Bonnie Turek

Eco 602 – Week 9 RQs

10/26/21

Q1. 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.

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

You might want to use classical "canned" methods due to their recognizable convention. The theory behind the familiar Least Squares method and other classical methods are well understood and there are a lot of published papers utilizing these methods. It would therefore be simple enough for readers to understand your statistical analysis. If you used a customized ML method, you might need to further explain in depth about the method, even though it may make more sense to fit your particular data. This is because it is less conventional, sometimes consider a "fancy" model, and people may not recognize the custom methods you used.

Classical models mean your data must fit a particular distribution, which makes them simpler, faster and easy to comprehend, however, your data may not always be a good match for one of these. R also makes it easy to compare a lot of classical models. On the contrary, you might want to use the customized maximum likelihood methods when you want to fit your specific collected data better. You may also want to use a customized maximum likelihood method when your data does not exactly fit a classical canned method. These customized max likelihood estimates may still closely resemble classical methods, just differently parameterized and these methods may require a bit more specification to your data.

Q2. Briefly (1 - 2 sentences) describe each of the four key assumptions of the general linear modeling approach.

• Independent observations

Key points related to the assumption of independent observations include that sampling is randomized and knowing something about X1 for example does not necessarily give us information about X2. Non-independence is the violation on this assumption and can often happen with pseudo replication or relate to close proximity in space or time of sampling.

• Constant variance, a.k.a. homogeneity

This assumption refers to how constant variance is required along the entire range of predictor variables. The key points of this assumption are that the stochastic model should follow a normal distribution, the spread parameter is constant, and the variability does not depend on the value of x.

• Fixed x:

This assumption of general linear modeling refers to requiring perfect accuracy in our measurements with no measurement error in our predictor variables. This assumption is frequently violated as this is very difficult to meet.

• Normality

normality refers to how under repeated sampling, the data would be normally distributed at each X. In group 1 models, general linear models, the residuals should be normally distributed, however the data itself is not normally-distributed.

Q3. 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)

In group 1 models (general linear models) the residuals should be normally distributed, however the data itself is not necessarily normally-distributed. In other words, the values for each variable in the dataset are not normally-distributed by themselves, but the residuals (or the difference between predicted and observed values) are normally-distributed in Group 1 models. This is how the Normality assumption would be met. For example, we looked at the Salamander dispersal data based on distance class. The deterministic model in this case, appears to best follow a Ricker model curve, however, the stochastic component of the model (aka the residuals) follow a more normal distribution.

You can conduct a Shapiro test on the residuals to determine the level of Normality and whether or not the assumption of general linear model is met. A low p-value in the Shapiro test of normality provides strong evidence that the data are not normally-distributed. We also looked at penguin bill length data. The bill length values observed appear non-normal and have a low p-value for the Shapiro test of normality. However, when we plot the residuals of predicted bill length as a function of body mass and species, we appear a normal distribution.

*I worked independently on this assignment.