Linear regression model (heart disease)

Answer submission

Student ID – 10899486

Name - S.C.S.Sandanayake

Degree – Data science(Plymouth)

Source code

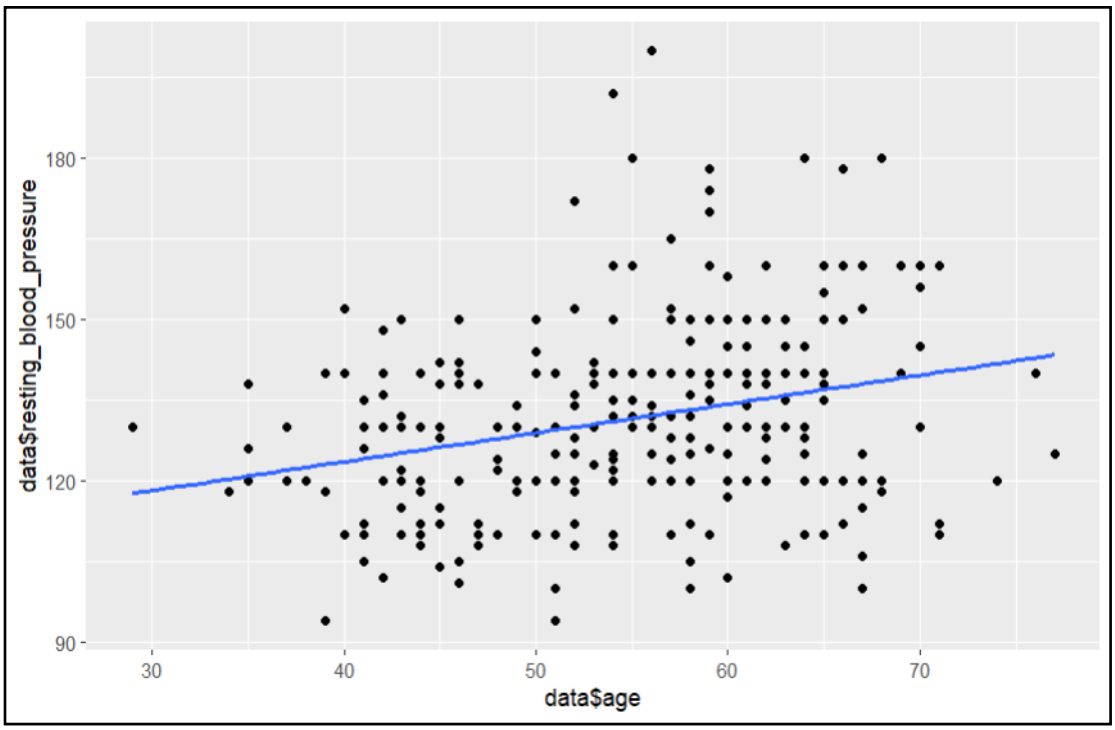
2.R* dataset

Source on Save

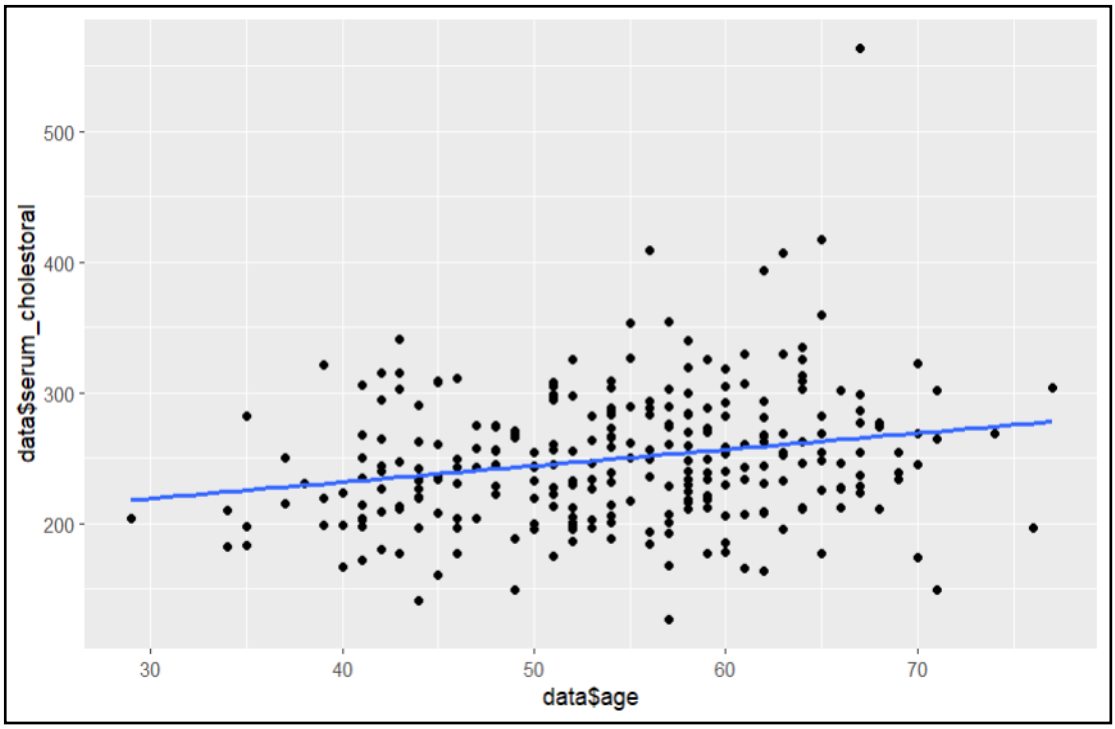
```
1  library(ggplot2)
2  library(dplyr)
3  data <- read.csv("D:\\R\\week 2\\heart disease\\heart_disease.csv")
4  summary(data)
5  str(data)
6
7  ggplot(data = data,aes(x=data$age,y=data$resting_blood_pressure))+
8    geom_point()+
9    geom_smooth(method = lm, se=FALSE)
10
11 ggplot(data = data,aes(x=data$age,y=data$serum_cholestoral))+
12   geom_point()+
13   geom_smooth(method = lm, se=FALSE)
14
15 #calculating the coeficeints and the constant
16
17 regmodel1 <- lm(resting_blood_pressure ~ age ,
18                 data = data)
19 coef(regmodel1)
20
21 regmodel2 <- lm(serum_cholestoral ~ age ,
22                 data = data)
23 coef(regmodel2)
```

Plots

1) Age vs. Resting Blood Pressure plot



2) Age vs. Serum Cholesterol plot



Summary and data structure

```
> summary(data)
      age             sex             chest       resting_blood_pressure serum_cholestoral fasting_blood_sugar
 Min.   :29.00   Min.   :0.0000   Min.   :1.000   Min.   : 94.0          Min.   :126.0     Min.   :0.0000
 1st Qu.:48.00   1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:120.0          1st Qu.:213.0     1st Qu.:0.0000
 Median :55.00   Median :1.0000   Median :3.000   Median :130.0          Median :245.0     Median :0.0000
 Mean   :54.43   Mean   :0.6778   Mean   :3.174   Mean   :131.3          Mean   :249.7     Mean   :0.1481
 3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:140.0          3rd Qu.:280.0     3rd Qu.:0.0000
 Max.   :77.00   Max.   :1.0000   Max.   :4.000   Max.   :200.0          Max.   :564.0     Max.   :1.0000
 resting_electrocardiographic_results maximum_heart_rate_achieved exercise_induced_angina    oldpeak
 Min.   :0.000                        Min.   : 71.0               Min.   :0.0000          Min.   :0.00
 1st Qu.:0.000                        1st Qu.:133.0               1st Qu.:0.0000          1st Qu.:0.00
 Median :2.000                        Median :153.5               Median :0.0000          Median :0.80
 Mean   :1.022                        Mean   :149.7               Mean   :0.3296          Mean   :1.05
 3rd Qu.:2.000                        3rd Qu.:166.0               3rd Qu.:1.0000          3rd Qu.:1.60
 Max.   :2.000                        Max.   :202.0               Max.   :1.0000          Max.   :6.20
     slope       number_of_major_vessels      thal           result
 Min.   :1.000   Min.   :0.0000          Min.   :3.000   Min.   :0.0000
 1st Qu.:1.000   1st Qu.:0.0000          1st Qu.:3.000   1st Qu.:0.0000
 Median :2.000   Median :0.0000          Median :3.000   Median :0.0000
 Mean   :1.585   Mean   :0.6704          Mean   :4.696   Mean   :0.4444
 3rd Qu.:2.000   3rd Qu.:1.0000          3rd Qu.:7.000   3rd Qu.:1.0000
 Max.   :3.000   Max.   :3.0000          Max.   :7.000   Max.   :1.0000
```

```
> str(data)
'data.frame':   270 obs. of  14 variables:
 $ age                                 : int  70 67 57 64 74 65 56 59 60 63 ...
 $ sex                                 : int  1 0 1 1 0 1 1 1 1 0 ...
 $ chest                               : int  4 3 2 4 2 4 3 4 4 4 ...
 $ resting_blood_pressure              : int  130 115 124 128 120 120 130 110 140 150 ...
 $ serum_cholestoral                   : int  322 564 261 263 269 177 256 239 293 407 ...
 $ fasting_blood_sugar                 : int  0 0 0 0 0 0 1 0 0 0 ...
 $ resting_electrocardiographic_results: int  2 2 0 0 2 0 2 2 2 2 ...
 $ maximum_heart_rate_achieved         : int  109 160 141 105 121 140 142 142 170 154 ...
 $ exercise_induced_angina             : int  0 0 0 1 1 0 1 1 0 0 ...
 $ oldpeak                             : num  2.4 1.6 0.3 0.2 0.2 0.4 0.6 1.2 1.2 4 ...
 $ slope                               : int  2 2 1 2 1 1 2 2 2 2 ...
 $ number_of_major_vessels             : int  3 0 0 1 1 0 1 1 2 3 ...
 $ thal                                : int  3 7 7 7 3 7 6 7 7 7 ...
 $ result                              : int  1 0 1 0 0 0 1 1 1 1 ...
```

1) coefficient for plot 1

```
> regmodel1 <- lm(resting_blood_pressure ~ age ,
+                 data = data)
> coef(regmodel1)
(Intercept)         age
102.1998345   0.5354184
>
```

2) coefficient for plot 2

```
> regmodel2 <- lm(serum_cholestoral ~ age ,
+                 data = data)
> coef(regmodel2)
(Intercept)         age
 181.691994    1.248633
>
```

Report

This data set includes information related to cardiovascular diseases. The main fields in the data table are patient age, resting blood pressure, serum cholesterol levels, and other variables. The summary and structure of the data set are generated, and the data visualization part was done using the ggplot2 library. The plots visually represent the relationship between age and two risk factors that mainly affect cardiovascular diseases. The plots are in “y=mx+c” type.

The first scatter plot shows how the resting blood pressure varies with age. The intercept is 102.1998345; the intercept shows estimated resting blood pressure when age equals 0. Practically, age is never becoming 0, as it is likely never 0 in the dataset. The coefficient for age is 0.5354184. That means for each one-unit increase in age, the estimated resting blood pressure is expected to increase approximately by 0.54 units. It’s a positive increase. Age and resting blood pressure are inversely proportional.

The second scatter plot shows how serum cholesterol levels vary with age. The intercept is 181.691994, which shows estimated serum cholesterol levels when age equals 0. Practically, age is never becoming 0, as it is likely never 0 in the dataset. The coefficient for age is 1.248633. That means for each one-unit increase in age, the estimated serum cholesterol level tends to positively increase approximately by 1.25 units. Age and serum cholesterol levels are inversely proportional.

In summary, both linear models show a positive relationship between age and the other variable (Y). Additionally, the reading of the intercept when age is 0 may not be practically meaningful. These conclusions are based solely on the relationships observed in the dataset. Other factors may increase resting blood pressure and serum cholesterol levels.

Thank you.