Problem 1
Below is a histogram of the lengths, in feet, of 44 Great White Sharks.
(Use this information to answer the following two problems).

1. What proportion of sharks were between 15 and 20 feet? (n=44)
   \( \frac{25}{44} \).

2. Which is the exact value of \( \bar{x} \)?
   \( \bar{x} \) cannot be estimated as we don’t have the values of the 44 observations

Problem 2

3. Which numerical summary is a good measure of center for the following distribution?
   a) Median
   b) Mean
   c) Mode
   d) Both median and mean
   e) Median, mean and mode
Problem 3

4. What is a reasonable action if an outlier is a legitimate data value and represents natural variability for the group and variable measured?
The value should not be discarded; in fact, it may be one of the more interesting values in the data set.

5. List statistics that give information only about the location of a dataset.
Mean, median and mode

6. List statistics that give information only about the spread of a dataset.
IQR, standard deviation and range.

Problem 4
In a survey, students are asked how many hours they study in a typical week. A five-number summary of the responses is: 2, 9, 14, 20, 60.
(Use this information to answer the following two problems).

7. Which of the following is the best estimate for the mean number of hours spent studying in a typical week of the students sampled?
   a) 4
   b) 14
   c) 15.1
   d) 20
   e) 60

8. Fill in the blank in the following sentence. About 75% of the students spent at least ____ hours studying in a typical week.
   a) 9
   b) 14
   c) 20
   d) 45
   e) 60

9. Which measure of center is not resistant to an outlier in the data?
Mean and mode

Problem 5
10. The head circumference (in centimeters) of 15 college-age males was obtained, resulting in the following measurements: 55, 56, 56, 56.5, 57, 57, 57, 57.5, 58, 58, 58, 58.5, 59, 59, 63. If the last measurement (63 cm’s) were incorrectly recorded as 73, which one of the following statistics would change?

a) Q1 (1st quartile)  
   b) **Standard deviation**  
   c) Median  
   d) Q3 (3rd quartile)

Problem 6

11. The following boxplot gives the distribution of the ratings of a new brand of peanut butter for 50 randomly selected consumers (100 points possible with higher points corresponding to a more favorable rating). Identify the five-number summary and the description of each of them using the example.

Min-60 points  
Q1 – 70 points  
Median – 75 points  
Q3 – 80 points  
Max – 90 points

Problem 7

This is a standard deviation contest. You must choose four numbers from the whole numbers 0 to 10, with repeats allowed.

12. Choose four numbers that have the smallest possible standard deviation.
   
a) 7, 7, 7, 8  
b) 3, 5, 7, 9  
c) 1, 1, 1, 1  
d) 1, 2, 3, 4

13. Choose four numbers that have the largest possible standard deviation.
a) 0, 3, 6, 10  
b) 9, 9, 10, 10  
c) 1, 4, 7, 10  
d) 0, 0, 10, 10

Problem 8

Researchers are interested in how crime rates are different for southern states.

14. The standard deviation of unemployment for 35 to 39-year-old for southern states is 8.01 and the standard deviation for non-southern states is 8.76. What do these standard deviations tell you about the southern and non-southern states?

a. The average unemployment rate for 35 to 39-year-old in southern states is about the same as in non-southern states.
b. The average unemployment rate for 35 to 39-year-old in southern states is slightly lower than in non-southern states.
c. The average spread from the mean rate for 35 to 39-year-old in southern states is about the same as in non-southern states.

Problem 9

15. Describe (in words) the distribution in the histograms below and match them to the box plots.

The histogram (a) can be matched to boxplot (2); the histogram (b) can be matched to boxplot (3); and the histogram (c) matched to boxplot (c).

Problem 10

16. College students in a statistics class were asked how many hours of television they watch per week, including online streaming services. This sample yielded an average of 8.28
hours, with a standard deviation of 7.18 hours. Is the distribution of number of hours students watch television weekly symmetric? If not, what shape would you expect this distribution to have? Explain your reasoning.

No, we would expect this distribution to be right skewed. There are two reasons for this: there is a natural boundary at 0 (it is not possible to watch less than 0 hours of TV) and the standard deviation of the distribution is very large compared to the mean.

Problem 11

Daily air quality is measured by the air quality index (AQI) reported by the Environmental Protection Agency. This index reports the pollution level and what associated health effects might be a concern. The index is calculated for five major air pollutants regulated by the Clean Air Act and takes values from 0 to 300, where a higher value indicates lower air quality. AQI was reported for a sample of 91 days in 2011 in Durham, NC. The histogram below shows the distribution of the AQI values on these days.

17. Estimate the median AQI value of this sample.
### About 30

18. **Would you expect the mean AQI value of this sample to be higher or lower than the median? Explain your reasoning.**

Since the distribution is right skewed the mean is higher than the median.

19. **Estimate Q1, Q3, and IQR for the distribution.**

Q1: between 15 and 20, Q3: between 35 and 40, IQR: about 20.

20. **Would any of the days in this sample be considered to have an unusually low or high AQI? Explain your reasoning.**

Values that are considered to be unusually low or high lie more than $1.5 \times \text{IQR}$ away from the quartiles. Upper fence: $Q3 + 1.5 \times \text{IQR} = 37.5 + 1.5 \times 20 = 67.5$; Lower fence: $Q1 - 1.5 \times \text{IQR} = 17.5 + 1.5 \times 20 = -12.5$; The lowest AQI recorded is not lower than 5 and the highest AQI recorded is not higher than 65, which are both within the fences. Therefore none of the days in this sample would be considered to have an unusually low or high AQI.