Statistical Inference Project Part 2

Andres P

29/01/2019

Analyze the ToothGrowth data in the R datasets package

required package:

```
library(ggplot2)
```

Load the ToothGrowth data and perform some basic exploratory data analyses

```
# Load the assignment data ToothGrowth
data(ToothGrowth)
# Look at the structure of the data
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 ...
```

```
# Investigate first 6 rows of the data
head(ToothGrowth, 6)
```

```
##    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
```

Provide a basic Summary of the Data

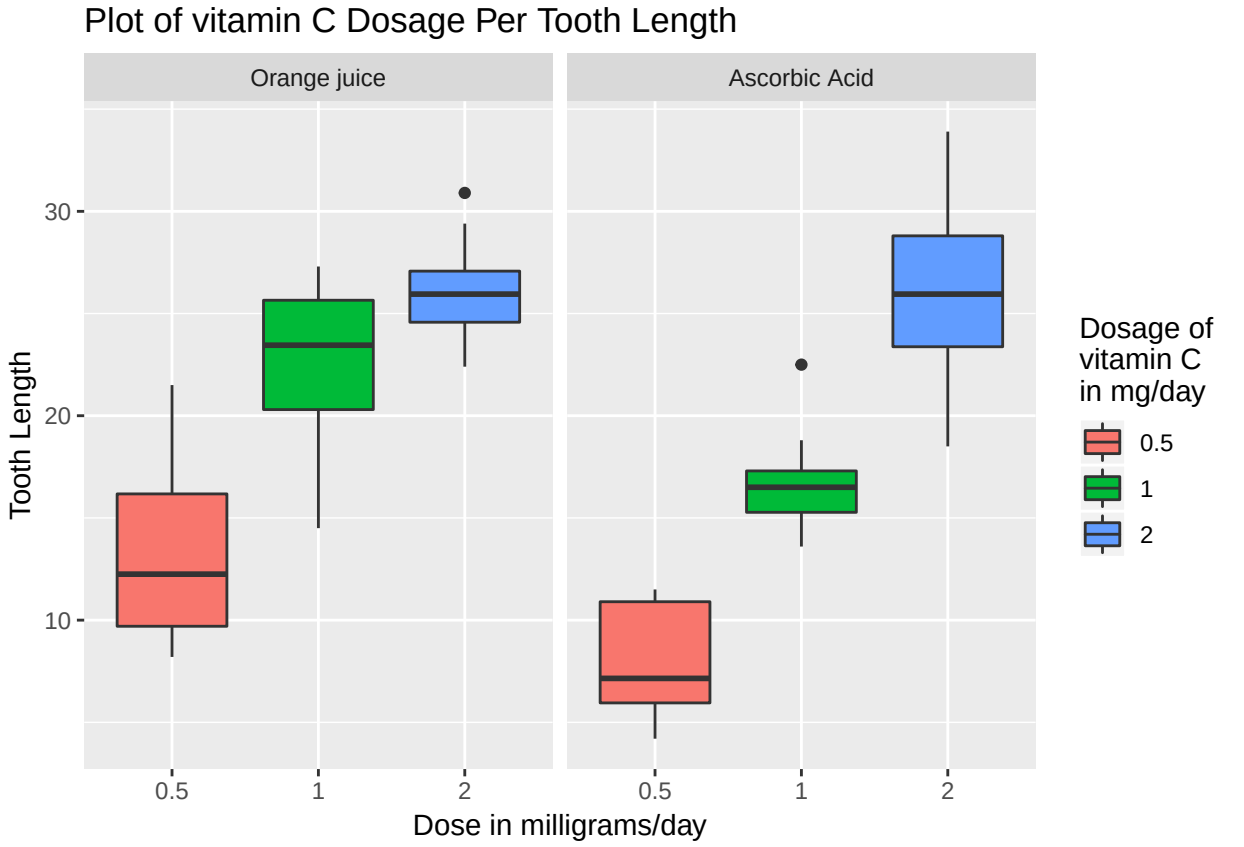
```
# Look at the summary of the data
summary(ToothGrowth)
```

```
##       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
```

```
# Compare means of the different delivery methods
tapply(ToothGrowth$len,ToothGrowth$supp, mean)
```

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

```
# Create a plot to look at data in a graphic format
ggplot(ToothGrowth, aes(factor(dose), len, fill = factor(dose))) +
      geom_boxplot() +
      # facet_grid(.~supp)+
      facet_grid(.~supp, labeller = as_labeller(
            c("OJ" = "Orange juice",
              "VC" = "Ascorbic Acid"))) +
      labs(title = "Plot of vitamin C Dosage Per Tooth Length",
           x = "Dose in milligrams/day",
           y = "Tooth Length") +
      scale_fill_discrete(name = "Dosage of\nvitamin C\nin mg/day")
```




```
tapply(ToothGrowth$len,ToothGrowth$supp, mean)
```

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

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

```
# Compare differences in delivery methods for the same dosage
t05 <- t.test(len ~ supp,
       data = rbind(ToothGrowth[(ToothGrowth$dose == 0.5) &
                                  (ToothGrowth$supp == "OJ"),],
                    ToothGrowth[(ToothGrowth$dose == 0.5) &
                                  (ToothGrowth$supp == "VC"),]),
```

```
       var.equal = FALSE)

t1 <- t.test(len ~ supp,
       data = rbind(ToothGrowth[(ToothGrowth$dose == 1) &
                                  (ToothGrowth$supp == "OJ"),],
                    ToothGrowth[(ToothGrowth$dose == 1) &
                                  (ToothGrowth$supp == "VC"),]),
       var.equal = FALSE)

t2 <- t.test(len ~ supp,
       data = rbind(ToothGrowth[(ToothGrowth$dose == 2) &
                                  (ToothGrowth$supp == "OJ"),],
                    ToothGrowth[(ToothGrowth$dose == 2) &
                                  (ToothGrowth$supp == "VC"),]),
       var.equal = FALSE)

# Summarize the conducted t.tests, which compare the delivery methods by dosage. Seperate p-values and
summaryBYsupp <- data.frame(
      "p-value" = c(t05$p.value, t1$p.value, t2$p.value),
      "Confidence.Low" = c(t05$conf.int[1],t1$conf.int[1], t2$conf.int[1]),
      "Confidence.High" = c(t05$conf.int[2],t1$conf.int[2], t2$conf.int[2]),
      row.names = c("Dose .05","Dose 1","Dose 2"))
# Diplay the data table
summaryBYsupp
```

```
##              p.value Confidence.Low Confidence.High
## Dose .05 0.006358607       1.719057        8.780943
## Dose 1   0.001038376       2.802148        9.057852
## Dose 2   0.963851589      -3.798070        3.638070
```

State your conclusions and the assumptions needed for your conclusions.

With 95% certainty it can be assumed that the dose does have an effect on toothlength. However, in comparison there is not 95% certainty with regards to the delivery method having an effect on toothlength.