Bonnie Turek

ECo 602 – Week 4 Reading Questions

9/20/21

**Q1.** Predictor variables for each model and their data type/scale:

A predictor variable is a variable that is being used to predict some other variable or outcome. Predictor variable and independent variable are both used to observe how they affect some other variable or outcome. The main difference between them is that independent variables can be used to determine if one variable is the cause of changes in another, whereas predictor variables cannot. Specifically, the researcher can better manipulate the independent variable than just a predictor variable.

In model #1, the predictor variable is late successional forest cover. It is defined as the percentage of the landscape covered by late-successional forest. This is of the data type numeric and is of a percentage scale from 0-100%.

In model #2, the predictor variable is the total basal area. Basal area is a measure of cross-sectional areas of trees. This is also of numeric data type and on a continuous integer or decimal scale from 0-200. This unit is most likely in $ft^2$/acre.

**Q2.** Response variables for each model and their data type/scale:

Response variable is the variable whose variation depends on other variables. I.e. it is subject to change based on other variables and the response variable is generally the major focus of a question in a study or experiment. It can often be compared to the ‘dependent variable.’ An explanatory variable is one that explains changes in that “outcome” or “response”. This can be compared to as an ‘independent variable.’

In model #1, the response variable is Brown creeper abundance.  This is represented on a numeric data type scale from 0.0-1.0 which leads me to believe that abundance is a fractional representation of a percent. In this model, the percent coverage of late-successional forest would be predicting the response of fractional percent of brown creeper abundance there.

In model #2, the response variable is Brown creeper presence/absence. This is of binary data type (so either 0 or 1 for absence and presence respectively). This holds less biological meaning/power since it is not an actual numeric count such as abundance.

**Q3.** How did the data type or scale of the data influence or constrain the choice of model in each of the two studies?

1. A model of Brown creeper abundance explained by late-successional forest percent.

In model #1, the response variable is Brown creeper abundance. (Percent of late-successional forest is the explanatory/predictor variable.) Brown creeper abundance in this case is represented as numeric fractional values. Therefore, there is a corresponding abundance numeric value for every given forest percent coverage value. That is why this model is best displayed as data in a scatter plot and/or a coplot. Since both parameters are numeric, there is more significance and biological meaning behind the values, as compared to a more limited presence/absence study. The model therefore is not very restrictive since we have actual numeric data with biological significance. However, a linear model is chosen to fit the scatter plot data points, and whether or not the plotted line is of best fit for the data is based on actual statistics that would be need to presented. A linear function using the y=mx+b equation is used. The line itself represents the deterministic portion of the model and the variance of the data points from that line represent the stochasticity of the model. Generally, the scatter plot shows a linear positive trend that as percent of late-successional forest increases, the abundance of brown creepers also increases.

2. A model of Brown creeper presence/absence explained total basal area (a measure of tree cover).

In model #2, the response variable is Brown creeper presence/absence. (Total basal area is the explanatory variable). Brown creeper presence/absence is of binary data type. This limits the choice of model for a study. That's because we only know whether the brown creeper was present or not present based on a single variable, total basal area in continuous numeric integer/decimal data type. We don't know a count of how many brown creepers were specifically present per each total basal area observation. Presence/Absence data is typically visualized best in tabular form. Unfortunately, there is not an easy way to aggregate or summarize this data, since we do not have specific count values. Given the binary data of presence/absence the best model to select is a simple logistic model. This is typical for presence/absence data in general and is of sigmoid shape. The logistic function uses this equation:

$$y = \frac{e^{a+b\cdot x}}{1+e^{a+b\cdot x}}$$

The data and logistic function shows a general trend that the proportion of presences of brown creepers increases at larger total basal area values.

**Q4.** What were the pros and cons of each of the two models (Ricker and quadratic models for density-dependent predator-prey interactions)?

The Ricker function is widely used as a phenomenological model for environmental variables that start at zero, increase to a peak, and decrease gradually back to zero, such as ecological population studies that are density dependent. The example given was striped bass recruitment based on stock size. A pro to the Ricker model and other logistic mechanistic curves (Bevholt and Holling 3), is that they are widely used and supported, and they have a biological/ecological theory backing them. This is based on previous knowledge of predator-prey interactions from many other previous studies. The cons of these models are that they don't necessarily fit your data the best since they are intended to fit a broad range of similar density-dependent studies. This relates to the researcher's concern of "goodness-of-fit."

The quadratic model, in this case actually fit the striped bass recruit-stock data better than the mechanistic Ricker and logistic models. The quadratic function is an example of a phenomenological model which often fits your specific dataset values better than more broad-reaching theory-based mechanistic models like the Ricker function. Therefore, using a phenomenological model may increase the goodness of fit to your data but be less theory-based. Overall, there are tradeoffs as to whether you would like your model to be significantly ecologically backed or to fit your data to a model in the best way possible, with less theoretical backing.