High Dimensional Exercise 4

Hamed Vaheb

2022-03-22

In an experiment on the analysis of the link between the beats per minute under stress and the age of a sample of 10 men, the following data have been collected:

Beats per minute	200	195	200	190	188	180	185	180	163	170
Age	10	20	21	25	29	30	31	40	45	50

1. Draw the scatterplot.
2. Build the linear regression model $Y_i = b_0 + b_1 t_i + \epsilon_i$ and estimate b_0 and b_1.
3. Verify the null hypothesis $H_0 : b_1 = 0$ against the alternative $H_1 : b_1 \neq 0$ with a significance level $\alpha = 0.05$.

Prepare

First we insert data of Beats per minute and age as columns of a new dataframe.

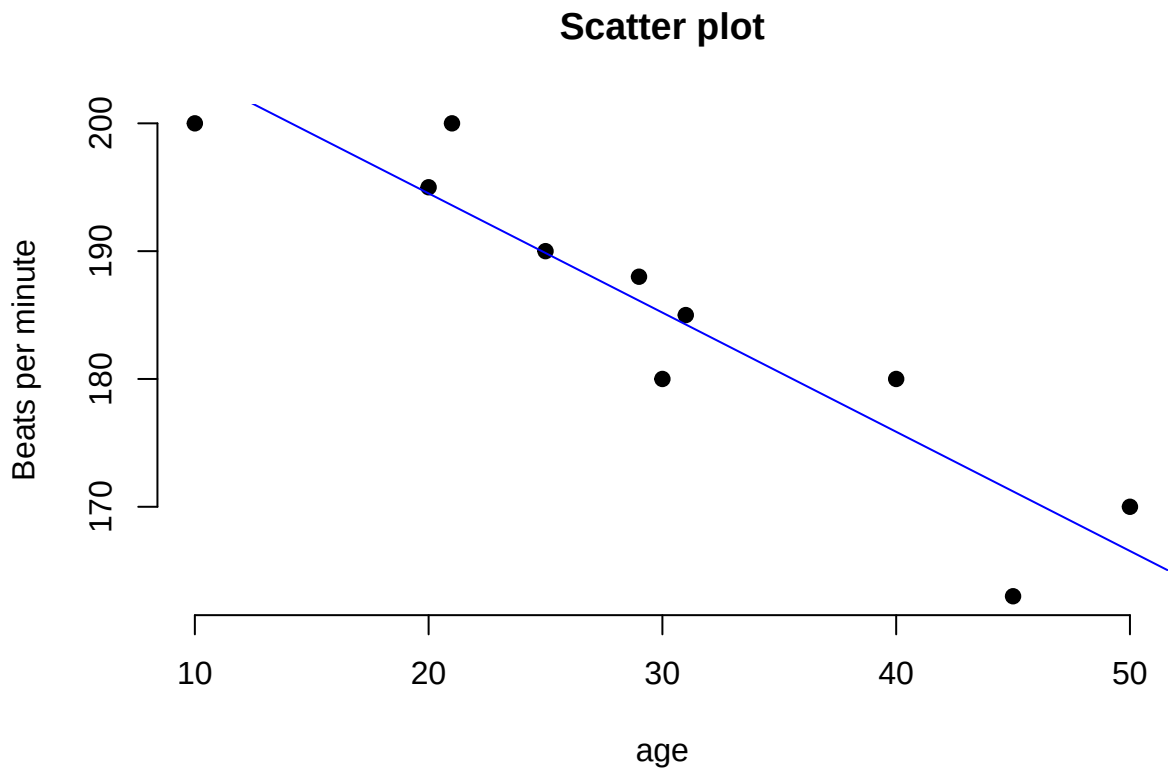
```
BeatsPerMin <- c(200, 195, 200, 190, 188, 180, 185, 180, 163, 170)
Age<- c(10, 20, 21, 25, 29, 30, 31, 40, 45, 50)

df <- data.frame(BeatsPerMin, Age)
```

Question 1: Drawing Scatterplot

Draw the scatterplot

```
x <- Age
y <- BeatsPerMin
# Plot with main and axis titles
# Change point shape (pch = 19) and remove frame.
# Add regression line
plot(x, y, main = "Scatter plot",
     xlab = "age", ylab = "Beats per minute",
     pch = 19, frame = FALSE)
abline(lm(y ~ x, data = df), col = "blue")
```



Question 2: Regression Model

Build the linear regression model $Y_i = b_0 + b_1 \cdot t_i + \epsilon_i$ and estimate b_0 and b_1.

$b_0 = 213.1721, b_1 = -0.9326$

```
model <- lm(BeatsPerMin ~ Age, data = df)
summary(model)
```

```
##
## Call:
## lm(formula = BeatsPerMin ~ Age, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2038 -2.8485  0.6099  3.0630  6.4131
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 213.1721     4.2296  50.400 2.66e-11 ***
## Age          -0.9326     0.1312  -7.108 0.000101 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.79 on 8 degrees of freedom
## Multiple R-squared:  0.8633, Adjusted R-squared:  0.8462
## F-statistic: 50.53 on 1 and 8 DF,  p-value: 0.0001011
```

Question 3: Hypothesis Testing

Verify the null hypothesis $H_0 : b_1 = 0$ against the alternative $H_1 : b_1 \neq 0$ with a significance level $\alpha = 0.05$.

The lm function provides p-values for significance of both intercept and slope. Since we are interested in the slope, we can compare the p-value (0.000101) with $\alpha = 0.05$, concluding that p-value is less than alpha, we reject the null hypothesis. This underlines that the dependent variable (age) has influence of target variable (beats per minute).