ToothGrowth data analysis

Ricardo Arango Giraldo

23 de marzo de 2018

Overview

We're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. 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)
4. State your conclusions and the assumptions needed for your conclusions

1. Exploratory Data Analysis

First, we load the libraries and data:

```
library(plyr)
library(ggplot2)
library(datasets)
library(grid)

data(ToothGrowth)
```

so, we look what the data reveal.

```
data<-data.frame(ToothGrowth)
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)
```

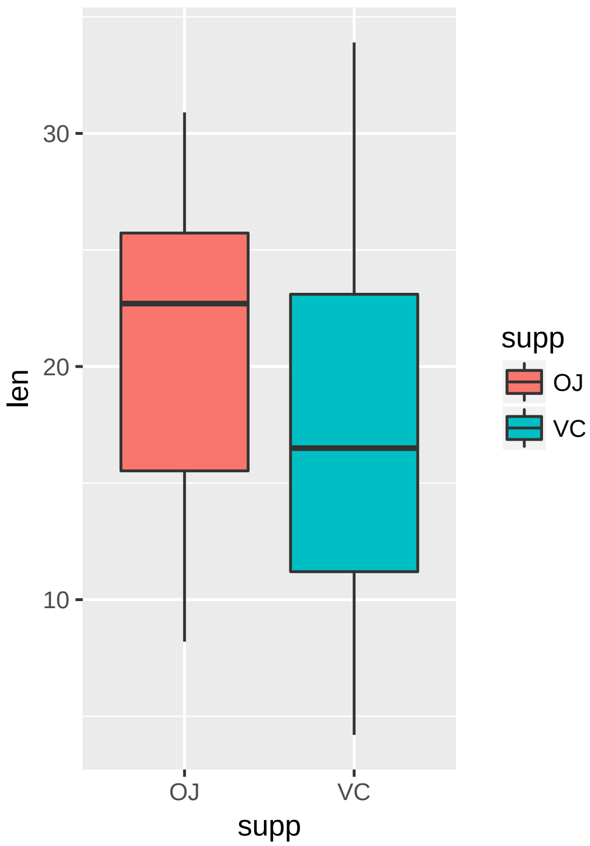
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

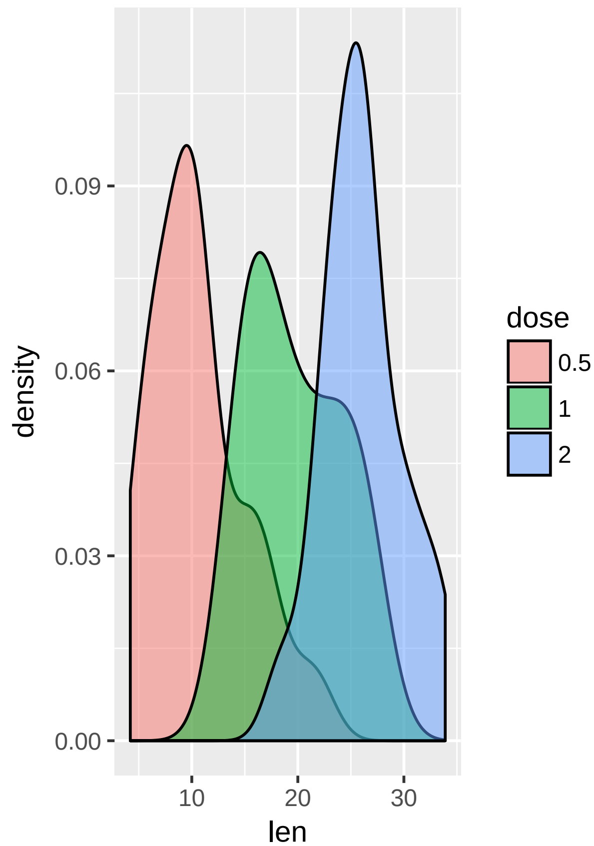
Next, we show the supplements and dose factors in box plot and histogram respectly:

```
data$supp <- factor(data$supp)
data$dose <- factor(data$dose)

p1 <- ggplot(data, aes(x=supp, y=len)) + geom_boxplot(aes(fill=supp))
p2 <- ggplot(data, aes(x=len, fill=dose)) + geom_density(alpha = 0.5)
```

```r
pushViewport(viewport(layout = grid.layout(1, 2)))
print(p1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
print(p2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
```



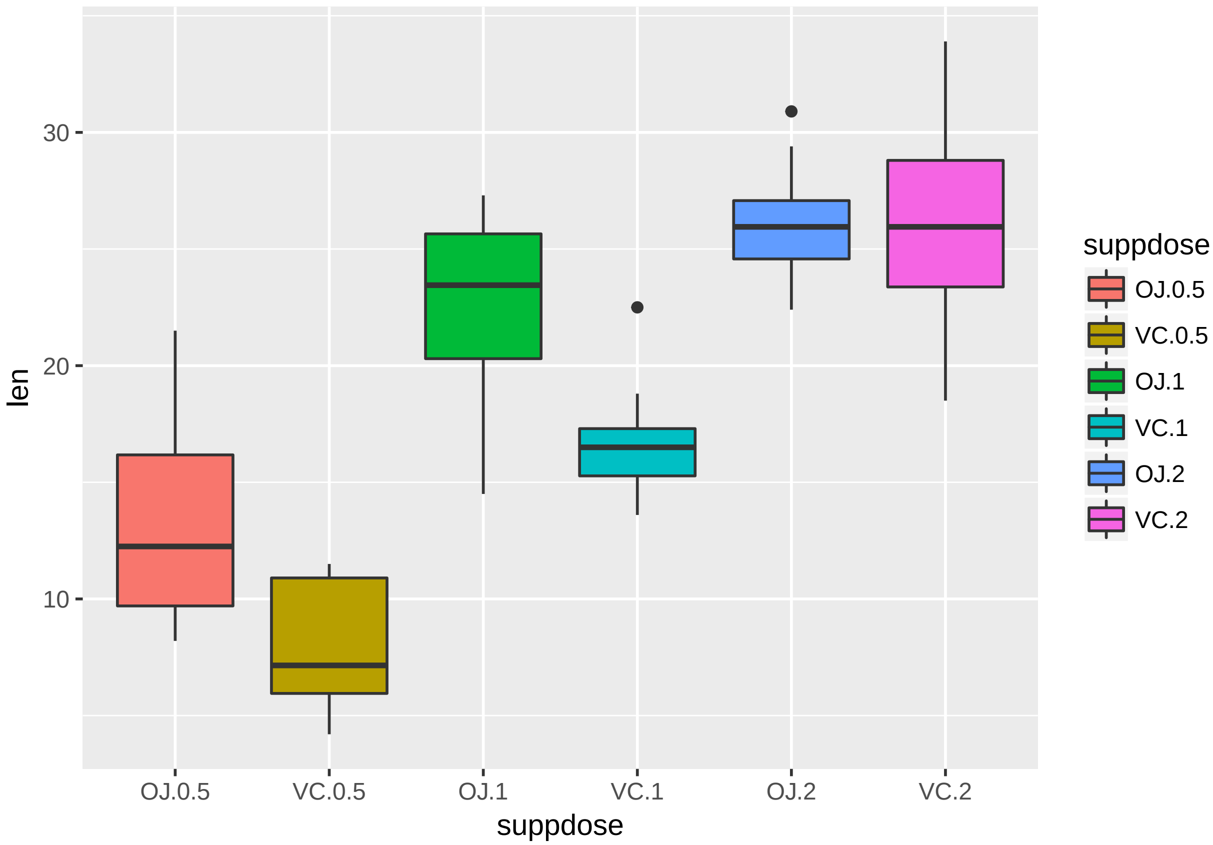


so, here are two looks: the boxplot of supplement versus length shows that their might not be a significant different. Looking at a plot of the histograms from dosage versus length, their might be a relationship there.

One last thing, lets check out the combination of a dose amount and a supplement to see their interaction:

```r
data$suppdose <- interaction(data$supp, data$dose)

ggplot(aes(y=len, x = suppdose), data = data) + geom_boxplot(aes(fill=suppdose))
```



looks like there might just be a relationship at the dosage level 1mg and the supplement.

2. Comparing supplement and dose to tooth length

We note two things:

1. There are small sample sizes, so the t-test is appropriate.

2.A general assumption will be that variances are not equal and just let R do the work to figure out the sample variance and apply it to the statistic.

NOTE: Confidence intervals, p-values, etc. will only be reported, saving the Conclusion section to summarize the results.

Supplement groups

Comparing the difference between supplement groups, independent of dose.

```
t1 <- t.test(len~supp, paired=F, var.equal=F, data=data)
t1.summary <- data.frame("p-value"=c(t1$p.value),"CI-Lower"=c(t1$conf[1]), "CI-Upper"=c(t1$conf[2]),
     row.names=c("OJ vs. VC:  "))
round(t1.summary,4)
```

```
##              p.value CI.Lower CI.Upper
## OJ vs. VC:    0.0606   -0.171    7.571
```

Dosage Groups

Looking at the different dosage groups requires three comparisons: (1) .5 to 1; (2) .5 to 2; (3) 1 to 2

```
df05 <- subset(data, data$dose==.5)
df10 <- subset(data, data$dose==1)
df20 <- subset(data, data$dose==2)

t0510<- t.test(df05$len, df10$len, paired=F, var.equal=F)
t0520<- t.test(df05$len, df20$len, paired=F, var.equal=F)
t1020<- t.test(df10$len, df20$len, paired=F, var.equal=F)

t2.summary <- data.frame("p-value"=c(t0510$p.value,t0520$p.value,t1020$p.value),
     "CI-Lower"=c(t0510$conf[1],t0520$conf[1],t1020$conf[1]),
     "CI-Upper"=c(t0510$conf[2],t0520$conf[2],t1020$conf[2]),
     row.names=c(".5mg vs 1mg: ", ".5mg vs 2mg: ","1mg vs 2mg: "))

round(t2.summary, 6)
```

```
##               p.value   CI.Lower   CI.Upper
## .5mg vs 1mg:  0.0e+00 -11.983781  -6.276219
## .5mg vs 2mg:  0.0e+00 -18.156167 -12.833833
## 1mg vs 2mg:   1.9e-05  -8.996481  -3.733519
```

Comparing supplement within each dosage group

Recall from our third graph, when the supplement was compared within each dosage group, it looked like there might have been a difference for the 1mg level. Let's look within the groups just to check.

```
t05 <- t.test(len~supp, paired=F, var.equal=F, data=df05)
t10 <- t.test(len~supp, paired=F, var.equal=F, data=df10)
t20 <- t.test(len~supp, paired=F, var.equal=F, data=df20)

t3.summary <- data.frame("p-value"=c(t05$p.value,t10$p.value,t20$p.value),
     "CI-Lower"=c(t05$conf[1],t10$conf[1],t20$conf[1]),
     "CI-Upper"=c(t05$conf[2],t10$conf[2],t20$conf[2]),
     row.names=c(".5mg OJ vs. VC: ", "1mg OJ vs. VC: ","2mg OJ vs. VC: "))

round(t3.summary, 6)
```

```
##                   p.value  CI.Lower CI.Upper
## .5mg OJ vs. VC:  0.006359  1.719057 8.780943
## 1mg OJ vs. VC:   0.001038  2.802148 9.057852
## 2mg OJ vs. VC:   0.963852 -3.798070 3.638070
```

Conclusions

Restating assumptions that small sample sizes lend themselves to t-tests and that variances were never treated as equal, allowing R to calculate the pooled variance for the test.

1. Overall, there appears to be no difference in supplement as the p-value was .061 and the confidence interval contained zero.
2. Appearances of no difference in supplement is false when looking at the dosage groups. For both .5mg and 1mg groups, a p-value of .006 and .001 respectively was obtained and both confidence intervals did

not contain zero. For 2mg, there was no difference in supplement. So, for lower dosages (.5mg, 1mg) the delivery mechanism of choice is OJ.

3. It was very apparent that higher dosages had a significant effect. In all cases, p-values were incredible small and no confidence interval contained zero.