Hazel Ortiz

ECO602 Week 12 Reading assignment

Worked together with Justin and Anastasia during office hours

**1. In the context of a dataset (real or made up), describe the inherent conflict between using a complicated model that minimizes the unexplained variation and using a simple model that is easy to communicate.**

The conflict between using a complicated model and a simple model can be simply explained in terms of pros and cons of each model. With a complicated model there are more details and information so as a result these models are more accurate. However, since these complicated models have all these details and information they are not as clear and easy to interpret when looking at the results. Simple models in contrast have less information and details so they are easier to interpret but the outcome is not nearly as accurate.

Looking at fish data for example, if we want to look at the relationship between total length and age, in a simple model the only two elements that we might record in a data set are height and age. However in a more complex and comprehensive model we may also add more elements, like weight, sex, and stomach content (for nutritional/feeding habits) that can give us a deeper understanding of the relationship between total length and age of a fish population.

**2. Which of the following predictor variables had slope coefficients that were significantly different from zero at a 95% confidence level? Select the correct answer(s)**

Both water (A) and nitrogen (B)

**3. Using the information in the model coefficient table above, calculate the expected biomass for a plant given:**

The calculated biomass would be equal to the intercept, so -1.7. Since there was no water, nitrogen, or phosphorus added to the plant, there were no changes and there was no added growth to the initial intercept value. We also used the following equation to describe the growth with the given values. We used the equation below to calculate this. Since there was no added value in nitrogen, water, or phosphorous, there was no added change to the intercept so the biomass would remain at -1.7,

Plant growth = 0.043(w) + 0.192(n) - 0.027(p) - 1.7

= 0.043(0) + 0.192(0) - 0.027(0) - 1.7

= 0 - 1.7 = -1.7

**4. Using the information in the model coefficient table above, what is the expected biomass for a plant given:**

Plant growth = 0.043(w(10)) + 0.192(n(30)) - 0.027(p(20)) - 1.7

0.43 + 5.76 - 0.54 - 1.7 = 3.95mg

Using the slope coefficients and intercept values provided in the estimate table, a linear model was created to assign values to the variables.

**5. Describe the key difference between a simple linear regression and a 1-way analysis of variance.**

Simple linear regression is a statistical model that is used to predict a continuous outcome based on one or more continuous predictor variables. In contrast, ANOVA is a statistical model that is used to predict a continuous outcome based on one or more categorical predictor variables.

**We often present the equation for a simple linear regression model as:**

**yi=α+β1xi+ϵ**

**6. Identify the *deterministic* component(s) of the model equation.**

A deterministic component explains the relationship between predictor and response value. In this equation, it would be the β which represents the slope that describes the relationship between the response variable Y and the predictor variable X.

**7. Identify the *stochastic* component(s) of the model equation.**

Stochastic components describe the randomness of a model so in this equation the variable ϵ would be the stochastic component which describes the random error of the equation.