Lab 1 - Working with Data   **Math 121**

The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey of 350,000 people in the United States conducted by the Center for Disease Control (CDC). In this lab we will take a look at a random sample of 200 individuals who completed the survey in 2000.

In particular, we will looked at the height, weight, age, gender of respondants, as well as their general health, whether or not they have a health plan, whether they have exercised in the past month, and whether they have smoked more than 100 cigarettes in their lives. Note that questions with yes or no answers use a 1 to represent yes and 0 to represent no. The data for this lab can be found at:

<div align="center">

`http://people.hsc.edu/faculty-staff/blins/mayterm14/Labs/cdc.xls`

</div>

1. How many variables are there? For each variable, determine whether it is categorical or quantitative.

2. In the sample, what proportion of people have health plans?

3. Make a pie chart and a bar graph showing the distribution of general health of people in this sample. Which chart do you think is more appropriate?

4. Find the five number summary for height, and then display the result as a box-and-whisker plot.

5. What is the interquartile range for height? What is the standard deviation?

6. One person in this sample is an outlier for height. Describe how the mean, median, standard deviation, and IQR for height change when the outlier is removed from the calculations. Which of the four statistics (mean, median, IQR, SD) appears to be the most sensitive to outliers?

7. Make a histogram for weight. Does the weight distribution look approximately normal, or is it skewed left or right?

8. Which is larger, the mean or median weight? Does this answer make sense in light of your description of the distribution above? Explain.