INTRODUCTION TO STATISTICAL REASONING

36-201 **Lab #4 -Partial Solutions**

Part I: Introduction to Scatterplots

Question #1 As time increased I would expect people to forget, so that the average score would decrease.

Questions #3 The relationship between score and time is a decreasing relationship. As time goes on, i.e., gets bigger, the score gets smaller. The relationship is *non*-linear (follows a curve instead of a straight line).

Question #4-5 The correlation coefficient is negative indicating that as time increases, score decreases, i.e., people make more errors. This agrees with the scatterplot. The relationship is pretty strong since the correlation coefficient ($r = -0.756$) is close to -1. Note, however, that although the correlation coefficient is descriptive in this case, it needs to be used cautiously since the scatterplot shows that the relationship is nonlinear (follows a curve instead of a straight line).

Part II: Scatterplots and Regression

Question #6 Gender is categorical. Height and Weight are both quantitative.

Question #7 (i) The mean is probably greater than the median, since the distribution of weights is skewed to the right. (ii) The mean is probably less than the median, because it is pulled down by the low outlier.

Question #8 Height is the explanatory variable and weight is the response.

Question #9 From the scatterplot we observe a positive association between Weight and Height. As height increases we see an increasing trend in weight. There is one student who is about 54 inches tall and about 120 pounds who appears to be an outlier relative to the general scatter of points. The form of the association appears linear. One interesting feature of the scatterplot is that as height increases the scatter seems to increase as well. In other words the points are more tightly clustered at shorter heights than they are are taller heights.

Questions #10–11 The median trace plot shows an increase in weight with an increase in height. This increase is fairly slight for low values of height and more distinct for higher values of height. The regression line seems to fit the data fairly well.

Questions #12–13 The regression equation is:

$$weight = -179 + 4.77(height).$$

For someone 65 inches tall, their predicted weight is about 131 pounds. The correlation coefficient between weight and height is 0.72, indicating a relatively strong linear association. The correlation coefficient is positive indicating a positive relationship between the two variables which agrees with what we see in the scatterplot, weight increases as height increases.

Questions #15–16 The plus signs in the scatterplot correspond to the females. We see, as we might have expected, that the taller, heavier students are predominantly males and the shorter, lighter students are predominantly females. Between 65 and 70 inches there is the most overlap among the male and the female students. For the females students it looks as if as height increases we do not see as steep an increase in weight as we do for the male students.