Answers Explained

1. **(A)** The set is {0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 3}, so the mean is $\frac{11}{19}$ and the median is 0.

2. **(E)** All elements of the sample are taken from the population, and so the smallest value in the sample cannot be less than the smallest value in the population; similarly, the largest value in the sample cannot be greater than the largest value in the population. The interquartile range is the full distance between the first quartile and the third quartile. Outliers are extreme values, and while they may affect the range, they do not affect the interquartile range when the lower and upper quarters have been removed before calculation.

3. **(A)** The value 50 seems to split the area under the histogram in two, so the median is about 50. Furthermore, the histogram is skewed to the left with a tail from 0 to 30.

4. **(B)** Looking at areas under the curve, Q_1 appears to be around 20, the median is around 30, and Q_3 is about 40.

5. **(C)** Looking at areas under the curve, Q_1 appears to be around 10, the median is around 30, and Q_3 is about 50.

6. **(C)** The boxplot indicates that 25% of the data lie in each of the intervals 10–20, 20–35, 35–40, and 40–50. Counting boxes, only histogram C has this distribution.

7. **(D)** The boxplot indicates that 25% of the data lie in each of the intervals 10–15, 15–25, 25–35, and 35–50. Counting boxes, only histogram D has this distribution.

8. **(E)** The boxplot indicates that 25% of the data lie in each of the intervals 10–20, 20–30, 30–40, and 40–50. Counting boxes, only histogram E has this distribution.

9. **(A)** Subtracting 10 from one value and adding 5 to two values leaves the sum of the values unchanged, so the mean will be unchanged. Exactly what values the outliers take will not change what value is in the middle, so the median will be unchanged.

10. **(C)** The high outlier is further from the mean than is the low outlier, so removing both will decrease the mean. However, removing the lowest and highest values will not change what value is in the middle, so the median will be unchanged.

11. **(C)** Adding the same constant to every value increases the mean by that same constant; however, the distances between the increased values and the increased mean stay the same, and so the standard deviation is unchanged.

Graphically, you should picture the whole distribution as moving over by a constant; the mean moves, but the standard deviation (which measures spread) doesn't change.

12. **(E)** Multiplying every value by the same constant multiplies both the mean and the standard deviation by that constant. Graphically, increasing each value by 25% (multiplying by 1.25) both moves and spreads out the distribution.

13. **(E)** The median is somewhere between 20 and 30, but not necessarily at 25. Even a single very large score can result in a mean over 30 and a standard deviation over 10.

14. **(B)** The median is less than the mean, and so the responses are probably skewed to the right; there are a few high guesses, with most of the responses on the lower end of the scale.

15. **(A)** Given that the empirical rule applies, a z-score of -1 has a percentile rank of about 16%. The first quartile Q_1 has a percentile rank of 25%.

16. **(C)** If the variance of a set is zero, all the values in the set are equal. If all the values of the population are equal, the same holds true for any subset; however, if all the values of a subset are the same, this may not be true of the whole population. If all the values in a set are equal, the mean and the median both equal this common value and so equal each other.

17. **(D)** Stemplots and histograms can show gaps and clusters that are hidden when one simply looks at calculations such as mean, median, standard deviation, quartiles, and extremes.

18. **(B)** There are a total of $10 + 17 + 25 + 38 + 27 + 21 + 12 = 150$ students. Their total salary is $10(15{,}000) + 17(20{,}000) + 25(25{,}000) + 38(30{,}000) + 27(35{,}000) + 21(40{,}000) + 12(45{,}000) = \$4{,}580{,}000$. The mean is $\frac{4{,}580{,}000}{150} = \$30{,}533$.

19. **(E)** The mean, standard deviation, variance, and range are all affected by outliers; the median and interquartile range are not.

20. **(C)** Because of the squaring operation in the definition, the standard deviation (and also the variance) can be zero only if all the values in the set are equal.

21. **(A)** The sum of the scores in one class is $20 \times 92 = 1840$, while the sum in the other is $25 \times 83 = 2075$. The total sum is $1840 + 2075 = 3915$. There are $20 + 25 = 45$ students, and so the average score is $\frac{3915}{45} = 87$.

22. **(A)** A distribution spread thinly on the high end is a skewed distribution with the mean greater than the median.

23. **(E)** None are reasonable because we are not looking at a random sample of states but rather only at the seven lowest values, that is, at the very low end of the tail of the whole distribution. (The distribution was actually skewed to the right, with California having 159,700 bankruptcies, followed by New York with 51,300, and all other states below 50,000.)

24. **(B)** Increasing every value by 5 gives 10% between 45 and 65, and then doubling gives 10% between 90 and 130.

25. **(A)** A distribution spread thinly on the high end is a skewed distribution with the mean greater than the median.

26. **(E)** All three statements are true.

27. **(A)** Dotplots and stemplots retain the identity of individual scores; however, histograms and boxplots do not.

28. **(A)** $206 + 2.69(35) = 300$; $206 - 1.13(35) = 166$.

29. **(B)** A calculator gives a mean of 18.875 with a standard deviation of 5.472, and so the z-score is $\frac{14.23-18.875}{5.472} = -0.85$. Note that in this case "top-ranked" brands have more negative z-scores.

30. **(A)** Bar charts are used for categorical variables.

31. **(C)** The median corresponds to the 0.5 cumulative proportion.

32. **(A)** The 0.25 and 0.75 cumulative proportions correspond to $Q_1 = 1.8$ and $Q_3 = 2.8$, respectively, and so the interquartile range is $2.8 - 1.8 = 1.0$.