Statistical Inference Part II - ToothGrowth dataset

Fabien Nugier

11/15/2019

This project consists of two parts, in Part I we study the properties of the exponential distribution and compare its properties with the Central Limit Theorem (CLT). In the Part II, we perform some basic inferential data analysis using the R dataset *ToothGrowth*. This is just an illustrative assignment.

This document contains part II.

Introduction

In this part we perform a simple exploratory analysis on the R dataset `ToothGrowth`. According to the description of the dataset, the data reports the study of the effect of Vitamin C on tooth growth (the tooth being used is the incisor tooth) in guinea pigs, more precisely:

> 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).

Part II - Inferential Data Analysis on the ToothGrowth dataset

We load the package as:

```
data(ToothGrowth)
```

and check information about the dataset:

```
data(ToothGrowth)
head(ToothGrowth,n=3)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
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(ToothGrowth$supp)
```

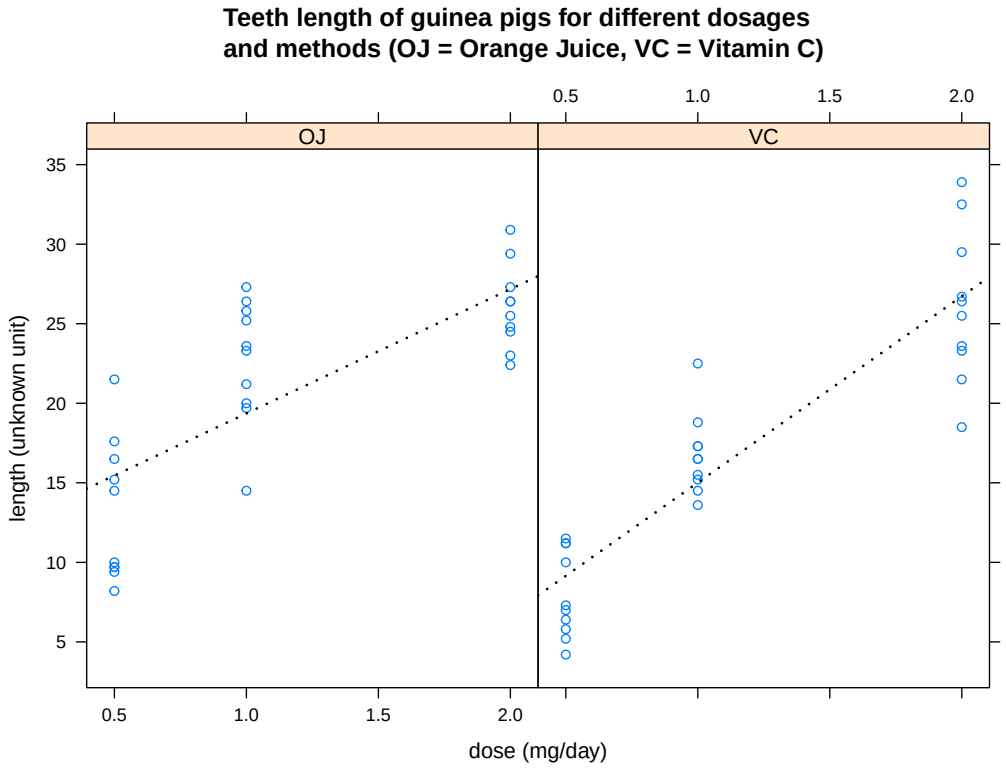
```
## OJ VC
## 30 30
```

```r
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

To get a better grasp on the data, we do a scatter plot separating the two delivery methods (OJ for "Orange Juice" and VC for "Ascorbic acid", a form of "Vitamin C") and different dosages:

```r
library(lattice)
pltt <- "Teeth length of guinea pigs for different dosages \n"
pltt <- paste(pltt,"and methods (OJ = Orange Juice, VC = Vitamin C)")
xl <- "dose (mg/day)"
yl <- "length (unknown unit)"
pan <- function(x,y,...) {
    panel.xyplot(x, y, ...)
    panel.lmline(x, y,type='l',lty=3,lwd=2, ...)
    }
xyplot(len~dose|supp,data=ToothGrowth,layout=c(2,1),xlab=xl,ylab=yl,main=pltt,panel=pan)
```



We perform a T-test on the data, assuming that the two guinea pigs test groups with OJ and VC treatments have the same underlying distribution and not pairing elements of the groups. We obtain the following results:

```r
t.test(len ~ supp, paired=FALSE, var.equal=TRUE, data=ToothGrowth)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
```

```
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The null hypothesis H_0 is that *both test groups have the same tooth length in the OJ and VC groups*. Considering the p-value which is 0.06039, we can see that a 5% p-value would not reject the null hypothesis. However, a 1% p-value would reject H_0, meaning the treatment has been more efficient on the OJ group, for which the average tooth length is longer, than in the VC group.

The t-test also returns us the 95% confidence interval of the difference between the two groups. As we can see the value 0 is not excluded, which confirms that the null hypothesis is not excluding a 5% confidence level.

Looking at the figures, we can also see that the dosage has a significant importance in the results. The more the dosage, the longer the tooth. This seems even more true for the VC group than it is for the OJ group.

Short conclusions

In this document we have studied a dataset of two different treatments and dosages applied to guinea pigs, with the associated measurement of their tooth length. It was shown that one treatment is better than the other, but a 5% p-value is not enough to reject the null hypothesis of equal mean between the two groups. However, dosage is clearly correlated with the tooth length.