Week 10 Reading Questions
Mercy Melo (no other collaborators)

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

The idea behind AIC scores is to determine how well your model fits with the naturally-occurring data you're trying to describe. While goodness-of-fit is one parameter considered when calculating this score, the number of parameters is also important. Adding more parameters allows the model to artificially become a great fit to the data because you are able to customize a function that can fit more of the datapoints than a standard two-parameter function. While adding more parameters allows us to achieve a perfect fit for the data, it also reduces the ecological relevance of the model. It can be quite simple to see the ecological relevance in a 2-parameter model (such as a linear model, where one metric is directly correlated with another metric), but adding more parameters can blur the correlations suggested by the model. Penalizing large numbers of parameters will keep models simple and ensure the most ecologically-relevant models are used over models with perfect fits but no ecological relevance.

Question 2: In 2 - 3 paragraphs, describe the meaning of the slope parameter β1 in the context of the relationship between the predictor variable, x, and the response variable y.

The slope of a linear model helps us understand how our two variables, our "cause" and our "effect", relate to one another. There are two main components to consider when looking at a slope: the magnitude and the sign. The magnitude of the slope (or how small or large the value is) tells us how great of an "effect" is caused by each addition of more "cause". For example, if we gave people a science test and additionally asked them how many years of college they've attended taking science classes and how many nature documentaries they've watched, we could come up with two different models: y (score on science test) = 1.6x (# of documentaries watched) + 50 and y (score on science test) = 8x (number of years of science-based college) + 50. In our first model, we can see that the slope is only 1.6, which means that for every nature documentary they watched, their score increased by about 1.6 points. In our second model, our slope was 8 which suggests that for every additional year in science-based college their score would increase by 8 points. The magnitude of these two slopes makes sense as attending a year of college science courses would likely help someone perform better on a science test than just watching a single nature documentary.

The other component we need to consider when evaluating the slope is the sign. The sign of the slope can be either negative or positive and tells us which direction the "effect" changes when more "cause" is added. For example, if people were surveyed and asked how happy their day was and how many pieces of chocolate they've eaten that day, we might fit a model that looks like this: y (happiness) = 2x(# of chocolates) + 4. In this case we can see that the slope has a value of positive two, which means that for every additional piece of chocolate someone ate, they likely had a 2-point increase in happiness. To contrast, if people were surveyed and asked the same happiness question but then additionally asked how many spam phone calls they received, the model would likely be more like: y (happiness) = -2x (# of spam calls) + 4. In this

case our slope takes on a value of -2, which suggests that for every extra 1 spam phone call, the person likely had a 2-point <u>decrease</u> in happiness.

Question 3: What is the *base case* water treatment?

The base case is low water level treatment.

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

2.4 grams. This is the "estimate" for the intercept treatment group (which is the low water level treatment).

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

3.7 grams. This is the "estimate" for the medium water level treatment (1.3g) plus the intercept value of 2.4g.

Question 6: Which of the following questions cannot be addressed with the model coefficient table?

B. Is water availability a significant predictor for plant biomass accumulation?