Practice Problems: Exam 1

1. Consider the diamonds data set, available on the class website.

   (a) Is the data set structured or unstructured? Why?

   (b) What are the cases?

   (c) What are the variables?

   (d) What kind of variable is each variable? Be specific!

   (e) What kind of graph(s) would you use to visualize each kind of variable?

   (f) If you wanted to use StatCrunch to visualize the distribution of the cut variable with a pie chart, would you choose "with data," or "with summary?" How would the data need to be structured to compel the other choice?

   (g) Are there any identifier variables in the diamonds data set?

2. What feature of a QQ plot (normal probability plot) indicates that the distribution of the variable in question is approximately normal.

3. Boxplots.

   (a) What feature of a boxplot indicates median?

   (b) What feature of a boxplot indicates Q1?

   (c) What feature of a boxplot indicates Q3?

   (d) What feature of a boxplot indicates the interquartile range (IQR)?

   (e) What feature of a boxplot indicates the maximum?

   (f) How are the fences determined?

4. Histograms.

   (a) What features of a histogram indicate that the distribution of the variable in question is symmetric?

   (b) What features of a histogram indicate that the distribution of the variable in question is skewed? Skewed left or right?

   (c) What features of a histogram indicate the distribution of the variable is unimodal? Bimodal? Multimodal? Uniform?

   (d) How can standardizing a variable (converting to z-scores) change the shape of a histogram?

5. Explain the difference between resistant and not resistant and give example statistics for each type.

6. Give the name for explain the condition listed in the book for using a normal model to describe what values to expect from a variable.

7. A standardized test is given to a large population of students across the USA. After an analysis of the data, researchers determine that a Normal model fits the scores well. The scores have a mean of 1300 points and a standard deviation of 50 points.

   (a) What percentile does a score of 1400 lie on?

   (b) What z-score does a score of 1200 have?