MODULE 6 R PRACTICE

Week 6

Abstract

In this assignment I will use regression modeling and subseting for some data

Mohammad Hossein Movahedi

Movahedi.m@northeastern.edu

The dataset I use is provided by our instructor and then I will create dummy variables to subset the dataset. Then rerun your regression line for dependent variable.

First Of all, I import and install required libraries

```
 print('Mohammad Hossein Movahedi')

#install and load packages

install.packages('FSA')

install.packages('FSAdata')

install.packages('magrittr')

install.packages('dplyr')

install.packages('tidyr')

install.packages('plyr')

install.packages('tidyverse')

install.packages('outliers')

install.packages('ggplot2')

install.packages('lubridate')

install.packages('corrplot')


library(ggplot2)

library(outliers)

library(FSA)

library(FSAdata)

library(magrittr)

library(dplyr)

library(tidyr)

library(dplyr)

library(tidyverse)

library(scales)

library(lubridate)

library(corrplot)
```

then I import the dataset

```
 #loading data

#loading data

data <-
read.csv("https://gist.githubusercontent.com/nstokoe/7d4717e96c21b8ad04ec91f361b000cb/raw/bf95a2
e30fceb9f2ae990eac8379fc7d844a0196/weight-height.csv")

head(data)
```

The results of head command is shown below

```
 > head(data)

  Gender   Height    Weight

1   Male 73.84702 241.8936

2   Male 68.78190 162.3105

3   Male 74.11011 212.7409

4   Male 71.73098 220.0425

5   Male 69.88180 206.3498

6   Male 67.25302 152.2122
```

now I assign 1 to males and 0 to females and then change the column name to Maleness.
This column will be my dummy variable

```
 #creating dummy variable

data$Gender[data$Gender == "Male"] = 1

data$Gender[data$Gender == "Female"] = 0

colnames(data)[1]="Maleness"
```

now we can procced with regression modeling

```
 #regression

reg <-lm(Height~Weight+Maleness ,data = data)

summary(reg)
```

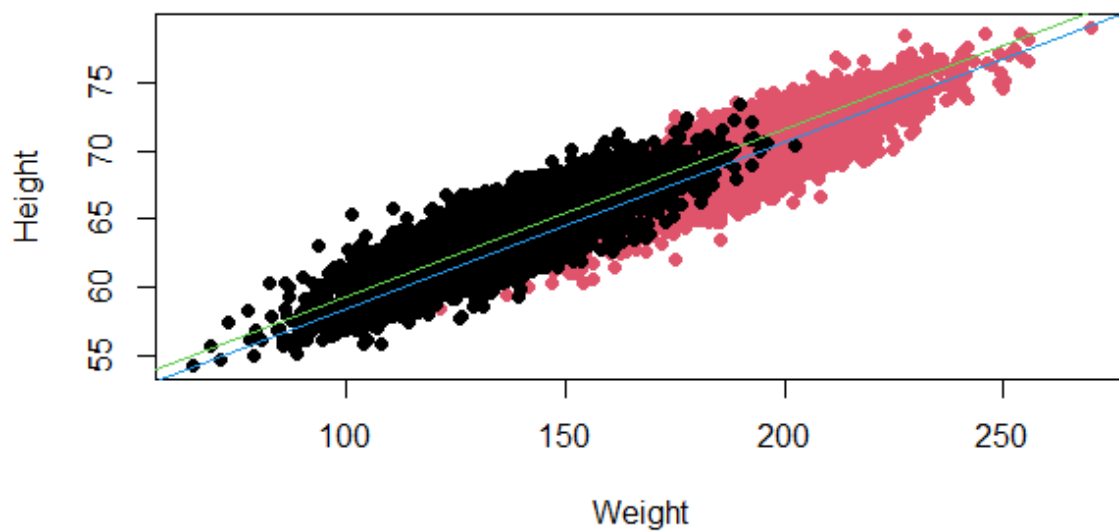then we plot the data and regression lines

```
 #ploting the result


coef <- coef(reg)


plot(Height~Weight, data, pch=16, col=as.numeric(Maleness)+1)

abline(a=coef[1], b=coef[2], col=3)

abline(a=coef[1] + coef[3], b=coef[2], col=4)
```

the resulting plot looks like this

As It can be seen, We created two plots that they show there is a difference between male and female samples

For the second part of the assignment, I subset data based on gender and then create separate regression lines for each subset.

First of all I subset the data

```
 #second part

#subseting based on gender

fdata <- filter(data,Maleness ==0)

mdata <- filter(data,Maleness ==1)
```

then I create a regression model for each subset

```
 #regression models

filter(data,Maleness ==0)

regf <-lm(Height~Weight ,data = fdata)

regm <-lm(Height~Weight ,data = mdata)

summary(regm)

summary(regf)
```

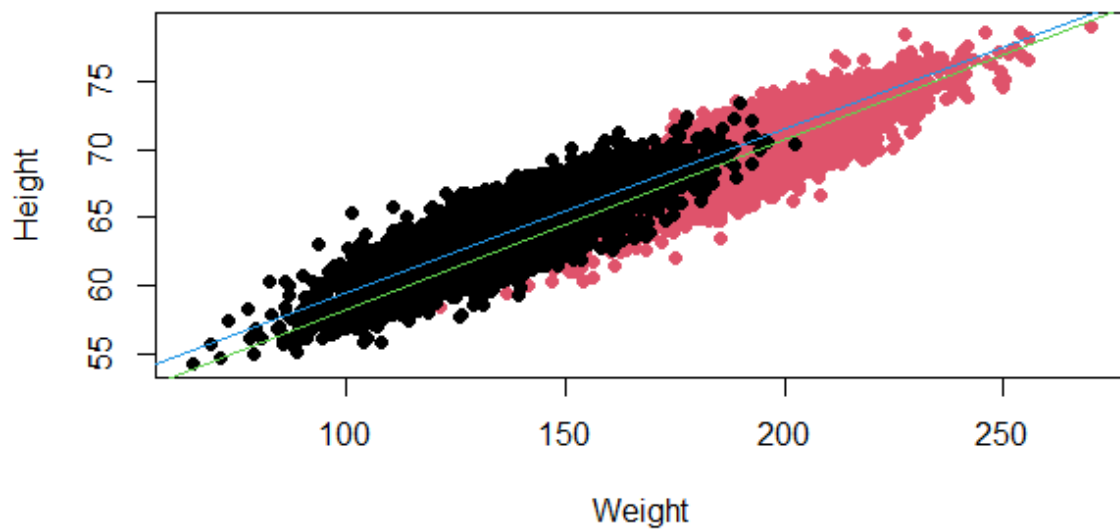And Finally I plot the results

```
 #ploting

plot(Height~Weight, data, pch=16, col=as.numeric(Maleness)+1)

abline(regm , col=3)

abline(regf , col=4)
```

the final plot is shown below

AS it's shown in the plot the regression lines aren't parallel like in the first plot which says that as the height increases the gender parameter losses its significance this finding is surprising for us considering human anatomy.

Bibliography

Lypsyl_123 (2021). Plot regression lines in r with multiple dummy variables. [online] Stack Overflow. Available at: https://stackoverflow.com/questions/67556086/plot-regression-lines-in-r-with-multiple-dummy-variables [Accessed 5 Apr. 2022].