Problems to be handed in:

1) Submit electronically exercises 67-71 from Unit A5 of CyberStats.

A5 Ex. 67. What is the median, approximately?

Solution(1): From the graph it appears to be about 12%

A5 Ex. 68. Are there any notable outliers?

Solution(1): There aren’t any evident in this plot.

A5 Ex. 69. Is there any skewness?

Solution(1): Yes, towards the higher values.

A5 Ex. 70. Now find the boxplot for the Employment variable. What is the median approximately?

Solution(1): From the graph it appears to be about 64%

A5 Ex. 71. Is there any skewness?

Solution(1): Maybe a little towards the lower values.

2) This question involves the same information on the poverty and employment in U.S. States as questions 67-71 from Unit A5 of CyberStats. There is some background information on the data set linked there. In brief, the data set contains several measurements on the 50 states plus D.C. collected in 1994/95. The Poverty variable is the percentage of people below the poverty line and Employment is percentage of people employed. This is an ongoing process - see http://www.census.gov/statab/www/ranks.html.

Create a grouped version of the Employment variable using the Data > Bin Columns menu option. This creates an ordinal categorical variable with three categories of Employment (50 to 60, 60 to 70, 70 to 80). This new variable is called Bin(Employment). Using the Graphics > Boxplot menu option, create a boxplot of Poverty, Grouped by: Bin(Employment). Save the boxplot in your homework report.

Is there a relationship between the levels of Employment and Poverty? How does the variation in Poverty depend on the level of Employment? What is the approximate ratio of IQR for the States with employment from 70 to 80 to the group of States with employment from 50 to 60?
**Solution (7):** There is a negative relationship between levels of Employment and Poverty - states with higher employment levels tend to have lower poverty levels. This is true whether you look at the mean or the median poverty for each unemployment level.

The variation in Poverty also depends on the level of Employment. The boxplot for the higher employment levels is less spread out than the boxplot for the lower unemployment level. That is, higher levels of employment are associated with lower variation in Poverty rate.

The approximate ratio of IQR for the States with employment from 70 to 80 to the group of States with employment from 50 to 60 is about 1.5 to 4 or about 0.3. The exact values are 1.45, 4.7 and 0.308. These can be calculated using the Stat > Columns menu option, for Poverty Grouped by: Bin(Employment).

3) Exercise 5.1 from page 161 in Chapter 5: “Relationships between Quantitative Variables” of MOS.

a) Amount of alcohol consumed and performance on a test of coordination

**Solution (1):** Negative, because coordination will decrease when amount of alcohol consumed is increased.

b) Height and grade point average for college students.

**Solution (1):** No association would be expected between height and grade point average.

c) Miles of running per week and time for a 5-kilometer run.

**Solution (1):** Negative because as amount of training for running increases, the average time will decrease.

d) Forearm length and foot length.

**Solution (1):** Positive, because bodies are generally proportional. Taller people are likely to have longer forearms and longer feet while shorter people are likely to have relatively shorter forearms and shorter feet.

4) Exercise 5.27 from page 164 in Chapter 5: “Relationships between Quantitative Variables” of MOS.

a) Which graph shows the strongest relationship between the two variables? Which graph shows the weakest?

**Solution (1):** Graph 2 shows the strongest relationship while Graph 3 shows the weakest.

b) In scrambled order, correlation values for these four graphs are -0.9, 0.0, +0.3, +0.6. Match these correlation values to the graphs.
Solution (2): Graph 1: +0.6; Graph 2: -0.9; Graph 3: 0; Graph 4: +0.3.

Exercise 5.29 from page 165 in Chapter 5: “Relationships between Quantitative Variables” of MOS.

Solution (4): The formula for $r^2$ is

$$r^2 = \frac{SSTO - SSE}{SSTO}.$$ 

So in the case mentioned,

$$r^2 = \frac{800 - 200}{800} = \frac{600}{800} = 0.75.$$

Exercise 5.38 from page 165 in Chapter 5: “Relationships between Quantitative Variables” of MOS.

Solution (5): The negative correlation might occur due to inappropriately combining groups (males and females). Perhaps females, who are generally shorter than males, did better on the memorization test. (See example 17.7, pages 529-30 of the text, for an example of a study where this was the case.) For the data in the following example sketch, the overall correlation is -0.685, but within each gender the correlation is close to 0.
7) Exercise 5.46 from page 166 in Chapter 5: “Relationships between Quantitative Variables” of MOS.

a) How much does average foot length increase for each 1-inch increase in height?

Solution: Average foot length increases 0.384 centimeters for each one-inch increase in height. This is the slope of the line. Consider person A with height 0 inches and person B with height 1 inch. Then the difference in heights between person A and person B is 1 inch. What’s the difference in their predicted foot lengths? Person A has foot length \( \hat{y} = 0.25 + 0.384 \times 0 = 0.25 \) cm. Person B has foot length \( \hat{y} = 0.25 + 0.384 \times 1 = 0.25 + 0.384 \) cm. So the difference in foot lengths is 0.384. (Notice that the intercept cancels out!)

b) Predict the difference in the foot lengths of men whose heights differ by 10 inches.

Solution: The predicted difference in foot lengths is \((10)(0.384)=3.84\) cm.

c) What is the predicted foot length for Max? What is the value of the prediction error (residual) for Max?

Solution: The predicted foot length is \( \hat{y} = 0.25 + 0.347 \times 70 = 27.13 \) cm, and the prediction error (residual) is \(28.5 - 27.13 = 1.37\) cm.