Statistical Inference Project

A Project By Plato Karageorgis

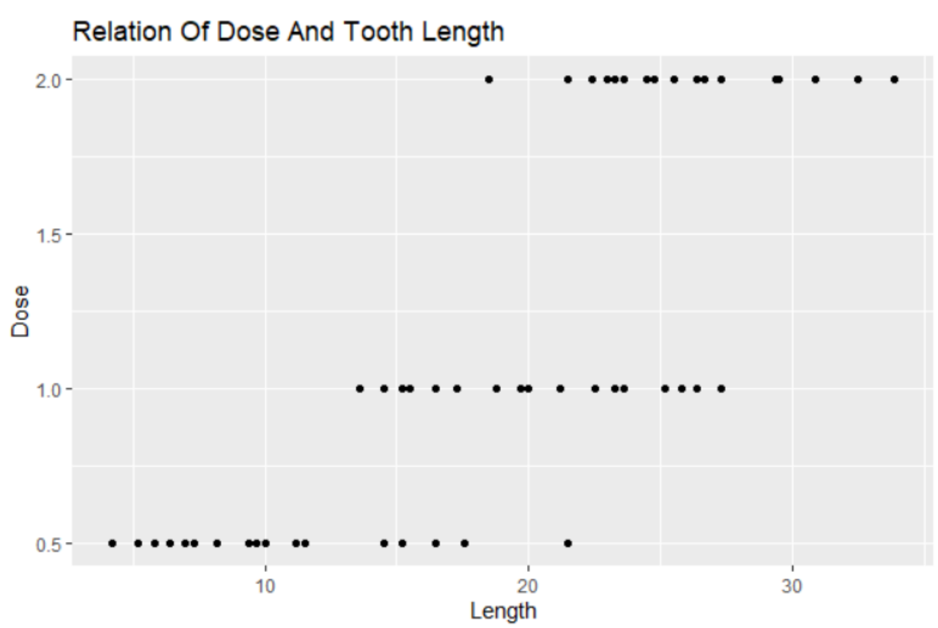
Part 2

The purpose of the 2nd part is to evaluate the data from the Tooth Growth dataset. We will then run some tests and find confidence intervals. Let's begin with the data loading.

- library(ggplot2)
- data(ToothGrowth)

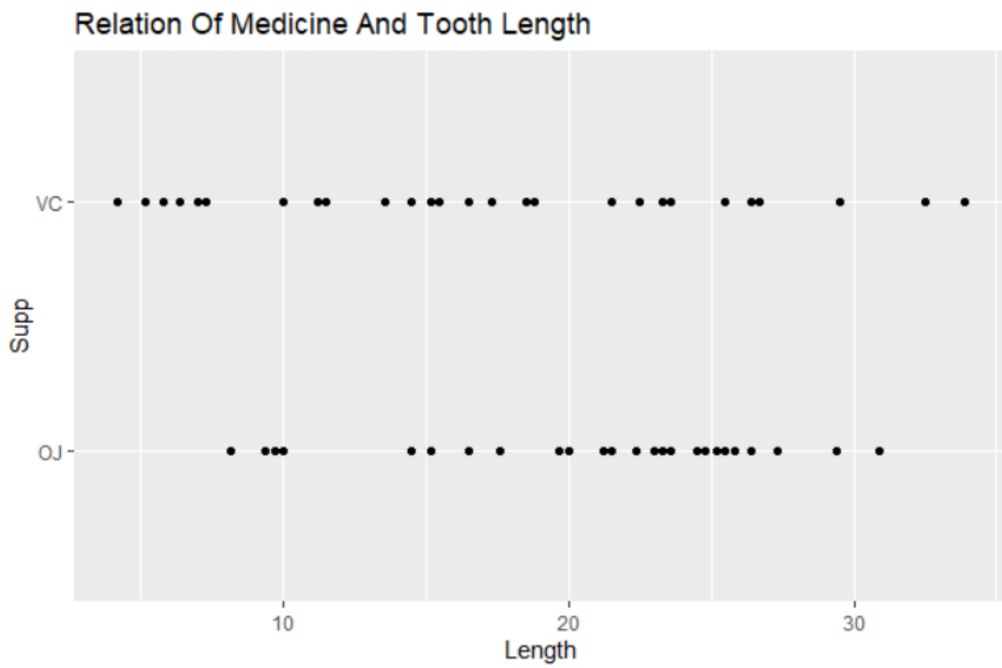
Now, for the first part the exercise tells us to perform some basic data analyses without specification. The data that I see here in R studio have 3 columns, the length of the teeth, the medicine that was administered and the dose of the medicine. From my point of view, we will have a good idea after creating 2 scatterplots, one comparing the length with the dose and one comparing the length with the medicine.

- qplot(ToothGrowth$len, ToothGrowth$dose, xlab = "Length", ylab = "Dose", main= "Relation Of Dose And Tooth Length")



In the scatterplot above we can see that the length increases with the amount of medicine that has been administered. That is a logical conclusion.

- qplot(ToothGrowth$len, ToothGrowth$supp ,xlab = "Length", ylab = "Supp", main= "Relation Of Medicine And Tooth Length")



In this one we can see that the most effective medicine is the “OJ” because the data seem to be more concentrated around 20-25, whereas the mean using the “VC” should be definitely below 20. We could easily perform these calculations but there is no reason for it yet, since we are just trying to get the idea for now.

<u>Tooth Length and Supp Comparison</u>

We will now attempt to get a deeper look in the comparison between the tooth length and the medicine. We make a hypothesis H0: The medicines have no difference at all. The alternative HA: The medicines are different. We set the alpha value 5%. We use the t.test known from the slides.

- t.test(ToothGrowth$len[ToothGrowth$supp=='OJ'], ToothGrowth$len[ToothGrowth$supp=='VC'])

We get the following result:

```
        Welch Two Sample t-test

data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$su
pp == "VC"]
t = 1.9153, df = 55.309, p-value = 0.06063
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1710156  7.5710156
sample estimates:
mean of x mean of y
 20.66333  16.96333
```

The pValue is 0.06 which is bigger than the alpha and so we fail to reject the null hypothesis that these 2 medicine are the same. If we modify our expectations and set alpha to be 10%, then we will be able to accept the null hypothesis but then we would be increasing the type I error rate.

Tooth Length and Dose Comparison

Furthermore, we will run a test that will prove that the increase of a dose increases the tooth length (that was our instinct before right?). We set H0: The increase of dose has no result and HA: The increase of dose is effective. We set once again alpha = 5%. In this situation we need to run 2 tests. One comparing the half dose with the complete dose and one comparing the complete dose with the double. For the first case:

- t.test(ToothGrowth$len[ToothGrowth$dose == 1], ToothGrowth$len[ToothGrowth$dose == 0.5], alternative = "greater")

```
	Welch Two Sample t-test

data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose
 == 0.5]
t = 6.4766, df = 37.986, p-value = 6.342e-08
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 6.753323      Inf
sample estimates:
mean of x mean of y 
   19.735    10.605 
```

The pValue is much smaller than the alpha and we can confidently say that we reject the null hypothesis in this first scenario.

- t.test(ToothGrowth$len[ToothGrowth$dose == 2], ToothGrowth$len[ToothGrowth$dose == 1], alternative = "greater")

```
	Welch Two Sample t-test

data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose
 == 1]
t = 4.9005, df = 37.101, p-value = 9.532e-06
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 4.17387     Inf
sample estimates:
mean of x mean of y 
   26.100    19.735 
```

The pValue is again much smaller than the alpha and so we can reject the null hypothesis again.

Consequently, we reject the null hypothesis that the increase of dose is irrelevant, since both of our tests showed that the increase does indeed matter.