Week # 12 Reading Questions Responses

McGarigal writes:

We expect a model with more parameters to fit better in the sense that the negative log-likelihood should be smaller if we add more terms to the model. But we also expect that adding more parameters to a model leads to increasing difficulty of interpretation.

**Q1 (2 pts.):** 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.

Consider the tradeoff between model complexity and interpretability.

Since your answer is targeted to a non-scientist audience, you should use narrative style using a concrete example.

The goal is to create a model complex enough that it has minimum unexplained variation while being simple enough through limiting the number of parameters it uses that it is interpretable. This inherent conflict between model complexity and interpretability can be illustrated with the Palmer Penguins data. If we are interested in predicting penguin body mass from flipper length then, bill length and bill depth could be included as parameters that could relate to flipper length. Likewise, penguin sex and island could be added as parameters by possibly relating to flipper length. Using all these variables, both categorical and continuous as parameters in our model would greatly limit the unexplained variation in predicting penguin body mass from flipper length because all variables that could affect flipper length would be accounted for. However, this model would be far too complex to interpret because all these parameters could have interactions not only with flipper length but among themselves in addition to their individual contributions. This would not be an interpretable model as a result.

**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)

1. water
2. nitrogen
3. phosphorus
4. None

Options A and B. Water and Nitrogen since Pr(>|t|) 0.05 for these predictors.

**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.

The expected biomass accumulation for a plant given the above conditions is -1.7 grams . This is also known as the base case, the absence of any predictor variables. It is the result of multiplying each  predictor by zero.

**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.

The expected biomass accumulation of 3.95 grams is calculated by summing the base case/intercept with the products of each coefficient value in the coefficient table with the corresponding given amount of water, nitrogen and phosphorus per week:

Converting to grams before calculation would suggest that adding water and nitrogen would result in expected biomass accumulation below the base case:

**Q5 (1 pt.):** 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.

Simple linear regression (SLR) requires one continuous predictor and one continuous response variable. A 1-way analysis of variance (ANOVA) requires a categorical predictor variable of three or more levels and a continuous response variable**. The key difference is therefore that SLR requires a continuous predictor and 1 -way ANOVA requires a categorical predictor**.

We often present the equation for a simple linear regression model as:

**Q6 (1 pt.):** Identify the *deterministic* component(s) of the model equation.

The deterministic components are .

**Q7 (1 pt.):** Identify the *stochastic* component(s) of the model equation.

The stochastic component is .