Johanna Ravenhurst
Submitted: November 25, 2022
I didn't work with other students.

ECO 602 – Week 10 Reading Questions

Q1: Model Selection - (1 pt.): **Why would we want a model selection criterion to penalize the number of parameters in a model?**
- "The first part of the AIC definition is a measure of goodness of fit. The second part is a penalty for the number of parameters in the model." - Zuur (2007)

We want to select a model that includes the fewest parameters possible because this makes it easier to interpret the results. It is also unlikely that all the possible predictors actually impact the response variable, so it's better to narrow down our set of predictors to the most important variables. We are looking to find a more precise estimate of the association between the predictors and response variable, rather than a fitted line that is derived from many different predictors but has a wide confidence interval.

Q2: Interpreting a Slope (3 pts.): In 2 - 3 short paragraphs, describe the meaning of the slope parameter *β*1 in the context of the relationship between the predictor variable, x, and the response variable y.

Suppose I am interested in studying the relationship between the density of Japanese Barberry plants and how that impacts the local tick population. I have data that was collected from 20 sample plots that includes the density of the Japanese Barberry plants and the number of ticks found at that same plot over the course of a year. I need to use a simple linear regression analysis to see if my data provides evidence of a relationship between the density of these plants and the number of ticks found in the same area.

Here is an example of the regression equation that I will use for my analysis: $y_i = \alpha + \beta_1 x_i + \epsilon$. The y value will be the number of ticks that I collected at that plot, and I am hypothesizing that this value is impacted by the density of Japanese Barberry plants, which will be included in the analysis as x. The symbol β_1 represents how much a one-unit change in the density of Japanese Barberry plants impacts the total tick population. For example, if the density of plants increases by one unit and this is related to an increase, on average, in the local tick population, then I would expect the β_1 to be equal to a positive value greater than 1, so perhaps it would be equal to something like 2.0. So in other words, an increase of one unit in density of Japanese Barberry would increase the average tick population by a factor of 2.

Q3-5 Interpreting a Coefficient Table

Q3 (1 pt.): Based on the model table, what is the *base case* water treatment?

The base case is when all the dummy variables have a value of 0, which is equivalent to the intercept in the linear regression model. The base case in this example is the low water level treatment.

Q4 (2 pts.): What is the average plant mass, in grams, for the **low** water treatment?

- How did you calculate this quantity?

The average plant mass for the low water treatment is just the estimate for the intercept, 2.4 g, because the low water treatment is used as the reference (or base case).

The intercept value is the plant mass when all the predictor values are set to zero. We are using dummy variables for the levels of water treatment and there is always one less dummy variable than the number of levels of the categorical variable. We can see from the table that the low water treatment was used as the reference level, so the value for low water treatment is found when the dummy variables for both medium and high water treatment are equal to zero (the intercept).

Q5 (2 pts.): What is the average plant mass, in grams, for the **medium** water treatment?

- How did you calculate this quantity?

The average plant mass for the medium water treatment = Intercept estimate + waterMed coefficient = 2.4 + 1.3 = 3.7 g

Q6: Coefficient Interpretation (1 pt.): Which of the following questions cannot be addressed with the model coefficient table? Select the correct answer or answers:

A. Is there a positive relationship between increased water availability and plant biomass accumulation?
B. Is water availability a significant predictor for plant biomass accumulation?
C. What is the average biomass of plants in the high water treatment?

We would need to look at the ANOVA table to see if water availability in general is a significant predictor of plant biomass accumulation.