Elyse McCormick

ECO 602

12/4/2022

**Week 12 Reading Questions**

**Q1 (2 pts.):** **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. Consider the trade-off between model complexity and interpretability. Since your answer is targeted to a non-scientist audience, you should use narrative style using a concrete example**.

The inherent trade-off between complexity and simplicity in a modeling framework is a “biological accuracy” of what’s actually happening in the model versus statistical power and accuracy of the statistics. For example, if I wanted to make a model of how landscape variables influence the number of pathogens in bumble bees in Massachusetts, I could incorporate a model that has many landscape variables, as well as environmental variables, to try to isolate biological causes and eliminate errors. However, the more things I put into my model, the less accurate the mathematics behind the statistics get. If I keep my model simple in examining the landscape factors, I’m much more likely to have a result I’m confident in mathematically. This is a difficult trade-off and is something for every scientist to keep in mind.

**Q2 (1 pt.):** **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)**

1. **water**
2. **nitrogen**
3. **phosphorus**
4. **None**

**Q3 (2 pts.):** **Using the information in the model coefficient table above, calculate the expected biomass for a plant given:**

1. **0 mL water per week**
2. **0 mg nitrogen per week**
3. **0 mg phosphorus per week**

**Explain how you made the calculation.**

(0 mL water/week)\* (0.043 (the estimated mean mass of the water treatment) = 0 mg

(0 mL nitrogen/week)\* (0.192 (the estimated mean mass of the nitrogen treatment) = 0 mg

(0 mL phosphorus/week)\* (-0.027 (the estimated mean mass of the phosphorous treatment) = 0 mg

**Q4 (2 pts.):** **Using the information in the model coefficient table above, what is the expected biomass for a plant given:**

* **10 mL water per week**
* **30 mg nitrogen per week**
* **20 mg phosphorus per week**

**Explain how you made the calculation**.

(10 mL water/week)\* (0.043 (the estimated mean mass of the water treatment) = 0.43 mg

(30 mg nitrogen/week)\* (0.192 (the estimated mean mass of the nitrogen treatment) = 5.76 mg

(20 mg phosphorus/week)\* (-0.027 (the estimated mean mass of the phosphorous treatment) = -0.54 mg

**Q5 (1 pt.):** **Describe the key difference between a simple linear regression and a 1-way analysis of variance**.

Mathematically, ANOVA and linear regression are identical, but the main difference is that ANOVA can has the independent variable as categorical, while regression can use both categorical and continuous.

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

**yi=****α+β1xi+ϵ**

**Q6 (1 pt.): Identify the *deterministic* component(s) of the model equation**.

The deterministic portions of the model are α+β1xi, this portion essentially tells you the slope, and the position of the data point x.

**Q7 (1 pt.): Identify the *stochastic* component(s) of the model equation**.

The stochastic portion of the model is the error term, ϵ.