Ian Eggleston

1. One of the benefits of using the canned methods is their versatility in R. There are easy ways to compare different types of models in R that are already built in; versus, using a custom ML model that would need to be debugged before integrated well. Another tradeoff of using custom ML models is that the algorithms are not optimized for R and might take longer to run. R has built in common models that can run faster because they have shortcuts already made to skip redundant steps like naming variables.
   1. The first assumption we make is that the observations are completely independent of one another and do not influence the outcome of another.
   2. Constant variance is the next assumption that shows there is no bias toward one side of the data that could have more or less variance. Constant variance means the data and experiment were well designed to have a consistent degree of error.
   3. Fixed x means that we have chosen a variable to control as our predictor variable. There is no error for these variables because they are controlled and chosen by us, but in reality it exists just not enough to affect the results.
   4. The normality assumption we make is that the data residuals are normally distributed, not the data points themselves. The residuals being normally distributed shows that there is constant variance and continues to backup the previous assumptions.
2. The normality assumption is met in general linear models by testing the residuals, not the response variables themselves. If the residuals are normally distributed then this means variance is constant and there is no bias in the experiment. The response variable may not seem normally distributed if there is an even distribution, however this is why we can then test the residuals to see if they follow a normal distribution.