Statistical Inference: ToothGrowth data set analysis

José Manuel Peñas

Contents

1 Overview

The present document will analyze the the **ToothGrowth** data set, giving basic information, exploratory plots and reaching conclusions about the effect of supplement delivery and the dosage over tooth growth.

2 Analysis

Getting this from the documentation, the **ToothGrowth** contains data from an experiment that measures the Effect of Vitamin C on Tooth Growth in Guinea Pigs.

2.1 Loading & some basic exploratory data analysis

We load the data set and check the data stored.

```
# Used in the plots later
library(ggplot2)
# loading the data
data(ToothGrowth)
# Getting the structure
str(ToothGrowth)
```

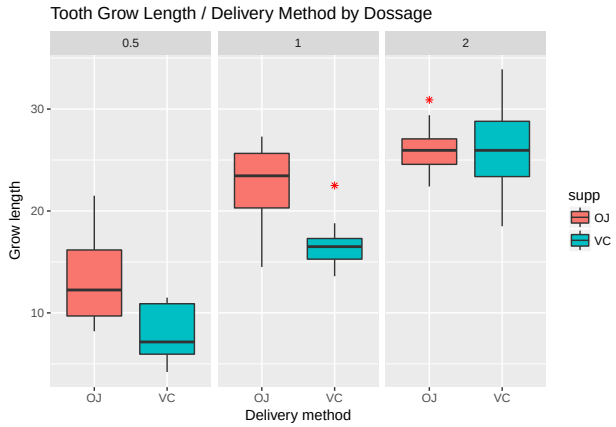
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
# Summary
summary(ToothGrowth)
```

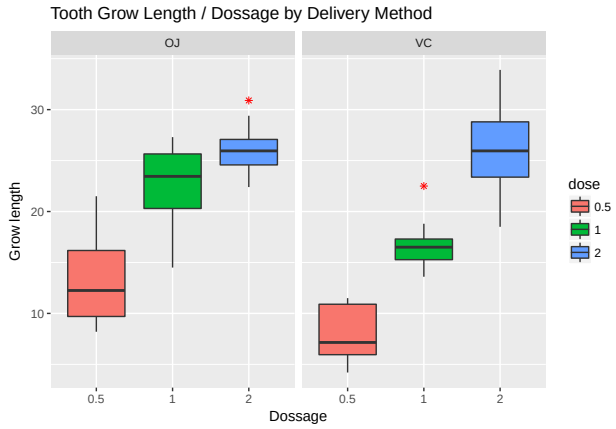
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Now, we are going to plot the distribution of the tooth growth by dosage and delivery method.

```
g <- ggplot(ToothGrowth, aes(x=supp, y=len)) +
        geom_boxplot(outlier.shape=8, outlier.colour = "red", aes(fill=supp))
g <- g + ggtitle("Tooth Grow Length / Delivery Method by Dossage") + xlab("Delivery method") +ylab("Grow
g <- g + facet_grid(.~ dose)
g
```




```
# Convert the variable dose from a numeric to a factor variable
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
g <- ggplot(ToothGrowth, aes(x=dose, y=len)) +
        geom_boxplot(outlier.shape=8, outlier.colour = "red", aes(fill=dose))
g <- g + ggtitle("Tooth Grow Length / Dossage by Delivery Method") + xlab("Dossage") +ylab("Grow length
g <- g + facet_grid(.~ supp)
g
```



The first exploratory analysis shows that it seems an increase of the dose could mean an increase on tooth length and that the delivery method is not a factor on that growth. Let's check it out these hypothesis.

2.2 Confidence intervals and hypothesis tests

The null Hypothesis in both cases is that so supp as dose has no effect over the grow. We will test them using t distribution test.

```
c(t.test(len ~ supp, data=ToothGrowth)$p.value,t.test(len ~ supp, data=ToothGrowth)$conf.int)
```

```
## [1]  0.06063451 -0.17101562  7.57101562
```

The *p-value* is over *5%* and the confidence interval contains zero. We **cannot reject the null hypotesis**.

```
# As the t.test only works with 2-level factors, we test for every combination
# and stored the p-value and confidence interval in a matrix
resultados <-
        c(
                "0.5 & 1",
                t.test(len ~ dose, data = subset(ToothGrowth, dose %in% c(.5, 1)))$p.value,
                t.test(len ~ dose, data = subset(ToothGrowth, dose %in% c(.5, 1)))$conf.int
        )
resultados <- rbind(resultados, c(
        "0.5 & 2",
        t.test(len ~ dose, data = subset(ToothGrowth, dose %in% c(.5, 2)))$p.value,
        t.test(len ~ dose, data = subset(ToothGrowth, dose %in% c(.5, 2)))$conf.int
))
resultados <- rbind(resultados, c(
        "1 & 2",
        t.test(len ~ dose, data = subset(ToothGrowth, dose %in% c(1, 2)))$p.value,
        t.test(len ~ dose, data = subset(ToothGrowth, dose %in% c(1, 2)))$conf.int
))
colnames(resultados) <- c("Combination","P.value","Conf.Int.Min","Conf.Int.Max")
rownames(resultados) <- NULL
data.frame(resultados)
```

Combination	P.value	Conf.Int.Min	Conf.Int.Max
0.5 & 1	1.26830072017385e-07	-11.9837812579016	-6.27621874209841
0.5 & 2	4.39752495936323e-14	-18.1561665388306	-12.8338334611694
1 & 2	1.9064295136718e-05	-8.99648051689202	-3.73351948310799

Neither of the three combinations of doses contains zero in the respective confidence interval and the p-value is extremely low on the 3 cases, way down the *5%* threshold. We can **reject the the null hypotesis**.

3 Conclusion

We are assuming that this sample is representative of the population of this experiment and any inference here could be applied to the total population.
1. Does the **delivery method** of the supplement has any effect on tooth growth?
* **No**. We cannot infer that answer as neither the box plot nor the t.test, where we couldn't reject the null hypothesis, could confirm it.
2. Does the **dossage** of the supplement has any effect on tooth growth?
* **Yes**. We are able to infer that answer as the box plot as the t.test, where we could reject the null hypothesis, confirm it.