Statistics Review:

In the paper "Reproduction in Laboratory colonies of Bank Vole," the authors presented the results of a study of litter size. (A vole is a small rodent with a stout body, blunt nose, and short ears.) As each new litter was born, the number of babies was recorded, and the accompanying results were obtained.

1	4	4	5	5	6	6	7	7	8
2	4	5	5	5	6	6	7	7	8
2	4	5	5	6	6	6	7	7	8
3	4	5	5	6	6	6	7	8	8
3	4	5	5	6	6	7	7	8	9
3	4	5	5	6	6	7	7	8	9
3	4	5	5	6	6	7	7	8	9
3	4	5	5	6	6	7	7	8	10
3	4	5	5	6	6	7	7	8	10
4	4	5	5	6	6	7	7	8	11

The authors also kept track of the color of the first born in each litter. (B = brown, G = gray, W = white, and T = tan)

B	B	T	W	T	G	G	G	B	B
W	B	W	B	T	T	G	B	T	B
B	T	B	B	B	G	W	B	B	G
G	G	G	B	B	T	B	W	T	T
B	T	B	B	T	W	W	B	G	B
B	B	B	G	T	B	B	T	T	G
G	B	B	B	B	G	W	G	T	G
B	B	B	B	G	G	T	T	W	G
G	W	T	G	T	B	B	G	B	B
B	G	T	W	B	G	T	W	G	W

1. Which variable is categorical, and which variable is quantitative?

2. Make a bar chart of the colors.

3. Make a histogram of the litter sizes.

4. Make a dotplot of the litter sizes.

5. Are there any outliers in the histogram or dotplot?

6. Describe the shape of the histogram (symmetric or skewed).

7. Find the mean of the litter sizes. Is the mean resistant to outliers?

8. Find the median of the litter sizes. Is the median resistant to outliers?

9. Find the range of the litter sizes.

10. Find the 5-number summary of the litter sizes.

11. What is the interquartile range?

12. Make a boxplot of the litter sizes.

13. Find the standard deviation of the litter sizes. Is standard deviation resistant to outliers?

14. If a distribution is skewed to the right, the (mean or median) will be further to the right than the (mean or median).

15. What is the difference between x-bar and μ?

16. What is the difference between s and σ?

17. How do you find the inflection points on a normal curve?

18. Sketch the graph of N(266, 16), the distribution of pregnancy length from conception to birth for humans.

19. What is the 68-95-99.7 rule?

20. Using the empirical rule (the 68-95-99.7 rule), find the length of the longest 16% of all pregnancies. Sketch and shade a normal curve for this situation.

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

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

23. What percentile rank is a pregnancy of 218 days?

24. What percentile rank is a pregnancy of 298 days?

25. What is the percentile of a pregnancy of 266 days?

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

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

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

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

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

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

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

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

34. Does the vole information from the beginning of this review seem to be normal? Justify by checking actual percentages within 1, 2, and 3 standard deviations of the mean.

35. Make a back-to-back split stemplot of the following data:

Reading Scores

4th Graders	12	15	18	20	20	22	25	26	28	29
	31	32	35	35	35	36	37	39	40	42
7th Graders	1	12	15	18	18	20	23	23	24	25
	27	28	30	30	31	33	33	33	35	36

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

37. What is the mode of each set of scores?

38. Is the score of "1" for one of the 7th graders an outlier? Test using the 1.5 IQR rule.

39. What is the difference between a modified boxplot and a regular boxplot? Why is a modified boxplot usually considered better?

40. Graph the following hot dog data:

Calories	Sodium (milligrams)
108	149
130	350
132	345
135	360
138	360
140	375
144	380
145	390
150	400
163	415
167	400
172	420
176	450
180	500
184	505
195	500
200	515

41. What is the response variable?

42. What is the explanatory variable?

43. What is the direction of this scatterplot? (positive, negative…)

44. What is the form of this scatterplot? (linear, exponential…)

45. What is the strength of this scatterplot? (strong, weak…)

46. Are there outliers? (Outliers in a scatterplot have large residuals.)

47. If there are outliers, are they influential?

48. Calculate the correlation.

49. Calculate the correlation without the point (108, 149).

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

51. If I change the units on sodium to grams instead of milligrams, what happens to the correlation?

52. What is the highest correlation possible?

53. What is the lowest correlation possible?

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

55. Is correlation resistant to outliers?

56. Does a high correlation indicate a strong cause-effect relationship?

57. Sketch a scatterplot with a correlation of about 0.8.

58. Sketch a scatterplot with a correlation of about –0.5.

59. Find the least-squares regression line (LSRL) for the calories-sodium data.

60. What is the slope of this line, and what does it tell you in this context?

61. Predict the amount of sodium in a hot dog with 155 calories.

62. Predict the amount of sodium in a hot dog with 345 calories.

63. Why is the prediction in problem 64 acceptable but the prediction in problem 65 not?

64. Find the error in prediction (residual) for a hot dog with 180 calories.

65. The point (x-bar, y-bar) is always on the LSRL. Find this point, and verify that it is on your scatterplot.

66. Find the standard deviation of the calories.

67. Find the standard deviation of the sodium.

68. Find the coefficient of determination for this data.

69. What does r^2 tell you about this data?

70. How can you use a residual plot to tell if a line is a good model for data?

71. If you know a scatterplot has a curved shape, how can you decide whether to use a power model or an exponential model to fit data?

72. Graph the following data:

Time (days)	Mice
0	6
30	19
60	60
90	195
120	597

73. Perform the appropriate logarithmic transformation (power or exponential) on the above data to get an equation.

74. Make a residual plot to support your choice for problem 73.

75. Graph the following data:

Diameter (inches)	Cost (dollars)
6	3.50
9	8.00
12	14.50
15	22.50
20	39.50

76. Perform the appropriate logarithmic transformation (power or exponential) on the above data to get an equation.

77. Make a residual plot to support your choice for problem 76.

78. What is the correlation for the equation you found in problem 76?

79. What is extrapolation, and why shouldn't we trust predictions using extrapolation?

80. What is interpolation?

81. What is a lurking variable?

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

Use this table for questions 83-88:

	Smoking Status		
Education	Never smoked	Smoked, but quit	Smokes
Did not complete high school	82	19	113
Completed high school	97	25	103
1 to 3 years of college	92	49	59
4 or more years of college	86	63	37

83. Fill in the marginal distributions for this table.

84. What percent of these people smoke?

85. What percent of never-smokers completed high school?
86. What percent of those with 4 or more years of college have quit smoking?

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

88. What conclusion can be drawn about smoking and education from this table?

89. What is Simpson's Paradox?

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

91. What is a voluntary response sample?

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

93. Why is convenience sampling biased?

94. SRS stands for what kind of sample? Name and define.

95. Discuss how to choose an SRS of 4 towns from this list:

Allendale	Bangor	Chelsea	Detour	Edmonton	Fennville
Gratiot	Hillsdale	Ionia	Joliet	Kentwood	Ludington

96. What is a stratified random sample?

97. What is a cluster sample?

98. What is undercoverage?

99. What is nonresponse?

100. What is response bias?

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

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

103. Explanatory variables in experiments are often called _____.

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

105. What is the placebo effect?

106. What is the purpose of a control group?

107. What are the two types of matched pairs used in experiments?

108. What are the three principles of experimental design?

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

110. What is block design?

111. 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.

112. 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.)

113. What is independence?

114. You are going to flip a coin three times. What is the sample space for each flip?

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

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

117. Make a tree diagram for the three flips.

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

119. If S is the sample space, P(S) = _____.

120. What are complements? Give an example and draw a Venn diagram.

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

Use the following chart for questions 122-124:

M&M Color	Brown	Red	Yellow	Green	Orange	Blue
Probability	0.3	0.2	0.2	0.1	0.1	?

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

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

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

125. 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?

126. What assumption are you making in problem 125? Do you think this assumption is valid?

127. Using two dice, what is the probability that you would roll a sum of seven or eleven?

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

129. 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?

130. What assumption are you making in problem 129? Do you think this assumption is valid?

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

132. What is the union of two events?

133. What is an intersection of two events?

134. How can we test independence?

135. Perform an independence test on the smoking/education chart from problem 83 to show that smoking status and education are not independent.

136. Make a Venn diagram for the following situation:

45% of kids like Barney
25% of kids like Blue
55% of kids like Pooh
15% of kids like Blue and Pooh
25% of kids like Barney and Pooh
5% of kids Barney, Blue, and Pooh
5% of kids like Blue but not Barney or Pooh

137. A dartboard has a circle with a 20-inch diameter drawn inside a 2-foot square. What is the probability that a dart lands inside the circle given that it at least lands inside the square? (Assume a random trial here.)

For problems 138-141 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.

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

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

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

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

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

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

144. Give an example of a discrete random variable.

145. Give an example of a continuous random variable.

146. Make a probability histogram of the following grades on a four-point scale:

Grade	0	1	2	3	4
Probability	0.05	0.28	0.19	0.32	0.16

147. Using problem 159, what is $P(X > 2)$?

148. Using problem 159, what is $P(X \geq 2)$?

149. How do your answers to problems 147 and 148 demonstrate a difference between continuous and discrete random variables?

150. Normal distributions are <u>(continuous or discrete)</u>.

151. Expected value is another name for _____.

152. Find the expected value of the grades in problem 146.

153. Find the variance of the grades in problem 146.

154. Find the standard deviation of the grades in problem 146.

155. 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?

156. If I charge \$2 per book and \$1.50 per CD in problem 155, what is my average amount of income per day?

157. Before you can use the rules for variances you must make sure the variables are _____.

For problems 158-166, 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.

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

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

160. What is the difference in the test averages?

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

162. If I cut the test scores on Test 2 in half and add 50, what is the new average?

163. What is the new standard deviation for Test 2 in problem 199?

164. If I add 7 points to every Test 1, what is the new standard deviation?

165. If I multiply every Test 1 by 2 and subtract 80, what is the new mean?

166. If I multiply every Test 1 by 2 and subtract 80, what is the new standard deviation?

Use the following situation for questions 169-183: The probability that a child born to a certain set of parents will have blood type AB is 25%.

167. The parents have four children. X is the number of those children with blood type AB. Is this binomial or geometric?

168. Using the situation in problem 167, find $P(X = 2)$.

169. Using the situation in problem 167, find $P(X < 3)$.

170. Using the situation in problem 167, find $P(X \geq 1)$.

171. Using the situation in problem 167, find $P(1 \leq X \leq 3)$.

172. Using the situation in problem 167, find $P(2 < X < 4)$.

173. What is the mean of the situation in problem 167?

174. What is the standard deviation of the situation in problem 167?

175. A set of parents continue having children until they have a child with type AB blood. X is the number of children they have to give birth to in order to have one child with type AB blood. Is this binomial or geometric?

176. Using the situation in problem 175, find $P(X = 1)$.

177. Using the situation in problem 175, find $P(X \leq 2)$.

178. Using the situation in problem 175, find $P(X > 5)$.

179. Using the situation in problem 175, find $P(2 \leq X < 4)$.

180. Using the situation in problem 175, find $P(2 < X \leq 5)$.

181. What is the mean of the situation in problem 175?