Quinn Knudsen

Week 1 Homework

**Getting started and Statistical Vocabulary**

**1)**

1. **Mean**-the average when taking all the sum of all values and dividing by the total number of values
2. **Median**-the middle value when lining up all the numbers from smallest to largest
3. **Mode-** the most common value in the dataset
4. **Variance-** the sum of squared deviations from the mean divided by the number of observations
5. **Standard deviation-** the square root of the variance (sum of squared deviations from the mean divided by the number of observations)
6. **Histogram**- a representation of a numeric data characterized by showing the frequency and distribution of the data
7. **Normal distribution-** the “bell curve” with the majority of values fall at the midpoint and equally tailing outward with each increase in standard deviations
8. **Poisson distribution-** a lesser known distribution best suited for arrival/time data that can have a skew left or right

**3)**

summary(BOD)

Time demand

Min. :1.000 Min. : 8.30

1st Qu.:2.250 1st Qu.:11.62

Median :3.500 Median :15.80

Mean :3.667 Mean :14.83

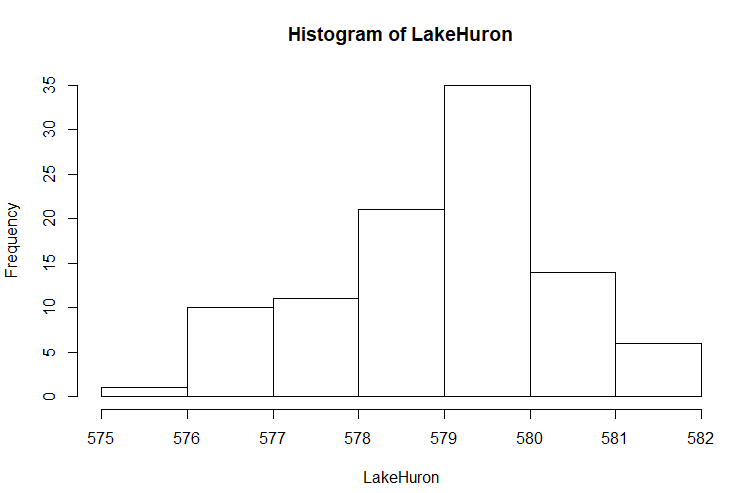
3rd Qu.:4.750 3rd Qu.:18.25

Max. :7.000 Max. :19.80

For the variable time, the median 3.5 is the halfway point for the 6 variables. In this instance, 3 and 4 are added and divided by 2 because there are an even number of variables. For demand, the median value is 14.83 which follows the same principle of add the two middle values and dividing by 2.

The mean of time is 3.667 which is the arithmetic average of the 6 variables summed and divided by 6. Demand has a mean of 14.83 following the same principles as described for the mean of time.

The practical meaning of the time central tendency metrics are less relevant for time than they are for demand. Time is not a numeric metric suited for this analysis. Demand, however, indicates that the average biochemical oxygen demand of the 6 measured points is 14.83 with a median value of 15.80 indicating a potential skew on the left tail.

**4)** The histogram is slightly skewed left. This looks to fit a normal distribution in it’s general bell curve shape. While this is a measurement over time, the annual measurements do not follow the logic of an arrival time adding further indication that this is likely a normal distribution and not a Poisson distribution. With more data collection, I would expect the distribution to be more bell shaped. 

#Code Used

summary(BOD)

length(BOD)

?BOD

summary(LakeHuron)

hist(LakeHuron)

?LakeHuron