

Analysis of ToothGrowth data in R datasets package

Erwin Lo

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

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
data(ToothGrowth)
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 ...
```

Break down the number of observations by supp and dose.

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

2. Provide a basic summary of the data.

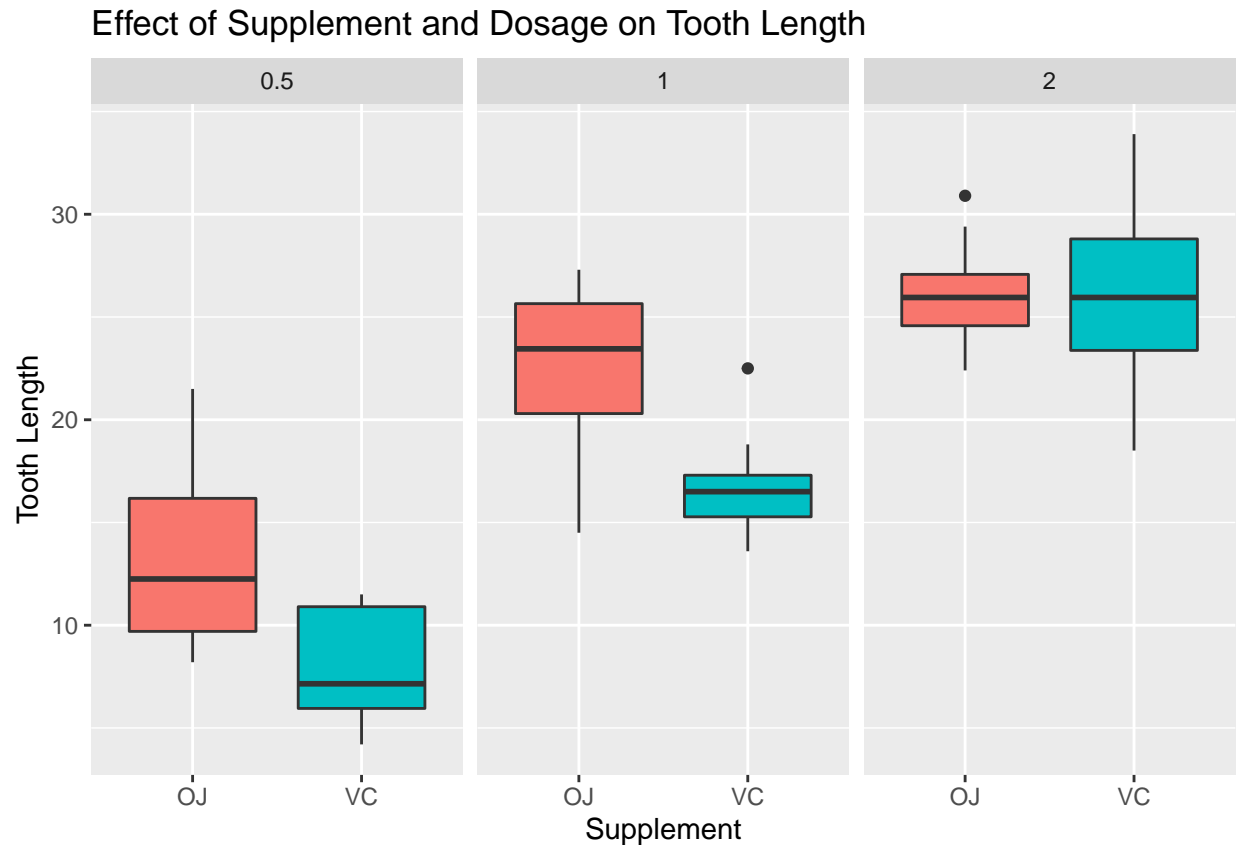
```
summary(ToothGrowth)
```

```
##      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
```

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)

Display data in graphical form

```
library(ggplot2)
ggplot(ToothGrowth, aes(x = supp, y = len, fill = supp)) +
  geom_boxplot() +
  facet_wrap(. ~ dose) +
  labs(title = "Effect of Supplement and Dosage on Tooth Length",
       x="Supplement", y="Tooth Length") +
  theme(legend.position = "none")
```



Let's split data into different supplements.

```
oj <- subset(ToothGrowth, supp == "OJ")
vc <- subset(ToothGrowth, supp == "VC")
```

We shall now do hypothesis tests to compare tooth growth by supp and dose.

Dose = 0.5

```
oj1 <- subset(oj, dose == 0.5)
vc1 <- subset(vc, dose == 0.5)
t.test(oj1$len - vc1$len)

##
## One Sample t-test
##
## data:  oj1$len - vc1$len
## t = 2.9791, df = 9, p-value = 0.01547
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.263458 9.236542
## sample estimates:
## mean of x
##      5.25
```

p-value is less than 0.02 and confidence interval is above 0. The effect of the two supplements is different and significant at dose 0.5.

Dose = 1.0

```
oj2 <- subset(oj, dose == 1.0)
vc2 <- subset(vc, dose == 1.0)
t.test(oj2$len - vc2$len)
```

```
##
## One Sample t-test
##
## data:  oj2$len - vc2$len
## t = 3.3721, df = 9, p-value = 0.008229
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.951911 9.908089
## sample estimates:
## mean of x
##      5.93
```

p-value is less than 0.01 and confidence interval is above 0. The effect of the two supplements is different and significant at dose 1.0.

Dose = 2.0

```
oj3 <- subset(oj, dose == 2.0)
vc3 <- subset(vc, dose == 2.0)
t.test(oj3$len - vc3$len)
```

```
##
## One Sample t-test
##
## data:  oj3$len - vc3$len
## t = -0.042592, df = 9, p-value = 0.967
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.328976 4.168976
## sample estimates:
## mean of x
##     -0.08
```

p-value is more than 0.9 and confidence interval is below 0. We can't say that one supplement has more effect than the other.

4. State your conclusions and the assumptions needed for your conclusions.

1. From the boxplot above, we can see that as dosage increases tooth length increases as well. Higher dosage results in longer tooth length.
2. For dosage of 0.5 and 1.0, supplement OJ will result in longer tooth length compared to supplement VC.
3. For dosage of 2.0, there is no difference in effect from either supplements.