Statistical Inference Project - ToothGrowth Data

Miriam Pavón

11/03/2018

Overview

This project is to investigate the effect of Vitamin C on Tooth Growth in Guinea Pigs. 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).

Click on links here to quickly view tasks completed in this assignment:

1. Load the ToothGrowth data and perform some basic exploratory data analyses.
2. Provide a basic summary of the data.
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).
4. State your conclusions and the assumptions needed for your conclusions.

Data

The data for this assignment comes from C. I. Bliss (1952). The Statistics of Bioassay. Academic Press.

The data frame is composed of 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

Requirements

For this assignment you will need some specific tools

 RStudio: You will need RStudio to publish your completed analysis document to RPubs. You can also use RStudio to edit/write your analysis.

 knitr: You will need the knitr package in order to compile your R Markdown document and convert it to HTML

Before beginning the project, be sure to load the required R libraries and set any environmental variables. Note that setting messages in markdown to false suppresses messages from library loading such as version number and dependencies. Updating to latest versions of these libraries may improve ability to obtain results fairly similar to the steps outlined here.

```
# load libraries
library(ggplot2)
library(knitr)
library(datasets)
```

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

```r
# Load the data ToothGrowth
data(ToothGrowth)
# Look at the structure of the data
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 ...
```

```r
# Dimension of the data
dim(ToothGrowth)
```

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

```r
#Show the head of the data
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
```

Back to top

2. Provide a basic summary of the data.

Now, the summary of the data:

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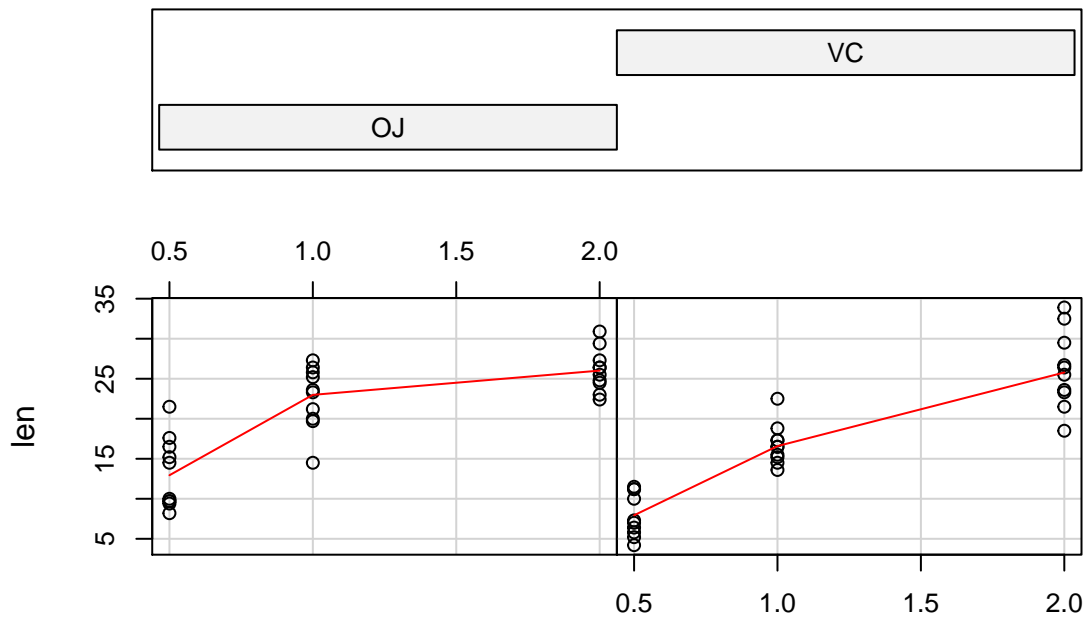
```r
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```
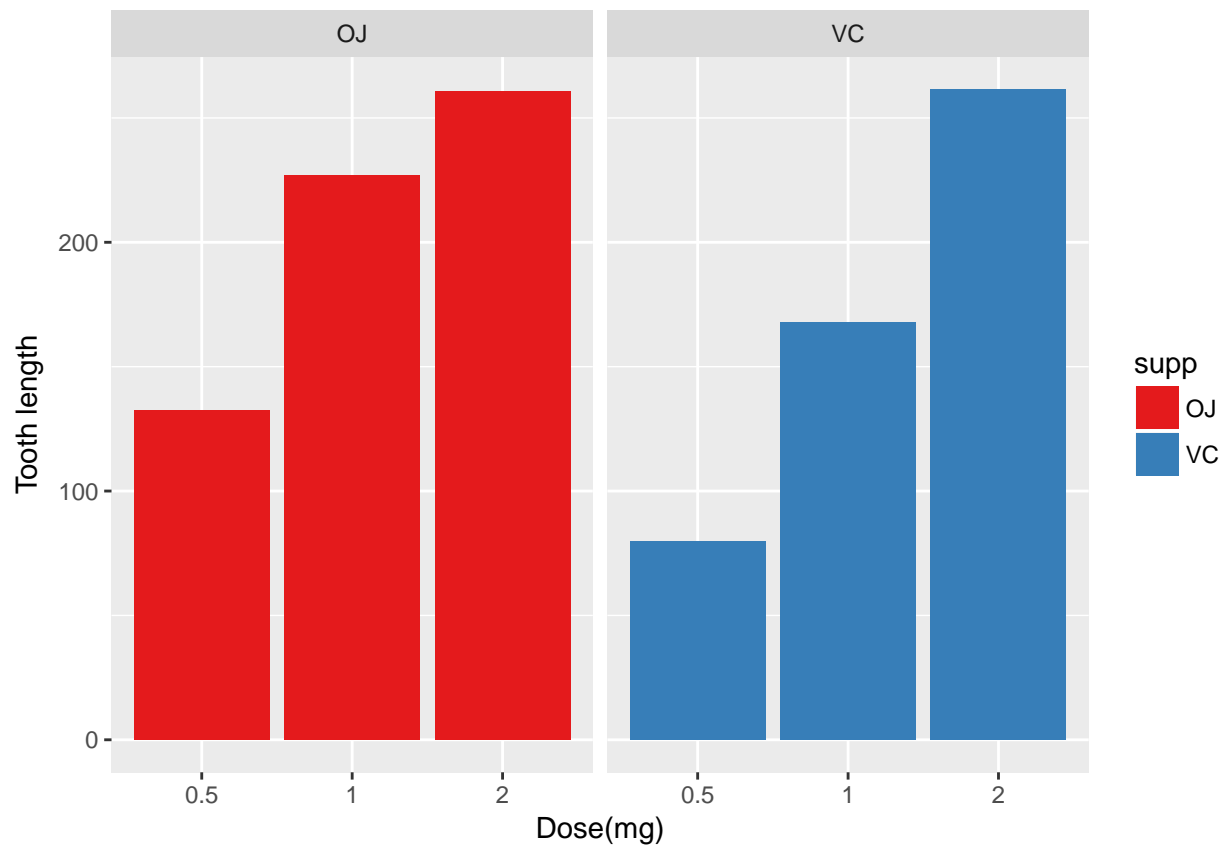
Given : supp

ToothGrowth data: length vs dose, given type of supplement

There are two types of supplement: OJ or VC. In the next figure, you can see the comparation of these two types of supplement.

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_bar(stat="identity") +
    scale_fill_brewer(palette="Set1") +
    facet_grid(. ~ supp) +
    xlab("Dose(mg)") +
    ylab("Tooth length")
```

Box plots are useful for identifying outliers and for comparing distributions. Figure provides a revealing summary of the data.

```
means <- aggregate(len ~  supp, ToothGrowth, mean)

ggplot(data=ToothGrowth, aes(x=supp, y=len, fill=supp)) + geom_boxplot() +
stat_summary(fun.y=mean, colour="darkred", geom="point", shape=18, size=3,show_guide = FALSE) + geom_tex
```
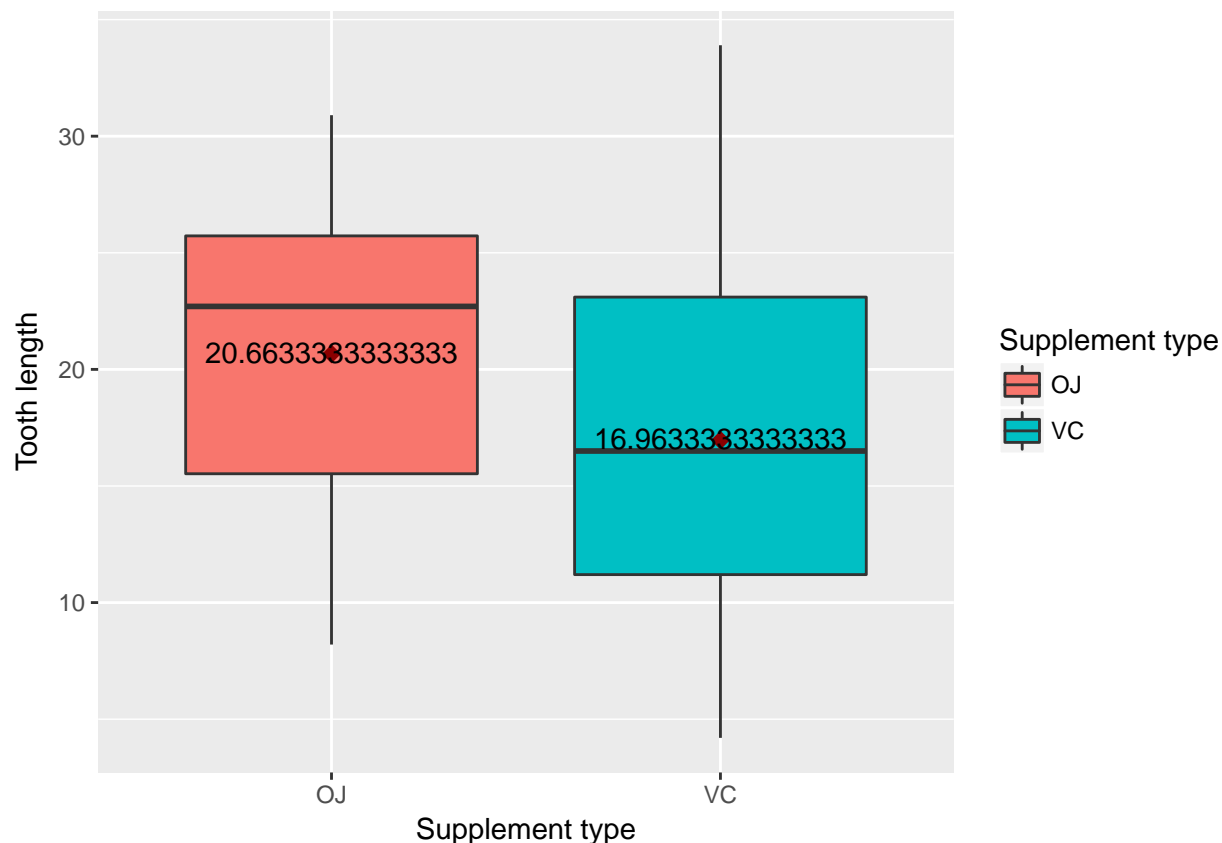
---

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).

3.1 Hypothesis tests to compare tooth growth by supp

```
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
```

So, as we can see, Null hypothesis can not be rejected as confindence intervals contain zero and p-value > 0.05 (0.06)

3.2 Hypothesis tests to compare tooth growth by dose levels

First, we divided the data in three subset:

```
ToothGrowth.doses_0.5_1.0 <- subset (ToothGrowth, dose %in% c(0.5, 1.0))
ToothGrowth.doses_0.5_2.0 <- subset (ToothGrowth, dose %in% c(0.5, 2.0))
ToothGrowth.doses_1.0_2.0 <- subset (ToothGrowth, dose %in% c(1.0, 2.0))
```

```
t.test(len ~ dose, data = ToothGrowth.doses_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
```

```
t.test(len ~ dose, data = ToothGrowth.doses_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
```

```
t.test(len ~ dose, data = ToothGrowth.doses_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
```

The p-value is all the cases is less than 0.05 and confindence intervals don't contain 0. We can reject the null hypothesis, this means that the tooth growth increase when the dose is increased.

Back to top

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

First, we can conclude that the difference in supplement type may not influence tooth growth. But, the dosage has an impact, the increase on the dosage seems to be have a direct impact with the tooth growth.

Back to top