Exponential distribution in R and Central Limit Theorem. Comparission

Miguel Angel Vallejo Gato

25 de octubre de 2015

Contents

1. Overview

In this project we're going to analyze the ToothGrowth data in the R datasets package.

We will:

1. Perform some basic exploratory data analyses.
2. Provide a basic summary of the data
3. Use confidence intervals and/or hypothesis test to compare tooth growth by supp and dose.

2. The dataset ToothGrowth

This dataset show 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

- Source

C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.

3. Exploratory data analysis

First we are going to load the dataset and the necessary libraries.

```
library(dplyr)
data(ToothGrowth)
dataset <- ToothGrowth
```

Now, we are going to do a basic analisis of the data.

1. Structure of the dataset

```
str(dataset)
```

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

2. Summary

```
summary(dataset)
```

```
##       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
```

3. First 3 rows of the dataset

```
head(dataset,3)
```

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

4. Last 3 rows of the dataset

```
tail(dataset,3)
```

```
##     len supp dose
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

5. Summary of each pair supp-dose.

```
summary(filter(dataset,supp=="OJ",dose==0.5))
```

```
##       len        supp         dose
##  Min.   : 8.20   OJ:10   Min.   :0.5
##  1st Qu.: 9.70   VC: 0   1st Qu.:0.5
##  Median :12.25           Median :0.5
##  Mean   :13.23           Mean   :0.5
##  3rd Qu.:16.18           3rd Qu.:0.5
##  Max.   :21.50           Max.   :0.5
```

```
summary(filter(dataset,supp=="VC",dose==0.5))
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ: 0   Min.   :0.5
##  1st Qu.: 5.95   VC:10   1st Qu.:0.5
##  Median : 7.15           Median :0.5
##  Mean   : 7.98           Mean   :0.5
##  3rd Qu.:10.90           3rd Qu.:0.5
##  Max.   :11.50           Max.   :0.5
```

```r
summary(filter(dataset,supp=="OJ",dose==1.0))
```

```
##       len        supp         dose
##  Min.   :14.50   OJ:10   Min.   :1
##  1st Qu.:20.30   VC: 0   1st Qu.:1
##  Median :23.45           Median :1
##  Mean   :22.70           Mean   :1
##  3rd Qu.:25.65           3rd Qu.:1
##  Max.   :27.30           Max.   :1
```

```r
summary(filter(dataset,supp=="VC",dose==1.0))
```

```
##       len        supp         dose
##  Min.   :13.60   OJ: 0   Min.   :1
##  1st Qu.:15.28   VC:10   1st Qu.:1
##  Median :16.50           Median :1
##  Mean   :16.77           Mean   :1
##  3rd Qu.:17.30           3rd Qu.:1
##  Max.   :22.50           Max.   :1
```

```r
summary(filter(dataset,supp=="OJ",dose==2.0))
```

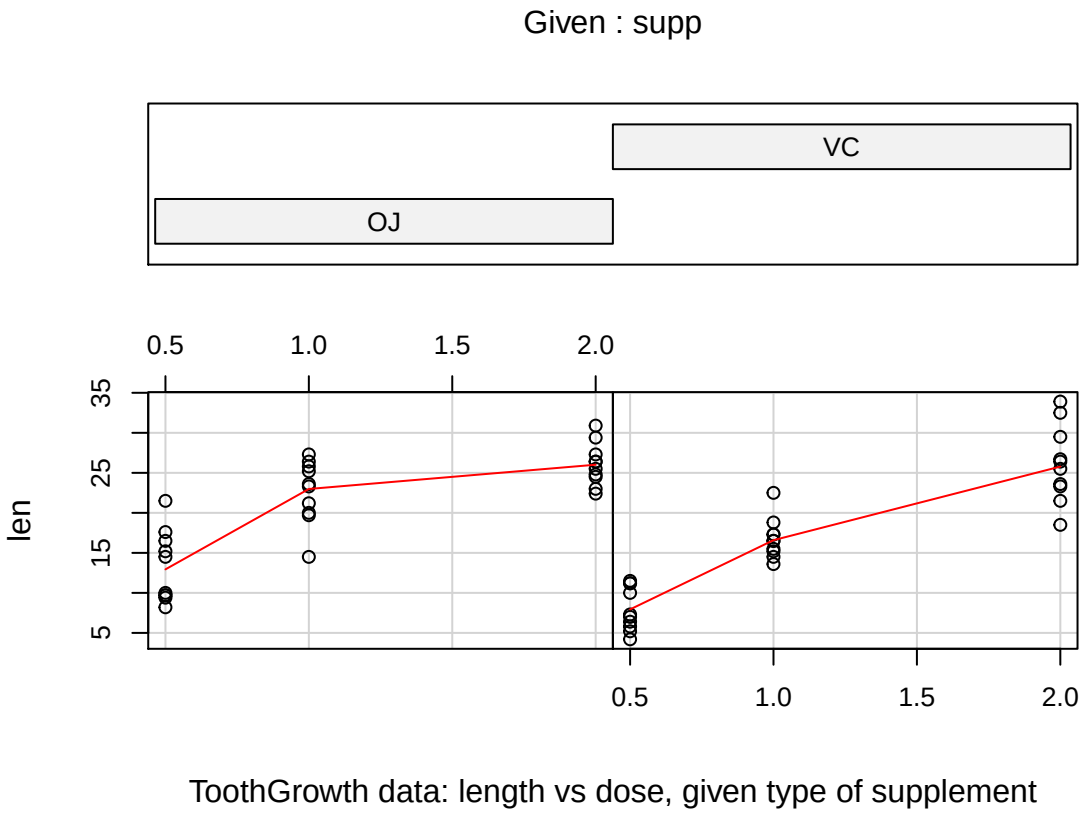
```
##       len        supp         dose
##  Min.   :22.40   OJ:10   Min.   :2
##  1st Qu.:24.57   VC: 0   1st Qu.:2
##  Median :25.95           Median :2
##  Mean   :26.06           Mean   :2
##  3rd Qu.:27.07           3rd Qu.:2
##  Max.   :30.90           Max.   :2
```

```r
summary(filter(dataset,supp=="VC",dose==2.0))
```

```
##       len        supp         dose
##  Min.   :18.50   OJ: 0   Min.   :2
##  1st Qu.:23.38   VC:10   1st Qu.:2
##  Median :25.95           Median :2
##  Mean   :26.14           Mean   :2
##  3rd Qu.:28.80           3rd Qu.:2
##  Max.   :33.90           Max.   :2
```

6. Plot showing lenght vs dose given the type of supplement.

```
coplot(len ~ dose | supp, data = dataset, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



4. Tests

1. Summary

Since we only have 30 obserbations, The stimations could not be so accurate to represent the entire population, so we are going to perform some tests.

We are going to use t.test in various subsets of the data to calculate the confidence interval and the p-value.

2. Assumptions

We will assume that: * Independence of the observations. * The distribution is nearly normal.

3. Supp influence in lenght

Null Hypothesis: Changing the supp does not affect the lenght.

Test:

```
test1 <- t.test(len~supp,dataset,var.equal = FALSE)
```

P-Value:

```
test1$p.value
```

```
## [1] 0.06063451
```

Confidence interval:

```
test1$conf.int[1:2]
```

```
## [1] -0.1710156  7.5710156
```

Since the confidence interval contains zero and p-value>0.05, we can't reject the null hypothesis.

4. Supp influence in lenght given a Dose level.

Null Hypothesis: Changing the supp does not affect the lenght given a dose level.

Test:

```
# We make a test with each dose level
test1 <- t.test(len~supp,filter(dataset,dose==0.5),var.equal = FALSE)
test2 <- t.test(len~supp,filter(dataset,dose==1.0),var.equal = FALSE)
test3 <- t.test(len~supp,filter(dataset,dose==2.0),var.equal = FALSE)
```

P-Value:

```
test1$p.value
```

```
## [1] 0.006358607
```

```
test2$p.value
```

```
## [1] 0.001038376
```

```
test3$p.value
```

```
## [1] 0.9638516
```

Confidence interval:

```
test1$conf.int[1:2]
```

```
## [1] 1.719057 8.780943
```

```
test2$conf.int[1:2]
```

```
## [1] 2.802148 9.057852
```

```
test3$conf.int[1:2]
```

```
## [1] -3.79807  3.63807
```

As we can see, we can reject the hyphotesis in the 2 first cases but not in the third case.

5. Dose influence in lenght

Null Hypothesis: Changing the dose does not affect the lenght.

Test:

```
#test with dose==0.5 or dose==1.0
test1 <- t.test(len~dose,filter(dataset,!dose==2.0),var.equal = FALSE)
#test with dose==1.0 or dose==2.0
test2 <- t.test(len~dose,filter(dataset,!dose==2.0),var.equal = FALSE)
```

P-Value:

```
test1$p.value
```

```
## [1] 1.268301e-07
```

```
test2$p.value
```

```
## [1] 1.268301e-07
```

Confidence interval:

```
test1$conf.int[1:2]
```

```
## [1] -11.983781  -6.276219
```

```
test2$conf.int[1:2]
```

```
## [1] -11.983781  -6.276219
```

In both cases, the confidence intervale don't cotains zero and the p-value is smaller than 0.05 so we can reject the null hyphotesis.

5 Conclusions

1. The supplement type alone doesn't affect the lenght.
2. The supplement type given a dose level affect the lenght if the dose level is 0.5 or 1.0 but don't affect the lenght if the dose level is 2.0.
3. The dose affects the lenght.