Julia Vineyard

Worked with: Doug Bishop

Week 9 Reading Questions

1. Using ML methods allows you to run your analysis faster and more efficiently than canned methods allow. This efficiency comes from reducing the likelihood that you will come across numerical issues and eliminates the task of selecting starting parameters. However, using ML methods creates more work when explaining the methodology. By not using the conventional, canned methods you have to describe to the reader how you completed the statistic analysis. Some readers may not be trusting of new ways to analyze data, or may just not be able to understand how you did it. This can also create a limitation if others attempt to reproduce your analysis but are unable to understand your methodology.
2. Briefly describe each of the 4 key assumptions of the general linear modeling approach.
   1. Normality: repeated sampling under the same environmental conditions will result in data that is normally distributed for each x. This refers to the model residuals, which are the vertically spread (distance between observed and predicted) values for each value of x.
   2. Homogeneity aka Constant Variance: the spread parameter (variance) is constant. In other words the variability does not depend on the value of x. For every value of x the variance of all possible values of the population is the same.
   3. Independence: sampling is randomized and knowing something about observation 1 gives us no information about observation 2. The y values of observation 1 doe not influence the y values of observation 2.
   4. Fixed X: there is no measurement error in our predictor variables and we know the exact values.
3. The response variable distribution does not matter when determining if a general linear model meets the normality assumption, only the variation of the residuals apply. The residuals are always the vertical distance between the observed and predicted values. The distribution of residuals can meet the assumption of normality even if the response variable, in this case bill length, is not normally distributed. This means that for a certain value of x the values of y can be normally distributed.