Peer-graded Assignment: Statistical Inference Course Project PART 2

Gabriela Ochoa

Resolution Part 2

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1. We load the ToothGrowth data in the R datasets package and perform some basic exploratory data analyses

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

EDA

Data: *The Effect of Vitamin C on Tooth Growth in Guinea Pigs*

Description:

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

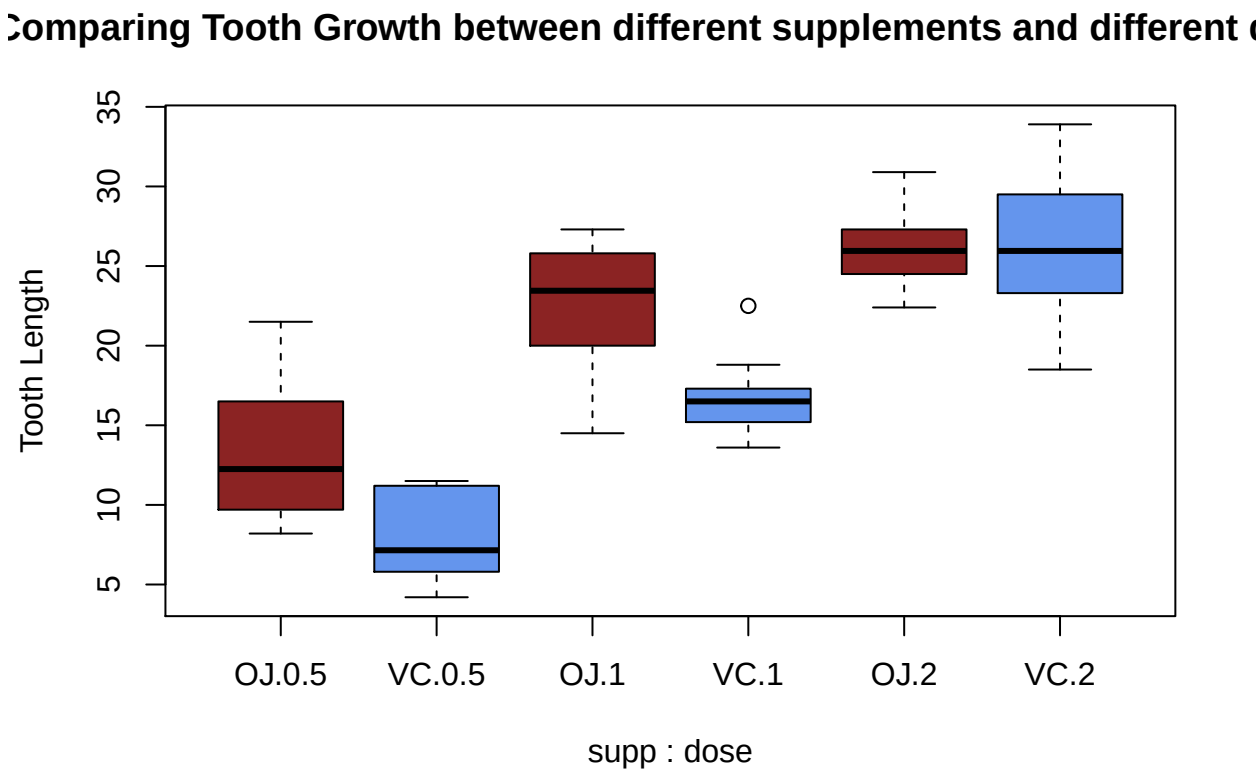
Format

A data frame with 60 observations on 3 variables.

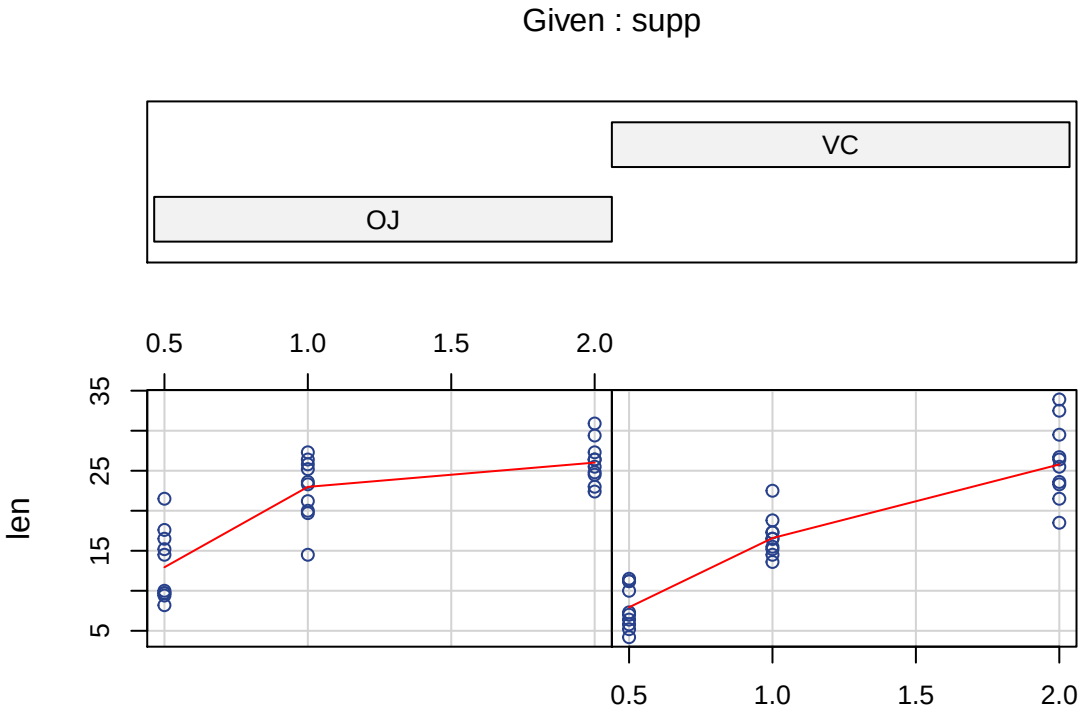
[,1]	len	numeric	Tooth length
[,2]	supp	factor	Supplement type (VC or OJ)
[,3]	dose	numeric	Dose in milligrams/day

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

```
boxplot(len ~ supp * dose, data=ToothGrowth, ylab="Tooth Length", main="Comparing Tooth Growth between
```




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



Comparing Tooth Growth between different supplements and different dosis

2. Provide a basic summary of the data.

As we can see in the plots, the average of the tooth length seems to increase with the supplement dosis. this can mean that it might be a relationship between applying a supplement dosis and the tooth growth.

Further data summaries:

```
# basic info
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
```

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

Sample size is not too big, we suggest T distribution

```
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 shows no strong 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 the different dosis level, since bigger dosis may yield contradicting evidence.

```
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 shows enough evidence to reject the null hypothesis, since the p-value aproximates to 0.

4. State your conclusions and the assumptions needed for your conclusions.

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. meaning **we fail to reject the *Null hypothesis* (H0).**

However, there is convincing evidence that there is a difference between the dosis level, and the growth. meaning **we reject the *Null hypothesis* (H0) in favour of the alternative hypothesis (Ha).**