AE Additional Exercises

Exercises

1. Jakob Nielsen, a website consultant, recently conducted an eye-tracking survey to determine whether people ignore purely decorative images when viewing web pages. An aspect of the study compared a set of products on Pottery Barn's furniture website and a page of televisions on Amazon.com. The study found that consumers tended to ignore the televisions on Amazon.com because they were generic, making the product image less inviting. When consumers viewed Pottery Barn's website, they were more inclined to view the photos of the bookcases for longer periods of time because they were images of the actual products for sale.

Source: New York Times

- **a.** Identify the population of interest.
- **b.** What characteristic of the population is being measured?
- **c.** Is the purpose of the data collection to perform descriptive or inferential statistics?

2. A recent study at Britain's Oxford and Exeter Universities explored whether a woman's diet before conception affects the gender of her child. Researchers studied the eating habits of 740 women during their first-time pregnancies and found that higher caloric intake prior to conception can significantly increase the chances of having a son while more restricted diets are more likely to produce daughters. They found that high potassium diets (eating bananas) and calcium rich diets (cereal and milk) were associated with having a baby boy. Researchers concluded that eating a bowl of cereal for breakfast can increase the chances of a male birth. "Of women eating cereals daily, 59 percent had boys, compared with only 43 percent who bore boys in the group eating less than a bowlful per week."

Source: CNNHealth.com

- **a.** Identify the population of interest.
- **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?
- **e.** What are some problems that could be associated with collecting data in this study?

3. Researchers at The Ohio State University and Zeppelin University Friedrichshafen, in Germany, recently conducted a study regarding the elderly and negative news coverage. The researchers presented 276 subjects with several stories (with photos) about either old or young people. Participants were presented with one of two versions of each story. In one version the main character was painted in a positive light and in the other the same character was described negatively. After the participants finished reading their self-esteem was measured. The study found that older readers were more inclined to read the negative stories about youth. In addition, they found that the more negative stories older people read about younger individuals, the higher their self-esteem tended to be. This could explain the prominence of negative media coverage on networks with an older audience such as Fox News and MSNBC.

Source: Huffpost Media

- **a.** Identify the population of interest.
- **b.** What characteristics of the population are being measured?
- **c.** Identify the sample.
- **d.** What are some problems that could be associated with this study?
- **e.** Is the purpose of this study to perform descriptive or inferential statistics?

4. Researchers at Pepperdine University's Graziadio School of Business and Management recently conducted a study to determine whether the capital crunch is affecting small businesses' abilities to expand. The companies studied included alternative lenders, venture capitalists, and private equity firms, among others. The project surveyed 559 privately held businesses and 1430 lenders and investors nationwide. The study found that 78 percent of businesses had solid growth strategies but only 40 percent had access to the resources needed to grow. Lenders and investors rejected 90 percent of loan applications or investment proposals that would be secured by a business's real estate holdings and 73 percent of loan applications or investment proposals that are based on a business's cash flow. According to the survey author John Paglia, "The study shows private business owners feel they are being constrained by access to financial capital. Owners currently expect a 10 percent revenue growth over the next 12 months. If they were to receive additional capital, they estimate their revenue growth rate to jump to 25 percent."

Source: smallbiztrends.com

- **a.** Identify the population of interest.
- **b.** What characteristics of the population are being measured?
- **c.** Identify the sample.
- **d.** Is the purpose of this study to perform descriptive or inferential statistics?

5. Ruder Finn, one of the world's largest independent public relations agencies, recently announced the *Mobile Intent Index* which studies mobile phone user habits and explores the underlying reasons that people have for accessing the internet on mobile devices. The *Mobile Intent Index* asked 500 American adults 18 years of age and older how often they use their mobile phones to access the internet for 295 reasons. The study found that 91% of mobile phone users go online to socialize compared to 79% of desktop users. In addition, 60% of mobile internet users go online to manage finances compared to only 45% of desktop users. They also found that mobile users were less likely to use the internet for educational purposes, only 42% compared to 92% of desktop users. Finally, unsurprisingly, mobile phones are not used for creative purposes; only 42% of mobile users personally express themselves online compared to 54% of desktop users.

Source: PR Newswire

- **a.** Identify the population of interest.
- **b.** What characteristic of the population is being measured?
- **c.** Identify the sample.
- **d.** What are some problems that could be associated with this study?
- **e.** Is the purpose of this study to perform descriptive or inferential statistics?

6. A personnel director is interested in determining how effective a new reading course will be in improving the reading comprehension of her company's employees. The director randomly selects twenty employees and determines the average reading comprehension both before and after instruction in the reading course.

- **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?

7. In 2009 and 2010 the market research firm Chadwick Martin Bailey conducted several studies to provide insights into recent dating behavior in the United States. The data were collected through research via an online Consumer Research Panel. In the "Marriage Survey" (a survey of recently married people), 7000 adults 18 years of age and older who were married in the past 5 years were polled. The study found that 17% of couples married within the last 3 years met each other on an online dating site. This is compared to 26% that met their significant other through a friend or family member, 36% who met through work or school, 4% who met through a church or place of worship, 11% who met through bars, clubs, or social events, and 7% who met in some other manner.

Source: Chadwick Martin Bailey/Match.com

- **a.** Identify the population of interest.
- **b.** What characteristics of the population are being measured?
- **c.** Identify the sample.
- **d.** What are some problems that could be associated with this study?

8. In a recent 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?

9. 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 continental 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?

10. 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?

AE Additional Exercises

1. Suppose you were 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:
   - **a.** Student learning
   - **b.** School discipline
   - **c.** Teacher preparation
   - **d.** Absenteeism (pupil and teacher)
   - **e.** Cafeteria food quality
2. 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.
   - **a.** What variables would you measure to explore this problem?
   - **b.** How would you collect measurements on these variables?
3. A family member has unexpectedly bequeathed you a sizable sum of money.
   - **a.** What criteria might you wish to evaluate in deciding how to invest the money?
   - **b.** What data might be useful in your considerations?
4. 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 they both be correct? Explain.
5. 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.
6. At the end of 2001, the United States had 32.9 million people living in poverty according to the Census Bureau (www.census.gov). 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?
7. 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?
8. Identify the following variables as discrete or continuous.
   - **a.** Average test score on a test ranging from 0 to 100
   - **b.** Number of boot errors on a computer
   - **c.** Investment ratios for 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. 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.

Source: U.S. News and World Report

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.

e. Why is some method of data summary necessary here?

11. Consider the world production of crude oil given in millions of barrels per day.

World Production of Crude Oil

Year	Total World Production (Millions of Barrels per Day)	Year	Total World Production (Millions of Barrels per Day)
1980	63.987	1995	70.274
1981	60.602	1996	71.919
1982	58.098	1997	74.160
1983	57.934	1998	75.656
1984	59.568	1999	74.853
1985	59.172	2000	77.768
1986	61.407	2001	77.686
1987	62.086	2002	76.994
1988	64.380	2003	79.598
1989	65.508	2004	83.105
1990	66.426	2005	84.595
1991	66.399	2006	84.661
1992	66.564	2007	84.543
1993	67.091	2008	85.507
1994	68.590	2009	84.389

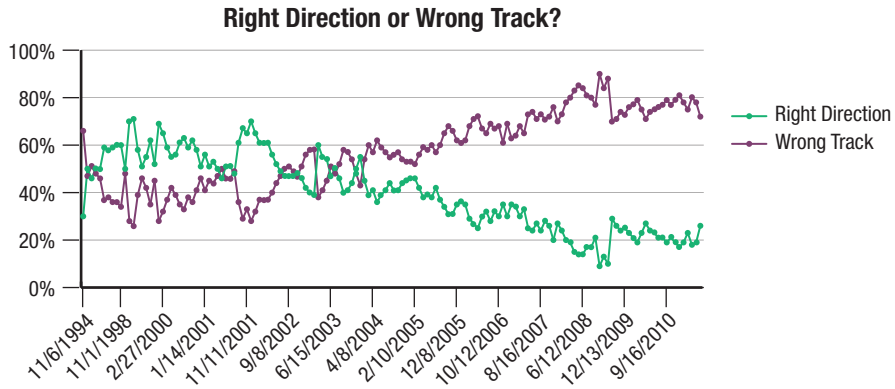
Source: Energy Information Administration

a. What is the level of measurement of the data?

b. Are the data time series or cross-sectional? If the data are 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 economy is going in the right direction. The data were 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?*

Source: Gallup Poll



a. Are the opinions on the outlook of the economy presented in time series or cross-sectional data? Justify your answer.

b. If the data are 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 1934 to 2005. If the data are 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.

Fertility Rates

Year	Fertility Rate	Year	Fertility Rate	Year	Fertility Rate
1934	2.294	1958	3.693	1982	1.829
1935	2.250	1959	3.705	1983	1.803
1936	2.207	1960	3.654	1984	1.806
1937	2.236	1961	3.629	1985	1.843
1938	2.288	1962	3.474	1986	1.836
1939	2.238	1963	3.333	1987	1.871
1940	2.301	1964	3.208	1988	1.933
1941	2.399	1965	2.928	1989	1.977
1942	2.628	1966	2.736	1990	2.081
1943	2.718	1967	2.573	1991	2.073
1944	2.568	1968	2.477	1992	2.065
1945	2.491	1969	2.465	1993	2.046
1946	2.943	1970	2.480	1994	2.036
1947	3.274	1971	2.267	1995	2.019
1948	3.109	1972	2.010	1996	2.040
1949	3.110	1973	1.879	1997	2.000
1950	3.091	1974	1.835	1998	2.030
1951	3.267	1975	1.774	1999	2.070
1952	3.355	1976	1.738	2000	2.056
1953	3.418	1977	1.790	2001	2.034
1954	3.537	1978	1.760	2002	2.013
1955	3.574	1979	1.808	2003	2.043
1956	3.682	1980	1.840	2004	2.046
1957	3.760	1981	1.815	2005	2.054

Source: U.S. National Center for Health Statistics

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. Are 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?

AE Additional Exercises

1. Median family income has grown substantially in recent years. The table below contains median household incomes in the United States for the years 1990 through 2009.

Median Family Income

Year	Income ($)	Percentage Change	Year	Income ($)	Percentage Change
1990	29,943	–	2000	41,990	3.2
1991	30,126	0.6	2001	42,228	0.6
1992	30,636	1.7	2002	42,409	0.4
1993	31,241	2.0	2003	43,318	2.1
1994	32,264	3.3	2004	44,334	2.3
1995	34,076	5.6	2005	46,326	4.5
1996	35,492	4.2	2006	48,201	4.0
1997	37,005	4.3	2007	50,233	4.2
1998	38,885	5.1	2008	50,303	0.1
1999	40,696	4.7	2009	49,777	–1.0

Source: U.S. Census Bureau

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 data represent income of households headed by adults 25 years and older, tabulated by educational attainment.

Income and Educational Attainment

Educational Attainment	Under $5000	$5000 – $9999	$10,000 – 14,999	$15,000 – 24,999	$25,000 – 34,999	$35,000 – 49,999	$50,000 – 74,999	$75,000 & Over	Median Income
Elementary	16.3%	26.7%	17.2%	19.7%	9.9%	6.4%	2.9%	0.8%	$11,730
< 8 Years	18.4	27.9	17.3	18.1	8.9	5.9	2.7	0.7	10,884
8 Years	13.7	25.2	17.1	21.7	11.0	7.0	3.1	1.0	12,999
High School	6.8	12.4	12.0	21.8	17.8	17.0	9.5	2.8	23,382
1 – 3 Years	10.9	19.4	14.9	22.2	14.2	11.2	5.6	1.6	16,727
4 Years	5.4	10.0	11.0	21.6	19.0	18.9	10.8	3.2	25,910
College	2.6	4.3	5.8	15.1	16.3	22.7	20.1	13.2	38,337
1 – 3 Years	3.6	6.6	8.1	18.7	18.1	22.6	16.1	6.1	31,865
4 Years +	1.8	2.5	4.1	12.4	14.9	22.8	23.0	18.5	43,952

Source: U.S. Census Bureau

a. What graphical methods would be useful in displaying the data?

b. Use a graphics program to display the data.

c. Discuss any conclusions you made from your graph(s).

3. Where do business school students most want to work? A CNNMoney list ranks companies that MBA students want to work for most after getting their degree. It can be seen that men and women have different desires when it comes to employment after business school. The table below shows seven of the top companies students want to work for and the percentage of students that ranked the particular company in their top 5.

Most Desired Companies

Company	Percentage of Men	Percentage of Women
Google	22.36	21.72
Goldman Sachs	17.76	5.81
Johnson & Johnson	5.68	11.41
Bain & Company	12.89	7.88
Apple Computer	12.80	11.01
Nike	8.06	9.04
J.P. Morgan	12.08	4.44

Source: CNNMoney.com

a. What graphical method do you think would be most useful in summarizing these data? Explain your answer.

b. Graph the data using the method you identified in part **a.**

c. Write a short paragraph describing the data, making conclusions from the graph you constructed in part **b.**

4. The following data give the percentage of people in the U.S. holding more than one job in a one-year period. Graph the data using a method that would contrast the difference between men and women in this situation.

Percentage Holding More Than One Job

Year	Men	Women
1970	7.0%	2.2%
1979	5.9%	3.5%
1989	6.4%	5.9%
1999	5.7%	5.6%
2003	5.2%	5.4%

Source: Bureau of Labor Statistics

5. *Billboard* magazine, in cooperation with Arbitron, produces a national radio format rating. The following data were gathered from radio listeners 12 and older.

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 20 global corporations, ranked by the amount spent on research and development in 2009.

Amount Spent on Research and Development (R&D) in 2009
(Millions of Dollars)

Rank	Company	R&D Spending	Spending as a Percentage of Sales	Headquarters Location	Industry
1	Roche Holding	9120	20.1	Europe	Healthcare
2	Microsoft	9010	15.4	N. America	Software and Internet
3	Nokia	8240	14.4	Europe	Computing and Electronics
4	Toyota	7822	3.8	Japan	Auto
5	Pfizer	7739	15.5	N. America	Healthcare
6	Novartis	7469	16.9	Europe	Healthcare
7	Johnson & Johnson	6986	11.3	N. America	Healthcare
8	Sanofi-Aventis	6391	15.6	Europe	Healthcare
9	GlaxoSmithKline	6187	13.9	Europe	Healthcare
10	Samsung	6002	5.5	S. Korea	Computing and Electronics
11	General Motors	6000	5.7	N. America	Auto
12	IBM	5820	6.1	N. America	Computing and Electronics
13	Intel	5653	16.1	N. America	Computing and Electronics
14	Merck	5613	20.5	N. America	Healthcare
15	Volkswagen	5359	3.7	Europe	Auto
16	Siemens	5285	5.1	Europe	Industrials
17	Cisco Systems	5208	14.4	N. America	Computing and Electronics
18	Panasonic	5143	6.4	Japan	Computing and Electronics
19	Honda	4996	5.4	Japan	Auto
20	Ford	4900	4.1	N. America	Auto

Source: Booz & Company

a. For comparative purposes, which of the two columns reporting R&D spending is more useful, and why?

b. What types of graphs would be useful in presenting these data? Explain your answers.

c. Develop a histogram for the spending as a percentage of sales.

d. Use computer software to develop pie charts for the headquarters location and industry categories of the top 20 global R&D spenders.

8. In New York, a group of women challenged the state's ban on topless sunbathing. The legal issue was whether the ban was discriminatory. During the controversy, the Gallup poll conducted a survey asking the following question: *Do you think women should be permitted to sunbathe topless on public beaches, if they choose to, or do you think topless sunbathing on public beaches should be banned?*

Responses to Survey Question				
	Permitted	Banned	No Opinion	Number of Interviews
National	33%	63%	4%	1001
Gender				
Male	50	45	5	500
Female	18	79	3	501
Age				
18 – 29	47	51	2	219
30 – 49	39	58	3	411
50 – 64	18	76	6	206
65 +	18	77	5	357
Region				
East	39	59	2	247
Midwest	34	62	4	254
South	25	71	4	301
West	38	57	5	199
Community				
Urban	42	55	3	345
Suburban	35	62	3	351
Rural	23	72	5	298
Race				
White	33	64	3	871
Non-white	39	57	4	121
Education				
College Grads	46	48	6	288
Some College	35	62	3	233
No College	28	69	3	475
Sex/Education				
Male/College	56	40	4	238
Male/ No College	45	49	6	238
Female/ College	26	70	4	264
Female/ No College	13	85	2	237

a. Suggest two different types of graphs that might be useful in graphing the data.

b. Create two different graphs using the data.

c. 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 Harris poll results on political identification from 1977 to 2008.

Nation's Political Identification
(Percentage of the Population) **1977 – 2008**

Year	Republican	Democrat	Independent	Year	Republican	Democrat	Independent
1977	21	48	25	1993	29	38	27
1978	22	43	30	1994	32	37	26
1979	22	41	31	1995	31	36	28
1980	24	41	29	1996	30	38	26
1981	28	39	28	1997	29	37	26
1982	26	40	28	1998	28	37	27
1983	26	41	27	1999	29	36	26
1984	27	40	24	2000	29	37	23
1985	30	39	26	2001	31	36	22
1986	30	39	25	2002	31	34	24
1987	29	38	28	2003	28	33	24
1988	31	39	25	2004	31	34	24
1989	33	40	23	2005	30	36	22
1990	33	38	25	2006	27	36	24
1991	32	37	26	2007	26	35	23
1992	30	36	29	2008	26	36	31

Source: Harris Interactive

a. What types of graphs would be useful in visualizing these data? Explain your answer.

b. Construct two different types of graphs from the data.

c. Examine the data and write a short paragraph on your conclusions.

10. Monaco is noted for having one of the highest population densities in the world, approximately 16,923 persons per square kilometer. Usually, dense urban areas have relatively high crime rates. This is not the case in Monaco. The following table gives crime data per 100,000 population for the year 2000 in Monaco as well as in other urban areas.

Crime Data per 100,000 Population					
	Monaco	**London**	**Chicago**	**New York**	**San Francisco**
Homicide	1.0	4.7	23.0	43.4	8.6
Forcible Rape	7.9	34.7	–	19.1	26.7
Robbery	–	625	635.6	352.5	409.9
Aggravated Assault	–	847.4	880.4	473.7	327.2
Burglary	–	938.9	895.4	394.6	764.8
Larceny/Theft	333.0	2675.3	3361.9	1674.2	803.0

Source: U.S. Department of Justice, CIA, BBC

a. What types of graphs would be useful in visualizing this data? Explain your answer.

b. Construct two different types of graphs from the data.

c. Examine the data and write a short paragraph on your conclusions.

AE Additional Exercises

1. A carpenter is attempting to repair a porch and needs twenty boards which are eight feet long. The salesman at the hardware store says he has twenty boards that "average" eight feet long. When the carpenter checks what he has bought, there are ten boards at six feet and ten boards at ten feet. Do you feel the salesman accurately represented the lengths? Discuss.

2. The maximum heart rates achieved while performing a particular aerobic exercise routine are measured (in beats per minute) for 9 randomly selected individuals.

Maximum Heart Rates (Beats per Minute)								
145	155	130	185	170	165	150	160	125

a. Calculate the sample variance of the maximum heart rate achieved.

b. Calculate the sample standard deviation of the maximum heart rate achieved.

c. Calculate the range of the maximum heart rate achieved.

d. What are some of the factors which might contribute to the variation in the observations?

3. A sample of teenagers was asked how many times they went to the movies in the past 3 months. The frequency distribution table summarizes the results.

Teenager Movie Visits	
Number of Visits	Frequency
0	13
1	18
2	11
3	7
4	4
5	3
6	0
7	3
8	3
9	0
10	2

a. What proportion of the sample visited the movies at least 3 times in the previous 3 months?

b. Find the mean and standard deviation of the number of visits 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? What is the reason for the discrepancy, if any?

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. Calculate the average number of defective circuit boards produced by each machine.

b. Calculate the variance of the number of defective circuit boards produced by each machine.

c. Calculate the standard deviation of the number of defective circuit boards produced by each machine.

d. Which machine do you think is better? Why?

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
7	31	29
8	33	missed
9	33	23
10	25	35
11	28	31
12	32	36
Total	**360**	**320**

a. What level of measurement does the data possess?

b. What statistical criteria might you use to select the better player? Justify your answer.

c. Calculate the statistics you proposed in **b.**

d. Which player is more consistent? Why?

7. Consider the literacy data given in the following table.

Literacy Rates

Country	Literacy Rate (%)	Country	Literacy Rate (%)
Australia	99.0	Luxembourg	99.0
Bolivia	90.7	Mexico	92.8
Canada	99.0	Netherlands	99.0
Denmark	99.0	Peru	89.6
France	99.0	Saudi Arabia	85.0
India	74.0	United States of America	99.0
Kenya	73.0	Zimbabwe	91.2

Source: United Nations Development Programme Report, 2009

a. What is the mean literacy rate for these selected countries?

b. What is the standard deviation of these literacy rates?

c. 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?

d. How many countries in this group actually have literacy rates between one standard deviation below the mean and one standard deviation above the mean?

e. What assumption did you make in answering part **c.** above?

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.

a. Calculate the sample proportion of defective items.

b. 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. 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)
1	0.5
2	0.7
3	0.6
4	0.7
5	0.8
6	0.8
7	0.9
8	0.6
9	0.9
10	1.0

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

i. Do the variables selected for measurement seem appropriate for answering the question the pharmacist is interested in?

ii. What biases or errors might be present in the data?

iii. What level of measurement (nominal, ordinal, interval, ratio) does the data possess?

b. Plot the data points on a scatterplot.

c. Based on the scatterplot in part **b.**, answer the following questions regarding the overall pattern of the data.

i. Does the pattern roughly follow a straight line?

ii. Is the pattern upward sloping or downward sloping? Are the data values tightly clustered in the pattern or widely dispersed?

iii. Are there significant deviations from the pattern?

10. Sometimes the following descriptions are assigned to the correlation coefficient.

$r = 0$	no linear relationship
$-0.5 < r < 0$	weak negative linear relationship
$0 < r < 0.5$	weak positive linear relationship
$-0.8 < r \leq -0.5$	moderate negative linear relationship
$0.5 \leq r < 0.8$	moderate positive linear relationship
$-1.0 < r \leq -0.8$	strong negative linear relationship
$0.8 \leq r < 1.0$	strong positive linear relationship
$r = 1$	exact positive linear relationship
$r = -1$	exact negative linear relationship

Describe the relationships indicated by the correlation coefficients below using the descriptions defined above.

a. $r = 0.9$ **c.** $r = -0.9$ **e.** $r = 0$

b. $r = 0.5$ **d.** $r = -0.5$

11. Describe the relationships indicated by the correlation coefficients below using the descriptions defined in problem 10 above.

a. $r = 0.8$ **c.** $r = -0.8$ **e.** $r = 0.1$

b. $r = 0.4$ **d.** $r = -0.4$

12. 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.

13. Consider the following data.

x	1	2	3	4	5	6	7
y	1.00	1.41	1.73	2.00	2.24	2.45	2.65

a. Plot the data points on a scatterplot.

b. Determine the correlation coefficient.

c. Describe the relationship between x and y.

AE 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. How exactly are odds related to probabilities? If the probability of event A occurring is p, then the odds in favor of A occurring are a to b such that $\frac{a}{b} = \frac{p}{(1-p)}$. The odds against A occurring are b to a.

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?

AE Additional Exercises

1. A statistics professor has determined the following probability distribution for X, the grade which a student will earn in a business statistics class.

Grade Distribution

Grade	x	$P(X = x)$
A	4.0	0.15
B	3.0	0.35
C	2.0	0.25
D	1.0	0.15
F	0.0	0.10

a. What is the average grade that a student will earn in a business statistics class?

b. Find the variance of the grades that students will earn in a business statistics class.

c. Find the standard deviation of the grades which students will earn in a business statistics class.

d. What is the probability that a student will earn a grade of 4.0?

e. Find the probability that a student will earn a grade of at least 2.0.

f. Find the probability that a student will earn a grade of at most 1.0.

g. Find the probability that a student will earn a grade of more than 3.0.

2. The U.S. 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?

3. 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?

4. 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

5. 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.

6. A manufacturer of digital cameras knows that a shipment of 30 cameras sent to a large discount store contains eight 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?

7. In a certain shipment of sixteen radios, four are defective. Eight of the radios are selected at random without replacement. What is the probability that at least one of the eight radios is defective?

8. According to the American Hotel and Lodging Association (AH&LA), women accounted for 31% of business travelers in the year 2009. Suppose that to attract these women business travelers, the AH&LA found that 80% of hotels offer hair dryers in the bathrooms. Consider a random and independent sample of 15 hotels.

Source: American Hotel & Lodging Association

- **a.** Based on the information given, how many of the 15 hotels are expected to offer hair dryers in the bathrooms?
- **b.** Find the probability that all of the hotels in the sample offer hair dryers in the bathrooms.
- **c.** Find the probability that more than 5 but less than 9 of the hotels in the sample offer hair dryers in the bathrooms.

9. 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?

10. 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 your first correct answer is within the first fifteen problems?

11. 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?

12. 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.

13. While on a shopping spree, you randomly select five portable music players from an electronics store that sells 20 portable music players. Of these 20 music players, 12 will last beyond the 1-year limited warranty and will not need to be replaced or repaired. What is the probability that at least three of the five portable music players selected will not last beyond the limited warranty period without needing to be replaced or repaired?

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.

a. What probability model would be appropriate for describing the number of buildings in the sample that have code violations? Explain your answer.

b. If L-Mart randomly selects four buildings to inspect, what is the probability that three of the buildings will have violations?

AE Additional Exercises

1. Joe is always between 5 and 35 minutes late for work. Assuming that X, the number of minutes that Joe is late for work, has a continuous uniform distribution, answer the following questions.

- **a.** Find the mean of X.
- **b.** Find the standard deviation of X.
- **c.** Find the probability that on any randomly selected day, Joe is at least a half hour late for work.
- **d.** What percent of the time will X fall within one standard deviation of its mean?
- **e.** Compare this percent to that given by the Empirical Rule of Chapter 4. Why the large discrepancy?

2. Using the standard normal tables, determine the following probabilities. Sketch the associated areas.

- **a.** $P(0 \le z \le 0.85)$
- **b.** $P(-1.25 \le z \le 2.25)$
- **c.** $P(z \ge 1.75)$
- **d.** $P(z \le -2.75)$

3. Using the standard normal tables, determine the following probabilities. Sketch the associated areas.

- **a.** $P(0 \le z \le 1.00)$
- **b.** $P(-2.50 \le z \le 3.01)$
- **c.** $P(z \ge 3.25)$
- **d.** $P(z \le -2.50)$

4. Find the value of z such that 0.99 of the area under the curve lies between $-z$ and z.

5. Find the value of z such that 0.80 of the area under the curve lies between $-z$ and z.

6. The weights of newborn baby boys 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 boy, born at the local hospital, is randomly selected, answer the following questions.

- **a.** Find the probability that the weight of the newborn baby boy will be more than 8 lb.
- **b.** Find the probability that the weight of the newborn baby boy will be less than 6 lb.
- **c.** Find the probability that the weight of the newborn baby boy will be between 6.5 lb and 8.5 lb.
- **d.** Find the weight that separates the lowest 10% of the weights from the highest 90% of the weights.
- **e.** If babies in the lowest 10 percent of weights are kept for observation, would a baby that weighed 5 lb be kept for observation?

7. 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.

- **a.** Find the expected number of patients in the sample that will have had a medication error.
- **b.** What is the standard deviation of the number of patients in the sample that will have had a medication error?
- **c.** What is the probability of observing one or more patients who have had medication errors in the sample?
- **d.** What is the probability of observing two or more patients who have had medication errors in the sample?
- **e.** Do you have any concerns about the accuracy of the probabilities you determined in parts **c.** and **d.**?

8. The number of violent crimes committed in a large city follows a Poisson distribution with an average rate of 10 per month.

- **a.** Find the expected number of violent crimes committed in a 3-month period.
- **b.** Find the standard deviation of the number of violent crimes committed in a 3-month period.
- **c.** Find the probability that at least 45 violent crimes will be committed in a 3-month period.
- **d.** Find the probability that between 25 and 40 (inclusive) violent crimes will be committed in a 3-month period.
- **e.** Find the probability that less than 20 violent crimes will be committed in a 3-month period.

9. 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. 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 verbal portion of the SAT Reasoning Test?

Source: U.S. Census Bureau

10. 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. 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.

Source: U.S. Census Bureau

- **a.** In 1980, what percent of the population consumed at least 100 pounds of red meat?
- **b.** In 2008, what percent of the population consumed at least 100 pounds of red meat?
- **c.** In 1980, what percent of the population consumed at most 130 pounds of red meat?
- **d.** In 2008, what percent of the population consumed at most 130 pounds of red meat?
- **e.** 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?

11. 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?

12. The manager of a retail store wants to determine the best method of staffing employees without being wasteful. The problem that is often encountered is having too many employees and too few customers, and vice versa. The manager realized that during the holiday season they attract, on average, 90 customers per hour. On some days, the number is higher, on others, the number is lower. The manager would like to determine the probability of at least 2 customers arriving in a given minute during holiday season. Use the normal approximation to find the probability.

13. 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.

14. 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?

15. 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?

16. 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?

17. 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. Calculate the mean and standard deviation of the income of families in the neighborhood.

18. Suppose that the 30^{th} percentile of a normal distribution is equal to 756 and that the 90^{th} percentile of this normal distribution is 996. Find the mean and standard deviation of the normal distribution.

AE Additional Exercises

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

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

a. What sampling technique was used for this survey?

b. What biases may be present in the responses?

c. Is 45% a reasonable estimate of the proportion of all Americans who favor a national healthcare 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 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 inquires 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.

d. Why might the data gathered from this sample misestimate the proportion of students who will actually 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 types 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. A report released by the U.S. Census Bureau in November of 2001 stated that fewer families fit the traditional family makeup in 2000 as compared with 1970. The report stated that in 1970, 5.6 million families were headed by women with no husband present, and 16.6% of the households consisted of people living alone. In 2000, these figures increased to 16.2 million and 26.7% respectively. The report also stated that in 1970, 40% of the households consisted of married couples with no children while in the year 2000 this decreased to 25%. Suppose that a random sample of 150 households is chosen in 2000, and the percent of these that consist of people living alone is determined.

- **a.** Find the probability that at least 30% of those households sampled consist of people living alone.
- **b.** Find the probability that at most 28% of those sampled households consist of people living alone.
- **c.** Find the probability that between 26% and 36% of those households consist of people living alone.

7. 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.

8. 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.

9. 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's 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?

10. 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.**?

11. 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.

a. What is the probability that a randomly selected game will generate noise in excess of 103 decibels at the stadium?

b. What is the probability that a randomly selected game will generate a noise level of exactly 103 decibels?

c. 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.

12. 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.

a. Assuming that the town manager is correct and that $p = 0.6$, describe the sampling distribution of $\hat{p}$.

b. What is the probability that between 50% and 70% of the residents will approve the stadium construction?

13. It is believed that 90% of all adults and 85% of all kids between the ages of 12 and 17 have cellular phones. Suppose a sample of 500 adults and 400 kids was taken.

a. Describe the sampling distribution of the proportion of adults that have cellular phones. Assume that the stated probabilities above are true.

b. Describe the sampling distribution of the proportion of kids that have cellular phones. Assume that the stated probabilities above are true.

c. What is the probability that the sample proportion of adults having cell phones will be within 2% of the true proportion?

d. What is the probability that the sample proportion of kids having cell phones will be within 4% of the true proportion?

14. 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?

15. 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?

16. 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?

17. Suppose that it has been reported by a group of researchers that the average number of hours of TV 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.

AE Additional Exercises

1. The owner of Sloppy Jack's bar is thinking about installing some video game machines. To estimate the profitability of the machines, he measures the number of times a competitor's machines are played over a randomly selected sample of days. The preliminary sample showed that the standard deviation of the number of times the machines are played is 10 times per day. Find the sample size (in days) necessary to estimate the average number of times the machines will be played in a day to within 5 plays with 99% confidence.

2. 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?

3. 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. If 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, 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 with 95% confidence.

4. According to a 2001 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.

5. 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.

6. In a population of non-unionized 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.

7. 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 88% confidence?

8. 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.

a. What is a 99% confidence interval for the mean account valuation of the population of customers? Interpret the interval in terms of the problem.

b. What assumption about the account distribution is necessary to solve this problem?

9. Direct Music 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 sample of size 36 was taken; the sample mean income is $31,100. The population standard deviation is estimated to be $4500. Construct a confidence interval using a confidence coefficient of 0.95.

10. 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 Virginia Tech employee. Also, suppose that from a previous experiment, we know that $s = \$6300$. What is the minimum sample size required?

11. 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 $650 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.

12. A stock market analyst wants to estimate the average return on a certain stock. A random sample of 15 days yields an average return of 10.37% and a standard deviation of 3.5%. Give a 90% confidence interval for the true average return on the stock. Assume that the stock returns are normally distributed.

13. 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.

14. Before beginning a pension program for its workers, a corporation wishes to estimate the proportion of its workers who have been employed at the company for at least 20 years. A random sample of 138 employees yielded 16 who have been with the corporation for at least 20 years. Construct a 92.2% confidence interval for the proportion of employees who have worked for this corporation for at least 20 years. Interpret your results.

AE Additional Exercises

1. A tire company has found that the mean time required for a mechanic to replace a set of four tires is 18 minutes. After instituting a new installation procedure, the company unfortunately 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 population standard deviation is 1.14 minutes. Do 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 900 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 900 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 920 hours and the population standard deviation is 80 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 population 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 the *P*-value approach with 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 ounce, 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. In mid-2002, 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. Do the sample data enable Techside Real Estate, Inc. to conclude that the population mean apartment rental rate now exceeds the level reported in 2002? 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 \$10,192. 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 \$9750 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. Do 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, Tommy Walker decides to conduct an experiment on his favorite brand of whiskey to determine if the proof rating on the bottle is accurate. He 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 Tommy conclude that the actual proof of whiskey is not equal to 80 at $\alpha = 0.05$?

e. What assumption did Tommy 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 a 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 and the population standard deviation is 3.4 pounds. Do you think that the airline's concern is justified? Use $\alpha = 0.01$.

13. Consider the following large sample hypothesis tests for the population mean. 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 small sample hypothesis tests for the population mean. 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. 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.

b. 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.

c. 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.

d. 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.

e. 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. Do 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.

a. Test the news magazine's claim at $\alpha = 0.08$.

b. 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.

a. Does this sample result support the claim of the union? Test using $\alpha = 0.10$.

b. 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 these 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 $\alpha = 0.05$.

AE 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	$\bar{x}$	σ
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?

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 the hospitals in a Southern state. 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 female and 40 male electrical engineers. It is of interest to determine if female electrical engineers tend to have higher starting salaries than their male counterparts.

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. It is intended to show that younger drivers are more likely than older drivers to have had a speeding ticket in the past 12 months.

e. Do women have a shorter reaction time than men when exposed to a certain stimulus? Random and independent samples of 10 men and 10 women are included in an experiment that measures reaction time to the stimulus.

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.

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

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. Do the data appear to satisfy the assumption described in part **b.**? Why or why not?

d. 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 approximately equal.

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

a. What assumptions must be made in order to perform the hypothesis test?

b. Will the design group recommend Composition A or Composition B for the monofilament cord at $\alpha = 0.10$?

5. Consider Example 11.3.1. If you were to perform a two-sample *t*-test, you would find that you would fail to reject the null hypothesis and conclude that there is no difference in average daily sales between the two restaurants. Of course, it would be difficult to believe such a test given that if we examine the data in Table 11.3.2, we see that each of the daily sales figures from Restaurant 2 is more than that from Restaurant 1. From this observation, it is clear that the average daily sales between the restaurants are different. Why, then, would the *t*-test be unable to detect this difference? The answer: **an independent samples *t*-test is not a valid procedure to use with paired data**. The *t*-test is inappropriate because the assumption of independent samples is invalid since the dependence between restaurants is a function of the days. Perform the two-sample *t*-test to verify that the independent samples test would lead the owner to fail to reject the null hypothesis and conclude that there is not a significant difference between average daily sales. The data from Table 11.3.2 are replicated below for your convenience. Use $\alpha = 0.01$.

Daily Food Sales for Two Restaurants ($)			
Day	Restaurant 1	Restaurant 2	Difference = Restaurant 1 – Restaurant 2
1	5828	7894	–2066
2	9836	11,573	–1737
3	3984	5319	–1335
4	5845	6389	–544
5	5210	6055	–845
6	9668	10,631	–963
7	6768	7866	–1098
8	6726	7976	–1250
9	4399	5652	–1253
10	6692	8083	–1391

6. Two independent random samples have been selected, 100 from Population 1 and 150 from Population 2. The sample mean from the first population is 1025 with a population standard deviation of 10. For the second sample, the mean is 1039 with a population standard deviation of 12.

a. Test that there is no difference between the groups using $\alpha = 0.01$.

b. Construct a 95% confidence interval for the true mean difference between Population 1 and Population 2.

7. The manufacturer of Brand 1 cigarettes claims that his cigarettes are no more harmful to health than Brand 2 (filtered) cigarettes. Assuming harmfulness is to be associated with nicotine content, the FDA took random samples of 125 cigarettes from Brand 1 and 180 cigarettes from Brand 2. The average nicotine content in the sample of Brand 1 was 24.6 mg with a population standard deviation of 1.4 mg; the average nicotine content in the sample of Brand 2 was 24.3 mg with a population standard deviation of 1.1 mg.

a. Is there evidence to refute the manufacturer's claim at $\alpha = 0.05$?

b. Construct an 85% confidence interval for the true mean difference in nicotine content between Brand 1 and Brand 2.

8. A team of organizational behavior managers investigated the effects of an orientation program on "first day of work" anxiety levels of new employees. 72 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.

a. 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.

b. Calculate the observed significance level (P-value) of this test.

9. A property manager of thousands of apartments wants to test the difference in the mean net annual income between two types of leasing arrangements. Arrangement A is to charge a lower rent but to require the tenants to make repairs. Arrangement B is to charge a higher rent and to state that the landlord will make the repairs. A sample of 25 apartments using Arrangement A had a mean net annual income of \$1532.50 with a standard deviation of \$400. A sample of 22 apartments using Arrangement B had a mean net annual income of \$1489.20 with a standard deviation of \$100. Test at $\alpha = 0.025$ that Arrangement B will have a lower mean net annual income by **at least \$10** than Arrangement A. Assume that the incomes are normally distributed and that the population variances are equal.

10. 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.

11. A random sample of 10 filled sports drink bottles is taken in one bottling plant, and the mean weight of the bottles is found to be 22 ounces with a variance of 0.09 ounces squared. At another plant, 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 $\alpha = 0.05$.

12. Employers implementing workplace wellness programs report reduced healthcare costs, mostly because of fitness center memberships and free preventative screenings for their employees, according to the findings of a survey. Survey results show that employers who invest in wellness programs see increased employee retention, attendance, and productivity. The numbers show that medical costs improve and absenteeism costs improve. The following data have been summarized from the 30 employers that implemented wellness programs in 2011 along with their costs in 2010. The standard deviation of the difference in cost between the two years is $1322. Assume that independent, random samples were used in the study.

Average Healthcare Cost per Employee (Dollars)		
	2010	2011
Mean	$11,151	$10,344
Sample Size	30	30

a. Calculate and interpret a 95% confidence interval for the true average difference in healthcare costs between 2010 and 2011.

b. Using the rejection region approach, test that there is a significant decrease in average healthcare costs between 2010 and 2011. Make your decision based on a significance level of 0.05.

c. Calculate the *P*-value for the test conducted in part **b.**

13. Travelers are increasingly using tablets and e-readers on planes, trains and buses instead of more basic electronic devices such as cell phones and iPods. According to a study, the use of tablets and e-readers such as iPads, Kindles and Nooks rose more than 50% in 2011 across all modes of transportation. A team of researchers observed 7781 passengers on various airplane, bus and train departures in December of 2011. Their findings were compared with data about technology use among travelers for the fourth quarters of 2009 and 2010. The following table represents the number of travelers surveyed that used tablets or e-readers.

Travelers That Used Tablets or e-Readers			
	2009	2010	2011
Number of Travelers	7545	7185	7781
Number using Tablets or e-Readers	330	570	792

a. Perform a hypothesis test using a significance level of 0.05 to determine if there is sufficient evidence to refute the claim that at least 10% of travelers used tablets or e-readers in each of the years surveyed.

b. Calculate and interpret a 95% confidence interval for the difference in the percentage of travelers using tablets or e-readers for 2010 and 2011.

c. Test that the proportion of travelers using tablets or e-readers has increased from 2010 to 2011. Make your decision based on a significance level of 0.05.

14. The University of Michigan Transportation Research Institute believes that a teenage rite of passage may be losing its allure among young people. They believe that teens are in no 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 2008, surveying a sample of 10,000 16-year-olds in the U.S., only 31% had their licenses. 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 2008. Use a significance level of 0.05.

15. The results of a recent study show that people blessed with intelligence are not necessarily free from financial woes. Research by an economist at Ohio State University's Center for Human Research, detailed in an upcoming issue of the journal *Intelligence*, indicated that the smart may still suffer monetarily, making their own financial lives difficult by doing things like maxing out their credit cards and missing bill payments. While past studies have shown a link between being smarter and earning more money, that doesn't mean that those who are intelligent and wealthy don't suffer from debt. Generally, respondents with higher IQ scores earn higher income, with each additional IQ point adding $202 to $616 in income annually. However, the study demonstrated that there was no link between total wealth and a person's IQ. With research showing that individuals with higher IQs are not significantly better off financially than those with average or below average intelligence, the suggestion is that those who are smarter may not be as good about saving money as others. In fact, the researchers discovered that more intelligent people were actually slightly more prone to financial problems, with more than 12 percent of those with an IQ of 90 having maxed out their credit cards, versus less than 8 percent of those individuals with an IQ of 75 and below. Among the highly intelligent, or those respondents with an IQ greater than 125, 6 percent had maxed out credit cards and 11 percent occasionally missed payments. Additionally, there appeared to be a financial sweet spot near the average IQ score, for which individuals had the lowest financial distress. The survey looked at information provided by 7400 respondents who took part in the National Longitudinal Survey of Youth, which considered baby boomers nationwide. The latest study is based on data from 2004, when participants were 40 to 47 years old. Respondents were polled about income and total wealth, as well as three measures of financial distress: whether they had maxed out credit cards, if they missed paying bills during the past five years, and if they had ever declared bankruptcy. The Armed Forces Qualification Test, long used to measure intelligence, was used to generate IQ scores. As a takeaway, the study's author, economist Jay Zagorsky, stated that people with lower intelligence should not feel like they are at a financial disadvantage, while those people with a high IQ should not feel they are at an advantage. In other words, every person has the ability to make smart financial decisions like avoiding debt, regardless of his or her intelligence or income. The following table contains some of the data from the study.

Income and Intelligence

IQ	Average Income ($)	Average Total Wealth ($)	*n*	Percentage with Maxed out Credit Cards
<75	55,000	500,000	2000	7.8
90	70,000	750,000	3000	12.0
>125	95,000	950,000	2400	6.0

Source: "Smarter doesn't mean richer," May 14, 2007, www.creditcards.com

a. Determine if there is a significant difference in the proportion with maxed out credit cards of those with an IQ of 90 versus those with an IQ less then 75. Make your decision based on a significance level of 0.05.

b. Calculate a 95% confidence interval for the difference in the percentage of people with maxed out credit cards between those with an IQ of 90 and those with an IQ more than 125. Interpret the interval.

c. Did you make any assumptions to carry out the test in **a.** or to calculate the confidence interval in **b.**? If so, what were they?

16. 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.

17. 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

AE 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.**? Do 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

a. Do the data indicate a difference among average fat contents per serving for the three brands of margarine at $\alpha = 0.01$?

b. What assumptions were made for the test in part **a.**? Do the data appear to satisfy these assumptions? Explain.

c. 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*. 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.

a. Compute the sums of squares and their degrees of freedom for treatments and error based upon the statistics given in the table.

b. Compute MST and MSE.

c. With $\alpha = 0.01$, can we conclude that there is a significant difference among the degrees of psychological reactance for the different age groups?

d. 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*. 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

a. Compute the grand mean and the sum of squares for treatments (SST).

b. 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.

c. With $\alpha = 0.01$, can we conclude that there is a significant difference among the groups in the perception of the umpire?

d. What assumptions are necessary for performing the test in part **c.**? Can they be checked in this instance?

6. Consider the following data regarding median starting and mid-career salaries for graduates from schools in different regions of the United States.

Starting and Mid-Career (Salaries by Region)		
School	**Starting Median Salary ($)**	**Mid-Career Median Salary ($)**
Midwest		
Notre Dame	52,900	107,000
Carleton College	42,800	98,300
Illinois Institute of Technology	52,000	96,000
Denison University	40,600	94,000
University of Chicago	46,900	92,700
Northwestern University	49,900	88,300
Washington University in St. Louis	51,200	87,700
Northeast		
Princeton	56,900	130,000
Harvard	54,100	116,000
Massachusetts Institute of Technology	69,700	115,000
Dartmouth	51,600	114,000
Bucknell University	52,600	108,000
Manhattan College	53,900	107,000
Williams College	51,800	105,000
South		
Duke University	54,400	113,000
Vanderbilt University	51,300	100,000
Washington and Lee University	48,600	99,800
Wake Forest University	46,000	98,800
Rice University	51,100	97,400
Georgetown University	50,300	96,900
College of William and Mary	45,000	96,500
West		
University of Colorado – Boulder	45,900	90,400
University of Washington	46,700	88,400
Gonzaga University	44,200	87,700
Brigham Young University	47,400	86,800
University of Arizona	45,400	81,600
University of Oregon	39,700	79,200
Santa Clara University	52,900	105,000

Source: Payscale.com

a. Using $\alpha = 0.05$, is there a significant difference among average starting salaries for the four different regions?

b. What is the value of F for this test?

c. Using $\alpha = 0.05$, is there a significant difference among average mid-career salaries for the four different regions?

d. State the assumptions made for the two hypothesis tests in parts **a.** and **c.**

e. Do you have any concerns about these data? Explain.

7. Consider the following data regarding the median starting and mid-career salaries by type of major.

Starting and Mid-Career Salaries by Major

Major	Starting Median Pay ($)	Mid-Career Median Pay ($)	Major	Starting Median Pay ($)	Mid-Career Median Pay ($)
Engineering			**Math and Science**		
Petroleum Engineering	97,900	155,000	Applied Mathematics	52,600	98,600
Chemical Engineering	64,500	109,000	Computer Science	56,600	97,900
Electrical Engineering	61,300	103,000	Statistics	49,000	93,800
Aerospace Engineering	60,700	102,000	Mathematics	47,000	89,900
Computer Engineering	61,800	101,000	Physics	49,800	101,000
Nuclear Engineering	65,100	97,800	Biochemistry	41,700	84,700
Biomedical Engineering	53,800	97,800	Food Science	43,300	83,700
Mechanical Engineering	58,400	94,500	Geology	45,300	83,300
Industrial Engineering	57,400	93,100	Molecular Biology	40,500	81,200
Civil Engineering	53,100	90,200	Chemistry	42,000	80,900
Environmental Engineering	51,700	88,600	**Other**		
Business			Economics	47,300	94,700
Finance	46,500	87,300	Film Production	41,600	80,700
Supply Chain Management	50,200	84,700	Political Science	39,900	80,100
International Business	41,600	83,700	International Relations	40,500	79,400
Accounting	44,700	75,700	Philosophy	39,800	75,600
Advertising	37,700	74,700	History	37,800	69,000
Marketing	38,200	73,500	Communications	38,000	66,900
Business	41,000	70,500	Journalism	36,100	66,400
Public Relations	35,500	65,700	Spanish	36,400	58,400

Source: Payscale.com

a. State the null and alternative hypotheses to test if there is a significant difference in starting salary among the four types of majors.

b. Use $\alpha = 0.01$ to test for a significant difference among average starting salaries for the different types of majors.

c. State the null and alternative hypotheses to test if there is a significant difference among average mid-career salaries for the four types of majors.

d. Use $\alpha = 0.01$ to test for a significant difference among average mid-career salaries for the different types of majors.

8. 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		

a. Complete the ANOVA table.

b. At the 0.05 level, is there evidence of significant interaction between A and B? Justify your answer.

c. At the 0.05 level, is there evidence of a significant Factor A effect? Justify your answer.

d. At the 0.05 level, is there evidence of a significant Factor B effect? Justify your answer.

e. Does the result of the test for interaction suggest further investigation? Justify your answer.

9. 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.

	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

a. Graphically plot the test scores by class level and testing method. Discuss the graph.

b. Perform an analysis of the data using the class-level blocks. Are blocking effects significant at the 0.05 level of significance? Explain.

c. Is the experiment useful having been analyzed as a completely randomized block design? Explain.

10. 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		

a. How many blocks are used in the experiment?

b. How many treatments are used in the experiment?

c. How many observations are used in the experiment?

d. What are the null and alternative hypotheses to test if there is a difference among the treatment means?

e. What test statistic should be used to conduct the test in part **d.**?

f. What is the rejection region for the test in parts **d.** and **e.**?

g. Carry out the test and state your conclusion based on a significance level of 0.05.

11. 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

a. Perform the appropriate analysis on the data given and generate the ANOVA table for the analysis.

b. Do 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.

c. What is the *P*-value for the test performed in part **a.**? Interpret this value.

12. The following table is a 3 × 3 factorial design with three observations for each factor level.

Factorial Design Data

Factor A	Factor B		
	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

a. Plot the treatment means using Factor A as the *x*-axis and Factor B as plotting symbols. Do the means appear to be different? Does interaction between factors A and B appear to be present? Justify your answers.

b. Perform the analysis using a software package, generating the ANOVA table.

c. Test for significant interaction using a 0.05 level of significance. Discuss your findings.

d. Test for A and B effects using a 0.05 level of significance. Discuss your findings.

13. 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 are given in the following table.

TSP Revenues by Marketing Strategy (Thousands of Dollars)

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

a. Specify the null and alternative hypotheses to determine if there is a significant difference among average revenues for the three advertising strategies.

b. Generate the ANOVA table to test the hypotheses in **a.**

c. Conduct the test in **a.** using a significance level of 0.05.

d. Was the variation among the observed revenues significantly reduced by blocking? Explain using $\alpha = 0.05$.

AE Additional Exercises

1. 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

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

$$\text{Reaction Time} = \beta_0 + \beta_1\left(\text{Amount of Drug}\right) + \varepsilon_i$$

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

The regression equation is
Reaction Time (Seconds) = 0.533 + 0.0394 Amount of Drug (mg)

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 (mg)
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 fraction 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 significance level and the 0.01 significance 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 of an individual who has 4 mg 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 mg of the drug in their bloodstreams.

i. Suppose a particular individual has 4 mg of the drug in the bloodstream. What would be the 95% prediction interval for the reaction time?

2. A sample of 11 lonely hearts advertisements, all placed by males, was selected from the local newspaper. In each of the selected ads, the males gave their heights, along with other physical characteristics and preferences. Some of the males obviously felt that being taller than average might result in more responses to the ad. Suppose that y, the number of responses to the ad 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
y	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 $\alpha = 0.01$ significance 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 (%)
1	756	6.32
2	679	7.85
3	527	10.20
4	839	5.52
5	677	7.30
6	686	7.37
7	512	9.67
8	590	8.40
9	765	5.82
10	502	10.01
11	819	5.86
12	630	8.51
13	704	6.83
14	679	7.72
15	663	7.68
16	542	9.53
17	575	6.86
18	508	9.65
19	689	7.75
20	750	6.89

a. Draw a scatterplot of the data. Does there appear to be a linear relationship between FICO score and interest rate?

b. Estimate the simple linear regression equation using statistical software.

c. What is the estimate of the mean square error? Interpret this value.

d. Test at the 5% significance level if a linear relationship exists between FICO scores and interest rates.

e. Interpret the regression coefficient corresponding to FICO score.

f. Construct a 95% confidence interval for the slope. Interpret the interval.

g. Compute the coefficient of determination. Interpret this value.

h. Calculate the correlation coefficient. Interpret this value.

i. What is the average interest rate for a credit score of 725?

j. Construct a 90% confidence interval for the average interest rate for people who have FICO scores of 725. Interpret this interval.

k. 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		

a. Draw a scatterplot of the data. Does there appear to be a linear relationship between age and the number of messages that one sends?

b. What is the estimated simple linear regression equation?

c. What is the estimate of the coefficient of determination? Interpret this value.

d. Test at the 5% significance level if a linear relationship exists between age and the number of messages sent via a cellular phone.

e. Interpret the regression coefficient corresponding to age.

f. Construct a 95% confidence interval for the slope. Interpret this interval.

g. Calculate the correlation coefficient. Interpret this value.

h. What is the average number of messages sent by a 15-year-old? Round your answer to the nearest whole number.

i. Construct a 95% confidence interval for the average number of messages sent by a 15-year-old. Interpret this interval.

j. 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.

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 a 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		

a. Draw a scatterplot of the data. Does there appear to be a linear relationship between home size and the amount of annual kWh used?

b. What is the estimated simple linear regression equation?

c. What is the estimate of the coefficient of determination? Interpret this value.

d. Test at the 5% significance level if a linear relationship exists between home size and the annual amount of kWh used.

e. Interpret the regression coefficient corresponding to home size.

f. Construct a 99% confidence interval for the slope. Interpret the interval.

g. Calculate the correlation coefficient. Interpret this value.

h. 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?

i. 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 U.S. 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.

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

a. Draw a scatterplot of the data. Does there appear to be a linear relationship between high school GPA and SAT score?

b. What is the estimated simple linear regression equation?

c. What is the coefficient of determination? Interpret this value.

d. Test at the 5% significance level if a linear relationship exists between high school GPA and SAT score.

e. Interpret the regression coefficient corresponding to high school GPA.

f. Construct a 95% confidence interval for the slope. Interpret the interval.

g. Calculate the correlation coefficient. Interpret this value.

h. What SAT score would you expect for students with a GPA of 3.5? Round your answer to the nearest whole number.

AE Additional Exercises

1. 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. Earning his master's degree will require him to take out approximately $20,000 worth of student loans. 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 $10,000, he will enroll in a degree program.

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Industry Salaries**.

Industry Salaries

Salary ($)	Years of Experience	Master's Degree
37,620	22	No
67,080	27	Yes
31,280	15	No
21,500	2	No
75,120	28	Yes
59,820	25	Yes
40,180	15	Yes
81,360	32	Yes
35,080	19	No
36,080	12	Yes
36,680	22	No
29,200	11	Yes
33,040	18	No
30,060	14	No
53,300	21	Yes
22,820	7	No
72,900	31	Yes
55,920	22	Yes
19,280	0	No
26,000	7	No

a. 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.

b. Using statistical software, estimate the following multiple regression model.

$$\text{Salary} = \beta_0 + \beta_1(\text{Experience}) + \beta_2(\text{Degree}) + \varepsilon_i$$

Write the estimated multiple regression equation.

c. According to the model, how much does salary increase on average with each additional year of experience?

d. According to this model, will Drew decide to enroll in a master's program? Explain your answer.

e. Why should Drew be cautious when using this model to make his decision?

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Sports Club Membership**.

2. A chain of sports clubs wishes to use regression analysis to help determine which features should be included in their new location. They believe that median income in the area is a significant factor in determining the number of people who join a neighborhood sports club. The CEO of the chain gathered data from existing sports clubs regarding the number of members each club had, the median income in the area in which they were located, and whether or not the clubs had a pool, racquetball courts, or group fitness classes. If management can determine with 90% confidence that a pool, racquetball courts, or group fitness classes produces significantly more memberships than sports clubs without those features, they will include them in the new location.

Sports Club Membership

Number of Members	Median Income ($)	Pool?	Racquetball Courts?	Fitness Classes?
1258	32,223	No	No	No
1479	34,975	No	No	No
1480	43,187	No	Yes	No
1701	44,337	No	No	No
2014	52,167	No	No	Yes
2271	57,521	No	No	Yes
2615	58,347	No	Yes	No
2632	60,960	Yes	No	No
2737	62,201	Yes	No	Yes
2810	67,993	No	No	Yes
3563	68,770	No	No	Yes
3765	81,289	Yes	Yes	Yes
3792	83,902	No	No	Yes
4069	84,594	Yes	No	Yes
4393	86,855	Yes	Yes	Yes
4787	88,381	Yes	Yes	Yes

a. What sign do you expect the coefficient of median income to have? Explain why.

b. Create three dummy variables, pool, courts, and classes, that are equal to 1 if the observation contains this feature and equal to 0 if the observation does not contain this feature.

c. Use statistical software to estimate the following regression models. In each case, write the estimated regression equation and state whether the coefficient of the independent variable is significant at the 0.10 level.

i. $\text{Members} = \beta_0 + \beta_1(\text{Pool}) + \varepsilon_i$

ii. $\text{Members} = \beta_0 + \beta_1(\text{Courts}) + \varepsilon_i$

iii. $\text{Members} = \beta_0 + \beta_1(\text{Classes}) + \varepsilon_i$

d. Estimate the following multiple regression model.

$$\text{Members} = \beta_0 + \beta_1(\text{Income}) + \beta_2(\text{Pool}) + \beta_3(\text{Courts}) + \beta_4(\text{Classes}) + \varepsilon$$

Write the estimated regression equation.

e. Are any of the coefficients of the indicator variables significant at the 0.10 level?

f. Explain why it is important to include the income variable in the regression model.

g. After studying these regression results, how would you suggest the management of the sports club chain go about building their new location? Should they use any of the regression models you have estimated? Explain why or why not.

3. 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

a. Assuming that yield is the dependent variable, plot yield against additive. Does the relationship appear to be linear?

b. Using statistical software, estimate the simple linear regression model. Identify R^2 and s_e^2.

c. 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 $\text{Estimated Yield} = -67.53 + 23.04(\text{Additive}) - 0.82(\text{Add\_sq})$. 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.

4. Suppose that an association of real estate professionals has reported home sales for 2011 in a data set titled Home Sales. The table contains the current sales by region and the inventory for existing-home sales (single-family and condos/co-ops). An excerpt of the full table is given below.

Home Sales							
Sale Price	**Region**	**Home Type**	**Inventory**	**Sale Price**	**Region**	**Home Type**	**Inventory**
$237,000	NE	Condo/Co-op	185,000	$239,600	NE	Single-Family	550,000
$225,400	NE	Condo/Co-op	188,000	$242,400	NE	Single-Family	550,000
$235,200	NE	Condo/Co-op	205,000	$244,600	NE	Single-Family	560,000
$144,900	MW	Condo/Co-op	80,000	$138,800	MW	Single-Family	850,000
$145,000	MW	Condo/Co-op	79,000	$138,900	MW	Single-Family	840,000
$139,200	MW	Condo/Co-op	82,000	$138,600	MW	Single-Family	900,000
$110,400	S	Condo/Co-op	194,000	$153,400	S	Single-Family	1,520,000
$108,100	S	Condo/Co-op	176,000	$153,100	S	Single-Family	1,520,000
$112,100	S	Condo/Co-op	200,000	$150,800	S	Single-Family	1,570,000
$154,600	W	Condo/Co-op	90,000	$223,100	W	Single-Family	940,000
$152,900	W	Condo/Co-op	91,000	$216,300	W	Single-Family	940,000
$146,800	W	Condo/Co-op	74,000	$216,900	W	Single-Family	1,050,000

Data

The full table can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Home Sales**.

a. Suggest a regression model that would allow you to predict sale price as a function of inventory, region, and whether you have a condo/co-op or a single-family home.

b. Estimate the model that you suggested in part **a.**

c. What is the estimated equation for predicting sale price by region?

d. What is the estimated equation for predicting sale price by type of home?

e. Is your model estimated in part **b.** statistically useful for predicting sale price at a 1% significance level? Explain your answer.

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Tablet Survey**.

5. Given the digital revolution in the United States and the trend towards textbooks and lecture materials being made available through electronic means, one wonders if having these technologies will help improve students' grades. To help answer this question, a survey was taken on a university campus inquiring if students used a tablet for their classes, which tablet was used, and their grade point average, household income, and the highest level of education attained by their parents. Using the data from the survey presented in the following table, answer the following questions.

Key for Student Survey

Tablet:	Household Income:	Parent's Highest Level of Education:
1 = Motorola Xoom	1 = < $30,000	1 = Some High School
2 = Samsung Galaxy	2 = $30,000–$49,999	2 = High School Diploma
3 = Apple iPad	3 = $50,000–$74,999	3 = Some College
4 = No Tablet	4 = > $75,000	4 = College Graduate

Student Data

GPA	Tablet	Income	Education	GPA	Tablet	Income	Education
3.9622	4	3	1	3.0530	2	3	4
2.9555	1	2	3	3.8034	2	3	3
3.2058	4	2	2	2.3986	2	2	4
3.6487	1	4	2	3.1191	1	3	2
3.4459	1	4	4	2.9556	1	3	3
3.5222	4	1	1	3.3100	2	3	1
3.9964	2	1	3	3.5477	1	3	2
2.4374	1	3	4	3.7710	3	3	1
2.9262	4	3	3	3.2706	1	3	1
3.8684	2	2	3	3.8039	2	1	3
3.2102	1	3	3	3.5149	1	1	2
3.4394	2	3	4	2.0290	2	3	1
3.8169	4	1	3	3.2985	3	1	4
3.6107	2	2	1	2.2877	2	2	2
2.7475	3	2	3	2.4569	3	2	1

a. Suggest a multiple regression model to predict GPA from tablet use and which type of tablet is being used.

b. What is the estimated regression equation for the model proposed in part **a.**?

c. Is the model useful in predicting GPA from whether a student uses a tablet or not and which type of tablet is being used at the 0.05 level? Justify your answer.

d. Discuss whether there is a difference between tablet use and household income regarding how the variables affect one's GPA.

e. Discuss whether there is a difference in the extent to which each type of tablet affects one's GPA.

6. 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.

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second 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	26	23.82	0.78	3.28
2	15.38	0.92	6.02	27	19.58	0.64	3.26
3	25.25	1.43	5.66	28	136.94	4.41	3.22
4	21.28	1.13	5.31	29	83.57	2.65	3.17
5	698.75	36.88	5.28	30	6.44	0.20	3.12
6	120.55	6.22	5.16	31	13.50	0.42	3.07
7	23.00	1.09	4.74	32	47.47	1.45	3.06
8	13.34	0.62	4.66	33	30.28	0.92	3.03
9	24.82	1.15	4.63	34	14.56	0.43	2.94
10	22.96	1.04	4.53	35	36.06	1.06	2.93
11	78.51	3.21	4.09	36	14.26	0.41	2.90
12	24.22	0.99	4.08	37	369.35	10.56	2.86
13	8.70	0.35	4.02	38	7.07	0.20	2.83
14	15.08	0.58	3.86	39	20.12	0.57	2.82
15	23.44	0.90	3.84	40	1.00	0.03	2.81
16	22.81	0.87	3.80	41	104.15	2.91	2.79
17	22.80	0.84	3.70	42	32.4	0.89	2.75
18	113.84	4.21	3.69	43	47.58	1.30	2.73
19	13.62	0.49	3.60	44	18.43	0.49	2.64
20	22.86	0.82	3.58	45	27.78	0.73	2.63
21	15.58	0.56	3.57	46	30.10	0.76	2.52
22	121.33	4.16	3.43	47	26.55	0.66	2.49
23	15.27	0.51	3.34	48	19.53	0.46	2.35
24	8.70	0.29	3.30	49	27.43	0.62	2.26
25	45.40	1.50	3.30	50	30.33	0.62	2.04

a. Using the data in the table, 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.

c. Can you detect any multicollinearity in the model containing share price and dividend per share? Explain your answer.

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > SNAP Benefits**.

7. 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. Eligible households can receive food assistance through regular SNAP or through the Louisiana Combined Application Project (LaCAP). Using the data in the table, 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
418.78	4	2736	642.77	4	3933
506.46	2	3274	364.81	8	1925
552.53	2	3480	619.30	6	3736
586.46	7	3741	238.71	1	1453
637.18	8	3684	378.94	4	2538
244.49	2	1476	302.58	1	1798
507.19	5	2835	231.74	8	1189
512.56	5	2873	428.67	6	2247
312.89	4	1618	286.99	5	1460
329.05	4	1565	268.81	1	1567
243.49	6	1582	329.81	6	1622
560.37	8	3380	627.25	3	3828
599.90	3	3922	421.52	6	2782
657.09	5	3845	656.38	2	3978
394.82	5	2233	400.64	3	2493

a. Suggest a regression model that will assist SNAP administrators in providing a monthly benefit to eligible households.

b. Fit the model that you suggested in part **a.** Is this model useful in predicting monthly benefits? Justify your answer.

c. Are all independent variables in the model helpful in explaining the variation in monthly benefits? Explain your answer.

d. Give a 95% confidence interval for average monthly benefits for a four-member household with a gross monthly income of $2500. Interpret this interval.

e. Provide a 99% prediction interval for a four-member household with a gross monthly income of $2500. Interpret this interval.

f. What is the difference between the intervals found in parts **d.** and **e.**?

AE Additional Exercises

1. Compute the Mean Absolute Deviation for the 3-month SMA forecast using the TixPixx data.
2. Compute the Mean Absolute Percentage Deviation for the 3-month SMA and WMA forecasts using the TixPixx data.
3. Compute the Cumulative Error for the 3-month SMA and WMA forecasts using the TixPixx data.
4. Compute the Mean Squared Error for the 3-month SMA and WMA forecasts using the TixPixx data.
5. Triplett Farms is a company that raises turkeys, which it sells to a meat-processing company throughout the year. However, the peak season obviously occurs during the fourth quarter of the year, October to December. Triplett Farms has experienced a demand for turkeys for the past 3 years shown in the following table:

Demand (1,000s)					
Year	**Quarter 1**	**Quarter 2**	**Quarter 3**	**Quarter 4**	**Total**
2019	12.6	8.6	6.3	17.5	45.0
2020	14.1	10.3	7.5	18.2	50.1
2021	15.3	10.6	8.1	19.6	53.6
Total	42.0	29.5	21.9	55.3	148.7

Compute the seasonally adjusted forecasts for each quarter of 2022.

AE 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.

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. Consider the following proportions of wives having affairs and husbands having affairs reported in *What the Odds Are* by Les Kranz.

Affairs

	Wives	Husbands
Have an Affair Before 2 Years of Marriage	0.13	0.14
Have an Affair Between 2 and 10 Years of Marriage	0.20	0.25
Have an Affair After 10 Years of Marriage	0.20	0.33
Have Not Had an Affair	0.47	0.28

a. If the results were based on a survey of 200 wives and 200 husbands, do the data suggest that the proportion of wives in the various categories who have had affairs differs significantly from the proportion of husbands in the various categories who have had affairs at $\alpha = 0.10$?

b. What assumptions were made in the test for part **a.**?

3. Do you think there are people somewhat like ourselves 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. Suppose that a sample of 100 people is chosen in 2011 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 of toward extraterrestrial life have changed since 1978?

4. 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?

5. According to the Statistical Abstract of the United States, 22.6% of those 18 and over in the U.S. in 1992 were never married, 61.1% were married, 7.5% were widowed, and 8.8% were divorced. Suppose the following table summarizes the marital status of 90 randomly chosen adults in 2011.

Marital Status				
	Never Married	**Married**	**Widowed**	**Divorced**
Frequency	27	42	10	11

With $\alpha = 0.01$, can we conclude that in 2011 the U.S. is different than in 1992, with respect to marital status?

6. The National Restaurant Association is interested in determining if there is a relationship between the type of pizza pie Americans prefer and the region of the country in which they live. The association randomly selects 285 Americans and records the category of pizza pie which best describes their preference and the region of the country in which they live with the following results.

Pizza Preference				
	Region			
Type of Pizza Pie Preferred	**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 pie 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.**?

7. Do you think marriages between homosexuals should or should not be recognized by law as valid with the same rights as traditional marriages? The responses to this question and the region of the country where the respondent lived were summarized in the *Gallup Poll Monthly*. The following contingency table summarizes the 799 responses.

Poll Results

	Response		
Region	**Should**	**Should Not**	**No Opinion**
East	78	136	16
Midwest	63	183	16
South	74	23	16
West	50	138	6

With $\alpha = 0.01$, can we conclude that attitude about homosexual marriages is dependent on region?

8. Consider the following data regarding the percentage of mobile application usage and smartphone operating system. Suppose the data were based off of 100 responses.

Mobile App Usage (%)

Operating System	**Multiple Times a Day**	**Once a Day**	**Several Times a Week**
Apple iOS	68	11	16
Android OS	60	12	21
Palm OS	48	14	25
Blackberry OS	45	13	29
Microsoft Windows Mobile	29	21	43
Other	47	13	21

a. Is there evidence that the frequency of mobile app usage is dependent on the smartphone operating system? Test at $\alpha = 0.05$.

b. What assumptions are made for the test in part **a.**? Are the assumptions reasonable for this problem? Why or why not?

AE 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, do 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, do 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.

Initial Public Offerings

Company	Initial Offer Price ($)	Price After 1st Day of Trading ($)
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

Initial Public Offerings		
Company	Initial Offer Price ($)	Price After 1st Day of Trading ($)
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

Source: IPOScoop.com

a. 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.

b. 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 $\alpha = 0.05$.

c. 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 $\alpha = 0.05$.

d. 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?

e. 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	Diet B
84	94
77	81
89	95
98	93
97	97
100	99
75	82
85	92
96	95
78	89

a. 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, do 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	37,000
3.20	38,000
3.21	42,000
3.39	40,000
3.55	44,000
3.57	48,000
3.76	50,000
3.77	60,000
3.79	59,000
3.90	55,000

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.

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
Coca-Cola	6	70
Amazon.com	7	78
FedEx	8	73
Microsoft	9	38
McDonald's	10	111

Source: CNN Money/Fortune Magazine

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
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.

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

Source: *U.S. News and World Report*, 2011

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.

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

Source: *Condé Nast Traveler*, 2011

a. 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.

b. What assumptions were made for the test performed in part **a.**?

AE Additional Exercises

1. Edwards Electrical Company produces voltage regulators designed to maintain 220 volts. Samples of 10 units are taken from production to monitor the process and the voltage regulators are found to maintain 220 volts as designed. The mean of the sample ranges is found to be 4 volts.

a. Determine the UCL, LCL, and centerline for an $\bar{x}$ chart.

b. Determine the UCL, LCL, and centerline for an R chart.

2. A steel products manufacturer makes 20-foot lengths of pipe that are later cut into smaller lengths in the production process. To monitor the production and to make sure the pipe is acceptable for the cutting state, a sample of 20 pipes is taken each hour of the day. Along with the pipe length, the range of pipe length is recorded for each sample. Determine the upper and lower control limits for an R chart and indicate which samples, if any, are out of control.

Pipe Length Ranges

Sample Number	Sample Range	Sample Number	Sample Range	Sample Number	Sample Range
1	0.03	9	0.08	17	0.05
2	0.07	10	0.06	18	0.06
3	0.07	11	0.05	19	0.07
4	0.03	12	0.06	20	0.05
5	0.02	13	0.06	21	0.05
6	0.08	14	0.03	22	0.06
7	0.07	15	0.03	23	0.04
8	0.03	16	0.02	24	0.04

3. The league director decides to sample 200 matches each season to study forfeits in league games. The numbers of forfeits for the last eight seasons are shown in the following table. Find the upper and lower control limits and construct a p chart. Indicate which seasons, if any, are out of control.

Game Forfeits

Season Number	Number Forfeited	Season Number	Number Forfeited
1	8	5	7
2	10	6	5
3	9	7	12
4	8	8	16

4. A tire manufacturer randomly samples 20 tires at the end of each shift to determine if the tires are defective. The numbers of defectives in 12 shifts are given in the following table. Construct an appropriate control chart to determine if the tire manufacturing process is in control. Identify any shifts that are out of control.

Defective Tires

Shift	Number of Defectives
1	4
2	2
3	0
4	5
5	2
6	3
7	14
8	2
9	3
10	4
11	12
12	3

5. The vice president of audit at a Fortune 500 firm customarily checks the financial statements for errors in 15 departments. The following table contains information about mistakes made on financial statements. Construct an appropriate control chart to determine if the VP should be concerned about the number of mistakes being made on the financial statements.

Mistakes on Financial Statements

Sample Size	Number of Mistakes	Sample Size	Number of Mistakes
15	0	15	0
15	0	15	3
15	3	15	8
15	2	15	7
15	6	15	0
15	3	15	1
15	4	15	5
15	8	15	4
15	2		

6. 15 samples of five items each were taken to monitor the amount of fill for a 12-ounce bottle of soda. As each bottle is selected from the line, it is measured. The data from the process is given in the following table. Create an $\bar{x}$ chart and an R chart to determine if the process is in control.

Fill Amounts for 12-Ounce Bottles

	Observations				
Sample	1	2	3	4	5
1	13.11	11.17	11.35	13.71	13.00
2	13.19	13.50	13.63	12.36	11.98
3	12.83	12.12	12.65	12.22	11.43
4	11.62	13.72	13.73	11.36	13.93
5	13.65	12.44	13.15	12.42	11.21
6	12.60	11.07	11.17	12.88	12.33
7	11.07	13.22	11.41	13.21	13.14
8	11.56	12.16	12.47	13.41	13.13
9	13.23	12.05	11.25	12.55	13.00
10	12.30	12.18	12.00	13.59	12.18
11	14.00	12.33	11.52	13.29	13.51
12	12.05	11.52	12.80	12.21	12.67
13	13.29	11.91	11.11	11.32	13.85
14	11.21	11.89	13.07	11.22	13.09
15	13.37	13.25	11.48	13.46	11.26

7. A company packages salt pellets for in-ground well water softeners in bags with a 40-pound label weight. During a typical day's operation of the filling process, 10 samples of five bag fills are selected and measured. Using the data in the following table, create an $\bar{x}$ and an R chart to determine if the process is in control.

Bag Weights (Pounds)

	Observations				
Sample	1	2	3	4	5
1	41.41	40.63	38.83	40.57	39.94
2	40.31	40.20	41.28	40.13	42.53
3	41.64	41.02	39.49	41.64	39.14
4	40.34	39.23	41.96	42.26	40.97
5	40.77	42.73	41.83	42.59	40.43
6	42.54	42.87	40.04	40.00	40.84
7	39.84	42.65	42.86	41.75	39.68
8	42.25	40.96	39.11	41.66	39.69
9	42.94	38.52	41.50	39.13	40.57
10	39.42	39.98	38.90	41.17	41.03