Johanna Ravenhurst
Submitted: November 20, 2022
I didn't work with other students

ECO 602 – Week 9 Reading Questions

this week's readings are:

- Slide Deck 7
- McGarigal Ch. 11a: Landscape of Statistical Methods: Part 1
    - Read sections 1 and 2 (we'll come back to the rest later)
- Bolker Ch. 9: Standard Statistics Revisited
    - Read sections 9.1 - 9.2
- Zuur Ch. 5: Linear Regression
    - Read sections 5.1 - 5.2

**Q1 – Modeling Approach- (1 pt.): Briefly (1 - 2 short paragraphs) describe at least two tradeoffs between the customized ML methods and the canned methods.**

One of the advantages of using a customized ML modeling approach is that you can adjust your model to the data that you collected. If the trends in your data are not easily approximated using one of the classical approaches, a custom model might provide insights that you wouldn't otherwise find. In contrast, if you can fit one of the classic models to your data, the modeling process is much more straightforward and doesn't involve as much work to determine the appropriate parameters.

An advantage of using a canned/classical modeling approach is the ease of explaining your methods and comparing your results with similar studies. When you use a classical modeling approach you can simply reference the name of the standard method, such as linear regression, and most people reading your paper will know what that means. This is an advantage because more people will understand the modeling method that you used and easily compare your results with their own study or the results from other studies. Using a customized modeling approach can sometimes make people skeptical of the validity of your approach.

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

1. **Independent observations –** We assume that the sample was random and the probability of each observation does not depend on another observation in the sample.
2. **Constant variance –** We assume that the variability in the observed values does not change depending on the value of x. In other words, we don't want to see increasing or decreasing variability in the observed values as the value of x increases.
3. **Fixed x (no measurement error in predictors) –** This assumption is generally ignored, but we assume that there is no more than minimal measurement error.
4. **Normality in residuals –** We assume that the stochastic part of the model includes values that are normally distributed because they were randomly sampled from a normally distributed population.

**Q3 – Normality -** (1 pt.): **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 is for the distribution of the residuals in the model, not the predictor or response variable. So it doesn't matter if the response variable is normally-distributed or not, we only need to pay attention to the distribution of the residuals. The residuals are the difference between the observed values and the mean or predicted value, and make up the stochastic portion of the model. As long as the residuals in the model are normally-distributed, the normality assumption is met for a general linear model. We can use the Shapiro test for normality to determine if we have enough evidence that the residuals are normally distributed.