

Appendix I

Result of logarithmic transformation of Number of Reviews variable

The logarithmic transformation applied to the "number of reviews" variable has significantly changed its distribution. Initially right-skewed, the transformation has successfully mitigated this skewness, resulting in a more symmetric distribution. This data normalization enhances its suitability for statistical analysis, as it reduces the impact of extreme values and facilitates the application of various modeling techniques. By transforming the variable logarithmically, we have effectively improved its distributional properties, making it more manageable for analysis and interpretation.

```
```{r}
transforming the variable to log
sephora$log_number_of_reviews <- log(sephora$number_of_reviews + 2)

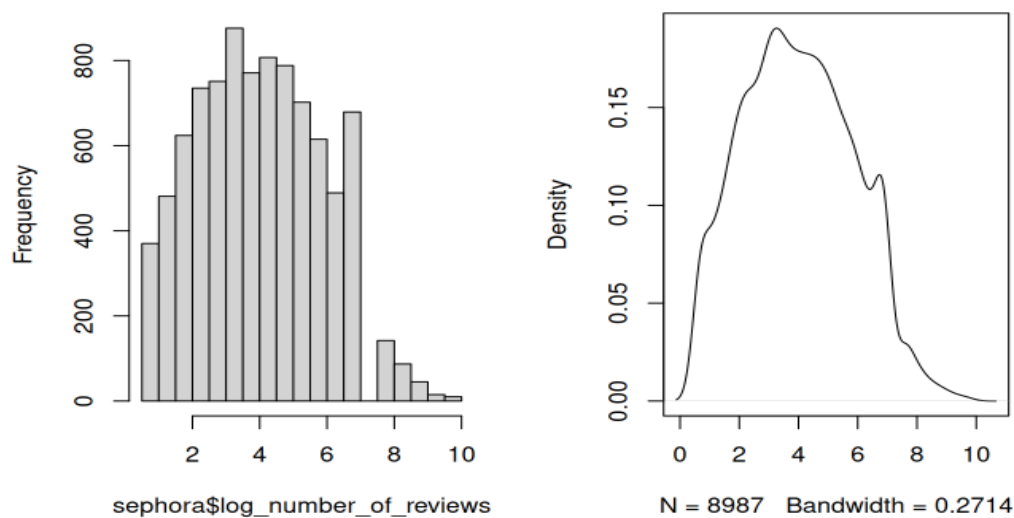
Skewness and kurtosis
skewness(sephora$log_number_of_reviews)
kurtosis(sephora$log_number_of_reviews)
```
```

Skewness: 0.187 **Kurtosis:** 2.344

```
```{r}
Set up the plotting layout
par(mfrow = c(1, 2))

Plot histogram and density
hist(sephora$log_number_of_reviews)
plot(density(sephora$log_number_of_reviews))
```
```

Histogram and density for value log Number of Reviews variable



```

```{r}
QQ plot and boxplot
qqnorm(sephora$log_number_of_reviews, main = "QQ Plot for
log_number_of_reviews variable")
qqline(sephora$log_number_of_reviews)

boxplot(sephora$log_number_of_reviews, horizontal = TRUE)
```
```{r}
```

```

QQ and box plots for log Number of Reviews variable

