Week # 9 Reading Questions Responses

“In the best case, your data will match a classical technique like linear regression exactly, and the answers provided by classical statistical models will agree with the results from your likelihood model.” - Bolker (2008)

Bolker describes custom-made analyses based on Maximum Likelihood, which often have a biological, ecological, or mechanistic justification.

He contrasts these with the familiar Least Squares, *canned* methods that we typically learn in our first statistics course.

1. Briefly (1 - 2 short paragraphs) describe at least two tradeoffs between the customized ML methods and the canned methods.

MLE involves finding parameters that make the observed data most likely to have occurred. This is achieved through using the deterministic and stochastic parts of the data find the probability of a particular outcome given a set of parameters. The set of parameters that yield the highest probability of this particular outcome are chosen as our best estimates of parameters that describe the data. Mechanistic, ecological and biological justifications often define a MLE which makes it a custom-made analysis appropriate for a particular set of data.

The full descriptive power of the MLE may be lost with canned methods but there are several conveniences to utilizing them instead. Canned methods have faster computational speeds and do not require one to specify starting parameters. Canned methods simplify parameter estimation. For example, a logistic model and threshold model may have biological justification, but the logistic model makes an easier fit due to smoother changes as parameters change and some generalized linear models like logistic regressions allow unconstrained parameter optimization. Using canned/standard methods limits confusion as it follows convention. Lastly, canned methods allow one to easily compare various methods and hypotheses such as switching between models incorporating effects of Nitrogen or Phosphorus only, Growth~N, Growth~P, their additive effects Growth~N+P, or interactions between them, Growth~N\*P.

1. Briefly (1 - 2 sentences) describe each of the four key assumptions of the general linear modeling approach.

There is perfect accuracy in measurement of predictor variables, all observations are independent with constant variance and residuals are normally distributed.

1. Explain how the normality assumption can be met in a general linear model, even if the response variable is not normally distributed. (1 - 2 paragraphs)

The normality assumption states that through repeated sampling, the data will be normally distributed around the predictor variables. This applies to the expected values, the residuals and therefore does not necessarily mean the data, i.e., the response variable is normally distributed. The normality assumption can be met in a general linear model, even if the response variable is not normally distributed, if residuals are normally distributed, through OLS which minimizes the squared residuals, the difference between observed and expected, i.e., errors. OLS can also be used to find optimal parameters for simple linear regression models, the mathematical relationship between independent and dependent variables. MLE is also equivalent for linear regression if the residuals are normally distributed.