Lab 2 - Associations in Data   **Math 121**

In this lab we will continue to investigate our random sample of 200 individuals who participated in the 2000 CDC Behavioral Risk Factor Surveillance System. The data for this lab can be found at:

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

1. Find the correlation coefficients for each of the followings pairs of variables.

    (a) Height and Weight

    (b) Height and Age

    (c) Age and Weight

2. Make a scatterplot with height on the x-axis and weight on the y-axis. Is there an outlier in the data? (You don't need to copy the scatterplot, but describe the height and weight of the outlier.)

3. Make a second scatterplot for height and weight with the outlier removed. Then use Excel to find the trendlines for each scatterplot. How much does the slope of the trendline change when you remove the outlier? Would you say that the outlier was influential?

4. Use Excel to compute the standard deviations for height and age and the correlation coefficient between height and age.

5. Find the formula for the least squares regression line for height as a function of age (age on x-axis, height on the y-axis). What are the units of the slope? Explain what the slope means in words.

6. Using the regression line from the last problem, estimate the average height of people who are 80 years old.

7. What percent of the variability in height is explained by age?

8. The two way table below compares general heath versus whether or not individuals have smoked at least 100 cigarettes in their lives.

	Smoked 100 Cigarettes	
General Health	No	Yes
Excellent	38	20
Very good	37	27
Good	24	32
Fair	7	12
Poor	0	3

Find the column proportions for the two way table, and then draw a segmented bar graph showing the results.

9. Based on the segmented bar graph above, does it look like there is an association between general health and smoking? Describe it.