

Week 9 Reading Questions

- 1) Three tradeoffs for using customized ML methods versus “canned” methods are (1) computational speed and stability, (2) stable definitions, and (3) convention. The first tradeoff is that customized methods use special case optimization algorithms that are faster and create less problems to run into. This tradeoff also creates little responsibility with choosing starting parameters. The second tradeoff is customized methods have often been chosen to simplify the parameter estimation. Then the third tradeoff being that you can use the method directly in your “methods” (explanation) section of your report, where if you used “canned” methods you would have to explain it in depth. The main difference between customized ML methods and “canned” methods is that customized methods are there to speed the process up and make your life easier in explain and producing results.
- 2) The four assumptions of general linear models are **normality** (assumptions that if we repeat the sampling many times under the same conditions, the observations will be normally distributed), **homogeneity** (assumption that the variance for all X values is the same), **independence** (the Y values for one observation should not influence the Y values for other observations), and **Fixed X** (X is not random and that we know the exact values of X, there is also no noise).
- 3) Normality is the assumption that the residuals follow a normal distribution because there are multiple Y observations for each X value there are also multiple residuals for every X value. In this assumption we assume that X is fixed, this implies that the errors are normally distributed at each X value. Instead of having a histogram of the response variable we will need a histogram of the residuals to view if it checks off normality. I think this quote from the McGarigal reading says it best, “normality applies to the variations around the expected value – the residuals – not to the whole data set” (McGargal Ch 11a, 4).