Community Contribution
STAT 5702
Prof. Joyce Robbins
Shanzhao Qiao, Yijia Jin

When I was learning my intro class to statistics in my sophomore year, I worked with my friends on a highly-skewed dataset. Our professor told us that we are not allowed to fit a linear prediction model with such shape of distribution because the linear regression assumes the normality of the data distribution. We were asked to perform a log transformation and magically our data looked much more normal after the transformation. I googled log transformation but there was no Wikipedia page for such term, and descriptions from other websites were not comprehensive or introductive. From then on, every time I encountered a skewed dataset I would perform a log transformation, whether suitable or not.

I have always been curious about the technique and would love to write a comprehensive guidance regarding log-transformation, because for the most time we just apply the technique but neglect the intuition behind it. In the guidance, there would be theoretical analysis, examples of manipulated dataset with plots and visualizations, real-world examples, and some thoughts about situations that can not apply on. In my research of log transformation, I saw some situations that I applied the transformation anyways but in fact I should not do so, which I would describe in the guidance as well. It is surprising to know that the log transformation could be used in adjusting images' grey scale as well as adjusting sound intensity level, which is much more sophisticated compared to just the calculation of log and a few lines of code to perform the transformation in R.