On OAKS, you will find the data file Earnings and Height, which contains data on earnings, height, and other characteristics of a random sample of U.S. workers from a paper published in economics. A detailed description of the data is also available on OAKS. In this problem, you will investigate the relationship between earnings and height.

Basic data statistics:

1. What is the average height in the sample? [2 points]
2. What is the average earnings for workers whose height is at most 67 inches? [2 points]
3. What is the average earnings for workers whose height is greater than 67 inches? [2 points]
4. Construct a scatterplot of annual earnings on height. Briefly explain the shape you see in

the graph (Hint: read the detailed data description to help with this part). [4 points]

Run a regression of Earnings on height.

1. What is the estimated slope? [2 points]
2. Use the estimated regression to predict earnings for a worker who is 67 inches tall, for a worker who is 79 inches tall, and for a worker who is 65 inches tall? [2 points]
3. What is the R2 and interpret its meaning. Show how you would calculate the R2  from the Analysis of Variance at the top of the SAS output [6 points]

Run a regression of Earnings on height, using data only for female workers.

1. What is the estimated slope? [2 points]
2. A randomly selected woman is 1 inch taller than the average woman in the sample. Would you predict her earnings were higher or lower than the average earnings for women in the sample? By how much? [5 points]

Run a regression of Earnings on height, using data only for male workers. [3 points for code that

correctly separates the sample]

1. What is the estimated slope? ( 2 points)
2. A randomly selected male is 1 inch taller than the average male in the sample. Would you predict his earnings were higher or lower than the average earnings for male in the sample? By how much? (5 points)
3. What is the intuition behind separating the sample into male and female samples? (5 points)
4. Do you think that height is uncorrelated with other factors that cause earnings? That is, do you
5. think that the regression error term ui has a conditional mean of zero, given height? Briefly
6. justify your answer. (5 points)