1. Which of the following statements defines the 75th percentile value for a data sample?

- ○ The value that is 25% larger than the sample mean
- ○ The value that is 75% of the largest observation in the sample.
- ◉ The value such that 25% of the sample observations are larger than this value, and 75% of the sample observations are smaller than this value.
- ○ The value such that 75% of the sample observations are larger than this value, and 25% of the sample observations are smaller than this value.

A researcher collects the following data on 50 subjects:

Bin	Number of Observations in Bin	Percentage of Observations in Bin	Data points
0-10	8	16%	1, 2, 5, 5, 7, 8, 9, 10
>10-20	15	30%	11, 11, 12, 13, 14, 14, 15, 15, 16, 17, 17, 17, 18, 19, 20
>20-30	9	18%	22, 22, 23, 24, 25, 26, 27, 27, 30
>30-40	7	14%	31, 33, 33, 33, 34, 37, 37, 38
>40-50	7	14%	42, 43, 44, 44, 45, 47, 49
>50-60	3	6%	51, 51, 52
>60-70	0		
>70-80	0		
>80-90	1	2%	89

A histogram representing this data would have the following shape:

- ○ left skewed
- ◉ right skewed
- ○ Symmetric and bell-shaped

A researcher collects the following data on 50 subjects:

Bin	Number of Observations in Bin	Percentage of Observations in Bin	Data points
0-10	8	16%	1, 2, 5, 5, 7, 8, 9, 10
>10-20	15	30%	11, 11, 12, 13, 14, 14, 15, 15, 16, 17, 17, 17, 18, 19, 20
>20-30	9	18%	22, 22, 23, 24, 25, 26, 27, 27, 30
>30-40	7	14%	31, 33, 33, 33, 34, 37, 37, 38
>40-50	7	14%	42, 43, 44, 44, 45, 47, 49
>50-60	3	6%	51, 51, 52
>60-70	0		
>70-80	0		
>80-90	1	2%	89

The value of the upper hinge of the boxplot for these data would fall in what data bin?

- ○ The upper hinge of the box would be in the ">50-60" bin
- ○ The upper hinge of the box would be in the ">0-10" bin
- ◉ The upper hinge of the box would be in the ">30-40" bin

4. A random sample of 200 students from first year medical students yielded a mean undergraduate GPA of 3.85 and standard deviation 0.35. If a larger sample of 500 students was taken from the sample population, which of the following would be true?

- ○ The mean undergraduate GPA of the larger sample would be systematically larger than the sample of 200 students.
- ○ The sample standard deviation of the larger sample would be systematically larger than the sample of 200 students.
- ○ The mean undergraduate GPA and the standard deviation of the larger sample would be systematically smaller than the sample of 200 students.
- ◉ None of the above

5. Which of the following is true about $\bar{x}$, the sample mean?

- ◉ The sample mean, $\bar{x}$, is the best estimate of the underlying population mean, based on the sample data.
- ○ The sample mean, $\bar{x}$, will always equal the underlying population mean.
- ○ The sample mean, $\bar{x}$, tends to decrease in value with increasing sample size.
- ○ The sample mean, $\bar{x}$, tends to increase in value with increasing sample size.

6. The distribution of GPA taken from a random sample of 200 first year medical students left skewed distribution. If a sample of 500 students were taken, how would the distribution differ?

- (●) The distribution of GPA from the larger sample would also be left-skewed, but would be more 'filled in', i.e. more densely populated with data points.
- ( ) The distribution could be left skewed or right skewed, since the skewness does not systematically increase or decrease based on sample size alone.
- ( ) The distribution of 500 students would be less skewed than the sample of 200 students.
- ( ) The distribution of 500 students would be more symmetric than the sample of 200 students.

7. 500 randomly selected university students were selected for a study on smoking status and stress levels. At the time of the study, students were asked to rate their stress levels on a scale of 0 – 100 with 0 indicating no stress. Students were also asked to disclose whether they were a non-smoker, a light smoker, or a heavy smoker.

Of the 500 students, 75 identified themselves as heavy smokers, 150 identified themselves as light smokers, and 275 identified themselves as non-smokers. The average stress level reported for heavy smokers was 60.4. The average stress level reported for light smokers was 45.7. The average stress level reported by non-smokers was 56.6. Which of the following statements is true?

- ( ) On average, heavy smokers have stress values of 3.8 times the stress levels of non-smokers.
- (◉) On average, heavy smokers have stress values of 3.8 units greater than non-smokers.
- ( ) On average, heavy smokers have stress values of 3.8 times less than the stress levels of non-smokers.
- ( ) On average, heavy smokers have stress values of 3.8 unit less than non-smokers.

8. The distributions of male and female blood glucose levels are similar in shape. The mean difference between male and female blood glucose levels was 10.4 (females are reference group). This difference represents which of the following (check all that apply)?

- [x] On average, males have 10.4 mg/dl higher blood glucose levels than females.
- [x] On average, females have 10.4 mg/dl lower blood glucose levels than males.
- [ ] All males have 10.4 higher mg/dl blood glucose than all females.
- [x] The distribution of female glucose values would be shifted approximately 10.4 units less than the mean of males.

9. The duration of time from first exposure to HIV infection to AIDS diagnosis is called the *incubation period.* The incubation periods of a random sample of 7 HIV infected individuals is given below (in years):

12.0 10.5 13.5 12.5. 9.5 6.3 7.2

Suppose another random sample of 100 persons is taken from the same population, and added to the sample of 7, for a total sample of 107 HIV infected individuals. How will the sample mean of 107 values compare in value to the sample mean from the original sample of n=7?

- [ ] It will be smaller
- [ ] It will be larger
- [ ] It will be exactly the same value
- [x] While the two values should be "similar", there is no way to predict exactly how the two values will compare

10. The duration of time from first exposure to HIV infection to AIDS diagnosis is called the *incubation period.* The incubation periods of a random sample of 7 HIV infected individuals is given below (in years):

    12.0 10.5 13.5 12.5 9.5 6.3 7.2

    Suppose another random sample of 100 persons is taken from the same population, and added to the sample of 7, for a total sample of 107 HIV infected individuals. How will the sample standard deviation of the 107 values compare in value to the sample standard deviation from the original sample of n=7?

- [ ] It will be larger
- [ ] It will be smaller
- [ ] It will be exactly the same value
- [x] While the two values should be "similar", there is no way to predict exactly how the two values will compare