

## Reading Questions Week 10

Sonja Glasser

1. Model selection criterion should penalize an increasing number of parameters to avoid overfitting the model too closely with sample dataset. This is helpful because when there is new data, the model is more flexible and easier to apply to more sample data sets as opposed to being too closely trained to the noise of the sample dataset from which the model was originally fitted. If there are too many parameters, then we could essentially explain all the variation of the data and make it very challenging to interpret the model. There is an optimum between model fit and model simplicity which is to find the least amount of parameters that explains a significant amount of the variation.
2. Drought conditions are raising river water temperatures potentially having physiological effects on river fish populations. To describe the meaning of the slope parameter in a regression model I will use an example of the effect of water temperature on body length of delta smelt fish. Suppose that we want to test if higher river water temperature is correlated with smaller sized delta smelt fish. We would need to go out and sample various locations along a river that represent potentially discrete habitats for various populations of delta smelt fish. We notice that some sites are deeper than others and therefore may have cooler river temperatures. We collect and measure at least 30 adult fish from each site.

If we were to fit a regression model to the question of the effect of water temperature on delta fish length we would set temperature as the predictor variable,  $x$ , and adult fish length as the response variable,  $y$ . Say that in fact we saw smaller fish in the delta smelt populations inhabiting warmer water. We would want to know numerically how  $1^{\circ}\text{C}$  increase in temperature is related how much smaller a delta fish may be. The  $\beta_1$  is the slope parameter which would describe exactly how much smaller the fish will be with an increase in 1 unit of temperature. Imagine that as  $x$  increases per unit the fish decreases in  $-0.05\text{ cm}$ , so  $\beta_1 = -0.05$ .
3. The *base case* for the water treatment is low.
4. The average plant mass in grams for the low water treatment is 2.4 grams, I did not need to calculate it because it is the intercept (base case) and the estimate given by the table is the actual weight given that treatment.
5. The average plant mass in grams for the medium water treatment is 3.7 grams. The other estimate values are relative to that of the base case, so I simply added  $2.4 + 1.7 = 3.7$ .
6. Which of the following questions cannot be addressed with the model coefficient table?  
Select the correct answer or answers:

**b. Is water availability a significant predictor for plant biomass accumulation?**