Statistical Inference Course Project Part2

Kaaltho

20/12/2020

Synopsis

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

Stages of the proyect.

1.Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data.

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

Part 2: Basic Inferential Data Analysis Instructions

The first step is to simulate the variables that we want to compare.

```
#Oppening libraries
library(ggplot2)
library(cowplot)

#Assigning variables names, the specified data is the following:
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
```

Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data.

The next plot contains the distribution of the variables and it is possible to see if it follows some distribution.

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)

mean.supp = split(ToothGrowth$len, ToothGrowth$supp)
sapply(mean.supp, mean)
```

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

```r
mean.dose = split(ToothGrowth$len, ToothGrowth$dose)
sapply(mean.dose, mean)
```

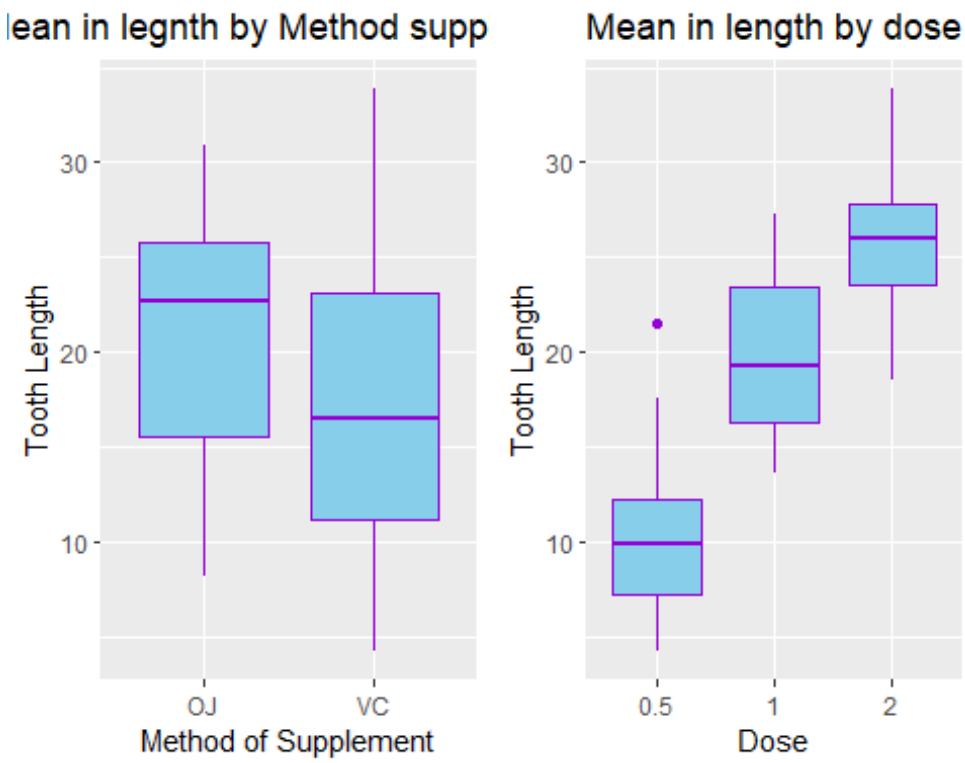
```
##    0.5      1      2
## 10.605 19.735 26.100
```

```r
#Ploting Type of Supplement versus the Tooth Length
#Changing as a factor
ToothGrowth$dose<-as.factor(ToothGrowth$dose)

plot1 <- ggplot(ToothGrowth, aes(x=supp, y=len)) +
  geom_boxplot(colour="darkviolet", fill="skyblue")+
  guides(fill=FALSE) +
  labs(x="Method of Supplement", y="Tooth Length", title=("Mean in legnth by Method
supplement"))+
  theme(plot.title = element_text(hjust = 0.5))


plot2 <- ggplot(ToothGrowth, aes(x=dose, y=len)) +
  geom_boxplot(colour="darkviolet", fill="skyblue")+
  guides(fill=FALSE) +
  labs(x="Dose", y="Tooth Length", title=("Mean in length by dose"))+
  theme(plot.title = element_text(hjust = 0.5))

#This plot is just one image with two plots on it
plot_grid(plot1, plot2, align = "v", ncol =2, rel_heights = c(1/2, 1/2))
```



As can be seen in the graph the distribution of the data of Method of supplement OJ the mean is 20.63 while the mean of the Method of supplement VC correspond to 16.93 length of teeth, apparently supplement OJ has more Tooth Length measure than the Supplement VC. Considering the distribution of Tooth Length relative to dose, it is possible to see that a Dose of 2 have better Tooth length. Could be possible to infer that an OJ method + 2 type dose could drive the results to a better tooth development.

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)

Comparing tooth growth by supplement using a t-test.

```
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
```

Due to the fact that the p-value is greater than 0.05 and the confidence interval contains zero, it could be said that the supplement types does not have a impact on Tooth growth.

Testing the tooth length by the group with vitamin C dosage

```
t.test(ToothGrowth$len[ToothGrowth$dose==2],ToothGrowth$len[ToothGrowth$dose==1], paired
= FALSE, var.equal = TRUE)

##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose ==
1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

State your conclusions and the assumptions needed for your conclusions.

Seeing that the p-value is approximately zero, can be said that there is evidence to reject the null hypothesis. It is possible to say that higher doses result in greater tooth growth.

Can be concluded that the Method of supplement doesn't have an noticeable impact on tooth growth, furthermore dose level appear to have an influence over tooth growth.