# Inferential Data Analysis

*Pravin Kumar Choudhary*

*December 27, 2015*

## Overiew

In this project, we will perform exploratory data analysis on ToothGrowth data (response is the length of odontoblasts (teeth) in each of 10 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)). First we will try to understand the data distribution and then using hypothesis / confidence interval test, we will try to find similarity between Tooth Growth using different supplement and/or doses.

## Data Setup

In this phase, we will load the data and create separate data sets for Tooth Growth length data for given Supplement and Dose.

You can also embed plots, for example:

```
library(ggplot2)
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 3.2.2
```

```
library(datasets)
data("ToothGrowth")

# Create seperate data set for given supplement type
ToothGrowthBysuppOJ <- subset(ToothGrowth, supp == 'OJ')
ToothGrowthBysuppVC <- subset(ToothGrowth, supp == 'VC')

# Create seperate data set for given supplement dose
ToothGrowthBysuppOJDose1 <- subset(ToothGrowthBysuppOJ, dose == .5)
ToothGrowthBysuppOJDose2 <- subset(ToothGrowthBysuppOJ, dose == 1)
```

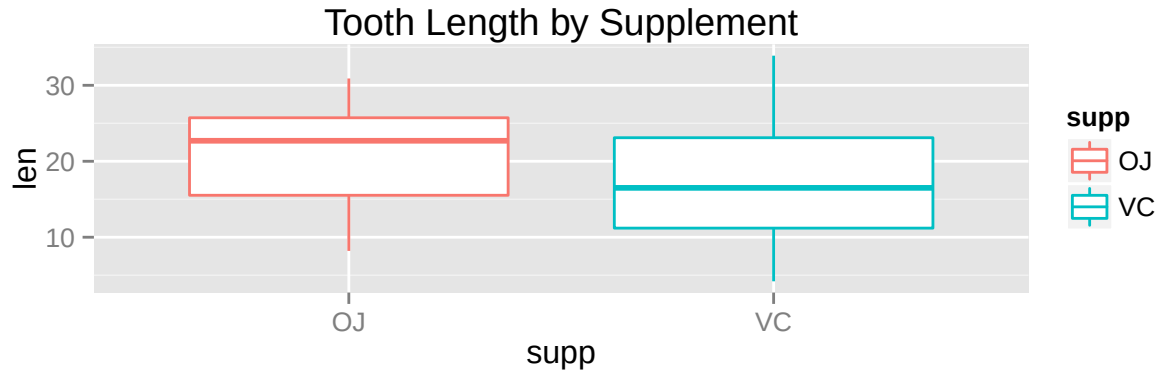
## Data Description

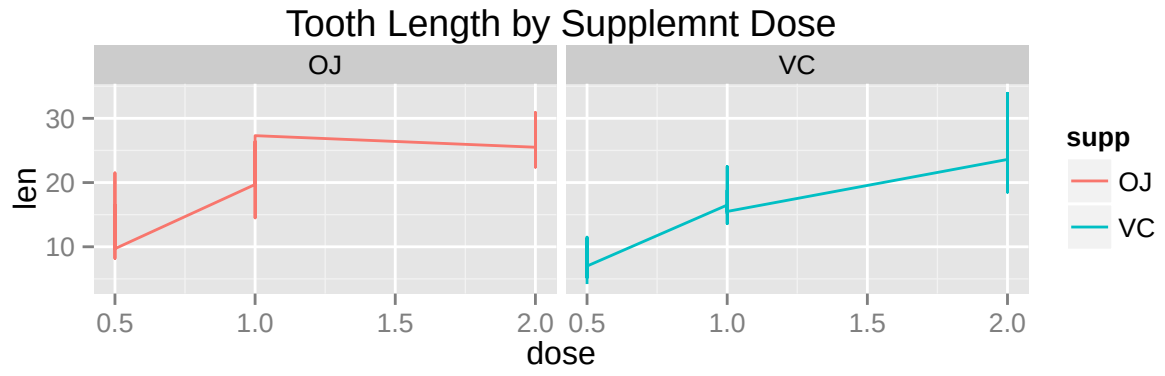
In this phase we will be be describing data attribute (e.g. range, comparative study for a given field)

```
p1<- ggplot(ToothGrowth, aes(x = supp, col = supp, y = len)) + geom_boxplot() +
    ggtitle("Tooth Length by Supplement")

p2<- ggplot(ToothGrowth, aes(x = dose, col = supp, y = len)) + facet_grid(. ~ supp) +
    geom_line() + ggtitle("Tooth Length by Supplemnt Dose")

grid.arrange(p1, p2, nrow=2)
```




### statistical Inference of Tooth Lengh by Supplement

We will use 'T' Statistics to understand impact of use of supplement on Tooth Growth. Here null hypothesis would be that there is change in Tooth Growth by Supplement, whereas Alternate Hypothesis would be otherwise.

Please note that here data is not paired and also variance of each sample list is not equal.

```
t.test(ToothGrowthBysuppOJ$len, ToothGrowthBysuppVC$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowthBysuppOJ$len and ToothGrowthBysuppVC$len
## 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 of x mean of y
##  20.66333  16.96333
```

### Conclusion

We conclude that Tooth Growth is similar for both types of Supplement. i.e. we Fail to reject Null Hypothesis as

- p-value is high (than 0.05) for 95% confidence interval
- Confidence interval contains 0

## statistical Inference of Tooth Lengh by Supplement Dose

We will use 'T' Statistics to understand impact different use of different dose for a given supplement on Tooth Growth. Since we are using two group interval, will compare two dose (0.5 & 1) for OJ supplement. Here null hypothesis would be that there is change in Tooth Growth by Dose, whereas Alternate Hypothesis would be otherwise. Please note that here data is not paired and also variance of each sample list is not equal.

```
t.test(ToothGrowthBysuppOJDose2$len, ToothGrowthBysuppOJDose1$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowthBysuppOJDose2$len and ToothGrowthBysuppOJDose1$len
## t = 5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.524366 13.415634
## sample estimates:
## mean of x mean of y
##     22.70     13.23
```

### Conclusion

We conclude that Tooth Growth higher for higher dose of a supplement. i.e. we reject Null Hypothesis as

- p-value is low (than 0.05) for 95% confidence interval
- Confidence interval is positive for different of Tooth Growth for dose of 1 and .5