Names:_____

MATH 232 · Introduction to Statistics | Spring 2017 | Instructor: E.M. Kiley

Week 4: Cartoon Guide Questions

Please read pages 19–26 of the Cartoon Guide with your group. The reading is a bit denser this time, and some of the concepts will be new to you, so do not rush through it; spend about 25 minutes on the initial reading, then answer the questions.

Question 1. Give an example consisting of two data sets that have the same mean, but different standard deviations. Show or explain how your example fits this description.

Question 2. On page 21, it is stated that John Tukey was the inventor of the box-and-whisker plot. You've seen one other example (from the Cartoon Guide!) of a type of plot invented by Tukey. What was it?

Question 3. On page 17, the median is referred to as the "midpoint" of the data. That is to say, half of the values in a data set are lower than the median, and half are higher than the median. This problem has several questions that ask you to explore similar interpretations for the quartiles defined on page 20. What proportion of data should be greater than the first quartile? Less than the first quartile? What proportion of data should lie between the first and third quartiles? What proportion of data should be greater than the third quartile? Less than the third quartile?

Question 4. Do the "whiskers" on box-and-whisker plots *always* extend 1.5x the interquartile range? If not, then what does it mean when two box plots look similar, but one has shorter whiskers than the other?

Question 5. Using your worksheet from last week to obtain the data set corresponding to the heights (in inches) of students in our class, compute the quartiles representing that data, and compute the interquartile range.

Question 6. On page 21, a definition of the term *outlier* is given; this definition is more precise than the one we had in the lecture notes this week. Restate that definition here, and if there are any outliers in the data set you used for Problem 5, state them here.

Question 7. What issue is resolved by defining the standard deviation as the square root of the variance?

Question 8. On a TI-brand calculator, the standard deviation can be computed easily using the corresponding function in the "STAT" menu. But for small data sets, it is sometimes easy enough to compute the standard deviation by hand. First, compute the mean; then, for each data point, find the distance from that point to the mean; square each of those values, and sum the squares; divide that sum by the number of data points minus one; take the square root of that quotient (this is what the formula on page 23 boils down to; an example is given there and on page 22). Compute the standard deviation of the following data set (either by hand, or with your calculator): $\{6, 2, 3, 1\}$.

Question 9. If two data points from the same set have z-scores of 1.26 and -0.4, respectively, then which point is lower? Why? Which point is further from the mean?

Question 10. Another definition of the term "outlier" is given on page 25. How does it differ from the previous definitions? Do you expect to see a great difference, in practice?