1.30. The following table gives the average attendance at interleague Major League baseball games for the 1999 to 2004 seasons.

<table>
<thead>
<tr>
<th>Season</th>
<th>Average Attendance at Interleague Games</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>33,352</td>
</tr>
<tr>
<td>2000</td>
<td>33,212</td>
</tr>
<tr>
<td>2001</td>
<td>33,692</td>
</tr>
<tr>
<td>2002</td>
<td>31,921</td>
</tr>
<tr>
<td>2003</td>
<td>30,894</td>
</tr>
<tr>
<td>2004</td>
<td>32,976</td>
</tr>
</tbody>
</table>

*Source: USA TODAY, July 8, 2004.*

Describe the meaning of a variable, a measurement, and a data set with reference to this table.

1.31. The following table lists the number of Americans who took cruises during 1995 to 2004.

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Passengers (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>4.4</td>
</tr>
<tr>
<td>1996</td>
<td>4.7</td>
</tr>
<tr>
<td>1997</td>
<td>5.1</td>
</tr>
<tr>
<td>1998</td>
<td>5.4</td>
</tr>
<tr>
<td>1999</td>
<td>5.9</td>
</tr>
<tr>
<td>2000</td>
<td>6.9</td>
</tr>
<tr>
<td>2001</td>
<td>6.9</td>
</tr>
<tr>
<td>2002</td>
<td>7.6</td>
</tr>
<tr>
<td>2003</td>
<td>8.2</td>
</tr>
<tr>
<td>2004</td>
<td>9.0</td>
</tr>
</tbody>
</table>

*Source: Cruise Lines International Association. USA TODAY, February 4, 2005.*

Describe the meaning of a variable, a measurement, and a data set with reference to this table.

1.32. Refer to Exercises 1.30 and 1.31. Classify these data sets as either cross-section or time-series.

1.33. Indicate whether each of the following examples refers to a population or to a sample.
   a. A group of 25 patients selected to test a new drug
   b. Total items produced on a machine for each year from 1995 to 2005
c. Yearly expenditures on clothes for 50 persons  
d. Number of houses sold by each of the 10 employees of a real estate agency during 2005

1.34. Indicate whether each of the following examples refers to a population or to a sample.  
a. Salaries of CEOs of all companies in New York City  
b. Allowances of 1500 sixth-graders selected from Ohio  
c. Gross sales for 2005 of four fast-food chains  
d. Annual incomes of all 33 employees of a restaurant

1.35. State which of the following is an example of sampling with replacement and which is an example of sampling without replacement.  
a. Selecting 10 patients out of 100 to test a new drug  
b. Selecting one professor to be a member of the university senate and then selecting one professor from the same group to be a member of the curriculum committee

1.36. State which of the following is an example of sampling with replacement and which is an example of sampling without replacement.  
a. Selecting seven cities to market a new deodorant  
b. Selecting a high school teacher to drive students to a lecture in March, then selecting a teacher from the same group to chaperone a dance in April

1.37. The number of shoe pairs owned by six women are 8, 14, 3, 7, 10, and 5. Let \( x \) denote the number of shoe pairs owned by a woman. Find  
a. \( \sum x \)  
b. \( (\sum x)^2 \)  
c. \( \sum x^2 \)

1.38. The number of restaurants in each of five small towns is 4, 12, 8, 10, and 5. Let \( y \) denote the number of restaurants in a small town. Find  
a. \( \sum y \)  
b. \( (\sum y)^2 \)  
c. \( \sum y^2 \)

1.39. The following table lists five pairs of \( m \) and \( f \) values.  

<table>
<thead>
<tr>
<th>( m )</th>
<th>3</th>
<th>16</th>
<th>11</th>
<th>9</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f )</td>
<td>7</td>
<td>32</td>
<td>17</td>
<td>12</td>
<td>34</td>
</tr>
</tbody>
</table>

Compute the value of each of the following:  
a. \( \sum m \)  
b. \( \sum f^2 \)  
c. \( \sum mf \)  
d. \( \sum m^2f \)  
e. \( \sum m^2 \)
1.40. The following table lists six pairs of $x$ and $y$ values.

<table>
<thead>
<tr>
<th>$x$</th>
<th>7</th>
<th>11</th>
<th>8</th>
<th>4</th>
<th>14</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>5</td>
<td>15</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>19</td>
</tr>
</tbody>
</table>

Compute the value of each of the following:

a. $\Sigma y$

b. $\Sigma x^2$

c. $\Sigma xy$

d. $\Sigma x^2y$

e. $\Sigma y^2$
CHAPTER 2

2.66. The following data give the political party of each of the first 30 U.S. presidents. In the data, D stands for Democrat, DR for Democratic Republican, F for Federalist, R for Republican, and W for Whig.

F F DR DR DR D D W W
D W W D D R R R R
R D R D R R R D D R R

a. Prepare a frequency distribution table for these data.
b. Calculate the relative frequency and percentage distributions.
c. Draw a bar graph for the relative frequency distribution and a pie chart for the percentage distribution.
d. What percentage of these presidents were Whigs?

2.67. In a survey conducted by Harris Interactive for Tylenol PM, people were asked about how they cope with afternoon drowsiness (USA TODAY, October 19, 2004). Of the respondents, 35% said they drink a caffeinated beverage (C). Other responses were: taking a nap (N), going for a walk (W), or eating a sugary snack (S). Suppose that in a recent poll, 30 people were asked which one of the above choices they preferred when dealing with drowsiness. Their responses are given below.

C C N C W N S C N C
S S W C C N S N C C
N C C W W C W N C S

a. Prepare a frequency distribution table for these data.
b. Calculate the relative frequencies and percentages for all classes.
c. Draw a bar graph for the frequency distribution and a pie chart for the percentage distribution.
d. What percentage of these respondents preferred to cope with afternoon drowsiness by taking a nap?

2.68. The following data give the numbers of television sets owned by 40 randomly selected households.

1 1 2 3 2 4 1 3 2 1
3 0 2 1 2 3 2 3 2 2
1 2 1 1 1 3 1 1 1 2
2 4 2 3 1 3 1 2 2 4

a. Prepare a frequency distribution table for these data using single-valued classes.
b. Compute the relative frequency and percentage distributions.
c. Draw a bar graph for the frequency distribution.

d. What percentage of the households own two or more television sets?

2.69. Twenty-four students from universities in Connecticut were asked to name the five current members of the U.S. House of Representatives from Connecticut. The number of correct names supplied by the students are given below.

4 2 3 5 5 4 3 1 5 4 4 3
5 3 2 3 1 3 2 5 2 1 5 0

a. Prepare a frequency distribution for these data using single-valued classes.
b. Compute the relative frequency and percentage distributions.
c. What percentage of the students in this sample named fewer than two of the representatives correctly?
d. Draw a bar graph for the relative frequency distribution.

2.70. The following data give the amounts spent on video rentals (in dollars) during 2005 by 30 households randomly selected from those who rented videos in 2005.

595 24 6 100 100 40 622 405 90
55 155 760 405 90 205 70 180 88
808 100 240 127 83 310 350 160 22
111 70 15

a. Construct a frequency distribution table. Take $1 as the lower limit of the first class and $200 as the width of each class.
b. Calculate the relative frequencies and percentages for all classes.
c. What percentage of the households in this sample spent more than $400 on video rentals in 2005?

2.71. The following data give the numbers of orders received for a sample of 30 hours at the Time-saver Mail Order Company.

34 44 31 52 41 47 38 35 32 39
28 24 46 41 49 53 57 33 27 37
30 27 45 38 34 46 36 30 47 50

a. Construct a frequency distribution table. Take 23 as the lower limit of the first class and 7 as the width of each class.
b. Calculate the relative frequencies and percentages for all classes.
c. For what percentage of the hours in this sample was the number of orders more than 36?

2.72. The following data give the amounts spent (in dollars) on refreshments by 30 spectators
randomly selected from those who patronized the concession stands at a recent Major League Baseball game.

<table>
<thead>
<tr>
<th>4.95</th>
<th>27.99</th>
<th>8.00</th>
<th>5.80</th>
<th>4.50</th>
<th>2.99</th>
<th>4.85</th>
<th>6.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.00</td>
<td>15.75</td>
<td>9.50</td>
<td>3.05</td>
<td>5.65</td>
<td>21.00</td>
<td>16.60</td>
<td>18.00</td>
</tr>
<tr>
<td>21.77</td>
<td>12.35</td>
<td>7.75</td>
<td>10.45</td>
<td>3.85</td>
<td>28.45</td>
<td>8.35</td>
<td>17.70</td>
</tr>
<tr>
<td>19.50</td>
<td>11.65</td>
<td>11.45</td>
<td>3.00</td>
<td>6.55</td>
<td>16.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Construct a frequency distribution table using the *less than* method to write classes. Take $0 as the lower boundary of the first class and $6 as the width of each class.
b. Calculate the relative frequencies and percentages for all classes.
c. Draw a histogram for the frequency distribution.

2.73. The following data give the repair costs (in dollars) for 30 cars randomly selected from a list of cars that were involved in collisions.

<table>
<thead>
<tr>
<th>2300</th>
<th>750</th>
<th>2500</th>
<th>410</th>
<th>555</th>
<th>1576</th>
</tr>
</thead>
<tbody>
<tr>
<td>2460</td>
<td>1795</td>
<td>2108</td>
<td>897</td>
<td>989</td>
<td>1866</td>
</tr>
<tr>
<td>2105</td>
<td>335</td>
<td>1344</td>
<td>1159</td>
<td>1236</td>
<td>1395</td>
</tr>
<tr>
<td>6108</td>
<td>4995</td>
<td>5891</td>
<td>2309</td>
<td>3950</td>
<td>3950</td>
</tr>
<tr>
<td>6655</td>
<td>4900</td>
<td>1320</td>
<td>2901</td>
<td>1925</td>
<td>6896</td>
</tr>
</tbody>
</table>

a. Construct a frequency distribution table. Take $1 as the lower limit of the first class and $1400 as the width of each class.
b. Compute the relative frequencies and percentages for all classes.
c. Draw a histogram and a polygon for the relative frequency distribution.
d. What are the class boundaries and the width of the fourth class?

2.74. Refer to Exercise 2.70. Prepare the cumulative frequency, cumulative relative frequency, and cumulative percentage distributions by using the frequency distribution table of that exercise.

2.75. Refer to Exercise 2.71. Prepare the cumulative frequency, cumulative relative frequency, and cumulative percentage distributions using the frequency distribution table constructed for the data of that exercise.

2.76. Refer to Exercise 2.72. Prepare the cumulative frequency, cumulative relative frequency, and cumulative percentage distributions using the frequency distribution table constructed for the data of that exercise.

2.77. Construct the cumulative frequency, cumulative relative frequency, and cumulative percentage distributions by using the frequency distribution table constructed for the data of Exercise 2.73.

2.78. Refer to Exercise 2.70. Prepare a stem-and-leaf display for the data of that exercise.

2.79. Construct a stem-and-leaf display for the data given in Exercise 2.71.
2.80. The following table gives the revenues (in millions of dollars) for the seven National Hockey League teams with the largest revenues during the 2003–04 season (Forbes, November 29, 2004).

<table>
<thead>
<tr>
<th>Team</th>
<th>Revenue (millions of dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York Rangers</td>
<td>118</td>
</tr>
<tr>
<td>Toronto Maple Leafs</td>
<td>117</td>
</tr>
<tr>
<td>Philadelphia Flyers</td>
<td>106</td>
</tr>
<tr>
<td>Dallas Stars</td>
<td>103</td>
</tr>
<tr>
<td>Detroit Red Wings</td>
<td>97</td>
</tr>
<tr>
<td>Colorado Avalanche</td>
<td>99</td>
</tr>
<tr>
<td>Boston Bruins</td>
<td>95</td>
</tr>
</tbody>
</table>

Draw two bar graphs for these data, the first without truncating the axis on which revenues are marked and the second by truncating this axis. In the second graph, mark the revenues on the vertical axis starting with $90 million. Briefly comment on the two bar graphs.

2.81. The following table lists the average price per gallon for unleaded regular gasoline in the United States from 1997 to 2004. Note that the average price for 2004 is for the months from January to June only.

<table>
<thead>
<tr>
<th>Year</th>
<th>Average Price Per Gallon (dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>1.234</td>
</tr>
<tr>
<td>1998</td>
<td>1.059</td>
</tr>
<tr>
<td>1999</td>
<td>1.165</td>
</tr>
<tr>
<td>2000</td>
<td>1.510</td>
</tr>
<tr>
<td>2001</td>
<td>1.461</td>
</tr>
<tr>
<td>2002</td>
<td>1.358</td>
</tr>
<tr>
<td>2003</td>
<td>1.591</td>
</tr>
<tr>
<td>2004</td>
<td>1.819</td>
</tr>
</tbody>
</table>

*Source: Energy Information Administration.*

Draw two bar graphs for these data, the first without truncating the axis on which the gasoline prices are marked and the second by truncating this axis. In the second graph, mark the prices on the vertical axis, starting with $1.00. Briefly comment on the two bar graphs.

2.82. Reconsider the data on the times (in minutes) taken to commute from home to work for 20 workers given in Exercise 2.53. Create a dotplot for those data.

2.83. Reconsider the data on the numbers of orders received for a sample of 30 hours at the Timesaver Mail Order Company given in Exercise 2.71. Create a dotplot for those data.
2.84. Twenty-four students from a university in Oregon were asked to name the five current members of the U.S. House of Representatives from their state. The following data give the numbers of correct names given by these students.

```
5 5 1 2 4 5 3 1 5 5 0 1
2 3 5 4 3 1 5 2 5 4 5 3
```

Create a dotplot for these data.

2.85. The following data give the numbers of visitors during visiting hours on a given evening for each of the 20 randomly selected patients at a hospital.

```
3 0 1 4 2 0 4 1 1 3
4 2 0 2 2 1 1 3 0
```

Create a dotplot for these data.

Advanced Exercises

2.86. The following frequency distribution table gives the age distribution of drivers who were at fault in auto accidents that occurred during a one-week period in a city.

<table>
<thead>
<tr>
<th>Age</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 to less than 20</td>
<td>7</td>
</tr>
<tr>
<td>20 to less than 25</td>
<td>12</td>
</tr>
<tr>
<td>25 to less than 30</td>
<td>18</td>
</tr>
<tr>
<td>30 to less than 40</td>
<td>14</td>
</tr>
<tr>
<td>40 to less than 50</td>
<td>15</td>
</tr>
<tr>
<td>50 to less than 60</td>
<td>16</td>
</tr>
<tr>
<td>60 and over</td>
<td>35</td>
</tr>
</tbody>
</table>

a. Draw a relative frequency histogram for this table.
b. In what way(s) is this histogram misleading?
c. How can you change the frequency distribution so that the resulting histogram gives a clearer picture?

2.87. Refer to the data presented in Exercise 2.86. Note that there were 50% more accidents in the 25 to less than 30 age group than in the 20 to less than 25 age group. Does this suggest that the older group of drivers in this city is more accident-prone than the younger group? What other explanation might account for the difference in accident rates?

2.88. Suppose a data set contains the ages of 135 autoworkers ranging from 20 to 53.
   a. Using Sturges's formula given in footnote 1 on page 36, find an appropriate number of classes for a frequency distribution for this data set.
   b. Find an appropriate class width based on the number of classes in part a.
2.89. Statisticians often need to know the shape of a population to make inferences. Suppose that you are asked to specify the shape of the population of weights of all college students.

a. Sketch a graph of what you think the weights of all college students would look like.

b. The following data give the weights (in pounds) of a random sample of 44 college students. (Here, F and M indicate female and male.)

<table>
<thead>
<tr>
<th>Weight</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>F</td>
</tr>
<tr>
<td>195</td>
<td>M</td>
</tr>
<tr>
<td>138</td>
<td>M</td>
</tr>
<tr>
<td>115</td>
<td>F</td>
</tr>
<tr>
<td>179</td>
<td>M</td>
</tr>
<tr>
<td>119</td>
<td>F</td>
</tr>
<tr>
<td>148</td>
<td>F</td>
</tr>
<tr>
<td>147</td>
<td>F</td>
</tr>
<tr>
<td>180</td>
<td>M</td>
</tr>
<tr>
<td>146</td>
<td>F</td>
</tr>
<tr>
<td>179</td>
<td>M</td>
</tr>
<tr>
<td>189</td>
<td>M</td>
</tr>
<tr>
<td>175</td>
<td>M</td>
</tr>
<tr>
<td>108</td>
<td>F</td>
</tr>
<tr>
<td>193</td>
<td>M</td>
</tr>
<tr>
<td>114</td>
<td>F</td>
</tr>
<tr>
<td>179</td>
<td>M</td>
</tr>
<tr>
<td>147</td>
<td>M</td>
</tr>
<tr>
<td>108</td>
<td>F</td>
</tr>
<tr>
<td>128</td>
<td>F</td>
</tr>
<tr>
<td>164</td>
<td>F</td>
</tr>
<tr>
<td>174</td>
<td>M</td>
</tr>
<tr>
<td>128</td>
<td>F</td>
</tr>
<tr>
<td>159</td>
<td>M</td>
</tr>
<tr>
<td>193</td>
<td>M</td>
</tr>
<tr>
<td>204</td>
<td>M</td>
</tr>
<tr>
<td>125</td>
<td>F</td>
</tr>
<tr>
<td>133</td>
<td>F</td>
</tr>
<tr>
<td>115</td>
<td>F</td>
</tr>
<tr>
<td>168</td>
<td>M</td>
</tr>
<tr>
<td>123</td>
<td>F</td>
</tr>
<tr>
<td>183</td>
<td>M</td>
</tr>
<tr>
<td>116</td>
<td>F</td>
</tr>
<tr>
<td>182</td>
<td>M</td>
</tr>
<tr>
<td>174</td>
<td>M</td>
</tr>
<tr>
<td>102</td>
<td>F</td>
</tr>
<tr>
<td>123</td>
<td>F</td>
</tr>
<tr>
<td>99</td>
<td>F</td>
</tr>
<tr>
<td>161</td>
<td>M</td>
</tr>
<tr>
<td>162</td>
<td>M</td>
</tr>
<tr>
<td>155</td>
<td>F</td>
</tr>
<tr>
<td>202</td>
<td>M</td>
</tr>
<tr>
<td>110</td>
<td>F</td>
</tr>
<tr>
<td>132</td>
<td>M</td>
</tr>
</tbody>
</table>

i. Construct a stem-and-leaf display for these data.

ii. Can you explain why these data appear the way they do?

c. Now sketch a new graph of what you think the weights of all college students look like. Is this similar to your sketch in part a?

2.90. Consider the two histograms given in Figure 2.23 that are drawn for the same data set. In this data set, none of the values are integers.

a. What are the endpoints and widths of classes in each of the two histograms?

b. In the first histogram, of the observations that fall in the interval that is centered at 8, how many are actually between the left endpoint of that interval and 8? Note that you have to consider both histograms to answer this question.

c. Observe the leftmost bars in both histograms. Why is the leftmost bar in the first histogram misleading?
2.91. Refer to the data on weights of 44 college students given in Exercise 2.88. Create a dotplot of all 44 weights. Then create stacked dotplots for the weights of male and female students. Describe the similarities and differences in the distributions of weights of male and female students. Using all three dotplots, explain why you cannot distinguish the lightest males from the heaviest females when you consider only the dotplot of all 44 weights.

2.92. The pie chart in Figure 2.24 shows the percentage distribution of ages (i.e., the percentages of all prostate cancer patients falling in various age groups) for men who were recently diagnosed with prostate cancer.
   a. Are more or fewer than 50% of these patients in their 50s? How can you tell?
   b. Are more or fewer than 75% of these patients in their 50s and 60s? How can you tell?
   c. A reporter looks at this pie chart and says, “Look at all these 50-year-old men who are getting prostate cancer. This is a major concern for a man once he turns 50.”
Explain why the reporter cannot necessarily conclude from this pie chart that there are a lot of 50-year-old men with prostate cancer. Can you think of any other way to present these cancer cases (both graph and variable) to determine if the reporter's claim is valid?

Figure 2.24 Pie Chart of Age Groups.
CHAPTER 3

3.109. Each year the faculty at Metro Business College chooses 10 members from the current graduating class that they feel are most likely to succeed. The data below give the current annual incomes (in thousands of dollars) of the 10 members of the class of 2005 who were voted most likely to succeed.

\[ 59 \ 68 \ 44 \ 68 \ 57 \ 104 \ 56 \ 44 \ 47 \ 40 \]

a. Calculate the mean and median.

b. Does this data set contain any outlier(s)? If yes, drop the outlier(s) and recalculate the mean and median. Which of these measures changes by a greater amount when you drop the outlier(s)?

c. Is the mean or the median a better summary measure for these data? Explain.

3.110. The following data give the weights (in pounds) of the nine running backs selected for PARADE magazine's 42nd annual All-America High School Football Team (PARADE, January 23, 2005). Note that because this All-America team included only nine running backs, it can be considered the population of running backs for this team.

\[ 225 \ 225 \ 210 \ 234 \ 218 \ 188 \ 190 \ 195 \ 185 \]

a. Calculate the mean and the median. Do these data have a mode? Why or why not? Explain.

b. Find the range, variance, and standard deviation.

3.111. The following table gives the total yards gained by each of the top 10 NFL pass receivers in a single game during the 2004 regular National Football League season. Note that the games included in this data set are the ones with the highest total yards for each pass receiver during that season.

<table>
<thead>
<tr>
<th>Player</th>
<th>Yards Gained</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. Bennett</td>
<td>233</td>
</tr>
<tr>
<td>R. Smith</td>
<td>208</td>
</tr>
<tr>
<td>J. Walker</td>
<td>200</td>
</tr>
<tr>
<td>R. Wayne</td>
<td>184</td>
</tr>
<tr>
<td>M. Muhammad</td>
<td>179</td>
</tr>
<tr>
<td>T. J. Houshmandzadeh</td>
<td>171</td>
</tr>
<tr>
<td>I. Bruce</td>
<td>170</td>
</tr>
<tr>
<td>A. Johnson</td>
<td>170</td>
</tr>
<tr>
<td>R. Gardner</td>
<td>167</td>
</tr>
<tr>
<td>J. Horn</td>
<td>167</td>
</tr>
</tbody>
</table>
3.112. The following data give the numbers of driving citations received by 12 drivers.

4 8 0 3 11 7 4 14 8 13 7 9

a. Calculate the mean and median. Do these data have a mode(s)? Why or why not? Explain.
b. Find the range, variance, and standard deviation.

c. Are the values of the summary measures in parts a and b population parameters or sample statistics?

3.113. The following table gives the distribution of the amounts of rainfall (in inches) for July 2005 for 50 cities.

<table>
<thead>
<tr>
<th>Rainfall</th>
<th>Number of Cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to less than 2</td>
<td>6</td>
</tr>
<tr>
<td>2 to less than 4</td>
<td>10</td>
</tr>
<tr>
<td>4 to less than 6</td>
<td>20</td>
</tr>
<tr>
<td>6 to less than 8</td>
<td>7</td>
</tr>
<tr>
<td>8 to less than 10</td>
<td>4</td>
</tr>
<tr>
<td>10 to less than 12</td>
<td>3</td>
</tr>
</tbody>
</table>

Find the mean, variance, and standard deviation. Are the values of these summary measures population parameters or sample statistics?

3.114. The following table gives the frequency distribution of the times (in minutes) that 50 commuter students at a large university spent looking for parking spaces on the first day of classes in the Spring semester of 2006.

<table>
<thead>
<tr>
<th>Time</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to less than 4</td>
<td>1</td>
</tr>
<tr>
<td>4 to less than 8</td>
<td>7</td>
</tr>
<tr>
<td>8 to less than 12</td>
<td>15</td>
</tr>
<tr>
<td>12 to less than 16</td>
<td>18</td>
</tr>
<tr>
<td>16 to less than 20</td>
<td>6</td>
</tr>
<tr>
<td>20 to less than 24</td>
<td>3</td>
</tr>
</tbody>
</table>

Find the mean, variance, and standard deviation. Are the values of these summary measures population parameters or sample statistics?

3.115. The mean time taken to learn the basics of a word processor by all students is 200
minutes with a standard deviation of 20 minutes.

a Using Chebyshev's theorem, find at least what percentage of students will learn the basics of this word processor in
   i. 160 to 240 minutes
   ii. 140 to 260 minutes

*b Using Chebyshev's theorem, find the interval that contains the time taken by at least 75% of all students to learn this word processor.

3.116. According to the *Statistical Abstract of the United States*, Americans were expected to spend an average of 1669 hours watching television in 2004 (*USA TODAY*, March 30, 2004). Assume that the average time spent watching television by Americans this year will have a distribution that is skewed to the right with a mean of 1750 hours and a standard deviation of 450 hours.

a Using Chebyshev's theorem, find at least what percentage of Americans will watch television this year for
   i. 850 to 2650 hours
   ii. 400 to 3100 hours

*b Using Chebyshev's theorem, find the interval that will contain the television viewing times of at least 84% of all Americans.

3.117. Refer to Exercise 3.115. Suppose the times taken to learn the basics of this word processor by all students have a bell-shaped distribution with a mean of 200 minutes and a standard deviation of 20 minutes.

a Using the empirical rule, find the percentage of students who learn the basics of this word processor in
   i. 180 to 220 minutes
   ii. 160 to 240 minutes

*b Using the empirical rule, find the interval that contains the time taken by 99.7% of all students to learn this word processor.

3.118. Assume that the annual earnings of all employees with CPA certification and 12 years of experience and working for large firms have a bell-shaped distribution with a mean of $134,000 and a standard deviation of $12,000.

a Using the empirical rule, find the percentage of all such employees whose annual earnings are between
   i. $98,000 and $170,000
   ii. $110,000 and $158,000

*b Using the empirical rule, find the interval that contains the annual earnings of 68% of all such employees.

3.119. Refer to the data of Exercise 3.109 on the current annual incomes (in thousands of dollars) of the 10 members of the class of 2005 of the Metro Business College who were voted most likely to succeed.

59 68 44 68 57 104 56 44 47 40
a. Determine the values of the three quartiles and the interquartile range. Where does the value of 40 fall in relation to these quartiles?

b. Calculate the (approximate) value of the 70th percentile. Give a brief interpretation of this percentile.

c. Find the percentile rank of 47. Give a brief interpretation of this percentile rank.

3.120. Refer to the data given in Exercise 3.111 on the total yards gained by the top 10 NFL pass receivers in single games during the 2004 regular National Football League season.

a. Determine the values of the three quartiles and the interquartile range. Where does the value of 179 lie in relation to these quartiles?

b. Calculate the (approximate) value of the 70th percentile. Give a brief interpretation of this percentile.

c. Find the percentile rank of 171. Give a brief interpretation of this percentile rank.

3.121. A student washes her clothes at a laundromat once a week. The data below give the time (in minutes) she spent in the laundromat for each of 15 randomly selected weeks. Here, time spent in the laundromat includes the time spent waiting for a machine to become available.

75  62  84  73  107  81  93  72
135  77  85  67  90  83  112

Prepare a box-and-whisker plot. Is the data set skewed in any direction? If yes, is it skewed to the right or to the left? Does this data set contain any outliers?

3.122. The following data give the lengths of time (in weeks) taken to find a full-time job by 18 computer science majors who graduated in 2005 from a small college.

10  3  12  21  15  8  4  2  16
8  9  14  33  7  24  11  42  15

Make a box-and-whisker plot. Comment on the skewness of this data set. Does this data set contain any outliers?

Advanced Exercises

3.123. Melissa's grade in her math class is determined by three 100-point tests and a 200-point final exam. To determine the grade for a student in this class, the instructor will add the four scores together and divide this sum by 5 to obtain a percentage. This percentage must be at least 80 for a grade of B. If Melissa's three test scores are 75, 69, and 87, what is the minimum score she needs on the final exam to obtain a B grade?

3.124. Jeffrey is serving on a six-person jury for a personal-injury lawsuit. All six jurors want to award damages to the plaintiff but cannot agree on the amount of the award. The jurors have decided that each of them will suggest an amount that he or she thinks should be awarded; then they will use the mean of these six numbers as the award to recommend to the plaintiff.

a. Jeffrey thinks the plaintiff should receive $20,000, but he thinks the mean of the other five jurors' recommendations will be about $12,000. He decides to suggest
an inflated amount so that the mean for all six jurors is $20,000. What amount would Jeffrey have to suggest?

b. How might this jury revise its procedure to prevent a juror like Jeffrey from having an undue influence on the amount of damages to be awarded to the plaintiff?

3.125. The heights of five starting players on a basketball team have a mean of 76 inches, a median of 78 inches, and a range of 11 inches.

a. If the tallest of these five players is replaced by a substitute who is two inches taller, find the new mean, median, and range.

b. If the tallest player is replaced by a substitute who is four inches shorter, which of the new values (mean, median, range) could you determine, and what would their new values be?

3.126. On a 300-mile auto trip, Lisa averaged 52 miles per hour for the first 100 miles, 65 mph for the second 100 miles, and 58 mph for the last 100 miles.

a. How long did the 300-mile trip take?

b. Could you find Lisa's average speed for the 300-mile trip by calculating \(\frac{52 + 65 + 58}{3}\)? If not, find the correct average speed for the trip.

3.127. A small country bought oil from three different sources in one week, as shown in the following table.

<table>
<thead>
<tr>
<th>Source</th>
<th>Barrels Purchased</th>
<th>Price Per Barrel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>1000</td>
<td>$51</td>
</tr>
<tr>
<td>Kuwait</td>
<td>200</td>
<td>64</td>
</tr>
<tr>
<td>Spot Market</td>
<td>100</td>
<td>70</td>
</tr>
</tbody>
</table>

Find the mean price per barrel for all 1300 barrels of oil purchased in that week.

3.128. During the 2004 winter season, a homeowner received four deliveries of heating oil, as shown in the following table.

<table>
<thead>
<tr>
<th>Gallons Purchased</th>
<th>Price Per Gallon</th>
</tr>
</thead>
<tbody>
<tr>
<td>198</td>
<td>$1.10</td>
</tr>
<tr>
<td>173</td>
<td>1.25</td>
</tr>
<tr>
<td>130</td>
<td>1.28</td>
</tr>
<tr>
<td>124</td>
<td>1.33</td>
</tr>
</tbody>
</table>

The homeowner claimed that the mean price he paid for oil during the season was \(\frac{1.10 + 1.25 + 1.28 + 1.33}{4} = $1.24\) per gallon. Do you agree with this claim? If not, explain why this method of calculating the mean is not appropriate in this case. Find the correct value of the mean price.

3.129. In the Olympic Games, when events require a subjective judgment of an athlete's performance, the highest and lowest of the judges' scores may be dropped. Consider a
gymnast whose performance is judged by seven judges and the highest and the lowest of the seven scores are dropped.

a. Gymnast A's scores in this event are 9.4, 9.7, 9.5, 9.5, 9.4, 9.6, and 9.5. Find this gymnast's mean score after dropping the highest and the lowest scores.

b. The answer to part a is an example of what percentage of trimmed mean?

c. Write another set of scores for a gymnast B so that gymnast A has a higher mean score than gymnast B based on the trimmed mean, but gymnast B would win if all seven scores were counted. Do not use any scores lower than 9.0.

3.130. A survey of young people's shopping habits in a small city during the summer months of 2005 showed the following: Shoppers aged 12–14 took an average of 8 shopping trips per month and spent an average of $14 per trip. Shoppers aged 15–17 took an average of 11 trips per month and spent an average of $18 per trip. Assume that this city has 1100 shoppers aged 12–14 and 900 shoppers aged 15–17.

a. Find the total amount spent per month by all these 2000 shoppers in both age groups.

b. Find the mean number of shopping trips per person per month for these 2000 shoppers.

c. Find the mean amount spent per person per month by shoppers aged 12–17 in this city.

3.131. The following table shows the total population and the number of deaths (in thousands) due to heart attack for two age groups in Countries A and B for 2005.

<table>
<thead>
<tr>
<th>Age 30 and Under</th>
<th>Age 31 and Over</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>A</td>
</tr>
<tr>
<td>40,000</td>
<td>25,000</td>
</tr>
<tr>
<td>Deaths due to heart attack</td>
<td>1000</td>
</tr>
</tbody>
</table>

a. Calculate the death rate due to heart attack per 1000 population for the 30 and under age group for each of the two countries. Which country has the lower death rate in this age group?

b. Calculate the death rates due to heart attack for the two countries for the 31 and over age group. Which country has the lower death rate in this age group?

c. Calculate the death rate due to heart attack for the entire population of Country A; then do the same for Country B. Which country has the lower overall death rate?

d. How can the country with lower death rate in both age groups have the higher overall death rate? (This phenomenon is known as Simpson's paradox.)

3.132. In a study of distances traveled to a college by commuting students, data from 100 commuters yielded a mean of 8.73 miles. After the mean was calculated, data came in late from three students, with distances of 11.5, 7.6, and 10.0 miles. Calculate the mean distance for all 103 students.

3.133. The test scores for a large statistics class have an unknown distribution with a mean of
70 and a standard deviation of 10.

a. Find $k$ so that at least 50% of the scores are within $k$ standard deviations of the mean.

b. Find $k$ so that at most 10% of the scores are more than $k$ standard deviations above the mean.

3.134. The test scores for a very large statistics class have a bell-shaped distribution with a mean of 70 points.

a. If 16% of all students in the class scored above 85, what is the standard deviation of the scores?

b. If 95% of the scores are between 60 and 80, what is the standard deviation?

3.135. How much does the typical American family spend to go away on vacation each year? Twenty-five randomly selected households reported the following vacation expenditures (rounded to the nearest hundred dollars) during the past year:

<table>
<thead>
<tr>
<th>2500</th>
<th>500</th>
<th>800</th>
<th>0</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>200</td>
<td>2200</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>900</td>
<td>321,500</td>
<td>400</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>0</td>
<td>8200</td>
<td>900</td>
</tr>
<tr>
<td>0</td>
<td>1700</td>
<td>1100</td>
<td>600</td>
<td>3400</td>
</tr>
</tbody>
</table>

a. Using both graphical and numerical methods, organize and interpret these data.

b. What measure of central tendency best answers the original question?

3.136. Actuaries at an insurance company must determine a premium for a new type of insurance. A random sample of 40 potential purchasers of this type of insurance were found to have suffered the following values of losses during the past year. These losses would have been covered by the insurance if it were available.

<table>
<thead>
<tr>
<th>100</th>
<th>32</th>
<th>0</th>
<th>0</th>
<th>470</th>
<th>50</th>
<th>0</th>
<th>14,589</th>
<th>212</th>
<th>93</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1127</td>
<td>421</td>
<td>0</td>
<td>87</td>
<td>135</td>
<td>420</td>
<td>0</td>
<td>250</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>309</td>
<td>0</td>
<td>177</td>
<td>295</td>
<td>501</td>
<td>0</td>
<td>143</td>
<td>0</td>
</tr>
<tr>
<td>167</td>
<td>398</td>
<td>54</td>
<td>0</td>
<td>141</td>
<td>0</td>
<td>3709</td>
<td>122</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

a. Find the mean, median, and mode of these 40 losses.

b. Which of the mean, median, or mode is largest?

c. Draw a box-and-whisker plot for these data, and describe the skewness, if any.

d. Which measure of central tendency should the actuaries use to determine the premium for this insurance?

3.137. A local golf club has men's and women's summer leagues. The following data give the scores for a round of 18 holes of golf for 17 men and 15 women randomly selected from their respective leagues.
a. Make a box-and-whisker plot for each of the data sets and use them to discuss the similarities and differences between the scores of the men and women golfers.

b. Compute the various descriptive measures you have learned for each sample. How do they compare?

3.138. Answer the following questions.
   a. The total weight of all pieces of luggage loaded onto an airplane is 12,372 pounds, which works out to be an average of 51.55 pounds per piece. How many pieces of luggage are on the plane?
   b. A group of seven friends, having just gotten back a chemistry exam, discuss their scores. Six of the students reveal that they received grades of 81, 75, 93, 88, 82, and 85, but the seventh student is reluctant to say what grade she received. After some calculation she announces that the group averaged 81 on the exam. What is her score?

3.139. Suppose that there are 150 freshmen engineering majors at a college and each of them will take the same five courses next semester. Four of these courses will be taught in small sections of 25 students each, whereas the fifth course will be taught in one section containing all 150 freshmen. To accommodate all 150 students, there must be six sections of each of the four courses taught in 25-student sections. Thus, there are 24 classes of 25 students each and one class of 150 students.
   a. Find the mean size of these 25 classes.
   b. Find the mean class size from a student's point of view, noting that each student has five classes containing 25, 25, 25, 25, and 150 students.
   Are the means in parts a and b equal? If not, why not?

3.140. The following data give the weights (in pounds) of a random sample of 44 college students. (Here F and M indicate female and male, respectively.)

<table>
<thead>
<tr>
<th>Men</th>
<th>87</th>
<th>68</th>
<th>92</th>
<th>79</th>
<th>83</th>
<th>67</th>
<th>71</th>
<th>92</th>
<th>112</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>75</td>
<td>77</td>
<td>102</td>
<td>79</td>
<td>78</td>
<td>85</td>
<td>75</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>101</td>
<td>100</td>
<td>87</td>
<td>95</td>
<td>98</td>
<td>117</td>
<td>107</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td></td>
<td>97</td>
<td>90</td>
<td>100</td>
<td>94</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 123 F| 195 M| 138 M| 115 F| 179 M| 119 F|
| 148 F| 147 F| 180 M| 146 F| 179 M| 189 M|
| 175 M| 108 F| 193 M| 114 F| 179 M| 147 M|
| 108 F| 128 F| 164 F| 174 M| 128 F| 159 M|
| 193 M| 204 M| 125 F| 133 F| 115 F| 168 M|
| 123 F| 183 M| 116 F| 182 M| 174 M| 102 F|
| 123 F| 99 F | 161 M| 162 M| 155 F| 202 M|
| 110 F| 132 M|     |     |     |     |
Compute the mean, median, and standard deviation for the weights of all students, of men only, and of women only. Of the mean and median, which is the more informative measure of central tendency? Write a brief note comparing the three measures for all students, men only, and women only.

3.141. The distribution of the lengths of fish in a certain lake is not known, but it is definitely not bell-shaped. It is estimated that the mean length is 6 inches with a standard deviation of 2 inches.
   a. At least what proportion of fish in the lake are between 3 inches and 9 inches long?
   b. What is the smallest interval that will contain the lengths of at least 84% of the fish?
   c. Find an interval so that fewer than 36% of the fish have lengths outside this interval.

3.142. The following stem-and-leaf diagram gives the distances (in thousands of miles) driven during the past year by a sample of drivers in a city.

   | 0 | 3 6 9 |
   | 1 | 2 8 5 1 0 5 |
   | 2 | 5 1 6 |
   | 3 | 8 |
   | 4 | 1 |
   | 5 | |
   | 6 | 2 |

   a. Compute the sample mean, median, and mode for the data on distances driven.
   b. Compute the range, variance, and standard deviation for these data.
   c. Compute the first and third quartiles.
   d. Compute the interquartile range. Describe what properties the interquartile range has. When would it be preferable to using the standard deviation when measuring variation?

3.143. Refer to the data in Problem 3.140. Two individuals, one from Canada and one from England, are interested in your analysis of these data but they need your results in different units. The Canadian individual wants the results in grams (1 pound = 435.59 grams), while the English individual wants the results in stone (1 stone = 14 pounds).
   a. Convert the data on weights from pounds to grams, and then recalculate the mean, median, and standard deviation of weight for males and females separately. Repeat the procedure, changing the unit from pounds to stones.
   b. Convert your answers from Problem 3.140 to grams and stone. What do you notice about these answers and your answers from part a?
   c. What happens to the values of the mean, median, and standard deviation when you convert from a larger unit to a smaller unit (e.g., from pounds to grams)? Does the same thing happen if you convert from a smaller unit (e.g., pounds) to a larger unit (e.g., stone)?
d. Figure 3.15 on the next page gives a stacked dotplot of these weights in pounds and stone. Which of these two distributions has more variability? Use your results from parts a to c to explain why this is the case.

e. Now consider the weights in pounds and grams. Make a stacked dotplot for these data and answer part d.

![Stacked Dotplot of Weights in Stone and Pounds](image)

3.144. Although the standard workweek is 40 hours a week, many people work a lot more than 40 hours a week. The data on the next page give the numbers of hours worked last week by 50 people.

40.5  41.3  41.4  41.5  42.0  42.2  42.4  42.4  42.6  43.3  
43.7  43.9  45.0  45.0  45.2  45.8  45.9  46.2  47.2  47.5  
47.8  48.2  48.3  48.8  49.0  49.2  49.9  50.1  50.6  50.6  
50.8  51.5  51.5  52.3  52.3  52.6  52.7  52.7  53.4  53.9  
54.4  54.8  55.0  55.4  55.4  55.4  56.2  56.3  57.8  58.7  

a. The sample mean and sample standard deviation for this data set are 49.012 and 5.080, respectively. Using the Chebyshev's theorem, calculate the intervals that contain at least 75%, 88.89%, and 93.75% of the data.

b. Determine the actual percentages of the given data values that fall in each of the intervals that you calculated in part a. Also calculate the percentage of the data values that fall within one standard deviation of the mean.

c. Do you think the lower endpoints provided by Chebyshev's Theorem in part a are useful for this problem? Explain your answer.
d. Suppose that the individual with the first number (54.4) in the fifth row of the data is a workaholic who actually worked 84.4 hours last week, and not 54.4 hours. With this change now $x = 49.61$ and $s = 7.10$. Recalculate the intervals for part a and the actual percentages for part b. Did your percentages change a lot or a little?

e. How many standard deviations above the mean would you have to go to capture all 50 data values? What is the lower bound for the percentage of the data that should fall in the interval, according to Chebyshev?

3.145. Refer to the women's golf scores in Exercise 3.137. It turns out that 117 was mistakenly entered. Although this person still had the highest score among the 15 women, her score was not a mild or extreme outlier according to the box-and-whisker plot, nor was she tied for the highest score. What are the possible scores that she could have shot?
CHAPTER 4

4.124. A car rental agency currently has 44 cars available, 18 of which have a GPS navigation system. One of the 44 cars is selected at random. Find the probability that this car
   a. has a GPS navigation system
   b. does not have a GPS navigation system

4.125. In a class of 35 students, 13 are seniors, 9 are juniors, 8 are sophomores, and 5 are freshmen. If one student is selected at random from this class, what is the probability that this student is
   a. a junior?
   b. a freshman?

4.126. A random sample of 250 juniors majoring in psychology or communications at a large university is selected. These students are asked whether or not they are happy with their majors. The following table gives the results of the survey. Assume that none of these 250 students is majoring in both areas.

<table>
<thead>
<tr>
<th></th>
<th>Happy</th>
<th>Unhappy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>Communications</td>
<td>115</td>
<td>35</td>
</tr>
</tbody>
</table>

a. If one student is selected at random from this group, find the probability that this student is
   i. happy with the choice of major
   ii. a psychology major
   iii. a communications major given that the student is happy with the choice of major
   iv. unhappy with the choice of major given that the student is a psychology major
   v. a psychology major and is happy with that major
   vi. a communications major or is unhappy with his or her major

b. Are the events “psychology major” and “happy with major” independent? Are they mutually exclusive? Explain why or why not.

4.127. A random sample of 250 adults was taken, and they were asked whether they prefer watching sports or opera on television. The following table gives the two-way classification of these adults.

<table>
<thead>
<tr>
<th>Prefer Watching Sports</th>
<th>Prefer Watching Opera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>96</td>
</tr>
<tr>
<td>Female</td>
<td>45</td>
</tr>
</tbody>
</table>

a. If one adult is selected at random from this group, find the probability that this
adult
  i. prefers watching opera
  ii. is a male
  iii. prefers watching sports given that the adult is a female
  iv. is a male given that he prefers watching sports
  v. is a female and prefers watching opera
  vi. prefers watching sports or is a male

b. Are the events “female” and “prefers watching sports” independent? Are they mutually exclusive? Explain why or why not.

4.128. A random sample of 80 lawyers was taken, and they were asked if they are in favor of or against capital punishment. The following table gives the two-way classification of their responses.

<table>
<thead>
<tr>
<th></th>
<th>Favors Capital Punishment</th>
<th>Opposes Capital Punishment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>32</td>
<td>24</td>
</tr>
<tr>
<td>Female</td>
<td>13</td>
<td>11</td>
</tr>
</tbody>
</table>

a. If one lawyer is randomly selected from this group, find the probability that this lawyer
  i. favors capital punishment
  ii. is a female
  iii. opposes capital punishment given that the lawyer is a female
  iv. is a male given that he favors capital punishment
  v. is a female and favors capital punishment
  vi. opposes capital punishment or is a male

b. Are the events “female” and “opposes capital punishment” independent? Are they mutually exclusive? Explain why or why not.

4.129. A random sample of 400 college students was asked if college athletes should be paid. The following table gives a two-way classification of the responses.

<table>
<thead>
<tr>
<th></th>
<th>Should Be Paid</th>
<th>Should Not Be Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student athlete</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>Student nonathlete</td>
<td>210</td>
<td>90</td>
</tr>
</tbody>
</table>

a. If one student is randomly selected from these 400 students, find the probability that this student
  i. is in favor of paying college athletes
  ii. favors paying college athletes given that the student selected is a nonathlete
  iii. is an athlete and favors paying student athletes
iv. is a nonathlete or is against paying student athletes

b. Are the events “student athlete” and “should be paid” independent? Are they mutually exclusive? Explain why or why not.

4.130. An appliance repair company that makes service calls to customers' homes has found that 5% of the time there is nothing wrong with the appliance and the problem is due to customer error (appliance unplugged, controls improperly set, etc.). Two service calls are selected at random, and it is observed whether or not the problem is due to customer error. Draw a tree diagram. Find the probability that in this sample of two service calls
a. both problems are due to customer error
b. at least one problem is not due to customer error

4.131. According to data from Watson Wyatt Worldwide, 51% of employees have confidence in their senior management (Business Week, January 24, 2005). Assume that the other 49% do not have confidence in their senior management. Further assume that these percentages are true for the current population of all employees. Two employees are selected at random and asked whether or not they have confidence in their senior management. Draw a tree diagram for this problem. Find the probability that in this sample of two employees
a. both have confidence in their senior management
b. at most one has confidence in his or her senior management

4.132. Refer to Exercise 4.124. Two cars are selected at random from these 44 cars. Find the probability that both of these cars have GPS navigation systems.

4.133. Refer to Exercise 4.125. Two students are selected at random from this class of 35 students. Find the probability that the first student selected is a junior and the second is a sophomore.

4.134. A company has installed a generator to back up the power in case there is a power failure. The probability that there will be a power failure during a snowstorm is .30. The probability that the generator will stop working during a snowstorm is .09. What is the probability that during a snowstorm the company will lose both sources of power?

4.135. Terry & Sons makes bearings for autos. The production system involves two independent processing machines so that each bearing passes through these two processes. The probability that the first processing machine is not working properly at any time is .08, and the probability that the second machine is not working properly at any time is .06. Find the probability that both machines will not be working properly at any given time.

Advanced Exercises

4.136. A player plays a roulette game in a casino by betting on a single number each time. Because the wheel has 38 numbers, the probability that the player will win in a single play is 1/38. Note that each play of the game is independent of all previous plays.
a. Find the probability that the player will win for the first time on the 10th play.
b. Find the probability that it takes the player more than 50 plays to win for the first time.
c. The gambler claims that because he has 1 chance in 38 of winning each time he plays, he is certain to win at least once if he plays 38 times. Does this sound reasonable to you? Find the probability that he will win at least once in 38 plays.

4.137. A certain state's auto license plates have three letters of the alphabet followed by a three-digit number.
   a. How many different license plates are possible if all three-letter sequences are permitted and any number from 000 to 999 is allowed?
   b. Arnold witnessed a hit-and-run accident. He knows that the first letter on the license plate of the offender's car was a B, that the second letter was an O or a Q, and that the last number was a 5. How many of this state's license plates fit this description?

4.138. The median life of Brand LT5 batteries is 100 hours. What is the probability that in a set of three such batteries, exactly two will last longer than 100 hours?

4.139. Powerball is a game of chance that has generated intense interest because of its large jackpots. To play this game, a player selects five different numbers from 1 through 53, and then picks a powerball number from 1 through 42. The lottery organization randomly draws five different white balls from 53 balls numbered 1 through 53, and then randomly picks a powerball number from 1 through 42. Note that it is possible for the powerball number to be the same as one of the first five numbers.
   a. If the player's first five numbers match the numbers on the five white balls drawn by the lottery organization and the player's powerball number matches the powerball number drawn by the lottery organization, the player wins the jackpot. Find the probability that a player who buys one ticket will win the jackpot. (Note that the order in which the five white balls are drawn is unimportant.)
   b. If the player's first five numbers match the numbers on the five white balls drawn by the lottery organization but the powerball number does not match the one drawn by the lottery organization, the player wins about $100,000 (or less if several winners must share the prize pool). Find the probability that a player who buys one ticket will win this prize.

4.140. A trimotor plane has three engines—a central engine and an engine on each wing. The plane will crash only if the central engine fails and at least one of the two wing engines fails. The probability of failure during any given flight is .005 for the central engine and .008 for each of the wing engines. Assuming that the three engines operate independently, what is the probability that the plane will crash during a flight?

4.141. A box contains 10 red marbles and 10 green marbles.
   a. Sampling at random from the box five times with replacement, you have drawn a red marble all five times. What is the probability of drawing a red marble the sixth time?
   b. Sampling at random from the box five times without replacement, you have drawn a red marble all five times. Without replacing any of the marbles, what is the probability of drawing a red marble the sixth time?
   c. You have tossed a fair coin five times and have obtained heads all five times. A friend argues that according to the law of averages, a tail is due to occur and,
hence, the probability of obtaining a head on the sixth toss is less than .50. Is he right? Is coin tossing mathematically equivalent to the procedure mentioned in part a or the procedure mentioned in part b? Explain.

4.142. A gambler has four cards—two diamonds and two clubs. The gambler proposes the following game to you: You will leave the room and the gambler will put the cards face down on a table. When you return to the room, you will pick two cards at random. You will win $10 if both cards are diamonds, you will win $10 if both are clubs, and for any other outcome you will lose $10. Assuming that there is no cheating, should you accept this proposition? Support your answer by calculating your probability of winning $10.

4.143. A thief has stolen Roger's automatic teller card. The card has a four-digit personal identification number (PIN). The thief knows that the first two digits are 3 and 5, but he does not know the last two digits. Thus, the PIN could be any number from 3500 to 3599. To protect the customer, the automatic teller machine will not allow more than three unsuccessful attempts to enter the PIN. After the third wrong PIN, the machine keeps the card and allows no further attempts.

a. What is the probability that the thief will find the correct PIN within three tries? (Assume that the thief will not try the same wrong PIN twice.)

b. If the thief knew that the first two digits were 3 and 5 and that the third digit was either 1 or 7, what is the probability of guessing the correct PIN in three attempts?

4.144. Consider the following games with two dice.

a. A gambler is going to roll a die four times. If he rolls at least one 6, you must pay him $5. If he fails to roll a 6 in four tries, he will pay you $5. Find the probability that you must pay the gambler. Assume that there is no cheating.

b. The same gambler offers to let you roll a pair of dice 24 times. If you roll at least one double 6, he will pay you $10. If you fail to roll a double 6 in 24 tries, you will pay him $10. The gambler says that you have a better chance of winning because your probability of success on each of the 24 rolls is 1/36 and you have 24 chances. Thus, he says, your probability of winning $10 is 24(1/36) = 2/3. Do you agree with this analysis? If so, indicate why. If not, point out the fallacy in his argument, and then find the correct probability that you will win.

4.145. A gambler has given you two jars and 20 marbles. Of these 20 marbles, ten are red and ten are green. You must put all 20 marbles in these two jars in such a way that each jar must have at least one marble in it. Then a friend of yours, who is blindfolded, will select one of the two jars at random and then will randomly select a marble from this jar. If the selected marble is red, you and your friend win $100.

a. If you put five red marbles and five green marbles in each jar, what is the probability that your friend selects a red marble?

b. If you put two red marbles and two green marbles in one jar and the remaining marbles in the other jar, what is the probability that your friend selects a red marble?

c. How should these 20 marbles be distributed among the two jars in order to give your friend the highest possible probability of selecting a red marble?

4.146. A screening test for a certain disease is prone to giving false positives or false negatives.
If a patient being tested has the disease, the probability that the test indicates a (false) negative is .13. If the patient does not have the disease, the probability that the test indicates a (false) positive is .10. Assume that 3% of the patients being tested actually have the disease. Suppose that one patient is chosen at random and tested. Find the probability that
   a. this patient has the disease and tests positive
   b. this patient does not have the disease and tests positive
   c. this patient tests positive
   d. this patient has the disease given that he/she tests positive

**Hint:** (A tree diagram may be helpful in part c.)

4.147. A pizza parlor has 12 different toppings available for its pizzas, and 2 of these toppings are pepperoni and anchovies. If a customer picks two toppings at random, find the probability that
   a. neither topping is anchovies
   b. pepperoni is one of the toppings

4.148. An insurance company has information that 93% of its auto policy holders carry collision coverage or uninsured motorist coverage on their policies. Eighty percent of the policy holders carry collision coverage, and 60% have uninsured motorist coverage.
   a. What percentage of these policy holders carry both collision and uninsured motorist coverage?
   b. What percentage of these policy holders carry neither collision nor uninsured motorist coverage?
   c. What percentage of these policy holders carry collision but not uninsured motorist coverage?

4.149. Many states have a lottery game, usually called a Pick-4, in which you pick a four-digit number such as 7359. During the lottery drawing, there are four bins, each containing balls numbered 0 through 9. One ball is drawn from each bin to form the four-digit winning number.
   a. You purchase one ticket with one four-digit number. What is the probability that you will win this lottery game?
   b. There are many variations of this game. The primary variation allows you to win if the four digits in your number are selected in any order as long as they are the same four digits as obtained by the lottery agency. For example, if you pick four digits making the number 1265, then you will win if 1265, 2615, 5216, 6521, and so forth, are drawn. The variations of the lottery game depend on how many unique digits are in your number. Consider the following four different versions of this game.
      i. All four digits are unique (e.g., 1234)
      ii. Exactly one of the digits appears twice (e.g., 1223 or 9095)
      iii. Two digits each appear twice (e.g., 2121 or 5588)
      iv. One digit appears three times (e.g., 3335 or 2722)
Find the probability that you will win this lottery in each of these four situations.

4.150. A restaurant chain is planning to purchase 100 ovens from a manufacturer provided that these ovens pass a detailed inspection. Because of high inspection costs, five ovens are selected at random for inspection. These 100 ovens will be purchased if at most one of the five selected ovens fails inspection. Suppose that there are eight defective ovens in this batch of 100 ovens. Find the probability that the batch of ovens is purchased. (Note: In Chapter 5 you will learn another method to solve this problem.)

4.151. A production system has two production lines; each production line performs a two-part process; and each process is completed by a different machine. Thus, there are four machines, which we can identify as two first-level machines and two second-level machines. Each of the first-level machines works properly 98% of the time, and each of the second-level machines works properly 96% of the time. All four machines are independent in regard to working properly or breaking down. Two products enter this production system, one in each production line.
   a. Find the probability that both products successfully complete the two-part process (i.e., all four machines are working properly).
   b. Find the probability that neither product successfully completes the two-part process (i.e., at least one of the machines in each production line is not working properly).
CHAPTER 5

5.2. Classify each of the following random variables as discrete or continuous.
   a. The time left on a parking meter
   b. The number of bats broken by a major league baseball team in a season
   c. The number of fish in a pond
   d. The total pounds of fish caught on a fishing trip
   e. The number of gumballs in a vending machine
   f. The time spent by a physician examining a patient

5.9. Each of the following tables lists certain values of \( x \) and their probabilities. Determine whether or not each one satisfies the two conditions required for a valid probability distribution.
   a.
   \[
   \begin{array}{c|c}
   x & P(x) \\
   \hline
   5 & .36 \\
   6 & .48 \\
   7 & .62 \\
   8 & .26 \\
   \end{array}
   \]
   b.
   \[
   \begin{array}{c|c}
   x & P(x) \\
   \hline
   1 & .27 \\
   2 & .24 \\
   3 & .49 \\
   \end{array}
   \]

5.17. According to data from the Pew Internet & American Life Project, 45% of American adults take prescription drugs regularly (Time, October 25, 2004). Assume that this result holds true for the current population of all adults. Suppose that two adults are selected at random. Let \( x \) denote the number of adults in this sample who take prescription drugs regularly. Construct the probability distribution table of \( x \). Draw a tree diagram for this problem.

5.21. In a group of 20 athletes, 6 have used performance-enhancing drugs that are illegal. Suppose that two athletes are randomly selected from this group. Let \( x \) denote the number of athletes in this sample who have used such illegal drugs. Write the probability distribution of \( x \). You may draw a tree diagram and use that to write the probability distribution. (Hint: Note that the draws are made without replacement from a small population. Hence, the probabilities of outcomes do not remain constant for each draw.)
5.36. An instant lottery ticket costs $2. Out of a total of 10,000 tickets printed for this lottery, 1000 tickets contain a prize of $5 each, 100 tickets have a prize of $10 each, 5 tickets have a prize of $1000 each, and 1 ticket has a prize of $5000. Let \( x \) be the random variable that denotes the net amount a player wins by playing this lottery. Write the probability distribution of \( x \). Determine the mean and standard deviation of \( x \). How will you interpret the values of the mean and standard deviation of \( x \)?

5.38. Refer to the probability distribution you developed in Exercise 5.21 for the number of athletes in a random sample of two who have used illegal performance-enhancing drugs. Calculate the mean and standard deviation of \( x \) for that distribution.

5.88. On average, 12.5 rooms stay vacant per day at a large hotel in a city. Find the probability that on a given day exactly three rooms will be vacant. Use the Poisson formula.

5.96. An average of .8 accidents occur per day in a large city.
   a. Find the probability that no accident will occur in this city on a given day.
   b. Let \( x \) denote the number of accidents that will occur in this city on a given day. Write the probability distribution of \( x \).
   c. Find the mean, variance, and standard deviation of the probability distribution developed in part b.
CHAPTER 6

6.11. For the standard normal distribution, find the area within 1.5 standard deviations of the mean—that is, the area between \( \mu - 1.5\sigma \) and \( \mu + 1.5\sigma \).

6.30. Find the following areas under a normal distribution curve with \( \mu = 12 \) and \( \sigma = 2 \).
   a. Area between \( x = 7.76 \) and \( x = 12 \)
   b. Area between \( x = 14.48 \) and \( x = 16.54 \)
   c. Area from \( x = 8.22 \) to \( x = 10.06 \)

6.36. Let \( x \) be a continuous random variable that is normally distributed with a mean of 65 and a standard deviation of 15. Find the probability that \( x \) assumes a value
   a. less than 43
   b. greater than 74
   c. greater than 56
   d. less than 71

6.43. According to the New York Times, working men spend an average of 48 minutes per day caring for their families (Time, September 27, 2004). Assume that the times that working men currently spend per day caring for their families are normally distributed with a mean of 48 minutes and a standard deviation of 11 minutes.
   a. Find the probability that a randomly selected working man spends more than 68 minutes per day caring for his family.
   b. What percentage of working men spend between 30 and 73 minutes per day caring for their families?

6.49. According to the Kaiser Family Foundation, the average amount of time spent watching TV by 8-to-18-year-olds is 231 minutes per day (USA TODAY, March 10, 2005). Suppose currently the times spent watching TV by all 8-to-18-year-olds have a normal distribution with a mean of 231 minutes and a standard deviation of 45 minutes. What percentage of the 8-to-18-year-olds watch TV for
   a. more than 290 minutes per day?
   b. less than 150 minutes per day?
   c. 180 to 320 minutes per day?
   d. 270 to 350 minutes per day?
6.52. The pucks used by the National Hockey League for ice hockey must weigh between 5.5 and 6.0 ounces. Suppose the weights of pucks produced at a factory are normally distributed with a mean of 5.75 ounces and a standard deviation of .11 ounces. What percentage of the pucks produced at this factory cannot be used by the National Hockey League?

6.61. According to the records of an electric company serving the Boston area, the mean electric consumption during winter for all households is 1650 kilowatt-hours per month. Assume that the monthly electric consumptions during winter by all households in this area have a normal distribution with a mean of 1650 kilowatt-hours and a standard deviation of 320 kilowatt-hours. The company sent a notice to Bill Johnson informing him that about 90% of the households use less electricity per month than he does. What is Bill Johnson's monthly electric consumption?
CHAPTER 7

7.12. How does the value of $\sigma \bar{x}$ change as the sample size increases? Explain.

7.17. For a population, $\mu = 125$ and $\sigma = 36$.
   a. For a sample selected from this population, $\mu \bar{x} = 125$ and $\sigma \bar{x} = 3.6$. Find the sample size. Assume $n/N \leq .05$.
   b. For a sample selected from this population, $\mu \bar{x} = 125$ and $\sigma \bar{x} = 2.25$. Find the sample size. Assume $n/N \leq .05$.

7.23. Suppose the standard deviation of recruiting costs per player for all female basketball players recruited by all public universities in the Midwest is $605. Let x be the mean recruiting cost for a sample of a certain number of such players. What sample size will give the standard deviation of $x$ equal to $55$?

7.45. Let $x$ be a continuous random variable that has a normal distribution with $\mu = 48$ and $\sigma = 8$. Assuming $n/N \leq .05$, find the probability that the sample mean, $\bar{x}$, for a random sample of 16 taken from this population will be
   a. between 49.6 and 52.2
   b. more than 45.7

7.54. According to the Kaiser Family Foundation, children aged 8 to 18 in the United States spend an average of 62 minutes per day using computers, not counting the time they spend using computers to do homework (Time, March 21, 2005). Suppose the probability distribution of such times for all children aged 8 to 18 is skewed to the right with a mean of 62 minutes and a standard deviation of 14 minutes. Find the probability that the mean time spent per day using computers by a random sample of 400 children aged 8 to 18 is
   a. between 60.5 and 63 minutes
   b. within 2 minutes of the population mean
   c. greater than the population mean by 1.5 minutes or more

7.81. According to a Newsweek poll, 17% of Americans believe that the end of the world would occur in their lifetime (Newsweek, May 24, 2004). Suppose that this percentage is true for the current population of Americans. Let $\hat{p}$ be the proportion of Americans in a random sample of 60 who hold this opinion. Find the mean and standard deviation of $\hat{p}$ and describe the shape of its sampling distribution.
CHAPTER 8

8.16. The standard deviation for a population is $\sigma = 7.14$. A random sample selected from this population gave a mean equal to 48.52.
   a. Make a 95% confidence interval for $\mu$ assuming $n = 196$.
   b. Construct a 95% confidence interval for $\mu$ assuming $n = 100$.
   c. Determine a 95% confidence interval for $\mu$ assuming $n = 49$.
   d. Does the width of the confidence intervals constructed in parts a through c increase as the sample size decreases? Explain.

8.25. Computer Action Company sells computers and computer parts by mail. The company assures its customers that products are mailed as soon as possible after an order is placed with the company. A sample of 25 recent orders showed that the mean time taken to mail products for these orders was 70 hours. Suppose the population standard deviation is 16 hours and the population distribution is normal.
   a. Construct a 95% confidence interval for the mean time taken to mail products for all orders received at the office of this company.
   b. Explain why we need to make the confidence interval. Why can we not say that the mean time taken to mail products for all orders received at the office of this company is 70 hours?

*8.35. You are interested in estimating the mean age of cars owned by all people in the United States. Briefly explain the procedure you will follow to conduct this study. Collect the required data on a sample of 30 or more cars and then estimate the population mean at a 95% confidence level. Assume that the population standard deviation is 2.4 years.

8.49. Suppose, for a sample selected from a population, $x = 25.5$ and $s = 4.9$.
   a. Construct a 95% confidence interval for $\mu$ assuming $n = 47$.
   b. Construct a 99% confidence interval for $\mu$ assuming $n = 47$. Is the width of the 99% confidence interval larger than the width of the 95% confidence interval calculated in part a? If yes, explain why.
   c. Find a 95% confidence interval for $\mu$ assuming $n = 32$. Is the width of the 95% confidence interval for $\mu$ with $n = 32$ larger than the width of the 95% confidence interval for $\mu$ with $n = 47$ calculated in part a? If so, why? Explain.

8.56. The high cost of health care is a matter of major concern for a large number of families. A random sample of 25 families selected from an area showed that they spend an average of
$143 per month on health care with a standard deviation of $28. Make a 98% confidence interval for the mean health care expenditure per month incurred by all families in this area. Assume that the monthly health care expenditures of all families in this area have a normal distribution.
9.1. Briefly explain the meaning of each of the following terms.
   a. Null hypothesis
   b. Alternative hypothesis
   c. Critical point(s)
   d. Significance level
   e. Nonrejection region
   f. Rejection region
   g. Tails of a test
   h. Two types of errors

9.5. Explain which of the following is a two-tailed test, a left-tailed test, or a right-tailed test.
   a. $H_0: \mu = 12, H_1: \mu < 12$
   b. $H_0: \mu \leq 85, H_1: \mu > 85$
   c. $H_0: \mu = 33, H_1: \mu \neq 33$

Show the rejection and nonrejection regions for each of these cases by drawing a sampling distribution curve for the sample mean, assuming that it is normally distributed.

9.9. Write the null and alternative hypotheses for each of the following examples. Determine if each is a case of a two-tailed, a left-tailed, or a right-tailed test.
   a. To test if the mean number of hours spent working per week by college students who hold jobs is different from 20 hours
   b. To test whether or not a bank's ATM is out of service for an average of more than 10 hours per month
   c. To test if the mean length of experience of airport security guards is different from three years
   d. To test if the mean credit card debt of college seniors is less than $1000
   e. To test if the mean time a customer has to wait on the phone to speak to a representative of a mail-order company about unsatisfactory service is more than 12 minutes

9.28. Consider the null hypothesis $H_0: \mu = 50$. Suppose a random sample of 24 observations is taken from a normally distributed population with $\sigma = 7$. Using $\alpha = .05$, show the rejection and nonrejection regions on the sampling distribution curve of the sample mean and find the critical value(s) of $z$ when the alternative hypothesis is
a. \( H_1: \mu < 50 \)

b. \( H_1: \mu \neq 50 \)

c. \( H_1: \mu > 50 \)

9.30. Consider \( H_0: \mu = 100 \) versus \( H_1: \mu \neq 100 \).

a. A random sample of 64 observations produced a sample mean of 98. Using \( \alpha = .01 \), would you reject the null hypothesis? The population standard deviation is known to be 12.

b. Another random sample of 64 observations taken from the same population produced a sample mean of 104. Using \( \alpha = .01 \), would you reject the null hypothesis? The population standard deviation is known to be 12.

Comment on the results of parts a and b.

9.42. A study conducted a few years ago claims that adult men spend an average of 11 hours a week watching sports on television. A recent sample of 100 adult men showed that the mean time they spend per week watching sports on television is 9 hours. The population standard deviation is given to be 2.2 hours.

a. Test at the 1% significance level whether currently all adult men spend less than 11 hours per week watching sports on television.

b. What will your decision be in part a if the probability of making a Type I error is zero? Explain.

9.48. For each of the following examples of tests of hypotheses about \( \mu \), show the rejection and nonrejection regions on the \( t \) distribution curve.

a. A two-tailed test with \( \alpha = .02 \) and \( n = 20 \)

b. A left-tailed test with \( \alpha = .01 \) and \( n = 16 \)

c. A right-tailed test with \( \alpha = .05 \) and \( n = 18 \)

9.60. The president of a university claims that the mean time spent partying by all students at this university is not more than 7 hours per week. A random sample of 40 students taken from this university showed that they spent an average of 9.50 hours partying the previous week with a standard deviation of 2.3 hours. Test at the 2.5% significance level whether the president's claim is true. Explain your conclusion in words.

9.85. Make the following hypothesis tests about \( p \).

a. \( H_0: p = .45, H_1: p \neq .45, n = 100, \hat{p} = .49 \), \( \alpha = .10 \)

b. \( H_0: p = .72, H_1: p < .72, n = 700, \hat{p} = .64 \), \( \alpha = .05 \)

c. \( H_0: p = .30, H_1: p > .30, n = 200, \hat{p} = .33 \), \( \alpha = .01 \)
11.13. The following table lists the frequency distribution for 60 rolls of a die.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>1-spot</th>
<th>2-spot</th>
<th>3-spot</th>
<th>4-spot</th>
<th>5-spot</th>
<th>6-spot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>7</td>
<td>12</td>
<td>8</td>
<td>15</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>

Test at the 5% significance level whether the null hypothesis that the given die is fair is true.

11.19. The following table lists the frequency distribution of cars sold at an auto dealership during the past 12 months.

<table>
<thead>
<tr>
<th>Month</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cars sold</td>
<td>23</td>
<td>17</td>
<td>15</td>
<td>10</td>
<td>14</td>
<td>12</td>
<td>13</td>
<td>15</td>
<td>23</td>
<td>26</td>
<td>27</td>
<td>29</td>
</tr>
</tbody>
</table>

Using the 10% significance level, will you reject the null hypothesis that the number of cars sold at this dealership is the same for each month?

11.37. In a Bernard Haldane Associates survey conducted in March 2004, white-collar workers who had changed jobs in the past 12 months were asked whether their new positions paid more, less, or the same as their previous jobs (USA TODAY, April 28, 2004). Assuming that the survey included randomly selected samples of 240 men and 240 women who had changed their jobs in the previous 12 months, the percentages given in the newspaper would yield the following table.

<table>
<thead>
<tr>
<th>Pay at New Job</th>
<th>More</th>
<th>Less</th>
<th>Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>140</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Women</td>
<td>93</td>
<td>104</td>
<td>43</td>
</tr>
</tbody>
</table>

Using the 1% significance level, test the null hypothesis that the changes in pay for workers who change jobs are similar for both men and women.

11.56. All shoplifting cases in the town of Seven Falls are randomly assigned to either Judge Stark or Judge Rivera. A citizens group wants to know whether either of the two judges is
more likely to sentence the offenders to jail time. A sample of 180 recent shoplifting
cases produced the following two-way table.

<table>
<thead>
<tr>
<th></th>
<th>Jail</th>
<th>Other Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge Stark</td>
<td>27</td>
<td>65</td>
</tr>
<tr>
<td>Judge Rivera</td>
<td>31</td>
<td>57</td>
</tr>
</tbody>
</table>

Test at the 5% significance level whether the type of sentence for shoplifting depends on
which judge tries the case.

11.77. Each of five boxes contains a large (but unknown) number of red and green marbles. You
have been asked to find if the proportions of red and green marbles are the same for each
of the five boxes. You sample 50 times, with replacement, from each of the five boxes
and observe 20, 14, 23, 30, and 18 red marbles, respectively. Can you conclude that the
five boxes have the same proportions of red and green marbles? Use a .05 level of
significance.
Suppose we run a regression of an LMU alumnus’s salary on his/her GPA for a group of 36 alumni, so the y variable is salary and the x variable is GPA. The graph below plots all the data for your 36 individuals.

Suppose that the slope of the regression line is $11,300 and that the intercept is $15,800.

1) Draw a reasonable regression line for this data.
2) The estimated equation for the regression line is $y = 15,800 + 11,300x + u$, where $u$ is the random error term. Suppose that you meet an alumnus named Henry, who tells you his GPA at LMU was 1.0. Based on the equation above, what is our estimate of Henry’s salary based on his GPA?
3) If the standard error of the estimate of the slope is $1500, can we conclude that GPA affects salary at the 1% significance level?