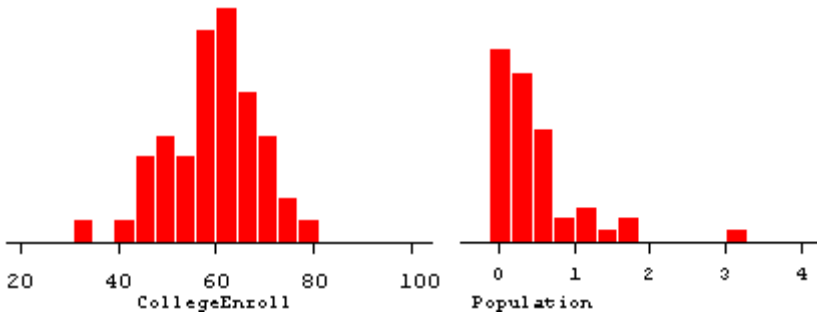


Fill in or circle the correct answer. **Correct answer is in bold italics.**

For questions 1 to 3: Here are data for two variables measured on each of the 50 states in the United States. The histogram on the right shows the populations of the states, in tens of millions. California is the one just above 3, with the largest population, over 30 million. The left-hand histogram shows the percentages of college-aged people who are enrolled in college.



1. For each of the two data sets, is the mean or the median higher? Explain.
College enrollment – is the mean or median higher? Circle one: Mean **Median**
Population – is the mean or median higher? Circle one: **Mean** Median
Explanation: **Outliers pull the mean towards them. (It is obvious in the case of the “Population” values that the outlier will cause the mean to be higher. It is less obvious in the “college enrollment” data that the mean will be lower.)**
2. For the college enrollment data, which of the following is closest to the standard deviation?
a. 1 b. 2 c. 4 **d. 10** e. 20 (Note: this is because the range should be about 4 to 6 standard deviations, based on the empirical rule. The range is just over 40, so the standard deviation is closer to 10 than to any of the other choices.)
3. For the “Population” data, there appears to be an outlier. The most likely reason for this outlier (from the possible explanations given in the text) and the appropriate action are:
 - a. A mistake was made while taking a measurement or entering it into the computer, so discard the value before further analyses.
 - b. The individual state in question belongs to a different group than the bulk of states; discard the value because it is misleading to include it.
 - c. The outlier is a legitimate data value and represents natural variability for the variable measured; don’t discard the value because to do so would be misleading.**
 - d. The outlier is a legitimate data value and represents natural variability for the variable measured, but discard the value because to leave it in would be misleading.
4. Which one of the following variables is not categorical?
 - a. Age of a person.**
 - b. Gender of a person: male or female.
 - c. Choice on a test item: true or false.
 - d. Marital status of a person (single, married, divorced, other)

5. The mean hours of sleep that students get per night is 7 hours, the standard deviation of hours of sleep is 1.5 hours, and the distribution is approximately normal. Give the range of values that should include approximately these percentages of students' sleep times:

68%: from 5.5 hours to 8.5 hours, 95%: from 4 hours to 10 hours, and 99.7%: from 2.5 hours to 11.5 hours.

6. Refer to the previous question. What would be the standardized score for someone who sleeps 4 hours a night? Show your work.

$$z = (4 - 7)/1.5 = -3/1.5 = -2.$$

7. Which one of these statistics is unaffected by outliers?

- Mean
- Interquartile range**
- Standard deviation
- Range

8. A list of 5 pulse rates is: 70, 64, 80, 74, 92. What is the median for this list?

- 74**
- 76
- 77
- 80

9. Which of the following would indicate that a dataset is not bell-shaped?

- The range is equal to 5 standard deviations.
- The range is larger than the interquartile range.
- The mean is much smaller than the median.**
- There are no outliers.

10. The stem and leaf plot below represents the ages of the CEOs of the 60 top-ranked small companies in America in 1993. (Data from the disk in the textbook.) To make it easier for you, the number of value in each part of the stem are shown. Create a 5-number summary to the right of the stem and leaf plot.

	<u>Number of values</u>
3 23	2
3 678	3
4 013344	6
4 555566677788889	14
5 000000112333	12
5 555666677889	12
6 0111223	7
6 99	2
7 04	2

Create 5-number summary here

Median		50	
Quartiles	45.5		57
Extremes	32		74