Week # 10 Reading Questions Responses

“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)

1. Why would we want a model selection criterion to penalize the number of parameters in a model?

A model selection criterion that penalizes the number of parameters in a model would allows to create a complex model that can be used to describe data but is understandable due to the limiting of the number of parameters. This is achieved by finding an optimal model which is a model with the smallest AIC (Akaike Information Criteria), which therefore has the least amount parameters which corresponds to a minimum .

1. Consider the regression equation for a simple linear regression:

In 2 - 3 paragraphs, describe the meaning of the slope parameter  in the context of the relationship between the predictor variable, x, and the response variable y.

Your answer must be in plain non-technical language. Your explanation will be most effective if you use a narrative approach, using a concrete example to illustrate the concept.

Regression is used when modeling to express a relationship between independent, predictor variables and dependent, response variables, . These values are our observations. represents the expected behavior while represents the uncertainty in this expected behavior.

In the linear regression equation above α, the value of the dependent variable if there are no independent variables. and represent quantities we want to estimate. For example, if we are using the Palmer Penguins data, would help us determine if we could predict body mass from flipper length. would be the expected change in flipper length due to a 1-unit change body mass.

1. What is the *base case* water treatment?

The base case which is analogous to the intercept term of a linear model, is low water level treatment.

1. What is the average plant mass, in grams, for the **low** water treatment?

How did you calculate this quantity?

Average plant mass, in grams, for the **low** water treatment is 2.4. Since this is the base case, the value of the response variable when there are nor predictor variables, I read this value from the Coefficient table.

1. What is the average plant mass, in grams, for the **medium** water treatment?

How did you calculate this quantity?

In a Coefficient table, expected behavior is relative to the base case. Therefore, the average plant mass, in grams, for the **medium** water treatment is calculated as:

grams

1. Which of the following questions cannot be addressed with the model coefficient table? Select the correct answer or answers:
   1. Is there a positive relationship between increased water availability and plant biomass accumulation?
   2. Is water availability a significant predictor for plant biomass accumulation?
   3. What is the average biomass of plants in the high-water treatment?

The model coefficient table tells the strength and significance of slope and individual slope coefficients but cannot tell us the significance of a categorical predictor. The model coefficient table therefore cannot address question b: Is water availability a significant predictor for plant biomass accumulation? Determining significance would require statistical tests.