

PH142 Midterm

Eddie Espejo

8/30/2018

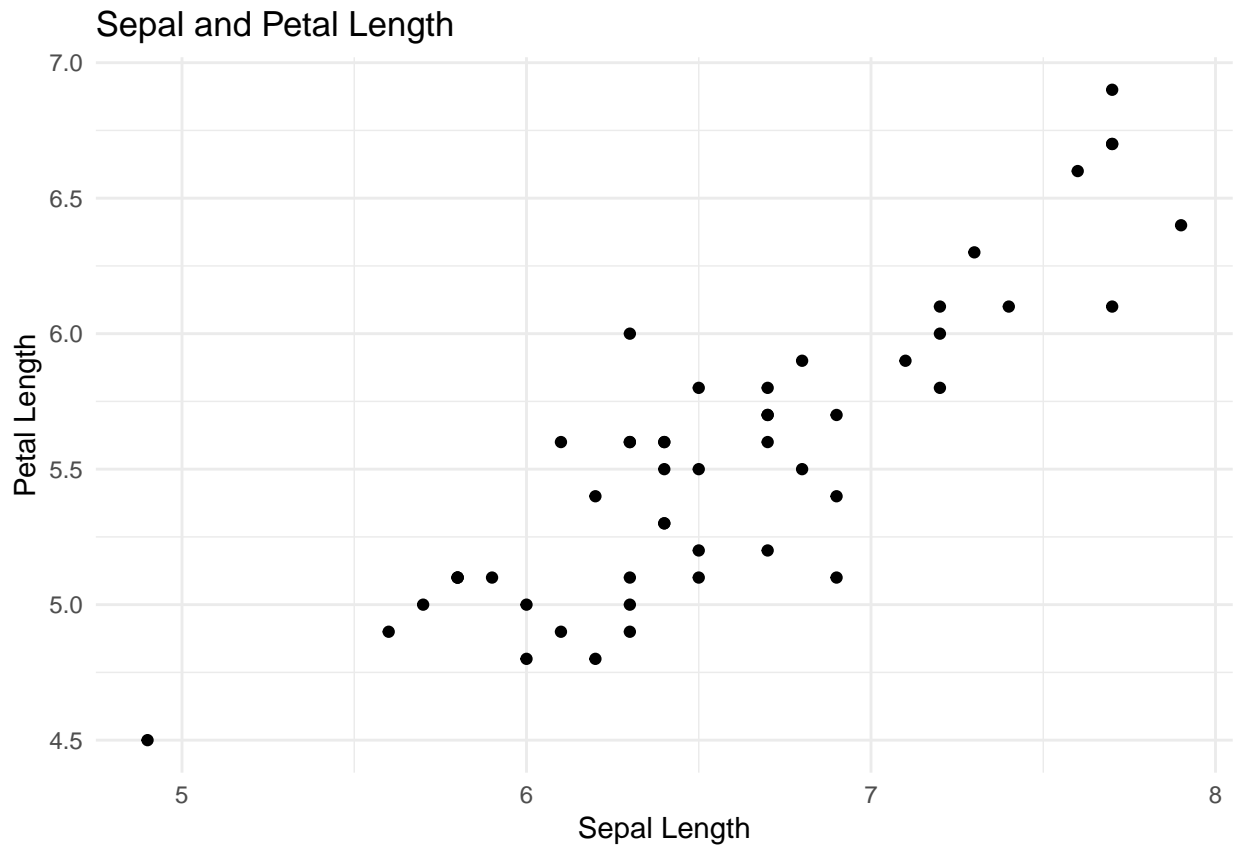
Problem 3

Part A

We have data on 50 flowers of the same species. We can safely assume there are no outliers in our dataset at this moment. The R output for `lm()` shows the following:

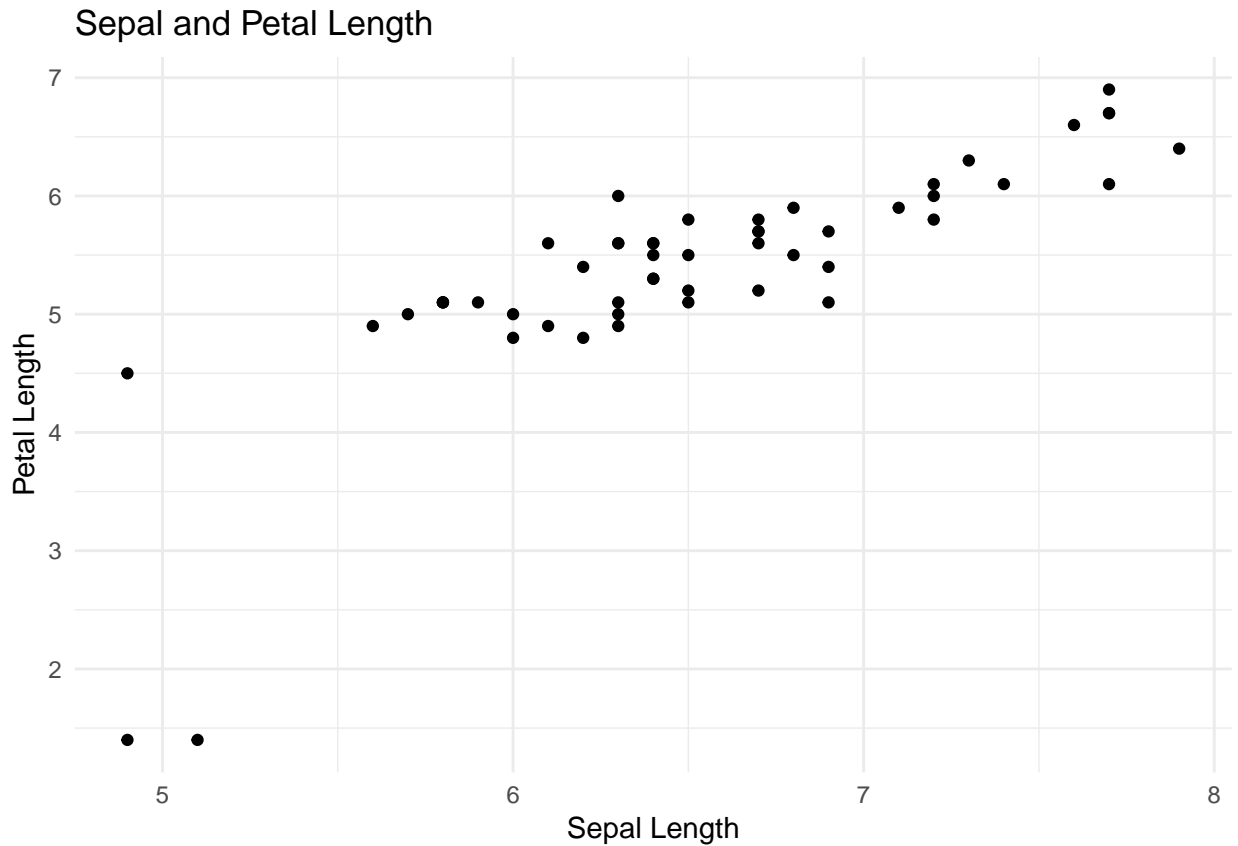
```
##  
## Call:  
## lm(formula = iris_subset$Petal.Length ~ iris_subset$Sepal.Length)  
##  
## Coefficients:  
## (Intercept)  iris_subset$Sepal.Length  
##          0.6105                0.7501
```

How would you write the formula for the least squares regression line? Draw the least squares regression line by hand on the plot below.



Part B

Now, you are given data on 2 more flowers that we think are the same species as the flowers we have data on. Redraw the an estimate for the line of best fit on the new plot.



Part C

The slope of the line for 3b is _____ r #lower than (or higher depending on the chosen example) than the slope of the line for 3a. This is because the slope is not _____ r #resistant to outliers.

Part D

On another occasion, you receive data on a flower that has a sepal length of 7.5. What do you expect its petal length to be? Is this prediction appropriate to do given the data you have?

Extra Question

Say you run the following code in your console.

```
x <- 4
y <- 2
```

Next, you write `z == x+y`. You do not have a variable yet assigned to `z`. Does `z` show up in your environment? If so, what does `z` contain?