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 6 6 6 7 7 8
2 4 5 5 6 6 6 7 7 8
3 4 5 5 6 6 6 7 7 8 8
3 4 5 5 6 6 7 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 B
W B W B T T G B T B
B T B B B G W B B G G
G G G B B T B W T T
B T B B T W W B G B B
B B B G T B B T T G G
G B B B B G W G T G
B B B B B G G T T W G
G W T G T B B G B B B
B G T W B G T W G W

1. Which variable is categorical, and which variable is quantitative?
   Color is categorical; litter size 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.  
   (not shown)

5. Describe the shape of the histogram (symmetric or skewed).  
   approximately symmetric

6. Find the mean of the litter sizes.  Is the mean resistant to outliers?  \( \bar{x} = 5.87 \)
   No, the mean is NOT resistant to outliers.

7. Find the median of the litter sizes.  Is the median resistant to outliers?  \( M = 6 \)
   Yes, the median is resistant to outliers.

8. Find the range of the litter sizes.  
   10

9. Find the 5-number summary of the litter sizes.
   \[ \text{min} = 1 \quad Q_1 = 5 \quad \text{median} = 6 \quad Q_3 = 7 \quad \text{max} = 11 \]

10. What is the interquartile range?  
    2

11. Are there any outliers?  Yes, \( Q_1 - 1.5(\text{IQR}) = 5 - 1.5(2) = 2 \), so 1 is an outlier.  Similarly, \( Q_3 + 1.5(\text{IQR}) = 7 + 1.5(2) = 10 \), so 11 is an outlier

12. Make a boxplot of the litter sizes.

13. Find the standard deviation of the litter sizes.  Is standard deviation resistant to outliers?  
    \( s_x = 1.8127 \)  
    No, standard deviation is not resistant to outliers.

14. What is the area under a density curve?  
    1

15. The \text{[mean or median]} of a density curve is the equal-areas point, the point that divides the area under the curve in half.  \text{median}

16. The \text{[mean or median]} of a density curve is the balance point, at which the curve would balance if made of solid material.  \text{mean}

17. If a density curve is skewed to the right, the \text{[mean or median]} will be further to the right than the \text{[mean or median]}.  \text{Mean, median}
18. What is the difference between x-bar and \( \mu \)? X-bar is the sample mean and \( \mu \) is the population mean.

19. What is the difference between \( s \) and \( \sigma \)? \( s \) is the standard deviation of the sample and \( \sigma \) is the standard deviation of the population.

20. How do you find the inflection points on a normal curve? They are located one standard deviation on either side of the mean, \( \mu \pm \sigma \).

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

22. What is the 68-95-99.7 rule? About 68% of the observations will fall within one standard deviation of the mean. About 95% of the observations will fall within two standard deviations of the mean. About 99.7% of the observations will fall within three standard deviations of the mean.

23. 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.

The longest 16% of all pregnancies are \( \geq 282 \) days.

24. Find the length of the middle 99.7% of all pregnancies. \( \text{between 218 and 314 days} \)

25. Find the length of the shortest 2.5% of all pregnancies. \( \leq 234 \) days

26. What percentile rank is a pregnancy of 218 days? 0.15th percentile

27. What percentile rank is a pregnancy of 298 days? 97.5th percentile

28. What is the percentile of a pregnancy of 266 days? 50th percentile

29. What z-score does a pregnancy of 257 days have? -0.5625

30. What percent of humans have a pregnancy lasting less than 257 days? 28.69%

31. What percent of humans have a pregnancy lasting longer than 280 days? 19.08%
32. What percent of humans have a pregnancy lasting between 260 and 270 days? 24.49%
   \( \text{Normalcdf} (260, 270, 266, 16) = 0.2449 \)

33. Would you say pregnancy length is a continuous or discrete variable? Justify.
   continuous, since the possible number of days can be any value in a given interval

34. How long would a pregnancy have to last to be in the longest 10% of all pregnancies?
   \( \text{invNorm} (0.9, 266, 16) = 286.5 \text{ days} \)

35. How short would a pregnancy be to be in the shortest 25% of all pregnancies?
   \( \text{invNorm} (0.25, 266, 16) = 255.2 \text{ days} \)

36. How long would a pregnancy be to be in the middle 20% of all pregnancies?
   Between \( \text{invNorm} (0.6, 266, 16) \) and \( \text{invNorm} (0.4, 266, 16) \)
   Between 261.9 and 270.1 days

37. Does the vole information from the beginning of this review seem to be normal? Justify by
   looking at a normal probability plot.
   Under STAT PLOT, choose the last graph, Data List: L1 Data Axis: X,
   Choose ZoomStat. The normal probability looks approximately linear so we conclude that
   the distribution is approximately normal.

38. Make a back-to-back split stemplot of the following data:
   \begin{tabular}{l|l}
   & \text{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 35 36 \\
   \end{tabular}

   \begin{tabular}{l|l}
   4th graders & \text{7th graders} \\
   0 & 0 \\
   2 & 1 \\
   8 & 5 1 5 8 8 \\
   2 0 0 & 2 0 3 3 4 \\
   9 & 8 6 5 2 5 7 8 \\
   2 1 3 0 0 1 3 3 3 \\
   9 7 6 5 5 5 & 3 5 6 \\
   2 0 & 4 \\
   \end{tabular}

   Key \( 1 \mid 2 = 12 \)

39. Make a comparison between 4th grade and 7th grade reading scores based on your
   stemplot. The distribution between 4th grade and 7th grade reading scores are similar. They
   are both slightly skewed left. The range is lower for the 4th grade scores and both the
   minimum and maximum values are higher for the 4th grade scores. The 4th grade scores
   peak in the upper 30's, while the 7th grade scores peak in the lower 30's.

40. What is the mode of each set of scores?
   The mode for 4th grade scores is 35; the mode for 7th grade scores is 33.

41. Is the score of “1” for one of the 7th graders an outlier? Test using the 1.5 IQR rule.
   \( 1.5(\text{IQR}) = 1.5(13) = 19.5 \)
   \( 19 - 19.5 = -0.5 \)
   No, 1 is not an outlier.
42. What is the difference between a modified boxplot and a regular boxplot? Why is a modified boxplot usually considered better? A modified boxplot is usually better because it shows all outliers.

43. Graph the following hot dog data:

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

44. What is the response variable? sodium
45. What is the explanatory variable? Calories
46. What is the direction of this scatterplot? (positive, negative…) positive
47. What is the form of this scatterplot? (linear, exponential…) linear
48. What is the strength of this scatterplot? (strong, weak…) strong
49. Are there outliers? (Outliers in a scatterplot have large residuals.) Yes, (108, 149) is an outlier.
50. If there are outliers, are they influential? Yes What do you do to determine if an outlier is influential? To determine if an outlier is influential, calculate the LSRL with this point and without this point. If there is a significant difference in the LSRL, then the outlier is influential. Find the equation of the LSRL including the point (108, 149). \( \hat{y} = -85.4072 + 3.1087x \) and find the equation of the LSRL without this point. \( \hat{y} = 24.4389 + 2.4594x \)
51. Calculate the correlation with the point (108, 149) \( r = 0.9195 \)
52. Calculate the correlation without the point (108, 149). \( r = 0.9587 \)
53. What two things does correlation tell us about a scatterplot? the strength and direction of a linear relationship
54. If I change the units on sodium to grams instead of milligrams, what happens to the correlation? It remains the same since $r$ is a standardized value.

55. What is the largest correlation possible, in a positive or a negative direction? 1 or -1.

56. What is the smallest correlation possible? 0.

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

58. Is correlation resistant to outliers? No, it is not resistant to outliers.

59. Does a high correlation indicate a strong cause-effect relationship? No, correlation does not necessarily imply causation.

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

61. Sketch a scatterplot with a correlation of about -0.5.

62. Find the least-squares regression line (LSRL) for the calories-sodium data. $\hat{y} = -85.4072 + 3.1087x$.

63. What is the slope of this line, and what does it tell you in this context? As the number of calories increases by 1, the sodium increases, on average, by 3.1087 milligrams.

64. Predict the amount of sodium in a hot dog with 155 calories. 396.44 milligrams.

65. Predict the amount of sodium in a hot dog with 345 calories. 987.09 milligrams.

66. Why is the prediction in problem 64 acceptable but the prediction in problem 65 not? What is the name for this error in prediction? 155 calories is within the domain of the data; 345 calories is not. Extrapolation.

67. Find the error in prediction (residual) for a hot dog with 180 calories. Residual = $y - \hat{y} = 500 - 474.16 = 25.84$.
68. The point \((x-bar, y-bar)\) is always on the LSRL. Find this point, and verify that it is on your scatterplot. What is the name of this point? The point is called the grand mean.  
\[
(x-bar, y-bar) = (156.4118, 400.8235)  
-85.4072 + 3.1087(156.4118) = 400.8301
\]

69. Find the standard deviation of the calories. \(25.6395\)

70. Find the standard deviation of the sodium. \(86.6799\)

71. Find the coefficient of determination for this data. \(r^2 = 0.8455\)

72. What does \(r^2\) tell you about this data? Approximately 84.55% of the variation in sodium can be explained by the linear relationship between calories and sodium.

73. How can you use a residual plot to tell if a line is a good model for data? The residuals should be randomly scattered and relatively close to zero. There should be about the same number of positive and negative residuals.

74. 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? Look at the scatterplot of \((x, \log y)\) and the scatterplot of \((\log x, \log y)\). If the first appears more linear then an exponential model is appropriate. If the second appears more linear then a power model is appropriate. Also look at the value of the correlation, \(r\), and the residual plot in each case to make your final determination.

75. Graph the following data:

<table>
<thead>
<tr>
<th>Time (days)</th>
<th>Mice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>30</td>
<td>19</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>90</td>
<td>195</td>
</tr>
<tr>
<td>120</td>
<td>597</td>
</tr>
</tbody>
</table>

76. Perform the appropriate logarithmic transformation (power or exponential) on the above data to get an equation. Let \(L3 = \log (L1)\) and \(L4 = \log (L2)\). Look at linear regression of \(L1\) vs \(L4\) (exponential model) and linear regression of \(L3\) vs \(L4\) (power model). In this case the exponential model appears more linear. \[
\log \hat{y} = 0.7688 + 0.0168x
\]

77. Make a residual plot to support your choice for problem 76. Under STAT PLOT choose scatterplot with \(Xlist = L1\) and \(Ylist = RESID\). Use ZOOM STAT to view the residual plot.

78. Graph the following data:

<table>
<thead>
<tr>
<th>Diameter (inches)</th>
<th>Cost (dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>3.50</td>
</tr>
<tr>
<td>9</td>
<td>8.00</td>
</tr>
<tr>
<td>12</td>
<td>14.50</td>
</tr>
<tr>
<td>15</td>
<td>22.50</td>
</tr>
<tr>
<td>20</td>
<td>39.50</td>
</tr>
</tbody>
</table>
79. Perform the appropriate logarithmic transformation (power or exponential) on the above data to get an equation. Let \( L_3 = \log(L_1) \) and \( L_4 = \log(L_2) \). Look at linear regression of \( L_1 \) vs \( L_4 \) (exponential model) and linear regression of \( L_3 \) vs \( L_4 \) (power model). In this case the power model appears more linear. 
\[
\log \hat{y} = -1.0127 + 2.0167 \log x
\]

80. Make a residual plot to support your choice for problem 79. Under STAT PLOT choose scatterplot with Xlist = \( L_1 \) and Ylist = RESID. Use ZOOM STAT to view the residual plot.

81. What is the correlation for the equation you found in problem 79? \( r = 0.9999 \)

82. What is extrapolation, and why shouldn’t we trust predictions using extrapolation? Extrapolation is making a prediction outside the domain of the data. It is not reliable.

83. What is interpolation? Interpolation is making a prediction within the domain of the data.

84. What is a lurking variable? A lurking variable is a variable that may influence the value of the variables in a study, although it is not part of the study.

85. Why should we avoid using averaged data for regression and correlation? Averaged data has less variability, which results in a higher correlation. This higher correlation may not be an accurate representation of the true correlation.

86. What is causation? Give an example. Changes in the explanatory variable cause changes in the response variable. Example: the amount of time since a pie was removed from the oven and the temperature of the pie.

87. What is common response? Give an example. Changes in the explanatory variable do not cause changes in the response variable; a lurking (third) variable causes changes in both the explanatory and the response variable. They are both exhibiting a common response to the lurking variable. Example: ice cream sales at Virginia Beach and the number of drownings at Virginia Beach. They both increase in response to the weather conditions.

88. What is confounding? Give an example. Changes in the explanatory cause changes in the response variable, but a confounding variable also causes changes in the response variable. Example: smoking during pregnancy may cause low birth weight, but there are other confounding variables such as poor nutrition that may also cause low birth weight.

89. Why is a two-way table called a two-way table? There are two variables.

Use this table for questions 90-95:

<table>
<thead>
<tr>
<th>Education</th>
<th>Smoking Status</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never smoked</td>
<td>Smoked, but quit</td>
</tr>
<tr>
<td>Did not complete high school</td>
<td>82</td>
<td>19</td>
</tr>
<tr>
<td>Completed high school</td>
<td>97</td>
<td>25</td>
</tr>
<tr>
<td>1 to 3 years of college</td>
<td>92</td>
<td>49</td>
</tr>
<tr>
<td>4 or more years of college</td>
<td>86</td>
<td>63</td>
</tr>
<tr>
<td>TOTAL</td>
<td>357</td>
<td>156</td>
</tr>
</tbody>
</table>
90. Fill in the marginal distributions for this table.

91. What percent of these people smoke? $\frac{312}{825} = 37.82\%$

92. What percent of never-smokers only completed high school? $\frac{97}{357} = 27.17\%$

93. What percent of those with 4 or more years of college have quit smoking? $\frac{63}{186} = 33.87\%$

94. What percent of smokers did not finish high school? $\frac{113}{312} = 36.22\%$

95. What conclusion can be drawn about smoking and education from this table? 
   - The more education a person has completed, the less likely they are to smoke: 
     - 53% of those who did not complete high school smoke, 
     - 45% of those who completed high school smoke, 
     - 30% of those with 1 to 3 years of college smoke, and 
     - 20% of those with 4 or more years of college smoke.

96. What is Simpson’s Paradox? 
   When data from several groups are combined to form a single group, the association may be reversed.

97. What is the difference between an observational study and an experiment? 
   In an experiment, a treatment is imposed on the units or subjects.

98. What is a voluntary response sample? 
   The subjects select themselves to be in the sample.

99. How are a population and a sample related but different? 
   A sample is a part (a subset) of the population.

100. Why is convenience sampling biased? 
   A convenience sample does not accurately represent the population; some groups are inevitably underrepresented.

101. SRS stands for what kind of sample? Name and define. 
   In a simple random sample, each member of the population is equally likely to be selected, and each possible sample of size $n$ is equally likely to be selected.

102. Discuss how to choose a SRS of 4 towns from this list: 
   Allendale  Bangor  Chelsea  Detour  Edmonton  Fennville  Gratiot  Hillsdale  Ionia  Joliet  Kentwood  Ludington 
   Assign each town a number 01-12. Read off pairs of digits, discarding repeats and pairs that are not between 01 and 12, until you have chosen 4 towns. Starting at line 139 we would choose: Detour, Joliet, Ludington, Detour.

103. What is a stratified random sample? 
   The population is divided into groups of similar individuals called strata. They may be similar by age, socio-economic class, grade in school, religious affiliation. . . Then choose a SRS from each stratum and combine these to form the sample.

104. What is a cluster sample? 
   The population is divided into groups, and may be further divided into subgroups. All members of one or more subgroups are chosen.
105. What is undercoverage?
One or more groups with similar characteristics do not have a chance to be chosen for the sample.

106. What is nonresponse?
A subject chosen to be in the sample does not respond or refused to participate.

107. What is response bias?
A subject is not truthful or is in some way influenced to respond differently than they normally would.

108. Why is the wording of questions important? Give an example.
Wording containing bias may influence answers. Example: Prohibiting children from praying in school is a violation of the Bill of Rights. Do you think children should be allowed to pray in school? This question contains bias, since it only presents one side of the issue.

109. How are experimental units and subjects similar but different?
Subjects are human.

110. Explanatory variables in experiments are often called _____. Factors

111. If I test a drug at 100 mg, 200 mg, and 300 mg, I am testing one variable at three _______ Levels.

112. What is the placebo effect?
Some individuals respond to any form of treatment, regardless of whether it is a “real” or “fake” treatment.

113. What is the purpose of a control group?
To reduce or eliminate the effects of lurking variables.

114. What are the two types of matched pairs used in experiments?
- each unit/subject receives both treatments, in a randomly assigned order and results are compared, or
- one of each pair of units/subjects is randomly assigned to receive treatment A (or B) and the other unit/subject receives treatment B (or A) and the results are compared.

115. What are the three principles of experimental design?
I. Control the effects of lurking variables by comparing several treatments.
II. Randomly assign subjects to treatment groups.
III. Replicate the experiment on many subjects to reduce chance variation.

116. What does double-blind mean, and why would we want an experiment to be double-blind?
In a double-blind experiment, neither the subjects nor the people recording the results knows which subject received which treatment. This reduces bias.

117. What is block design?
A block design divides the sample into groups of similar characteristics to reduce the effects of lurking variables. Within each block, units/subjects are randomly assigned to each of the treatment groups.
118. 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.

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 men</td>
<td>random</td>
<td>5 men</td>
</tr>
<tr>
<td>allocation</td>
<td>Treatment A</td>
<td>5 men</td>
</tr>
<tr>
<td></td>
<td>Treatment B</td>
<td>5 women</td>
</tr>
</tbody>
</table>

119. What are the five steps of a simulation?
I. State the problem
II. State the assumptions
III. Assign digits to represent outcomes
IV. Simulate many repetitions
V. State your conclusions

120. 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.) Assign 0 to girl and 1 to boy. Use the command randInt(0, 1). For one trial, press enter until you have two 1’s. Count the number of tries to get two 1’s. Record in table.

121. What is independence?
Two events are independent if knowing that one occurs does not change the probability that the other occurs. \( P(A \mid B) = P(A) \) and \( P(B \mid A) = P(B) \).

122. You are going to flip a coin three times. What is the sample space for each flip? \( S = \{ H, T \} \)

123. You are going to flip a coin three times and note how many heads you get. What is the sample space? \( S = \{ 0, 1, 2, 3 \} \)

124. You are going to flip a coin three times and note what you get on each flip. What is the sample space? \( S = \{ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT \} \)

125. Make a tree diagram for the three flips. (not shown)

126. There are three ways I can drive from Fremont to Grand Rapids and four ways I can drive from Grand Rapids to my home. How many different ways can I drive from Fremont to my home through Grand Rapids? 12

127. How many different four-digit numbers can you make? \( 10^4 = 10,000 \)

128. How many different four-digit numbers can you make without repeating digits? \( 10^4 \times 9 \times 8 \times 7 = 5,040 \)
129. What is an event in probability theory? An outcome or a set of outcomes from a sample space.

130. Any probability is a number between (and including) __0__ and __1__. 

131. All possible outcomes together must have probability of __1__. 

132. If S is the sample space, P(S) = __1__. 

133. What are complements? Give an example and draw a Venn diagram. If A is the event that something occurs, then A complement is the event that it does not occur. Example: rolling a die and landing on an even number is the complement of rolling a die and landing on an odd number. 

134. What are disjoint events? Give an example and draw a Venn diagram. Disjoint events have no outcomes in common. Knowing that one event occurs imply that the other event will not occur. 

Use the following chart for questions 135-137:

<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>

135. What is the probability that an M & M is blue? 0.1 

136. What is the probability that an M & M is red or green? 0.3 

137. What is the probability that an M & M is yellow and orange? 0 

138. 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? 

\[(0.09)(0.92) = 0.0828\] 

139. What assumption are you making in problem 138? Do you think this assumption is valid? Independence, maybe, maybe not.
140. Using two dice, what is the probability that you would roll a sum of seven or eleven?
8/36 = 0.2222

141. Using two dice, what is the probability that you would roll doubles? 1/6 = 0.1667

142. 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? (8/36)(1/6) = 0.0370

143. What assumption are you making in problem 142? Do you think this assumption is valid? Independence. Yes, because what you get on the first roll does not change the probability of what you get on the second roll.

144. Using two dice, what is the probability that you would roll a sum of 7 or 11 that is also doubles? 0

145. What is the union of two events? The event that either one or both occurs.

146. What is an intersection of two events? The event that both occur.

147. How can we test independence? If P(A|B) = P(A) then A and B are independent. Or if P(A and B) = P(A)P(B)

148. Perform an independence test on the smoking/education chart from problem 90 to show that smoking status and education are not independent.
Let A = smokes and B = 4 or more years of college
P(A|B) = 0.1989 P(A) = 0.3782

149. 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
150. 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.) \[
\frac{\pi 10^2}{24(24)} = 0.5454
\]

For problems 151-154 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.

151. Are the events A and B disjoint? Explain. No, the king of hearts is a member of A and B.

152. Are the events A and B independent? Explain. Yes. \(P(\text{Heart}) = \frac{13}{52} = \frac{1}{4}\). \(P(\text{Heart} | \text{a king}) = \frac{1}{4}\). The probability does not change.

153. Are the events A and C disjoint? Explain. Yes, no card can be both a heart and a spade.

154. Are the events A and C independent? Explain. No. \(P(\text{Spade}) = \frac{13}{52}. P(\text{Spade} | \text{Heart}) = 0\).

155. What does the symbol \(\cup\) mean? Union means “or”

156. What does the symbol \(\cap\) mean? Intersection means “and”

157. Give an example of a discrete random variable. The number of students absent per week.

158. Give an example of a continuous random variable. The height of students in a class.

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

<table>
<thead>
<tr>
<th>Grade</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.05</td>
<td>0.28</td>
<td>0.19</td>
<td>0.32</td>
<td>0.16</td>
</tr>
</tbody>
</table>

160. Using problem 159, what is \(P(X > 2)\)? 0.48

161. Using problem 159, what is \(P(X \geq 2)\)? 0.67

162. What is a uniform distribution? Draw a picture. A uniform distribution has constant height.

163. In a uniform distribution with \(0 < X < 1\), what is \(P(0.2 < X < 0.6)\)? 0.4
164. In a uniform distribution with $0 < X < 1$, what is $P(0.2 \leq X \leq 0.6)$? 0.4

165. How do your answers to problems 160, 161, 163, and 164 demonstrate a difference between continuous and discrete random variables? For a discrete random variable, the probability that $X = k$, where $k$ is a constant, could be any number between 0 and 1, inclusive. For a continuous random variable, the probability that $X = k$ is equal to 0.

166. Normal distributions are (continuous or discrete). continuous

167. Expected value is another name for _____. mean

168. Find the expected value of the grades in problem 159. 2.26

169. Find the variance of the grades in problem 159. 1.3724

170. Find the standard deviation of the grades in problem 159. 1.1715

171. What is the law of large numbers? As the number of observations increases, the sample mean $\bar{X}$ approaches the population mean $\mu$, the expected value of $X$ approaches the population mean $\mu$.

172. 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? 12

173. If I charge $2 per book and $1.50 per CD in problem 192, what is my average amount of income per day? $20.50

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

For problems 175-183, 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.

175. What is the average for the two tests added together? 150

176. What is the standard deviation for the two tests added together? 15.6205

177. What is the difference in the test averages? 10

178. What is the standard deviation for the difference in the test averages? 15.6205

179. If I cut the test scores on Test 2 in half and add 50, what is the new average? 85

180. What is the new standard deviation for Test 2 in problem 179? 6

181. If I add 7 points to every Test 1, what is the new standard deviation? 10

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

183. If I multiply every Test 1 by 2 and subtract 80, what is the new standard deviation? 20