974 | 59.1 | 1999 | 71.6 |
| 1975 | 59 | 2000 | 72.2 |
| 1976 | 60.9 | 2001 | 73.2 |
| 1977 | 62.7 | 2002 | 72.8 |
| 1978 | 63.6 | 2003 | 73.6 |
| 1979 | 63.9 | 2004 | 73.1 |

| World Production of Crude Oil | | | |
|-------------------------------|---|------|---|
| Year | World Crude Oil Production (Million barrels per day) | Year | World Crude Oil Production (Million barrels per day) |
| 1980 | 63 | 2005 | 73 |
| 1981 | 59 | 2006 | 72.2 |
| 1982 | 58.4 | 2007 | 73.3 |
| 1983 | 58.7 | 2008 | 74.2 |
| 1984 | 59.8 | 2009 | 68.1 |
| 1985 | 60.7 | 2010 | 68.7 |
| 1986 | 62.1 | 2011 | 70.3 |
| 1987 | 62.2 | 2012 | 71.5 |
| 1988 | 62.7 | 2013 | 72.2 |
| 1989 | 62.8 | 2014 | 73 |
| 1990 | 63.3 | 2015 | 73.7 |
| 1991 | 64.3 | 2016 | 77.4 |
| 1992 | 66.9 | 2017 | 80.9 |
| 1993 | 67.8 | 2018 | 82.2 |
| 1994 | 68.5 | 2019 | 80.6 |
| 1995 | 68.9 | 2020 | 91.1 |
| 1996 | 69.8 | | |

- a. What is the level of measurement of the data?
- b. Is the data time series or cross-sectional? If the data is time series, plot the data. Does the series appear to be stationary or nonstationary? Explain your answer.

12. Consider the graph of the number of respondents (in percentages) who think things in the U.S. are now on the wrong track versus those that think the country is going in the right direction. The data was collected using a survey asking the question, *In general, are you satisfied or dissatisfied with the way things are going in the United States at this time?*⁹¹

American's Views of the Way Things are Going



- a. Are the opinions on the outlook of the country presented as time series or cross-sectional data? Justify your answer.
 - b. If the data is time series data, does the series appear to be stationary or nonstationary? Explain your answer.
13. Can you think of a process that would yield measurements that did not have any variability? Would studying such a process be very interesting?
14. One of the measurements that population experts use in predicting trends in population growth is the fertility rate. The total fertility rate is sometimes defined as the number of likely births one woman will have in her lifetime. The accompanying table gives the fertility rate from 1960 to 2020.⁹² If the data is time series data, plot the data in a line chart. Make observations based on the graph as to whether the series is stationary or nonstationary. If the time series is nonstationary, identify any noticeable trends.

Data

The data set is available on stat.hawkeslearning.com under **Discovering Statistics and Data, Fourth Edition > Data Sets > Fertility Rates.**

| Fertility Rates | | | | | |
|-----------------|----------------|------|----------------|------|----------------|
| Year | Fertility Rate | Year | Fertility Rate | Year | Fertility Rate |
| 1960 | 3.6540 | 1981 | 1.8120 | 2002 | 2.0205 |
| 1961 | 3.6200 | 1982 | 1.8275 | 2003 | 2.0475 |
| 1962 | 3.4610 | 1983 | 1.7990 | 2004 | 2.0515 |
| 1963 | 3.3190 | 1984 | 1.8065 | 2005 | 2.0570 |
| 1964 | 3.1900 | 1985 | 1.8440 | 2006 | 2.1080 |
| 1965 | 2.9130 | 1986 | 1.8375 | 2007 | 2.1200 |
| 1966 | 2.7210 | 1987 | 1.8720 | 2008 | 2.0720 |
| 1967 | 2.5580 | 1988 | 1.9340 | 2009 | 2.0020 |
| 1968 | 2.4640 | 1989 | 2.0140 | 2010 | 1.9310 |
| 1969 | 2.4560 | 1990 | 2.0810 | 2011 | 1.8945 |
| 1970 | 2.4800 | 1991 | 2.0625 | 2012 | 1.8805 |
| 1971 | 2.2660 | 1992 | 2.0460 | 2013 | 1.8575 |
| 1972 | 2.0100 | 1993 | 2.0195 | 2014 | 1.8625 |
| 1973 | 1.8790 | 1994 | 2.0015 | 2015 | 1.8435 |
| 1974 | 1.8350 | 1995 | 1.9780 | 2016 | 1.8205 |
| 1975 | 1.7740 | 1996 | 1.9760 | 2017 | 1.7655 |
| 1976 | 1.7380 | 1997 | 1.9710 | 2018 | 1.7295 |
| 1977 | 1.7900 | 1998 | 1.9990 | 2019 | 1.7060 |
| 1978 | 1.7600 | 1999 | 2.0075 | 2020 | 1.6375 |
| 1979 | 1.8080 | 2000 | 2.0560 | | |
| 1980 | 1.8395 | 2001 | 2.0305 | | |

15. One of the problems associated with the management of solid waste is the NIMBY (not in my backyard) syndrome. In separate surveys taken in 1988, 1989, and 1990 the National Solid Waste Management Association asked, *Would you object to a new landfill in your community?* The percentage response is given in the table below.

| Survey Results | | | |
|----------------|--------------|--------|----------|
| Survey Date | Don't Object | Object | Not Sure |
| March 1990 | 36 | 59 | 5 |
| February 1989 | 23 | 65 | 12 |
| February 1988 | 30 | 62 | 8 |

- a. What is the level of measurement of the survey data?
 - b. Is the data time series or cross-sectional?
 - c. What other information would be useful in evaluating the results of the study?
16. In a recent study of four leading anesthetics, three hundred patients were randomly selected and assigned to be given one of the four products during a surgery. One of the products performed significantly better than the rest. Is this an observational study or a controlled experiment?
17. In the fall of 2022, Hurricane Ian swept through Florida causing billions of dollars of damage. Suppose you were the finance manager for one of the insurance companies which insured primarily residential housing. You need to be sure that adequate funds are available to pay the anticipated claims. The morning after the hurricane you sit in your office in Hartford, Connecticut, and begin wondering what kind of data you might collect in order to anticipate the company's financial obligations resulting from the hurricane. What variable(s) would you measure, and how would you collect the measurements?
18. Do seat belts affect the types and degree of injuries sustained in an automobile crash? What kind of data should you obtain to answer this question? Will the data be experimental or observational?
19. An article on the BBC website titled "Higher temperatures linked to EU asylum figures" discusses a study which found that as temperatures rise above average in agricultural areas, more people are seeking refuge abroad.⁹³ The researchers believe these "weather shocks" happen due to decreased agricultural yield, which in turn damages national GDP. They also believe the heat increases aggressive behavior. The lead author of the study told BBC News "I feel very confident that what we discovered for 2000-2014 is a causal relationship between weather and asylum applications." What variables would you measure to try and establish this link?

Additional Exercises

1. The table below contains median household incomes (adjusted to 2022 US dollars) in the United States for the years 1997 through 2021.²⁹

| Median Household Income | | | | | |
|-------------------------|-------------|----------------|------|-------------|----------------|
| Year | Income (\$) | Percent Change | Year | Income (\$) | Percent Change |
| 1997 | 37,005 | – | 2010 | 49,276 | –1.0 |
| 1998 | 38,885 | 5.1 | 2011 | 50,054 | 1.6 |
| 1999 | 40,696 | 4.7 | 2012 | 51,017 | 1.9 |
| 2000 | 41,990 | 3.2 | 2013 | 53,585 | 5.0 |
| 2001 | 42,228 | 0.6 | 2014 | 53,657 | 0.1 |
| 2002 | 42,409 | 0.4 | 2015 | 56,516 | 5.3 |
| 2003 | 43,318 | 2.1 | 2016 | 59,039 | 4.5 |
| 2004 | 44,334 | 2.3 | 2017 | 61,136 | 3.6 |
| 2005 | 46,326 | 4.5 | 2018 | 63,179 | 3.3 |
| 2006 | 48,201 | 4.0 | 2019 | 68,703 | 8.7 |
| 2007 | 50,233 | 4.2 | 2020 | 68,010 | –1.0 |
| 2008 | 50,303 | 0.1 | 2021 | 70,784 | 4.1 |
| 2009 | 49,777 | –1.0 | | | |

- a. What graphical methods would be useful in displaying the data?
 - b. Graph the data.
 - c. Write a short paragraph describing the data.
2. The following table displays the income of households headed by adults 25 years and older, tabulated by educational attainment.³⁰ The percentages shown in the table represent the proportion of households within the same level of education that fall into the given income range.

| Income and Educational Attainment, 2022 | | | | | | | | | | |
|---|----------------|-------------------|-------------------|-------------------|-------------------|--------------------|---------------------|---------------------|------------------|---------------|
| Educational Attainment | Under \$10,000 | \$10,000–\$29,999 | \$30,000–\$49,999 | \$50,000–\$69,999 | \$70,000–\$89,999 | \$90,000–\$109,999 | \$110,000–\$129,999 | \$130,000–\$149,999 | \$150,000–& Over | Median Income |
| Less than 9th grade | 14.3% | 37% | 19.8% | 10.6% | 6.4% | 4.3% | 1.9% | 2.0% | 3.6% | 28,294 |
| 9th to 12th grade, no diploma | 13.6% | 34.6% | 20.4% | 11.7% | 7.4% | 4.6% | 2.7% | 1.5% | 3.4% | 31,162 |
| High school graduate (includes equivalency) | 6.9% | 23.3% | 19.4% | 15.2% | 10.8% | 7.4% | 5.1% | 3.5% | 8.5% | 50,401 |
| Some college, no degree | 5.2% | 18.6% | 17.3% | 15.0% | 12.0% | 8.6% | 6.2% | 4.8% | 12.2% | 60,980 |

| Income and Educational Attainment, 2022 | | | | | | | | | | |
|---|----------------|-------------------|-------------------|-------------------|-------------------|--------------------|---------------------|---------------------|------------------|---------------|
| Educational Attainment | Under \$10,000 | \$10,000–\$29,999 | \$30,000–\$49,999 | \$50,000–\$69,999 | \$70,000–\$89,999 | \$90,000–\$109,999 | \$110,000–\$129,999 | \$130,000–\$149,999 | \$150,000–& Over | Median Income |
| Associate degree | 4.4% | 15.0% | 16.1% | 14.1% | 12.0% | 10.5% | 7.2% | 6.0% | 14.6% | 70,450 |
| Bachelor's degree | 2.7% | 6.9% | 10.0% | 12.0% | 10.4% | 10.1% | 8.4% | 6.9% | 32.7% | 105,552 |
| Master's degree | 2.2% | 5.4% | 7.0% | 9.5% | 10.1% | 9.2% | 8.8% | 7.4% | 40.4% | 124,341 |
| Doctorate or Professional degree | 1.8% | 5.2% | 5.5% | 7.2% | 7.3% | 6.8% | 6.3% | 6.6% | 53.4% | 235,413 |

- What graphical methods would be useful in displaying the data?
 - Use a graphics program to display the data.
 - Discuss any conclusions you made from your graph(s).
3. The Paycheck Protection Plan (PPP) was a business loan program implemented by the United States government in response to the COVID-19 pandemic. Its purpose was primarily to help small businesses keep their workforce employed during the economic downturn caused by the pandemic. The following data summarizes the percentage of businesses within certain sectors that requested PPP assistance and the percentage that received PPP loans.³¹ There are various reasons that businesses were denied PPP loans including ineligibility, insufficient application documentation, non-compliance, and depleted funds

| COVID-19 Financial Assistance | | |
|--|---------------|--------------|
| Business Sector | Requested PPP | Received PPP |
| Construction | 57.1% | 54.4% |
| Manufacturing | 69.3% | 65.7% |
| Retail trade | 69.9% | 66.3% |
| Arts, entertainment, and recreational services | 58.8% | 56.2% |
| Accommodation and food services | 74.1% | 67.7% |

- What graphical method(s) do you think would be most useful in summarizing this data? Explain your answer.
- Graph the data using the method you identified in part **a**.
- Write a short paragraph describing the data, making conclusions from the graph you constructed in part **b**.

4. The following data gives the total annual amount (in billions of dollars) of exports and imports for the United States. By examining the changes in US trade, we can gain insights into the country's economic relationship with other nations and the impact on domestic industries. Graph the data using a method that would contrast the difference between US exports and imports.³²

| Trade Balance of the United States (in billions of dollars) | | |
|--|--------|--------|
| Year | Export | Import |
| 1960 | 25.9 | 22.4 |
| 1970 | 56.6 | 54.3 |
| 1980 | 271.8 | 291.2 |
| 1990 | 535.2 | 616.1 |
| 2000 | 1083.0 | 1452.7 |
| 2010 | 1872.3 | 2375.4 |
| 2020 | 2160.1 | 2813.0 |

5. *Billboard* magazine, in cooperation with Arbitron, produces a national radio format rating. The following data were gathered from radio listeners 12 and older on the percentage of people listening to a specific radio format during a specified time period.

| Radio Formats | | | | | | |
|--------------------|---------------------------|---------------------------|--------------------------|---------------------------|---------------------------|---------------------------|
| | Mon - Fri 6 AM - 10 AM | Mon - Fri 10 AM - 3 PM | Mon - Fri 3 PM - 7 PM | Mon - Fri 7 PM - 12 AM | Mon - Sun 12 AM - 6 AM | Mon - Sun 6 AM - 12 AM |
| Adult Contemporary | 17.2 | 19.7 | 17.7 | 15.0 | 16.2 | 20.0 |
| News/Talk | 17.9 | 13.1 | 12.5 | 14.3 | 5.3 | 15.6 |
| Country | 13.0 | 13.2 | 13.2 | 10.3 | 11.7 | 14.3 |
| Album Rock | 10.0 | 10.4 | 10.9 | 9.8 | 18.7 | 10.2 |
| Top 40 | 8.9 | 9.7 | 10.9 | 12.9 | 14.3 | 4.7 |
| Urban | 7.5 | 7.6 | 8.9 | 14.1 | 11.8 | 7.1 |
| Oldies | 6.0 | 6.8 | 6.9 | 6.5 | 4.3 | 10.2 |
| Classic Rock | 4.7 | 3.6 | 3.7 | 3.9 | 6.1 | 2.9 |
| Spanish | 4.5 | 4.2 | 3.7 | 2.2 | 4.9 | 4.2 |
| Adult Standards | 3.4 | 4.2 | 3.7 | 2.7 | 0.3 | 2.8 |
| Religious | 2.1 | 1.7 | 1.8 | 1.8 | 1.3 | 2.5 |
| Classical | 1.4 | 1.7 | 1.7 | 1.9 | 0.5 | 2.3 |
| Easy Listening | 0.9 | 1.1 | 0.9 | 0.8 | 0.2 | 1.2 |
| Modern Rock | 1.0 | 1.1 | 1.3 | 1.6 | 2.4 | 0.4 |
| Adult Alternative | 1.5 | 1.9 | 2.2 | 2.2 | 2.0 | 1.6 |

- a. What kinds of graphs would be appropriate for displaying the data? Explain your choices.

- b. Graph a column of the data. Briefly analyze your graph.
- c. Create a graph that would be useful in visually comparing two columns of the data. Briefly analyze your graph.
6. The Caribbean has been a favorite vacation spot for affluent North Americans and Europeans, especially during the winter months. The following table lists the number of tourists during the first six months of the year for a number of Caribbean destinations.

| Number of Tourists | | | |
|--------------------|---------|--------|--------|
| | U.S. | Canada | Europe |
| Antigua & Barbuda | 53,811 | 10,709 | 18,591 |
| Aruba | 94,028 | 1320 | 4681 |
| Barbados | 105,236 | 51,830 | 34,562 |
| Bermuda | 250,390 | 21,241 | 11,715 |
| Bonaire | 12,210 | 352 | 2266 |
| Cayman Islands | 81,180 | 3791 | 3025 |
| Curacao | 15,186 | 572 | 6543 |
| Guadeloupe | 15,596 | 10,654 | 25,409 |
| Trinidad & Tobago | 29,110 | 12,470 | 11,820 |

- a. Create a stacked bar graph that shows where tourists from the U.S., Canada, and Europe travel in the Caribbean.
- b. Create three separate bar charts, one for American tourists, one for Canadian tourists, and one for European tourists, that show the number of people traveling to each Caribbean destination.
7. The following table contains a list of the top global corporations, ranked by the amount spent on research and development in 2020.³³

| Amount Spent on Research and Development (R&D) in 2020 (Millions of Dollars) | | | | | |
|---|------------|--------------|--------------------------|-----------------------|---------------------------|
| Rank | Company | R&D Spending | Spending as a % of Sales | Headquarters Location | Industry |
| 1 | Amazon | 42.74 | 11.1 | N. America | Software and Internet |
| 2 | Alphabet | 27.57 | 15.1 | N. America | Software and Internet |
| 3 | Huawei | 22.04 | 15.9 | Asia | Computing and Electronics |
| 4 | Microsoft | 19.27 | 13 | N. America | Software and Internet |
| 5 | Apple | 18.75 | 7.0 | N. America | Computing and Electronics |
| 6 | Samsung | 18.75 | 9.0 | Asia | Computing and Electronics |
| 7 | Facebook | 18.45 | 21.0 | N. America | Software and Internet |
| 8 | Volkswagen | 14.3 | 7.6 | Europe | Auto |

| Amount Spent on Research and Development (R&D) in 2020 (Millions of Dollars) | | | | | |
|---|-------------------|--------------|--------------------------|-----------------------|---------------------------|
| Rank | Company | R&D Spending | Spending as a % of Sales | Headquarters Location | Industry |
| 9 | Intel | 13.56 | 17.4 | N. America | Computing and Electronics |
| 10 | Roche | 13.01 | 20.6 | Europe | Healthcare |
| 11 | Johnson & Johnson | 12.2 | 14.7 | N. America | Healthcare |

- For comparative purposes, which of the two columns reporting R&D spending is more useful, and why?
 - What types of graphs would be useful in presenting the variable that you chose in the previous part? Create the graph and write two statements describing the data.
 - Use computer software to develop pie charts for the headquarters location and industry categories of the top global R&D spenders.
8. In 2021, the Pew Research Center conducted a poll among adult individuals in the United States to gather information about their usage of different social media sites.³⁴ The poll results are displayed in the table below, indicating the percentages of respondents from various demographic groups who reported using each social media platform.

| Percent of US Adults Who Say They Use Various Social Media Platforms | | | | | |
|--|---------|----------|-----------|----------|---------|
| | YouTube | Facebook | Instagram | Snapchat | Twitter |
| National | 81 | 69 | 40 | 25 | 23 |
| Gender | | | | | |
| Male | 82 | 61 | 36 | 22 | 25 |
| Female | 80 | 77 | 44 | 28 | 22 |
| Age | | | | | |
| 18 – 29 | 95 | 70 | 71 | 65 | 42 |
| 30 – 49 | 91 | 77 | 48 | 24 | 27 |
| 50 – 64 | 83 | 73 | 29 | 12 | 18 |
| 65 + | 49 | 50 | 13 | 2 | 7 |
| Income | | | | | |
| <\$30K | 75 | 70 | 35 | 25 | 12 |
| \$30K–\$49,999 | 83 | 76 | 45 | 27 | 29 |
| \$50K–\$74,999 | 79 | 61 | 39 | 29 | 22 |
| \$75K+ | 90 | 70 | 47 | 28 | 34 |
| Community | | | | | |
| Urban | 84 | 45 | 45 | 28 | 27 |
| Suburban | 81 | 41 | 41 | 25 | 23 |
| Rural | 74 | 25 | 25 | 18 | 18 |

| Percent of US Adults Who Say They Use Various Social Media Platforms | | | | | |
|--|---------|----------|-----------|----------|---------|
| | YouTube | Facebook | Instagram | Snapchat | Twitter |
| Education | | | | | |
| HS or less | 70 | 64 | 30 | 21 | 14 |
| Some College | 86 | 71 | 44 | 32 | 26 |
| College+ | 89 | 73 | 49 | 23 | 33 |

- Suggest two different types of graphs that might be useful in graphing the data.
 - Create two different graphs using the data.
 - Write a short paragraph describing the data.
9. The nation's political identification (Republican, Democrat, or Independent) changes over time. The data in the following table represent Pew Research Center results on political identification from 1994 to 2018/2019.³⁵

| US Political Identification Based on Registered Voters (% of population) | | | | | | | |
|--|------------|----------|-------------|------------|------------|----------|-------------|
| Year | Republican | Democrat | Independent | Year | Republican | Democrat | Independent |
| 1994 | 33 | 33 | 30 | 2007 | 28 | 35 | 32 |
| 1995 | 33 | 33 | 31 | 2008 | 28 | 38 | 29 |
| 1996 | 33 | 34 | 29 | 2009 | 26 | 36 | 33 |
| 1997 | 31 | 35 | 29 | 2010 | 29 | 34 | 33 |
| 1998 | 31 | 36 | 28 | 2011 | 28 | 34 | 34 |
| 1999 | 30 | 36 | 30 | 2012 | 29 | 35 | 33 |
| 2000 | 31 | 35 | 27 | 2013 | 27 | 34 | 35 |
| 2001 | 32 | 36 | 26 | 2014 | 27 | 34 | 35 |
| 2002 | 33 | 34 | 26 | 2015 | 28 | 32 | 36 |
| 2003 | 33 | 34 | 28 | 2016 | 29 | 34 | 33 |
| 2004 | 33 | 35 | 27 | 2017 | 26 | 33 | 37 |
| 2005 | 33 | 35 | 28 | *2018/2019 | 29 | 33 | 34 |
| 2006 | 31 | 35 | 29 | | | | |

Notes: The percentages don't add up to 100—the remaining can be considered as other, neither, or don't know responses.

*Due to smaller sample sizes in 2018 and 2019, the data from those years has been combined.

- What types of graphs would be useful in visualizing this data? Explain your answer.
- Construct two different types of graphs from the data.
- Examine the data and write a short paragraph on your conclusions.

Additional Exercises

1. The following data shows the age (in years) at inauguration of all United States presidents from 1900 to 2022.

42 43 46 47 51 51 51 52 54 54
55 55 56 56 60 61 62 64 69 70 78

- a. Determine the mean, median, and mode for the ages of the U.S. presidents data.
 - b. Determine the 5-number summary of the data.
 - c. Determine the interquartile range.
 - d. Does the data set contain any outliers?
2. The maximum heart rates achieved while performing a particular aerobic exercise are measured (in beats per minute) for 9 randomly selected individuals.

| Maximum Heart Rates (BPM) | | | | | | | | |
|---------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| 145 | 155 | 130 | 185 | 170 | 165 | 150 | 160 | 125 |

- a. Determine the sample variance of the maximum heart rate achieved.
 - b. Determine the sample standard deviation of the maximum heart rate achieved.
 - c. Determine the range of the maximum heart rate achieved.
 - d. What are some of the factors which might contribute to the variation in the observations?
 - e. Create a box plot of the data.
 - f. What is the percentile associated with a BPM of 145?
3. A sample of executives from small businesses were asked how many vacation days they used during the previous calendar year. The frequency distribution table summarizes the results.

| Number of Vacation Days Last Year | | | |
|-----------------------------------|-----------|----------------|-----------|
| Number of Days | Frequency | Number of Days | Frequency |
| 0 | 13 | 6 | 0 |
| 1 | 18 | 7 | 3 |
| 2 | 11 | 8 | 3 |
| 3 | 7 | 9 | 0 |
| 4 | 4 | 10 | 2 |
| 5 | 3 | | |

- a. What proportion of the executives used at least 3 vacation days in the previous year?
- b. Find the sample mean and sample standard deviation of the number of vacation days using the formulas for grouped data.

- c. Compute the interval one standard deviation about the mean.
 - d. Find the percent of data falling in the interval one standard deviation about the mean.
 - e. Is the percent of the data falling in the interval one standard deviation about the mean close to what the Empirical Rule predicts? If there is a discrepancy, what is a likely reason for the discrepancy?
4. A high school math teacher summarized the 35 math SAT scores for the students in her calculus class. The mean for the class was 521 and the median was 535. The range of the scores was 235 and the highest score in the entire class was 675. Approximately 40% of the class scored higher than 562. State whether each of the following is true or false.
- a. The 45th percentile exceeds 540.
 - b. The lowest score in the class was 440.
 - c. The z-score for a score of 510 is a negative number.
 - d. The third quartile exceeds 562.
 - e. The percentile rank of 562 is 40.
5. Consider the following number of defective circuit boards produced by two different machines on seven randomly selected days.

| Defective Circuit Boards | | | | | | | |
|--------------------------|---|---|---|---|---|---|---|
| Machine A | 2 | 3 | 7 | 4 | 5 | 1 | 0 |
| Machine B | 2 | 3 | 4 | 3 | 4 | 2 | 4 |

- a. Determine the average number of defective circuit boards produced by each machine.
 - b. Determine the variance of the number of defective circuit boards produced by each machine.
 - c. Determine the standard deviation of the number of defective circuit boards produced by each machine.
 - d. If both machines produce approximately the same number of circuit boards each day, which machine do you think is better? Why?
 - e. Is there another way of approaching this problem?
6. A basketball coach has one remaining scholarship to offer and has narrowed his choice to two players. Listed in the following table are the points scored per game over the last season for each player.

| Points Scored | | |
|---------------|-----------|---------|
| Game Number | Braudrick | Douglas |
| 1 | 27 | 35 |
| 2 | 34 | 21 |
| 3 | 29 | 50 |
| 4 | 25 | 28 |
| 5 | 28 | missed |
| 6 | 35 | 32 |

| Points Scored | | |
|---------------|------------|------------|
| Game Number | Braudrick | Douglas |
| 7 | 31 | 29 |
| 8 | 33 | missed |
| 9 | 33 | 23 |
| 10 | 25 | 35 |
| 11 | 28 | 31 |
| 12 | 32 | 36 |
| TOTAL | 360 | 320 |

- What level of measurement do the data possess?
 - What statistical criteria might you use to select the better player? Justify your answer.
 - Compute the statistics you proposed in **b**.
 - Which player is more consistent? Why?
 - What biases or errors might be present in the data?
 - Is there another way of approaching this problem?
7. Consider the literacy data from recent years that is compiled in the following table.¹⁸

| Literacy Rates | | | |
|----------------|-------------------|-------------|-------------------|
| Country | Literacy Rate (%) | Country | Literacy Rate (%) |
| Argentina | 99.00 | Ecuador | 93.60 |
| Chile | 96.40 | Suriname | 94.40 |
| Brazil | 93.20 | Mexico | 95.20 |
| Colombia | 95.60 | Guatemala | 80.80 |
| Uruguay | 98.80 | Costa Rica | 97.90 |
| Paraguay | 94.50 | El Salvador | 89.10 |
| Bolivia | 92.50 | Honduras | 88.50 |
| Peru | 94.50 | Nicaragua | 82.60 |
| Venezuela | 97.10 | Panama | 95.70 |
| Guyana | 88.50 | | |

- What is the mean literacy rate for these selected countries?
- What is the standard deviation of these literacy rates?
- Using the empirical rule, how many countries in this group would we expect to have literacy rates between one standard deviation below the mean and one standard deviation above the mean?
- How many countries in this group actually have literacy rates between one standard deviation below the mean and one standard deviation above the mean?
- What assumption did you make in answering part **c**. above?
- Using the Literacy Rate data from the companion website, compare the literacy rates for Asia to those in South America.

Data

8. A manufacturer considers her production process to be “in control” if the proportion of defective items is less than 3%. She randomly selects 200 items and determines that 9 of the items are defective.
- Determine the sample proportion of defective items.
 - Based on the sample, do you think it is reasonable for the manufacturer to conclude that the production process is “out of control”? Why or why not?
9. Late in the summer of 1996, Tiger Woods became a professional golfer. This highly publicized event followed a sensational college career at Stanford University, where Tiger won three United States Amateur championships. Tiger was not a professional very long before he had his first win on the pro tour, the Las Vegas Invitational. He received a total of \$297,000 for his accomplishment. The prize money (in thousands) for the top 40 finishers in the tournament are given in the data set Las Vegas Golf Invitational and shown below.¹⁹

Data
stat.hawkeslearning.com
 Discovering Statistics and Data,
 Fourth Edition > Data Sets >
 Las Vegas Golf Invitational

| Las Vegas Invitational Prize Money (Thousands of Dollars) | | | | | | | |
|--|------|------|------|------|------|------|-----|
| 297.0 | 60.2 | 46.2 | 31.3 | 21.4 | 14.5 | 10.7 | 8.5 |
| 178.2 | 60.2 | 31.3 | 31.3 | 14.5 | 14.5 | 10.7 | 8.5 |
| 95.7 | 46.2 | 31.3 | 24.7 | 14.5 | 14.5 | 8.5 | 8.5 |
| 95.7 | 46.2 | 31.3 | 21.4 | 14.5 | 14.5 | 8.5 | 8.5 |
| 60.2 | 46.2 | 31.3 | 21.4 | 14.5 | 10.7 | 8.5 | 8.5 |



- Find the mean.
 - Find the median.
 - Find the mode.
 - Find the 10% trimmed mean and compare it to the mean and the median.
 - Comment on the skewness of the distribution.
10. A pharmacist is interested in studying the relationship between the amount of a particular drug in the bloodstream (in mg) and reaction time (in seconds) of subjects taking the drug. Ten subjects are randomly selected and administered various doses of the drug. The reaction times (in seconds) are measured 15 minutes after the drug is administered with the following results.

| Reaction Times | | | |
|---------------------|-------------------------|---------------------|-------------------------|
| Amount of Drug (mg) | Reaction Time (Seconds) | Amount of Drug (mg) | Reaction Time (Seconds) |
| 1 | 0.5 | 6 | 0.8 |
| 2 | 0.7 | 7 | 0.9 |
| 3 | 0.6 | 8 | 0.6 |
| 4 | 0.7 | 9 | 0.9 |
| 5 | 0.8 | 10 | 1.0 |

Analyze the data collected for the study by answering the following questions:

- Do the variables selected for measurement seem appropriate for answering the question the pharmacist is interested in?
- What biases or errors might be present in the data?
- What level of measurement (nominal, ordinal, interval, ratio) do the data possess?

11. Our galaxy, the Milky Way, is 120,000 light years across and contains more than 100 billion stars. In the 1930s it was unknown whether the galaxies were moving away from one another (i.e., the universe was expanding) or moving toward one another (contracting). The astronomer Edwin Hubble attempted to measure the rate at which the galaxies were moving and discovered they were moving apart. He originally estimated they were moving apart at the rate of 500 km/s/Mpc or about 160 km/sec per million-light-years from Earth, which was called Hubble's constant. As new telescopes have come online and new methods and technologies have become available, Hubble's constant has been reestimated. The current contenders are a team at the European Space Agency which estimates the constant at 67.8 km/s/Mpc, while another team has estimated the constant at 73.24 km/s/Mpc. A recent *Science* magazine article said, "the two values are separated by a gulf of 3.4 sigma."²⁰
- What is the value of sigma in this case?
 - What questions would you like to ask about this conclusion?

| Key Formulas | |
|--------------|--|
| Section | |
| 5.3 | <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>Mean Square Error</p> $s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{\sum (e_i)^2}{n-2} = \frac{\text{SSE}}{n-2}$ <p>Total Sum of Squares</p> $\text{Total SS} = \sum (y_i - \bar{y})^2$ $\text{Total SS} = \text{SSE} + \text{SSR}$ <p>Coefficient of Determination</p> $R^2 = \frac{\text{SSR}}{\text{Total SS}} = 1 - \frac{\text{SSE}}{\text{Total SS}}$ <p style="text-align: center;">or</p> $R^2 = \left(\frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n \sum x_i^2 - (\sum x_i)^2)(n \sum y_i^2 - (\sum y_i)^2)}} \right)^2$ </div> <div style="width: 45%;"> <p>Standard Error</p> $s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\text{SSE}}{n-2}}$ <p>Sum of Squares of Regression</p> $\text{SSR} = \text{Total SS} - \text{SSE}$ <p>Correlation Coefficient</p> $r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n \sum x_i^2 - (\sum x_i)^2)(n \sum y_i^2 - (\sum y_i)^2)}}$ </div> </div> |

Additional Exercises

1. Consider the following data:

| | | | | | | | |
|----------|---|---|---|----|----|----|----|
| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| y | 1 | 4 | 9 | 16 | 25 | 36 | 49 |

- a. Plot the data points on a scatterplot.
- b. Determine the correlation coefficient.
- c. Describe the relationship between x and y .
- d. Determine the least squares line. Use x as the independent variable.
- e. Plot the least squares line on the scatterplot.
- f. Use the model to compute the error for each data point.
- g. Determine the average value of the model's errors.
- h. Determine the variance of the errors.

2. Consider the following data:

| | | | | | | | |
|----------|---|------|------|---|------|------|------|
| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| y | 1 | 1.41 | 1.73 | 2 | 2.24 | 2.45 | 2.65 |

- Plot the data points on a scatterplot.
 - Determine the correlation coefficient.
 - Describe the relationship between x and y .
 - Determine the least squares line. Use x as the independent variable.
 - Plot the least squares line on the scatterplot.
 - Use the model to compute the error for each data point.
 - Determine the average value of the model's errors.
 - Determine the variance of the errors.
3. The Road Warrior Trucking Company has kept careful records on ten hauls. The traffic manager has recorded the haul weight of each truck and its miles per gallon during ten runs with the intent of building a regression model. He wants to predict the miles/gallon for a haul based on the haul weight. The haul weights and miles/gallon information are given below. Haul weights are given in thousands of pounds.

| Trucking | |
|--------------|----------------------------|
| Miles/Gallon | Haul Weight (in thousands) |
| 4.6 | 36 |
| 4.8 | 33 |
| 5.1 | 31 |
| 4.0 | 42 |
| 4.7 | 33 |
| 5.2 | 30 |
| 4.5 | 37 |
| 4.6 | 37 |
| 4.2 | 40 |
| 4.5 | 36 |

- What level of measurement do the two variables in the table possess?
- What is the dependent variable in the model? [Hint: which variable does the traffic manager want to predict?]
- What is the independent variable in the model?
- Draw a scatterplot of the data. Based on the scatterplot, does a linear model seem appropriate?
- Write the model in symbolic form. (Assume the parameters of the model have not been estimated.)
- Use the data provided and estimate the coefficients of the linear model.
- Interpret the coefficient of the independent variable.
- Use the model to predict the miles/gallon for a truck hauling 38,000 pounds.

- i. Do you believe there is a causal relationship between haul weight and the miles/gallon? If so, which direction is the causality? Do greater haul weights cause reduced mileage, or vice versa? Does the regression analysis prove the causality?
 - j. Compute the correlation coefficient of the data.
 - k. What percentage of the variation in miles per gallon is explained by the haul weight?
4. An agricultural research station is trying to determine the relationship between the yield of sunflowers and the amount of fertilizer applied. To determine the relationship, three different fields were planted. In each field four different plots were defined. In each plot a different amount of fertilizer was used. The plot assignments for the fertilizer application were randomly selected in each field.

| Agricultural Research | | | |
|---------------------------------|--------------------------------------|---------------------------------|--------------------------------------|
| Pounds of Fertilizer (per acre) | Pounds of Sunflower Seeds (per acre) | Pounds of Fertilizer (per acre) | Pounds of Sunflower Seeds (per acre) |
| 200 | 420 | 600 | 580 |
| 200 | 445 | 600 | 600 |
| 200 | 405 | 600 | 610 |
| 400 | 580 | 800 | 630 |
| 400 | 540 | 800 | 620 |
| 400 | 550 | 800 | 626 |

- a. What level of measurement do the two variables in the table possess? Is the data developed through controlled experiment or is the data observational?
 - b. Draw a scatterplot of the data.
 - c. If a linear model is developed, which of the variables will be the dependent variable? Why?
 - d. Use least squares techniques to estimate the appropriate model.
 - e. Interpret the meaning of the slope coefficient in the model.
 - f. What percentage of the variation in pounds of sunflower seeds per acre can be explained by the amount of fertilizer used?
 - g. Predict the sunflower seed yield per acre if 500 pounds of fertilizer is applied.
5. Ten games were recently played in the National Football League. The number of passing yards and rushing yards, along with the winner (w) and loser (l) of the game, for the 20 participating teams is given below.

| National Football League | | | | | | | | | | | | | | |
|--------------------------|------|------|------|---------|------|------|------|------|---------|------|------|------|------|---------|
| Game | Team | Rush | Pass | Outcome | Game | Team | Rush | Pass | Outcome | Game | Team | Rush | Pass | Outcome |
| 1 | 1 | 60 | 320 | 1 | 4 | 8 | 88 | 190 | w | 8 | 15 | 125 | 310 | w |
| 1 | 2 | 158 | 124 | w | 5 | 9 | 75 | 183 | l | 8 | 16 | 99 | 184 | l |
| 2 | 3 | 53 | 183 | l | 5 | 10 | 110 | 182 | w | 9 | 17 | 187 | 189 | w |
| 2 | 4 | 127 | 164 | w | 6 | 11 | 43 | 328 | w | 9 | 18 | 21 | 366 | l |
| 3 | 5 | 60 | 115 | l | 6 | 12 | 95 | 192 | l | 10 | 19 | 171 | 87 | w |
| 3 | 6 | 343 | 50 | w | 7 | 13 | 56 | 237 | l | 10 | 20 | 60 | 337 | l |
| 4 | 7 | 148 | 242 | l | 7 | 14 | 200 | 132 | w | | | | | |

Before computing any statistics, if you compared winners and losers with respect to rushing and passing yardage, what would you expect to find? Compare the rushing and passing yards of winners and losers.

- a. Make a scatterplot of rushing yards versus passing yards.
 - b. Compute the correlation coefficient for these two variables. Does the sign of the correlation coefficient make sense? Explain your answer.
 - c. Compute the coefficient of determination and interpret its meaning.
 - d. How do the correlation coefficient and the coefficient of determination relate?
6. Experimental results comparing two methods for estimating maximum aerobic speed were summarized in ‘Comparison of Two Field Tests to Estimate Maximum Aerobic Speed’, and reported in *The Journal of Sports Sciences*.¹⁵ The techniques compared were the treadmill test and the track test. Data from the 17 participants is given in the table below.

| Maximum Aerobic Speed | | | | | |
|-----------------------|-------|-----------|---------|-------|-----------|
| Subject | Track | Treadmill | Subject | Track | Treadmill |
| 1 | 15 | 14 | 10 | 16 | 16 |
| 2 | 16 | 16 | 11 | 16 | 16 |
| 3 | 16 | 16 | 12 | 16 | 16 |
| 4 | 13 | 12 | 13 | 13 | 12 |
| 5 | 13 | 12 | 14 | 17 | 18 |
| 6 | 13 | 14 | 15 | 19 | 20 |
| 7 | 16 | 16 | 16 | 18 | 20 |
| 8 | 16 | 16 | 17 | 18 | 20 |
| 9 | 18 | 16 | | | |

- a. Make a scatterplot of the results.
- b. Does there appear to be a negative or positive relationship between the variables?
- c. Compute the correlation coefficient for these two variables.
- d. Create two new variables, V and W , where $V = \text{Track}/2$ and $W = \text{Treadmill}/5$.
- e. Compute the correlation coefficient between V and W .
- f. What appears to be the effect on the correlation coefficient of multiplying each variable by a constant value?

7. The following data is obtained on two quantitative variables.

| | | | | | | | | | |
|----------|------|------|------|------|------|------|------|------|------|
| x | 876 | 516 | 598 | 789 | 734 | 667 | 682 | 714 | 598 |
| y | 50.1 | 88.2 | 80.7 | 39.6 | 20.5 | 40.9 | 30.6 | 22.9 | 34.8 |

- Make a scatterplot of the data.
 - Does there appear to be a negative or positive relationship between the variables?
 - Compute the correlation coefficient.
 - Create two new variables, V and W , where $V = X - 700$ and $W = Y - 50$.
 - Compute the correlation coefficient between V and W .
 - What appears to be the effect on the correlation coefficient of subtracting a constant value from each of the variables?
8. The FBI releases crime statistics for cities categorized by population size. The table below gives data for cities containing different population groups.¹⁶ The robbery and murder rates are given for these cities together with their crime index and violent crime index. The trends for the crime index indicate the percent change in offenses known to law enforcement.

| Crime Statistics | | | | |
|--------------------|--------------|-------------|-------------|---------------------|
| City Population | Robbery Rate | Murder Rate | Crime Index | Violent Crime Index |
| Over 1,000,000 | 1.1 | 4.0 | 0.6 | 1.5 |
| 500,000 to 999,999 | 5.7 | 0.2 | 0.9 | 4.1 |
| 250,000 to 499,999 | 0.7 | 3.6 | 1.9 | 1.8 |
| 100,000 to 249,999 | 2.1 | 2.6 | 0.7 | 1.8 |
| 50,000 to 99,999 | 0.2 | 6.7 | 1.2 | 0.9 |
| 25,000 to 49,999 | 1.4 | 0.1 | 1.2 | 1.2 |
| 10,000 to 24,999 | 2.6 | 9.8 | 0.9 | 1.7 |
| Under 10,000 | 1.4 | 14.7 | 1.2 | 1.8 |

- In general, if we have k quantitative variables measured in a study, how many different correlation coefficients can be computed?
- How many different correlation coefficients can be computed from the above data?
- Which pair of variables appears to be most correlated?
- Which pair of variables appears to be least correlated?
- Do any pairs of variables appear to be negatively correlated?

Additional Exercises

1. A couple plans to have three children.
 - a. List all possible outcomes for the sexes of the three children.
 - b. Find the probability that the couple will have three girls.
 - c. Find the probability that the couple will have at least one boy.

2. 671 registered voters were surveyed and asked their political affiliation and whether or not they favor a national healthcare policy. The results of the survey are displayed in the table below.

| Survey Results | | | |
|---------------------------------|----------|-------------|------------|
| Position on National Healthcare | Democrat | Independent | Republican |
| Favor | 161 | 40 | 130 |
| Do Not Favor | 110 | 40 | 190 |

- If one of the surveyed voters is randomly selected, answer the following questions.
- a. What is the probability that the voter will be a Republican?
 - b. What is the probability that the voter will not favor a national healthcare policy?
 - c. What is the probability that the voter will be a Democrat or an Independent?
 - d. What is the probability that the voter will be a Democrat and favor a national healthcare policy?
 - e. Given that the voter is a Republican, what is the probability that the voter will favor a national healthcare policy?
 - f. If the voter does not favor a national healthcare policy, what is the probability that the voter is an Independent?
 - g. Are the events {voter is a Democrat} and {voter favors national healthcare policy} independent? Explain.
-
3. A roulette wheel has 38 outcomes labeled 1 through 36 plus 0 and 00. The wheels are supposed to be designed so that each outcome is equally likely. The numbers 0 and 00 are often referred to as house numbers because the only way that a player can win when these outcomes are observed is by directly betting on the numbers. A great deal of the money wagered on a roulette wheel is wagered on odd or even numbers, or columns or rows of numbers. The numbers 0 and 00 are not in any row or column, nor are they odd or even.
 - a. What is the probability of observing an even number (0 and 00 are neither odd nor even)?
 - b. What is the probability of observing a number between 1 and 12, inclusive?
 - c. What is the probability of observing 0 or 00?
 - d. What is the probability of observing a 4?
 - e. What is the probability of not observing 7, 13, or 21?

4. A survey of customers in a particular retail store showed that 10% were dissatisfied with the customer service. Half of the customers who were dissatisfied dealt with Bill, the senior customer service representative. If Bill responds to 40% of all customer service inquiries in the retail store, find the following probabilities.
 - a. The probability that a customer will be unhappy, given that the representative was Bill.
 - b. The probability that the service representative was not Bill, given that the customer complained.
5. A package of documents needs to be sent to a given destination, and it is important that it arrive within one day. To maximize the chances of on-time delivery, three copies of the documents are sent via three different delivery services. Service A is known to have a 90% on-time delivery record, Service B has an 88% on-time delivery record, and Service C has a 91% on-time delivery record. Assuming that the delivery services and their records are independent, what is the probability that at least one copy of the documents will arrive at its destination on time?
6. A boxcar contains six complex electronic systems. Two of the six are to be randomly selected for thorough testing and then classified as defective or not defective. If two of the six systems are actually defective:
 - a. find the probability that at least one of the two systems tested will be defective.
 - b. find the probability that both are defective.
7. Odds in favor of and odds against are often used to express chances of occurrences. For example, if the odds are 5 to 2 that it will rain tomorrow then we would be wise to carry an umbrella with us.
 - a. What are the odds of rolling a six when a single die is thrown?
 - b. What are the odds against getting a head when a coin is tossed?
 - c. What are the odds against getting 3 consecutive heads when a coin is tossed 3 times?
 - d. Suppose the odds in favor of your favorite athletic team winning this weekend are 8 to 3. What is the probability that they will win?
8. Consider a well-shuffled deck of cards with 13 hearts, 13 spades, 13 clubs, and 13 diamonds.
 - a. Find the probability that the first card dealt is a heart.
 - b. Find the probability that the first card dealt is a spade.
 - c. Find the probability that the first card dealt is not a spade.
 - d. If you know that the first card dealt will not be a spade, find the probability that it will be a heart.
 - e. Suppose you saw the bottom card, and it was the queen of hearts. What is the probability that the first card dealt will be a heart?
9. A box contains eighteen large marbles and ten small marbles. Each marble is either green or white. Twelve of the large marbles are green and four of the small marbles are white. If a marble is randomly selected from the box, what is the probability that it is white or large?

10. User passwords for a certain computer network consist of four letters followed by two numbers. How many different passwords are possible?
11. Hydraulic assemblies for landing coming from an aircraft rework facility are each inspected for defects. Historical records indicate that 8% have defects in shafts only, 6% have defects in bushings only, and 2% have defects in both shafts and bushings. One of the hydraulic assemblies is selected randomly. What is the probability that:
 - a. the assembly has a bushing defect?
 - b. the assembly has a shaft or bushing defect?
 - c. the assembly has exactly one of the two types of defects?
 - d. the assembly has neither type of defect?
12. Robin is assigned the task of setting the passwords for every computer at her office. Depending on the computer, the guidelines differ as to how the password can be set. How many different four-digit passwords using the digits 0–9 can she create under each of the following guidelines?
 - a. The passwords must be odd and greater than 3000.
 - b. The passwords must be even and greater than 4000.
 - c. The passwords must be even and less than 5000.
13. Suppose that there are eight employees at the farmer's market in downtown Charleston, SC. One employee is needed to serve as a cashier, one is needed to wash produce, and one is needed to restock. In how many ways can these tasks be assigned?
14. Argentina's soccer coach has ten strikers on his roster, and he wants to take four of them to the World Cup. In how many ways can he select the strikers for the next World Cup?
15. A machine at the local department store is filled with children's toys. The machine is filled with 28 dinosaurs, 35 balls, 18 rubber snakes, and 41 small stuffed animals. Suppose your four children want to each get a toy from the machine, and they want 2 dinosaurs and 2 small stuffed animals. Find the probability that you get what you want on the first 4 tries. (This requires some thought. Don't undercount!)

| Key Formulas | |
|--------------|--|
| Section | |
| 7.4 | <p>Variance and Standard Deviation of a Binomial Random Variable</p> $\sigma^2 = V(X) = np(1-p)$ $\sigma = \sqrt{V(X)} = \sqrt{np(1-p)}$ |
| 7.5 | <p>Poisson Probability Distribution Function</p> $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$ <p>where $e = 2.71828\dots$ and $\lambda =$ the mean number of successes.</p> |
| 7.6 | <p>Hypergeometric Probability Distribution Function</p> $P(X = x) = \frac{{}_r C_x {}_{N-r} C_{n-x}}{{}_N C_n}$ <p>where $r =$ the total number of successes possible, $N =$ the size of the total population, $n =$ the size of the sample drawn, $x =$ the number of successes in the sample of size n, and maximum of $(0, n+r-N) \leq x \leq$ minimum of (r, n).</p> <p>Expected Value of a Hypergeometric Random Variable</p> $\mu = E(X) = n \left(\frac{r}{N} \right)$ <p>Variance of a Hypergeometric Random Variable</p> $\sigma^2 = V(X) = n \left(\frac{r}{N} \right) \left(1 - \frac{r}{N} \right) \left(\frac{N-n}{N-1} \right).$ |

Additional Exercises

1. Since 1930, the FIFA World Cup has been held nearly every four years with Qatar hosting in 2022. The following table shows the probability distribution for X , the goal differential for the World Cup Final.

| Goal differential for the World Cup Final | 0 | 1 | 2 | 3 |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| Probability | $\frac{7}{20} = 0.35$ | $\frac{4}{20} = 0.20$ | $\frac{6}{20} = 0.30$ | $\frac{3}{20} = 0.15$ |

- a. What is the average goal differential in a World Cup Final?
- b. Find the variance of the goal differential in a World Cup Final.
- c. Find the standard deviation of the goal differential in a World Cup Final.
- d. What is the probability that the goal differential will be 1 for a World Cup Final?
- e. Find the probability that the goal differential will be at least 1 in a World Cup Final.

- f. Find the probability that the goal differential will be at most 1 in a World Cup Final.
 - g. What is the most likely outcome of the World Cup Final according to the empirical data shown in the probability distribution table?
2. Consider the random variable X = outcome of rolling a 100-sided fair die. Then the possible values of X are $x = 1, 2, 3, \dots, 98, 99, 100$.
- a. What type of probability distribution does X follow?
 - b. Describe the probability distribution of X both verbally and in mathematical form.
 - c. Determine the expected value and standard deviation of the random variable X .
 - d. Define another random variable Y as the numerical value of the outcome of rolling a 6-sided fair die. Compare the standard deviation of X to the standard deviation of Y . Which is larger? Explain why this is the case.
3. The US Department of Labor has issued a new set of guidelines governing certain work practices for employees. It estimates that only 20% of all firms will be subject to the new guidelines. To validate the estimate of the number of firms that will be affected by the new guidelines, the department randomly selects a sample of twenty firms for a study. Assuming their initial estimate of 20% is correct, answer the following questions.
- a. What is the probability that 1 or fewer of the sampled firms will be subject to the new rules?
 - b. What is the probability that between 15 and 25 percent of the sampled firms will be subject to the rules?
 - c. One of the directors in the department remarked he thought that ten firms out of the sample would be subject to the rules. If the initial estimate is correct, what is the chance of this occurring?
4. Historically, the probability that a library book will be returned in one week is $p = 0.50$. The head librarian for the University Staff Hospital library is monitoring a random sample of 10 books to determine if the historical proportion of the books returned within one week, 0.50, has changed. Assuming the historical return rate is still the same, answer the following questions.
- a. What is the probability that between four and six books will be returned in one week?
 - b. What is the chance that eight or more books will be returned in one week?
 - c. What is the probability that only one book will be returned in one week?
5. The number of fatalities resulting from automobile accidents for a 10-mile stretch of an interstate highway averages 1 per 100,000 automobiles. During a particular holiday weekend, 500,000 automobiles traveled over the 10-mile segment. Using a Poisson distribution, find the probability of each of the following.
- a. No fatalities
 - b. 3 fatalities
 - c. At least one fatality

6. Compute the mean and variance for the following random variables.
 - a. The number of sixes obtained in 10 rolls of a single die.
 - b. The number of hearts in a 13-card bridge hand. (Draw 13 cards from a standard deck without replacement.)
 - c. The number of free throws made by a professional basketball player in his next 10 attempts. (Assume the player makes 88% of his free throws in the long run.)
 - d. The number of cracked eggs selected when randomly selecting 5 eggs from a 12-egg carton containing 2 cracked eggs.
 - e. The number of dots on the upper face when a single die is thrown.
7. A manufacturer of digital cameras knows that a shipment of 30 cameras sent to a large discount store contains three defective cameras. The manufacturer also knows that the store will choose two of the cameras at random, test them, and accept the shipment if neither one is defective.
 - a. Find the probability that at least one is defective.
 - b. What is the probability that the shipment is accepted?
8. On a roster of sixteen softball players, four are pitchers. Nine of the softball players are selected at random without replacement. What is the probability that exactly one of the nine players is a pitcher?
9. According to the American Hotel and Lodging Association (AH&LA), women accounted for 31% of business travelers last year.² Suppose that to attract these women business travelers, the AH&LA found that 80% of hotels offered thick, comfortable bath robes and towels. Consider a random and independent sample of 15 hotels.
 - a. Based on the information given, how many of the 15 hotels are expected to offer thick, comfortable bathrobes and towels?
 - b. Find the probability that all of the hotels in the sample offer thick, comfortable bathrobes and towels.
 - c. Find the probability that more than 5 but less than 9 of the hotels in the sample offer thick, comfortable bathrobes and towels.
10. A carnival has a game of chance: a fair coin is tossed. If it lands heads, you win \$1, and if it lands tails, you lose \$0.50. How much should a ticket cost to play this game if the carnival wants to break even?
11. You are working on a multiple-choice test which consists of 15 problems. Each of the problems has five answers, only one of which is correct. If you are totally unprepared for the test and are guessing, what is the probability that you don't get a zero on the test?
12. An automobile manufacturer is always trying to improve the quality of its vehicles. Assume that the number of defects per vehicle follows a Poisson distribution. If these defects occur randomly at an average rate of five per vehicle, what is the probability that a randomly selected vehicle will have at least one defect?

13. When proofreading a statistics textbook, one can expect to find a number of errors, whether they are typographical, symbolic, or even incorrect mathematical calculations. On average, a statistics textbook will contain 30 errors. What is the probability that when proofreading a text, one finds at least three errors? Assume that the number of errors found follows a Poisson distribution.
14. A jeweler was given a collection of twelve diamonds, of which three were synthetic (fake). If the jeweler selected two of these diamonds at random (without replacement), what is the probability that neither jewel is found to be synthetic?
15. L-Mart Inspections is a building inspection company. There were ten new commercial construction buildings completed in the last month and the sites are now available for inspection. L-Mart plans to inspect some of the new constructions for code violations and believes that half of the buildings will have violations.
- What probability model would be appropriate for describing the number of buildings in the sample that have code violations? Explain your answer.
 - If L-Mart randomly selects four buildings to inspect, what is the probability that three of the buildings will have violations?
16. In the casino game of roulette, a wheel is spun and a ball is set in motion, ultimately coming to rest in one of the 38 slots on the wheel. Any slot is as likely as any other to capture the ball. Of the 38 slots, 18 are red, 18 are black, and 2 are green. Suppose the entry fee to play a single game is \$1 and the participant bets on red. If the ball comes to rest in one of the red slots, he wins \$1 in addition to getting back the original \$1 entry fee. If the ball does not end up in a red slot, the \$1 entry fee is lost. Let X denote the monetary gain when betting \$1 on red, in a single game of roulette. Gain is defined as the amount won minus the fee to play.
- What are possible values of X ?
 - Is X a discrete or a continuous random variable? Explain.
 - Construct the probability distribution of X .
 - Find the expected value of X and interpret this number.
 - Do you feel that in any casino games you would have a positive expected gain? Why?
17. An experiment consists of tossing two coins and a die simultaneously.
- List the 24 equally likely simple events.
 - Define the random variable X as the sum of the number of heads on the two coins and the number of dots on the die. What are the possible values of X ?
 - Construct the probability distribution of X in the form of a table.
 - Find the expected value of X .

- 18.** A pediatric nurse is studying the number of babies born at a hospital with congenital defects. The nurse believes that the number of babies born during a month with congenital defects follows a Poisson distribution with an average of one baby born with a congenital defect per month.
- Find the average number of babies born in one year with congenital defects.
 - Find the standard deviation of the number of babies born in one year with congenital defects.
 - Find the probability that no babies are born with congenital defects in one year.
 - Find the probability that at least 1 baby is born with a congenital defect in a year.
 - Find the probability that exactly 12 babies are born with a congenital defect in a year.
- 19.** A lifelong beer drinker claims to be able to distinguish his favorite brand of beer, Sudriser, from all others. Two of his friends, doubtful of the claim, set up a taste test involving 4 brands of beer. Sudriser is poured into two mugs and each of the other 3 brands is also poured into 2 mugs, giving a total of 8 mugs of beer. The brand in each mug is written on a slip of paper taped to the bottom of the mug. The mugs are then arranged in a random order, and our friend is told to sample each mug and choose which 2 mugs contain Sudriser. Cheerfully, our friend complies. Answer the following, assuming that the beer drinker really can't distinguish between beers.
- What is the probability that the 2 mugs containing Sudriser are chosen?
 - What is the probability that at least one of the 2 mugs containing Sudriser is chosen?
 - What is the probability that the 2 mugs chosen contain the same brand?
 - What is the probability that the 2 mugs chosen contain different brands?

| Key Formulas | |
|--------------|---|
| Section | |
| 8.2 | Normal Probability Density Function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ |
| 8.4 | Standardizing a Normal Random Variable $z = \frac{x - \mu}{\sigma}$ |
| 8.6 | Normal Approximation for the Binomial Distribution If X is a binomial random variable where $np \geq 10$ and $n(1-p) \geq 10$, then X can be approximated by a normal distribution with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$. |

Additional Exercises

- Using the standard normal tables, determine the following probabilities and draw the corresponding diagram.
 - $P(0 \leq z \leq 0.85)$
 - $P(-1.25 \leq z \leq 2.25)$
 - $P(z \geq 1.75)$
 - $P(z \leq -2.75)$
- Using the standard normal tables, determine the following probabilities and draw the corresponding diagram.
 - $P(0 \leq z \leq 1.00)$
 - $P(-2.50 \leq z \leq 3.01)$
 - $P(z \geq 3.25)$
 - $P(z \leq -2.50)$
- Find the value of z such that 0.99 of the area under the curve lies between $-z$ and z .
- Find the value of z such that 0.80 of the area under the curve lies between $-z$ and z .
- The weights of newborn babies born at a local hospital are believed to have a normal distribution with an average weight of 7.25 lb. and a standard deviation of 1 lb. If a newborn baby born at the local hospital is randomly selected, answer the following questions.
 - Find the probability that the weight of the newborn baby will be more than 8 lb.
 - Find the probability that the weight of the newborn baby will be less than 6 lb.
 - Find the probability that the weight of the newborn baby will be between 6.5 lb. and 8.5 lb.
 - Find the weight that separates the lowest 10% of the weights from the highest 90% of the weights. If babies in the lowest 10 percent of weights are kept for observation, would a baby that weighed 5 lb. be kept for observation?

6. Medication errors in a hospital can be dangerous and expensive. Medication errors are defined as giving a patient a non-prescribed medication in any quantity or the improper dosage of a prescribed medication. Suppose the national average for medication errors is one out of every 1000 patients. A hospital believes that their medication error rate is comparable to the national average. If the hospital randomly selects 5000 patients, answer the following questions.
- Find the expected number of patients in the sample that will have had a medication error.
 - What is the standard deviation of the number of patients in the sample that will have had a medication error?
 - What is the probability of observing one or more patients who have had medication errors in the sample?
 - What is the probability of observing two or more patients who have had medication errors in the sample?
 - Do you have any concerns about the accuracy of the probabilities you determined in parts **c.** and **d.**?
7. According to the 2011 Statistical Abstract of the United States, 20.5% of the scores on the critical reading portion of the SAT Reasoning Test exceeded 600.¹¹ Approximately 17.4% of the scores were less than 400, according to the same reference. Assuming that scores on the critical reading portion of the SAT are approximately normally distributed, what are the mean and the standard deviation of the scores on the critical reading portion of the SAT Reasoning Test?
8. The annual average per capita consumption of red meat in the United States in 2008 was 108.3 pounds, according to the 2011 Statistical Abstract of the United States.¹² This figure was down from a per capita average of 126.4 pounds in 1980. Assume that both in 2008 and 1980 the per capita amount of red meat consumed was a normal random variable with a standard deviation of 15 pounds.
- In 1980, what percent of the population consumed at least 100 pounds of red meat?
 - In 2008, what percent of the population consumed at least 100 pounds of red meat?
 - In 1980, what percent of the population consumed at most 130 pounds of red meat?
 - In 2008, what percent of the population consumed at most 130 pounds of red meat?
 - Do you feel that it is reasonable to assume that the per capita amount of red meat consumed has a normal distribution? Why or why not?
9. A cell phone manufacturer has developed a new type of battery for its phones. Extensive testing indicates that the population battery life (in days) obtained by all batteries of this new type is normally distributed with a mean of 700 days and a standard deviation of 100 days. The manufacturer wishes to offer a guarantee providing a discount on batteries if the original battery purchased does not exceed the days stated in the guarantee. What should the guaranteed battery life be (in days) if the manufacturer desires that no more than 5% of the batteries will fail to meet the guaranteed number of days?

10. A machine used to regulate the amount of dye dispensed for mixing shades of paint can be set so that it discharges an average of μ milliliters of dye per can of paint. The amount of dye discharged is known to have a normal distribution with a variance equal to 0.0160. If more than 6 milliliters of dye are discharged when making a particular shade of blue paint, the shade is unacceptable. Determine the setting of μ so that no more than 1% of the cans of paint will be unacceptable.
11. The length of time required to complete a college achievement test is found to be normally distributed with a mean of 75 minutes and a standard deviation of 15 minutes. When should the test be terminated if we wish to allow sufficient time for 95% of the students to complete the test?
12. A manufacturing plant utilizes 3000 electric light bulbs that have a length of life that is normally distributed with a mean of 500 hours and a standard deviation of 50 hours. To minimize the number of bulbs that burn out during operation hours, all the bulbs are replaced after a given period of operation. How often should the bulbs be replaced if we want not more than 2% of the bulbs to burn out between replacement periods?
13. Howe's Finance Corporation provides financing for customers at an automotive dealership. The average loan amount is \$24,000 with a standard deviation of \$8000. Assuming that the loan amount is normally distributed, what is the probability that a randomly selected consumer buying a car will want to finance at least \$20,000?
14. Suppose that the income of families in a large community follows a normal distribution. Two families are randomly selected and their incomes are \$55,000 and \$85,000, respectively. The two incomes correspond to z -scores of -0.5 and 2.0 respectively. Determine the mean and standard deviation of the income of families in the neighborhood.
15. Suppose that the 30th percentile of a normal distribution is equal to 756 and that the 90th percentile of this normal distribution is 996. Find the mean and standard deviation of the normal distribution.

Additional Exercises

1. A national news network is interested in the opinion which Americans have regarding a national health care policy. During the evening news they display a 900-phone number and ask their viewers to call in and respond to the question: *Do you favor a national health care policy in the U.S.?*

| Survey Responses | |
|--|----------------|
| Category | % of Responses |
| Yes | 45 |
| No | 45 |
| Do not have enough information to decide | 10 |

- a. What type of sampling technique was used for this survey?
 - b. What type of biases may be present in the responses?
 - c. Is 45% a reasonable estimate of the proportion of all Americans who favor a national health care policy? Explain.
2. An entrepreneur wants to open a new Indian restaurant in a resort community. To determine if there is a market for the new restaurant, the entrepreneur decides to conduct a survey.
- a. What is the population of interest to the entrepreneur?
 - b. Can you think of any good sources for a sampling frame?
 - c. What are the shortcomings (if any) of the sources you picked for the sampling frame?
3. A paint manufacturer is developing a new type of paint. Test panels were exposed to various corrosive conditions to measure the protective ability of the paint. Based on the results of the test, the manufacturer has concluded that the mean life before corrosive failure for the new paint is 168 hours with a standard deviation of 30 hours. If the manufacturer's conclusions are correct, find the probability that the paint on a sample of 60 test panels will have a mean life before corrosive failure of less than 150 hours.
4. Seventy-five percent of the students graduating from high school in a small Iowa farm town attend college. The town's Chamber of Commerce randomly selects 30 recent graduates and determines whether or not they will attend college.
- a. Find the probability that at least 80% of the surveyed students will be attending college.
 - b. Find the probability that at most 70% of the surveyed students will be attending college.
 - c. Find the probability that between 65% and 85% of the surveyed students will be attending college.

5. A biology professor is interested in the proportion of students at his college who are pre-med majors. In his next class he asks for the students who are pre-med majors to raise their hands. Fifty percent of the students raised their hands.
 - a. What type of sampling technique was used for this survey?
 - b. What type of biases may be present in the responses?
 - c. Is 50% a reasonable point estimate of the proportion of students at the college who are pre-med. majors? Explain.

6. It is known that the percentage return for a group of stocks in the technology sector is normally distributed with a mean of 15 percent and a standard deviation of 22 percent. Suppose you selected a random sample of 10 stocks from this sector.
 - a. What are the mean and standard deviation of \bar{x} ?
 - b. Find the interval containing 68.26% of all possible sample mean returns.

7. A restaurant wants to determine the average time to prepare meals for its customers. To aid in this process, the restaurant randomly selects the meal preparation time of 150 of its customers and finds that the average preparation time is 18 minutes with a standard deviation of eight minutes. Describe the distribution of the sample mean of preparation time for its customers.

8. With such a large number of people using text messages as a means of communication, a company is interested in determining the number of work hours lost due to text messaging. Based on a survey of 30 randomly selected employees (anonymously, of course), the company has determined that the average amount of time spent texting over a one-month period is 180 minutes with a standard deviation of 60 minutes.
 - a. What is the probability that the average amount of time spent using text messages is more than 210 minutes in this one-month period?
 - b. Thinking that it is practically impossible for her employees to spend, on average, three hours a month texting while at work, the manager conducts another survey. She randomly samples 45 employees and finds that the average amount of time spent texting while at work over a one-month period is less than 180 minutes. Is it reasonable to conclude that the average amount of time spent using text messaging has decreased since the initial survey? Justify your answer.
 - c. How might the data gathered from this sample not accurately depict the loss of productivity from text messaging?

9. Suppose that a random sample of size 64 was selected and the researcher found that the mean was 30 and the standard deviation was 4.
 - a. What is the probability that the sample mean is more than 31.25?
 - b. What assumptions were made in part a.?

10. A town is considering building a high school football stadium approximately one-half mile from a well-established housing development. The residents of the development opposed the stadium construction due to the noise coming from the stadium during games. In presenting their argument, the residents indicated that any noise more than 103 decibels would be unacceptable. Using a sample of 35 games previously played in the old arena, the town found that the average decibels were 100 with a standard deviation of 8 decibels.
- What is the probability that a randomly selected game will generate noise in excess of 103 decibels at the stadium?
 - What is the probability that a randomly selected game will generate a noise level of exactly 103 decibels?
 - Suppose a compromise was made that required the noise level to be lower than 103 decibels 95% of the time. Will the mean level of the noise have to be lowered to comply with the new regulation? If so, by how much? Assume that the standard deviation remains at 8 decibels.
11. The town manager believes that 60% of the residents will approve the construction of the proposed high school football stadium. A random sample of 100 residents will be used to estimate the proportion of residents that will approve the construction.
- Assuming that the town manager is correct and that $p = 0.6$, describe the sampling distribution of \hat{p} .
 - What is the probability that between 50% and 70% of the residents will approve the stadium construction?
12. It is believed that 90% of all adults and 85% of all kids between the ages of 12 and 17 have cell phones. Suppose a sample of 500 adults and 400 kids was taken.
- Describe the sampling distribution of the proportion of adults that have cell phones. Assume that the stated probabilities above are true.
 - Describe the sampling distribution of the proportion of kids that have cell phones. Assume that the stated probabilities above are true.
 - What is the probability that the sample proportion of adults having cell phones will be within 2% of the true proportion?
 - What is the probability that the sample proportion of kids having cell phones will be within 4% of the true proportion?
13. A survey of college students was conducted to learn about their attitudes toward alcohol abuse on college campuses. Sixty-two percent of student respondents indicated that they believe there was a high rate of alcohol abuse on college campuses. Suppose that a sample of 250 college students was taken. What is the probability that more than seventy percent believed that there was a high rate of alcohol abuse on college campuses?

14. A credit card issuer believes that 75% of college students between the ages of 18 and 22 have more than \$5000 of credit card debt. The credit card issuer conducted a survey of 500 college students between the ages of 18 and 22.
 - a. What is the probability that at least 70% of college students between the ages of 18 and 22 have credit card debt in excess of \$5000?
 - b. Assuming that the credit card issuer is correct, what is the probability that the proportion will be within three percent of the population proportion?
 - c. What is the probability that the proportion will not be within three percent of the population proportion?
15. A marketing firm conducts a survey by mail with a 20% response rate. If the firm mailed 1000 surveys for a new study, what is the probability that at least 220 individuals will respond?
16. Suppose that it has been reported by a group of researchers that the average number of hours of YouTube viewing per household per week in the United States is 50.4 hours. Suppose the standard deviation is 11.8 hours, and a random sample of 42 U.S. households is taken.
 - a. What is the probability that the sample average is more than 35 hours? If the sample average is actually more than 35 hours, what would it mean in terms of the figures presented by the researchers?
 - b. Suppose the population standard deviation is unknown. If 71% of all sample means are greater than 49 hours and the population mean is still 50.4 hours, what is the value of the population standard deviation? Use a sample size of 42.
17. Are you grumpy, anxious, irritable? Do you view the past with regret and the future with dread? If so you may have a shorter version of the gene responsible for the brain's ability to use an important neurochemical, serotonin. Researchers, whose results were reported in *Science*, found that there is a positive relationship between having a generally pessimistic attitude and having the shorter version of this gene. Approximately 70% of the population suffer from this deficiency, according to the article. Assume a random sample of 200 people is chosen and for each person included in the sample it is determined whether or not the person has the shorter version of the gene.
 - a. Find the probability that at least three-fourths of those sampled have the shorter version of the gene.
 - b. Find the probability that at most 72% of those sampled have the shorter version of the gene.
 - c. Find the probability that between 72% and 82% of those sampled have the shorter version of the gene.

Additional Exercises

1. A random sample of fifteen eleven-year-old boys is selected in order to estimate the mean height for boys belonging to that age group. The resulting measurements in inches are given in the table below.

| Heights (Inches) | | | | |
|------------------|----|----|----|----|
| 55 | 58 | 52 | 58 | 54 |
| 57 | 56 | 54 | 58 | 56 |
| 52 | 59 | 55 | 61 | 57 |

- a. Calculate the sample mean and the sample standard deviation of the heights.
 - b. Construct a 95% confidence interval for the mean height of all eleven-year-old boys.
 - c. What assumption did you make about the heights in constructing your interval?
2. R. Cramden, chief development officer for Fontana Area Transport bus company, is concerned about the declining use of the bus system. He wishes to estimate the percentage of Fontana residents who consider safety a significant factor in their decision about whether or not to ride a bus. This will be a preliminary study, so he is willing to develop an estimate with an error of 10% at a confidence level of 90%.
- a. What sample size will be needed?
 - b. Suppose 150 residents in a random sample of 500 Fontana residents say that they consider safety a significant factor in their decision about whether or not to ride a bus. With 95% confidence, estimate the true proportion of Fontana residents who think safety is a significant factor in their decision about whether or not to ride a bus.
3. According to a study conducted by the American Stock Exchange, 87% of 500 young Americans surveyed said that they can't count on Social Security as a source of income when they retire. Construct a 90% confidence interval for the proportion of young Americans who feel they can't count on Social Security as a source of income when they retire. (Assume the sample was randomly selected.)
4. The State Bureau of Standards must inspect gasoline station pumps on a regular basis to be sure they are operating properly. A recent survey of a randomly selected group of 61 pumps produced a sample mean of 9.75 gallons dispensed for a pump reading ten gallons. If the sample had a standard deviation of 1.12 gallons, find the 80% confidence interval for the mean amount of gas dispensed when a gas pump reads ten gallons.
5. In a population of nonunionized employees, 55% are sympathetic toward unionization. The American Federation of Labor has drawn a random sample of 250 persons selected from this population to investigate union interest. Construct a 90% confidence interval for the proportion of the sample that will be sympathetic toward unionization.

6. Suppose a study designed to collect data on smokers and nonsmokers uses a preliminary estimate of the proportion that smoke of 22%. How large a sample should be taken to estimate the proportion of smokers in the population with a margin of error of 0.02 with 80% confidence?
7. As part of an annual review of its accounts, a discount brokerage firm selects a random sample of 15 customers. Their accounts are reviewed for a total account valuation, which showed a mean of \$32,000 with a sample standard deviation of \$8200.
- What is a 99% confidence interval for the mean account valuation of the population of customers? Interpret the interval in terms of the problem.
 - What assumption about the account distribution is necessary to solve this problem?
8. Direct Shoes has 250 retail outlets throughout the United States. The firm is evaluating a potential location for a new outlet, based in part, on the mean annual income of the individuals in the marketing area of the new location. A random sample of size 36 was taken from the marketing area; the sample mean income is \$31,100. The population standard deviation is estimated to be \$4500. Construct a confidence interval using a confidence level of 0.95.
9. Suppose we want to determine the sample size required to give us a 95% confidence interval that estimates, to within \$500, the average salary of a state employee. Also, suppose that from a previous experiment, we know that $s = \$6300$. What is the minimum sample size required?
10. A reporter for a student newspaper is writing an article on the cost of off-campus housing. A sample of 16 efficiency apartments within a half-mile of campus resulted in a sample mean of \$1100 per month and a sample standard deviation of \$55. Construct a 95% confidence interval estimate of the mean rent per month for the population of efficiency apartments within a half-mile of campus. We will assume that this population is normally distributed.
11. Voting, Inc. specializes in voter polls and surveys designed to keep political office seekers informed of their position in a race. Using telephone surveys, interviewers ask registered voters who they would vote for if the election were held that day. In a current election campaign, Voting, Inc. has found that 220 registered voters, out of 500 contacted, favor a particular candidate. Find a 95% confidence interval estimate for the proportion of the population of registered voters that favor the candidate.
12. According to the National Funeral Directors Association, the nation's 19,000 funeral homes take in \$13 billion a year, with the average cost of a funeral being \$7181 in a recent survey. Suppose that the costs of ten random funerals taking place in an affluent suburb of a Pennsylvania city are given below:

| | | | | |
|--------|--------|----------|----------|--------|
| \$8206 | \$5819 | \$10,225 | \$6450 | \$8575 |
| \$5450 | \$7335 | \$9775 | \$10,455 | \$8225 |

- Construct a 90% confidence interval for the average funeral cost in this suburb.
- Is the average funeral cost in this suburb higher than in the U.S. as a whole?
- What population assumption needs to be made here?

| Key Formulas | |
|--------------|--|
| Section | |
| 11.6 | <p>Test Statistic for a Population Variance</p> $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \text{ with } n-1$ <p>degrees of freedom</p> <p>Assumptions:</p> <ol style="list-style-type: none"> 1. The data is obtained via a random sample of size n. 2. The population is normally distributed with mean μ and standard deviation σ. |

Additional Exercises

1. A tire company has found that the mean time required for a mechanic to replace a set of four tires is 62 minutes. After instituting a new installation procedure, the company believes that the expected time required to replace the set of four tires remains unchanged. A test of the company's belief will be performed.
 - a. What are the null and alternative hypotheses for the test of the company's belief?
 - b. Describe, in terms of the problem, how a Type I error could occur.
 - c. Describe, in terms of the problem, how a Type II error could occur.
2. Tech Transit wishes to test whether the mean number of passenger miles on a particular route exceeds 66,000 passenger miles, the number of passenger miles the company needs on that route to cover all allocated costs. A random sample of 25 trips on the route yields a mean of 70,250 miles and a standard deviation of 9000 miles. It is desired to control the significance level at 1%.
 - a. State the appropriate hypotheses for this problem.
 - b. Describe, in terms of the problem, how a Type I error could occur.
 - c. Describe, in terms of the problem, how a Type II error could occur.
3. A pain reliever currently being used in a hospital is known to bring relief to patients in a mean time of 3.5 minutes. To compare a new pain reliever with the one currently being used, the new drug is administered to a random sample of 50 patients. The mean time to relief for the sample of patients is 2.8 minutes and the standard deviation is 1.14 minutes. Does the data provide sufficient evidence to conclude that the new drug was effective in reducing the mean time until a patient receives relief from pain? Test using $\alpha = 0.10$.
4. Tech uses thousands of fluorescent light bulbs each year. The brand of bulb it currently uses has a mean life of 25,000 hours. A manufacturer claims that its new brand of bulbs, which cost the same as the brand the university currently uses, has a mean life of more than 25,000 hours. The university has decided to purchase the new brand if, when tested, the test evidence supports the manufacturer's claim at the 0.05 significance level. Suppose 64 bulbs were tested and they were found to have an average life of 25,550 hours. The population standard deviation is known to be 2000 hours. Will the university purchase the new brand of fluorescent bulbs?

5. The daily wages in a particular industry are normally distributed with a mean of \$13.20 and a standard deviation of \$2.50. If a company in this industry employing 40 workers pays these workers, on average, \$12.20, can this company be accused of paying inferior wages? Use a significance level of 1%.
6. A coin-operated soft drink machine was designed to discharge, on average, 12 ounces of beverage per cup. In a test of the machine, ten cupfuls of beverage were drawn from the machine and measured. The mean and standard deviation of the ten measurements were 12.1 ounces and 0.12 ounces, respectively. Do these data present sufficient evidence to indicate that the mean discharge differs from 12 ounces? Test using $\alpha = 0.10$.
7. Techside Real Estate, Inc. is a research firm that tracks the cost of apartment rentals in Southwest Virginia. Five years ago, the regional average apartment rental rate was \$895 per month. Assume that, based on the historical quarterly surveys, it is reasonable to assume that the population standard deviation is \$225. In a current study of apartment rental rates, a sample of 180 apartments in the region provided the apartment rental rates. Does the sample data enable Techside Real Estate, Inc. to conclude that the population mean apartment rental rate now exceeds the level reported five years ago? The sample mean is \$915 and the sample standard deviation is \$227.50. Make your decision based on $\alpha = 0.10$.
8. Suppose that the national average price for used cars is \$28,205. A manager of a local used car dealership reviewed a sample of 25 recent used car sales at the dealership in an attempt to determine whether the population mean price for the used cars at this particular dealership differed from the national mean. The prices for the sample of 25 cars are given in the data with a mean of \$27,750 and standard deviation of \$1400. Test using $\alpha = 0.05$ whether a difference exists in the mean price for used cars at the dealership.
9. Suppose you are responsible for auditing invoices. Historically, about 0.003 of the invoices possessed material errors. During the last audit cycle a number of suggestions were made and implemented, and you hope that the next audit will provide evidence of improvement in the error rate. An audit of 6000 recent invoices reveals 12 material errors.
 - a. Does the data suggest an improvement in the invoice error rate at $\alpha = 0.05$?
 - b. Compute the P -value of the test statistic.
 - c. Based on the P -value, would the decision change at $\alpha = 0.10$?
10. After completing Chemistry 101, a student decides to conduct an experiment on their favorite brand of whiskey to determine if the proof rating on the bottle is accurate. The student selects eight small eighty-proof bottles from different stores around town and measures the percent of alcohol in each bottle. (**Note:** 80-proof alcohol contains 40% alcohol.) The resulting measurements are as follows.

| Percent of Alcohol per Bottle | | | | | | | |
|-------------------------------|-----|-----|-----|-----|-----|-----|-----|
| 38% | 40% | 42% | 41% | 39% | 38% | 40% | 38% |

- a. What is the population being studied?
- b. What is the variable being measured?

- c. What level of measurement does the data possess?
- d. Can the student conclude that the actual proof of whiskey is not equal to 80 at $\alpha = 0.05$?
- e. What assumption did the student make in performing the test in part d.?
11. You have decided to become a professional gambler specializing in roulette. If the roulette wheel is fair (each number has a $\frac{1}{38}$ chance) then you will lose in the long run. However, you plan to locate wheels that are not balanced properly. An unbalanced wheel will produce some numbers more often than expected. You believe that you have found such a wheel and have started keeping track of the number 29. After 420 spins of the wheel, the number 29 has been observed 14 times. Is this overwhelming evidence at the $\alpha = 0.05$ level that you should start betting heavily on the number 29?
12. A commercial airline is concerned over the increase in weight of carry-on luggage. In the past, the airline has estimated that the average piece of carry-on luggage will weigh 12 pounds. A random selection of 148 pieces of carry-on luggage has an average weight of 14.2 pounds with a standard deviation of 3.4 pounds. Do you think that the airline's concern is justified? Use $\alpha = 0.01$.
13. Consider the following hypothesis tests for the population mean with σ known. Compute the P -value for each test and decide whether you would reject or fail to reject the null hypothesis at $\alpha = 0.01$.
- a. $H_0: \mu = 15, H_a: \mu > 15, z = 2.50$
- b. $H_0: \mu = 80, H_a: \mu < 80, z = -1.95$
- c. $H_0: \mu = 1200, H_a: \mu \neq 1200, z = 3.70$
14. Deli Delivery delivers sandwiches to neighboring office buildings during lunch time in New York City. The deli claims that the sandwiches will be delivered within 20 minutes from receiving the order. Given the hectic schedules of their customers, consistent delivery time is a must. The owner has decided that the standard deviation of delivery times should be at most 4 minutes. To determine how consistently the sandwiches are being delivered, the manager randomly selects 27 orders and measures the time from receiving the order to delivery of the sandwich. The average time to delivery of the sample was 20 minutes with a standard deviation of 4.5 minutes.
- a. Will the manager conclude at $\alpha = 0.10$ that the delivery times vary more than the owner desires?
- b. What assumption did you make about the delivery times in performing the test in part a.?
15. Consider the following hypothesis tests for the population mean with σ unknown. Compute the P -value for each of the tests and decide whether you would reject or fail to reject the null hypothesis at $\alpha = 0.05$.
- a. $H_0: \mu = 12, H_a: \mu > 12, t = 1.75, n = 25$
- b. $H_0: \mu = 0.12, H_a: \mu < 0.12, t = -2.95, n = 16$
- c. $H_0: \mu = 55, H_a: \mu \neq 55, t = 2.35, n = 8$

16. In each of the following experimental situations, give the appropriate null and alternative hypotheses to be tested. Define all terms that appear in these hypotheses.
- A random sample of 100 customers in a bank are selected and their times to be served are noted. The bank has recently retrained its tellers to be more efficient with the hope of decreasing its average time in servicing its customers, which has been 4 minutes in the past.
 - A local driver training school claims that at least 75% of its pupils pass the driving test on their first attempt. A sample of 60 students from the school are selected, and their performances on the driving test are noted. Based upon the data collected, we would like to refute the claim of the school.
 - A spokesperson for a popular diet claims that the average weight lost for someone on the diet will be at least 15 pounds over a two-month period. The amount of weight lost for each person in a sample of 10 people on the diet is determined in order to try to refute the claim of the diet spokesperson.
 - A tire company tests 68 of its new premium tires to determine if the average lifespan of the tire is more than the average lifespan of its major competitor's best tire. The average lifespan of the competitor's tire is 63,000 miles.
 - An elementary statistics student conducts an experiment in order to show that a coin from a magic kit is biased. The student flips the coin 500 times.
17. It is essential in the manufacture of machinery to utilize parts that conform to specifications. In the past, diameters of the ball bearings produced by a certain manufacturer had a variance of 0.00156. To cut costs, the manufacturer instituted a less expensive production method. The variance of the diameters of 101 randomly sampled bearings produced by the new process was 0.0021. Does the data provide sufficient evidence to indicate that the diameters of ball bearings produced by the new process are more variable than those produced by the old process? Test using $\alpha = 0.10$.
18. A national news magazine is interested in the proportion of counties in which the cost of living has decreased in the past 24 months. The news magazine believes that the true proportion is less than 30%. In a random sample of 100 counties, 20 counties had cost of living decreases.
- Test the news magazine's claim at $\alpha = 0.08$.
 - Find the P -value for this test.
19. An increasing number of businesses are offering child-care benefits for their workers. However, one union claims that more than 90% of firms in the manufacturing sector still do not offer any child-care benefits to their workers. A random sample of 350 manufacturing firms is selected, and only 28 of them offer child-care benefits.
- Does this sample result support the claim of the union? Test using $\alpha = 0.10$.
 - Calculate the P -value associated with this test.
20. During the holiday season, law enforcement officials estimated that 500 people would be killed and 25,000 injured on the nation's roads. They claimed that more than 50% of the accidents would be caused by drunk driving. A sample of 120 accidents showed that 67 were caused by drunk driving. Use this data to test their claim with $\alpha = 0.05$.

21. The quality control supervisor of a cannery is concerned about the variance of fill per can. Regulatory agencies specify that the standard deviation of the amount of fill should be less than 0.1 ounce. To determine whether the process is meeting this specification, the supervisor randomly selects ten cans, weighs the contents of each, and finds that the sample standard deviation of these measurements is 0.04. Do these data provide sufficient evidence to indicate that the variability is as small as desired? Test using $\alpha = 0.05$.

Additional Exercises

1. Black Bark, a Colorado based company, makes wood burning stoves. They are interested in comparing two designs to determine which design will produce a stove with a greater average burning time. Several prototypes of each design are tested and the time required to burn 15 pounds of wood was measured (the burning time is measured in hours). The results of the test are as follows.

| Burning Time for Stoves (Hours) | | | |
|---------------------------------|-----|-----------|------|
| | n | \bar{x} | s |
| Stove A | 32 | 9.35 | 0.50 |
| Stove B | 35 | 9.75 | 0.75 |

Is there sufficient evidence at $\alpha = 0.05$ for Black Bark to conclude that the mean burning time for Stove B is greater than for Stove A? Assume that the population standard deviations are equal. Let Population 1 be Stove A burning times and let Population 2 be Stove B burning times.

2. In each of the following experimental situations give the appropriate null and alternative hypotheses to be tested. Define all terms that appear in these hypotheses.
- a. Independent random samples of 50 male nurses and 50 female nurses are selected from a hospital. Each nurse is asked whether he or she is satisfied with the working conditions in the hospital. It is of interest to see if there is a difference between male nurses and female nurses on satisfaction with working conditions.
 - b. A group of 45 high school seniors take the SAT reasoning test both before and after a 3-month training course, which is designed to improve SAT scores. We wish to determine if the training course is effective.
 - c. Starting salaries are determined for 40 computer science majors and 40 computer engineering majors. It is of interest to determine if computer science majors tend to have higher starting salaries than computer engineers.
 - d. Random and independent samples of younger (age ≤ 30) and older (age > 30) automobile drivers are chosen and asked whether they have had a speeding ticket in the past 12 months. Are younger drivers more likely than older drivers to have had a speeding ticket in the past 12 months?
 - e. Do adjunct faculty receive lower ratings than tenured faculty on course evaluations? Random and independent samples of 10 course evaluations of adjunct faculty members and 10 tenured faculty members are collected and compared.
3. A nutritionist is interested in determining the decrease in cholesterol level which a person can achieve by following a particular diet that is low in fat and high in fiber. Seven subjects are randomly selected to try the diet for six months, and their cholesterol levels are measured both before and after the diet. The results of the study are as follows. Let Population 1 be the cholesterol levels before the diet and Population 2 be the cholesterol levels after the diet.

| Cholesterol Levels | | | | | | | |
|--------------------|-----|-----|-----|-----|-----|-----|-----|
| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Before Diet | 155 | 170 | 145 | 200 | 162 | 180 | 160 |
| After Diet | 152 | 168 | 148 | 195 | 162 | 178 | 157 |

- Is a paired design appropriate for the above experiment? Explain.
 - What assumption must be made in order to perform the test of hypothesis?
 - Does the data appear to satisfy the assumption described in part **b.**? Why or why not?
 - Can the nutritionist conclude that there is a significant decrease in average cholesterol level when the diet is used? Use $\alpha = 0.01$.
4. The design group for a monofilament cord manufacturer is testing two possible compositions of the cord for tensile strength. Composition A is more difficult to manufacture than Composition B, so the design group has decided that it will recommend Composition A only if the mean tensile strength for Composition A is shown to be significantly greater than the mean tensile strength for Composition B. Several monofilament cords of each sample are tested and the tensile strengths are measured in pounds per square inch. Assume that the population variances are unequal. Let Population 1 be the tensile strengths of Composition A and Population 2 be the tensile strengths of Composition B.

| Tensile Strength (Pounds per Square Inch) | | | |
|---|-----|-----------|------|
| | n | \bar{x} | s |
| Composition A | 20 | 52,907 | 2575 |
| Composition B | 20 | 50,219 | 1210 |

- What assumptions must be made in order to perform the hypothesis test?
 - Will the design group recommend Composition A or Composition B for the monofilament cord at $\alpha = 0.10$?
5. Typically, the fastest players on a football team are the cornerbacks and wide receivers. The primary responsibility of wide receivers is to catch passes down the field, while cornerbacks must possess the speed necessary to prevent them from making significant plays. To obtain a representative sample of all NFL players, we can examine the data from the participants of the NFL combine, an annual showcase event held prior to the draft. A sample of 77 wide receivers had an average 40-yard dash time of 4.50 seconds, while 60 cornerbacks had an average of 4.47 seconds. The standard deviations of these samples were 0.10 and 0.09 seconds, respectively, and assume the population standard deviations are equal. Let Population 1 be wide receivers and Population 2 be cornerbacks.
- Test whether there is a significant difference in speed between the two groups of football players using $\alpha = 0.05$.
 - Construct a 95% confidence interval for the true mean difference between the speed of the wide receivers and the cornerbacks. Interpret this interval in context.

6. The manufacturer of Big Bang energy drinks claims that their drink has less caffeine, so is therefore healthier than their competitor Black Mustang. Twenty 16-oz. energy drinks are sampled from each manufacturer. The average caffeine content in the sample of Big Bang energy drinks was 162 mg with a standard deviation of 7 mg. The average caffeine content in the sample of Black Mustang energy drinks was 173 with a standard deviation of 5 mg. Assume the population standard deviations are equal. Let Population 1 be Big Bang energy drinks and Population 2 be Black Mustang energy drinks.
- Is there evidence to refute the manufacturer's claim at $\alpha = 0.05$?
 - Construct an 85% confidence interval for the true mean difference in caffeine content between Big Bang and Black Mustang.
7. A team of organizational behavior managers investigated the effects of an orientation program on "first day of work" anxiety levels of new employees. Seventy-two new employees were randomly assigned to receive or not to receive a two-day company orientation program prior to their first day at work. Two hours after beginning work, each employee was given a test to measure his or her level of anxiety. The mean score was 1002 for the 37 receiving orientation and 1018 for the 35 who did not receive the orientation. Scores of employees who attended similar orientation programs in the past have had a standard deviation of 142. Scores of employees that did not attend the orientation program had this same standard deviation. Let Population 1 be employees who attended the orientation program and Population 2 be employees who did not attend the orientation program.
- Test to see if there is evidence of a difference in the mean test scores between those who participate in an orientation program and those who do not. Use a level of significance equal to 0.05.
 - Calculate the observed significance level (P -value) of this test.
8. For a consumer product, the mean dollar sales per retail outlet last year in a sample of 25 stores were \$3425 with a standard deviation of \$400. For a second product, the mean dollar sales per outlet in a sample of 16 stores were \$3250 with a standard deviation of \$175. The sales amounts per outlet are assumed to be approximately normally distributed for both products. Test to see if the first product has a better mean dollar sales record than the second product. Use the P -value approach and base your decision on a significance level of 0.01. Assume that the population variances are not equal.
9. A random sample of 10 filled sports drink bottles is taken in one bottling plant (Plant 1), and the mean weight of the bottles is found to be 22 ounces with a variance of 0.09 ounces squared. At another plant (Plant 2), 10 randomly selected bottles have a mean weight of 21 ounces with a variance of 0.04 ounces squared. Assuming the weights in both populations are normally distributed and the population variances are equal, test whether there is a difference between the average weights of the bottles being filled at the two plants. Use $\alpha = 0.05$.
10. The teenage rite of passage of rushing to the DMV on your birthday to get your driver's license may be losing its allure among young people. Social researchers believe that teens are no longer in a hurry to get their driver's licenses. In a national study conducted in 1983 with a sample of 10,000 16-year-olds in the U.S., 46% had their licenses. In a similar study in 2018, surveying a sample of 10,000 16-year-olds

in the U.S., only 25% had their licenses.⁴ Using the information provided, test to determine if there is a significant difference between the proportion of 16-year-olds that had their licenses in 1983 and 2018. Use a significance level of 0.05. Let Population 1 be 16-year-olds in 1983 and Population 2 be 16-year-olds in 2018.

11. A new method for temporarily relieving the lung congestion of cystic fibrosis patients has been introduced. The traditional method of relieving the congestion involves a series of manual techniques where the chest and back area are pounded and massaged. The new method is a mechanical vest which has been designed to perform the manual techniques. A study is conducted to measure the effectiveness of the new vest. Five cystic fibrosis patients are randomly selected and the diameter of the blood vessels in their lungs are measured after using the traditional treatment and after using the vest treatment. The larger the diameter of the blood vessels within the lungs the better the treatment. If the study provides conclusive evidence that the vest is at least as effective as the manual method in increasing the diameter of the blood vessels, the hospital will recommend the vest to its patients because the vest allows the patients to be much more independent. The results of the study are shown in the following table. Let Population 1 be patients using the traditional method and Population 2 be patients using the vest method.

| Diameter of Lung Blood Vessels (in mm) | | | | | |
|--|-----|-----|-----|-----|-----|
| Subject | 1 | 2 | 3 | 4 | 5 |
| After Traditional Method | 0.5 | 0.4 | 0.7 | 0.6 | 0.2 |
| After Vest Method | 0.6 | 0.6 | 0.7 | 0.7 | 0.5 |

- a. Is a paired design appropriate for the above experiment? Explain.
- b. What assumption must be made in order to perform the test of hypothesis?
- c. Does the data appear to satisfy the assumption described in part b.? Why or why not?
- d. Based on the data will the hospital recommend the Vest method to its cystic fibrosis patients? Use $\alpha = 0.05$.
12. A company believes that the variance in revenue from products produced in two facilities, measured in millions of dollars, is greater for Facility A than for Facility B. The sample standard deviation of a random sample of 19 products from Facility A is 1.5984 million of dollars. The sample standard deviation for a random sample of 18 products from Facility B is 1.0426 million of dollars. Assume that both population distributions are approximately normal and test the company's claim using a 0.10 level of significance. Does the evidence support the company's claim? Let the products produced in Facility A be Population 1 and let the products produced in Facility B be Population 2.
13. Shirley is analyzing her family's budget regarding how much they spend when eating out. She believes that the variance in expenditures when eating out is less when she uses cash as compared to when she uses her credit card. The following data represent a random sample of her family's cash and credit card purchases when eating out last month. Assume that both population distributions are approximately normal and test Shirley's claim using a 0.05 level of significance. Let Population 1 be the cash purchases and let Population 2 be the credit card purchases.

| | | | | | | | | | | |
|-------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Cash | \$24.24 | \$26.96 | \$22.48 | \$26.45 | \$26.74 | \$23.99 | \$25.70 | \$26.73 | \$25.12 | \$24.23 |
| Credit Card | \$20.46 | \$25.02 | \$26.36 | \$23.95 | \$25.84 | \$24.96 | \$20.82 | \$23.41 | \$24.70 | \$23.58 |

Additional Exercises

1. A pharmacist is interested in studying the relationship between the amount of a particular drug in the bloodstream in nanograms (ng) and reaction time in seconds of subjects taking the drug. Ten subjects are randomly selected and administered various doses of the drug. The reaction times are measured 15 minutes after the drug is administered with the following results.

| Reaction Times | | | |
|---------------------|-------------------------|---------------------|-------------------------|
| Amount of Drug (ng) | Reaction Time (Seconds) | Amount of Drug (ng) | Reaction Time (Seconds) |
| 1 | 0.5 | 6 | 0.8 |
| 2 | 0.7 | 7 | 0.9 |
| 3 | 0.6 | 8 | 0.6 |
| 4 | 0.7 | 9 | 0.9 |
| 5 | 0.8 | 10 | 1.0 |

A regression analysis has been performed to estimate the model, and the following output was produced.

$$\text{Reaction Time} = \beta_0 + \beta_1 \text{Amount of Drug} + \varepsilon_i$$

Regression Analysis: Reaction Time (Seconds) versus Amount of Drug (ng)

The regression equation is

$$\text{Reaction Time (Seconds)} = 0.533 + 0.0394 \text{ Amount of Drug (ng)}$$

| Predictor | Coef | SE Coef | T | P |
|---------------------|---------|---------|------|-------|
| Constant | 0.53333 | 0.07521 | 7.09 | 0.000 |
| Amount of Drug (mg) | 0.03939 | 0.01212 | 3.25 | 0.012 |

S = 0.110096 R-Sq = 56.9% R-Sq(adj) = 51.5%

Analysis of Variance

| Source | DF | SS | MS | F | P |
|----------------|----|---------|---------|-------|-------|
| Regression | 1 | 0.12803 | 0.12803 | 10.56 | 0.012 |
| Residual Error | 8 | 0.09697 | 0.01212 | | |
| Total | 9 | 0.22500 | | | |

Predicted Values for New Observations

| New Obs | Fit | SE Fit | 95% CI | 95% PI |
|---------|--------|--------|------------------|------------------|
| 1 | 0.6909 | 0.0393 | (0.6003, 0.7815) | (0.4214, 0.9605) |

Values of Predictors for New Observations

| New Obs | Amount of Drug (ng) |
|---------|---------------------|
| 1 | 4.00 |

- a. Draw a scatterplot of the data. Describe the relationship you observe between the reaction time and the amount of drug in the bloodstream. Are there any unusual observations?
- b. Find and interpret the standard deviation of the error terms in the output.
- c. Interpret the slope coefficient.

- d. What proportion of the variation in reaction time is explained by the amount of drug in the bloodstream? What other factors might affect reaction time?
 - e. Is there evidence of a linear relationship between the amount of drug in the bloodstream and reaction time? Test at the 0.05 level and the 0.01 level.
 - f. Construct and interpret a 95% confidence interval for β_1 , the slope of the line.
 - g. Find the predicted value of the reaction time for an individual who has 4 ng of the drug in the bloodstream.
 - h. Find and interpret a 95% confidence interval for the average reaction time of all individuals who have 4 ng of the drug in their bloodstreams.
 - i. Suppose a particular individual has 4 ng of the drug in the bloodstream. What would be the 95% prediction interval for the reaction time?
2. A sample of 11 online dating profiles, all placed by males, was selected from a popular dating website. In each of the selected profiles, the males gave their heights, along with other physical characteristics and preferences. Suppose that y , the number of responses to the profile over the next 30 days, was determined for each male. The following table contains the data.

| Height and Response | | | | | | | | | | | |
|----------------------------|----|----|----|----|----|----|----|----|----|----|----|
| Height (Inches) | 70 | 62 | 67 | 75 | 78 | 69 | 70 | 64 | 66 | 69 | 75 |
| Number of Responses | 14 | 7 | 10 | 18 | 17 | 12 | 15 | 9 | 12 | 14 | 17 |

- a. Draw a scatterplot of the data. Does the relationship appear to be linear?
- b. Estimate the slope and intercept of the regression equation using statistical software.
- c. Is there evidence of a linear relationship between the number of responses and height? Test at the $\alpha = 0.01$ level.
- d. Interpret the regression coefficient corresponding to height.
- e. Construct a 95% confidence interval for the slope.
- f. Compute R^2 and interpret this value.
- g. Estimate the number of responses for a male 6 feet tall. Round your answer to the nearest whole number.
- h. Construct and interpret a 95% prediction interval for the number of responses for a male who is 6 feet tall.
- i. Construct and interpret a 95% confidence interval for the average number of responses for a male who is 6 feet tall.

3. It is believed that when one is in the process of buying a home, the interest rate that is given on the loan is a function of his or her credit score. The Fair Isaac Corporation (FICO) is a major producer of credit scores. They have collected data from major lenders about buyers' history of borrowing and paying back credit. The following table contains 20 randomly selected loan applicants along with their FICO scores and the interest rate that they were given when financing their homes. With the data given, answer the following questions.

| Credit Scores and Interest Rates | | | | | |
|----------------------------------|------------|-------------------|-------------|------------|-------------------|
| Observation | FICO Score | Interest Rate (%) | Observation | FICO Score | Interest Rate (%) |
| 1 | 756 | 6.32 | 11 | 819 | 5.86 |
| 2 | 679 | 7.85 | 12 | 630 | 8.51 |
| 3 | 527 | 10.20 | 13 | 704 | 6.83 |
| 4 | 839 | 5.52 | 14 | 679 | 7.72 |
| 5 | 677 | 7.30 | 15 | 663 | 7.68 |
| 6 | 686 | 7.37 | 16 | 542 | 9.53 |
| 7 | 512 | 9.67 | 17 | 575 | 6.86 |
| 8 | 590 | 8.40 | 18 | 508 | 9.65 |
| 9 | 765 | 5.82 | 19 | 689 | 7.75 |
| 10 | 502 | 10.01 | 20 | 750 | 6.89 |

Data

stat.hawkeslearning.com Data Sets > Credit Scores and Interest Rates

- Draw a scatterplot of the data. Does there appear to be a linear relationship between FICO score and interest rate?
- Estimate the simple linear regression equation using statistical software.
- What is the estimate of the mean square error? Interpret this value.
- Test at the 5% level if a linear relationship exists between FICO scores and interest rates.
- Interpret the regression coefficient corresponding to FICO score.
- Construct a 95% confidence interval for the slope. Interpret the interval.
- Compute the coefficient of determination. Interpret this value.
- Calculate the correlation coefficient. Interpret this value.
- What is the average interest rate for a credit score of 725?
- Construct a 90% confidence interval for the average interest rate for people who have FICO scores of 725. Interpret this interval.
- Construct a 90% prediction interval for the interest rate for a person with a FICO score of 725. Interpret this interval.

4. It appears that many cellular phone service providers are making huge profits from customers using their messaging services such as text and multimedia messaging services (MMS). To that end, the cellular phone companies are using their marketing campaigns to target kids rather than adults. The belief is that kids tend to utilize their messaging services much more than adults. In fact, it is the belief that the younger one is, the more texts and MMS sent via his or her cell phone. Using the data given which reports the number of monthly messages sent by age, formulate a simple linear regression model to answer the following questions.

| Age and Message Use | | | |
|---------------------|--------------------|-----|--------------------|
| Age | Number of Messages | Age | Number of Messages |
| 78 | 7 | 37 | 1541 |
| 36 | 1607 | 69 | 6 |
| 11 | 3037 | 69 | 25 |
| 69 | 26 | 55 | 517 |
| 56 | 491 | 39 | 1439 |
| 74 | 0 | 20 | 2505 |
| 22 | 2373 | 14 | 2845 |
| 74 | 5 | 10 | 3048 |
| 10 | 3059 | 80 | 0 |
| 26 | 2155 | 59 | 295 |
| 18 | 2619 | 40 | 1374 |
| 68 | 17 | 67 | 35 |
| 10 | 3067 | | |

- Draw a scatterplot of the data. Does there appear to be a linear relationship between age and the number of messages that one sends?
- What is the estimated simple linear regression equation?
- What is the estimate of the coefficient of determination? Interpret this value.
- Test at the 5% level if a linear relationship exists between age and the number of messages sent via a cellular phone.
- Interpret the regression coefficient corresponding to age.
- Construct a 95% confidence interval for the slope. Interpret this interval.
- Calculate the correlation coefficient. Interpret this value.
- What is the average number of messages sent by a 15-year-old? Round your answer to the nearest whole number.
- Construct a 95% confidence interval for the average number of messages sent by a 15-year-old. Interpret this interval.
- Suppose Jacob's parents are contemplating giving him a cell phone but with a limited messaging plan at 500 per month. Eager to get the cell phone, Jacob, at 15 years old, promises that he won't send more than 500 messages per month and he'll also limit the number of friends that will have his phone number. In spite of Jacob's honesty and loyalty, should his parents believe that he won't send more than 500 messages per month? Explain your answer.

 **Data**

stat.hawkeslearning.com **Data Sets > Cell Phone Message Use**

 **Data**
stat.hawkeslearning.com Data

Sets > Home Size and Energy Usage

5. For the last 10 years, the Virginia Department of Mines, Minerals, and Energy (VDMME) has been promoting Energy Star, a resource for energy-efficient products and solutions. VDMME wants all energy consumers to take responsibility and exercise leadership by practicing conservation and efficiency on a daily basis. The average annual energy usage for an 1800 square foot home is 18,000 kilowatt hours. VDMME believes that this number can be significantly reduced if consumers started using Energy Star appliances. Answer the following questions based on data of 25 randomly selected homes with Energy Star appliances built within the last five years.

| Home Size and Energy Usage | | | |
|----------------------------|---------------------------|-------------------------|---------------------------|
| Home Size (Square Feet) | Annual Energy Usage (kWh) | Home Size (Square Feet) | Annual Energy Usage (kWh) |
| 2895 | 15,200 | 2180 | 13,227 |
| 3650 | 17,333 | 4492 | 19,492 |
| 2927 | 15,050 | 6450 | 25,353 |
| 6289 | 24,763 | 1583 | 11,075 |
| 7252 | 27,098 | 4170 | 18,557 |
| 4147 | 18,291 | 4189 | 18,636 |
| 6505 | 25,028 | 3920 | 18,210 |
| 1413 | 11,099 | 6833 | 26,075 |
| 2279 | 13,110 | 4469 | 19,232 |
| 3251 | 15,844 | 6141 | 24,225 |
| 2992 | 14,904 | 5084 | 21,530 |
| 6912 | 26,329 | 6746 | 26,333 |
| 2503 | 13,765 | | |

- Draw a scatterplot of the data. Does there appear to be a linear relationship between home size and the amount of annual kWh used?
- What is the estimated simple linear regression equation?
- What is the estimate of the coefficient of determination? Interpret this value.
- Test at the 5% level if a linear relationship exists between home size and the annual amount of kWh used.
- Interpret the regression coefficient corresponding to home size.
- Construct a 99% confidence interval for the slope. Interpret the interval.
- Calculate the correlation coefficient. Interpret this value.
- Suppose the James family constructed a 3200 square foot home using all Energy Star appliances. How many kilowatt hours should they expect to use in their first year in the home?
- Construct a 95% confidence interval for the average number of kWh that will be used by the James family. Interpret this interval.

6. With grade inflation being a major problem in many US high schools, college admissions offices are beginning to look at other performance measures when evaluating student applications. It is believed that many students with high grade point averages in high school will not necessarily score high on the SAT. Using the data of 30 randomly selected students that took the SAT, answer the following questions to determine if there is a linear relationship between high school GPA and SAT score.

 Data

stat.hawkeslearning.com Data
Sets > High School GPA and SAT Score

| High School GPA and SAT Score | | | |
|-------------------------------|-----------|-----------------|-----------|
| High School GPA | SAT Score | High School GPA | SAT Score |
| 3.21 | 1448 | 4.95 | 1960 |
| 2.23 | 1435 | 4.69 | 1717 |
| 2.89 | 1411 | 2.49 | 1365 |
| 1.84 | 1291 | 2.45 | 1561 |
| 3.34 | 1462 | 2.57 | 1474 |
| 2.42 | 1357 | 1.28 | 1328 |
| 2.75 | 1396 | 1.94 | 1302 |
| 2.35 | 1549 | 4.75 | 1622 |
| 4.80 | 1829 | 1.91 | 1499 |
| 1.98 | 1508 | 4.25 | 1566 |
| 2.92 | 1514 | 1.15 | 1413 |
| 4.18 | 1658 | 2.17 | 1428 |
| 4.50 | 1694 | 4.73 | 1720 |
| 4.42 | 1686 | 4.39 | 1783 |
| 4.78 | 1840 | 2.92 | 1614 |

- Draw a scatterplot of the data. Does there appear to be a linear relationship between high school GPA and SAT score?
- What is the estimated simple linear regression equation?
- What is the coefficient of determination? Interpret this value.
- Test at the 5% level if a linear relationship exists between high school GPA and SAT score.
- Interpret the regression coefficient corresponding to high school GPA.
- Construct a 95% confidence interval for the slope. Interpret the interval.
- Calculate the correlation coefficient. Interpret this value.
- What SAT score would you expect for students with a GPA of 3.5? Round your answer to the nearest whole number.

| Key Formulas | |
|--------------|--|
| Section | |
| 14.1 | <p>Estimated Multiple Regression Equation</p> $\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$ <p>where $b_0, b_1, b_2, \dots, b_k$ are estimates of their population counterparts.</p> |
| | <p>Sum of Squared Errors</p> $SSE = \sum (y - \hat{y})^2$ |
| 14.2 | <p>Coefficient of Determination</p> $R^2 = \frac{SSR}{\text{Total SS}} = 1 - \frac{SSE}{\text{Total SS}}$ |
| | <p>Adjusted R^2</p> $R_a^2 = 1 - \left(\frac{n-1}{n-k-1} \right) \frac{SSE}{\text{Total SS}}$ <p>where n is the number of observations and k is the number of independent variables in the model.</p> |
| 14.3 | <p>F-Statistic</p> $F = \frac{\frac{\text{Sum of Squares of Regression}}{k}}{\frac{\text{Sum of Squared Errors}}{n-(k+1)}}$ $= \frac{\frac{SSR}{k}}{\frac{SSE}{n-(k+1)}} = \frac{\text{Mean Square Regression}}{\text{Mean Square Error}}$ |
| | <p>100(1 - α)% Confidence Interval for an Individual Coefficient, β_i</p> $b_i \pm t_{\alpha/2, n-(k+1)} s_{b_i} \quad i = 1, \dots, k$ <p>Test Statistic for Testing the Hypothesis $\beta_i \neq 0$</p> $t = \frac{b_i - 0}{s_{b_i}} = \frac{b_i}{s_{b_i}} \quad i = 1, \dots, k$ <p>where b_i is the estimated coefficient and s_{b_i} is the standard deviation (standard error) of the estimated coefficient.</p> |

Additional Exercises

- Drew is undecided about whether to go back to school and get his master's degree. He is trying to perform a cost-benefit analysis to determine whether the cost of attending the school of his choice will be outweighed by the increase in salary he will receive after he attains his degree. He does research and compiles data on annual salaries in the industry he currently works in (he has been working for 10 years), along with the years of experience for each employee and whether or not the employee has a master's degree. He has decided that if the multiple regression model shows, with 95% confidence, that earning a master's degree is significant in predicting annual salary, and the estimated increase in salary is at least \$20,000, he will enroll in a degree program.

Data

stat.hawkeslearning.com
Discovering Statistics and Data,
Fourth Edition > Data Sets >
Industry Salaries

| Industry Salaries | | | | | |
|-------------------|---------------------|-----------------|-------------|---------------------|-----------------|
| Salary (\$) | Years of Experience | Master's Degree | Salary (\$) | Years of Experience | Master's Degree |
| 75,240 | 22 | No | 73,360 | 22 | No |
| 134,160 | 27 | Yes | 58,400 | 11 | Yes |
| 62,560 | 15 | No | 66,080 | 18 | No |
| 43,000 | 2 | No | 60,120 | 14 | No |
| 150,240 | 28 | Yes | 106,600 | 21 | Yes |
| 119,640 | 25 | Yes | 45,640 | 7 | No |
| 80,360 | 15 | Yes | 145,800 | 31 | Yes |
| 162,720 | 32 | Yes | 111,840 | 22 | Yes |
| 70,160 | 19 | No | 38,560 | 0 | No |
| 72,160 | 12 | Yes | 52,000 | 7 | No |

- Create an indicator variable, *degree*, that is equal to 1 if the employee has a master's degree and equal to 0 if the employee does not have a master's degree.
- Using statistical software, estimate the following multiple regression model.

$$\text{Salary} = \beta_0 + \beta_1 \text{Experience} + \beta_2 \text{Degree} + \varepsilon$$

Write the estimated multiple regression equation.

- According to the model, how much does salary increase on average with each additional year of experience?
 - Interpret the meaning of the coefficient of the indicator variable *Degree* in the estimated multiple regression equation.
 - According to this model, will Drew decide to enroll in a master's program? Explain your answer.
 - Why should Drew be cautious when using this model to make his decision?
2. The amount of a certain additive injected into a chemical process has a direct effect on the yield. The following table contains data on the amount of additive and yield.

| Amount of Additive and Yield | | | | | | | | | | | |
|------------------------------|------|-----|-----|------|-----|-----|------|------|-----|-----|-----|
| Additive | 12.0 | 6.7 | 5.6 | 13.2 | 8.9 | 7.8 | 12.9 | 16.4 | 4.5 | 9.6 | 5.8 |
| Yield | 96 | 50 | 42 | 82 | 76 | 70 | 89 | 94 | 15 | 75 | 32 |

- Assuming that *Yield* is the dependent variable, plot *Yield* against *Additive*. Does the relationship appear to be linear?
- Using statistical software, estimate the simple linear regression model. Identify R^2 and s_e^2 .
- In instances such as this where linearity does not hold, polynomial regression can be used to provide a better fit to the data. Polynomial regression is a special case of multiple regression where new predictor variables are formed by raising other predictor variables to integral powers. In this exercise, a new predictor will be formed by squaring the values of additive (*Add_sq*). *Yield* will then be fitted to the predictors *Additive* and *Add_sq*. The prediction equation based upon the polynomial regression is $Yield = -67.53 + 23.04 \text{ Additive} - 0.82 \text{ Add_sq}$. R^2 and s_e^2 are 0.95 and 47.53, respectively. Predict the yield when *Additive* = 16. Make this prediction using both the linear and polynomial fits. Compare your results.

- d. Compare the linear and polynomial fits to the data by the values for R^2 and s_e^2 .
- e. Which model do you believe is best to use for estimation and prediction? Explain your answer.
3. The following table contains a list of high-dividend exchange-traded funds (ETFs). Exchange-traded funds are investment funds traded on stock exchanges, much like stocks. ETFs are traditionally index funds, but ETFs can hold assets such as stocks, commodities, or bonds, and trade at approximately the same price as the net asset value of their underlying assets over the course of the trading day. ETFs may be attractive as investments because of their low costs, tax efficiency, and stock-like features. The full data set can be found on the companion site.

 **Data**

stat.hawkeslearning.com
Discovering Statistics and Data,
Fourth Edition > Data Sets >
Exchange-Traded Funds

| Exchange-Traded Funds | | | | | | | |
|-----------------------|------------------|-------------------------|--------------------|-----|------------------|-------------------------|--------------------|
| ETF | Share Price (\$) | Dividend Per Share (\$) | Dividend Yield (%) | ETF | Share Price (\$) | Dividend Per Share (\$) | Dividend Yield (%) |
| 1 | 4.32 | 0.28 | 6.49 | 6 | 120.55 | 6.22 | 5.16 |
| 2 | 15.38 | 0.92 | 6.02 | 7 | 23.00 | 1.09 | 4.74 |
| 3 | 25.25 | 1.43 | 5.66 | 8 | 13.34 | 0.62 | 4.66 |
| 4 | 21.28 | 1.13 | 5.31 | 9 | 24.82 | 1.15 | 4.63 |
| 5 | 698.75 | 36.88 | 5.28 | 10 | 22.96 | 1.04 | 4.53 |
| ... | | | | | | | |

- a. Using the Exchange-Traded Funds data set, can dividend yield be predicted by share price and dividend per share? Is it a useful model? Justify your answers.
- b. Which variable explains the greatest amount of variability in dividend yield? Explain your answer.

Data

stat.hawkeslearning.com

Discovering Statistics and Data,
Fourth Edition > Data Sets > SNAP

4. The Supplemental Nutrition Assistance Program (SNAP) provides monthly benefits that help eligible low-income households buy the food they need for good health. For most households, SNAP funds account for only a portion of their food budgets, so they must also use their own funds to buy enough food to last throughout the month. Using the SNAP data set, answer the following questions to help predict monthly benefits to eligible households.

| SNAP Benefits | | | | | |
|----------------------|-------------|---------------------------|----------------------|-------------|---------------------------|
| Monthly Benefit (\$) | Family Size | Gross Monthly Income (\$) | Monthly Benefit (\$) | Family Size | Gross Monthly Income (\$) |
| 603.41 | 5 | 3753 | 556.42 | 1 | 3098 |
| 560.69 | 3 | 3778 | 569.05 | 8 | 3707 |
| 623.24 | 6 | 3609 | 365.80 | 8 | 2071 |
| 416.12 | 5 | 2262 | 489.08 | 5 | 3166 |
| 323.90 | 1 | 1966 | 495.86 | 4 | 3126 |
| ... | | | | | |

- Suggest a regression model that will assist SNAP administrators in providing a monthly benefit to eligible households.
- Fit the model that you suggested in part **a**. Is this model useful in predicting monthly benefits? Justify your answer.
- Are all independent variables in the model helpful in explaining the variation in monthly benefits? Explain your answer.
- Give a 95% confidence interval for average monthly benefits for a four-member household with a gross monthly income of \$2500. Interpret this interval.
- Provide a 99% prediction interval for a four-member household with a gross monthly income of \$2500. Interpret this interval.
- What is the difference between the intervals found in parts **d**. and **e**.?

Note

To complete parts **d**. and **e**., you will need to use Minitab, R, or Rguroo.

Additional Exercises

1. An experimenter often uses a randomized block design to reduce variation by comparing the treatments in homogeneous groups of experimental units called blocks. In many cases, differences in treatments are more likely to be detected with such a design than if the blocking factor were ignored. In each of the following situations, give an example of how one would run a randomized block design to make the comparison in each case. Be aware that there may be more than one correct answer in each example.
 - a. Three different methods of teaching science are to be analyzed by comparing final exam scores from classes taught by the different methods. Assume that the same final exam is given to each class.
 - b. Five hypertensive treatments are to be compared based on their ability to reduce systolic blood pressure. It is felt that the performance of the drugs is affected by the weight of the participant in the study.
 - c. It is desired to compare three different ethnic groups on their knowledge of American history. Each person in the study will be given a 50-question multiple choice test to determine their overall knowledge of the subject.

2. A pharmacist is interested in studying the rate at which three different sinus headache drugs are absorbed into the bloodstream. She randomly selects 12 people, and then randomly assigns four people to try each drug. She administers the drug to each participant and measures the time it takes for the drug to be absorbed into the patient's bloodstream (in minutes). The results of the study are as follows.

| Drug Absorption Time (Minutes) | | |
|--------------------------------|--------|--------|
| Drug 1 | Drug 2 | Drug 3 |
| 5 | 10 | 6 |
| 4 | 11 | 7 |
| 6 | 9 | 5 |
| 3 | 8 | 5 |

- a. Can the pharmacist conclude at $\alpha = 0.01$ that there is a significant difference among the average times required for absorption into the bloodstream for the three drugs?
 - b. What assumptions did the pharmacist make in performing the test procedure in part a.? Does the data appear to satisfy these assumptions? Explain.
 - c. Describe an alternate design that the pharmacist could have used for the above analysis. What are the advantages and disadvantages of this design?

3. An FDA representative is interested in knowing if there is a difference in the average fat contents of three different brands of margarine. The representative randomly selects six samples of each of the brands of margarine and measures the average fat contents per serving. The results of the study are displayed in the following table.

| Fat Content per Serving (Grams) | | |
|---------------------------------|--------------|--------------|
| Margarine #1 | Margarine #2 | Margarine #3 |
| 6 | 5 | 9 |
| 7 | 6 | 8 |
| 6 | 5 | 7 |
| 8 | 4 | 8 |
| 6 | 6 | 9 |
| 8 | 5 | 7 |

- Does the data indicate a difference among average fat contents per serving for the three brands of margarine at $\alpha = 0.01$?
 - What assumptions were made for the test in part **a.**? Does the data appear to satisfy these assumptions? Explain.
 - Why wouldn't a randomized block design be appropriate for this experiment?
4. Psychological reactance may be viewed as the motivational state resulting when someone's freedom is threatened or eliminated. A study relating psychological reactance to one's age was reported in "Psychological Reactance: Effects of Age and Gender" in the *Journal of Social Psychology*.⁶ In order to determine the degree of psychological reactance, participants were asked to fill out a questionnaire which was then scored. The higher the score, the more acute the degree of psychological reactance. The means, standard deviations, and group sizes (for different age groups) are given in the following table.

| Psychological Reactance | | | |
|-------------------------|------|--------------------|------------|
| Age Group | Mean | Standard Deviation | Group Size |
| 18–24 | 3.36 | 0.60 | 1011 |
| 24–29 | 3.28 | 0.65 | 321 |
| 30–40 | 3.16 | 0.64 | 385 |

Although the summary statistics were given in the article, the actual data values upon which the statistics were based were not listed. This is standard procedure in many scientific journals.

- Compute the sums of squares and their degrees of freedom for treatments and error based upon the statistics given in the table.
 - Compute MST and MSE.
 - With $\alpha = 0.01$, can we conclude that there is a significant difference among the degrees of psychological reactance for the different age groups?
 - What assumptions are necessary for performing the test in part **c.**? Can they be checked in this instance?
5. Interviews of fans following an Australian Football League game were summarized in the article "On Being a Sore Loser: How Fans React to Their Team's Failure" in the *Australian Journal of Psychology*.⁷ The study divided the fans interviewed into losers (those who supported the losing team), winners (those who supported the winning team), and non-partisans (those who were indifferent to the outcome of the game). Each fan was asked several questions, all dealing with the fan's perceptions of the game. The purpose of the study was to see if the groups differed

on their responses to any of the questions. One question asked the fans to rate the umpire's performance on a five-point scale from *very bad* (1) to *very good* (5). The mean responses and group sizes associated with this question are given in the following table.

| Umpire Performance | | |
|--------------------|------|------------|
| Group | Mean | Group Size |
| Losers | 2.8 | 49 |
| Winners | 3.7 | 35 |
| Non-Partisans | 3.5 | 57 |

- Compute the grand mean and the sum of squares for treatments (SST).
 - Compute the F -statistic for testing for equality of the group means. The sum of squares for error (SSE) was given in the article as 39.3.
 - With $\alpha = 0.01$, can we conclude that there is a significant difference among the groups in the perception of the umpire?
 - What assumptions are necessary for performing the test in part **c.**? Can they be checked in this instance?
6. Consider the following partially completed ANOVA table for a 3×4 factorial experiment with two replications.

| ANOVA | | | | |
|---------------------|--------|----|----|---|
| Source of Variation | SS | df | MS | F |
| Factor A | 0.800 | 2 | | |
| Factor B | 5.300 | 3 | | |
| Interaction | 9.600 | | | |
| Within | | | | |
| Total | 17.000 | 26 | | |

- Complete the ANOVA table.
 - At the 0.05 level, is there evidence of significant interaction between A and B? Justify your answer.
 - At the 0.05 level, is there evidence of a significant Factor A effect? Justify your answer.
 - At the 0.05 level, is there evidence of a significant Factor B effect? Justify your answer.
 - Does the result of the test for interaction suggest further investigation? Justify your answer.
7. In an experiment to determine the best method by which to assess college students, a group of students were exposed to one of three types of tests. The three methods were: all multiple choice questions, all free-response questions, and mixed questions (a mixture of multiple choice and free-response questions). The scores were recorded for each test taken. Fifteen students were used in the study and were grouped by class level (freshman, sophomore, junior, senior, and graduate). The following table contains the results of the experiment.

| Test Scores | | | |
|-------------|-----------------|---------------|-------|
| | Testing Methods | | |
| Class Level | Multiple Choice | Free-Response | Mixed |
| Freshman | 78 | 84 | 90 |
| Sophomore | 82 | 90 | 95 |
| Junior | 90 | 94 | 98 |
| Senior | 88 | 96 | 100 |
| Graduate | 95 | 98 | 99 |

- Identify the dependent variable, the treatment variable, and the blocking variable.
 - Graphically plot the test scores by class level and testing method. Discuss the graph.
 - Perform an analysis of the data using the class-level blocks. Are blocking effects significant at the 0.05 level of significance? Explain.
 - Is the experiment useful having been analyzed as a randomized block design? Explain.
8. A randomized block design yielded the following ANOVA table.

| ANOVA | | | | |
|---------------------|---------|----|-------|-------|
| Source of Variation | SS | df | MS | F |
| Treatment | 500.000 | 5 | 100 | 7.502 |
| Block | 230.000 | 3 | 76.67 | 5.752 |
| Error | 120.000 | 9 | 13.33 | |
| Total | 850.000 | 17 | | |

- How many blocks are used in the experiment?
 - How many treatments are used in the experiment?
 - How many observations are used in the experiment?
 - What are the null and alternative hypotheses to test if there is a difference among the treatment means?
 - What test statistic should be used to conduct the test in part **d.**?
 - What is the rejection region for the test in parts **d.** and **e.**?
 - Carry out the test and state your conclusion based on a significance level of 0.05.
9. JAS & Associates, a commercial developer, usually gets three cost estimates for many of the jobs for their building projects. Even though one contractor normally works on each potential job, it is in the best interest of the company to get additional estimates and compare them for consistency, no matter who gets the job. To check the consistency of the estimates, several projects are selected and three contractors are asked to submit estimates. The estimates (in thousands of dollars) for the 10 jobs are given in the following table.

| Contractor Cost Estimates (Thousands of Dollars) | | | |
|--|--------------|--------------|--------------|
| Job | Contractor A | Contractor B | Contractor C |
| 1 | 27 | 26 | 28 |
| 2 | 20 | 18 | 22 |
| 3 | 14 | 13 | 17 |
| 4 | 18 | 21 | 20 |
| 5 | 23 | 20 | 22 |
| 6 | 19 | 17 | 19 |
| 7 | 12 | 14 | 15 |
| 8 | 10 | 12 | 13 |
| 9 | 16 | 20 | 19 |
| 10 | 40 | 42 | 47 |

- Perform the appropriate analysis on the data given and generate the ANOVA table for the analysis.
 - Does the data provide sufficient evidence to conclude that there is a difference among the cost estimates supplied by the contractors? Use a significance level of 0.05 to make your decision.
 - What is the P -value for the test performed in part **a.**? Interpret this value.
10. The following table is a 3×3 factorial design with three observations for each factor level.

| Factorial Design Data | | | |
|-----------------------|----------|----|----|
| | Factor B | | |
| Factor A | 1 | 2 | 3 |
| 1 | 30 | 47 | 36 |
| | 30 | 42 | 37 |
| | 30 | 42 | 38 |
| 2 | 12 | 27 | 35 |
| | 14 | 24 | 31 |
| | 15 | 22 | 33 |
| 3 | 10 | 34 | 24 |
| | 13 | 31 | 20 |
| | 12 | 31 | 22 |

- Make an interaction plot of the treatment means using Factor A along the x -axis and each level of Factor B as a different plotting symbol (see Figure 15.4.1). Do the means appear to be different? Does interaction between factors A and B appear to be present? Justify your answers.
- Perform the analysis using a software package, generating the ANOVA table.
- Test for significant interaction using a 0.05 level of significance. Discuss your findings.
- Test for A and B effects using a 0.05 level of significance. Discuss your findings.

11. Tech SportsPlex (TSP) is conducting a study to determine the effectiveness of three types of marketing/advertising methods: e-coupons, newspaper ads, and price discounts. Three counties (believed to be of equal size and close driving distance to TSP) were selected for the marketing campaign. Each strategy was used for a three-month period. It is known that the sales would be seasonal (i.e., TSP's management expects less activity during the summer months). The revenue data (in thousands of dollars) from the study is given in the following table.

| Quarter | e-Coupons | Newspaper Ads | Price Discounts |
|---------|-----------|---------------|-----------------|
| 1 | 48 | 42 | 37 |
| 2 | 25 | 18 | 21 |
| 3 | 20 | 15 | 18 |
| 4 | 40 | 30 | 24 |

- Identify the dependent variable, the treatment variable, and the blocking variable.
 - Specify the null and alternative hypotheses to determine if there is a significant difference among average revenues for the three advertising strategies.
 - Generate the ANOVA table to test the hypotheses in **b**.
 - Conduct the test in **c**. using a significance level of 0.05.
 - Was the variation among the observed revenues significantly reduced by blocking? Explain using $\alpha = 0.05$.
12. In "Death Anxiety in Malaysian and Australian University Students" (The Journal of Social Psychology, 128(1)),⁸ the results of an experiment measuring death anxiety was reported. In the study, a total of 153 students were administered the Templer Death Anxiety Scale, a questionnaire which assesses the extent to which the respondent is preoccupied with issues surrounding death. Scores ranged from 0 to 15, with higher scores indicating greater death anxiety. The students were classified into one of three ethnic groups: Chinese, Indian, or Australian. The means, standard deviations (s.d.) and group sizes are given in the following table.

| Group | mean | s.d. | size |
|------------|------|------|------|
| Chinese | 6.0 | 3.1 | 62 |
| Indian | 4.8 | 3.1 | 15 |
| Australian | 6.5 | 2.7 | 76 |

Although, the above summary statistics were given in the article, the actual data values upon which the statistics were based were not listed. This is standard procedure in many scientific journals.

- Compute the sums of squares and their degrees of freedom for treatments and error based upon the statistics given in the above table. Compute MST and MSE.
- With $\alpha = 0.05$, can we conclude that there is a difference in the degree of death anxiety for the 3 groups?
- What assumptions are necessary for performing the test in part **b**? Can they be checked in this instance?

Additional Exercises

1. A sales manager for an insurance company believes that customers have the following preferences for life insurance products: 50% prefer Whole Life, 25% prefer Universal Life, and 25% prefer Life Annuities. A survey of 250 customers produced the following results.

| Life Insurance Preferences | |
|----------------------------|--------|
| Product | Number |
| Whole Life | 60 |
| Universal Life | 100 |
| Life Annuities | 90 |

- a. Is the sales manager's claim refuted by the data at $\alpha = 0.05$?
- b. What assumptions were made in the test for part a.?
2. Do you think there are people somewhat like us living on other planets in the universe? The responses to this question, which was asked in several different calendar years, were summarized in the *Gallup Poll Monthly*. For the year 1978, 51% answered "Yes," 33% "No," and the remainder had no opinion.² Suppose that a sample of 100 people is chosen in 2018 in order to determine if opinions have changed concerning extraterrestrial life. Assume that the same question is asked and that 42 answer "Yes," 30 "No," and that the rest have no opinion. Assuming that the percentages given above accurately represent the attitudes of the people in 1978, can we conclude with $\alpha = 0.05$ that people's attitudes toward extraterrestrial life have changed since 1978?
3. A traffic engineer feels that on a certain four-lane highway, the probability of being in the innermost lane is twice as great as any of the other lanes. Assume the other lanes have equal probabilities. A random sample of 200 motorists is chosen and the lanes in which they are traveling in are noted. The results (Lane 1 is the innermost lane) are given in the following table.

| Traffic Lanes | | | | |
|---------------|----|----|----|----|
| Lane | 1 | 2 | 3 | 4 |
| Frequency | 55 | 45 | 62 | 38 |

- a. Find the probabilities implied by the engineer's claim that a randomly chosen motorist will be in each of the four lanes.
- b. With $\alpha = 0.05$, can we refute the claim of the traffic engineer?
4. The National Restaurant Association is interested in determining if there is a relationship between the type of pizza Americans prefer and the region of the country in which they live. The association randomly selects 285 Americans and records the category of pizza which best describes their preference and the region of the country in which they live with the following results.

| Pizza Preference | | | | |
|-------------------------|--------|-------|------|------|
| Type of Pizza Preferred | Region | | | |
| | North | South | East | West |
| Thin Crust | 40 | 30 | 35 | 45 |
| Thick Crust | 17 | 15 | 21 | 22 |
| Pan Pizza | 15 | 15 | 15 | 15 |

- a. Can the association conclude that the type of pizza Americans prefer and the region of the country in which they live are dependent at $\alpha = 0.10$?
 - b. What assumptions were made in the test for part a.?
5. Amazon asked consumers to assign star ratings to different brands of staplers. The results are summarized in the table below.

| Amazon Stapler Ratings | | | | | |
|--------------------------------|-------------|--------|--------|--------|--------|
| Brand | Star Rating | | | | |
| | 5 Star | 4 Star | 3 Star | 2 Star | 1 Star |
| Swingline Optima 40 Compact | 589 | 60 | 22 | 15 | 60 |
| PaperPro inPower | 485 | 60 | 40 | 20 | 60 |
| Rapid Classic K1 Plier Stapler | 312 | 26 | 11 | 7 | 15 |
| EcoElectronix Electric Stapler | 361 | 45 | 14 | 5 | 27 |
| Bostitch Ascend 3 in 1 | 378 | 66 | 44 | 22 | 38 |

- a. Can it be concluded that star rating and stapler brand are dependent at $\alpha = 0.10$?
- b. What assumptions were made in the test for part a.?

Key Formulas

Section

17.6 Test Statistic for the Kruskal-Wallis Test

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

$$\text{where } R_i = \sum_{j=1}^{n_i} r_{ij} \text{ and } N = \sum_{i=1}^k n_i.$$

Additional Exercises

1. A new method for temporarily relieving the lung congestion of cystic fibrosis patients has been introduced. The traditional method of relieving the congestion involves a series of manual techniques where the chest and back area are pounded and massaged. The new method is a mechanical vest which has been designed to perform the manual techniques. A study is conducted to measure the effectiveness of the new vest. Five cystic fibrosis patients are randomly selected and the diameter of the blood vessels in their lungs is measured after using the traditional treatment and after using the vest treatment. The larger the diameter of the blood vessels within the lungs, the better the treatment. If the study provides conclusive evidence that the vest is more effective than the manual method in increasing the diameter of the blood vessels, the hospital will recommend the vest to its patients because the vest allows the patients to be much more independent. The results of the study are as follows.

| Diameter of Lung Blood Vessels (in mm) | | | | | |
|--|-----|-----|-----|-----|-----|
| Subject | 1 | 2 | 3 | 4 | 5 |
| After Traditional Treatment | 0.5 | 0.4 | 0.7 | 0.6 | 0.2 |
| After Vest Treatment | 0.6 | 0.6 | 0.7 | 0.7 | 0.5 |

- a. What assumption must be made in order to perform the test of hypothesis using the paired difference t -test?
- b. Using the Sign Test, does the data provide conclusive evidence that the median diameter of blood vessels in the lungs is significantly larger after using the vest treatment than after using the traditional treatment at $\alpha = 0.01$?
- c. What assumptions were made in performing the Sign Test?
- d. Using the Signed-Rank Test, does the data provide conclusive evidence that the median diameter of blood vessels in the lungs is significantly larger after using the vest treatment than after using the traditional treatment at $\alpha = 0.01$?
- e. What assumptions were made in performing the Signed-Rank Test?
- f. Which test do you think produces more accurate results? Why?
- g. Perform a paired difference t -test. How do the results of the Sign Test and the Signed-Rank Test compare with the results of the t -test?

2. As private companies prepare to go public, many analysts attempt to predict whether the stock will have a positive or negative return in the first day of trading. One particular analyst believes the return after the first day of trading is positive, on average. Another analyst wishes to test this analyst's claim using 15 recently offered public stocks.²

| Initial Public Offerings | | |
|---|--------------------------|---|
| Company | Initial Offer Price (\$) | Price After 1 st Day of Trading (\$) |
| Enduro Royalty Trust (NDRO) | 22.00 | 21.26 |
| ZELTIQ Aesthetics (ZLTQ) | 13.00 | 15.50 |
| Ubiquiti Networks (UBNT) | 15.00 | 17.50 |
| Tudou Holdings Limited (TUDO) | 29.00 | 25.56 |
| Carbonite (CARB) | 10.00 | 12.35 |
| SandRidge Permian Trust (PER) | 18.00 | 18.00 |
| American Capital Mortgage Investment (MTGE) | 20.00 | 18.41 |
| C&J Energy Services (CJES) | 29.00 | 30.50 |
| Chefs Warehouse Holdings (CHEF) | 15.00 | 17.50 |
| Spirit Airlines (SAVE) | 12.00 | 11.55 |
| Pandora Media (P) | 16.00 | 17.42 |
| Wesco Aircraft Holdings (WAIR) | 15.00 | 14.92 |
| American Midstream Partners (AMID) | 21.00 | 20.95 |
| Dunkin Brands Group (DNKN) | 19.00 | 27.85 |
| Skullcandy (SKUL) | 20.00 | 20.00 |

- Use the Sign Test to test the hypothesis that the return after the first day of trading is positive. Test at the 0.05 level.
 - Suppose that an analyst claimed that the median return on a stock in the first day of trading is +\$1.00. Perform a test of hypothesis to determine if the median return is different than what the analyst claims. Use $\alpha = 0.05$.
 - Use the Wilcoxon Signed-Rank Test to determine if the price after the first day of trading is generally greater than the initial offer price. Test at $\alpha = 0.05$.
 - Analyze the results of the Sign Test performed in part **a.** and the Signed-Rank Test performed in part **c.** in terms of the problem. Which test do you think yields more accurate results?
 - What concerns do you have with the tests performed in this problem?
3. A weight loss center is trying to determine which of its diets results in higher client satisfaction. The center polled 20 clients (10 were on Diet A and 10 on Diet B) and had them rate their satisfaction in the diets from 1 to 100.

| Diet Ratings | | | | | | | | | | |
|--------------|----|----|----|----|----|-----|----|----|----|----|
| Diet A | 84 | 77 | 89 | 98 | 97 | 100 | 75 | 85 | 96 | 78 |
| Diet B | 94 | 81 | 95 | 93 | 97 | 99 | 82 | 92 | 95 | 89 |

- Which nonparametric test do you think is most appropriate to test the claim that Diet B results in higher client satisfaction than Diet A? Explain why.

- b. Write the null and alternative hypotheses for a Rank-Sum Test to determine if Diet B results in higher client satisfaction than Diet A.
 - c. Using the Rank-Sum Test, does the data provide sufficient evidence at $\alpha = 0.05$ that Diet B results in greater client satisfaction than Diet A?
 - d. What assumptions were made in the test performed in part c.?
4. Are students with higher GPAs more likely to get a higher paying job upon graduation? Consider the following data regarding student GPA and starting salary.

| GPA and Starting Salary | |
|-------------------------|----------------------|
| GPA | Starting Salary (\$) |
| 2.37 | 47,000 |
| 3.20 | 48,000 |
| 3.21 | 52,000 |
| 3.39 | 50,000 |
| 3.55 | 54,000 |
| 3.57 | 58,000 |
| 3.76 | 60,000 |
| 3.77 | 70,000 |
| 3.79 | 69,000 |
| 3.90 | 65,000 |

- a. Determine the ranks for the x -variable, GPA.
 - b. Determine the ranks for the y -variable, Starting Salary.
 - c. Calculate the Spearman rank correlation coefficient.
 - d. Interpret the value of the coefficient. Is the relationship between these two variables positive or negative? Is this relationship what you expected? Explain.
 - e. Comment on the strength of the relationship between these two variables.
 - f. Is there evidence at $\alpha = 0.10$ that these two variables are related?
5. *Fortune* magazine releases a list of the world's most admired companies. In the survey they ask business people to vote for companies that they admire most from various industries. The table below lists the top 10 most admired companies of 2011 along with each company's Fortune 500 ranking. Note that the Fortune 500 ranks companies based on revenues.³

| Top 10 Most Admired Companies and Profits | | |
|---|----------------------|---------------------|
| Company | Most Admired Ranking | Fortune 500 Ranking |
| Apple | 1 | 35 |
| Google | 2 | 92 |
| Berkshire Hathaway | 3 | 7 |
| Southwest Airlines | 4 | 205 |
| Procter & Gamble | 5 | 26 |

| Top 10 Most Admired Companies and Profits | | |
|---|----------------------|---------------------|
| Company | Most Admired Ranking | Fortune 500 Ranking |
| Coca-Cola | 6 | 70 |
| Amazon.com | 7 | 78 |
| FedEx | 8 | 73 |
| Microsoft | 9 | 38 |
| McDonald's | 10 | 111 |

- a. Compute the Spearman rank correlation coefficient. Interpret this value.
 - b. With 95% confidence, can we conclude that there is an association between company admiration and revenue?
6. The given table shows key dates of the Dow-Jones Industrial Average. Apply the runs test to check for randomness at the 0.05 level. (**Hint:** First find the median of the values, then label each value by A if it is above the median and B if it is below the median.)

| Significant Levels on the Dow (December 1974 to February 2009) | | |
|---|------------------------------|--|
| Date | Dow Jones Industrial Average | Significance |
| December 6, 1974 | 577 | The last Bear Market bottom |
| July 12, 1976 | 1011 | Highest point between January, 1973 and October, 1982 |
| August 12, 1982 | 776 | The start of the "Reagan Bull" |
| August 25, 1987 | 2722 | The 1987 high |
| October 19, 1987 | 1738 | The (508 point) crash of 1987 |
| February 2, 1994 | 3975 | The top of the post 1987 crash recovery |
| November 23, 1994 | 3674 | The start of the Clinton "super bull" |
| March 29, 1999 | 10,006 | The first Dow close above 10,000 |
| January 14, 2000 | 11,723 | The "Clinton bull" high |
| March 17, 2000 | 10,630 | The biggest one day gain (499 points) |
| March 20, 2001 | 9720 | Dow closes below previous year low for the first time since 1982 |
| September 11-14, 2001 | 9605 | Terrorist attack closed the Dow for four days |
| September 17, 2001 | 8920 | The biggest one day fall (685 points) |
| September 21, 2001 | 8235 | The Dow's second worst week ever (-14.26%) |
| December 31, 2001 | 10,021 | Dow up 21.7% from September 21 low but down 7.2% on the year |
| September 30, 2002 | 7591 | New 2002 low – all treasury yields (except 30-year bond) at 2002 lows |
| October 9, 2002 | 7286 | New 2002 low – Dow down 37.8% from the January, 2000 all-time high |
| October 31, 2002 | 8397 | Dow up 806 points (10.6%) for October – first positive month since March |
| November 6, 2002 | 8771 | Federal reserve cuts rates for the first time since December, 2001 |

| Significant Levels on the Dow (December 1974 to February 2009) | | |
|---|------------------------------|---|
| Date | Dow Jones Industrial Average | Significance |
| December 31, 2002 | 8341 | Dow down 16.8% for 2002 – first three consecutive year loss since 1939-41 |
| May 23, 2003 | 8601 | Senate passes bill raising the treasury debt limit |
| June 25, 2003 | 9011 | Federal reserve cuts rates by 0.25% |
| December 31, 2003 | 10,453 | Dow up 25.32% in 2003 |
| October 3, 2006 | 11,727 | Dow exceeds the previous all-time high in January, 2000 |
| October 9, 2007 | 14,164 | New all-time high on the Dow |
| July 2, 2008 | 11,215 | Dow closes more than 20% below the October, 2007 high |
| February 27, 2009 | 7062 | Dow closes more than 50% below the October, 2007 high |

7. The irrational number π can be approximated by the rational number $\frac{22}{7}$. Test the randomness of odd and even digits in $\frac{22}{7}$ at the 0.05 level using the first nine digits.
8. A polling agency conducts exit interviews after an election. If R = Republican and D = Democrat, the first 20 voter responses in a random sample are as follows.
- D D D R D R R D D D D R D D R R R D R D
- Test for non-randomness using $\alpha = 0.05$.
9. Consider the following *U.S. News and World Report* college rankings for schools in the Big Ten, the Big 12, and the Atlantic Coast conferences.⁴

| College Rankings by Conference | | |
|--------------------------------|---------|--------|
| Atlantic Coast | Big Ten | Big 12 |
| 29 | 55 | 101 |
| 10 | 62 | 45 |
| 101 | 42 | 58 |
| 71 | 28 | 143 |
| 68 | 45 | 90 |
| 31 | 71 | 94 |
| 55 | 45 | 101 |
| 25 | 12 | 75 |
| 38 | 68 | 101 |
| 101 | 71 | 132 |
| 36 | 75 | 160 |
| 25 | | 97 |

With $\alpha = 0.10$, use the Kruskal-Wallis Test to determine if there is a significant difference in *U.S. News and World Report* rankings among the three athletic conferences.

10. Consider the following scores reported by *Condé Nast Traveler* in their annual list of the top cities to visit around the world. The results were determined from more than 8 million votes cast in the Readers' Choice Awards survey.^{5,6,7}

| Readers' Choice City Rankings, 2011 | | | | | |
|-------------------------------------|------|-----------|------|-------------------|------|
| Asia | | Europe | | United States | |
| Kyoto | 82.3 | Florence | 85.0 | Charleston, SC | 84.7 |
| Bangkok | 81.6 | Barcelona | 82.8 | San Francisco, CA | 83.7 |
| Hong Kong | 81.1 | Rome | 82.4 | Santa Fe, NM | 83.0 |
| Chiang Mai | 80.8 | Paris | 81.9 | Chicago, IL | 82.2 |
| Ubud | 80.0 | Bruges | 81.7 | Honolulu, HI | 80.9 |
| Singapore | 78.4 | Venice | 81.7 | New York, NY | 80.8 |
| Tokyo | 76.8 | Salzburg | 81.4 | Savannah, GA | 79.1 |
| Luang Prabang | 76.4 | Vienna | 81.0 | Carmel, CA | 78.5 |
| Thimphu | 75.1 | Prague | 79.7 | Seattle, WA | 78.4 |
| Shanghai | 74.9 | Siena | 79.7 | Boston, MA | 78.0 |

- Using the Kruskal-Wallis Test and $\alpha = 0.05$, test to determine if there is a difference in rankings between Asia, Europe, and the United States.
- What assumptions were made for the test performed in part a.?