STAT 312 Homework 1

Max Tjen

7/11/21

Problem 1:

An article in a women's magazine reported that women who nurse their babies feel warmer and more receptive towards their infants than mothers who bottle feed. This conclusion was based on the following scores from a questionnaire designed to measure warmth and receptivity, given to a random sample of 10 women from each group taken from recent hospital birth records (higher score represents more warmth and reception).

Nurse: 48.4, 51.5, 45.0, 47.0, 46.9, 47.7, 49.9, 48.5, 45.6, 48.9 Bottle feed: 42.5, 43.8, 44.5, 47.5, 44.5, 45.6, 43.4, 44.6, 47.1, 45.8

a)

Is this an observational study or a controlled experiment? Justify your answer. (2pt)

This is an observational study because the data was collected from a survey of mothers rather than having a control group that is given a treatment, which in this case would be involve the mother bottle feeding.
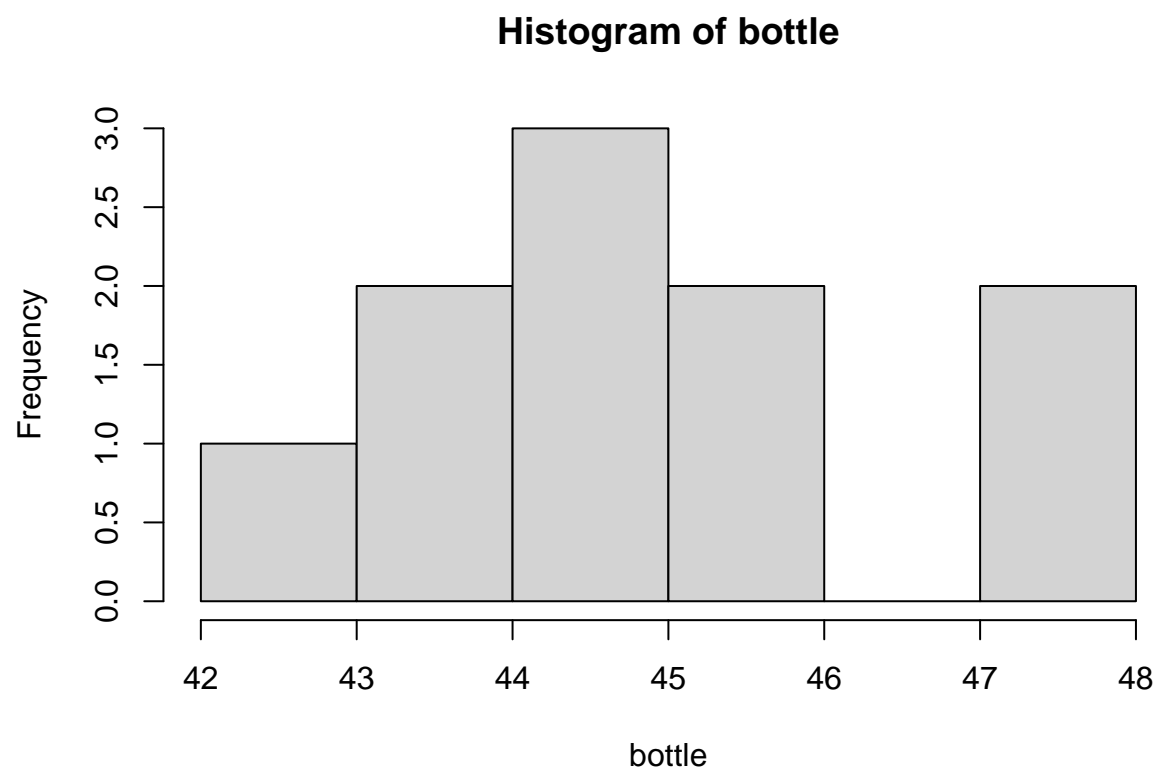
b)

Using SOCS, summarize the data for the two groups. (4pts)

```
nurse  = c(48.4, 51.5, 45.0, 47.0, 46.9, 47.7, 49.9, 48.5, 45.6, 48.9)
bottle = c(42.5, 43.8, 44.5, 47.5, 44.5, 45.6, 43.4, 44.6, 47.1, 45.8)

# see value distribution
hist(nurse)
```
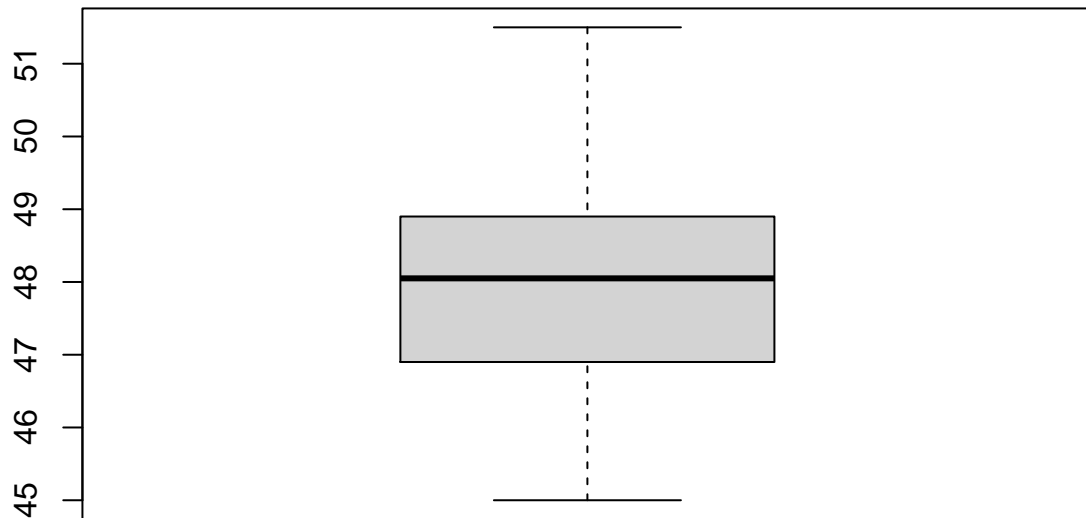
Histogram of nurse


```
hist(bottle)
```

Histogram of bottle


```r
# look for outliers
boxplot(nurse)
```

```r
summary(nurse)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   45.00   46.92   48.05   47.94   48.80   51.50
```

```r
nurseIQR = summary(nurse)[5] - summary(nurse)[2]
summary(nurse)[5] + nurseIQR
```
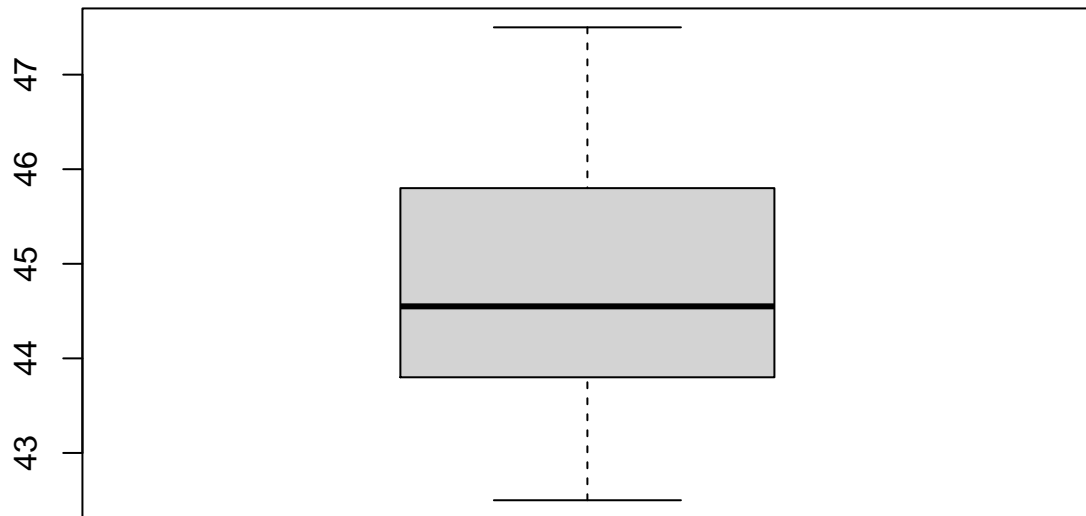
```
## 3rd Qu.
##  50.675
```

```r
summary(nurse)[2] - nurseIQR
```

```
## 1st Qu.
##   45.05
```

```r
sort(nurse)
```

```
##  [1] 45.0 45.6 46.9 47.0 47.7 48.4 48.5 48.9 49.9 51.5
```

```r
boxplot(bottle)
```

```
summary(bottle)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   42.50   43.98   44.55   44.93   45.75   47.50
```

```
bottleIQR = summary(bottle)[5] - summary(bottle)[2]
summary(bottle)[5] + bottleIQR
```

```
## 3rd Qu.
##  47.525
```

```
summary(bottle)[2] - bottleIQR
```

```
## 1st Qu.
##    42.2
```

```
sort(bottle)
```

```
##  [1] 42.5 43.4 43.8 44.5 44.5 44.6 45.6 45.8 47.1 47.5
```

```
# find center representative values
mean(nurse)
```

```
## [1] 47.94
```

```r
median(nurse)
```

```
## [1] 48.05
```

```r
mean(bottle)
```

```
## [1] 44.93
```

```r
median(bottle)
```

```
## [1] 44.55
```

```r
# find the spread of data
var(nurse)
```

```
## [1] 3.789333
```

```r
sd(nurse)
```

```
## [1] 1.946621
```

```r
range(nurse)
```

```
## [1] 45.0 51.5
```

```r
range(nurse)[2]-range(nurse)[1]
```

```
## [1] 6.5
```

```r
summary(nurse)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   45.00   46.92   48.05   47.94   48.80   51.50
```

```r
var(bottle)
```

```
## [1] 2.502333
```

```r
sd(bottle)
```

```
## [1] 1.581877
```

```
range(bottle)[2]-range(bottle)[1]
```

```
## [1] 5
```

```
summary(bottle)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   42.50   43.98   44.55   44.93   45.75   47.50
```

nurse SOCS
left skewed, two outliers (45.0, 51.5), mean = 47.94, median = 48.05, variance = 3.79, standard deviation = 1.95, range = 6.50
5 point summary: min = 45, Q1 = 46.92, Q2 = 48.05, Q3 = 48.8, max = 51.5

bottle SOCS
right skewed, no outliers, mean = 44.93, median = 44.55, variance = 2.50, standard deviation = 1.58, range = 5.0
5 point summary: min = 42.5, Q1 = 43.98, Q2 = 44.55, Q3 = 45.75, max = 47.5

c)

Using your descriptions in part b, compare the scores in the two groups. (2pts)

The two datasets provide data that seems to point show that nursing an infant rather than bottlefeeding them makes the mother more warm and receptive to the baby. This can be seen by the mother's having both a higher mean and median score when nursing rather than bottlefeeding, as the nurse data had a mean score that was 3.01 higher than the bottlefeeding mean and the nurse median score was 3.5 higher than the bottlefeeding median respectively. The nurse data had two outliers, which were the minimum and maximum values, while the bottle data had no outliers, albeit the maximum value was close to being so. The standard deviations of both datasets are pretty small, with the nurse data having a standard deviation of 1.95 and the bottle data having a standard deviation of 1.58. Given that there are two outliers in the nurse data, it makes sense that the standard deviation is a bit bigger than the bottle data. The range values for each was 6.5 for the nurse data and 5.0 for the bottle data. With the ranges and standard deviations of each dataset, it seems like the bottle data is more similar than the nurse data as there is less spread across the data. These values of data spread reinforce the idea that both datasets have pretty precise data.

d)

If you or your partner were thinking about breast feeding, what effect would this study have on your decision? Why? (2 pts)

This study would somewhat affect the decision because it seems like for the most part, mothers who nursed their babies felt warmer and more receptive towards them relative to mothers who decided to bottle feed their baby. As such, it would probably be better for the relationship between the mother and their baby if there was more warmth and reception between them.

Problem 2:

Do the following problem from Chapter 6 in MR (Textbook: Montgomery & Runger): 6.S17(2+4+3+1=10pts)

One of the authors (DCM) has a Mercedes-Benz 500 SL Roadster. It is a 2003 model and has fairly low mileage (currently 60,112 miles on the odometer). He is interested in learning how his car's mileage compares with the mileage on similar SLs. The table below contains the mileage on 100 Mercedes-Benz SLs from the model years 2003–2009 taken from the Cars.com website.

a)

Calculate the sample mean and standard deviation of the odometer readings.

```
mileage = read.csv('hw1.csv')  # file is just the column of data
mean(mileage$Mileage.Data)
```

```
## [1] 34232.05
```
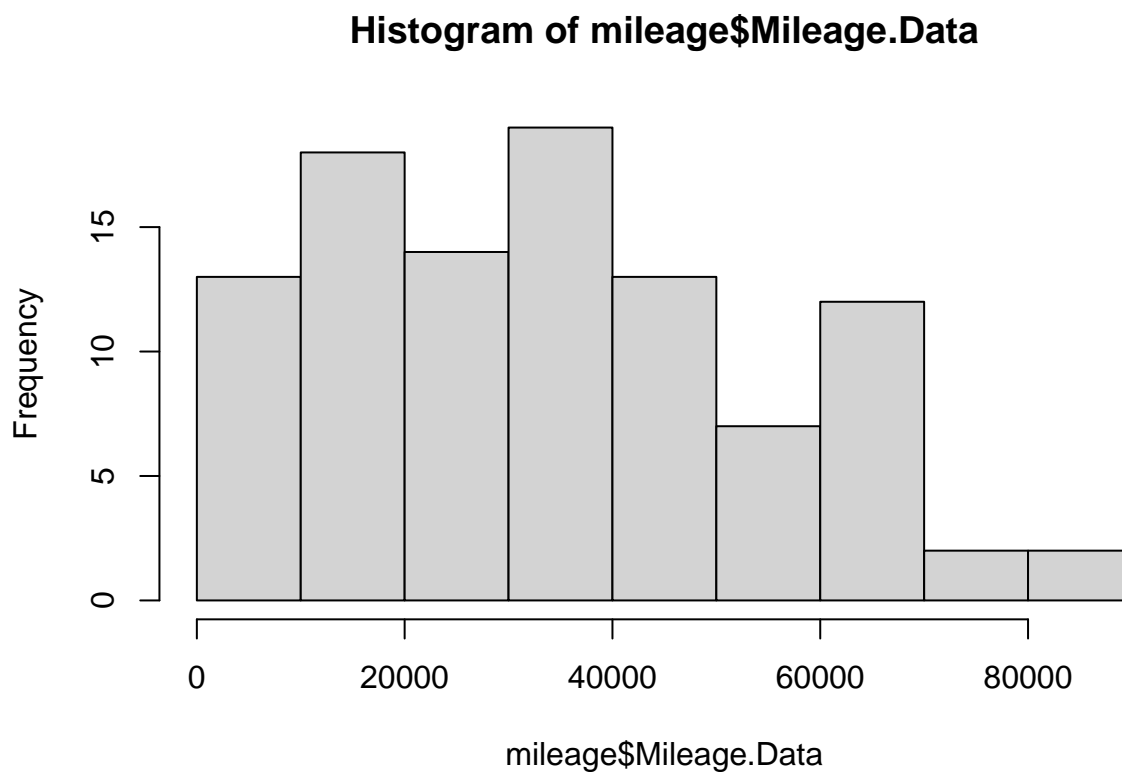
```
sd(mileage$Mileage.Data)
```

```
## [1] 20414.52
```

mean = 34,232.05, standard deviation = 20,414.52

b)

Construct a histogram of the odometer readings and comment on the shape of the data distribution.

```
hist(mileage$Mileage.Data)
```


Histogram of mileage$Mileage.Data

```
summary(mileage$Mileage.Data)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1698   18238   31756   34232   48767   85475
```

The shape of the odometer readings appear to be right skewed. This can be confirmed by finding the mean and median, which shows that the mean is greater than the median.

c)

Construct a stem-and-leaf diagram of the odometer readings.

```
stem(mileage$Mileage.Data, scale = 2)
```

```
##
##   The decimal point is 4 digit(s) to the right of the |
##
##   0 | 222444
##   0 | 5668999
##   1 | 024
##   1 | 5566667888899
##   2 | 001223
##   2 | 6667999999
##   3 | 0001222333444
##   3 | 678888
##   4 | 011333
##   4 | 5678999
##   5 | 1124
##   5 | 689
##   6 | 00133334
##   6 | 5677
##   7 | 2
##   7 | 8
##   8 | 4
##   8 | 5
```

d)

What is the percentile of DCM's mileage?

```
quantile(mileage$Mileage.Data, probs = seq(.1, .9, by = .1))
```

```
##     10%     20%     30%     40%     50%     60%     70%     80%     90%
##   8803.8 15953.4 19779.6 28783.6 31756.5 37208.2 43850.6 52374.0 63250.1
```

```
quantile(mileage$Mileage.Data, probs = seq(.8, .9, by = .01))
```

```
##       80%       81%       82%       83%       84%       85%       86%       87%
## 52374.00 54541.72 56618.56 58094.40 58833.68 60456.50 60510.76 60832.21
##       88%       89%       90%
## 62552.80 62973.99 63250.10
```

DCM's mileage is in the 85th percentile of this data