Data Inferences Answer Key

1. **C.** If the study was repeated many times with randomly selected groups of African elephants, they would expect the true mean weight within the *margin of error around* the sample mean 98% of the time. Thus, it is highly likely that the true mean is within the bounds of the *margin of error around* the mean in this particular random sample. The interval around the sample mean is called the "confidence interval." In this case, the confidence interval is 11982 ± 26.

$$11982 + 26 = 12008$$
$$11982 - 26 = 11956$$

It is highly likely that the true mean is between 12008 pounds and 11956 pounds. Only option C of 11978 falls in this range. Choices A, B, and D are incorrect because none of them fall within the confidence interval.

2. **C.** Since the random sample showed that 82% of the rides arrived within 10 minutes of their scheduled arrival times, it is appropriate to believe that somewhere around 82% of *all* the bus rides did so too. It is highly unlikely that the percent from the sample exactly matches the percent from the full populations, however. The margin of error lets us know an interval around the approximation in which we can be reasonably sure that the true population percentage falls. We call that the "confidence interval." The confidence interval for this study can be calculated. We are given that the margin of error is 2%, so we know that the confidence interval extends 2% above and below the percentage from the sample, which was 82%.

$$82\% - 2\% = 80\%$$
$$82\% + 2\% = 84\%$$

With 94% confidence, we can say that between 80% and 84% of all LA bus rides arrived within 10 minutes of their scheduled times. Choice A and B are incorrect because the answer should be a range represented by a confidence interval. Choice D is incorrect because the confidence interval is calculated from the sample mean, not from the confidence level.

3. **A**. If the poll was repeated many times with randomly selected groups of high schoolers, we would expect the percentage of high schoolers who report that they eat breakfast in the morning to be within the margin of error around the sample percentage 90% of the time. Thus, it is reasonable to claim that the true percentage is within the margin of error around the sample percentage in this particular poll. This interval is called the "confidence interval," which can be calculated with the equation 33.7 ± 4.5.

$$33.7 + 4.5 = 38.2$$
$$33.7 - 4.5 = 29.2$$

It is highly likely that the true percentage of high school students who report to eat breakfast in the morning is between 29.2% and 38.2%. Only option A of 38% falls within this range. Choices B, C, and D are incorrect because none of those answers fall within the confidence interval.

4. **A**. If a random sample was selected multiple times and this study was repeated, the true mean can be expected to be within the margin of error around the sample 97% of the time. Thus, the true mean is likely within the margin of error around the mean in this particular study in an interval called the "confidence interval," calculated by 19.9 ± 2.9.

$$19.9 + 2.9 = 22.8$$
$$19.9 - 2.9 = 17$$

It is very likely for the true mean to be between 17 mpg and 22.8 mpg. Only option A of 17.3 mpg falls within this range. Choices B, C, and D are incorrect because none of those answers fall within the confidence interval.

5. **D.** The average of two number is the sum of the two numbers divided by 2. From what we are given, we can establish the equations $a = \frac{4x+19}{2}, b = \frac{\left(\frac{3x}{2}\right)+8}{2}, c = \frac{\frac{13x}{2}+27}{2}$. The average of a, b, c are given by

$\frac{a+b+c}{3}$. By substitution, we get the expression $\frac{\frac{4x+19}{2}+\frac{\left(\frac{3x}{2}\right)+8}{2}+\frac{\frac{13x}{2}+27}{2}}{3}$ which can be simplified to $\frac{12x+54}{6}=2x+9$. Choices A, B, and C are incorrect and may be due to algebra errors.

6. **D**. Since the mean for lion who have been living at the preserve for less than 1 year is lower than that of lions who have been living at the preserve for a year or more, then the combined mean cannot be greater than or equal to that of lions who have been living at the preserve for a year of more. By the same logic, the combined mean also cannot be less than or equal to the mean of the lions who have been living at the preserve for less than 1 year. Therefore, the combined mean must be between the two separate means. Only choice D gives a correct range without making incorrect assumptions. Choice A is incorrect and results from finding the mean of the two means. This answer makes an unjustified assumption that there are an equal number of lions who have been living at the preserve for less than a year and for lions who have been living at the preserve more than a year. Choice B is incorrect and makes an unjustified assumption that there are more lions living at the preserve for less than a year than lions living at the preserve for more than a year. Choice C is incorrect and makes an unjustified assumption that there are less lions living at the preserve for less than a year than lions living at the preserve for more than a year.

7. **C.** The mean of a data set is the sum of the values divided by the number of values. The mean of 7.5 is obtained by finding the sum of the first 10 ratings and dividing it by 10. Thus, the sum of the first 10 ratings was 75. In order for the mean of the first 20 ratings to be at least 8.5, the sum of the first 20 ratings must be at least $(8.5)(20)=170$. Therefore, the sum of the next 10 ratings must be at least $170-75=95.$ The maximum rating is 10, so the maximum possible value of the sum of the 12th through 20th rating is $9\times10=90$. Therefore, for the professor to be able to have an average rating of at least 8.5 for the first 20 ratings, the least possible value for the 11th rating is $95-90=5.$ Choices A, B, and D are incorrect and result from algebra errors.

8. **C**. The mean and median values of a data set are only equal when there is a symmetrical distribution, such as in a normal distribution. When the mean and median values are not equal, then this indicates that the distribution is not symmetrical because there exist outliers that pull the mean in either direction (smaller or larger) while the median remains the same. In this question, the mean is larger than the median. Thus, it is reasonable to conclude that there are large outliers in the data set that are pulling the mean towards a larger value. In context, that means there are a few cars that are valued at a higher price than the rest. Choice A is incorrect because the closeness of the values does not contribute to the difference between the mean and median values. Choice B is incorrect because there are a few cars that are valued much higher, not lower, than the rest. Choice D is incorrect because the number of values of car prices between \$18,000 and \$25,000 does not contribute to the difference between the mean and median values.

9. **B**. We know that the mean of the 4 players is calculated by taking the sum of the 4 individual scores and dividing the sum by 4. We can set up the equation $\frac{sum\,of\ 4\ players}{4}=16.8$. We know that $sum\,of\ 4\ players=(16.8)(4)=67.2$. We also know $\frac{sum\,of\ 3\ players}{3}=12.2$, $sum\,of\ 3\ players=(12.2)(3)=36.6.$ To find the highest score, we subtract $67.2-36.6=30.6$. Choice A is incorrect because 12.2 is the new mean. Choice C is incorrect because it is the sum of the three players. Choice D is incorrect because it is the sum of the four players. The question is asking for the score of the player who left, and thus the difference between the sum of four players and the sum of three players.

10. **B**. If the study was repeated many times with randomly selected measurements, they would expect the true mean weight within the *margin of error around* the sample mean 90% of the time. Thus, it is highly likely that the true mean is within the bounds of the *margin of error around* the mean in this particular random sample.

The interval around the sample mean is called the "confidence interval." In this case, the confidence interval is 2500 ± 160.

$$2500 + 160 = 2660$$
$$2500 - 160 = 2340$$

It is likely that the true mean would be between 2340 and 2660 pounds of vegetation per year. With 90% confidence, we can say that the annual pounds of vegetation consumed by deer in that particular area is between 2340 and 2660 pounds. Choices A, C, and D are incorrect because none of them give the correct confidence interval which may have resulted from algebra errors.