Tooth Growth Dataset Analysis

Carlos Martinez Reyes

4/10/2020

Part 2 - Basic Inferential Data Analysis

Overview

In the second part of the project, we want to analyze the effectiveness of vitamin c on teeth growth in guinea pigs. We'll do this by using t-tests, comparing teeth length by supplement type and dose level.

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

```
# Load the Data
library(dplyr, warn.conflicts = F)
library(ggplot2)
library(datasets)
data(ToothGrowth)

# Explore the dataset
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 ...
```

2. Basic summary of the data

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
# Variables
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  
```

The statistical summary shows that average length of tooth growth over all the observations is 18.81mm between the range of 4.2mm and 33.9mm.

```
# The  variable dose contains 3 values: 0.5, 1, 2. Convert dose to factor
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

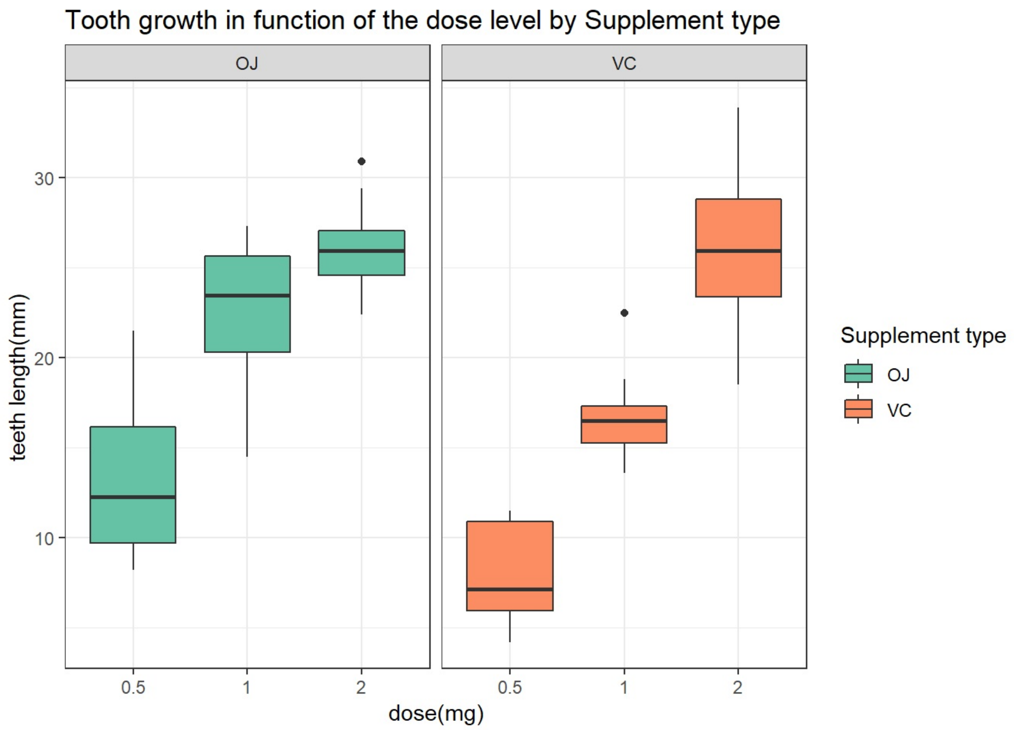
```
# Analysis and visualization of tooth growth in function of the dose level

table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##      
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

```
ToothGrowth %>%


ggplot(aes(x=dose, y=len, fill = supp)) +
geom_boxplot() +
facet_grid(. ~ supp) +
scale_fill_brewer(palette = "Set2") +
theme_bw() +
ggtitle("Tooth growth in function of the dose level by Supplement type") +
labs(x="dose(mg)", y= "teeth length(mm) ") +
guides(fill=guide_legend(title="Supplement type"))
```




```
#Supplement type mean and range:
suppmean = split(ToothGrowth$len, ToothGrowth$supp)
sapply(suppmean, mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

```
sapply(suppmean, range)
```

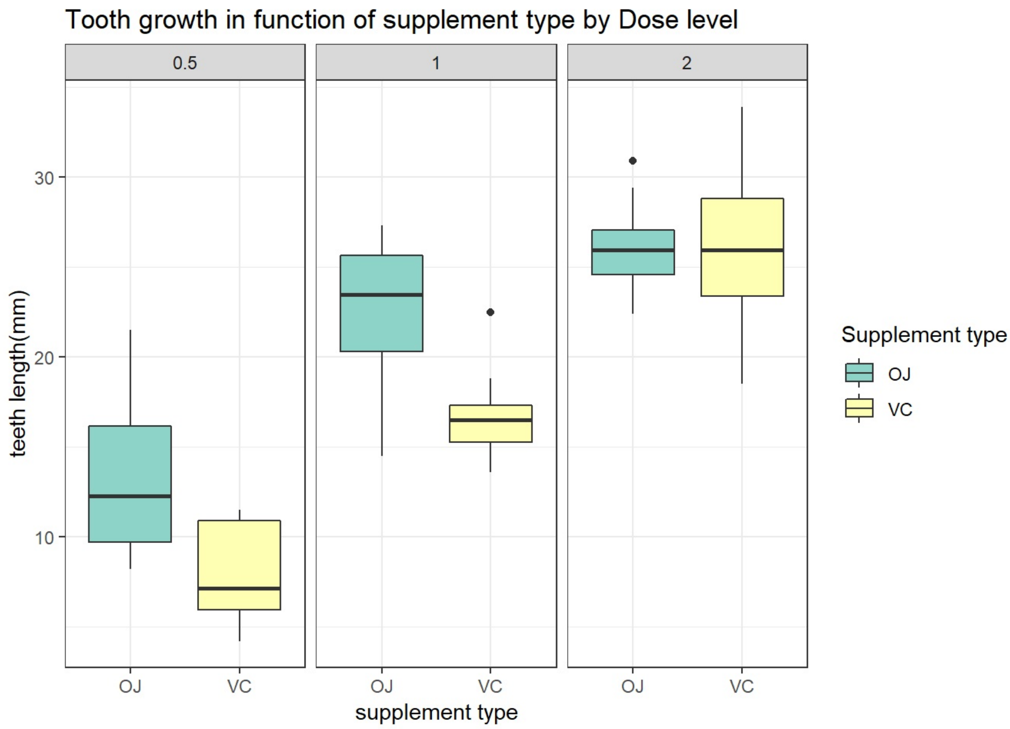
```
##        OJ   VC
## [1,]  8.2  4.2
## [2,] 30.9 33.9
```

```
# visualization of tooth growth in function of supplement type

table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

```
ToothGrowth %>%
ggplot(aes(x = supp, y = len)) +
geom_boxplot(aes(fill = supp)) +
facet_wrap(~ dose) +
scale_fill_brewer(palette = "Set3") +
theme_bw() +
ggtitle("Tooth growth in function of supplement type by Dose level") +
labs(x="supplement type", y= "teeth length(mm)") +
guides(fill=guide_legend(title="Supplement type"))
```




```
#Dosage level mean and range:
dosmean = split(ToothGrowth$len, ToothGrowth$dose)
sapply(dosmean, mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

```
sapply(dosmean, range)
```

```
##       0.5    1    2
## [1,]  4.2 13.6 18.5
## [2,] 21.5 27.3 33.9
```

In summary, when comparing the box plots, the means and their ranges, it is possible to observe that the dose level influences tooth growth more significantly than the type of supplement. To test this claim, we are going to perform hypothesis tests.

3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

From the conclusions at the end of the previous section we are going to perform a two-sample t-test for the supplement factor and a two-sample t-test for each possible pair of the 3 levels of the dose factor, in total we will run 4 t-tests.

a). Test 1: dose = 0.5 and dose = 1

```
len_a <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(len) %>% unlist()
dose_a <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(dose) %>% unlist()

#Test
(Test.a <- t.test(len_a~dose_a, paired = FALSE))
```

```
## 
##  Welch Two Sample t-test
## 
## data:  len_a by dose_a
## 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
```

b). Test 2: dose = 0.5 and dose = 2

```
len_b <- ToothGrowth %>% filter(dose %in% c(0.5,2)) %>% select(len) %>% unlist()
dose_b <- ToothGrowth %>% filter(dose %in% c(0.5, 2)) %>% select(dose) %>% unlist()

#Test
(Test.b <- t.test(len_b~dose_b, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_b by dose_b
## 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
```

c). Test 3: dose = 1 and dose = 2

```
len_c <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(len) %>% unlist()
dose_c <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(dose) %>% unlist()
#Test c
(Test.c <- t.test(len_c~dose_c, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_c by dose_c
## 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
```

Testing all possible combinations of levels of the dose factor, we obtain that in all cases the p-value is less than the significance level 0.05. Therefore, we reject the null hypothesis Ho. In other words, there is a significant difference in average tooth length due to the dosage level.

d). Test 4: by Supplement type

```
len <- ToothGrowth %>% select(len) %>% unlist()
supp <- ToothGrowth %>% select(supp) %>% unlist()
#Test
t.test(len~supp, paired=F)
```

```
##
##  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
```

We can see that the p-value of the test is 0.06 and since it is greater than 0.05, the null hypothesis Ho is accepted, that is, there is not enough evidence to reject it, so we cannot affirm that on average the levels of supplements have different impact on tooth growth. In other words, there is no statistically significant difference between them.

4. Conclusions and the assumptions.

i). The samples used are random and iid.

ii). The population distribution of each sample should be approximately normal.

If the above assumptions are met, we can conclude the following: There is a statistically significant difference between the length of the teeth and the dose, so we can expect that as the dose increases, the length of the teeth also increases. On the other hand, there does not appear to be a statistically significant difference between the methods of administration.