1. Name the four kinds of useful information that you can get about a set of measurement data once it has been organized and summarized.
   ANSWER: 1) THE CENTER; 2) THE VARIABILITY; 3) THE SHAPE; AND 4) UNUSUAL VALUES (OUTLIERS).

   Narrative: Quiz scores
   Bob has taken 6 quizzes so far in his statistics class. Each quiz has a possible of 10 total points. Bob’s scores are the following: 10, 8, 9, 7, 2, and 9.

2. Find the three measures of center for Bob’s quiz scores.
   ANSWER: THE MEAN (AVERAGE) IS 7.5; THE MODE IS 9; THE MEDIAN IS 8.5.

3. Explain how (if) Bob’s lowest quiz score affects the mean and the median of this data set.
   ANSWER: THE 2 BRINGS DOWN THE MEAN (AVERAGE); IT DOES NOT AFFECT THE MEDIAN.

4. Which of the following measures of center is affected by an outlier?
   a. Mean
   b. Median
   c. Mode
   d. All of the above
   ANSWER: A

5. The mode is most meaningful for which type of data?
   a. Measurement data
   b. Categorical data
   c. Biased data
   d. None of the above
   ANSWER: B

6. The amount of spread in the data is a measure of what characteristic of a data set?
   a. Center
   b. Variability
   c. Shape
   d. None of the above
   ANSWER: B

7. One or two scores that are far removed from the rest of the data are called __________.
   ANSWER: OUTLIERS

   Narrative: School costs
   Suppose a random sample of liberal arts schools was taken, and the average cost per student was measured for each school. The data are pictured in the histogram below:
8. {School cost narrative} What is the shape of this data set?
ANSWER: SKewed TO THE RIGHT. DATA ALSO HAS AN OUTLIER ON THE HIGH END.

9. {School cost narrative} How many schools were sampled in this study?
ANSWER: 25

10. {School cost narrative} Two measures of center were calculated for this data set and were found to be $26,668 and $34,832. One of them is the mean, and the other is the median. Which one is which, and how do you know that?
ANSWER: $26,668 IS THE MEDIAN AND $34,832 IS THE MEAN. WHY: THE DATA IS SKewed TO THE RIGHT AND HAS AN OUTLIER, BOTH OF WHICH DRIVE UP THE MEAN AND LEAVE THE MEDIAN UNAFFECTED.

11. Which of the following pictures of a data set allows you to retrieve the actual data (assuming no digits are dropped)?
   a. A histogram
   b. A stemplot
   c. Both a) and b)
   d. Neither a) nor b)
   ANSWER: B

12. If the bars of a histogram represent the proportion of the total count that falls into each interval, what must the heights of the bars sum to?
   a. The total number of numbers in the data set.
   b. One.
   c. 1 divided by the total number of intervals used in the histogram.
   d. Not enough information to tell.
   ANSWER: B

13. Which of the following statements is true?
   a. If a data set is skewed to the right, that means there is bias in the results; the data are higher than they should be.
   b. If a data set is skewed to the right, then the higher values are more spread out than the lower values.
   c. If a data set is skewed to the right, then the lower values are more spread out than the higher values.
   d. None of the above.
   ANSWER: B
14. Which of the following statements regarding stemplots is false?
   a. A stemplot allows you to retrieve the original data (assuming no digits are dropped).
   b. A stemplot can never reuse the same stem digit twice.
   c. If a certain value in your data set is repeated three times, it must appear 3 times in appropriate stem and leaf of the stemplot.
   d. None of the above statements are false.
   ANSWER: **B**

15. A __________ is a quick and easy way to put a list of numbers into order while getting a picture of their shape.
   ANSWER: STEM PLOT

16. A data set is __________ if the two halves of the data set (when cut down the middle) are mirror images of each other.
   ANSWER: SYMMETRIC

17. What are the five numbers used in a five-number summary?
   ANSWER: THE LOWEST (MINIMUM); THE HIGHEST (MAXIMUM); THE MEDIAN; THE LOWER QUARTILE; AND THE UPPER QUARTILE.

Narrative: Acceptance rates
A random sample of 50 colleges and universities in the U.S. was selected, and acceptance rates were recorded for each school (percentage of student applicants who were accepted to the school). The following five-number summary was calculated for this data set: lowest = 17; lower quartile = 25.75; median = 36; upper quartile = 47.75; highest = 67.

18. {Acceptance rates narrative} Describe the shape of this data based on the five-number summary.
   ANSWER: THE DATA SET LOOKS FAIRLY EVENLY DISTRIBUTED EXCEPT FOR THE LAST QUARTER OF THE DATA, WHICH IS MORE SPREAD OUT THAN THE REST, INDICATING SKEWNESS TO THE RIGHT.

19. {Acceptance rates narrative} Describe the center of this data set in words a prospective student would understand.
   ANSWER: THE MEDIAN (MIDDLE) OF THE ACCEPTANCE RATES IS 36%. ABOUT HALF OF THE SCHOOLS ACCEPT 36% OR MORE OF APPLICANTS, AND ABOUT HALF ACCEPT 36% OR LESS.

20. {Acceptance rates narrative} Find the range of the acceptance rates and give one possible reason that it is such a high number.
   ANSWER: RANGE = 50; EXPLANATION: ANY REASONABLE ANSWER OK. EXAMPLES: AN OUTLIER WOULD AFFECT THE RANGE; THERE ARE MANY VARIABLES AFFECTING ACCEPTANCE RATES, INCLUDING TYPE OF INSTITUTION (LIBERAL ARTS VS. FOUR-YEAR UNIVERSITY, ETC); REGION OF THE U.S.; QUALITY OF EDUCATION; ETC.

21. Suppose that in a five-number summary you find that a larger gap exists between the third quartile and the highest value than between the lowest value and the first quartile. What does this mean about the shape of the data set?
   a. Symmetric
   b. Skewed right
   c. Skewed left
   d. Not enough information to tell.
   ANSWER: **B**
22. Which of the following does not require the data to be ordered before you can get the right answer?
   a. Mean
   b. Median
   c. Quartiles
   d. Range
   e. All of the above require the data to be ordered.
   ANSWER: A

23. The ________ are the medians of the two halves of an ordered data set.
   ANSWER: QUARTILES

24. A five-number summary involves the lower quartile, the upper quartile, the lowest number, the highest number, and the ________.
   ANSWER: MEDIAN

25. Describe which numerical summaries (statistics) of a dataset a boxplot is based upon.
   ANSWER: THE FIVE-NUMBER SUMMARY (LOWEST NUMBER, LOWER QUARTILE, MEDIAN, UPPER QUARTILE, AND HIGHEST NUMBER).

Narrative: Liberal arts costs
A random sample of 25 liberal arts colleges in the U.S. was selected, and the average cost per student was recorded for each school. The following five-number summary was calculated for this data set: lowest = $17,554; lower quartile = $23,115; median = $26,668; upper quartile = $45,879; highest = $102,262.

26. {Liberal arts costs narrative} Make a boxplot of this data set and use it to discuss the shape of the data.
   ANSWER: BOXPLOT SHOULD HAVE A BOX AROUND THE UPPER AND LOWER QUARTILES WITH A LINE DOWN THE MIDDLE FOR THE MEDIAN, AND LINES GOING OUT TO THE LOWEST AND HIGHEST VALUES, CONNECTING WITH THE BOX, ALL IN THE PROPER SCALE. SHAPE: SKEWED RIGHT.

27. {Liberal arts costs narrative} Using the statistical rule discussed in your textbook, is there an outlier in this data set? Explain your answer.
   ANSWER: YES. THE HIGHEST AMOUNT, $102,262 IS MORE THAN 1.5 TIMES THE IQR AWAY FROM THE UPPER END OF THE BOX.

28. How do you calculate the interquartile range for a data set?
   a. Take the highest value minus the lowest value and divide it by four.
   b. Subtract the value of the lower quartile from the upper quartile.
   c. Subtract the value of the upper quartile from the lower quartile.
   d. Divide the data set into four equal parts and find the range between each of the resulting quarters.
   ANSWER: B

29. If the width of a box in a boxplot is very large, compared to the rest of the boxplot, what does that mean about the shape of the data set?
   a. The data are very spread out in the middle.
   b. The data are clumped tightly in the middle.
   c. The data are not symmetric.
   d. Not enough information to tell.
   ANSWER: A
30. Suppose you look at two boxplots comparing the weights of male cats vs. female cats, and you find that the box for the males is much wider than the box for the females. What does this mean about the data sets?
   a. Male cats weigh more than female cats overall.
   b. Male cats have more variability in their weights than female cats.
   c. Weights of male cats are more skewed than for female cats.
   d. None of the above.
   **ANSWER:** B

31. A(n) ________ is any value that is more than 1.5 times the IQR from the closest end of the box in a boxplot.
   **ANSWER:** OUTLIER

Narrative: Create data
Suppose you can create your own data set by choosing from the numbers 1, 2, 3, 4, and 5. You can repeat a number as many times as you wish, as long as your final data set contains four numbers in it. Here are two examples of data sets you could create: \{1, 2, 3, 4\} or \{1, 1, 5, 5\}.

32. {Create data narrative} Create a data set that has the lowest possible standard deviation.
   **ANSWER:** ANY DATA SET CONTAINING THE SAME 4 NUMBERS IS ACCEPTABLE.
   **EXAMPLE:** 1, 1, 1, 1.

33. {Create data narrative} Create a data set that has mean 3 and standard deviation 0.
   **ANSWER:** 3, 3, 3, 3.

34. Which of the following two data sets has the larger standard deviation: Data Set A= \{1, 1, 5, 5\} or Data Set B= \{1, 3, 3, 5\}?
   **ANSWER:** DATA SET A

35. What is the relationship between the variance and the standard deviation?
   a. The variance is the square root of the standard deviation.
   b. The variance is the square of the standard deviation.
   c. The variance is twice the standard deviation.
   d. There is no relationship between them.
   **ANSWER:** B

36. Suppose a data set is skewed left. What is the most likely relationship between the mean and the median?
   a. The mean is larger than the median.
   b. The mean is smaller than the median.
   c. The mean and the median are not related to each other at all.
   d. The mean and the median are essentially equal.
   **ANSWER:** B

37. Which of the following statements is false?
   a. If the standard deviation is positive, the mean must be positive.
   b. The standard deviation can be negative.
   c. If the mean is large, the standard deviation will be large also.
   d. All of the above are false.
   **ANSWER:** D
38. Which of the following is *not* a measure of spread or variability in a data set?
   a. Standard deviation
   b. IQR
   c. Range
   d. All of the above are measures of spread or variability in a data set.
   **ANSWER:** D

39. Because the ________ can be distorted by high outliers, the center of a data set involving incomes or prices is usually summarized using the ________.  
   **ANSWERS (RESPECTIVELY):** MEAN, MEDIAN

40. If the shape of a data set is ________ then the mean and the median should be about equal.  
   **ANSWER:** SYMMETRIC

41. If you just use an average to describe a set of measurements, is this enough? Explain your answer.  
   **ANSWER:** NO; YOU ALSO NEED TO ADDRESS THE AMOUNT OF VARIABILITY IN THE DATA.