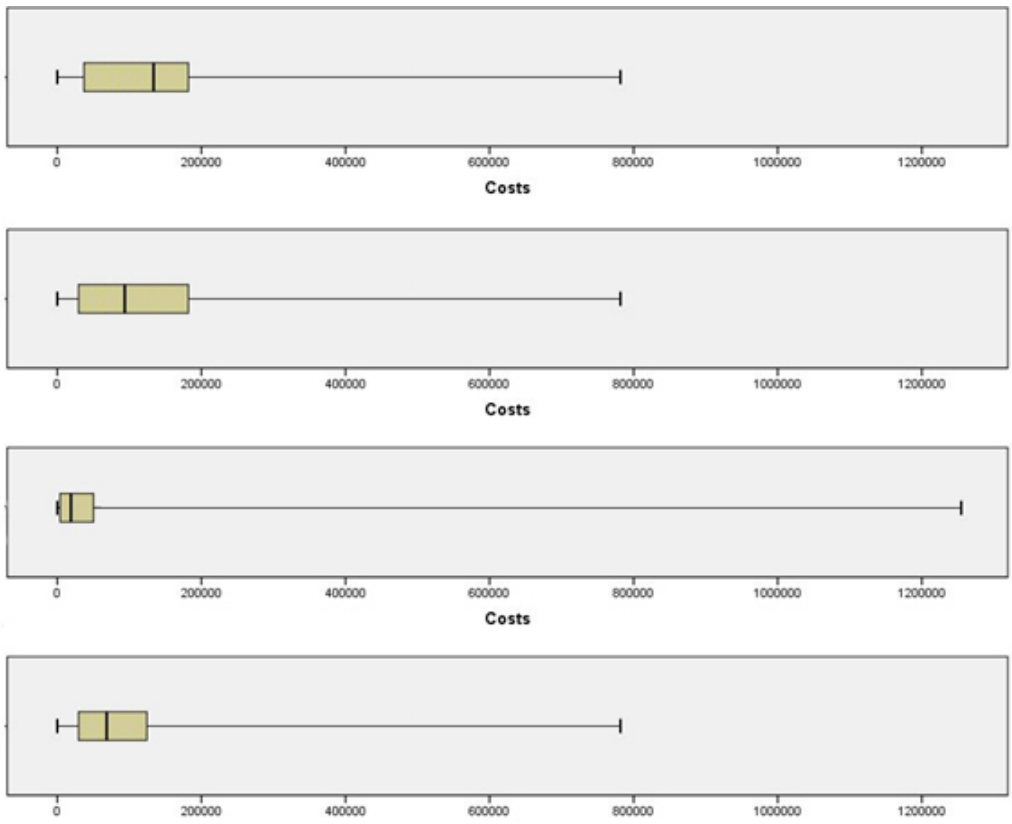
L03: Homework Answer Key

Instructions: You are encouraged to collaborate with other students on the homework, but it is important that you do your own work. Before working with someone else on the assignment, you should attempt each problem on your own.

1. Please explain the term Standard Deviation in your own words.
   The standard deviation is a measure of how spread out the data are. A larger standard deviation indicates that data are more spread out and less consistent than data that have a smaller standard deviation.

On rare occasions, a medical procedure is performed on the wrong body part of a patient's body or on the wrong patient. These are called wrong-site and wrong-patient mistakes. Such errors occur hundreds of times each year across the United States. The medical community is trying to eliminate these errors but have had difficulty reducing their frequency. In a small percentage of these cases, the patient files a lawsuit against the hospital. Philip Stahel et al. conducted a study on these mistakes and the lawsuits that follow. The data in the file WrongSiteWrongPatient represent the amount (in US dollars) hospitals have been required to pay in wrong-site and wrong-patient lawsuits. Some of the values equal zero, indicating that the hospital won the legal battle. Use this information to answer questions 2 through 6.

2. Which of the following boxplots illustrates the wrong-patient lawsuit data?



Box Plot C

3. What is the shape of the distribution of wrong-patient lawsuit costs?
   a. Right Skewed
   b. Symmetric
   c. Left Skewed
   d. Bell Shaped

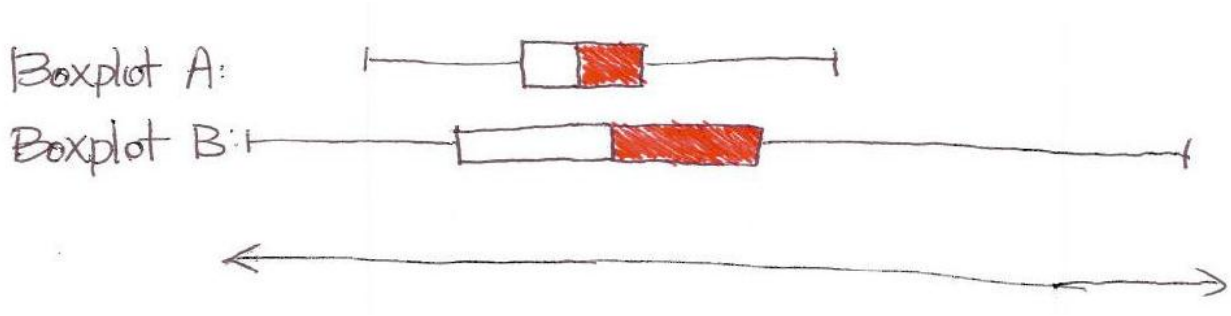
4. Find the mean amount hospitals had to pay in wrong-patient lawsuits. Round your answer to the nearest whole dollar.
   $46,172

5. Find the median amount hospitals had to pay in wrong-patient lawsuits.
   $18,882

6. Find the standard deviation of the amount paid in wrong-patient lawsuits. Round your answer to the nearest whole dollar.
   $105,987

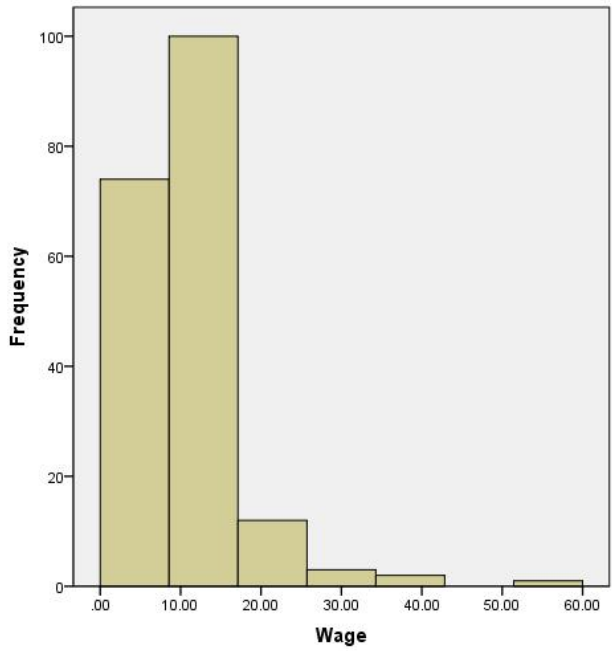
7. If the distribution of the data is left-skewed, we expect the boxplot will:
   a. be symmetric
   b. have a longer whisker on the right side than on the left side
   c. have a longer whisker on the left side

8. The boxplots for two different data sets are given below. Which of these boxplots has a greater percentage of data in the area that is highlighted in red?



   a. Boxplot A
   b. Boxplot B
   c. The percentage of data is the same for both.
   d. There is not enough information to answer this question. We need the original data to make this determination.

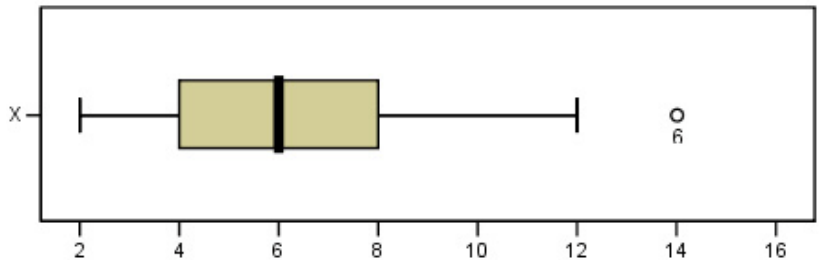
Use the following information to answer Questions 9 and 10. The graph below illustrates typical student hourly wages for BYU-Idaho.



9. Approximately, how many students are in this survey?
   This is just a rough estimate, but somewhere between 190 and 200

10. How would you describe the shape of the distribution of wages?
    Right-skewed

The number of hours students spent studying for an exam were recorded. The data are represented by the boxplot below. Use this boxplot to answer Questions 11 through 13.



11. Find Q1 for the study time data summarized in the boxplot above.
    4 hours

12. One of the observations represented in the boxplot above is a suspected outlier. How long did that student spend studying?
    14 hours

13. The lowest 25% of hours spent on an exam are approximately between what two numbers?
    2 and 4 hours

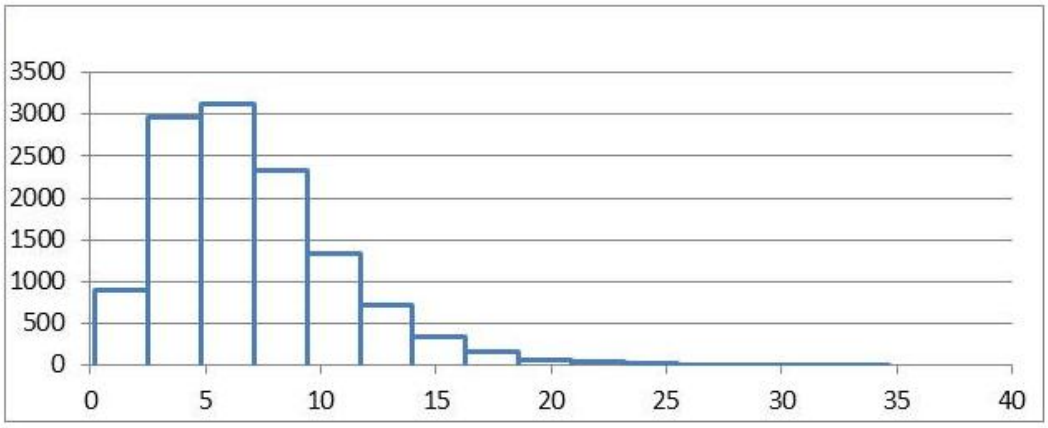
14. What is the mean of the data illustrated in the boxplot?

There is not enough information to answer this question. We need the original data to make this determination.

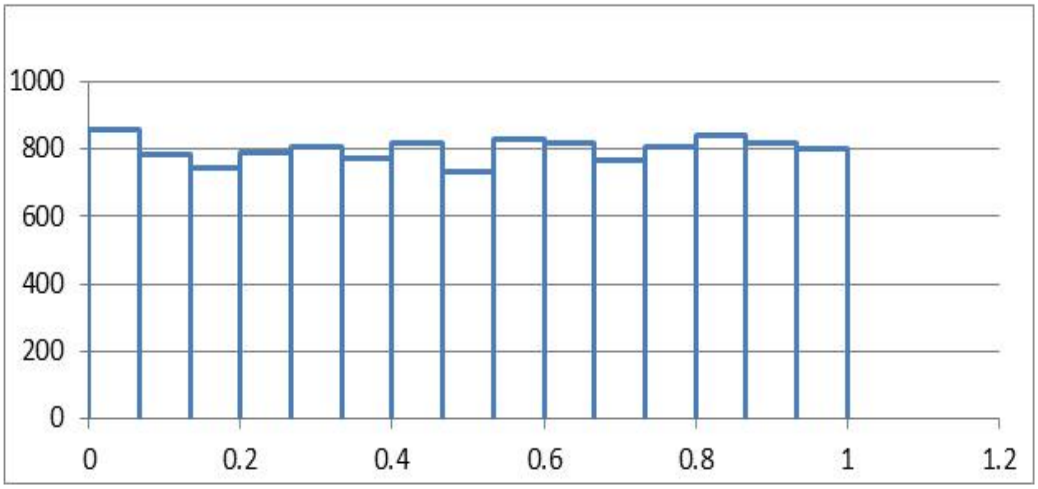
15. Match each graph below to the description of its shape.

Uniform = b
Bell-shaped = d
Right-skewed = a
Left-skewed = e
Symmetric, but not bell-shaped or uniform = c

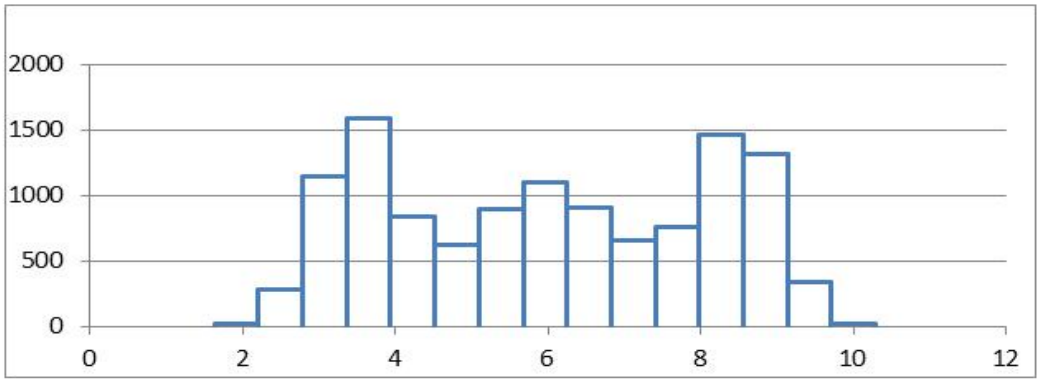
a.



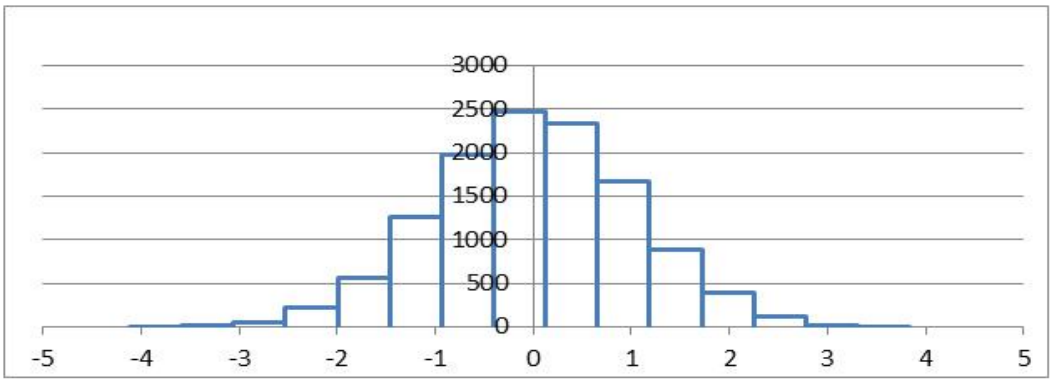
b.



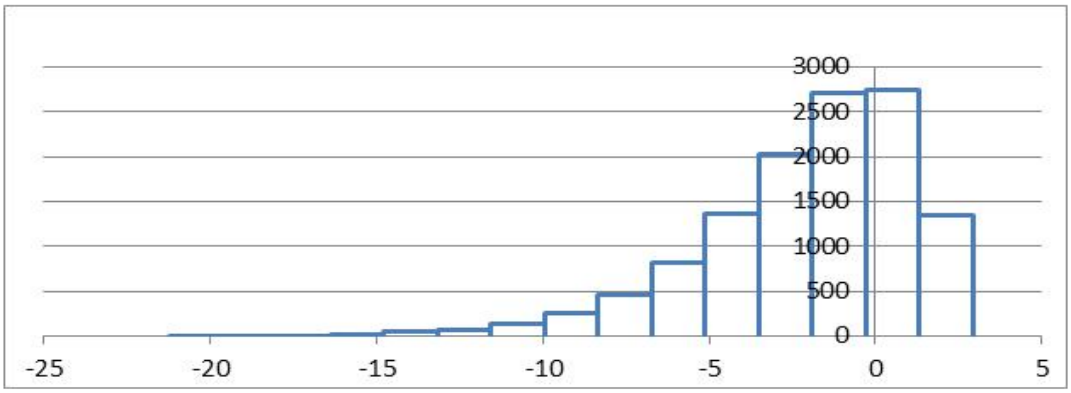
c.



d.



e.



16. A reporter creates a histogram of baseball player salaries and finds that the distribution of salaries is right-skewed. Which measure of center would be greater, the mean or the median?

    Mean

A scientist tested for the presence of many hazardous elements for nuclear reactors. He considered the concentration of Plutonium-238. Plutonium-238 is a radioactive waste generated by a nuclear reactor. He wants there to be as little Plutonium-238 in the tank as possible.

The data below give the concentration of Plutonium-238 in nanocuries per liter (nCi/L) in his sample.

Complete the following table and answer questions 17 through 21 below.

Concentration Level (nCi/L)	Deviation from the mean	Squared Deviations
x	$(x - \bar{x})$	$(x - \bar{x})^2$
9.4		
70.7		"B"
7.8		
4.6	"A"	
50.2		

17. What is the mean of the concentration levels? (Round to 1 decimal place)
    28.5 nCi/L

18. What is the value of the number that goes in the position marked with an "A" in the table above? (Round to 1 decimal place)
    -23.9 nCi/L

19. What is the value of the number that goes in the position marked with a "B" in the table above? (Round to 1 decimal place)
    1777.5nCi/L

20. What is the sample variance of these concentration levels? (Round to 1 decimal place)
    $904.1\ (nCi/L)^2$

21. What is the sample standard deviation of these concentration levels? (Round to 1 decimal place)
    30.1 nCi/L

22. Which of the following sets of numbers has the largest standard deviation? (No calculations are required.)

    a. {7, 8, 9, 10}
    b. {10, 10, 10, 10}
    c. {0, 0, 10, 10}
    d. {0, 1, 2, 3}

For a Math 221 project, one group of students studied the ages of students on the BYU-Idaho campus. They collected data from a random sample of n = 100 students. The sample mean was 21.2 and the sample standard deviation was 2.61. An excerpt of their data is given below.

ID	Gender	Age
1	Female	21
2	Male	18
3	Male	12 (error)
4	Female	20
:	:	:
99	Male	25

The group notices an error in their data. The age of one of the males (ID=3) was entered incorrectly. He is actually 21 years old.

23. When the error is corrected, what will happen to the sample standard deviation?
    a. The standard deviation will increase.
    b. The standard deviation will decrease.
    c. The standard deviation will stay the same.
    d. It is not possible to determine this without the full data set.