MATH1324 Assignment 1

Code

Modeling Body Measurements

Student Details

Student Name (s3835204)

Problem Statement

Write a clear and concise problem statement that guides your investigation. Explain which variable you choose and outline the approaches taken for normal distribution fitting.

Males and Females tend to have different body dimensions. The variable of interest is (che.di) which indicates Respondent's chest diameter in centimeters, measured at nipple level, mid-expiration. The investigation includes selecting a particular variable to determine if the selected variable (che.di) fits Normal Distribution separately in both Male and Female.

Investigation begins by loading the required packages and dataset bdims.csv. Then, a factor is used to further define variable sex with appropriate levels and labels to it. The summary of statistical information is undertaken followed by visually plotting the histogram for analysis. The investigation ends by comparing the empirical data with the theoretical Normal Distribution.

Load Packages

Hide

```
# This is a chunk where you can load the necessary packages required to reproduce the report
library(readr)
library(magrittr)
library(dplyr)
library(ggplot2)
library(rmarkdown)
```

Data

Import the body measurements data and prepare it for analysis. Show your code.

bdims.csv is the datatset. Data type variable "sex" is numeric with values 0 and 1, it is then converted to factors of two levels i.e Female and Male.

Hide

```
# This is a chunk for your Data section.
bdims <- read_csv("F:/MS/Sem 1/AA/Assignment1/bdims.csv")
```

```
Parsed with column specification:
cols(
  .default = col_double()
)
See spec(...) for full column specifications.
```

Hide

```
# Variable sex is defined as a factor with two levels labelled as Male and Female
bdims$sex <-
  bdims$sex %>% factor(levels = c(0,1),
                       labels = c("Female","Male"))
```

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.

Hide

```
# This is a chunk for your Summary Statistics section.
bdims$che.di %>% summary()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  22.20   25.65   27.80   27.97   29.95   35.60
```

Hide

```
bdims$che.di %>% sd() # Standard Deviation
```

```
[1] 2.74165
```

Hide

```
bdims$che.di %>% IQR() # Interquartile Range
```

```
[1] 4.3
```

Hide

```
# Summary statistics for Male
bdimsmale <- bdims %>% filter(sex == "Male")
bdimsmale$che.di %>% summary()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  24.70   28.65   29.90   29.95   31.40   35.60
```

Hide

```
bdimsmale$che.di %>% sd() # Standard Deviation
```

```
[1] 2.083108
```

Hide

```
bdimsmale$che.di %>% IQR() # Interquartile Range
```

```
[1] 2.75
```

Hide

```
# Summary statistics for Female
bdimsfemale <- bdims %>% filter(sex == "Female")
bdimsfemale$che.di %>% summary()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   22.2    24.9    25.9    26.1    27.1    33.2
```

Hide

```
bdimsfemale$che.di %>% sd() # Standard Deviation
```

```
[1] 1.818808
```

Hide

```
bdimsfemale$che.di %>% IQR() # Interquartile Range
```

```
[1] 2.2
```

Hide

```
# Summary statistics grouped by Sex for Male and Female
bdims %>% group_by(sex) %>% summarise(Min = min(che.di, na.rm = TRUE),
                                      Q1 = quantile(che.di, probs = .25, na.rm = TRUE),
                                      Median = median(che.di ,na.rim = TRUE),
                                      Q3 = quantile(che.di, probs = .75, na.rm = TRUE),
                                      Max = max(che.di, na.rm = TRUE),
                                      Mean = mean(che.di, na.rm = TRUE),
                                      SD = sd(che.di , na.rm = TRUE),
                                      InterquartileRange = IQR(che.di, na.rm = TRUE),
                                      N = n())
```

sex	Min	Q1	Median	Q3	Max	Mean	SD	InterquartileRange	N
<fctr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
Female	22.2	24.90	25.9	27.1	33.2	26.09731	1.818808	2.20	260
Male	24.7	28.65	29.9	31.4	35.6	29.94899	2.083108	2.75	247

2 rows

Distribution Fitting

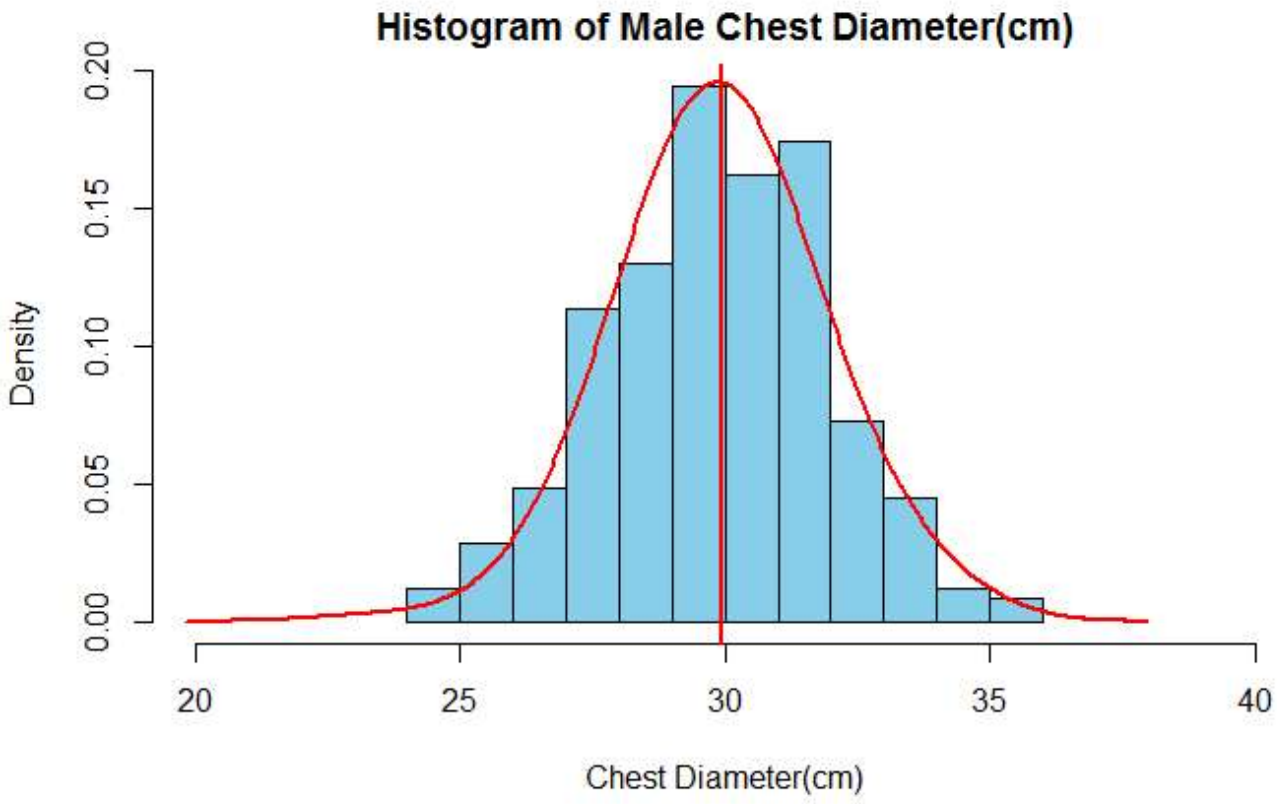
Compare the empirical distribution of selected body measurement to a normal distribution separately in men and in women. You need to do this visually by plotting the histogram with normal distribution overlay. Show your code.

Hide

```
# This is a chunk for your Distribution Fitting section.

# Histogram of Male Chest Diameter in centimeters, measured at nipple level,      mid-expirat
ion.
hist(bdimsmale$che.di, breaks = 15, xlim = c(20,40),
      xlab = "Chest Diameter(cm)",
      ylab = "Density",
      main = "Histogram of Male Chest Diameter(cm)",
      col = "skyblue", probability = TRUE)


## Overlay a Normal Distribution curve on histogram

# Generate a Normal Distribution with mean, standard deviation and the Number of observations
bdimsov <- rnorm(length(bdimsmale$che.di), mean(bdimsmale$che.di), sd(bdimsmale$che.di))

# Adding the Normal Distribution to histogram using density and lines function
lines(density(bdimsov, adjust = 2), col = "red", lwd = 2)
```

Hide

```
# Adding a vertical line with the mean to histogram
bdimsmale$che.di %>% mean() %>% abline(v = .,col = 'red',lw = 2)
```

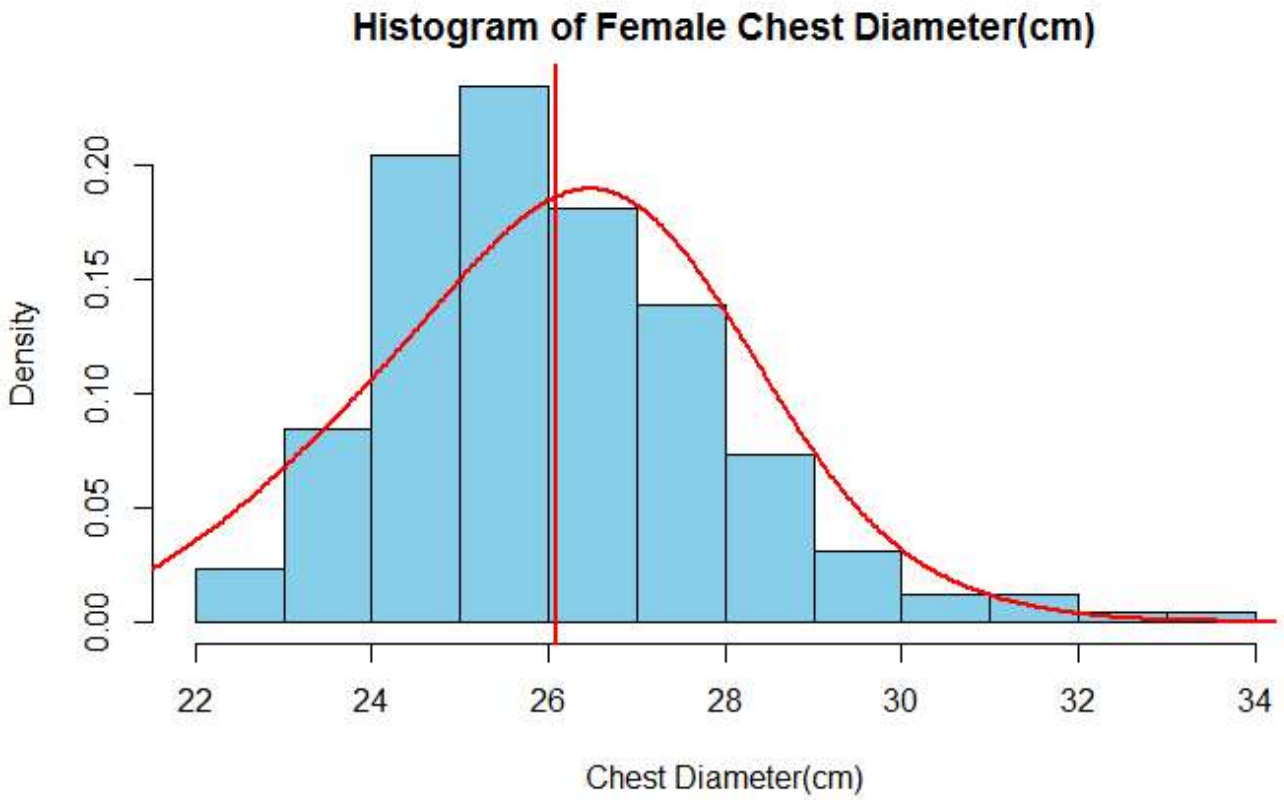


Hide

```
# Histogram of Female Chest Diameter in centimeters, measured at nipple level,      mid-expir
ation.
hist(bdimsfemale$che.di, breaks = 15,
     xlab = "Chest Diameter(cm)",
     ylab = "Density",
     main = "Histogram of Female Chest Diameter(cm)",
     col = "skyblue", probability = TRUE)


## Overlay a Normal Distribution curve on histogram

# Generate a Normal Distribution with mean, standard deviation and the Number of observations
bdimsovf <- rnorm(length(bdimsfemale$che.di), mean(bdimsfemale$che.di), sd(bdimsfemale$che.d
i))

# Adding the Normal Distribution to histogram using density and lines function
lines(density(bdimsovf, adjust = 2), col = "red", lwd = 2)
```

Hide

```
# Adding a vertical line with the mean to histogram
bdimsfemale$che.di %>% mean() %>% abline(v = .,col = 'red',lw = 2)
```



Interpretation

Going back to your problem statement, what insight has been gained from the investigation? Discuss the extent to how your theoretical normal distribution fits the empirical data.

Taking into account the statistics and visually analyzing the histogram, it can be said that Male chest diameter(che.di) fits normal distribution as it matches Normal distribution characteristics. Chest Diameter measurements for the Male population include continuous random values from 24.7 cm to 35.6 cm. The

calculated values of both Mean = 29.95 and Median = 29.9 are equal. The Normal Distribution curve forms a bell-shaped curve over the histogram. Also, the distribution curve is symmetrical on both sides. The standard deviation for the Chest Diameter of the Male population is 2.08 cm. Adding the ± 2 standard deviation to the mean includes 95% of the empirical data. The total area under the Normal Distribution curve is equal to one.

As a result, it can be implied that only the Male population of the selected variable fits the normal distribution.