Matt Fertakos (worked alone)

1) Why would we want a model selection criterion to penalize the number of parameters in a model?

   A model selection criterion that penalizes the number of parameters in a model is helpful because it makes the results easier to interpret. Furthermore, as you increase the number of parameters you increase the chance of parameters providing repeated information, known as collinearity. Collinearity is bad for a model because it reduces the precision of the estimated coefficients, and weakens statistical power. One of the parameters that correlates to another needs to be removed.

2) Consider the regression equation for a simple linear regression: $y_i=\alpha+\beta 1 x_i+\epsilon$. In 2 - 3 paragraphs, describe the meaning of the slope parameter $\beta 1$ in the context of the relationship between the predictor variable, x, and the response variable y. Your answer must be in plain non-technical language. Your explanation will be most effective if you use a narrative approach, using a concrete example to illustrate the concept.

   The slope parameter states that for every 1-unit change in the value of the predictor variable x_i, we expect a $\beta 1$ change in the response variable y_i. In non-technical language, this means that when we increase the value of x, y will increase at a rate that is equal to $\beta 1$. An example would be measuring the height of a plant based on the amount of water you give it. After performing an experiment with three different plants in which we gave them either 10, 20, or 30mL of water once per week, we measure their heights. In this case, the x value would be the mL of water given to the plant and the y value would be its height. Using this data, and the equation, we can solve for a rate $\beta 1$, that will allow us to predict the height of a plant based on the amount of water it was given.

3) What is the *base case* water treatment?

   The base case water treatment is “low” water level.

4) What is the average plant mass, in grams, for the **low** water treatment? How did you calculate this quantity?

   The average plant mass for the low water treatment is 2.4g. This is taken from the intercept row of the coefficient table, as the intercept row represents the base case water treatment which is “low”.

5) What is the average plant mass, in grams, for the **medium** water treatment? How did you calculate this quantity?

   The average plant mass for the medium water treatment is 3.7g. This was calculated as the intercept estimate + the waterMed estimate on the coefficient table. Specifically, since this is equivalent to adding 1 ‘water treatment’ unit the equation would be

2.4+(1x1.3)+(0x13.6)=3.7. Adding 1 ‘water treatment’ unit increases the plant biomass by 1.3g on average.

6) Which of the following questions cannot be addressed with the model coefficient table? Select the correct answer or answers: (bold = cannot be addressed)
    a. Is there a positive relationship between increased water availability and plant biomass accumulation?
    b. Is water availability a significant predictor for plant biomass accumulation?
    c. What is the average biomass of plants in the high water treatment?