Statistical Inference - Course Project - Part 2

Michael Elfellah

September 27, 2015

Part 2: Basic inferential data analysis.

In part 2, we're going to analyze the ToothGrowth data in the R datasets package.

The steps are:

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
4. State our conclusions and the assumptions needed for our conclusions.

The evaluation criteria are:

1. Exploratory data analysis of at least a single plot or table highlighting basic features of the data?
2. Some relevant confidence intervals and/or tests?
3. Results of the tests and/or intervals interpreted in the context of the problem?
4. Description of the assumptions needed to arrive at the appropriate conclusions?

Loading the required library and dataset

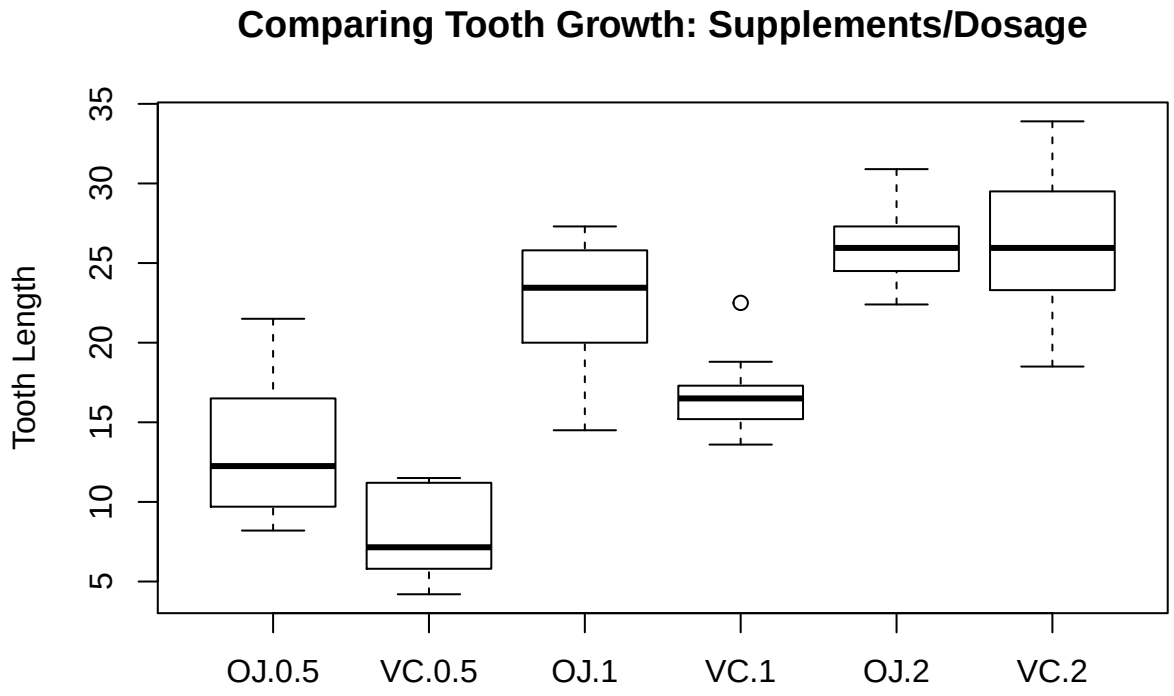
```
library(datasets)
data(ToothGrowth)
```

Exploratory analysis

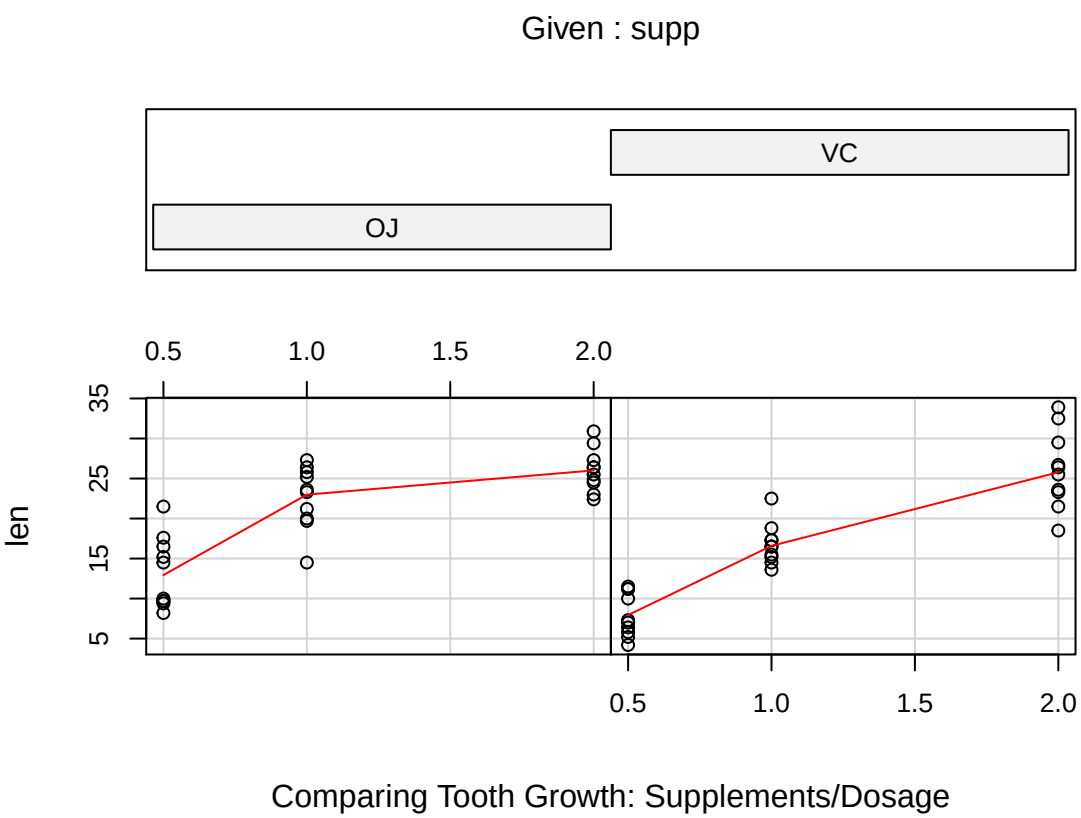
The command data() provides a quick summary about each dataset. In this case, we can get more information about the ToothGrowth dataset as follows: "The Effect of Vitamin C on Tooth Growth in Guinea Pigs". Furthermore, "The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid)."

We can visualize the dataset easily with a boxplot and coplot:

```
boxplot(len ~ supp * dose, data=ToothGrowth, ylab="Tooth Length",
        main="Comparing Tooth Growth: Supplements/Dosage")
```




```
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "Comparing Tooth Growth: Supplements/Dosage")
```



Basic summary of the data.

The average of the tooth length seems to increase with the supplement dosage. In other words, there seems to be a relationship between applying a supplement dosage and the tooth growth. Further data summaries:

```
head(ToothGrowth)
```

```
##    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
```

```
nrow(ToothGrowth)
```

```
## [1] 60
```

```
summary(ToothGrowth)
```

```
##       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
```

```
table(ToothGrowth$supp,ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

Using confidence intervals and/or hypothesis tests to compare tooth growth by supp and dosage. Since the sample size is small, a T distribution is recommended.

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

Comparing the difference between the two supplements yields no convincing evidence to reject the null hypothesis, since the p-value is bigger than the 5% significance level.

However, for this dataset it is important to also compare the differences between the different dosage levels, since bigger dosage may yield opposing results.

```
t.test(ToothGrowth$len, ToothGrowth$dose)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len and ToothGrowth$dose
## t = 17.81, df = 59.798, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  15.66453 19.62881
## sample estimates:
## mean of x mean of y
## 18.813333  1.166667
```

Comparing the difference between the two supplements yields conclusive evidence to reject the null hypothesis, since the p-value aproximates to 0, and is thus substancially smaller in comparison to the significance level.

As already stated, there is no convincing evidence that there is a difference between the two type of supplements based on the existing datasets and T statistics. Thus we fail to reject the Null hypothesis.

However, there is convincing evidence that there is a difference between the dosage levels, and the growth. Thus we reject the Null hypothesis in favour of the alternative hypothesis.