Question 1

1. R codes
2. R codes
3. The mean is ~0 and the standard deviation is close to 1. Moreover, the histogram and boxplot is symmetric. This corresponds with the assumptions that x.n is a sample of size 50,000 of a normal distribution.

* Mean = 1.993904
* Median = 1.41199
* Sd = 2.601173

The sample median is smaller than the sample mean. Therefore, the distribution is skewed to the right (not symmetric). This corresponds with Pareto distribution. Therefore, we cannot use the sample mean and standard deviation to predict new variables.

Question 2

1. R codes
2. Yes? – based on summary, histogram and boxplot, there’s 2 outliers? ***How many should be removed?***
   * + Negative – row 415
     + Too high (>4,000) – row 804
3. Mean = 6.564636

Median = 2.017111

Median is different from mean because of the high observations.

1. Plot => median more stable => use median???