MATH1324 Introduction to Statistics Assignment 2

Code

Phil Steinke s3725547

1. Problem Statement:

- This paper examins the wrist girth of male and females from the dataset Heinz G, Peterson LJ, Johnson RW, Kerk CJ. 2003. Exploring Relationships in Body Dimensions. Journal of Statistics Education 11(2). (https://ww2.amstat.org/publications/jse/v11n2/datasets.heinz.html).
- The examined the factors are `wri.di` wrist girth (in cm) and `sex` (male or female)
- I compare the differences in body dimensions between male and females against the normal disntribution of the population of the dataset.

```
suppressPackageStartupMessages(library(readr))
suppressPackageStartupMessages(library(base))
suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library(readxl))
suppressPackageStartupMessages(library("mosaic"))
library("gridExtra"))
suppressPackageStartupMessages(library("latticeExtra"))
```

3. Data: Import the body measurements data and tidy it up

```
body_measurements <- read_excel("data/bdims.csv (1).xlsx", sheet = "bdims.csv") %>% as.d
ata.frame()
wrist_measurements <- select(body_measurements, wri.gi, sex) %>% as.data.frame()
colnames(wrist_measurements) <- c("wrist.girth", "sex")
wrist_measurements <- mutate(wrist_measurements, sex=factor(sex, labels=c("female", "mal
e")))
wrist_measurements %>% head()
```

	wrist.girth <dbl>	sex <fctr>
1	16.5	male
2	17.0	male
3	16.9	male
4	16.6	male
5	18.0	male
6	16.9	male

6 rows

4. Summary Statistics: Calculate descriptive statistics

Summary of wrist measurements for all sexes

```
wrist_summary <- wrist_measurements$wrist.girth
summary(wrist_summary)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   13.0    15.0    16.1    16.1    17.1    19.6 
```

```
summary_mean <- mean(wrist_summary) %>% print()
```

```
[1] 16.09744
```

```
summary_sd <- sd(wrist_summary) %>% print()
```

```
[1] 1.380931
```

```
IQR(wrist_summary) # Interquartile Range
```

```
[1] 2.1
```

Male Summary Statistics

```
male <- subset(wrist_measurements, sex == "male")$wrist.girth
summary(male)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  14.60   16.50   17.10   17.19   17.85   19.60 
```

```
male_mean <- mean(male) %>% print()
```

```
[1] 17.19028
```

```
male_sd <- sd(male) %>% print()
```

```
[1] 0.9079967
```

```
IQR(male) # Interquartile Range
```

```
[1] 1.35
```

Female Summary Statistics

```
female <- subset(wrist_measurements, sex == "female")$wrist.girth
summary(female)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  13.00   14.50   15.00   15.06   15.60   18.20 
```

```
female_mean <- mean(female) %>% print()
```

```
[1] 15.05923
```

```
female_sd <- sd(female) %>% print()
```

```
[1] 0.8494101
```

```
IQR(female) # Interquartile Range
```

```
[1] 1.1
```

5. Distribution Fitting:

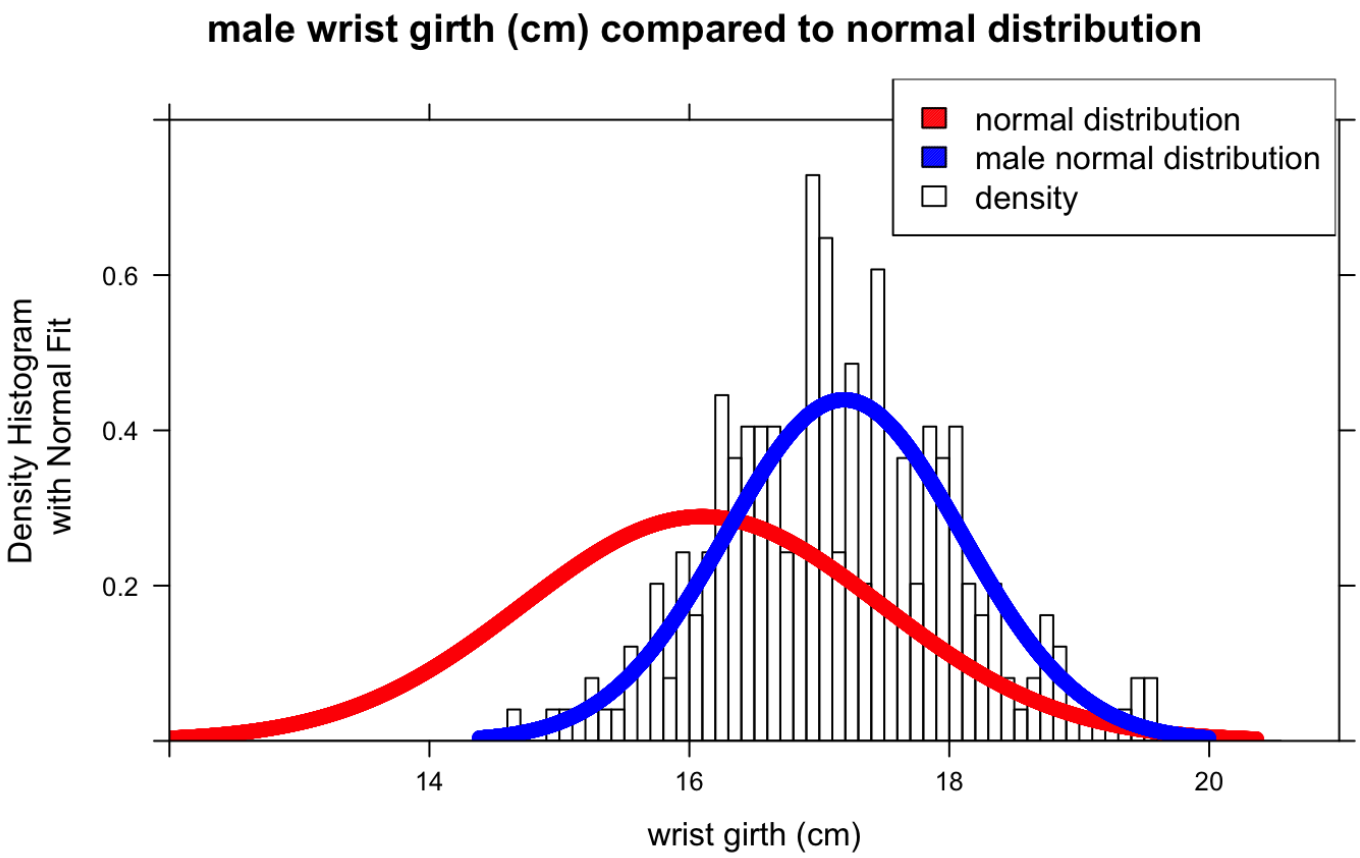
For males

```
plot.new()
```

```
histogram(
  x = male,
  type="density",
  main = "male wrist girth (cm) compared to normal distribution",  col=rgb(0,0,0,alpha=
0.0), xlim=c(12,21), ylim = c(0,0.80),
  breaks = 40,
  xlab = "wrist girth (cm)",
  ylab = "Density Histogram\n with Normal Fit")
```

```
plotDist(
  "norm",
  mean = summary_mean,
  sd = summary_sd,
  kind= "density",
  type = "b",
  col = "red",
  lwd=2,
  add=TRUE,
  xlab = "wrist girth (cm)")
```

```
plotDist(
  "norm",
  mean = male_mean,
  sd = male_sd,
  kind= "density",
  type = "b",
  col = "blue",
  lwd=2,
  add=TRUE,
  xlab = "wrist girth (cm)")
legend("topright",
       c("normal distribution", "male normal distribution", "density"),
       density = c(100, 100, 100),
       fill=c("red","blue","white"))
```



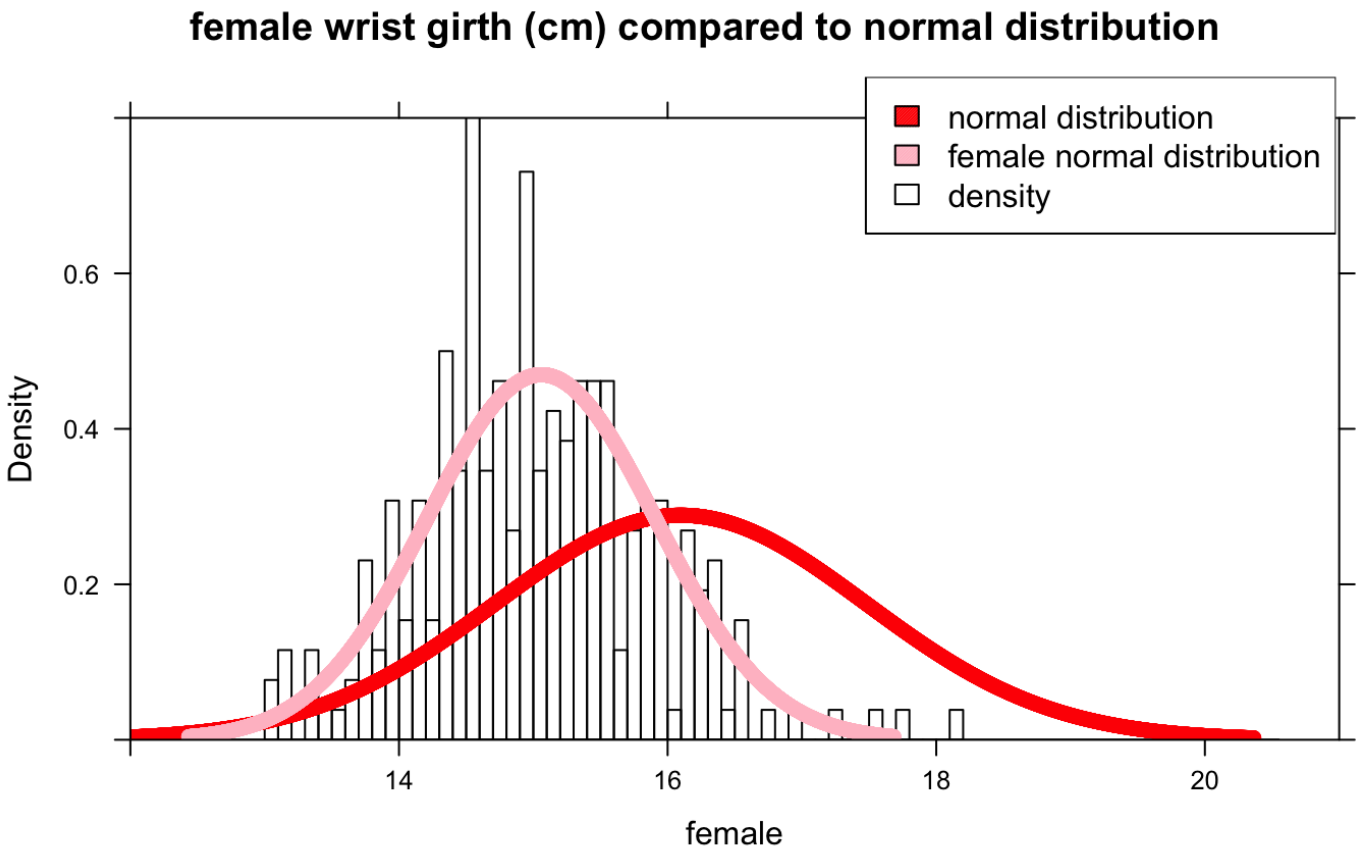
For females

```
plot.new()
```

```
histogram(
  x = female,
  type="density",
  main = "female wrist girth (cm) compared to normal distribution",
  col=rgb(0,0,0,alpha=0.0), xlim=c(12,21), ylim = c(0,0.80),
  breaks = 40)
```

```
plotDist(
  "norm",
  mean = summary_mean,
  sd = summary_sd,
  kind= "density",
  type = "b",
  col = "red",
  lwd=2,
  add=TRUE)
```

```
plotDist(
  "norm",
  mean = female_mean,
  sd = female_sd,
  kind= "density",
  type = "b",
  col = "pink",
  lwd=2,
  add=TRUE)
legend("topright",
       c("normal distribution", "female normal distribution", "density"),
       density = c(100, 100, 100),
       fill=c("red","pink","white"))
```



Interpretation:

What insight has been gained from the investigation?

- The mean of males is `~17.2cm` and females are `~15cm`, and the median of males is `~17.1cm` and females is `15cm`. The girth of male wrists are larger than female wrists by an average of `2cm`

Discuss the extent to how your theoretical normal distribution fits the empirical data

- The data fits a similar distribution
- There are very few female cases above `17cm`
- Male, female and both sexes fit a normal bell curve

Make recommendations regarding the modelling of this body measurement

- Further analysis to check against other distributions
- Check against the fit with greater and fewer bins