Julia Vineyard

Week 10 Reading Questions

1. We want to reduce the number of parameters because having more than necessary can make our model more complex than it needs to be. Having a more complex model is not necessarily a benefit as you may overfit the model to your data. This can result in the p-values and coefficients to be misleading. Additionally, a model that is overfit to the data may not be reflective of the population.
2. Any change in the slope, β1, results in a change in the sensitivity off the response variable, y, on the values of the predictor variable, x. This means how fast or how slow the value of y changes when a unit of x is changed depends on the slope. If you have a slope equal to 0 it does not matter what the value of x is, your value of y will always be equal to the sum of the constant and error.

If we wanted to analyze the assumption that household income impacts the number of books a family has we would set income as our predictor variable, x, and number of books as our response variable, y. To obtain our data we surveyed 50 houses to ask for the annual income of the household and the number of books. While this is not a perfect scenario since other predictor variables could influence the number of books a household has, such as size of the family and number of members old enough to read, we will just look at income. After running a statistical analysis on the data we have, we found that for every 1 more unit of income a household has our data shows that they will have 0.25 more units of books. To put it in real terms, if a household has $500 in income they will have 125 books.

If we were to complete the survey in a different part of town and got a different set of 50 responses, our data would most likely be different. For this analysis we found that for every 1 more unit of income a household will have 0.02 more units of books. Keeping with the same example, if a household as $500 in income they will have only 10 books. While we kept our income, predictor variable/x, the same our slope, β1, changed from 0.25 to 0.02. This drastically changed the number of books that a household would have, which is y/the response variable. This example does not take into consideration the constant, α, nor the error, ϵ.

1. The base case treatment is low.
2. The mean plant mass for the low water treatment is 2.4. This is taken directly from the table.
3. The mean plant mass for the medium water treatment is 3.7. Add the (Intercept) mean, the mean of the low water treatment, to the estimate of the medium treatment.
4. The coefficient table can not address B. Is water availability a significant predictor for plant biomass accumulation?