

Analysis of Environmental Data – Reading questions 12

Olivia Dinkelacker

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 trade off between model complexity and interpretability.

The advantage of using a complicated model is that it can minimize the amount of unexplained variation or error, but on the other hand this complex model is less easy to interpret/understand. Simple models are easier to communicate and interpret, but have more unexplained error or variation.

For example, when modeling penguin distribution, a simple model might only include penguin species to model bill length. A more complex model might include more predictor variables like island, sex, and diet. This second model could likely explain more of the variation in penguin distribution, but it isn't as simple to interpret how the predictor variables are effecting the bill length.

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)

water
nitrogen

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

The biomass would be -1.7, since we don't add any additional values.

$$-1.7 + (0.043 * 0) + (0.192 * 0) - (0.027 * 0) = 0$$

Q4 (2 pts.): Using the information in the model coefficient table above, what is the expected biomass for a plant given:

First calculate gram in mg and then.

$$-1.7 + (0.043 * 10) + (0.192 * 30) + (-0.027 * 20) = 3.95$$

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.

A simple linear regression is used to predict a continuous response on the basis of a continuous predictor variable (flipper length in terms of body mass).

Whereas, ANOVA is used to test for one categorical predictor that has 3 or more levels (species) and a continuous response.

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

Alpha and beta

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

ϵ