statistical inference : assignment 2

vincent trouillet

7 mars 2018

Overview

The purpose of the this data analysis is to analyze the ToothGrowth data set by comparing the guinea tooth growth by supplement and dose. A rapid exploratory data analysis on the data set and a comparison with confidence intervals will be made in order to make conclusions about the tooth growth.

1) The data : basic exploratory data analyses

```
library(datasets)
library(ggplot2)

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
```

```
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 ...
```

```
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
```

Total Observations are 60 with 3 columns - len, supp, dose

Tooth Length (len) is of type numeric values.

Supplementary (supp) is of type OJ(orange juice) and VC(vitamin C)

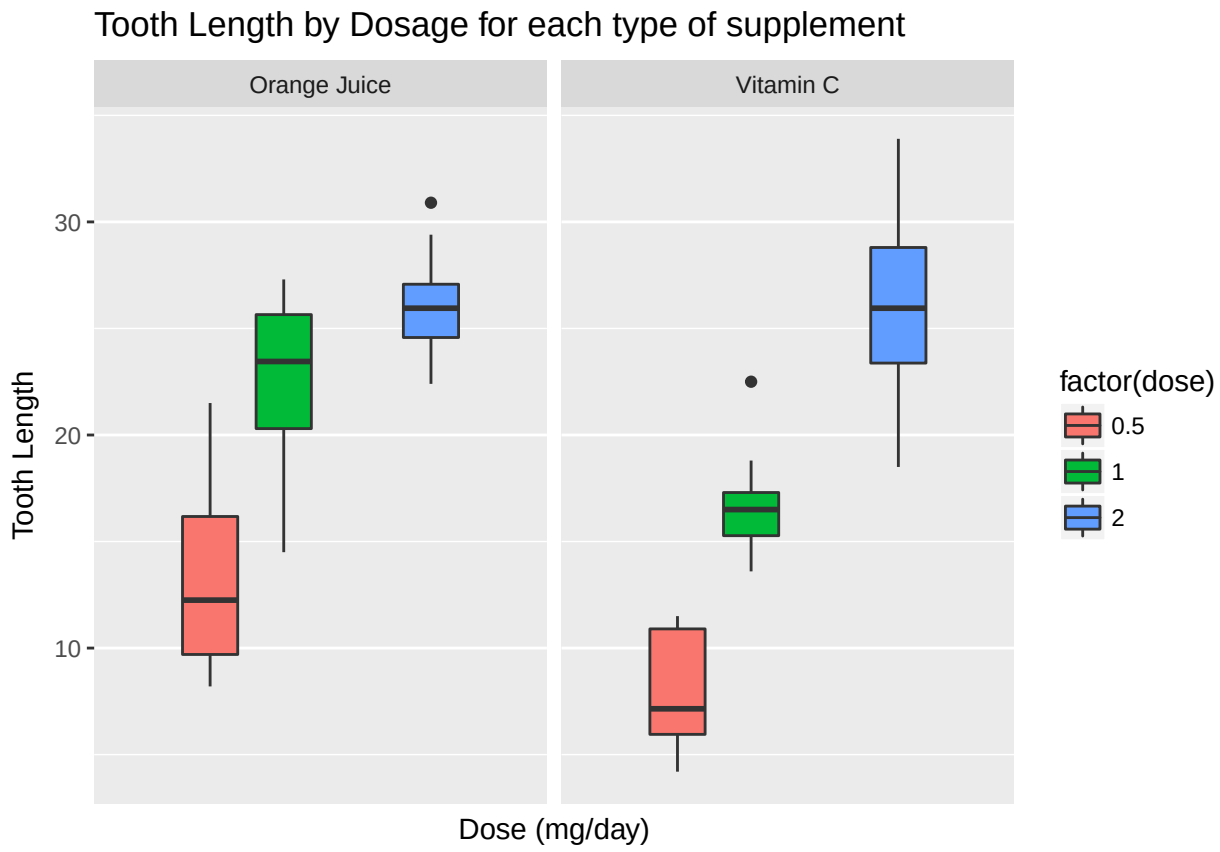
Dosage (dose) is of 3 Types - 0.5, 1 and 2

3) Comparison of tooth growth by supp and dose

use boxplot for a graphical comparison

```
levels(ToothGrowth$supp)<-c("Orange Juice","Vitamin C")
plot.bxpl <- ggplot(ToothGrowth, aes(x=dose,y=len,fill=factor(dose))) +
    facet_grid(.~supp)+ geom_boxplot() +
    scale_x_discrete("Dose (mg/day)") +
    scale_y_continuous("Tooth Length") +
    ggtitle("Tooth Length by Dosage for each type of supplement")

plot.bxpl
```



We can see that there is a relation between dose and tooth growth. Orange Juice seems more effective for dosages 0.5 & 1.0. For a dose amount 2.0 mg/day, both supplement look equivalent.

let's explore hypothesis.

We assume that no other unmeasured factors are affecting tooth length and the tooth lengths follow a normal distrubution.

Hypothese 1 : orange juice & vitamin C deliver the same tooth growth across the dataset

```
hypoth1 <- t.test(len ~ supp, data = ToothGrowth)
hypoth1
```

```
##
##  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 Orange Juice    mean in group Vitamin C
##                   20.66333                   16.96333
```

The confidence intervals includes 0 and the p-value is greater than the threshold of 0.05. The null hypothesis cannot be rejected.

Hypothese 2 : orange juice & vitamin C deliver the same tooth growth for a dosage of 0.5 mg/day

```
hypoth2<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 0.5))
hypoth2
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group Orange Juice    mean in group Vitamin C
##                      13.23                      7.98
```

The confidence interval does not include 0 and the p-value is below the 0.05 threshold. The null hypothesis can be rejected. The alternative hypothesis that 0.5 mg/day dosage of orange juice delivers more tooth growth than ascorbic acid is accepted.

Hypothese 3 : orange juice & vitamin C deliver the same tooth growth for a dosage of 1 mg/day

```
hypoth3<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 1))
```

The confidence interval does not include 0 and the p-value is smaller than the 0.05 threshold. The null hypothesis can be rejected. The alternative hypothesis that 1 mg/day dosage of orange juice delivers more tooth growth than ascorbic acid is accepted.

Hypothese 4 : orange juice & vitamin C deliver the same tooth growth for a dosage of 2 mg/day

```
hypoth4<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 2))
hypoth4
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group Orange Juice    mean in group Vitamin C
##                      26.06                      26.14
```

The confidence interval does include 0 and the p-value is larger than the 0.05 threshold. The null hypothesis cannot be rejected.

4) Conclusions

Orange juice delivers more tooth growth than ascorbic acid for dosages 0.5 & 1.0. Orange juice and ascorbic acid deliver the same amount of tooth growth for dose amount 2.0 mg/day. For the entire data set we cannot conclude orange juice is more effective that ascorbic acid.