Statistical Inference Course Project - Part 2

FChemin

November 3, 2016

Overview

In this project, we're going to analyze the ToothGrowth data in the R datasets package.

Loading the dataset and performing basic exploratory data analysis

```
data <- ToothGrowth
head(data)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

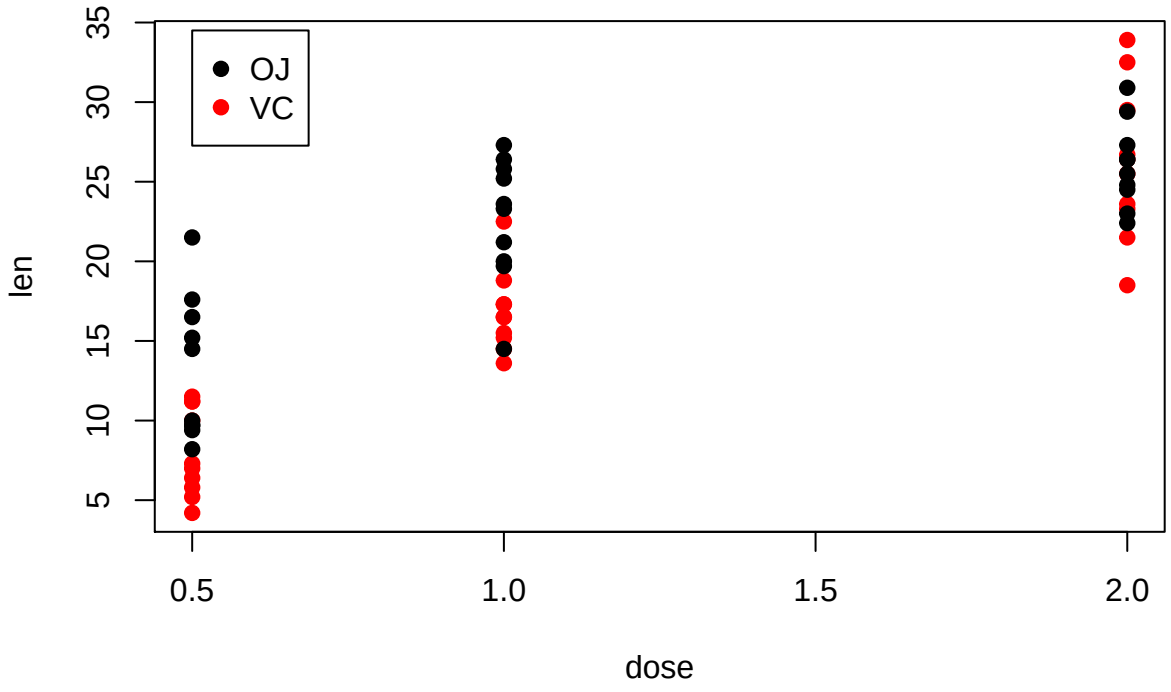
```
tail(data)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

```
str(data)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
plot(data$dose, data$len, xlab = "dose", ylab = "len", pch=19, col=data$supp)
legend(x=0.5, y=34.5, legend = levels(data$supp), col = c(1,2), pch=19)
```



Providing a basic summary of the data

```
summary(data)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Comparing tooth growth by supp and dose using confidence intervals and/or hypothesis tests

First, we perform a t-test on the supp and check if p-value < 0.05 and if 0 is in the confidence interval:

```
t_test_supp <- t.test(ToothGrowth$len[ToothGrowth$supp == "OJ"], ToothGrowth$len[ToothGrowth$supp == "V

t_test_supp
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "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 p-value is equal to 0.0606345 and larger than 0.05, and the 95 percent confidence interval ranges from -0.1710156 to 7.5710156. Zero is inside this interval so **we fail to reject the null hypothesis**.

Then, we perform another t-test but this time on the dose variable:

```
t_test_dose <- t.test(ToothGrowth$len[ToothGrowth$dose == 1], ToothGrowth$len[ToothGrowth$dose == 2], pa

t_test_dose
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 2]
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

The p-value is equal to 0.00001810829 so that's really close to zero, and the 95 percent confidence interval ranges from -8.9943868 to -3.7356132. Zero is not inside this interval so **we reject the null hypothesis**.

Conclusions

We conclude here that the supplement type has no effect on tooth growth but, the act of giving supplement and the dosage of it have clearly effects, particularly on the higher dosage. This is true if we assume a normal distribution of the means, that the sample is representative of the population, and dosage and supplement were given randomly.