Week 12 Reading Questions

Mercy Melo (no collaborators)

Question 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.

When fitting a model to a dataset, we have to make choices regarding how specific that model is going to be. Adding more predictor variables allows us to make the model fit very closely to the data points, and having fewer predictor variables keeps the model more loosely-fit. Although it seems like adding more predictor variables would be the best option because the model will fit the data more precisely, the addition of these extra variables can make the model over-complicated. Let's say we're trying to model how parental care factors influence duckling development. It would be easy to interpret a model where we relate the number of times the mother duck feeds the duckling to the duckling's weight (a model with only 1 predictor variable). If we try to involve more predictor variables (feeding rate, ambient air temperature, incubation time, time swimming, etc.) the model can become very complicated. Limiting the number of predictor variables, in this case, would allow you to more easily interpret the model and communicate the important factors that impact duckling development.

Question 2: Which of the following predictor variables had slope coefficients that were significantly different from zero at a 95% confidence level?

Water and Nitrogen

Question 3: Using the information in the model coefficient table above, calculate the expected biomass for a plant given: 0 mL water, 0 mg nitrogen, and 0 mg phosphorus per week. Explain how you made the calculation.

The plant would have 0g of biomass. To achieve this answer, you would plug a "0" in for every value of x in the equation: biomass = $0.043x_1 + 0.192x_2 - 0.027x_3$. Since you are multiplying each model coefficient by 0, the plant biomass is therefore 0 as well.

Question 4: Using the information in the model coefficient table above, what is the expected biomass for a plant given: 10 mL water, 30 mg nitrogen, and 20 mg phosphorus per week. Explain how you made the calculation.

The plant would have 96.05g of biomass. Similar to question 3, you would plug in the values of each input into the equation with the model coefficients which results in a weekly biomass growth rate of 5.65g. To get the accumulation of biomass over 17 weeks, you would simply have to multiply this value by 17 to get a value of 96.05g.

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

Linear models and 1-way ANOVAs go hand-in-hand, but have one key difference: the ability to test for significance in the model's parameters. While a linear regression can simply correlate two variables, an ANOVA goes one step further to test if that variable has a significant effect on the response variable (which is shown in the ANOVA table).

Question 6: Identify the *deterministic* component(s) of the model equation ($yi=\alpha+\beta_1 x_i+\epsilon$).

The deterministic component of the model is the "$\alpha+\beta_1 x_i$" portion.

Question 7: Identify the *stochastic* component(s) of the model equation ($yi=\alpha+\beta 1xi+\epsilon$).

The stochastic component of the model is the "ϵ" portion.