Exercise 1.7

2024-03-11

Exercise 1.7

Setting up R

```
testscores <- data.frame(hours = c(5,7.5,3,10.5,6,8,4,9), score = c(72,70,55,80,68,75,60,78))
```

(a)

Estimate the relationship between the test score as a dependent variable and the hours studied as an explanatory variable by computing the corresponding least-squares estimators beta1 and beta2.

beta1 is a measure of how much the dependent variable changes when you change the independent variable by one unit.
beta2 is the "starting point" for the relationship between the independent and dependent variables, for example if beta2 is 5 that means when the independent variable is zero the dependent variable is 5.

```
model <- lm(score ~ hours, data = testscores)
summary(model)
```

```
##
## Call:
## lm(formula = score ~ hours, data = testscores)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5526 -1.9030 -0.7305  0.9360  7.2695
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  49.2857     3.7561  13.121 1.21e-05 ***
## hours         3.0889     0.5329   5.797  0.00115 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.629 on 6 degrees of freedom
## Multiple R-squared:  0.8485, Adjusted R-squared:  0.8232
## F-statistic:  33.6 on 1 and 6 DF,  p-value: 0.001155
```

The depended variable is score and the explanatory variable is hours.
I created a data frame (testscores) containing these variables.
I used the lm() function to fit a linear regression model.
The summary() function provides the summary of the linear regression model, which includes the estimated coefficients beta1 and beta2.

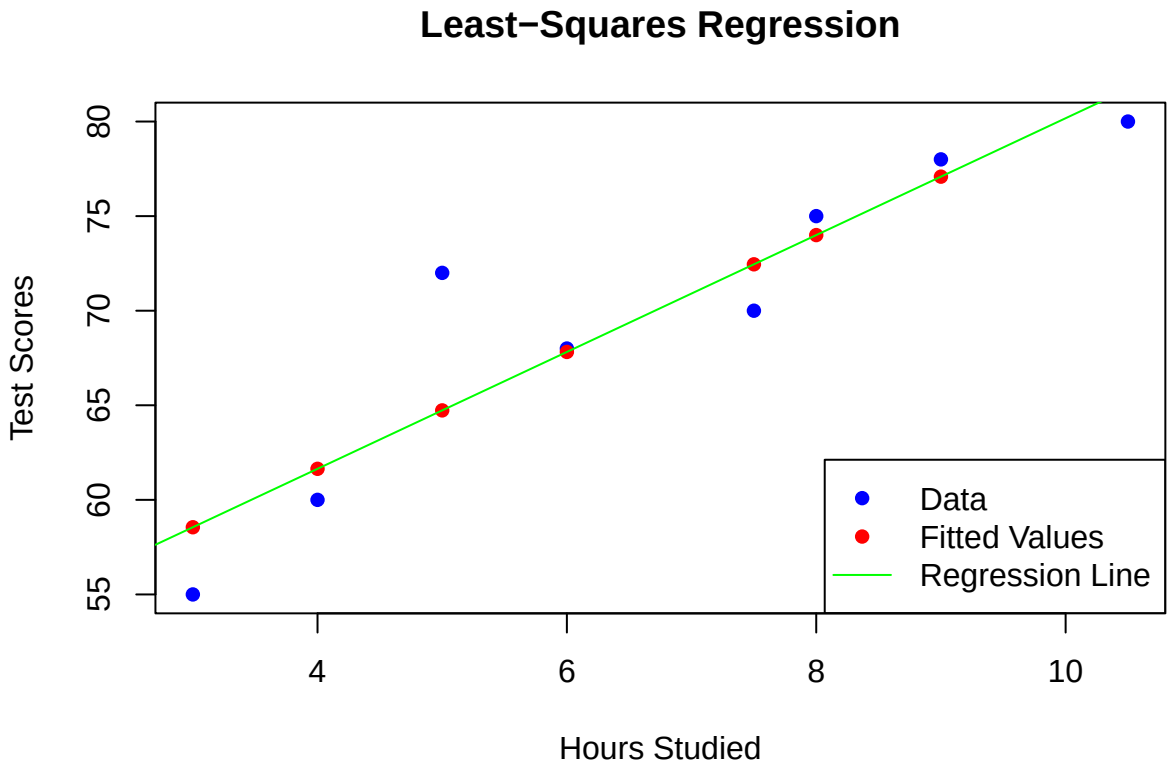
The coefficient for hours represents the estimate for beta1 and the intercept term represents the estimator of beta2.

beta1 = 3.0889
beta2 = 49.2857

(b)

Produce a plot of your results: plot the least-squares regression line together with the points (xi, yi) and (xi, ^yi). Comment on the "direction" of the relationship, i.e., the sign of the estimated slope parameter beta2

```
fitted_values <- predict(model)
plot(testscores, xlab = "Hours Studied", ylab = "Test Scores", main = "Least-Squares Regression",
col = "blue", pch = 16)
points(testscores$hours, fitted_values, pch = 16, col = "red")
abline(model, col = "green")
legend("bottomright", legend = c("Data", "Fitted Values", "Regression Line"),
col = c("blue", "red", "green"), pch = c(16,16, NA), lty = c(NA, NA, 1))
```



Since beta2 is positive it indicates a positive relationship.

(c)

What is the expected increase in the test score if the student studied an extra hours?

See beta1: If a student studied an extra hour he/she would gain 3.0889 Testpoints.