MATH1324 Assignment 1

Modeling Body Measurements

Student Details

Rinaldo Gagiano (S3870806)

Problem Statement

Determine if the knee diameter measurements fit a normal distribution for both males and females. The body measurement explored will be the respondent's knee diameter in centimetres, measured as the sum of two knees. The approach taken to compare the theoretical normal distribution to the empirical data is to overlay a normal distribution density curve produced by the function 'dnorm()'.

Load Packages

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------- tidyverse 1.3
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  2.1.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------- tidyverse_conflicts
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readxl)
library(dplyr)
library(lattice)
library(ggplot2)
```

Data

Data Importation:

```
df <- read_xlsx("bdims.xlsx")
```

Data Cleaning and Preperation:

```
knee_df <- cbind(df[,18],df[,25])                    #Original Data Subset
colnames(knee_df)[1]<-c("Knee_Diameter")             #Column Rename
knee_df$sex <- ifelse(knee_df$sex==1,'M','F')        #Reassigning labels to numerical values
knee_df$sex <- factor(knee_df$sex,levels=c('F','M'))  #Factoring of Sex Variable
```

Data Preview:

```
head(knee_df,10)   #Preview of the first 10 observations
```

```
##    Knee_Diameter sex
## 1           34.5   M
## 2           36.5   M
## 3           37.0   M
## 4           37.0   M
## 5           37.7   M
## 6           36.6   M
## 7           40.1   M
## 8           39.2   M
## 9           35.5   M
## 10          38.3   M
```

Summary Statistics

Calculate descriptive statistics (i.e., mean, median, standard deviation, first and third quartile, interquartile range, minimum and maximum values) of the selected measurement grouped by sex.

```
knee_df %>% group_by(sex) %>% summarise(Min = min(Knee_Diameter,na.rm = TRUE),
                                        Q1 = quantile(Knee_Diameter,probs = .25,na.rm
                                                      = TRUE),
                                        Median = median(Knee_Diameter, na.rm = TRUE),
                                        Q3 = quantile(Knee_Diameter,probs = .75,na.rm
                                                      = TRUE),
                                        Max = max(Knee_Diameter,na.rm = TRUE),
                                        Mean = mean(Knee_Diameter, na.rm = TRUE),
                                        SD = sd(Knee_Diameter, na.rm = TRUE),
                                        IQR = quantile(Knee_Diameter,probs = .75,na.rm
                                                       = TRUE)
                                        -quantile(Knee_Diameter,probs = .25,na.rm = TRUE),
                                        n = n(),
                                        Missing = sum(is.na(Knee_Diameter)))
```

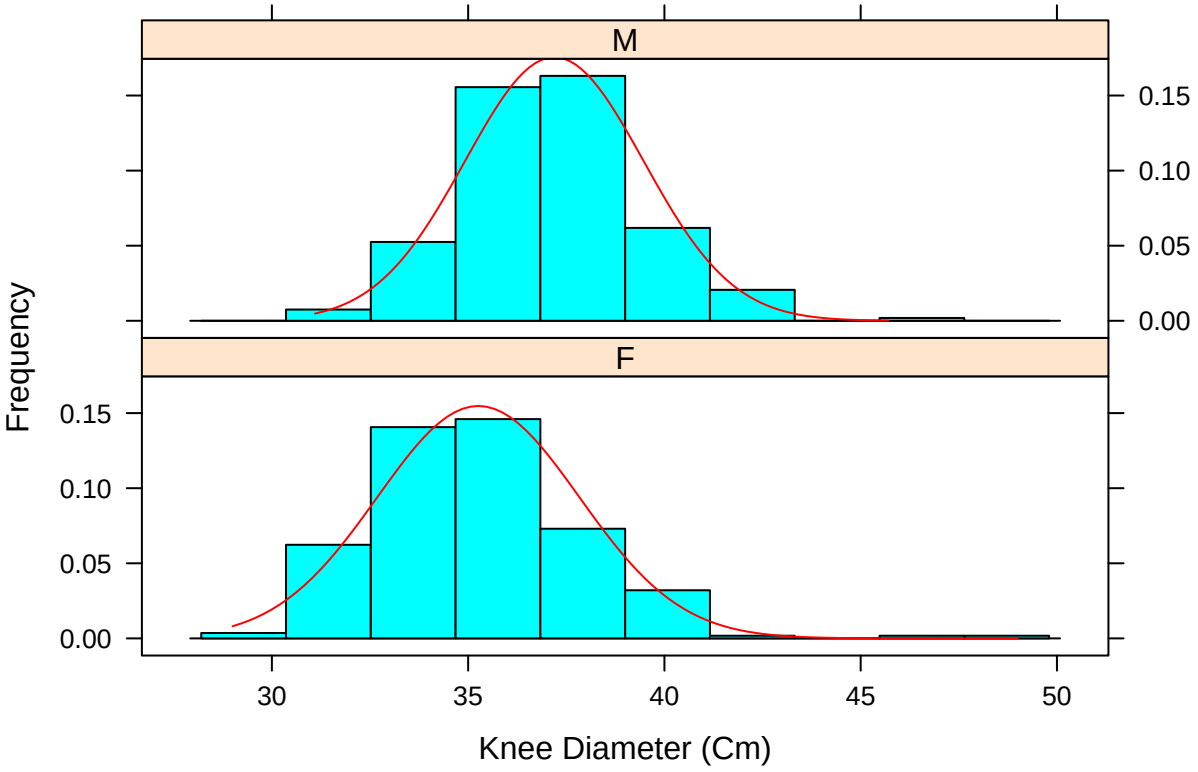
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 11
##   sex     Min    Q1 Median    Q3   Max  Mean    SD   IQR     n Missing
##   <fct> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <int>   <int>
## 1 F      29    33.5     35  36.8  49    35.3  2.58  3.30   260       0
## 2 M      31.1  35.8     37  38.4  45.7  37.2  2.27  2.7    247       0
```

Distribution Fitting

Compare the empirical distribution of selected body measurement to a normal distribution separately in men and in women. You need to do this visually by plotting the histogram with normal distribution overlay. Show your code.

```
knee_df %>% histogram(~Knee_Diameter | sex, data = .,layout=c(1,2),type = "density",
                      ylab = "Frequency",xlab="Knee Diameter (Cm)",
                      panel=function(x, ...){
                        panel.histogram(x, ...)
                        xn <- seq(min(x), max(x), length.out = 100)
                        yn <- dnorm(xn, mean(x), sd(x))
                        panel.lines(xn, yn, col = "red")
                      })
```



Interpretation

The problem at hand was to determine if knee diameters, for both males and females, were normally distributed. Looking at the graphs, it can be seen that both groups of genders' empirical data in blue, generally fit the normal distribution overlay in red. Males appear to fit the expected normal distribution a bit better, as their range was more contained at 14.6 compared to the females' range at 20. Both sexes have a median almost equal to their respective mean and both have quite small standard deviations. With all this in mind, it is a comfortable assumption to make that knee diameter, in both males and females, will be found to fall into a normal distribution.

References

Iowa, University. Histograms and Density Plots, homepage.divms.uiowa.edu/~luke/classes/STAT4580/histdens.html.