Chapter 1

Literary scholars sometimes use the distribution of word lengths in a work as a test of authenticity. Here are the words lengths for the first 25 words on a randomly-selected page from Toni Morrison's *Song of Solomon*.

\[2 \quad 3 \quad 4 \quad 10 \quad 2 \quad 11 \quad 2 \quad 8 \quad 4 \quad 3 \quad 7 \quad 2 \quad 7 \]
\[5 \quad 3 \quad 6 \quad 4 \quad 4 \quad 2 \quad 5 \quad 8 \quad 2 \quad 3 \quad 4 \quad 4 \]

1. Make a dotplot of the lengths of the words.

2. Make a histogram of the lengths of the words.

3. Describe the histogram. Remember SCOSCU.

4. Find the mean of the lengths of the words. Is the mean resistant to outliers?

5. Find the standard deviation of the lengths of the words. Is standard deviation resistant to outliers?
6. Find the median of the lengths of the words. Is the median resistant to outliers?

7. Find the range of the lengths of the words.

8. Find the 5-number summary of the lengths of the words.

9. What is the interquartile range of the lengths of the words?

10. Make a boxplot of the lengths of the words that is not modified.

11. Make a modified boxplot of lengths of the words.

12. What is Simpson’s Paradox?
13. Make a back-to-back split stemplot of the following data:

**Reading Scores**

<table>
<thead>
<tr>
<th>4th Graders</th>
<th>12</th>
<th>15</th>
<th>18</th>
<th>20</th>
<th>20</th>
<th>22</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>26</td>
<td>28</td>
<td>29</td>
<td>31</td>
<td>32</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>7th Graders</td>
<td>1</td>
<td>12</td>
<td>15</td>
<td>18</td>
<td>18</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>27</td>
<td>28</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>33</td>
<td>33</td>
<td>35</td>
<td>36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

14. Make a comparison between 4th grade and 7th grade reading scores based on your stemplot.

Below is some information about films for the “Best Movie” Academy Award (Oscar) in 2009.

<table>
<thead>
<tr>
<th>Name</th>
<th>Genre</th>
<th>Budget (millions of dollars)</th>
<th>Total # of Oscar Nominations</th>
<th>Running time (minutes)</th>
<th>MMPA Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>Adventure</td>
<td>237</td>
<td>9</td>
<td>162</td>
<td>PG-13</td>
</tr>
<tr>
<td>The Blind Side</td>
<td>Drama</td>
<td>29</td>
<td>2</td>
<td>128</td>
<td>PG-13</td>
</tr>
<tr>
<td>District 9</td>
<td>Action</td>
<td>30</td>
<td>4</td>
<td>112</td>
<td>R</td>
</tr>
<tr>
<td>An Education</td>
<td>Drama</td>
<td>7</td>
<td>3</td>
<td>95</td>
<td>PG-13</td>
</tr>
<tr>
<td>The Hurt Locker</td>
<td>Action</td>
<td>11</td>
<td>9</td>
<td>131</td>
<td>R</td>
</tr>
<tr>
<td>Inglourious Basterds</td>
<td>Drama</td>
<td>70</td>
<td>8</td>
<td>153</td>
<td>R</td>
</tr>
<tr>
<td>Precious</td>
<td>Drama</td>
<td>10</td>
<td>6</td>
<td>110</td>
<td>R</td>
</tr>
<tr>
<td>A Serious Man</td>
<td>Comedy</td>
<td>7</td>
<td>2</td>
<td>106</td>
<td>R</td>
</tr>
<tr>
<td>Up</td>
<td>Animated</td>
<td>175</td>
<td>5</td>
<td>96</td>
<td>PG</td>
</tr>
<tr>
<td>Up in The Air</td>
<td>Comedy</td>
<td>30</td>
<td>6</td>
<td>109</td>
<td>R</td>
</tr>
</tbody>
</table>

15. List the categorical variable(s) and the quantitative variable(s).

16. Make a bar chart that displays the number of Oscar nominations received by the ten films listed.
<table>
<thead>
<tr>
<th>Education</th>
<th>Never smoked</th>
<th>Smoked but quit</th>
<th>Smokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not complete high school</td>
<td>82</td>
<td>19</td>
<td>113</td>
</tr>
<tr>
<td>Completed high school only</td>
<td>97</td>
<td>25</td>
<td>103</td>
</tr>
<tr>
<td>1 to 3 years of college</td>
<td>92</td>
<td>49</td>
<td>59</td>
</tr>
<tr>
<td>4 or more years of college</td>
<td>86</td>
<td>63</td>
<td>37</td>
</tr>
</tbody>
</table>

17. Fill in the marginal distributions for this table.

18. What percent of these people still smoke?

19. What percent of never-smokers completed high school only?

20. What percent of those with 4 or more years of college have quit smoking?

21. What percent of smokers did not finish high school?

22. Create three segmented bar graphs to show the distribution of education level among the different smoking statuses.

23. What conclusion can be drawn about smoking and education from this table?
Chapter 2

24. What is the area under a density curve?

25. The (mean or median) of a density curve is the equal-areas point, the point that divides the area under the curve in half.

26. The (mean or median) of a density curve is the balance point, at which the curve would balance if made of solid material.

27. If a density curve is skewed to the right, the (mean or median) will be further to the right than the (mean or median).

28. If a density curve is skewed to the left, the (mean or median) will be further to the right than the (mean or median).

29. What is the difference between $\bar{x}$ and $\mu$?

30. What is the difference between $s_x$ and $\sigma$?

31. Sketch the graph of $N(266, 16)$, the distribution of pregnancy length from conception to birth for humans. Be sure to include the empirical rule.

32. Would you say pregnancy length is a continuous or discrete variable? Justify.

33. Find the length of the longest 16% of all pregnancies.

34. Find the length of the middle 99.7% of all pregnancies.

35. Find the length of the shortest 2.5% of all pregnancies.

36. What percentile rank is a pregnancy of 250 days?
37. What percentile rank is a pregnancy of 298 days?

38. What z-score does a pregnancy of 257 days have?

39. What percent of humans have a pregnancy lasting less than 257 days?

40. What percent of humans have a pregnancy lasting longer than 280 days?

41. What percent of humans have a pregnancy lasting between 260 and 270 days?

42. How long would a pregnancy have to last to be in the longest 10% of all pregnancies?

43. How short would a pregnancy be to be in the shortest 25% of all pregnancies?

44. How long would a pregnancy be to be in the middle 20% of all pregnancies?

45. Are the lengths of the words presented at the beginning of the review normally distributed? Create a normal probability plot to justify your answer.
Chapter 3 and Section 12.2

46. What two things does correlation tell us about a scatterplot?

47. What is the highest value the correlation coefficient can take?

48. What is the lowest value the correlation coefficient can take?

49. What value represents the worst correlation?

50. Correlation only applies to what type(s) of relationship(s)?

51. Is the correlation coefficient resistant to outliers?

52. Does a high correlation indicate a strong cause-effect relationship? Explain your answer.

53. Why should we avoid using averaged data for regression and correlation?

54. What is extrapolation, and why shouldn’t we trust predictions using extrapolation?
Below is some data on the relationship between the price of a certain manufacturer’s flat-panel LCD televisions and the area of the screen. We would like to use these data to predict the price of televisions based on size.

<table>
<thead>
<tr>
<th>Screen Area (sq. inches)</th>
<th>Price (dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>154</td>
<td>250</td>
</tr>
<tr>
<td>207</td>
<td>265</td>
</tr>
<tr>
<td>289</td>
<td>330</td>
</tr>
<tr>
<td>437</td>
<td>375</td>
</tr>
<tr>
<td>584</td>
<td>575</td>
</tr>
<tr>
<td>683</td>
<td>650</td>
</tr>
</tbody>
</table>

55. Create a scatterplot to display this data.

56. What is the response variable?

57. What is the explanatory variable?

58. Describe the scatterplot. Be sure to address the strength, form, and direction.

59. Does it appear that there are outliers in this data set? Explain your answer.

60. If there are outliers, are they influential?

61. Calculate the correlation coefficient and interpret its meaning in context.
62. If screen area is measured in square centimeters, what happens to the correlation?

63. Find the least-squares regression line (LSRL) for the data.

64. What is the slope of this line? Interpret its meaning in context.

65. What is the y-intercept of this line? Interpret its meaning in context.

66. Predict the price of a television with a screen area of 361 square inches.

67. Predict the price of a television with a screen area of 3600 square inches.

68. Why is the prediction in problem 66 acceptable but the prediction in problem 67 not?

69. Find the residual value for a television with a screen area of 584 square inches.

70. Find the coefficient of determination for this data. Interpret this value in context.

71. Is a linear model appropriate for this data? Create a residual plot to support your answer.
Some college students collected data on the intensity of light at various depths in a lake.

<table>
<thead>
<tr>
<th>Depth (meters)</th>
<th>Light Intensity (lumens)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>168</td>
</tr>
<tr>
<td>6</td>
<td>120.42</td>
</tr>
<tr>
<td>7</td>
<td>86.31</td>
</tr>
<tr>
<td>8</td>
<td>61.87</td>
</tr>
<tr>
<td>9</td>
<td>44.34</td>
</tr>
<tr>
<td>10</td>
<td>31.78</td>
</tr>
<tr>
<td>11</td>
<td>22.78</td>
</tr>
</tbody>
</table>

72. Find the least squares regression line for the data above.

73. Find the value of the correlation coefficient and the coefficient of determination.

74. Is a linear model appropriate for this data set? Use a residual plot to justify your answer.

75. Take the log of light intensity and record those values in the chart.

76. Find the LSRL for the transformed data.

77. Change the LSRL back into the original variables.

78. Predict the light intensity at a depth of 12 meters.
Chapter 4

79. What is the difference between an observational study and an experiment?

80. What is a lurking variable?

81. What is a voluntary response sample?

82. How are a population and a sample related but different?

83. Why is convenience sampling biased?

84. SRS stands for what kind of sample? Name and describe.

85. Use the digits below to choose a SRS of 4 towns from this list. Explain your steps.

<table>
<thead>
<tr>
<th>Allendale</th>
<th>Bangor</th>
<th>Chelsea</th>
<th>Detour</th>
<th>Edmonton</th>
<th>Fennville</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gratiot</td>
<td>Hillsdale</td>
<td>Ionia</td>
<td>Joliet</td>
<td>Kentwood</td>
<td>Ludington</td>
</tr>
<tr>
<td>29732</td>
<td>00116</td>
<td>96309</td>
<td>13108</td>
<td>35600</td>
<td>86132</td>
</tr>
<tr>
<td>99735</td>
<td>89621</td>
<td>01012</td>
<td>81546</td>
<td>93450</td>
<td>21966</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18353</td>
<td>47128</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73301</td>
<td>00254</td>
</tr>
</tbody>
</table>
86. What is common response? Give an example.

87. What is confounding? Give an example.

88. What is a stratified random sample?

89. What is a cluster sample?

90. What is undercoverage?

91. What is nonresponse?

92. What is response bias?

93. Why is the wording of questions important? Give an example.

94. How are experimental units and subjects similar but different?

95. Explanatory variables in experiments are often called _____.

96. If I test a drug at 100 mg, 200 mg, and 300 mg, I am testing one variable at three _____.

97. What is the placebo effect?

98. What is the purpose of a control group?
99. What are the two types of matched pairs used in experiments?

100. What are the three principles of experimental design?

101. What does double-blind mean, and why would we want an experiment to be double-blind?

102. What is block design?

103. I want to test the effects of aerobic exercise on resting heart rate. I want to test two different levels of exercise, 30 minutes 3 times per week and 30 minutes 5 times per week. I have a group of 20 people to test, 10 men and 10 women. I will take heart rates before and after the experiment. Draw a chart for this experimental design. Be sure to explain how randomization is used.

104. What are the five steps of a simulation?

105. Design and perform a simulation of how many children a couple must have to get two sons. (A simulation involves many trials. For this simulation, perform 10 trials.)
Chapter 5

106. Any probability is a number between (and including) _____ and _____.

107. All possible outcomes together must have a sum of ______.

108. If S is the sample space, P(S) = ______.

109. What is the law of large numbers?

110. What is the complement of an event? Draw a Venn diagram to illustrate your answer.

111. What are disjoint events? Give two examples and draw a Venn diagram.

112. What is the union of two events?

113. What is an intersection of two events?

114. What does the symbol $\cup$ mean?

115. What does the symbol $\cap$ mean?

116. What does it mean if two events are independent?

117. What are the three ways to test for independence?
118. You are going to flip a coin three times. What is the sample space for each flip?

119. You are going to flip a coin three times and note how many heads and tails you get. What is the sample space?

120. You are going to flip a coin three times and note what you get on each flip. What is the sample space?

121. How many different four-digit numbers can you make?

122. How many different four-digit numbers can you make without repeating digits?

Use the following chart for questions 123 – 125:

<table>
<thead>
<tr>
<th>M&amp;M Color</th>
<th>Brown</th>
<th>Red</th>
<th>Yellow</th>
<th>Green</th>
<th>Orange</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>?</td>
</tr>
</tbody>
</table>

123. What is the probability that an M & M is blue?

124. What is the probability that an M & M is red or green?

125. What is the probability that an M & M is yellow and orange?

126. Bre can beat Erica in tennis 9% of the time. Erica can swim faster than Bre 8% of the time. What is the probability that Bre would beat Erica in a tennis match and in a swimming race?

127. What assumption are you making in problem 138? Do you think this assumption is valid?
128. Using two dice, what is the probability that you would roll a sum of seven or eleven?

129. Using two dice, what is the probability that you would roll doubles?

130. Using two dice, what is the probability that you would roll a sum of 7 or 11 on the first roll and doubles on the second roll?

131. What assumption are you making in problem 142? Do you think this assumption is valid?

132. Using two dice, what is the probability that you would roll a sum of 7 or 11 that is also doubles?

133. Perform an independence test on the smoking/education chart from problem 90 to show that smoking status and education are not independent.

A study was conducted at a local day care to see which television show(s) the students watched on Saturday. It was found that 45% of students watched Bubble Guppies at 8 AM, 25% of students watched Mickey's Clubhouse at 9 AM, 55% of students watched Sophia the First at 10 AM, 15% watched Bubble Guppies and Mickey's Clubhouse, 25% watched Bubble Guppies and Sophia the First, 10% watched Mickey's Clubhouse and Sophia the First, and 5% watched all three cartoons.

134. Create a Venn diagram to represent this situation.

135. What percent of students watched none of these shows?

136. What percent of students watched exactly two of these shows?

137. What percent of students watched Bubble Guppies and nothing else?

138. What percent of students did not watch Sophia the First?
For problems 139 – 142 consider the process of a drawing a card from a standard deck and replacing it. Let A be drawing a heart, B be drawing a king, and C be drawing a spade.

139. Are the events A and B disjoint? Explain.

140. Are the events A and B independent? Explain.

141. Are the events A and C disjoint? Explain.

142. Are the events A and C independent? Explain.

Chapter 6 Part 1

143. Give an example of a discrete random variable.

144. Give an example of a continuous random variable.

145. Normal distributions are (continuous or discrete).

146. What is a uniform distribution? Draw a picture.

147. Expected value is another name for __________.
The chart below shows the probability distribution of AP scores for Mr. Barry’s AP Statistics class.

<table>
<thead>
<tr>
<th>AP Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.0455</td>
<td>0.1212</td>
<td>0.1515</td>
<td>0.4697</td>
<td>0.2121</td>
</tr>
</tbody>
</table>

148. What is the probability a randomly selected AP student scored more than a 3?

149. What is \( P(X > 3) \)?

150. Find the expected value of the AP scores.

151. Find the variance of the AP scores.

152. Find the standard deviation of the AP scores.

153. If I sell an average of 5 books per day and 7 CDs per day, what is the average number of items I sell per day?

154. If I charge $2 per book and $1.50 per CD in problem 153, what is my average amount of income per day?

155. Before you can use the rules for variances you must make sure the variables are __________.
A probability distribution connects the points (0, 0.1), (1, 0.1), (2, 0.6), (3, 0.1), and (5, 0.1). Use this information to answer the following questions.

156. Draw the probability distribution function on the axis provided.

157. Find \( P(x > 2) \).

158. Find \( P(x \geq 2) \).

159. Find \( P(x = 2) \).

160. Find \( P(1 < x < 3) \).

161. Find \( P(x \geq 5) \).
For problems 156 – 164, use the following situation: For Test 1, the class average was 80 with a standard deviation of 10. For Test 2, the class average was 70 with a standard deviation of 12. A student’s grade is based on the total points earned on the tests.

162. What is the average for the two tests added together?

163. What is the standard deviation for the two tests added together?

164. What is the difference in the test averages?

165. What is the standard deviation for the difference in the test averages?

166. If each test score on Test 2 in halved and 50 points are added, what is the new average?

167. What is the new standard deviation for Test 2 in problem 179?

168. If 7 points are added to every Test 1 score, what is the new standard deviation?

169. If every Test 1 score is doubled and 80 points are subtracted, what is the new mean?

170. If every Test 1 score is doubled and 80 points are subtracted, what is the new standard deviation?