# MATH1324 Assignment 1

Code

## Modeling Body Measurements

# STUDENT DETAILS

AKSHAT VIJAYVARGIA (s3826627)

# PROBLEM STATEMENT

This examination looks to compare the hip girth of male and female and to check whether this dataset has null hypothesis to fit in the normal distribution or the alternative hypothesis that makes the dataset not to fit in normal distribution.

# LOAD PACKAGES

Hide

```
library(readr)
library(dplyr)
library(magrittr)
```

# DATA

This is the step, where we will import the database in Rstudio using readr function, factoring the sex column and labelling it and creating separate table for male as well as for female.

Hide

```


# IMPORTING DATASET
library(readr)
EXCEL <- read_csv("C:/Users/61422/OneDrive/Desktop/bdims.csv - bdims.csv")

# DEFINING THE SEX VARIABLE
EXCEL$sex <- factor(EXCEL$sex, levels=c(1 ,0), labels=c("MALE","FEMALE"))

# SEPARATING MALE AND FEMALE
EXCEL_FEMALE <- EXCEL %>% filter(sex=='FEMALE')
EXCEL_MALE <- EXCEL %>% filter(sex=='MALE')
```

# SUMMARY STATISTICS

In this step, we will calculate the descriptive statistics separately for female and male. The coding is below-

Hide

```
# CODING TO CALCULATE HIP GIRTH IN FEMALES
EXCEL_FEMALE$hip.gi %>% summary()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  78.80   90.75   94.95   95.65   99.50  128.30 
```

Hide

```
EXCEL_FEMALE$hip.gi %>% sd()
```

```
[1] 6.940728
```

Hide

```
EXCEL_FEMALE$hip.gi %>% IQR()
```

```
[1] 8.75
```

Hide

```
EXCEL_FEMALE$hip.gi %>% range()
```

```
[1]  78.8 128.3
```

Hide

```
# CODING TO CALCULATE HIP GIRTH IN MALES
EXCEL_MALE$hip.gi %>% summary()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  81.50   93.25   97.40   97.76  101.55  118.70 
```

Hide

```
EXCEL_MALE$hip.gi %>% sd()
```

```
[1] 6.228043
```

Hide

```
EXCEL_MALE$hip.gi %>% IQR()
```

```
[1] 8.3
```

Hide

```
EXCEL_MALE$hip.gi %>% range()
```

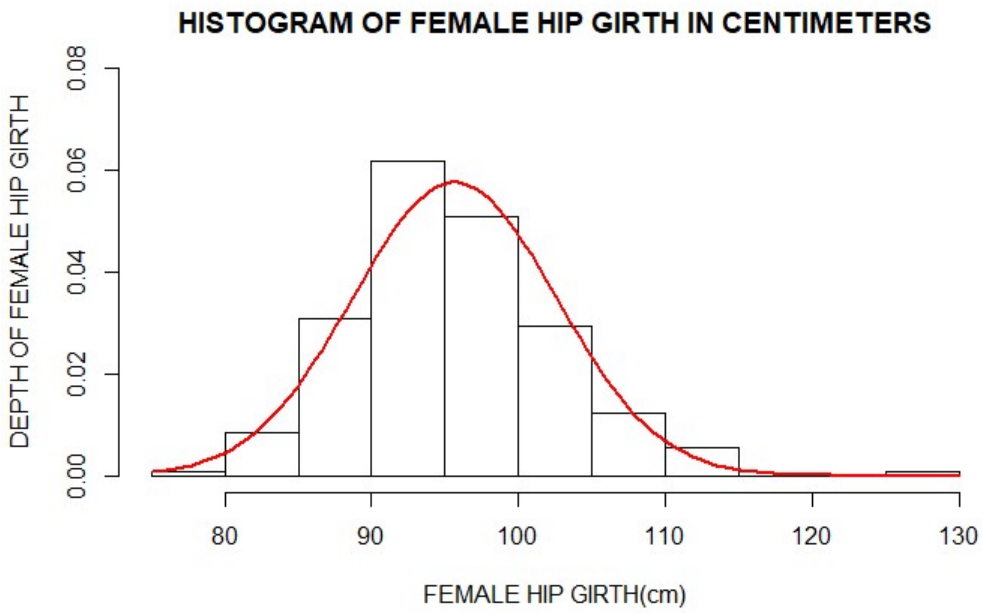
```
[1]  81.5 118.7
```

# DISTRIBUTION FITTING

And this is the main step, where we will differentiate the measurement of hip girth of men and women by plotting two histograms one with breaks 10 and 50 each for both the sex.

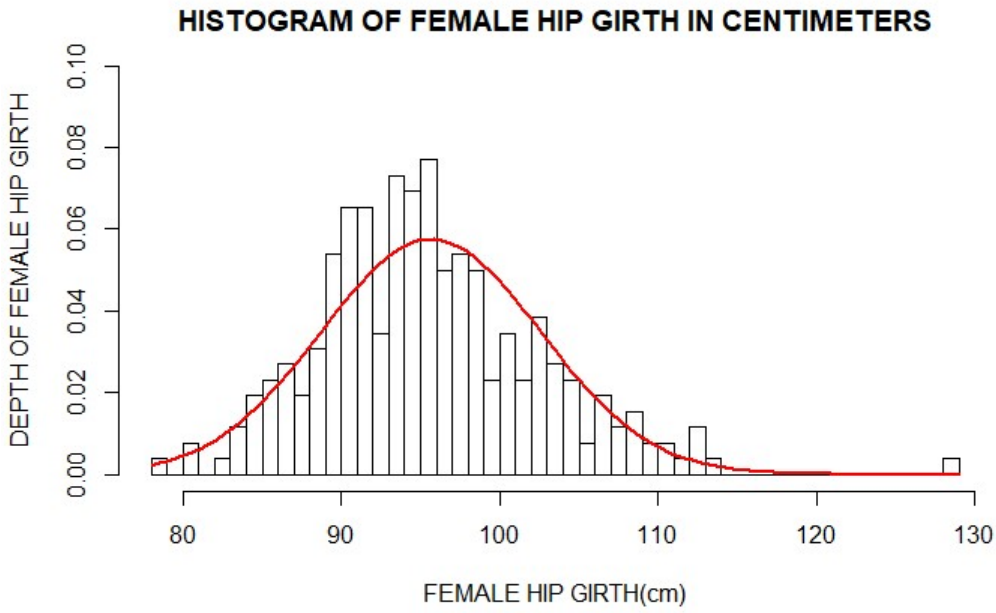
Hide

```
# CODING FOR FEMALE HIP GIRTH(WITH BREAKS=10)
FH <- EXCEL_FEMALE$hip.gi
MFH <- mean(FH)
STDF <- sqrt(var(FH))
hist(FH, prob = TRUE, breaks = 10, xlab = "FEMALE HIP GIRTH(cm)", ylab = "DEPTH OF FEMALE HIP GIRTH", main = "HISTOGRAM OF
FEMALE HIP GIRTH IN CENTIMETERS", ylim = c(0,.08))
curve(dnorm(x, mean=MFH, sd=STDF),
      col="red", lwd=2, add=TRUE, yaxt="n")
```



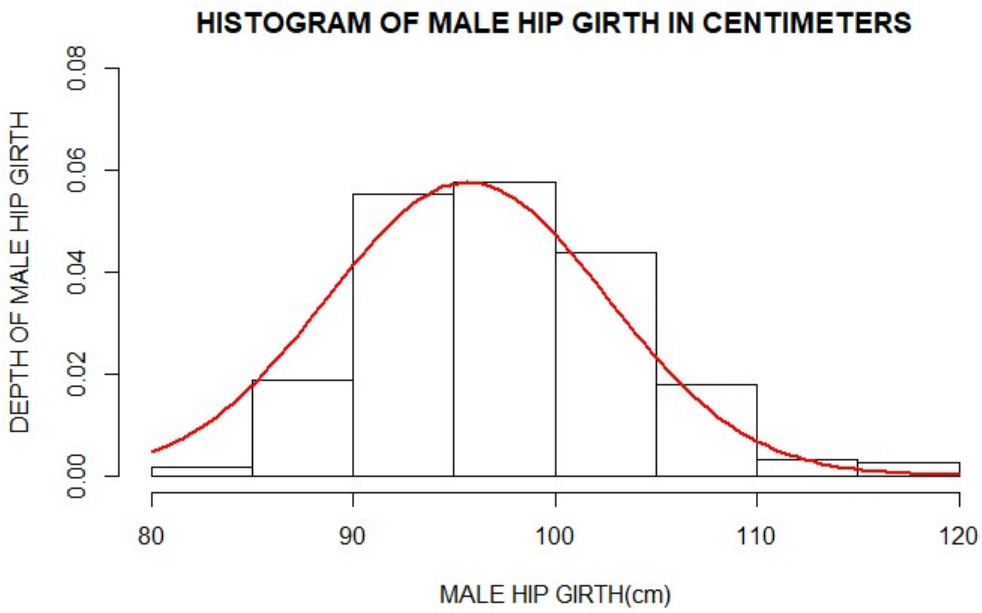

Hide

```
# CODING FOR FEMALE HIP GIRTH(WITH BREAKS=50)
FH <- EXCEL_FEMALE$hip.gi
MFH <- mean(FH)
STDF <- sqrt(var(FH))
hist(FH, prob = TRUE, breaks = 50, xlab = "FEMALE HIP GIRTH(cm)", ylab = "DEPTH OF FEMALE HIP GIRTH", main = "HISTOGRAM OF
FEMALE HIP GIRTH IN CENTIMETERS", ylim = c(0,.1))
curve(dnorm(x, mean=MFH, sd=STDF),
      col="red", lwd=2, add=TRUE, yaxt="n")
```



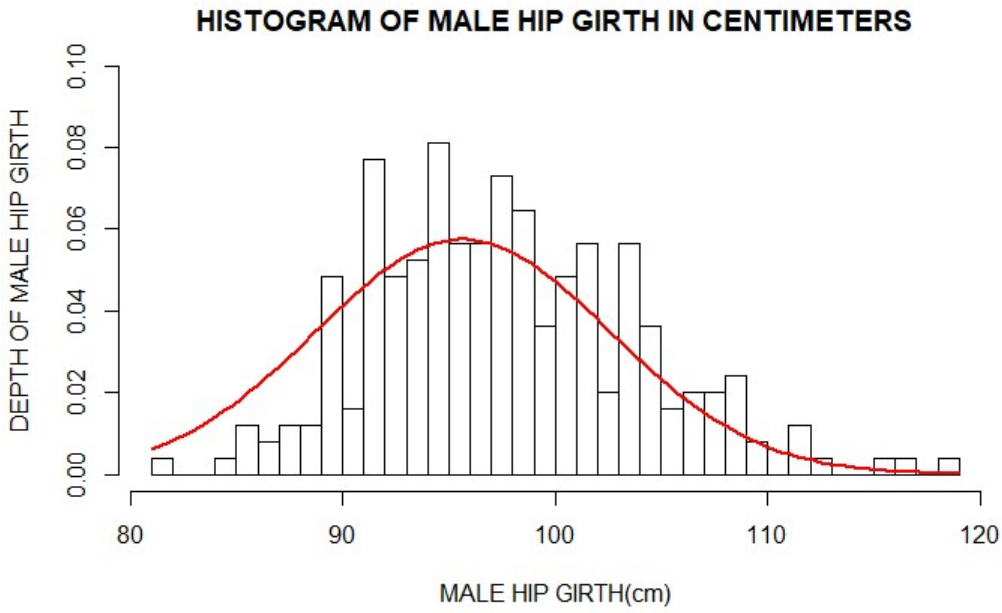

Hide

```
# CODING FOR MALE HIP GIRTH(WITH BREAKS=10)
MH <- EXCEL_MALE$hip.gi
MMH <- mean(FH)
STDF <- sqrt(var(FH))
hist(MH, prob = TRUE, breaks = 10, xlab = "MALE HIP GIRTH(cm)", ylab = "DEPTH OF MALE HIP GIRTH", main = "HISTOGRAM OF MAL
E HIP GIRTH IN CENTIMETERS", ylim = c(0,.08))
curve(dnorm(x, mean=MMH, sd=STDF),
      col="red", lwd=2, add=TRUE, yaxt="n")
```




Hide

```
# CODING FOR MALE HIP GIRTH(WITH BREAKS=50)
MH <- EXCEL_MALE$hip.gi
MMH <- mean(FH)
STDF <- sqrt(var(FH))
hist(MH, prob = TRUE, breaks = 50, xlab = "MALE HIP GIRTH(cm)", ylab = "DEPTH OF MALE HIP GIRTH", main = "HISTOGRAM OF MAL
E HIP GIRTH IN CENTIMETERS", ylim = c(0,.1))
curve(dnorm(x, mean=MMH, sd=STDF),
      col="red", lwd=2, add=TRUE, yaxt="n")
```




# INTERPRETATION

Now, lets check and come to the solution of the problem statement.

1. MALE HIP GIRTH- If we look at the histogram with breaks=10, we can notice that our data somewhat goes with the normal distribution and the peak is relatively around our mean, that is 97.And if look at the histogram with 50 breaks, the larger data is above our mean. With the visual plotting, its still unclear whether that data fits the normal distribution or not.

To have a clear answer, we will use Shapiro test- For Males: H0=Male Hips Girth fits the normal distribution, HA=Male Hips Girth does not fit the normal distribution.

Hide

```
shapiro.test(EXCEL_MALE$hip.gi)
```

```
    Shapiro-Wilk normality test

data:  EXCEL_MALE$hip.gi
W = 0.98791, p-value = 0.03579
```

The result of Shapiro test is p-value<0.05(the p-value is less than 0.05), this means that the there is huge contrast between our data and the normal distribution. Therefore, we can reject the null hypothesis, and now we can conclude that male hip girth does not fit the normal distribtion, with 95% confidence.

2. FEMALE HIP GIRTH- Now lets see on the histogram of female hip girth with 10 breaks, there is a slight resemblance between our data and the normal distribution and again the peak of the curve is near to our mean, that is 95. And if we inspect the histogram with breaks=50, we can say that broadly the data is below the normal distribution except some high lines.

To have a concrete answer, we have conduct Shapiro test- For Females: H0=Female Hips Girth fits the normal distribution, HA=Female Hips Girth does not fit the normal distribution.

Hide

```
shapiro.test(EXCEL_FEMALE$hip.gi)
```

```
    Shapiro-Wilk normality test

data:  EXCEL_FEMALE$hip.gi
W = 0.97522, p-value = 0.0001692
```

The test result of shapiro test is p-value<0.05(the p-value is less than 0.05), this indicates that there is enormous deviation between the data of female hip girth and the normal distribution.Now, it can be concluded, that the null hypothesis is dropped and the measurement of female hip girth does not fit the normal distribution.

OVERALL-

With the above results, it is clear that neither male hip girth nor female hip girth fit in a normal distribution, It may be possible that this data comes under 5% error category. In the end, I would say that the researchers should gather some more information or samples as this is worth exploring.

# REFERENCES

1)https://www.quora.com/How-do-I-know-my-data-fits-a-normal-distribution (https://www.quora.com/How-do-I-know-my-data-fits-a-normal-distribution) 2)https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/shapiro.test (https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/shapiro.test)