# Statistical Inference Course Project - Part 2

*Joshua Chua*

*Sunday, August 23, 2015*

# Overview

In this part of the project we analyse the ToothGrowth data in the R datasets package by doing the following:

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.

# Data analysis

From the documentation:https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html (https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html)

**Description:** The response is the length of odontoblasts (teeth) in 60 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

**Format:** A data frame with 60 observations on 3 variables.

| Variable | Class | Description |
|---|---|---|
| le | numeric | Tooth length |
| supp | factor | Supplement type (VC or OJ). |
| dose | numeric | Dose in milligrams. |

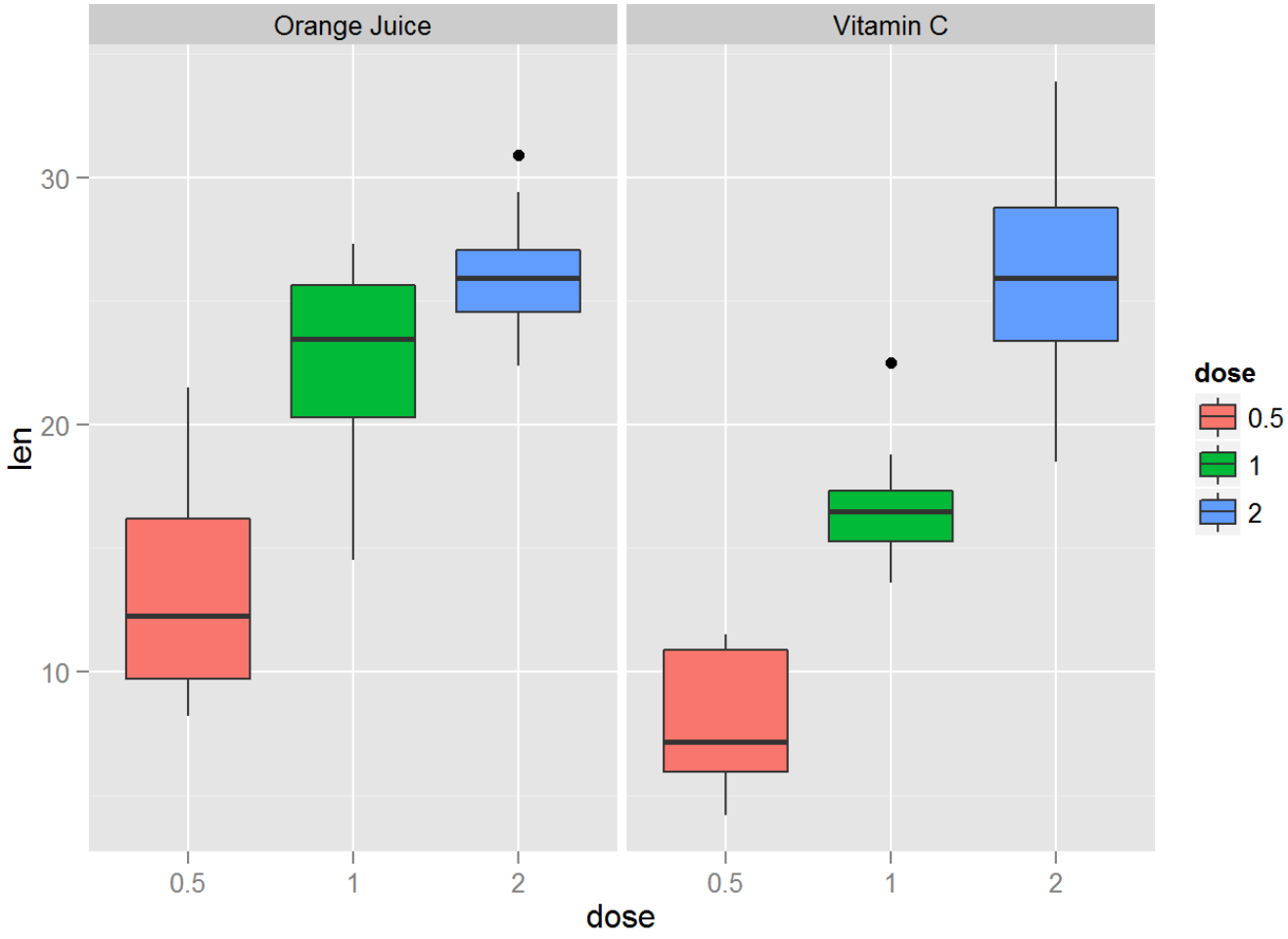
We can see that there are effectively only 3 dosages, so we will convert this to a factor variable.

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose) #dose as a factor
levels(ToothGrowth$supp)<-c("Orange Juice","Vitamin C") #rename supplement type
summary(ToothGrowth)
```

```
##       len                  supp     dose
##  Min.   : 4.20   Orange Juice:30   0.5:20
##  1st Qu.:13.07   Vitamin C   :30   1  :20
##  Median :19.25                     2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Plotting the data, let's compare the growth in tooth length given the varying dosage for orange juice and ascorbic acid.

```
ggplot(ToothGrowth, aes(x=dose, y=len))+
  geom_boxplot(aes(fill=dose))+
  #geom_smooth(aes(group=1),method = "lm", se = TRUE)+
  facet_grid(.~supp)
```




We can observe that Vitamin C seems to have an effect on teeth length since a higher does is associated with longer teeth in both delivery methods. Also, orange juice as a method of delivery seems result in longer teeth than Vitamin C, with the exception of the 2mg dose where the lengths are very similar. This might indicate a maximum effect point where doses above 2mg have no additional positive effect on tooth growth.

# Hypothesis testing

From these observations we propose the following hypotheses and check them by means of two-group t-tests.

### 1. H0, mean tooth length corresponding to two different dosages is the same.

Comparing the dosages 0.5mg to 1mg, 1mg to 2mg, 0.5 to 2mg, we get the following results.

| Dosage | Confidence Interval | p-value | Outcome |
|---|---|---|---|
| 0.5 - 1 | [-11.983781 -6.276219] | 1.268e-07 | Reject H0 |
| 1 - 2 | [-8.996481 -3.733519] | 1.906e-05 | Reject H0 |
| 0.5 - 2 | [-18.15617 -12.83383] | 4.398e-14 | Reject H0 |

### 2. H0, mean tooth length corresponding to two different delivery methods is the same.

Comparing the delivery methods for the 3 different doses we get the following results.

| Dosage | Confidence Interval | p-value | Outcome |
|---|---|---|---|
| 0.5 | [1.719057 8.780943] | 0.006359 | Reject H0 |
| 1 | [2.802148 9.057852] | 0.001038 | Reject H0 |
| 2 | [-3.79807 3.63807] | 0.9639 | Fail to reject H0 |

# Conclusions and assumptions

Based on the above, we can conclude that:

1. Increased vitamin C dosages (in either orange juice or pure ascorbic acid form) does promote of tooth growth.
2. Delivery method has an effect on tooth growth when dosages is between 0.5 and 1mg, but has little influence at a 2mg dose.

We assume that the data is not paired as there are only 60 observations from 10 guinea pigs, 2 delivery methods and 3 dosages, and also that double blind reasearch methods were followed when obtaining the data.

# Appendix 1: Codes and outputs

```
#subsetting the data by dose
dose05_1 <- subset(ToothGrowth, dose %in% c(0.5,1))
dose1_2 <- subset(ToothGrowth, dose %in% c(1,2))
dose05_2 <- subset(ToothGrowth, dose %in% c(0.5,2))

#0.5mg vs 1mg
t.test(len ~ dose, data=dose05_1, paired=FALSE, var.equal=FALSE)
```

```
##
##  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
```

```
#1mg vs 2mg
t.test(len ~ dose, data=dose05_2, paired=FALSE, var.equal=FALSE)
```

```
##
##  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
```

```
#0.5mg vs 2mg
t.test(len ~ dose, data=dose1_2, paired=FALSE, var.equal=FALSE)
```

```
##
##  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
```

```
#delivery method t-tests
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ],paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group Orange Juice    mean in group Vitamin C
##                      13.23                       7.98
```

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group Orange Juice    mean in group Vitamin C
##                      22.70                      16.77
```

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group Orange Juice    mean in group Vitamin C
##                      26.06                      26.14
```