tooth_growth

Kyle Scully
October 25, 2014

Introduction

This report will analyze the ToothGrowth data in the R datasets package.

Load the ToothGrowth data

```
library(ggplot2)
data(ToothGrowth)
```

Basic exploratory data analyses

```
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.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```
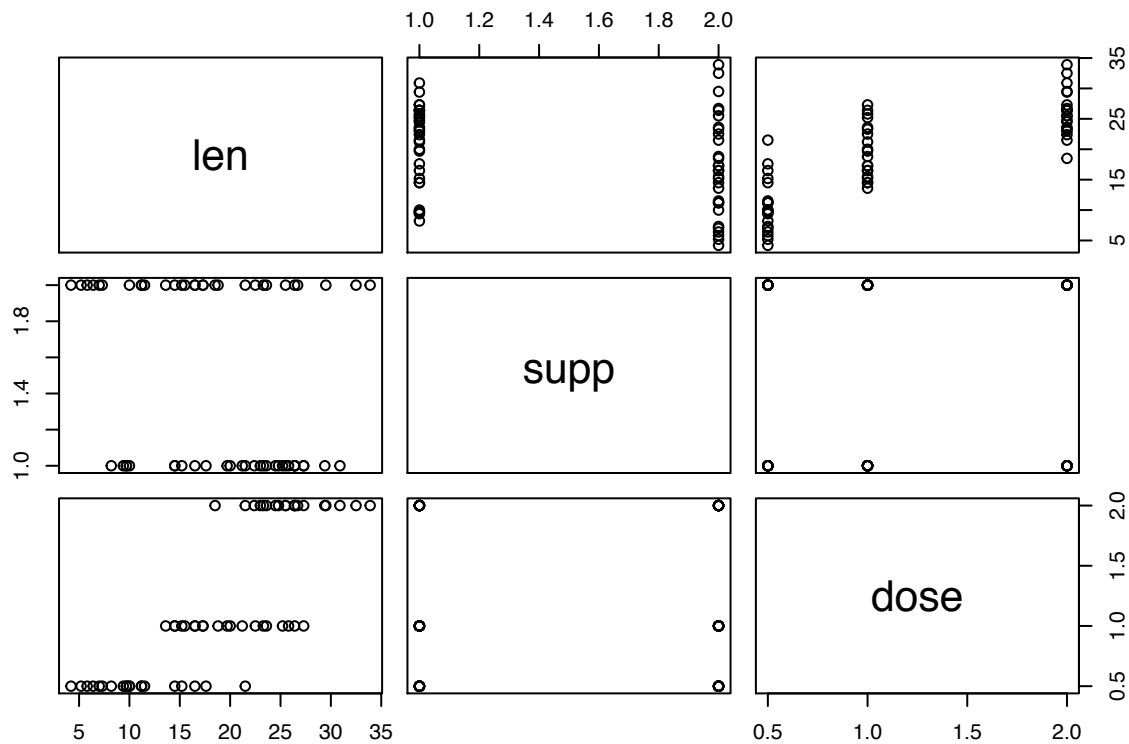
```
head(ToothGrowth)
```

```
##    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
```

```
tail(ToothGrowth)
```

```
##      len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

```
plot(ToothGrowth)
```


Basic summary of the data

From the plot and data summaries in the previous section we can see that there are three variables:

- Tooth Length (len)
- Supplement (supp)
- Dosage (dose)

We also observe that there are only two Supplements tested:

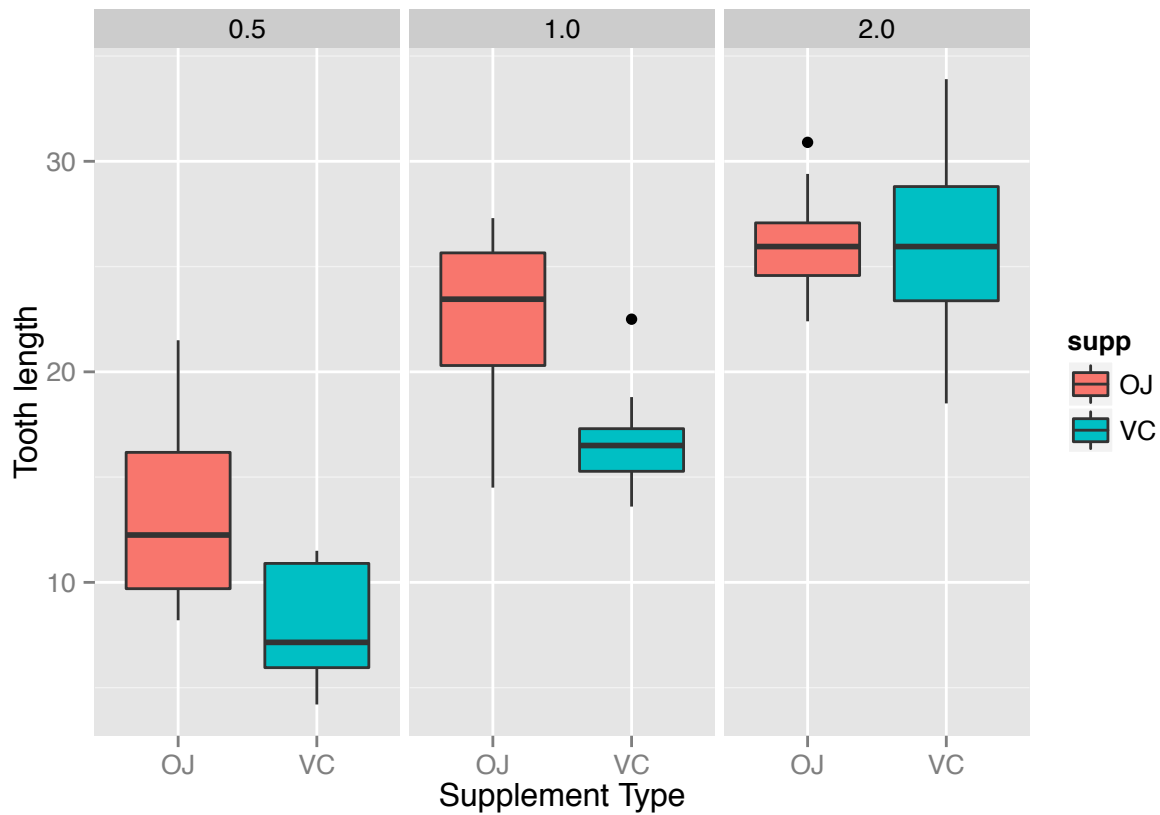- OJ - Orange Juice
- VC - Vitamin C

These Supplements are tested at 3 dosage amounts:

- 0.5 mg
- 1.0 mg
- 2.0 mg

Also there seems to be some sort of positive correlation between dosage and tooth length from the plot.

Confidence intervals and hypothesis tests

```
ggplot(aes(x = supp, y = len),
       data = ToothGrowth) +
  geom_boxplot(aes(fill = supp)) +
  facet_wrap(~ dose) +
  xlab("Supplement Type") +
  ylab("Tooth length")
```



```
ggplot(aes(x=supp, y=len),
       data=ToothGrowth) +
  geom_boxplot(aes(fill=supp)) +
  xlab("Supplement Type") +
  ylab("Tooth Length")
```
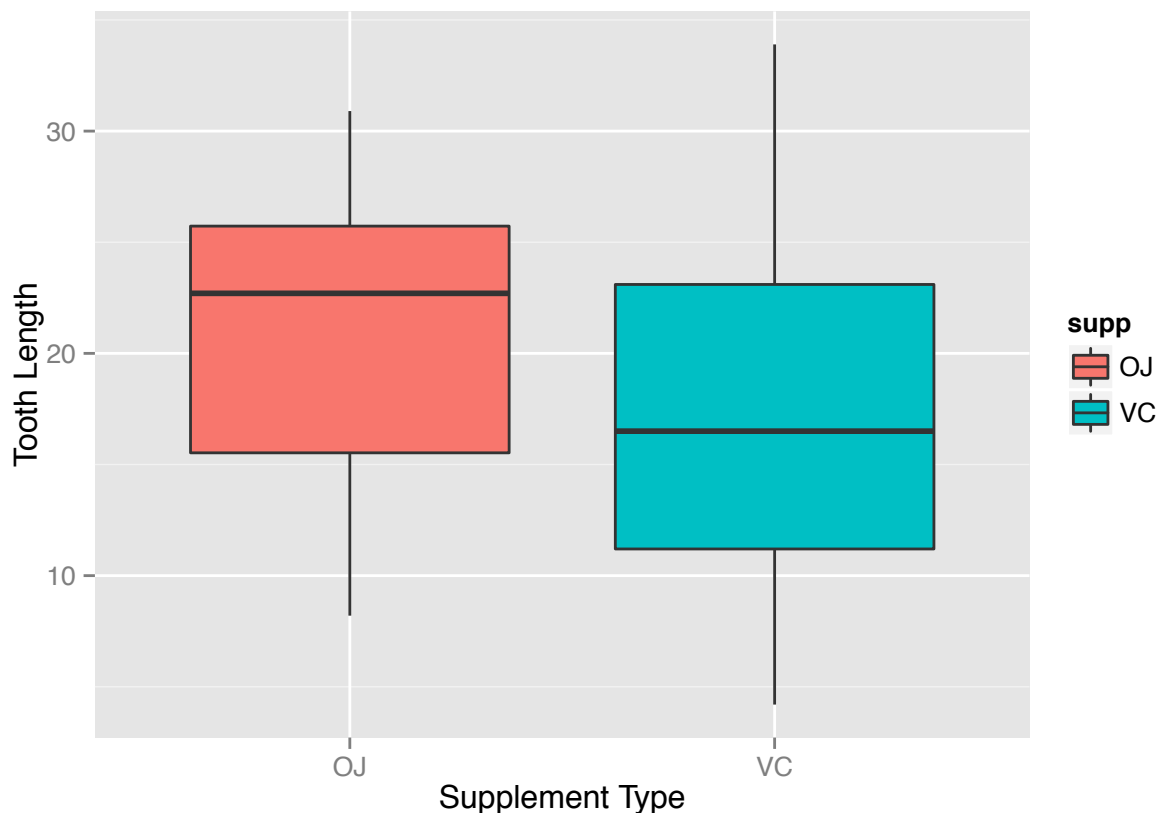
Our suspicion from the above plots is that tooth growth will only have significant impact from dosage amount. Lets setup some null hypothesis and put the data to the test.

First null hypothesis: The variances of tooth length between each Supplement are not equal:

```
t.test(ToothGrowth$len[ToothGrowth$supp=="OJ"],
       ToothGrowth$len[ToothGrowth$supp=="VC"],
       paired = FALSE,
       var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "VC"]
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean of x mean of y
##     20.66     16.96
```

We find the p-value to be 0.06063, which means there is enough evidence to reject the null hypothesis. This essentially means that we cannot assume that the Supplement provided has an effect on tooth growth as the variances between supplements tested are essentially equal.

Second null hypothesis: The variances of Dosage=0.5 mg and Dosage=1.0 mg are equal

```
t.test(ToothGrowth$len[ToothGrowth$dose==1],
       ToothGrowth$len[ToothGrowth$dose==0.5],
       paired = FALSE,
       var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## t = 6.477, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276 11.984
## sample estimates:
## mean of x mean of y
##     19.73     10.61
```

We find the p-value to be 1.266e-07 which is very small, which means there is not enough evidence to reject the null hypothesis. This essentially means it is safe to assume there is some correlation between tooth growth and dosages between tests at value 0.5 mg and value 1.0 mg.

Similarly, we can test between dosages 1.0 mg and 2.0 mg.

Third null hypothesis: The variances of Dosage=1.0 and Dosage=2.0 are equal

```
t.test(ToothGrowth$len[ToothGrowth$dose==2],
       ToothGrowth$len[ToothGrowth$dose==1],
       paired = FALSE,
       var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 1]
## t = 4.901, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.736 8.994
## sample estimates:
## mean of x mean of y
##     26.10     19.73
```

We find the p-value to be 1.811e-05 which is very small, which means there is not enough evidence to reject the null hypothesis. This essentially means it is safe to assume there is some correlation between tooth growth and dosages between tests at value 1.0 mg and value 2.0 mg.

Conclusions and Assumptions

While quite obvious from the graphs we can conclude that supplement delivery (via juice or pill) had no significant effect on tooth growth. Additionally, we can conclude that there is a positive correlation between dosage amount taken and tooth growth.

Assumptions:

- This assumes that the data collected took the appropriate steps to randomize the population so that the sample selected was representative of the whole.
- There are no other variables in the way the test was conducted that had confounding effects on tooth growth
- The amounts of supplement administered were true to the amounts recorded.