Title: “Assignment 1(Intro to stats)”

Author: “S3807007 Riya Minesh Amin”

Date: “30/08/2019”

#The package dplyr is imported here before. The structure of the data is checked with str(). Data types consist of num and int. The categories male (m) and female (f) has been given the int values 1 and 0, respectively.

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(readxl)
bdims_csv <- read_excel("C:/Users/s3807007/bdims (1).xlsx")
 View(bdims_csv)
 str(bdims_csv)

## Classes 'tbl_df', 'tbl' and 'data.frame':    507 obs. of  25 variables:
##  $ bia.di: num  42.9 43.7 40.1 44.3 42.5 43.3 43.5 44.4 43.5 42 ...
##  $ bii.di: num  26 28.5 28.2 29.9 29.9 27 30 29.8 26.5 28 ...
##  $ bit.di: num  31.5 33.5 33.3 34 34 31.5 34 33.2 32.1 34 ...
##  $ che.de: num  17.7 16.9 20.9 18.4 21.5 19.6 21.9 21.8 15.5 22.5 ...
##  $ che.di: num  28 30.8 31.7 28.2 29.4 31.3 31.7 28.8 27.5 28 ...
##  $ elb.di: num  13.1 14 13.9 13.9 15.2 14 16.1 15.1 14.1 15.6 ...
##  $ wri.di: num  10.4 11.8 10.9 11.2 11.6 11.5 12.5 11.9 11.2 12 ...
##  $ kne.di: num  18.8 20.6 19.7 20.9 20.7 18.8 20.8 21 18.9 21.1 ...
##  $ ank.di: num  14.1 15.1 14.1 15 14.9 13.9 15.6 14.6 13.2 15 ...
##  $ sho.gi: num  106 110 115 104 108 ...
##  $ che.gi: num  89.5 97 97.5 97 97.5 ...
##  $ wai.gi: num  71.5 79 83.2 77.8 80 82.5 82 76.8 68.5 77.5 ...
##  $ nav.gi: num  74.5 86.5 82.9 78.8 82.5 80.1 84 80.5 69 81.5 ...
##  $ hip.gi: num  93.5 94.8 95 94 98.5 95.3 101 98 89.5 99.8 ...
##  $ thi.gi: num  51.5 51.5 57.3 53 55.4 57.5 60.9 56 50 59.8 ...
##  $ bic.gi: num  32.5 34.4 33.4 31 32 33 42.4 34.1 33 36.5 ...
##  $ for.gi: num  26 28 28.8 26.2 28.4 28 32.3 28 26 29.2 ...
##  $ kne.gi: num  34.5 36.5 37 37 37.7 36.6 40.1 39.2 35.5 38.3 ...
##  $ cal.gi: num  36.5 37.5 37.3 34.8 38.6 36.1 40.3 36.7 35 38.6 ...
##  $ ank.gi: num  23.5 24.5 21.9 23 24.4 23.5 23.6 22.5 22 22.2 ...
```

```
##  $ wri.gi: num  16.5 17 16.9 16.6 18 16.9 18.8 18 16.5 16.9 ...
##  $ age   : num  21 23 28 23 22 21 26 27 23 21 ...
##  $ wgt   : num  65.6 71.8 80.7 72.6 78.8 74.8 86.4 78.4 62 81.6 ...
##  $ hgt   : num  174 175 194 186 187 ...
##  $ sex   : num  1 1 1 1 1 1 1 1 1 1 ...
```

#Here the rows of male and female are confirmed. rows from 2 to 248 are males and rows from 249 to 507 are females.

```
bdims_csv$sex[c(1,247,248,507)]

## [1] 1 1 0 0

table(bdims_csv$sex[2:247])

##
##   1
## 246

table(bdims_csv$sex[248:507])

##
##   0
## 260
```

Here Summary Statistics of the dataset is shown

#Mean (All, Male, Female)

```
mean(bdims_csv$che.de)

## [1] 19.22604

mean(bdims_csv$che.de[1:247])

## [1] 20.80648

mean(bdims_csv$che.de[248:507])

## [1] 17.72462
```

#Median (All, Male, Female)

```
median(bdims_csv$che.de)

## [1] 19

median(bdims_csv$che.de[1:248])

## [1] 20.6

median(bdims_csv$che.de[249:507])
```

```
## [1] 17.5
```

#Range (All, Male, Female) - Gives min and max

```
range(bdims_csv$che.de)

## [1] 14.3 27.5

range(bdims_csv$che.de[1:247])

## [1] 14.4 27.5

range(bdims_csv$che.de[248:507])

## [1] 14.3 26.8
```

#Varience (All, Male, Female)

```
var(bdims_csv$che.de)

## [1] 6.329637

var(bdims_csv$che.de[1:247])

## [1] 4.595161

var(bdims_csv$che.de[248:507])

## [1] 3.356457
```

#Standard Deviation (All, Male, Female)

```
sd(bdims_csv$che.de)

## [1] 2.515877

sd(bdims_csv$che.de[1:247])

## [1] 2.143633

sd(bdims_csv$che.de[248:507])

## [1] 1.832064
```

#Quartile (All, Male, Female)

```
quantile(bdims_csv$che.de)

##   0%  25%  50%  75% 100% 
## 14.3 17.3 19.0 20.9 27.5

quantile(bdims_csv$che.de[1:247])

##    0%   25%   50%   75%  100% 
## 14.40 19.35 20.60 22.10 27.50
```

```r
quantile(bdims_csv$che.de[248:507])

##   0%  25%  50%  75% 100% 
## 14.3 16.5 17.5 18.7 26.8
```

#interquartile

```r
IQR(bdims_csv$che.de)

## [1] 3.6

IQR(bdims_csv$che.de[1:247])

## [1] 2.75

IQR(bdims_csv$che.de[248:507])

## [1] 2.2
```

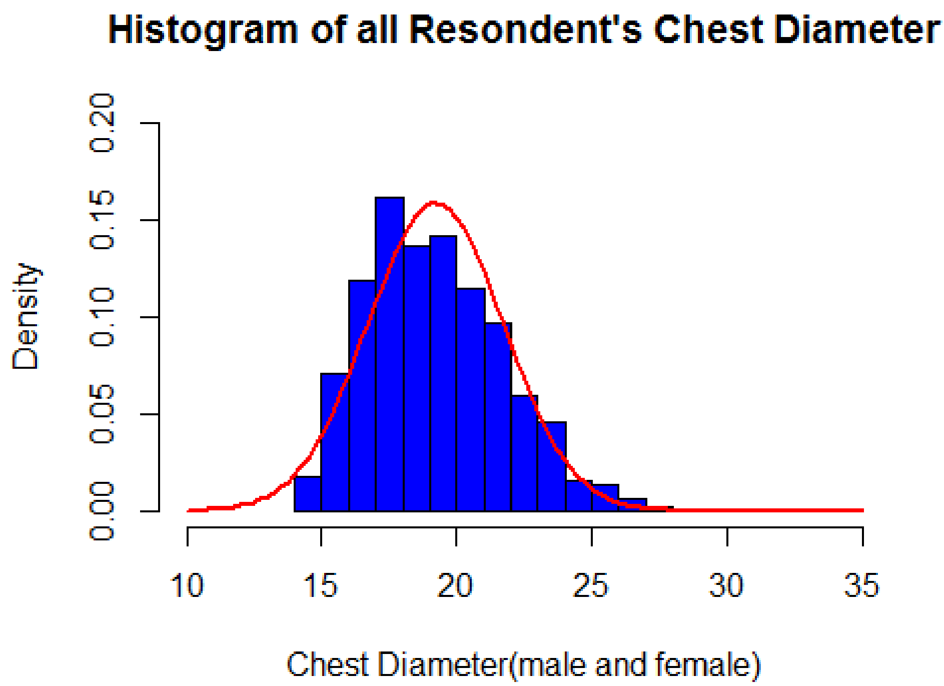
here grouping is done.

```r
bdims_csv %>% group_by(sex) %>% summarise(Min = min(bdims_csv$che.de,na.rm = 
TRUE),
                Q1 = quantile(bdims_csv$che.de,probs= .25,na.rm = TRUE),
                Median = median(bdims_csv$che.de, na.rm = TRUE),
                 Q3 = quantile(bdims_csv$che.de,probs = .75,na.rm = TRUE),
                Max = max(bdims_csv$che.de,na.rm = TRUE),
                Mean = mean(bdims_csv$che.de, na.rm = TRUE),
                 SD = sd(bdims_csv$che.de, na.rm = TRUE),n = n(),
                   Missing = sum(is.na(bdims_csv$che.de)))

## # A tibble: 2 x 10
##     sex   Min    Q1 Median    Q3   Max  Mean    SD     n Missing
##   <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <int>   <int>
## 1     0  14.3  17.3     19  20.9  27.5  19.2  2.52   260       0
## 2     1  14.3  17.3     19  20.9  27.5  19.2  2.52   247       0
```

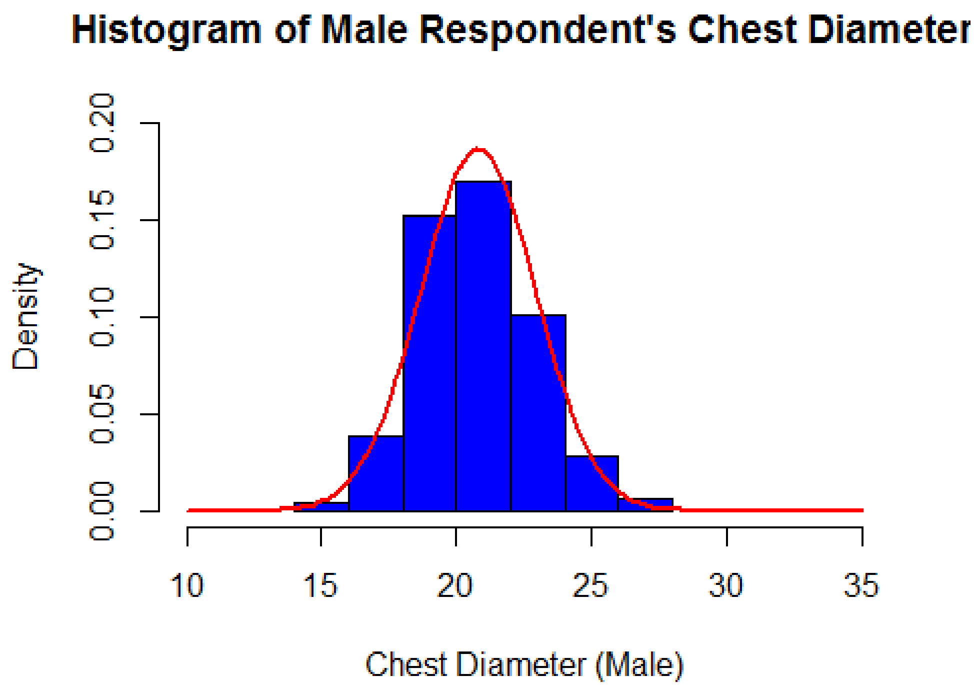
#Distribution Fitting #Histogram (All)

```r
mean_all <- mean(bdims_csv$che.de)
sd_all <- sd(bdims_csv$che.de)
hist(bdims_csv$che.de, xlim = c(10, 35), ylim = c(0, 0.2), col="blue", xlab="
Chest Diameter(male and female)",
                         main="Histogram of all Resondent's Chest Diameter",
freq = FALSE)
    curve(expr = dnorm(x, mean=mean_all, sd = sd_all), col="red", lwd=2, add=
TRUE)
```



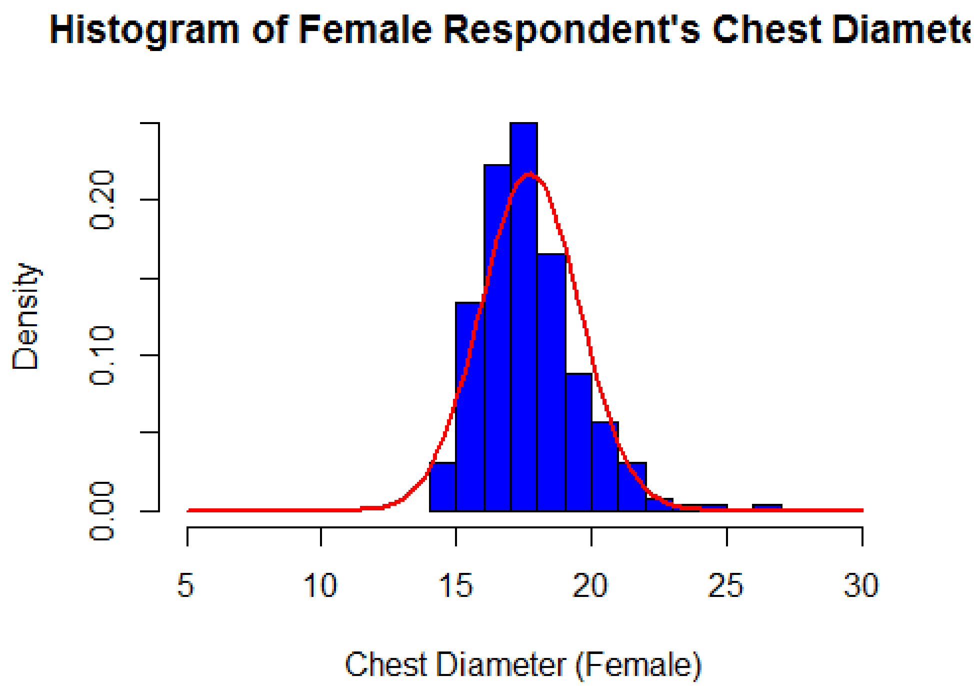
Histogram (Male)

```
mean_male <- mean(bdims_csv$che.de[1:247])
sd_male <- sd(bdims_csv$che.de[1:247])
hist(bdims_csv$che.de[1:247], xlim = c(10, 35), ylim = c(0, 0.2), col="blue",
xlab="Chest Diameter (Male)",
                         main="Histogram of Male Respondent's Chest Diameter"
, freq = FALSE)
    curve(expr = dnorm(x, mean=mean_male, sd = sd_male), col="red", lwd=2, ad
d=TRUE)
```



#histogram female

```
mean_female <- mean(bdims_csv$che.de[248:507])
sd_female <- sd(bdims_csv$che.de[248:507])
hist(bdims_csv$che.de[248:507], xlim = c(5, 30), ylim = c(0, 0.25), col="blue
", xlab="Chest Diameter (Female)",
                         main="Histogram of Female Respondent's Chest Diamete
r", freq = FALSE)
    curve(expr = dnorm(x, mean=mean_female, sd = sd_female), col="red", lwd=2
, add=TRUE)
```



By visualizing the results, we initially saw a histogram with a symmetrical spread (all respondents). This histogram can be seen to poorly fit the overlaid normalized curve. When the data was split into sex, we saw that the histogram (empirical) for females' che.de gave a slightly right skewed unimodal plot. And for males' che.de it gave almost perfect skewness.The dispensation and subsequent visualization shows that for the variable che.de, splitting the respondent's into male and female gives a more significant statistical result.