MATH1324 Assignment 1

Code

Modeling Body Measurements

Student Details

Student Abdul Raheman Nasir Gazi (s3705448)

Problem Statement

The purpose of this assignment is to determine if che.di(Respondent's chest diameter in centimeters) measurement of bdims.csv datafile fits it's normal distribution. The aim is to check how well the selected measurement is approximated by normal distribution. This will be done by creating histogram for empirical data with a bell curve represnting normal distribution for the selected measurement for male and female differently. The normality will get checked and it's fitting will be interpreted based on the visualization provided.

Packages & Data

Hide

```
# This is a chunk where you can load the necessary packages required to reproduce the report
library(dplyr)
library(magrittr)
library(readr)
# This is a chunk for your Data section.
# read_csv was used to import the dataset
# factor() was used to label 0,1 with Male and Female
# subsetting was done for plotting purposes so that accurate results are provided for Male and Fe
male chest diameter.
body <- read_csv("/home/literalmoniker/Downloads/bdims.csv")
```

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

Hide

```
body$sex <- factor(body$sex, levels = c(0,1), labels = c("Female","Male"), ordered = TRUE)
bodymale <- body %>% filter(body$sex == "Male")
bodyfemale <- body %>% filter(body$sex == "Female")
body
```

bia.di <dbl>	bii.di <dbl>	bit.di <dbl>	che.de <dbl>	che.di <dbl>	elb.di <dbl>	wri.di <dbl>	kne.di <dbl>	ank.di <dbl>	sho.gi <dbl>
42.9	26.0	31.5	17.7	28.0	13.1	10.4	18.8	14.1	106.2
43.7	28.5	33.5	16.9	30.8	14.0	11.8	20.6	15.1	110.5
40.1	28.2	33.3	20.9	31.7	13.9	10.9	19.7	14.1	115.1
44.3	29.9	34.0	18.4	28.2	13.9	11.2	20.9	15.0	104.5
42.5	29.9	34.0	21.5	29.4	15.2	11.6	20.7	14.9	107.5
43.3	27.0	31.5	19.6	31.3	14.0	11.5	18.8	13.9	119.8

bia.di <dbl>	bii.di <dbl>	bit.di <dbl>	che.de <dbl>	che.di <dbl>	elb.di <dbl>	wri.di <dbl>	kne.di <dbl>	ank.di <dbl>	sho.gi <dbl>
43.5	30.0	34.0	21.9	31.7	16.1	12.5	20.8	15.6	123.5
44.4	29.8	33.2	21.8	28.8	15.1	11.9	21.0	14.6	120.4
43.5	26.5	32.1	15.5	27.5	14.1	11.2	18.9	13.2	111.0
42.0	28.0	34.0	22.5	28.0	15.6	12.0	21.1	15.0	119.5

1-10 of 507 rows | 1-10 of 25 columns Previous **1** 2 3 4 5 6 … 51 Next

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.
#group_by was used to find the descriptive statics of male and female differently
# na.rm was used to remove any na values and is.na was used inside sum() to find number of na val
ues for both male and female respectively
body %>% group_by(body$sex) %>% summarise(Min = min(che.di,na.rm = TRUE),
                                          Q1 = quantile(che.di,probs = .25,na.rm = TRUE),
                                          Median = median(che.di, na.rm = 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),
                                          IQR(che.di),
                                          n = n(),
                                          Missing = sum(is.na(che.di)))
```

body$sex <ord>	Min <dbl>	Q1 <dbl>	Median <dbl>	Q3 <dbl>	Max <dbl>	Mean <dbl>	SD <dbl>	IQR(che.di) <dbl>	n <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

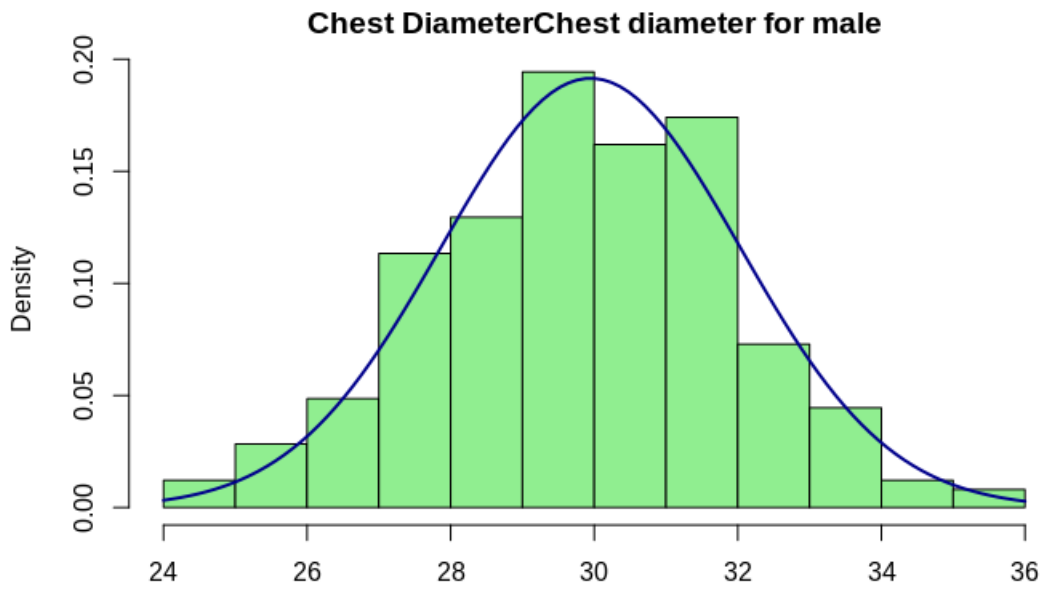
2 rows | 1-10 of 11 columns

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.

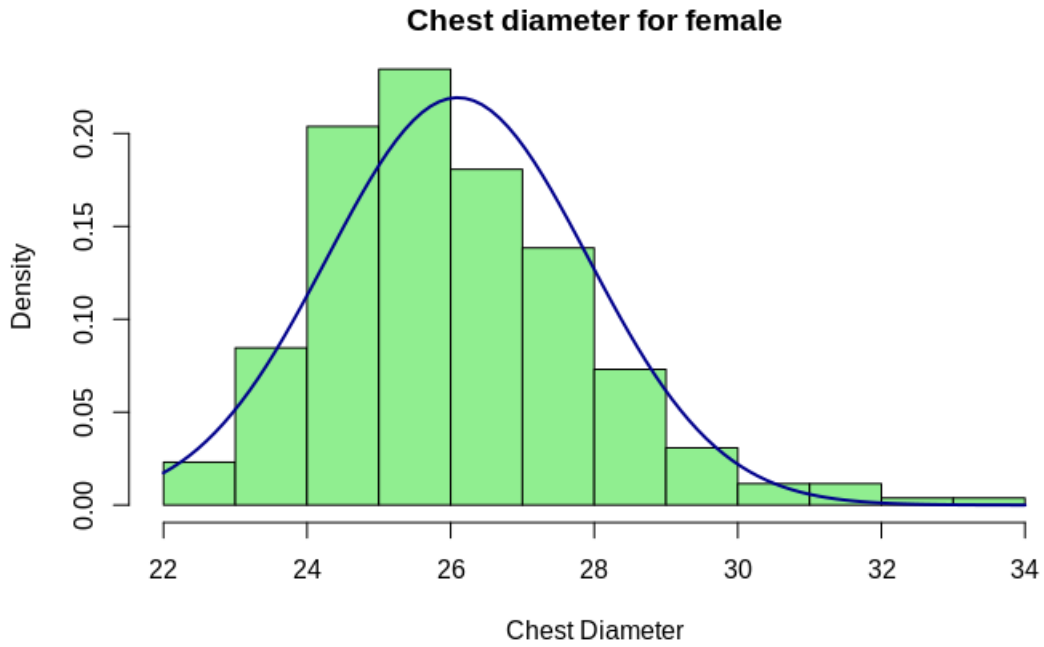
Hide

```
# This is a chunk for your Distribution Fitting section.
hist(bodymale$che.di, freq=FALSE, xlab="", main="Chest DiameterChest diameter for male", col="lig
htgreen")
curve(dnorm(x, mean=mean(bodymale$che.di), sd=sd(bodymale$che.di)), add=TRUE, col="darkblue", lwd
=2)
```



Hide

```
hist(bodyfemale$che.di, freq=FALSE, xlab="Chest Diameter", main="Chest diameter for female", col="lightgreen")
curve(dnorm(x, mean=mean(bodyfemale$che.di), sd=sd(bodyfemale$che.di)), add=TRUE, col="darkblue", lwd=2)
```



Interpretation

As per the histogram with curve diagrams above for male you can see that the bell curve and the histogram fits which leads us to the conclusion that empirical data of the chest diameter fits with it's normal distribution. For the chest diameter for female the histogram also apears to be in shape of it's bell curve which indicates that the empirical data fits it's normal distribution.

Both the results satisfy the problem statement provided at the start of the assignment. Chest diameter's normal distribution for both male and female fits their respective empirical data provided.