# Statistical Inference Project2 - Tooth Data

***Diane Clow***

***August 22, 2015***

## Introduction

This is the Second Course Project for Statistical Inference. The assignment asks students to analyse the data in the ToothGrowth dataset in R and run some analysis.

To complete this assignment I will:

- load the ToothGrowth data set and perform exploritory analysis
- provide a basic summary
- compare tooth growth by supp and dose
- state my conclusion and assumptions needed

## Exploratory Analysis

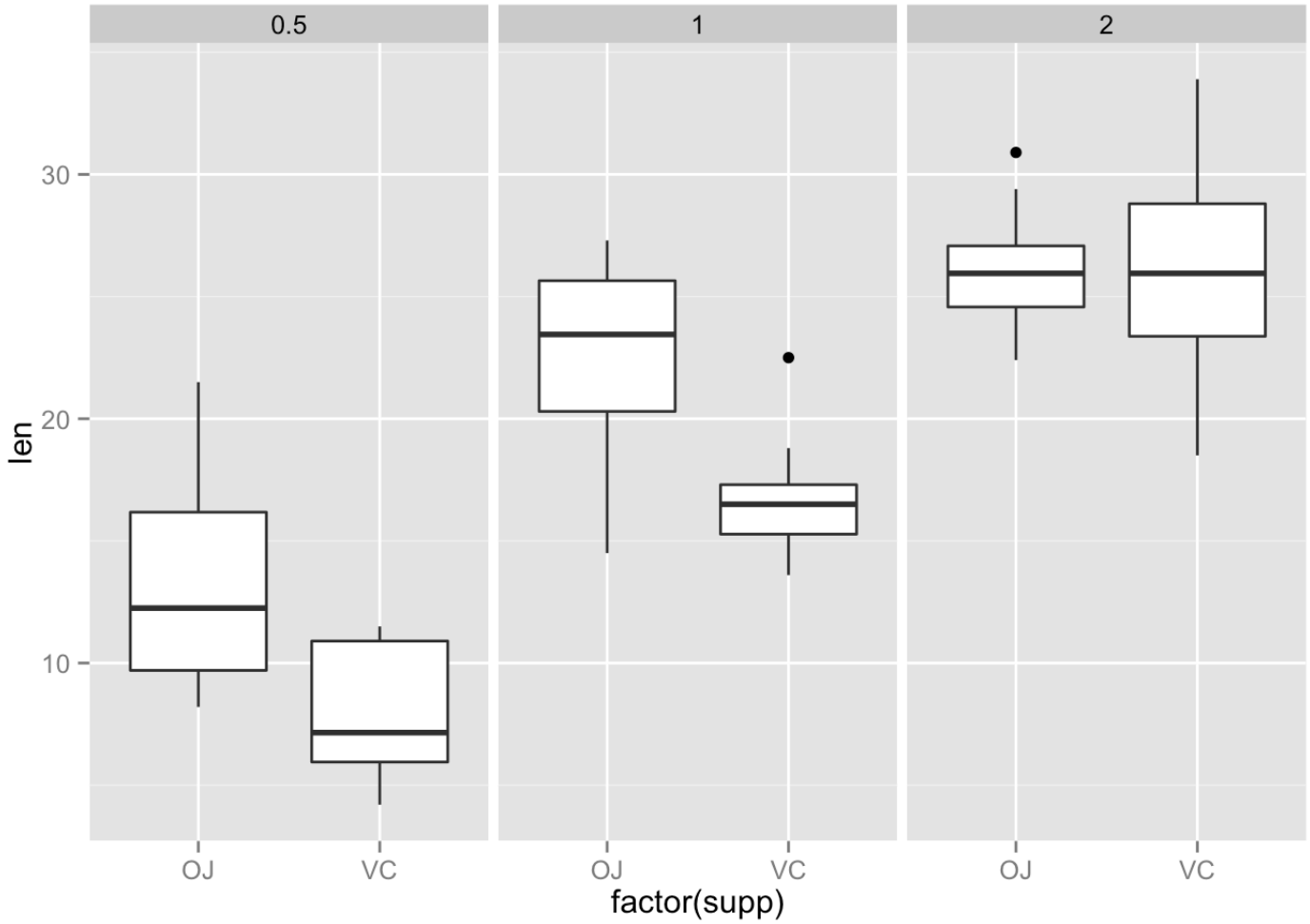
```
library(ggplot2)
data(ToothGrowth)
head(ToothGrowth)
```

```
##    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
```

The three variables in this data set are: len - numeric, the length of the tooth, supp - factor, the supplement type (VC or OJ), dose - numeric, dose in milligrams/day. VC is ascorbic acid, a form of vitamin C and OJ stands for Orange Juice.

The following graph shows box plots broken up by supplement and dose.

```
ggplot(ToothGrowth, aes(factor(supp), len)) + geom_boxplot() + facet_grid(. ~ dose)
```




## Basis Summary

Shown below is the summary information for the data set ToothGrowth. Below that is the data aggregrated for mean and standard deviation on both supp and dose.

```
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
```

```
aggregate(ToothGrowth$len, by=list(ToothGrowth$supp, ToothGrowth$dose), FUN=mean)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 13.23
## 2      VC     0.5  7.98
## 3      OJ     1.0 22.70
## 4      VC     1.0 16.77
## 5      OJ     2.0 26.06
## 6      VC     2.0 26.14
```

```
aggregate(ToothGrowth$len, by=list(ToothGrowth$supp, ToothGrowth$dose), FUN=sd)
```

```
##   Group.1 Group.2        x
## 1      OJ     0.5 4.459709
## 2      VC     0.5 2.746634
## 3      OJ     1.0 3.910953
## 4      VC     1.0 2.515309
## 5      OJ     2.0 2.655058
## 6      VC     2.0 4.797731
```

## Comparison

```
OJ0.5 <- subset(ToothGrowth, supp == "OJ" & dose == 0.5)
VC0.5 <- subset(ToothGrowth, supp == "VC" & dose == 0.5)
OJ1 <- subset(ToothGrowth, supp == "OJ" & dose == 1)
VC1 <- subset(ToothGrowth, supp == "VC" & dose == 1)
OJ2 <- subset(ToothGrowth, supp == "OJ" & dose == 2)
VC2 <- subset(ToothGrowth, supp == "VC" & dose == 2)
```

For each dose size a t-test will be preformed. For all these test, I will be testing if the length with OJ is the same as the length with VC. If the lengths are the same, I will accept my null hypothosis and 0 will be within the confidence interval. If 0 is not within the confidence interval, the alternate hypothosis will be accepted, that OJ has a greater impact on teeth length. The confidance level for the following tests is 0.95. This means that there is a 5% chance they will produce false results.

H_0 <- Both supplements have the same impact

H_a <- OJ has a larger impact that VC

Sig <- .95

```
t.test(OJ0.5$len - VC0.5$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  One Sample t-test
##
## data:  OJ0.5$len - VC0.5$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
```

```
t.test(OJ1$len - VC1$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  One Sample t-test
##
## data:  OJ1$len - VC1$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
```

```
t.test(OJ2$len - VC2$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  One Sample t-test
##
## data:  OJ2$len - VC2$len
## t = -0.0426, 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
```

## Conclusion

In analysing this data set a couple of assumptions were made. The guinea pigs were a representative sample of the total population, and that the change in the supplamant is the only variations between the two samples. There is nothing in the documentation to say that the samples are paired, so it is assumed that the samples are independant.

As shown in the tests above only with the dosage = 2, is the difference between OJ and VC statistically insignificant. For both of the lower dosages, OJ has an impact on the length.