Ollie Murphy
Week 12 Reading Questions
ECo 602

1. When fitting a model to a data set, it is not uncommon to have to choose between using a complex model with many parameters, which fits the data closely, and using a simpler model with fewer parameters that may have a worse fit. Each type of model has its pros and cons, and each inherently conflicts with the other. For example, consider a fictional dataset examining the population of birds in a particular neighborhood versus the number of free-roaming outdoor cats. A more complex model that minimizes unexplained variation would likely have a large number of variables-- accounting, hypothetically, for factors like cat age, sex, and whether or not they are fixed, as well as environmental factors like weather and proximity to other feeding locations. This model could provide highly accurate predictions, but would be difficult to understand to someone not familiar with the dataset. Additionally, collecting more data increases the likelihood of measurement error. On the other hand, you could use a much simpler model--hypothetically only accounting for the number of cats and their sex. This model would likely produce predictions with much more significant variation; however, the data would be simpler to collect, and it would be much easier to explain to a lay person.

2. Nitrogen was the only predictor variable with a slope coefficient that was significantly different from zero.

3. In this case, the expected biomass for a plant receiving no water, nitrogen, or phosphorus would be -1.7 g. The calculation in this case is very simple-- because there are 3 variables and 4 values in the table including the intercept, we can assume that the intercept in this case is a lack of any of the three variables. In other words, the intercept is the plant receiving 0 ml/mg of the three substances.

4. In this case, the expected biomass can be calculated by multiplying the amount of each variable by its value in the coefficient table and adding it to the intercept value. The equation would be -1.7+(10)*.043+(30)*1.92+(20)*-.027 = -1.7+.43+5.76-.54 = 3.95 g on average.

5. The key difference between a simple linear regression and a 1-way ANOVA is that while ANOVAs are used for categorical predictor variables, linear regressions are used for continuous predictor variables. The exception to this is the use of dummy variables with regressions, which essentially converts categorical data to continuous for the purpose of regression analysis. Both models have a continuous response variable.

6. The deterministic component of the model equation is a+bx

7. The deterministic component of the model equation is epsilon.