

ToothGrowth Inferential Analysis

By: Syed Musharraf Ali

Overview

In this dataset we are given 3 variables len,supp,dose. Where as **len** is a numeric variable, **supp** is a character variable which has two categories VC , OJ and **dose** is a numerical variable which have 3 different categories 0.5 , 1 , 2

In this project eploratory data analysis is done by compairing **len** with **dose** and **supp** and after that hypothesis testing is done to determine whether this differences are by chance or are actually there.

Assumptions

In the dataset there are total 60 observation. These observations are further divided into six part on the basis of supp and dose. For example, there are 10 observations for supp = VC and dose = 0.5 . we are assuming that each observations are independent and because each group have 10 observations we will use t test for comparison. We are also assuming that each group have different variance and significance level of 5%.

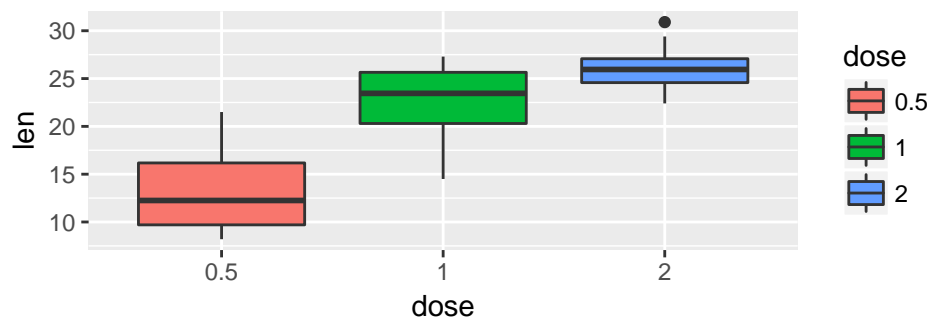
Comparison between doses

To check whether different dose have any effect on len , we will do hypothesis t test.our null hypothesis is that their is no different in len because of the change in dose. whereas our alternative hypothesis is that there is different in len because of change in dose

comparison of dose within supp “OJ”

From the below figure we see that due to different doses there is a different in len.

fig 1: Comparison of doses of OJ



Now checking whether this is by chance or not

by compairing dose of 0.5 and 1 we get

```
t.test(oj0.5,oj1,paired = FALSE,var.equal = FALSE)$p.value
```

```
## [1] 8.784919e-05
```

Since the P-value is less than 5% therefore we will reject our null hypothesis. Similarly for comaprison between dose of 1 and 2 we get

```
## [1] 0.03919514
```

and for dose 0.5 and 2 we have,

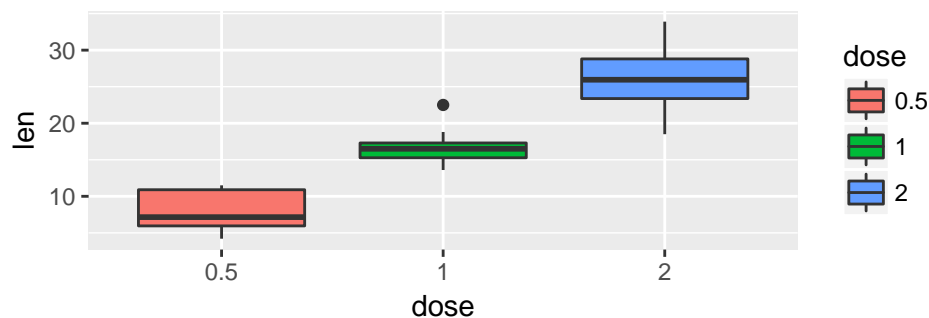
```
## [1] 1.323784e-06
```

Since all of the p-value are less than 0.05 there for we reject our null hypothesis and conclude that different doses have different effect on length when using supp **OJ**.

comparison of dose within supp “VC”

From the below figure we see that due to different doses there is a different in len.

fig 2: Comparison of doses of VC



Now checking whether this is by chance or not

by compairing dose of 0.5 and 1 we get

```
t.test(vc0.5,vc1,paired = FALSE,var.equal = FALSE)$p.value
```

```
## [1] 6.811018e-07
```

Since the P-value is less than 5% therefore we will reject our null hypothesis. Similarly for comparison between dose of 1 and 2 we get

```
## [1] 9.155603e-05
```

and for dose 0.5 and 2 we have,

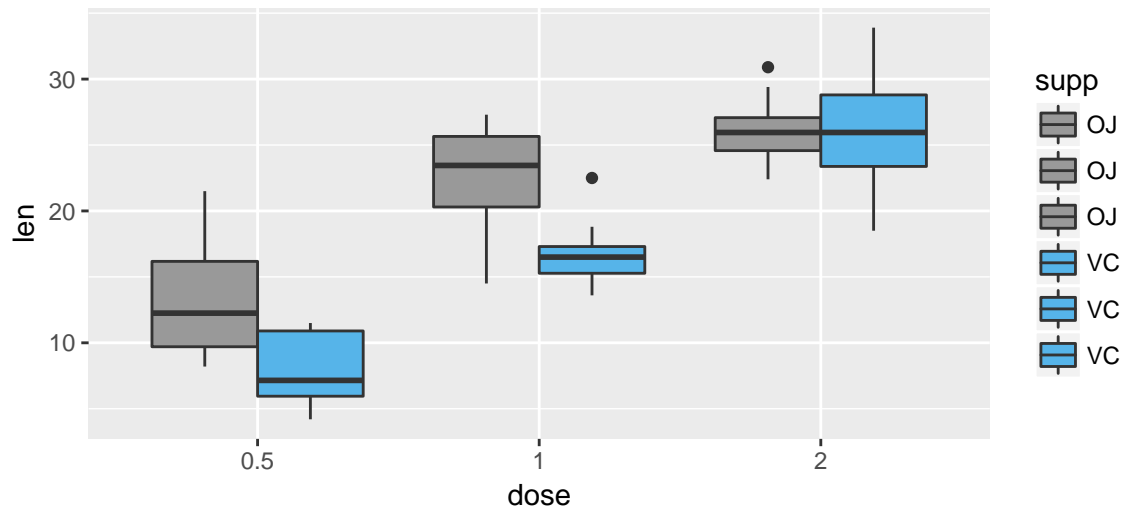
```
## [1] 4.681577e-08
```

Since all of the p-value are less than 0.05 there for we reject our null hypothesis and conclude that different doses have different effect on length when using supp **VC**.

Finding effect on length when different supp is used

From the exploratory data analysis we see that there is difference between OJ and VC. It shows that the median of OJ is greater than VC when the dose is 0.5 and 1

fig 3: Comparison of OJ and VC



So our null hypothesis is that there is no difference in dose of 0.5 when we use either **OJ** or **VC**. similarly same for the dose of 1.0.

Doing t test for comparing VC and OJ when Dose of 0.5 is used.

```
t.test(oj0.5,vc0.5,paired = FALSE,var.equal = FALSE)$p.value
```

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

Now doing t test for comparing VC and OJ when dose of 1.0 is used

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

Since both of the P values are less than 5% ,we reject the null hypothesis and conclude that the difference in length do occur when different kind of **supp** is used for the quantity of dose of 0.5 and 1.0.

APPENDIX

Library used

ggplot2,dplyr

fig 1

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
ToothGrowth %>% filter(supp=="OJ")%>% ggplot(aes(dose,len,group=dose))+
geom_boxplot(aes(fill=dose))
+ggtitle("fig 1: Comparison of doses of OJ")
```

fig 2

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
ToothGrowth %>% filter(supp=="VC")%>% ggplot(aes(dose,len,group=dose))+
```

```
geom_boxplot(aes(fill=dose))
+ggtitle("fig 2: Comparison of doses of VC")
```

fig 3

```
xz<-ToothGrowth
xz$supp<-as.character(xz$supp)
for( i in 1:length(xz$len))
{

  if(xz$supp[i]=="VC"){
    if((xz$dose[i]==0.5))
      xz$supp[i]="VC_0.5"

    if(xz$dose[i]==1)
      xz$supp[i]="VC_1"

    if(xz$dose[i]==2)
      xz$supp[i]="VC_2"

  }
  if(xz$supp[i]=="OJ"){
    if((xz$dose[i]==0.5))
      xz$supp[i]="OJ_0.5"

    if(xz$dose[i]==1)
      xz$supp[i]="OJ_1"

    if(xz$dose[i]==2)
      xz$supp[i]="OJ_2"

  }}

xz %>% ggplot(aes(dose,len,group=supp))+geom_boxplot(aes(fill=supp))

xz %>% ggplot(aes(dose,len,fill=supp))
+geom_boxplot(aes(fill=supp))
+scale_fill_manual(labels = c("OJ","OJ","OJ","VC","VC","VC"),
  values=c("#999999","999999","999999", "#56B4E9","56B4E9","56B4E9"))
+ggtitle("fig 3:Comparison of OJ and VC")
```

seperating data

```
oj0.5<-ToothGrowth %>% filter(supp=="OJ" & dose==0.5) %>% .$len
oj1<-ToothGrowth %>% filter(supp=="OJ" & dose==1)%>% .$len
oj2<-ToothGrowth %>% filter(supp=="OJ" & dose==2)%>% .$len
vc0.5<-ToothGrowth %>% filter(supp=="VC" & dose==0.5) %>% .$len
vc1<-ToothGrowth %>% filter(supp=="VC" & dose==1)%>% .$len
vc2<-ToothGrowth %>% filter(supp=="VC" & dose==2)%>% .$len
```

T tests

```
t.test(oj0.5,oj1,paired = FALSE,var.equal = FALSE)$p.value
t.test(oj1,oj2,paired = FALSE,var.equal = FALSE)$p.value
t.test(oj0.5,oj2,paired = FALSE,var.equal = FALSE)$p.value
t.test(vc0.5,vc1,paired = FALSE,var.equal = FALSE)$p.value
t.test(vc1,vc2,paired = FALSE,var.equal = FALSE)$p.value
t.test(vc0.5,vc2,paired = FALSE,var.equal = FALSE)$p.value
t.test(oj0.5,vc0.5,paired = FALSE,var.equal = FALSE)$p.value
t.test(oj1,vc1,paired = FALSE,var.equal = FALSE)$p.value
```