

# Statistical Inference Course Project

Tonya MacDonald

## Purpose

Analyze Tooth Growth Dataset.

## Libraries

```
library("data.table")
library("ggplot2")
#library(dplyr)
#data(ToothGrowth)
```

## Basic summary of the data

```
# show what column names are
names(ToothGrowth)

## [1] "len" "supp" "dose"

# show a bit of the data
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

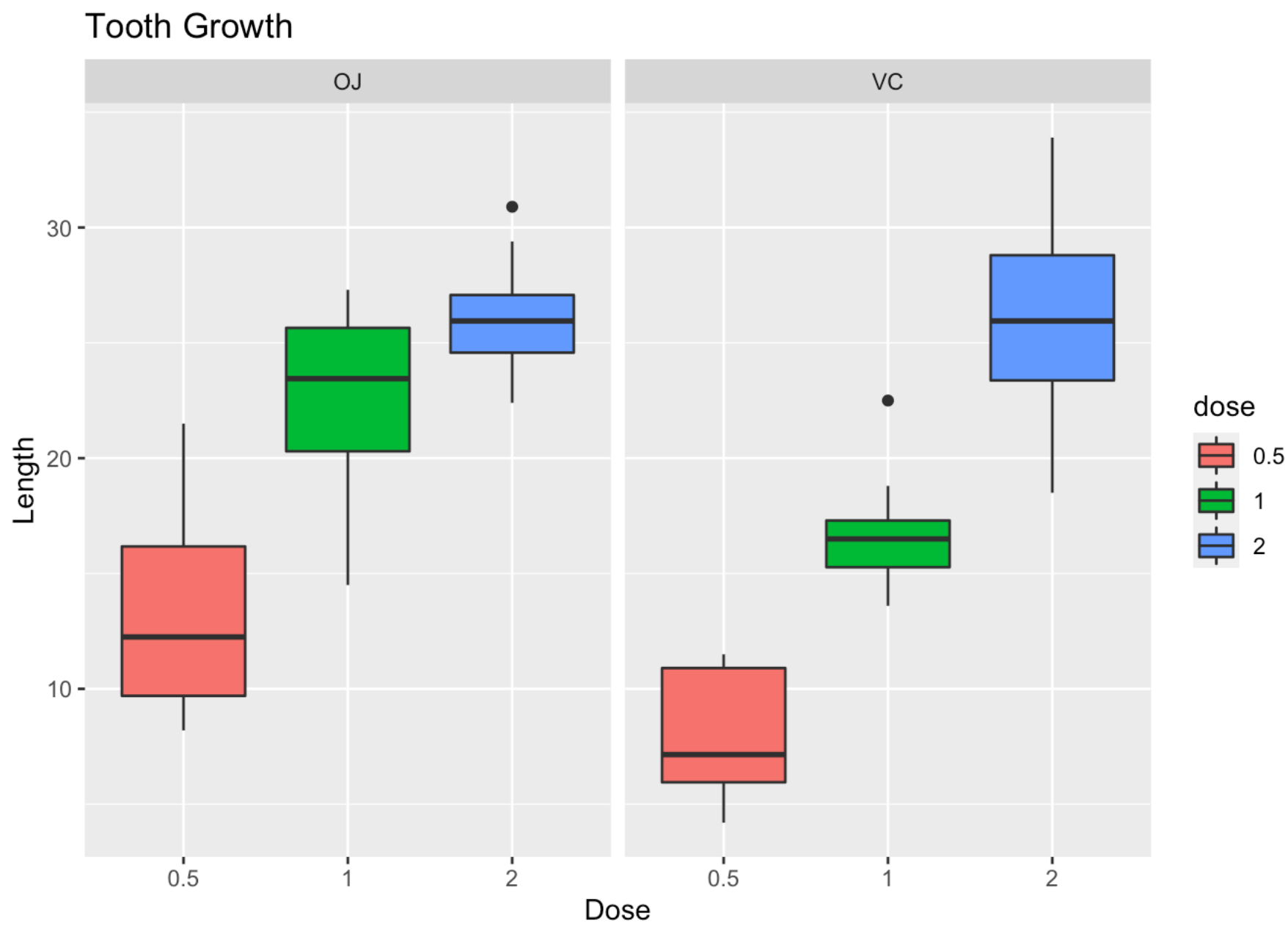
# show summary of the data
summary(ToothGrowth)

##      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
```

## Compare tooth growth

```
# make dose a factor
ToothGrowth$dose<-as.factor(ToothGrowth$dose)

# show box plot
ggplot(aes(x=dose, y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=dose)) +
  xlab("Dose") +
  ylab("Length") +
  facet_grid(~ supp) +
  ggtitle("Tooth Growth")
```



```
# run t-test
t.test(len~supp,data=ToothGrowth)

##
## Welch Two Sample t-test
##
## data: len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##      20.66333      16.96333
```

The p-value is ~ 0.06

```
# run t-test using dose amounts 0.5 and 1.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5))
t.test(len~dose,data=ToothGrowth_sub)

##
## Welch Two Sample t-test
##
## data: len by dose
## 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 in group 0.5 mean in group 1
##      10.605      19.735
```

The p-value is ~ 0, and the confidence interval does not include 0

```
# run t-test using dose amounts 1.0 and 2.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(2.0,1.0))
t.test(len~dose,data=ToothGrowth_sub)

##
## Welch Two Sample t-test
##
## data: len by dose
## 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 in group 1 mean in group 2
##      19.735      26.100
```

The p-value is ~ 0, and the confidence interval does not include 0

```
# run t-test using dose amounts 0.5 and 2.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))
t.test(len~dose,data=ToothGrowth_sub)

##
## Welch Two Sample t-test
##
## data: len by dose
## 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 in group 0.5 mean in group 2
##      10.605      26.100
```

The p-value is ~ 0, and the confidence interval does not include 0

## Summary

For each of the two sample t-tests, the p-value is ~ 0, and the confidence interval does not cross over 0. Also the means increase with dose.

We can conclude that as the dose increases, then the tooth growth increases. Delivery method does not have an effect on the tooth growth.