STA610 Case Study 1

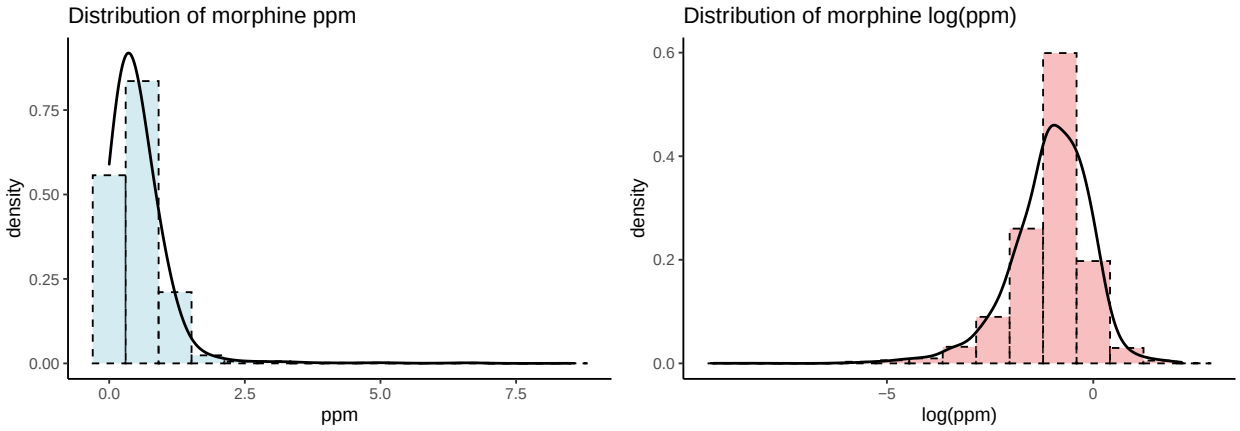
Emily Gentles, Weiyi Liu, Jack McCarthy, Qinzhe Wang

9/24/2021

Introduction

EDA

First, a look at the distributions of the response variable "ppm". Observations with ppm < 10 were considered so as to avoid the influence of extreme outliers.



The distribution of the raw ppm value is right-skewed, so we may consider applying a log transformation. The distribution of the resulting log(ppm) values is shown on the right, and it appears closer to the desired normal. Furthermore, ppm is nonnegative, so we will not lose any observations should we wish to apply the log transformation. As such, it may be beneficial to conduct our analysis on log(ppm) rather than ppm.