Tooth Growth

Alex Petkovski

Sunday, May 24, 2015

Overview

The following report investigates **R** dataset **Tooth Growth**. Some preliminary exploratory data analysis is conducted along with a summary of the key statistics. Finally, we construct some hypothesis tests to compare tooth growth by supplement factor and dose in order to state some conclusions and assumptions needed for those conclusions.

Data Exploration

Load the data, get the structure and summarize

```
data(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 ...
```

```
##       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
```

From the structure and review of `?ToothGrowth` help file, we have the following structure and description:

Variable	Data Type	Description
len	numeric	Tooth length
supp	factor	Supplement type (VC or OJ)
dose	numeric	Dose in milligrams

Let's make **dose** a factor since we will be analyzing tooth length by dose factor

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

Finally, let's review the data in a Supplement type panel box plot
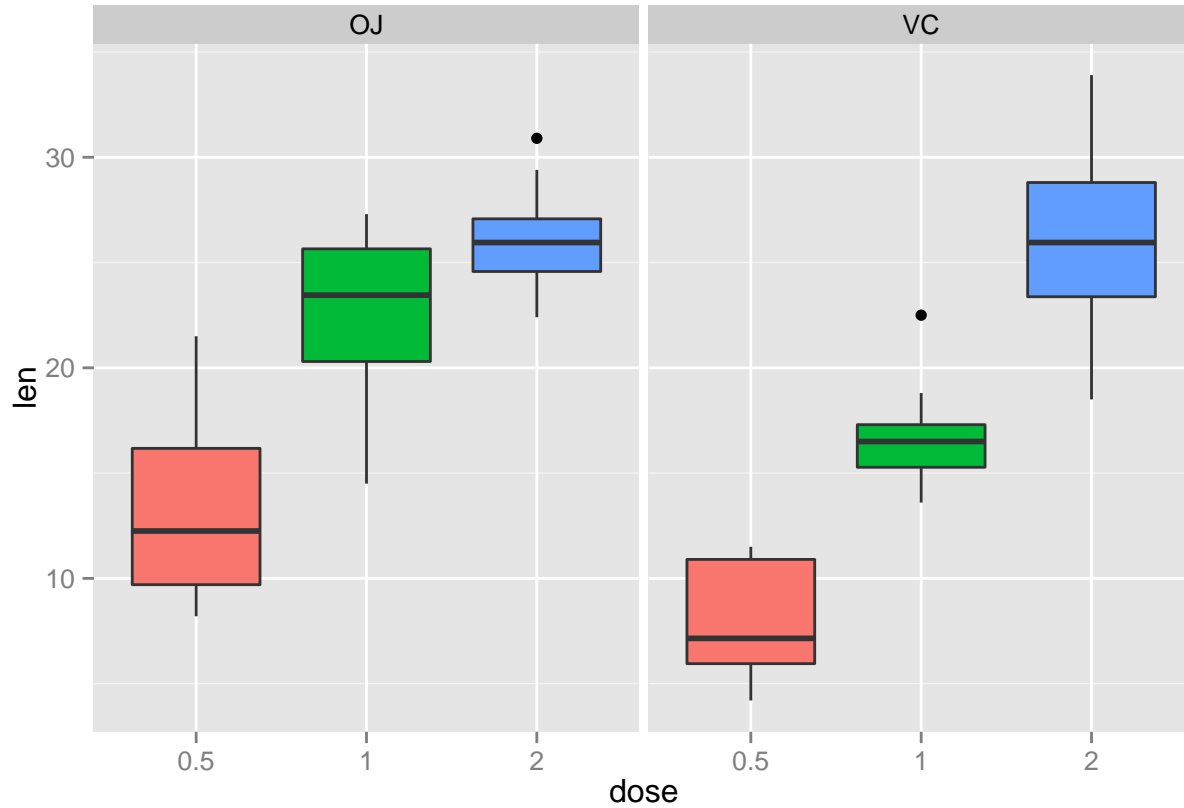
Summary

Assuming a normal distribution, we'll construct a table that calculates the various statistics such as X_bar, S and n.

```
##   dose supp   x.X_bar       x.S       x.n
## 1  0.5   OJ 13.230000  4.459709 10.000000
## 2    1   OJ 22.700000  3.910953 10.000000
## 3    2   OJ 26.060000  2.655058 10.000000
## 4  0.5   VC  7.980000  2.746634 10.000000
## 5    1   VC 16.770000  2.515309 10.000000
## 6    2   VC 26.140000  4.797731 10.000000
```

Confidence Intervals and Hypothesis Tests

Supplement Type Investigation

Does supplement type (VC or OJ) have any impact on tooth length? - H0: different supplement types have no effect on tooth length

```
##
```

```
##   Paired t-test
##
## data:  len by supp
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   1.408659 5.991341
## sample estimates:
## mean of the differences
##                      3.7
```

t is high, confidence does not contain 0 and p-value is less that 5%, thus we reject H0.

Dose Investigation

Does dose have any impact on tooth length? - H0: different doses have no effect on tooth length

```
##     t.value  CI.low CI.high p.value
## 1   -6.967 -11.873  -6.387   0.000
## 2 -11.291 -18.367 -12.623   0.000
## 3   -4.605  -9.258  -3.472   0.000
```

t is very low, confidence does not contain 0 and p-value is less that 5%, thus we reject H0 across all tests.

Conclusions

Assuming that t-tests are paired which is based on description of Tooth Growth data-set, we conclude the following: - Reject hypothesis that supplement types have no effect on tooth growth. - Reject hypothesis that doses have no effect on tooth growth. - Both conclusions also assume that two variances in all t-tests are not equal.

Appendix

Code used to get structure and summary of data

```
str(ToothGrowth)
summary(ToothGrowth)
```

Code used to generate box-plot

```
library(ggplot2)
ggplot(ToothGrowth, aes(x=dose, y=len, fill=dose)) +
    geom_boxplot() +
    facet_grid(. ~ supp) +
    guides(fill=FALSE)
```

Code used to generate summary table

```
attach(ToothGrowth)
dsum <- aggregate(len,list(dose,supp),FUN=function(x) c(X_bar = mean(x)
                                                        , S = sd(x)
                                                        , n = length(x)))
names(dsum) <- c("dose","supp","x")
dsum
detach(ToothGrowth)
```

Code for supplement test

```
t.test(len ~ supp, paired = TRUE, data = ToothGrowth)
```

Code for dose test

```
d_05 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
d_10 <- ToothGrowth$len[ToothGrowth$dose == 1]
d_20 <- ToothGrowth$len[ToothGrowth$dose == 2]

t_05v10 <- t.test(d_05, d_10, paired = TRUE)
t_05v20 <- t.test(d_05, d_20, paired = TRUE)
t_10v20 <- t.test(d_10, d_20, paired = TRUE)

t1 <- data.frame(
    "t.value" = sprintf("%.3f", t_05v10$statistic),
    "CI-low" = sprintf("%.3f", t_05v10$conf[1]),
    "CI-high" = sprintf("%.3f", t_05v10$conf[2]),
    "p.value" = sprintf("%.3f", t_05v10$p.value)
)
t2 <- data.frame(
    "t.value" = sprintf("%.3f", t_05v20$statistic),
    "CI-low" = sprintf("%.3f", t_05v20$conf[1]),
    "CI-high" = sprintf("%.3f", t_05v20$conf[2]),
    "p.value" = sprintf("%.3f", t_05v20$p.value)
)
t3 <- data.frame(
    "t.value" = sprintf("%.3f", t_10v20$statistic),
    "CI-low" = sprintf("%.3f", t_10v20$conf[1]),
    "CI-high" = sprintf("%.3f", t_10v20$conf[2]),
    "p.value" = sprintf("%.3f", t_10v20$p.value)
)

result <- rbind(t1, t2, t3)
result
```