

MATH1324 Assignment 1

Code ▼

Modeling Body Measurements

Student Details

Junaid Ahmed Syed (s3731300)

Problem Statement

Investigation for normality assumption of the variable (thi.gi) and to determine if it fits the normal distribution separately in men and women.

Approach for the above problem will be: *Calculation of descriptive statistics of Thigh girth for both Male and Female. Creating Histogram. Generation of theoretical normal distribution.* Examination between empirical and theoretical values, Which is used to verify wheatear the data fits in normal distribution or not.

Load Packages

Hide

```
# This is a chunk where you can load the necessary packages required to reproduce the report
library(readxl) #Useful for importing data to R
library(magrittr) #Useful for using pipe operator
library(dplyr) #For manipulation of data
```

Data

Hide

```
# This is a chunk for your Data section.
body_ms <- read_csv("bdims.csv")
```

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

Hide

```
body_ms$sex <- body_ms$sex %>% factor(levels = c(1,0), labels =
c("Male", "Female"))
levels(body_ms$sex)
```

```
[1] "Male"    "Female"
```

Summary Statistics

Calculation of descriptive statistics of Thigh Grith

[Hide](#)

```
body_ms %>% group_by(sex) %>% summarise(Mean = mean(thi.gi),
                                         Median = median(thi.gi),
                                         SD = sd(thi.gi),
                                         First_Quartile = quantile(thi.gi, probs = 0.2
5, na.rm = TRUE),
                                         Third_Quartile = quantile(thi.gi, probs = 0.7
5, na.rm = TRUE),
                                         Interquartile_Range = IQR(thi.gi, na.rm = TRU
E), Minimum = min(thi.gi),
                                         Maximum = max(thi.gi))
```

Distribution Fitting

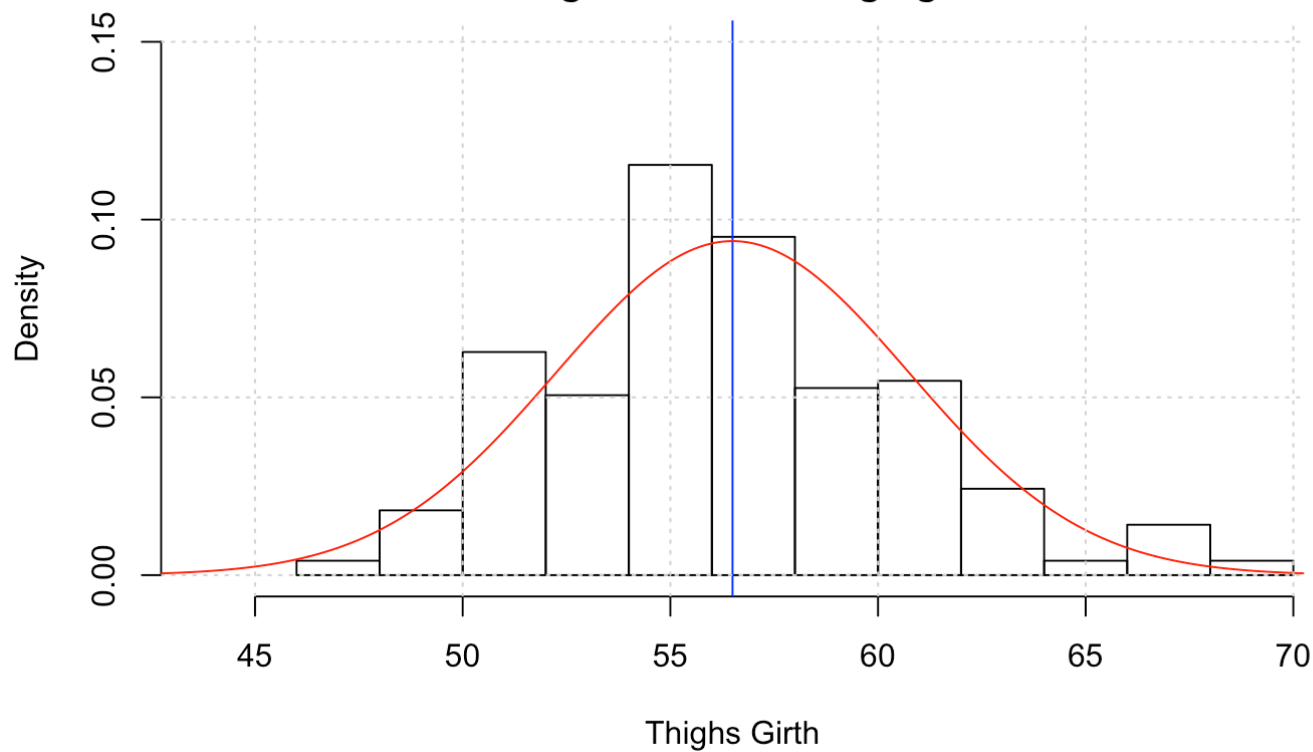
[Hide](#)

```
# This is a chunk for your Distribution Fitting section.
# Distribution Fitting section
#Filtering the dataframe based on sex
female = body_ms %>% filter(sex == 'Female')
male = body_ms %>% filter(sex == 'Male')
# Calculation of mean and Standard deviation for Female
femalemean = mean(female$thi.gi)
femalesd = sd(female$thi.gi)
# Calculation of mean and standard deviation for Male
malemean = mean(male$thi.gi)
malesd = sd(male$thi.gi)
# Generation of the theoretical normal distribution for Female
x_female = seq(-4, 4, length=1000) * femalesd + femalemean
femalenorm = dnorm(x_female, mean = femalemean, sd = femalesd)
# Generating of the theoretical normal distribution for Male
x_male = seq(-4, 4, length=1000) * malesd + malemean
malenorm = dnorm(x_male, mean = malemean, sd = malesd)
# Plotting histogram with normal distribution overlay
hist(male$thi.gi, ylim = c(0, 0.15), xlim = c(malemean - 3 * malesd, malemean + 3 * m
alesd), ylab = 'Density', xlab = 'Thighs Girth', main = 'Histogram of male thigh girt
h', probability = TRUE)
abline(v = malemean, col = 'blue')
```

[Hide](#)

```
lines(x_male, malenorm, col = 'red')
grid()
```

Histogram of male thigh girth

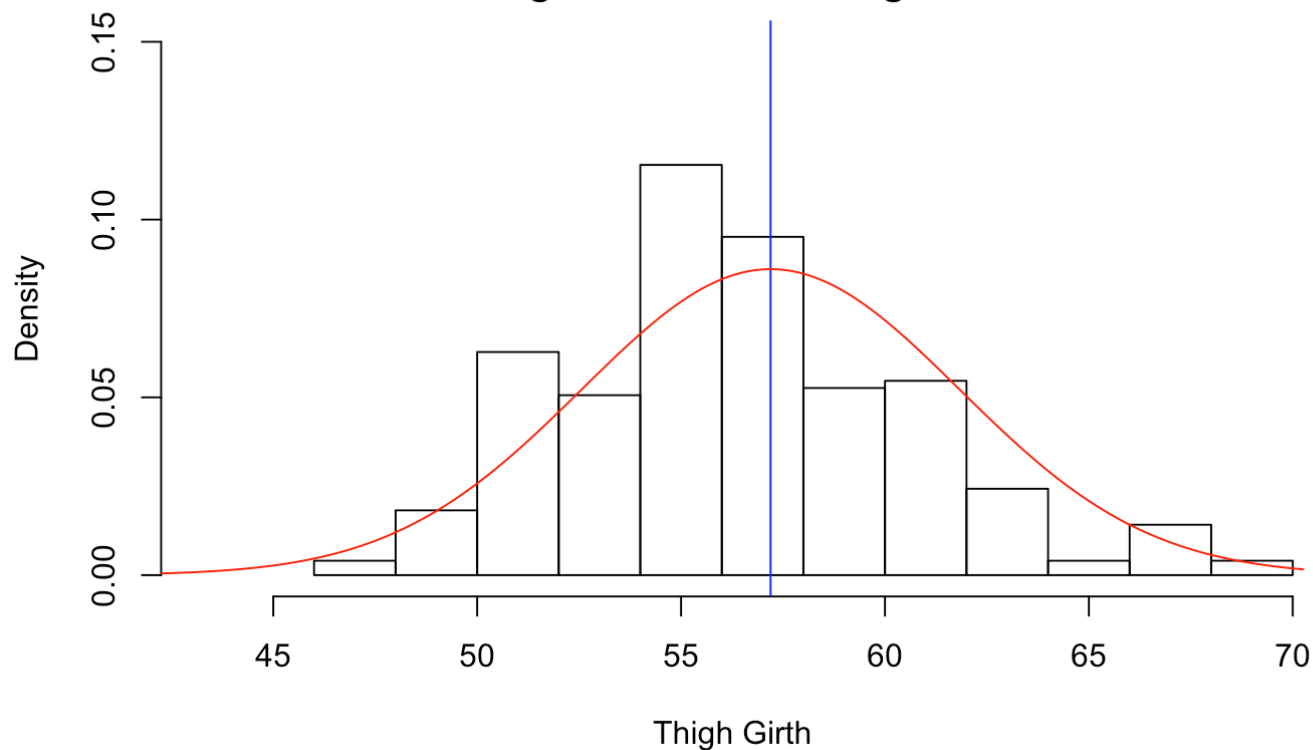
[Hide](#)

```
hist(male$thi.gi, ylim = c(0, 0.15), xlim = c(femalemean - 3 * femalesd, malemean +  
                                              3 * malesd), ylab = 'Density', xlab =  
      'Thigh Girth', main = 'Histogram of Female Thigh Girth', probability = TRUE)  
abline(v = femalemean, col = 'blue')
```

[Hide](#)

```
lines(x_female, femalenorm, col = 'red')
```

Histogram of Female Thigh Girth

[Hide](#)

```
qnorm(p = 0.25, mean = malemean, sd = malesd)
```

```
[1] 53.63364
```

[Hide](#)

```
qnorm(p = 0.75, mean = malemean, sd = malesd)
```

```
[1] 59.36231
```

[Hide](#)

```
qnorm(p = 0.25, mean = femalemean, sd = femalesd)
```

```
[1] 54.06883
```

[Hide](#)

```
qnorm(p = 0.75, mean = femalemean, sd = femalesd)
```

```
[1] 60.32271
```

Interpretation

Male Thigh Girth:-

It can be seen from the above histogram that it has a bell-shaped curve as the empirical data fits in the normal distribution. From the descriptive statistics the values of mean and median obtained are 56.49 and 56.0, It has a difference of 0.4 which is less than 0.5. *The First and Third Quartile from Theoretical normal distribution are 53.6,59.36 and from the empirical data are 53.7,59.15. Their difference doesn't surpass 0.5.

Female Thigh Girth:-

It can be seen from the above histogram that it doesn't have a bell-shaped curve as the empirical data doesn't fits in the normal distribution. From the descriptive statistics the values of mean and median obtained are 57.19 and 56.4, It has a difference greater than 0.5. *The First and Third Quartile from Theoretical normal distribution are 54.06,60.32 and from the empirical data are 53.7,59.80. There difference exceeded 0.5.

Conclusion:-

The conclusion I draw upon observing the above values is, That the thigh girth measurements for the male is normally distributed whereas for the female it's not.