Peer-graded Assignment: Statistical Inference Course ProjectPart 2

Alejandro Coy

2019-02-15

Part 2: Basic inferential data analysis. :

The datased is imported and basic data information is obtained with the following functions:

```
library(datasets)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
data <- ToothGrowth
glimpse(data)

## Observations: 60
## Variables: 3
## $ len  <dbl> 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16.5,…
## $ supp <fct> VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC,…
## $ dose <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1.0,…

str(data)

## '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 ...

head(data)

##    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 of the data:

Description

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Usage

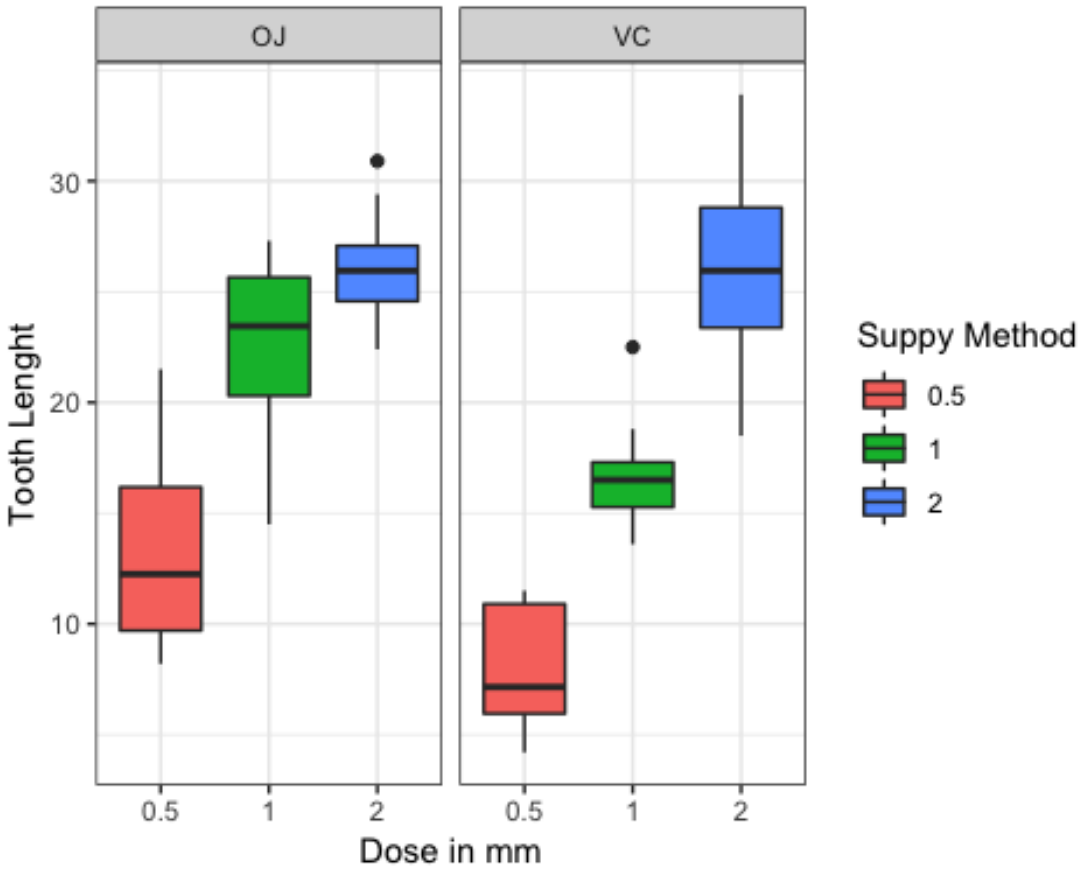
ToothGrowth Format

A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length [,2] supp factor Supplement type (VC or OJ). [,3] dose numeric Dose in milligrams/day

Exploratory data analysis

```
ggplot(data,aes(factor(dose),y=len,fill=factor(dose)))+
  geom_boxplot()+
  facet_wrap(~supp)+
  theme_bw()+
  labs(x="Dose in mm",y="Tooth Lenght",fill="Suppy Method")
```



From the plot it can be seen the dose seems to have an effect in the tooth groth. However, it is not clear that the supplement type have an effect.

Hypothesis Testing

To compare the effect of the doses we will be using hypothesis testing between the three doses. First we have to filter the data by dose

```
dose_0.5 <- data %>%
  filter(dose == 0.5)
dose_1 <- data %>%
  filter(dose == 1)
dose_2 <- data %>%
  filter(dose == 2)
#t-tes Between doses
t.test(dose_0.5$len,dose_1$len,paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  dose_0.5$len and dose_1$len
## 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
```

```
t.test(dose_0.5$len,dose_2$len,paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_0.5$len and dose_2$len
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

```
t.test(dose_1$len,dose_2$len,paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_1$len and dose_2$len
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

From these results it can be infered that there is a significant difference between each on of the doses. The interval of confidence can be esay compute using conf

```
t.test(dose_0.5$len,dose_1$len,paired = FALSE)$conf
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(dose_0.5$len,dose_2$len,paired = FALSE)$conf
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(dose_1$len,dose_2$len,paired = FALSE)$conf
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

Now we can compare betwen supplement type. Since we have stablished that the dose has a significant effect, the suplement type will be compare for each of doses and not between them.

```
vc_0.5 <- data %>%
  filter(dose == 0.5, supp =="VC")
vc_1 <- data %>%
  filter(dose == 1, supp =="VC")
vc_2 <- data %>%
  filter(dose == 2, supp =="VC")
oj_0.5 <- data %>%
  filter(dose == 0.5, supp =="OJ")
oj_1 <- data %>%
  filter(dose == 1, supp =="OJ")
oj_2 <- data %>%
  filter(dose == 2, supp =="OJ")
#Hypothesis testin
t.test(vc_0.5$len,oj_0.5$len,paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  vc_0.5$len and oj_0.5$len
## 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:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98     13.23

t.test(vc_1$len,oj_1$len,paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  vc_1$len and oj_1$len
## t = -4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70

t.test(vc_2$len,oj_2$len,paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc_2$len and oj_2$len
## 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:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

```
# Confidence Intervals
t.test(vc_0.5$len,oj_0.5$len,paired = FALSE)$conf
```

```
## [1] -8.780943 -1.719057
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(vc_1$len,oj_1$len,paired = FALSE)$conf
```

```
## [1] -9.057852 -2.802148
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(vc_2$len,oj_2$len,paired = FALSE)$conf
```

```
## [1] -3.63807  3.79807
## attr(,"conf.level")
## [1] 0.95
```

From the result we can see that the supplmenet type has a signifcant effect for doses of 0.5 a 1 mg/day. In the case the 2 mg/day there is not change in the tooth growth by changing the supplment type.

Conclusions

As conclusion the dose has a significant effect in the tooth growth. The confidence didn't contain 0 and they were negative. This indicate that increase of doses yield greater growth.

In the case of the supplement type, the effect was just seen in the 0.5 and 1 mg.day dose. The intervals were negative, indicating that Orange Juice was a better delivery method. For the case of dose 2, there is not significant difference in the growth.