

## 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.

Transformations help to linearize the relationship and stabilize the variance. But transformed model coefficients can be more difficult to interpret, since have to read them in terms of proportional increases/decreases instead of constant amounts now.

Can't say: "Every 1% increase in CO2 in a tank was associated with 2 dead fish."

Say: "1% increase in CO2 was associated with a 6% proportional increase in mortality rate of fish."

**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.

$$-1.7 + (43 * 0) + (192 * 0) - (27 * 0) = 0$$

**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.

$$0.043\text{g} = 43 \text{ mg}$$

$$0.192\text{g} = 192\text{mg}$$

$$-0.027\text{g} = -27\text{mg}$$

$$10 \text{ mL water per week: } -1.7 + (43 * 10) = 428,3\text{mg}$$

$$30 \text{ mg nitrogen per week: } -1.7 + (192 * 30) = 5758,3\text{mg}$$

$$20 \text{ mg phosphorus per week: } -1.7 - (27 * 20) = -541.7\text{mg}$$

$$428.3\text{mg} + 5758.3\text{mg} - 541.7\text{mg} = 5644.9\text{mg} = 5.64\text{g}$$

**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.

$\beta$

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

$\epsilon$