

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

Two major tradeoffs between customized ML and canned methods are the amount of information and processing power needed to perform the analyses. Whereas customized MLs are parametric, meaning they require starting parameters about the sample in question, a method like Least Squares does not require any information about the systems at work; instead, you are simply working phenomenologically from the data collected. Similarly, methods like Least Squares tend to be less computationally intensive, meaning that they are faster to perform and use less processing power.

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

Normality assumption: the residuals of the y values for any particular x in a data set should be normally distributed with repeated sampling.

Homogeneity assumption: stochasticity/error should be evenly distributed across the dataset, not dependent on the value of x.

Independent observations assumption: sampling is randomized, and the value of one data point does not give us information about any other data point

Fixed X assumption: all explanatory variables are measured with perfect accuracy.

3. 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)

A common misunderstanding of the normality assumption for group 1 models is that it means that the data itself must be normally distributed. If this were true though, the group 1 models would be basically useless, as very few data sets are actually normally distributed. Instead, the normality assumption refers to the distribution of the error for data at each x value. In other words, it states that if you were to repeatedly resample a population and calculate the residuals between each observed y value and the model's predicted value for one particular x, a histogram of those residuals would be normally distributed. This can be achieved no matter the distribution of the original data.

