Part 2: Basic Inferential Data Analysis

Kumar Shaket

28/12/2020

## Load the ToothGrowth Data

library(datasets)  
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 ...

data("ToothGrowth")  
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

## Summary

#Identifying if there is any NA value in dataset  
sum(!complete.cases(ToothGrowth))

## [1] 0

This suggest there is no NA values in the dataset

#Computing summary of ToothGrowth Dataset  
summary(ToothGrowth)

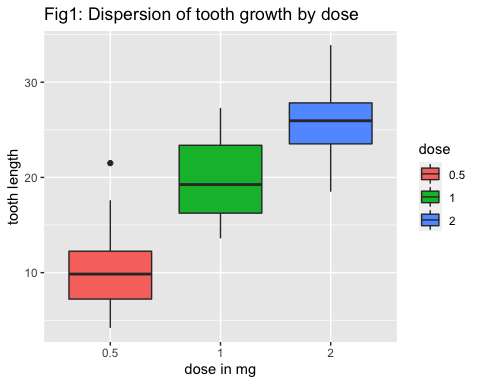
## len supp dose   
## Min. : 4.20 OJ:30 Min. :0.500   
## 1st Qu.:13.07 VC:30 1st Qu.:0.500   
## Median :19.25 Median :1.000   
## Mean :18.81 Mean :1.167   
## 3rd Qu.:25.27 3rd Qu.:2.000   
## Max. :33.90 Max. :2.000

##Exploratory Analysis

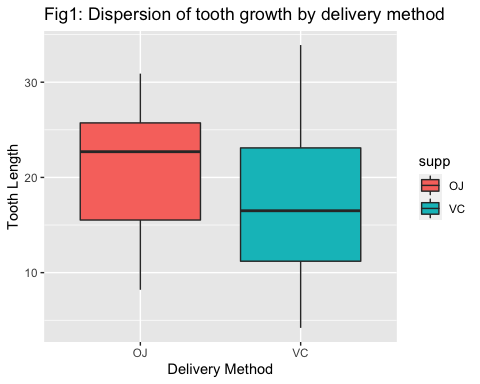
library(ggplot2)

#Computing dose factor  
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

#Plotting boxplot to understand how dose impact tooth length  
p <- ggplot(ToothGrowth,aes(x=dose,y=len,fill=dose))+ geom\_boxplot() + ggtitle("Fig1: Dispersion of tooth growth by dose")+xlab("dose in mg")+ylab("tooth length")  
p

 We can see that the higher the dose is, the longer the teeth are. We can notice that for a dose of 1 mg, the mean is nearly twice than for dose 0.5 mg. The progression is then lower when the dose is 2 mg. The position of the boxes are really different; this is a first clue for suggesting that the tooth length depends a lot of the dose.

#plotting boxplot of delivery method of these doses impacting toothlength  
p1 <- ggplot(ToothGrowth,aes(x=supp,y=len,fill=supp))+ geom\_boxplot() + ggtitle("Fig1: Dispersion of tooth growth by delivery method")+xlab("Delivery Method")+ylab("Tooth Length")  
p1

 The boxes are quite similar. However, the median is much higher for Orange Juice than for Vitamin C; that means that for Orange Juice, the high values are more numerous than the low values, in comparison with Vitamin C. ## Hypothesis Tests we will test if the delivery mode has an influence on the tooth growth. nul hypothesis H0 could be formulated as follows :

#H0 : The delivery mode of Vitamin C does not have any influence on the tooth growth

dose <-ToothGrowth$dose  
supply <- ToothGrowth$supp  
len <- ToothGrowth$len  
t.test(len[supply=="VC"],len[supply=="OJ"],paired = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: len[supply == "VC"] and len[supply == "OJ"]  
## 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:  
## -7.5710156 0.1710156  
## sample estimates:  
## mean of x mean of y   
## 16.96333 20.66333

This first test shows the following :

* The p-value is 0.06, i.e. nearly the significance level α = 0.05.
* So we do not reject the null hypothesis, but as 0.05≦ p-value ≦0.1, it is not clearly obvious that we can reject the null hypothesis.
* The confidence interval contains 0, so the test is not really significant.

Now let’s try to test the influence of the dose on the tooth growth

t.test(len[dose==0.5],len[dose==1],paired = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: len[dose == 0.5] and len[dose == 1]  
## 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 of x mean of y   
## 10.605 19.735

This second test shows the following :

* The p-value is nearly 0 and hence we can reject H0
* The confidence interval doesn’t contain 0

An identical conclusion can be taken comparing the length of dose = 1 and dose = 2. This could already be detected from the boxplot above.

## Conclusion

* The dose of Vitamin C is clearly a factor of growth of the teeth.
* The delivery mode does not have much impact on the teeth growth.