Johanna Ravenhurst
Submitted: December 6, 2022
I didn't work with other students.

ECO 602 – Week 12 Reading Questions

Q1: Model Comparison - 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.

> Suppose I have a dataset with many different types of measurements from a set of plots in Western Massachusetts. I am interested in tick-borne diseases, so these data include counts of ticks found at the plots, density of vegetation, type of vegetation, diversity of potential animal hosts present, humidity, and temperature. I could include all of these variables in my model to show how many different factors contribute to the number of ticks found at the plots. However, it might be difficult to use the results of that model to provide information about which elements of an environment have the biggest impact on the local tick population. Therefore, it would be better to reduce the model to include fewer elements and provide clear recommendations for how to alter a habitat to reduce potential tick bite exposures.

Q2-Q4: Interpreting a Coefficient Table

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)

A. water
B. nitrogen
C. phosphorus
D. None

Both water and nitrogen had significant p-values.

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

- 0 mL water per week
- 0 mg nitrogen per week
- 0 mg phosphorus per week

Explain how you made the calculation.

> Plant biomass accumulation (g) = intercept + $\beta_{water}(x_{water}) + \beta_{nitrogen}(x_{nitrogen}) + \beta_{phosphorus}(x_{phosphorus})$
>
> Plant biomass accumulation (g) = -1.7 + (0.043)(0) + (0.192)(0) + (-0.027)(0)
>
> Plant biomass accumulation = -1.7 g.

The expected biomass accumulation is that the plant would lose 1.7 g per week. I made this calculation by writing out the model equation and then filling in the coefficient values from the table along with values of 0 for each of the predictor variables.

However, we know that the intercept is the expected value of plant biomass accumulation when all of the predictors are equal to 0, so I could have just used the intercept coefficient as the answer instead of writing out the equation.

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.

Plant biomass accumulation (g) = intercept + $\beta_{water}(x_{water}) + \beta_{nitrogen}(x_{nitrogen}) + \beta_{phosphorus}(x_{phosphorus})$

Plant biomass accumulation (g) = -1.7 + (0.043)(10) + (0.192)(30) + (-0.027)(20)

Plant biomass accumulation = 3.95 g

The expected biomass accumulation is that the plant would gain 3.95 g per week. I made this calculation by writing out the model equation and then filling in the coefficient values from the table along with values for each of the predictor variables. The coefficient values are how much the plant biomass changes for every one-unit change in the continuous predictor variables, so they just need to be multiplied by the number of units of the predictors.

Q5: Regression and ANOVA - Describe the key difference between a simple linear regression and a 1-way analysis of variance.

- Consider the data types/scales of the predictor and response variables.

The key difference between these two methods is the type of predictor variables.

Simple linear regression is used when the response variable is continuous and there is a single continuous predictor variable.

1-way ANOVA is used when the response variable is continuous and the predictor variable is categorical and contains 3 or more levels. ANOVA tells us if any of the group means are different from the others, but we need a follow up test to give us information about which levels are different.

Q6: Regression Equation as a Dual Model - Identify the *deterministic* component(s) of the model equation.

The deterministic components of the model include the intercept term and the slope/coefficient for the predictor. $\alpha + \beta_1 x_i$

Q7: Regression Equation as a Dual Model - Identify the *stochastic* component(s) of the model equation.

The stochastic component is the error term at the end. This is the residual, or the difference between what we observe in our sample data and the predicted/expected value from the deterministic model. - $+ \epsilon$