Course Project - Basic inferential data analysis

Rafael Lavagna

4 de abril de 2018

Overwiew

The objective of this project is to analize the ToothGrow data in the R datasets package. This dataset contains information about the effect of vitamin C on tooth growth in Guinea Pigs.

Description of the data

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).

Summary of the data

```
library(ggplot2)
library(dplyr)
data("ToothGrowth")
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
```

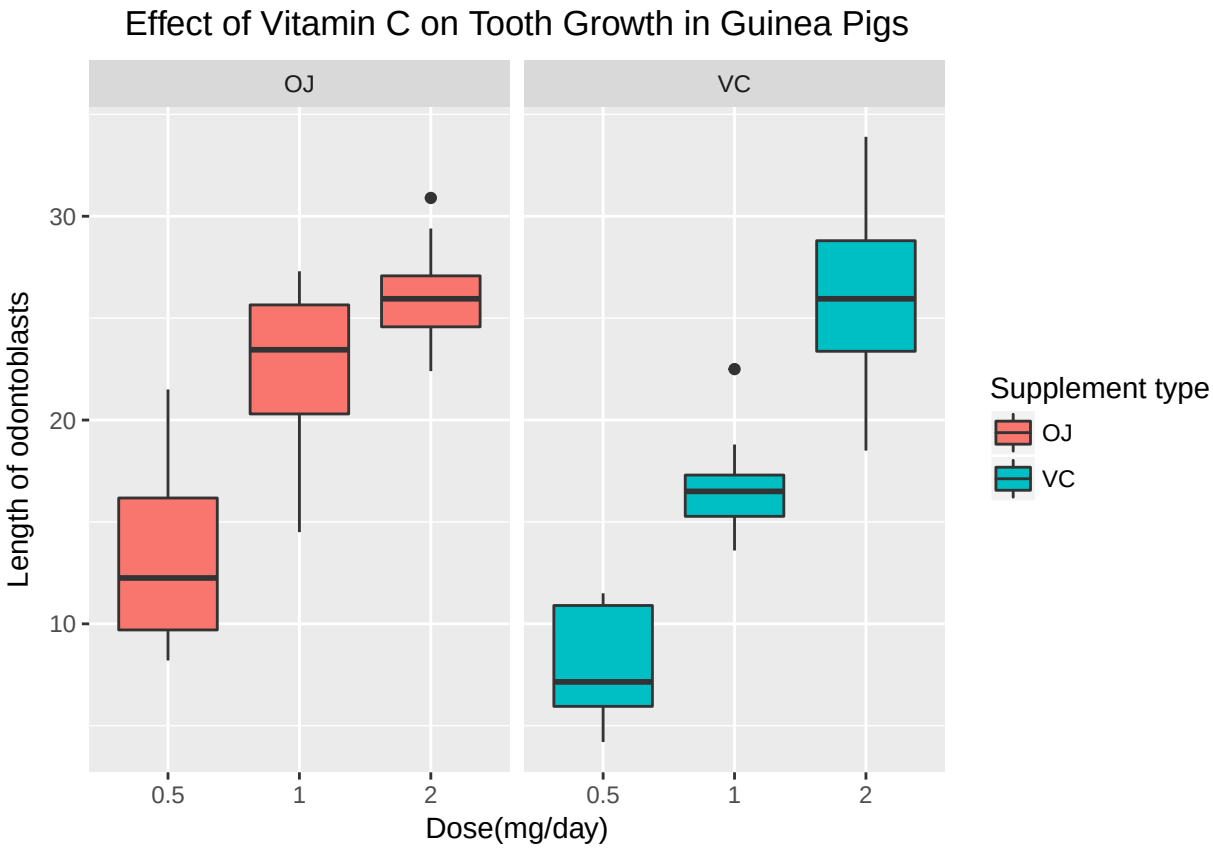
```
str(ToothGrowth)
```

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

Exploratory Data Analyisis

In order to show the variation of the tooth length as the dose of the different supplements increase we are going to transform the dose variable into a factor one. Afterwards, we are going to build a boxplot showing the mentioned variation.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
g <- ggplot(ToothGrowth, aes(x=dose, y=len)) + geom_boxplot(aes(fill=supp))+
facet_grid(.~supp)+ggtitle("Effect of Vitamin C on Tooth Growth in Guinea Pigs")+xlab("Dose(mg/day)")+
ylab("Length of odontoblasts") + theme(plot.title = element_text(hjust = 0.5))+
labs(fill="Supplement type")
g
```



At first glance, it seems that for both supplements, increasing the dose results in an increase in the length of the odontoblasts (i.e. tooth grow). Another thing that we can observe in the boxplot is that for dosis lower than 2 mg/day the orange juice supplement seems to have a stronger impact than the ascorbic acid. However, for a 2 mg/day dose, the length of both groups appers to be the same. In order to verify this observations we are going to perform different hypothesis tests to draw firmer conclusions.

Statistical inference

First of all, we are going to build 6 subsets of data. Each of them will contain the observations of the different dosis for each supplement.

```
OJ_0.5 <- subset(ToothGrowth,dose=="0.5" & supp=="OJ")
OJ_1 <- subset(ToothGrowth,dose=="1" & supp=="OJ")
OJ_2 <- subset(ToothGrowth,dose=="2" & supp=="OJ")
VC_0.5 <- subset(ToothGrowth,dose=="0.5" & supp=="VC")
VC_1 <- subset(ToothGrowth,dose=="1" & supp=="VC")
VC_2 <- subset(ToothGrowth,dose=="2" & supp=="VC")
```

We are going to perform some hypothesis tests to compare the means of the different groups to infer whether the difference between their means is statistically significant or not.

In order to perform the corresponding hypothesis tests we are going to assume that the data are iid normal (that is independent and identical distributed). We are also going to consider that variances between populations are not equal.

First of all, we are going to see if increasing the dose of each supplement results in an increase of the tooth length. We consider the null hypothesis: difference in means equal to 0; and the alternative hypothesis: true

difference in means is greater than 0.

```
t.test(OJ_1$len,OJ_0.5$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 4.39246e-05
```

```
t.test(OJ_2$len,OJ_0.5$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 6.618919e-07
```

```
t.test(OJ_2$len,OJ_1$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 0.01959757
```

```
t.test(VC_1$len,VC_0.5$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 3.405509e-07
```

```
t.test(VC_2$len,VC_0.5$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 2.340789e-08
```

```
t.test(VC_2$len,VC_1$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 4.577802e-05
```

In every case we obtain p-values lower than 0.05. This means that we have enough statistical evidence to reject the null hypothesis. In other words, in every case, the probability of commiting an error when saying that increasing the dose of any of the supplements considered results in an increase in the length of the odontoblasts is lower than 0.05.

We are now going to analyse if any of the supplements has a stronger impact on the tooth grow than the other one. We will perform a test for each dose.

```
t.test(OJ_0.5$len,VC_0.5$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 0.003179303
```

```
t.test(OJ_1$len,VC_1$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 0.0005191879
```

```
t.test(VC_2$len,OJ_2$len,paired = FALSE,var.equal = FALSE,alternative = "greater")$p.value
```

```
## [1] 0.4819258
```

From the results obtained we can infer that there is enough statistical evidence to reject the null hypothesis and say that for a daily dose of 0.5 and 1 mg the orange juice has a stronger impact on the tooth grow than the ascorbic acid (p-values of 0.003179 and 0.0005192 respectively). However, for a daily dose of 2 mg, we obtain a p-value of 0.4819 and so we fail to reject the null hypothesis. This means that we do not have enough evidence to infer that any of the supplements is better than the other one when the daily dose is 2 mg.

Conclusions

Summing up the results presented above, we can draw the following conclusions from the ToothGrow data:

- Increasing the dose of vitamin C leads to an increase in the length of odontoblasts, no matter which supplement is being supplied
- For a daily dose lower than 2 mg the orange juice supplement administration results in a higher increase in the length of odontoblasts
- For a daily dose of 2 mg there is no statistical evidence to determine if any of the supplements has a stronger imapct on the increase of the length of odontoblasts than the other one.