Statistical Inference Course Project. Part 2

Pedro Villalba

12 de junio de 2018

1. Exploratory data analysis

Load data ToothGrowth and perform some basic EDA.

```
# Load data
data(ToothGrowth)
# Data structure
str(ToothGrowth)
```

```
## '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 ...
```

```
# Frequency of supp and dose
print(table(ToothGrowth$supp,ToothGrowth$dose))
```

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

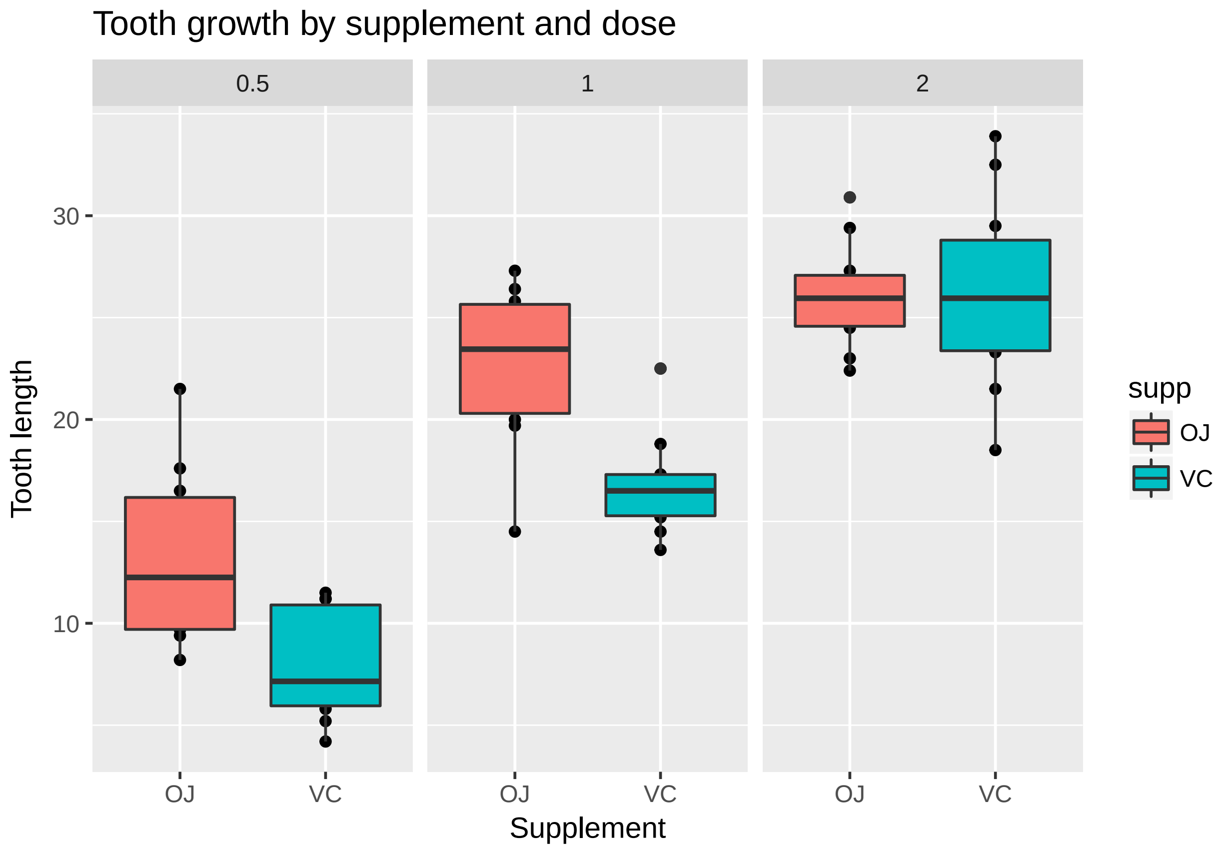
2. Data summary

```
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
```

```
# Plot sup vs.dose
p <- qplot(supp,len,data=ToothGrowth, facets=~dose,
           main="Tooth growth by supplement and dose",
           xlab="Supplement", ylab="Tooth length") +
    geom_boxplot(aes(fill = supp))

print(p)
```



3. Hypothesis testing

We will use T test to compare the tooth length for differents supplement types and dose. We will consider inequal variances in all cases.

First, check the effect of supplement type which can have the values "OJ" and "VC".

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, 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
```

The p-Value is greater than 0.05, so we cannot reject the null hypothesis that the tooth growth means are equal due to supplementary type.

The factor "dose" has 3 possible values ("0.5", "1.0" and "2.0") so we have to perform 3 T tests. Study the effect of dose "0.5" and "1.0"

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE,
       data = ToothGrowth[ToothGrowth$dose %in% c(0.5,1.0),])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

Study the effect of dose "1.0" and "2.0"

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE,
       data = ToothGrowth[ToothGrowth$dose %in% c(1.0,2.0),])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

Study the effect of dose "0.5" and "2.0"

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE,
       data = ToothGrowth[ToothGrowth$dose %in% c(0.5,2.0),])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

The p-Value is less than 0.05 in all three cases, so we accept the null hypothesis that the tooth growth means are different for every test.

4. Conclussions

We can conclude that only dose has effect in tooth growth.