# ToothGrowth data analysis.

*Bosah Chude*

*Friday, November 21, 2014*

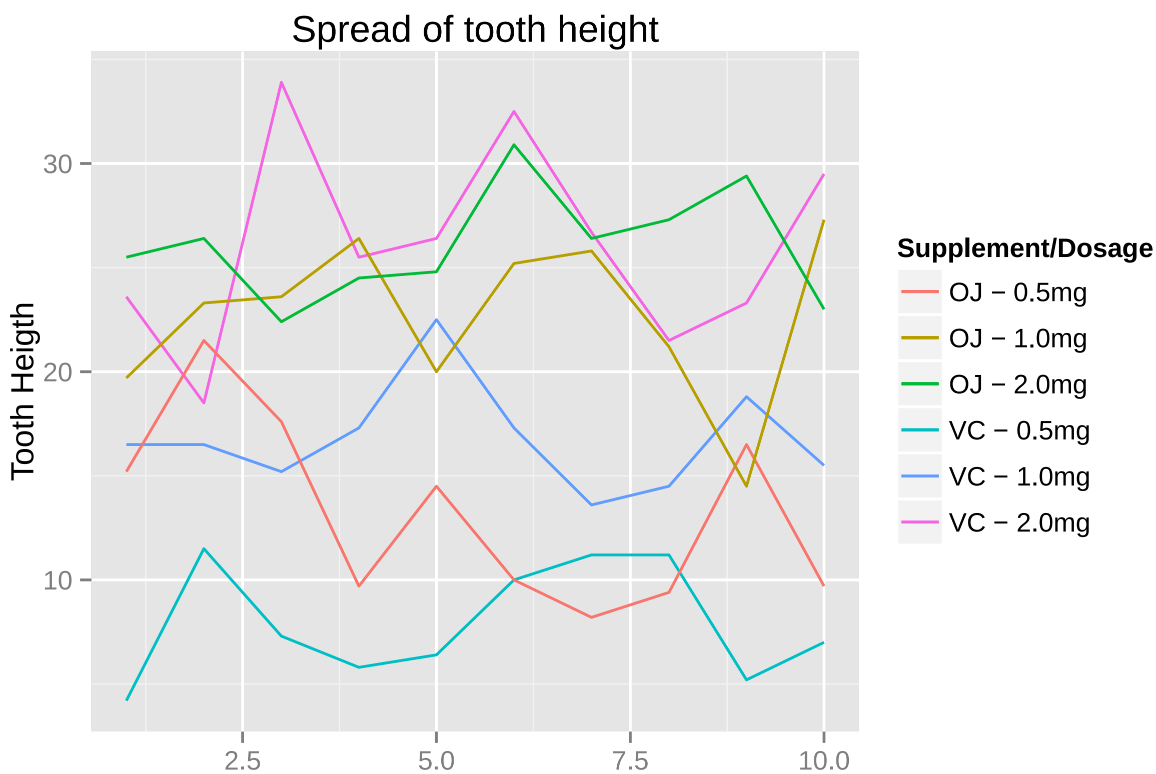
Solution to Statistical Inference course project (Section 2).

### Q1. Load the ToothGrowth data and perform some basic exploratory data analyses

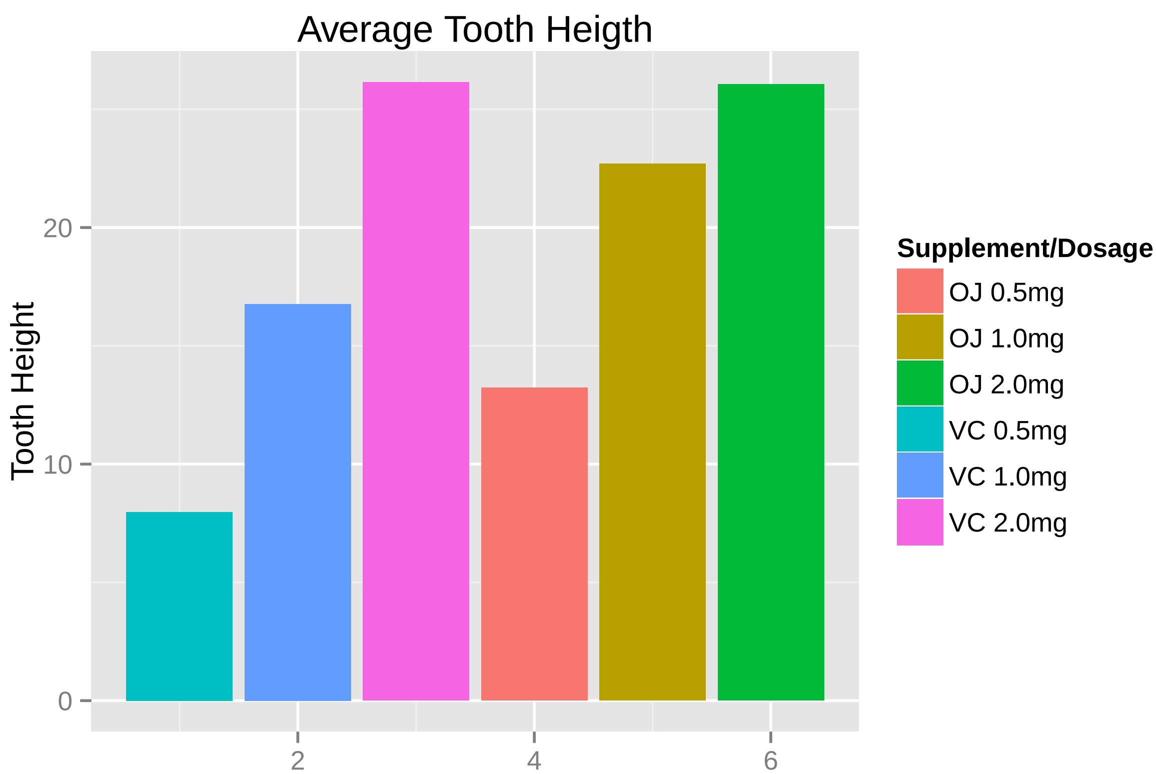
```r
#Loading required library and datasets.
library(datasets)
library(ggplot2)
data(ToothGrowth)

#Preview 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
```

```r
#Subset the data by supplement and dosage
suppVC1 <- ToothGrowth$len[1:10]
suppVC2 <- ToothGrowth$len[11:20]
suppVC3 <- ToothGrowth$len[21:30]

suppOJ1 <- ToothGrowth$len[31:40]
suppOJ2 <- ToothGrowth$len[41:50]
suppOJ3 <- ToothGrowth$len[51:60]

exPlot <- qplot(x = 1:10, y = suppVC1, geom = 'line', color = "VC - 0.5mg",
                ylab = "Tooth Heigth", xlab ="", main="Spread of tooth height") +
     geom_line(aes(y = suppVC2, color = "VC - 1.0mg")) +
     geom_line(aes(y = suppVC3, color = "VC - 2.0mg")) +
     geom_line(aes(y = suppOJ1, color = "OJ - 0.5mg")) +
     geom_line(aes(y = suppOJ2, color = "OJ - 1.0mg")) +
     geom_line(aes(y = suppOJ3, color = "OJ - 2.0mg")) +
     guides(color=guide_legend(title="Supplement/Dosage"))
exPlot
```




```
suppMeans <- c(mean(suppVC1), mean(suppVC2), mean(suppVC3),
               mean(suppOJ1), mean(suppOJ2), mean(suppOJ3))
names(suppMeans) <- c("VC 0.5mg", "VC 1.0mg", "VC 2.0mg", "OJ 0.5mg", "OJ 1.0mg", "OJ 2.0mg")

#Plot the average tooth heigth for each supplemt and dosage.
meanPlot <- qplot(seq_along(suppMeans), suppMeans, geom = "bar", stat="identity",
          fill = names(suppMeans), xlab ="", ylab = "Tooth Height",
          main = "Average Tooth Heigth") +
          guides(fill=guide_legend(title="Supplement/Dosage"))
meanPlot
```




### Q2 . Provide a basic summary of the data.

The mean and variance for each dataset is given below.

```
#Average heigth in each case.
suppMeans
```

```
## VC 0.5mg VC 1.0mg VC 2.0mg OJ 0.5mg OJ 1.0mg OJ 2.0mg
##     7.98    16.77    26.14    13.23    22.70    26.06
```

```
#Standerd Deviations
suppSD <- c(sd(suppVC1), sd(suppVC2), sd(suppVC3), sd(suppOJ1), sd(suppOJ2), sd(suppOJ3))
names(suppSD) <- c("VC 0.5", "VC 1.0", "VC 2.0", "OJ 0.5", "OJ 1.0", "OJ 2.0")
suppSD
```

```
##   VC 0.5   VC 1.0   VC 2.0   OJ 0.5   OJ 1.0   OJ 2.0
## 2.746634 2.515309 4.797731 4.459709 3.910953 2.655058
```

From the Exploratory graphs, the following observations can be drawn.

- OJ at 2.0 doses is the most effective supplement and dosage combination
- VC at 0.5 doses is the least effective supplement and dosage combination
- OJ and VC both show increase in Tooth growth as the dosage increased

### Q3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Assuming that they are not paired and their variance differ.

```r
#Comparing OJ and VJ for 0.5 dosage
t.test(suppOJ1, suppVC1, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  suppOJ1 and suppVC1
## 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 of x mean of y
##     13.23      7.98
```

```r
#Comparing OJ and VJ for 0.5 dosage
t.test(suppOJ2, suppVC2, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  suppOJ2 and suppVC2
## 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 of x mean of y
##     22.70     16.77
```

```r
#Comparing OJ and VJ for 0.5 dosage
t.test(suppOJ3, suppVC3, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  suppOJ3 and suppVC3
## t = -0.0461, 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 of x mean of y
##     26.06     26.14
```

In the first two cases (dosage 0.5mg and 1.0mg), the p-value is very low indicating that both samples are very different. The null hypothesis is not true.

In the third case, the p-value is high, this suggest that if the dosage is amplified to 2.0mg, the results are similar for both supplements.

For the first and second scenario, where the dosage is 0.5mg and 1.0mg respectively, the confidence intervals are above zero. This tells us that OJ is always better for those doses.

In the third case (dosage 2.0mg) where the confidence interval is -3.79807 to 3.63807, there is mot much to choose between OJ and VC as they both produced similar results.

## Q4.State your conclusions and the assumptions needed for your conclusions.

From the above analysis, the following conclusions can be drawn.

- OJ is more efficient that VS overall
- OJ is a much better supplement to VC especially in low dosages.

These Conclusions are based on the following assumptions

- The guinea pigs were selected by random draw and the sampling is not biased
- The samples are not paired
- The variances are inequal