# Basic Inferential Data Analysis.

*Aravind Sesetty*

*Saturday, January 24, 2015*

---

## Overview

**ToothGrowth Dataset Description:**
The 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).

We're going to analyze the ToothGrowth data in the R datasets package.
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.

**Setting Global Option for knitr**

- Set `echo=TRUE` so that the R code is displayed in the Report.
- Set `results='asis'` so that the result from R code is displayed in the Report.
- Set `fig.width=5` (Plot Width) & `fig.height=5` (Plot Height).

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

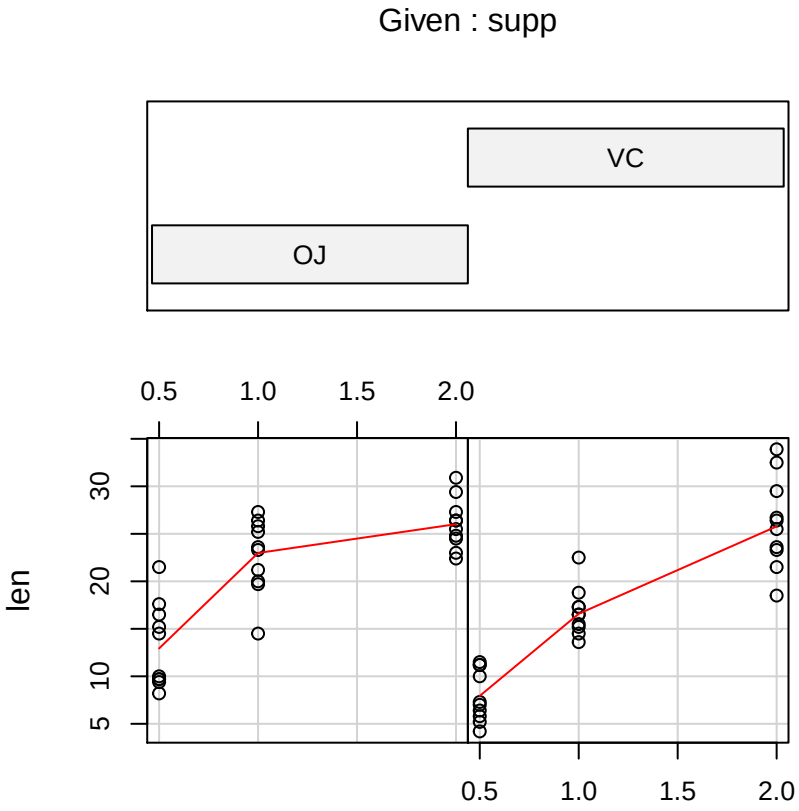
```
data(ToothGrowth)
str(ToothGrowth)
```

'data.frame': 60 obs. of 3 variables: $ len : num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ... $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ... $ dose: num 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...

**Observation**
The ToothGrowth dataset has 60 observations with 3 variables i.e., Length, Supplement(OJ, VC), Dose(0.5, 1, 2)

```
require(graphics)

coplot(len ~ dose | supp,
       data = ToothGrowth,
       panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```




ToothGrowth data: length vs dose, given type of supplement

**Observation**
Given the Supplements (OJ, VC), the exploratory graph shows that the length of teeth of three dose levels of Vitamin C (0.5, 1, and 2 mg) is better with VC

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

```
xtable(summary(ToothGrowth), type="pdf")
```

% latex table generated in R 3.1.1 by xtable 1.7-4 package % Sun Jan 25 14:07:53 2015

| | len | supp | dose |
|---|---|---|---|
| 1 | Min. : 4.2 | OJ:30 | Min. :0.50 |
| 2 | 1st Qu.:13.1 | VC:30 | 1st Qu.:0.50 |
| 3 | Median :19.2 | | Median :1.00 |
| 4 | Mean :18.8 | | Mean :1.17 |
| 5 | 3rd Qu.:25.3 | | 3rd Qu.:2.00 |
| 6 | Max. :33.9 | | Max. :2.00 |

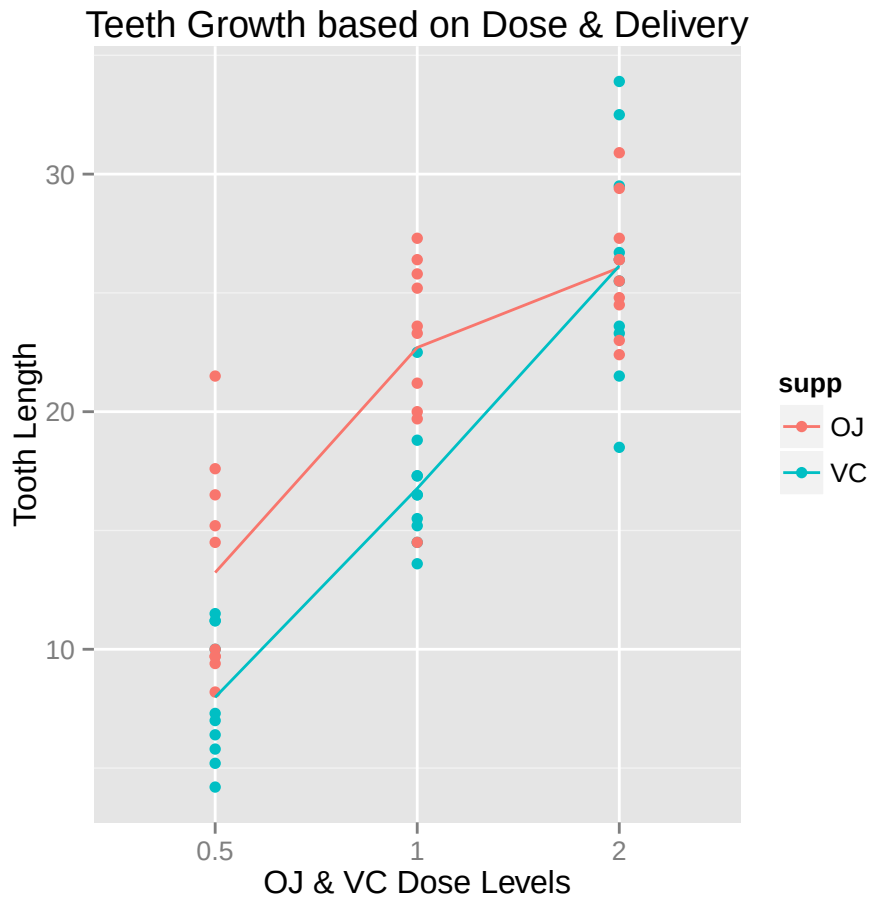
```
xtable(table(ToothGrowth$dose, ToothGrowth$supp), type="pdf")
```

% latex table generated in R 3.1.1 by xtable 1.7-4 package % Sun Jan 25 14:07:53 2015

| | OJ | VC |
|---|---|---|
| 0.5 | 10 | 10 |
| 1 | 10 | 10 |
| 2 | 10 | 10 |

```
meanLength <- aggregate(formula = len ~ ., data = ToothGrowth, FUN = mean)

ggp <- ggplot(ToothGrowth, aes(x = factor(dose), y = len)) +
        geom_point(aes(colour = supp)) +
        labs(title = "Teeth Growth based on Dose & Delivery") +
        labs(x = "OJ & VC Dose Levels", y = "Tooth Length") +
        geom_line(data = meanLength, aes(group=supp, color=supp))
print(ggp)
```




**Observation**

The summary of ToothGrowth shows that
-Tooth Length Min = 4.20, Mean = 18.81, Max = 33.90
-OJ Supplement samples 30, VC Supplement samples 30
-OJ, VC Doseage Min = 0.5, Mean = 1.2, Max = 2

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

### Tooth Growth by Supplement Type

```
# Hypothesis 1 that VC improves tooth growth more than OJ
t.test(ToothGrowth$len[ToothGrowth$supp == "OJ"],
       ToothGrowth$len[ToothGrowth$supp == "VC"],
       paired = FALSE)
```

```
Welch Two Sample t-test
```

data: ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "VC"] t = 1.915, df = 55.31, p-value = 0.06063 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.171 7.571 sample estimates: mean of x mean of y 20.66 16.96

### Tooth Growth by Dose Level

```
# Hypothesis 2 that dosage improves growth
t.test(ToothGrowth$len,
       ToothGrowth$dose,
       paired = TRUE)
```

```
Paired t-test
```

data: ToothGrowth$len and ToothGrowth$dose t = 19.11, df = 59, p-value < 2.2e-16 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: 15.80 19.49 sample estimates: mean of the differences 17.65

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

**Hypothesis 1: Welch Two Sample t-test**
An independent-sample t-test was used to check the effectiveness of supplements (OJ, VC) in greater tooth length, t(55)=1.91, p=0.06 with supplement (OJ, VC) associated with more Tooth growth using VC than OJ is accepted as the p-value is >0.05

**Hypothesis 2: Paired t-test**
An paired-sample t-test was used to check the effectiveness of Dose (0.5, 1, 2) in greater tooth length, t(59)=19.1, p< 2.2e-16 with Dose (0.5, 1, 2) associated with more Tooth growth is rejected as the p-value is <0.05.