Name: Nikhil H

Modelling Body Measurements

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.

The aim is to find if the biacromial diameter fits the normal distribution seperately in men and women as per data provided by Data source(Heinz G, Peterson LJ, Johnson RW, Kerk CJ. 2003. Exploring Relationships in Body Dimensions. Journal of Statistics Education 11(2)). To check for normal distribution using the shapiro wilk test and the QQ plots and analyse the normal distribution data with resepect to the empirical distribution.

Load Packages

Hide

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

Data

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

Hide

```
# This is a chunk for your Data section.  bdims <- read.csv(file =
'bdims.csv',sep=",", dec=".", fill = TRUE,strip.white =FALSE,
stringsAsFactors = default.stringsAsFactors())
# defining the categorical variable sex as factors
bdims$sex <- factor(bdims$sex, levels = c(0,1),ordered = TRUE )
#labelling the sex factors as 0 = 'Female' and 1 = 'Male'
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.
## select variable bia.di(biacromial diameter in centimeters) and gender
bia_gender <- bdims %>% dplyr::select(bia.di,sex)
## filter for male values
bia_male <- bia_gender %>% filter(sex =='Male')
## summary for male
bia_male %>% summary()
```

```
     bia.di          sex
 Min.   :34.10   Female:  0
 1st Qu.:40.00   Male  :247
 Median :41.20
 Mean   :41.24
 3rd Qu.:42.60
 Max.   :47.40
```

Hide

```
##std dev for biacromial diameter values in male
sd(bia_male$bia.di)
```

```
[1] 2.087164
```

Hide

```
# From the summary details we observe that the mean and median is approximately equal
```

Hide

```
## filter for female values
bia_female <- bia_gender %>% filter(sex =='Female')
## summary for female
bia_female %>% summary()
```

```
     bia.di          sex
 Min.   :32.40   Female:260
 1st Qu.:35.17   Male  :  0
 Median :36.40
 Mean   :36.50
 3rd Qu.:37.80
 Max.   :42.60
```

Hide

```
## std dev for biacromial diameter values in Female
sd(bia_female$bia.di)
```

```
[1] 1.779221
```

Hide

```
# From the summary details we observe that the mean and median is approximately equal
```

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

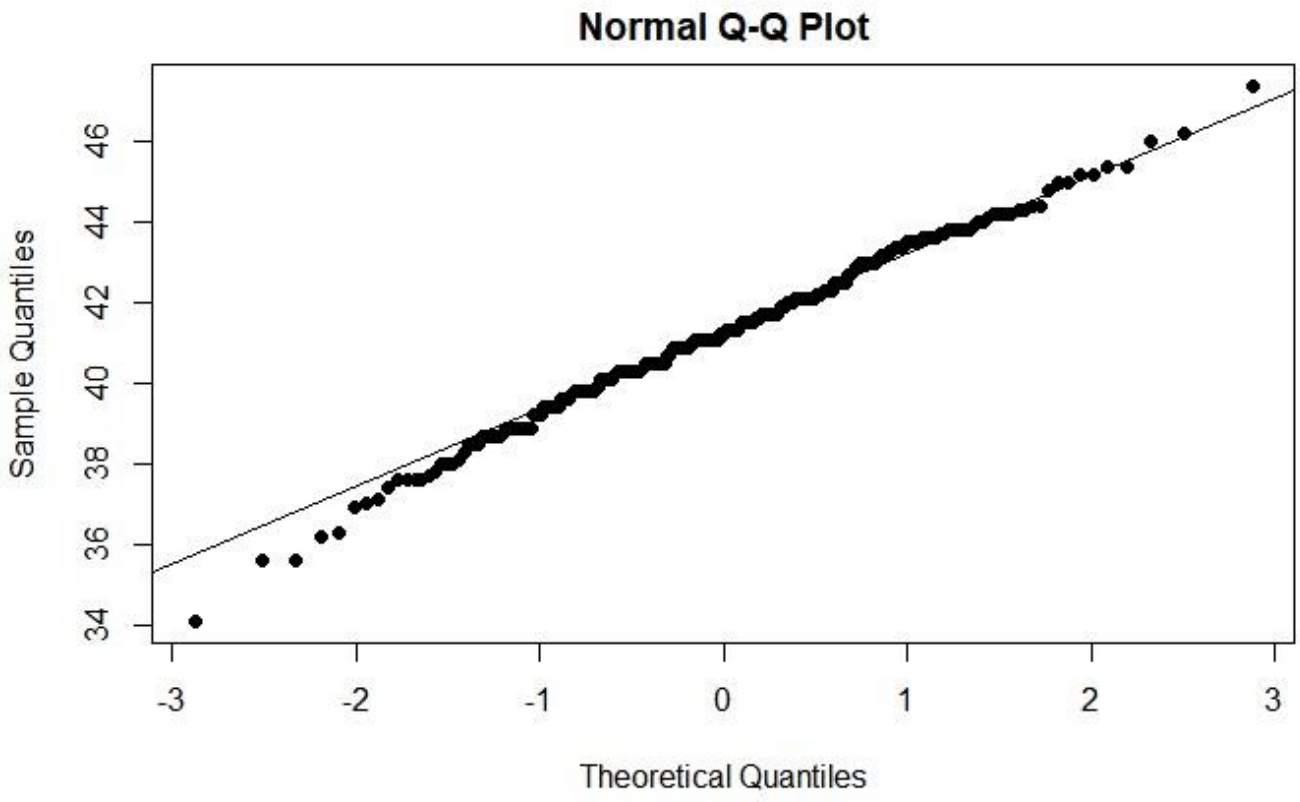
```
# This is a chunk for your Distribution Fitting section.
# verify that the male biiliac diameter fits the normal distribution by using the Shapiro Wil
k
# test and qq plots
# Reference - (for shapiro wilk test and QQ plots)
#https://www.sheffield.ac.uk/polopoly_fs/1.579191!/file/stcp-karadimitriou-normalR.pdf
# Null Hypothesis : Data is normally distributed. If p>0.05 then it is normally distributed
shapiro.test(bia_male$bia.di)
```

```
    Shapiro-Wilk normality test

data:  bia_male$bia.di
W = 0.99408, p-value = 0.4437
```

Hide

```
# as displayed in the output the p-value = 0.4437 which is greater than 0.05
# The data is approximately normal distributed with null hypothesis accepted
# QQ plots
qqnorm(bia_male$bia.di,pch=19)
qqline(bia_male$bia.di)
```



Hide

```
# The QQ plot the lines displays a perfect normal distribution
# So the data values should lie as close to the line as possible.
# Hence the male biiliac diameter fits the normal distribution
```

Hide

```
# Null Hypothesis : Data is normally distributed. If p>0.05 then it is normally distributed
shapiro.test(bia_female$bia.di)
```

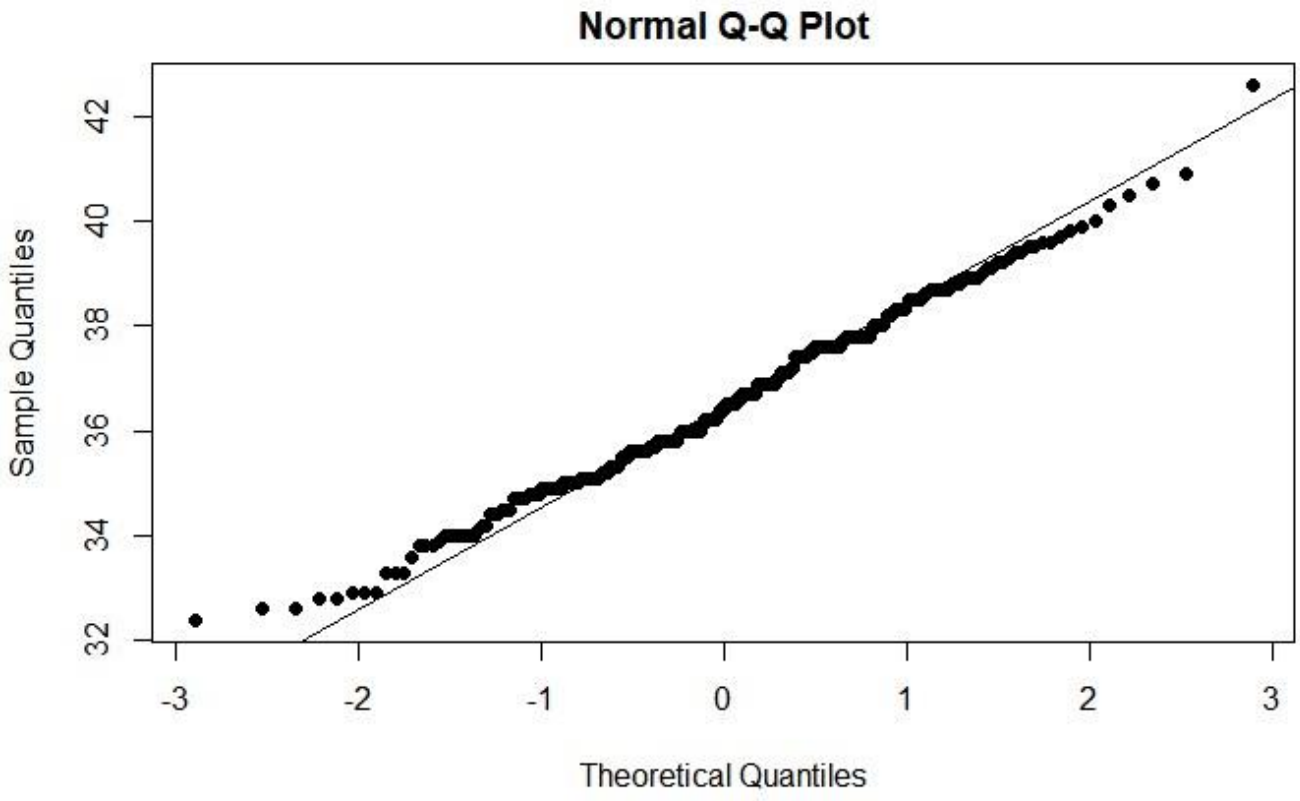
```

    Shapiro-Wilk normality test

data:  bia_female$bia.di
W = 0.99271, p-value = 0.2318
```

Hide

```
# as displayed in the output the p-value = 0.2318 which is greater than 0.05
# The data is approximately normal distributed with null hypothesis accepted
# QQ plots
qqnorm(bia_female$bia.di,pch=19)
qqline(bia_female$bia.di)
```

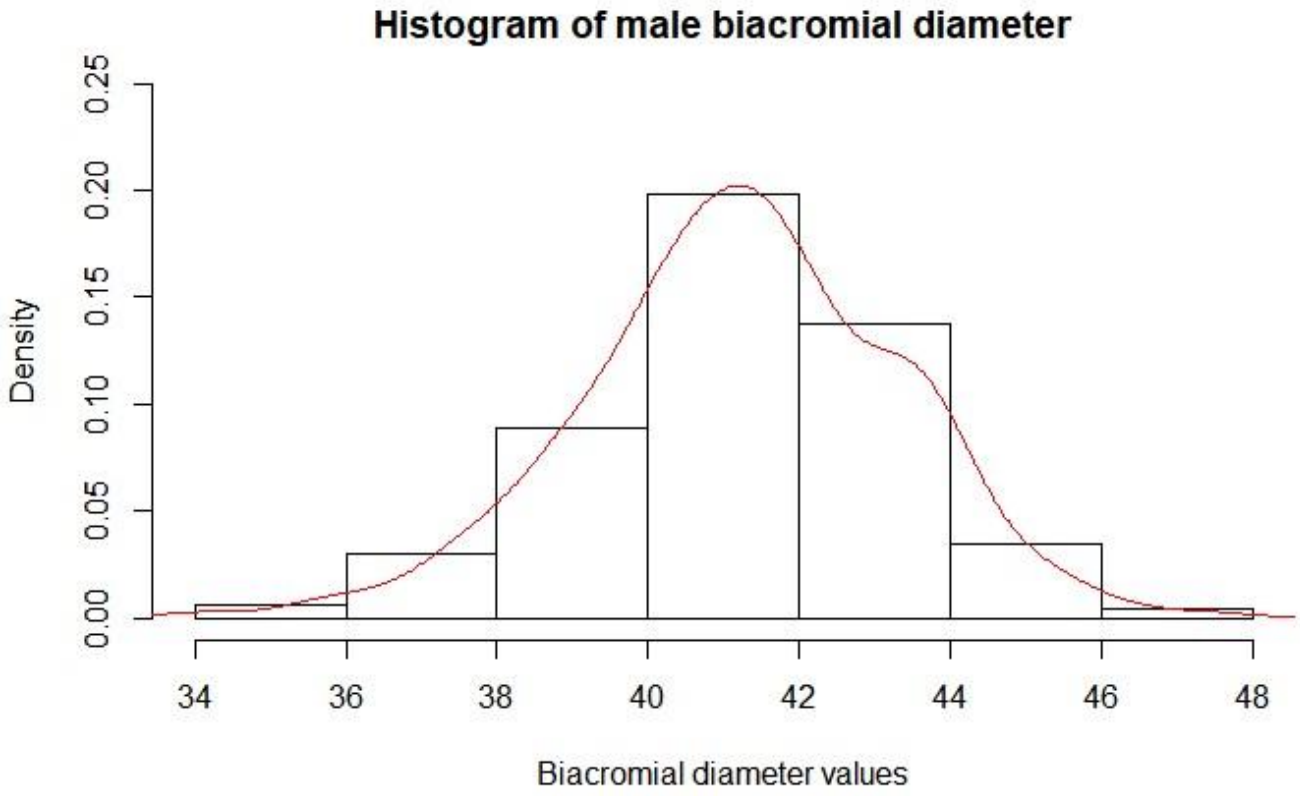


Hide

```
# The QQ plot the lines displays a perfect normal distribution
# So the data values should lie as close to the line as possible.
# Hence the female biiliac diameter fits the normal distribution
```

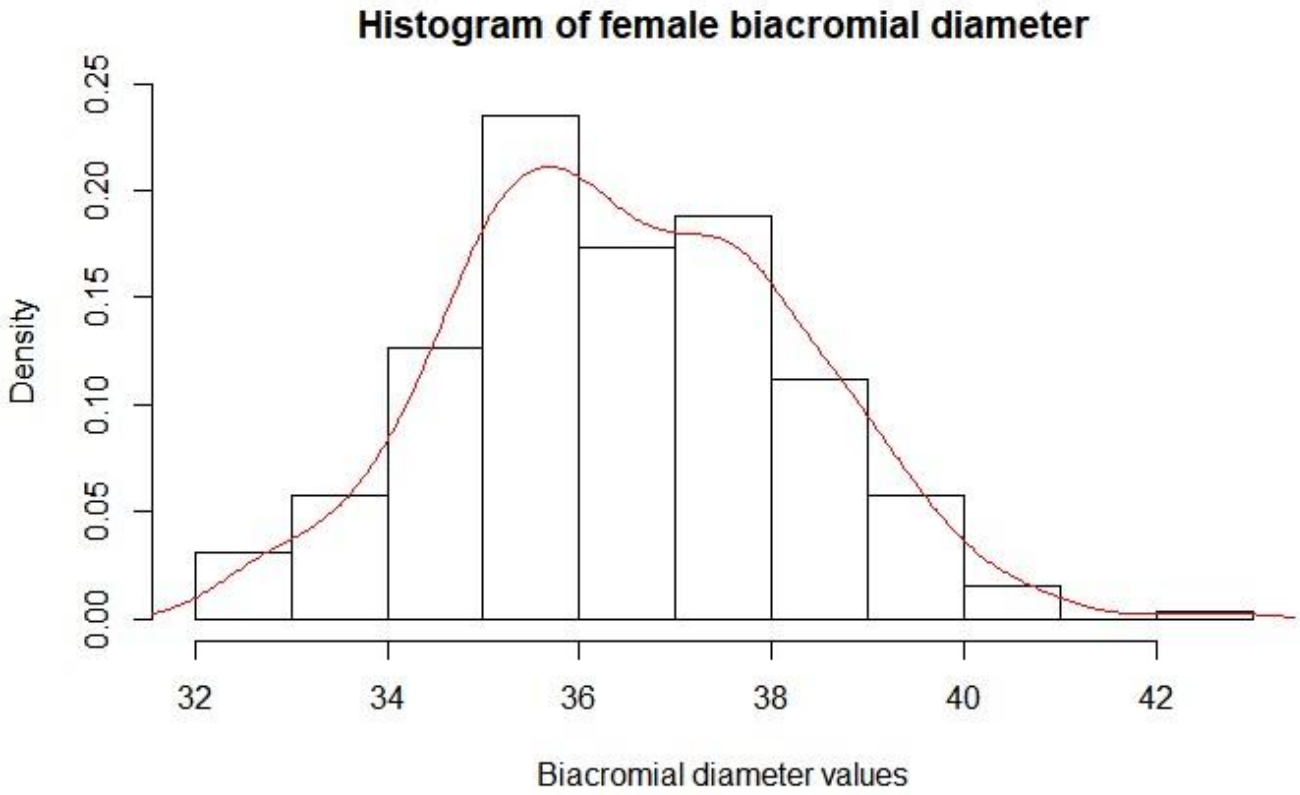
Hide

```
# the Historam with density curve oervlay for male biiliac diameter
# the density curve depcits the empirical values which is approximately a bell shaped curve
bia_male$bia.di %>%  hist(probability=, main="Histogram of male biacromial diameter",
                          xlab="Biacromial diameter values",ylim=c(0,0.25))
lines(density(bia_male$bia.di),col=2)
```



Hide

```
# the Historam with density curve oervlay for female biiliac diameter
# the density curve depcits the empirical values which is approximately a bell shaped curve
bia_female$bia.di %>%  hist(probability=, main="Histogram of female biacromial diameter",
                            xlab="Biacromial diameter values",ylim=c(0,0.25))
lines(density(bia_female$bia.di),col=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.

After the analysis of the shapiro wilk test and qq plots we can conclude that the biacromial diameter approximately fits the normal distribution seperately in men and women.The two histogram plots with density curve prove that empirical distribution of biacromial diamter approximately fits the normal distribution seperately in men and women as per data provided.