IST 772 Homework 1

Due October 12, 2021 at 8:00AM EDT

Homework 1 by Nora Lin: I produced the material below with no assistance.

Excercise 1 p.20:

Mean: The mean is more commonly known as the average. You measure the mean by summing all the values in a vector and dividing by the number of entries.

Median: The median is the value that is in the midpoint of the data set. That means that half the data falls above this median and the other half fall below the median.

Mode: The mode is the value that appears the most in your vector.

Variance: The variance is one of the statistical descriptions that describes the spread of the data. The variance is the sum of the squared deviations devided by the number of oberservations.

Standard Deviation: The standard deviation also gives us another measurement of spread. The standard deviation is the square root of the variance

Histogram: A histogram is one way to visualize data. It falls under an x and y graph where the data is broken into buckets noting ranges. For example, in

Normal Distribution: A normal distribution is typically identified by its bell curve. The data is not skewed towards the right nor the left.

Poisson Distribution: The Poisson distribution doesn't have a distinct identifying shape. It is not a normal distribution but a probability distribution. It shows how many times an event is likely to happen.

Exercise 3 p.20:

```
#looking through all the data set options:
#data()

#looking at example data set for textbook:
#data(BOD)

#investigating data:
#airquality
summary(airquality)
```

```
##      Ozone           Solar.R          Wind             Temp
##  Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00
##  1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00
##  Median : 31.50   Median :205.0   Median : 9.700   Median :79.00
##  Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88
##  3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00
##  Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
##  NA's   :37       NA's   :7
##      Month            Day
##  Min.   :5.000   Min.   : 1.0
##  1st Qu.:6.000   1st Qu.: 8.0
##  Median :7.000   Median :16.0
```

```
##  Mean   :6.993   Mean   :15.8
##  3rd Qu.:8.000   3rd Qu.:23.0
##  Max.   :9.000   Max.   :31.0
##
```

Table 1: Summary statistics for "airquality" data set

Analysis: The data set "airquality" describes New York Air Quality Measurements. No year is specified. The median or middle month in the data set is 7.000. The average month in the data set is 6.993. The median or middle day in the data set is 16.0 and the average of the days in the data set is 15.8. The average ozone measurement in New York was 42.13 and the median or midpoint of data was 31.5. The average of the Solar R measurements is the mean of 185.9 and the midpoint of all the Solar R measurements has a value of 205.0. The wind measurements in New Yrok average to the mean value of 9.958 and all the measurements of wind have a midpoint of 9.700. The average of temperature measurements in New York is the mean of 77.88 and the midpoint of all the temperature measurements in New York is 79.00. The units of all these variables is unclear.

Exercise 4 p.20:

```
#data()

#looking at example data set for textbook:
#data(LakeHuron)

#investigating data:
#airmiles
summary(airmiles)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     412    1580    6431   10528   17532   30514
```

Table 2: Summary statistics for "airmiles" data set

Analysis: The data set "airmiles" is described as the "Passenger Miles on Commercial US Airlines, 1937 - 1960". The mean is the average. The average number of miles that passengers took on commercial US airlines between 1937 and 1960 was 10528 miles. The median is the value that is in the midpoint of the data. The median and the mean are not the same in this data set. The middle of data set is 6431.

End of Homework Assignment #1