

# Statistical Inference – Data Analysis Part 2

## The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Author Name: DataRacer11

9-26-2015

### References:

- McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.
- Crampton, E. W. (1947) The growth of the odontoblast of the incisor teeth as a criterion of vitamin C intake of the guinea pig. The Journal of Nutrition 33(5): 491–504. <http://jn.nutrition.org/content/33/5/491.full.pdf>
- C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.
- R Tutorials - William B. King, Ph.D, Coastal Carolina University, <http://ww2.coastal.edu/kingw/statistics/R-tutorials/factorial.html>

## Overview:

"ToothGrowth is a library dataset available in R. The dataset includes the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC))." The goal of this report is to analyze the ToothGrowth data of different dosage amounts of the Vitamin C. There are two supplement types of Vitamin C tested, Orange Juice and Ascorbic Acid, and they are given in three different milligram dosage amount, 0.5, 1.0, and 2.0. Tests involving some 650 pigs, were aimed at establishing an assay procedure that could achieve greater precision, simplicity and reliability. Archived data are available from studies focused on the length of odontoblasts of the incisor teeth as a criterion of vitamin C intake of guinea pigs.

This report has been developed to respond to the following questions: 1. Was exploratory data analysis performed with at least a single plot or table highlighting basic features of the data? 2. Were relevant confidence intervals and/or tests performed on the data? 3. Were results of the tests and/or intervals interpreted in the context of the problem correctly? 4. Were assumptions described as needed to make conclusions?

### Load the ToothGrowth data and perform some basic exploratory data analyses:

```
library(datasets)
```

### Obtain a basic summary of the data. A snapshot of the data.frame:

```
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 ...

summary(ToothGrowth)
> lensupp 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
```

Observe that the data.frame after the library is loaded in R highlights 60 observations on 3 sets with 3 or more variables.

1. len | numeric | Tooth length
2. supp | factor | Supplement type (VC or OJ).
3. dose | numeric | Dose in milligrams/day

### Investigate the data table further

```
dt<-data.frame(ToothGrowth)
```

## Exploratory Data Analysis

```
head(dt)
  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

summary(dt)
> 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
```

### Correlating averages

The dataset ToothGrowth contains measurements on the effects on tooth length (len). These are based on variations of the Vitamin C dosages given to guinea pigs with the goal to gain insight to a more precise conclusion regarding researcher's idea that an increase in Vitamin C would promote tooth growth. This can be checked by correlating the ToothGrowth (len) with ToothGrowth (dose).

```
cor(ToothGrowth$dose, ToothGrowth$len)
> [1] 0.8026913
```

For each level of dosage, there are several experimental units, as the experiment is replicated. Researchers use this as a common experimental design to average and then take the correlation of the averaged data. This can change the correlation significantly. The data can be split into three dosages groups and the dosages values can then be computed as well as the group averages.

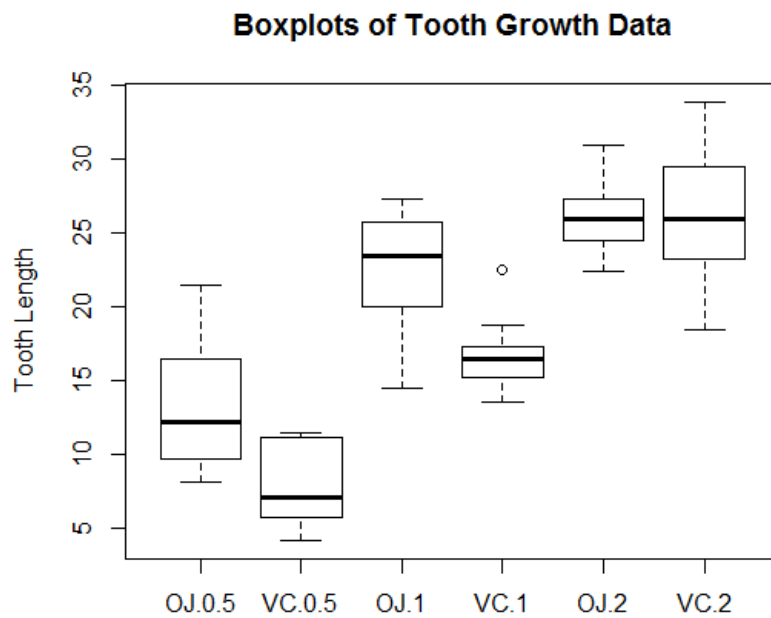
```
l <-split(ToothGrowth$len, ToothGrowth$dose)
group_means <- c(mean(l[[1]]), mean(l[[2]]), mean(l[[3]]))
cor(c(0.5, 1, 2), group_means)
> [1] 0.9574428
```

The value of 0.95 for the aggregated data is higher than the value 0.80 for the individual data. In general, correlations formed from averages are typically close to 1 or -1 than when all of the data is considered individually.

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

**There are several ways that we can plot for data visualization such as this boxplot.**

```
boxplot(len ~ supp * dose, data=ToothGrowth, ylab="Tooth Length", main="Boxplots of Tooth Growth Data")
```

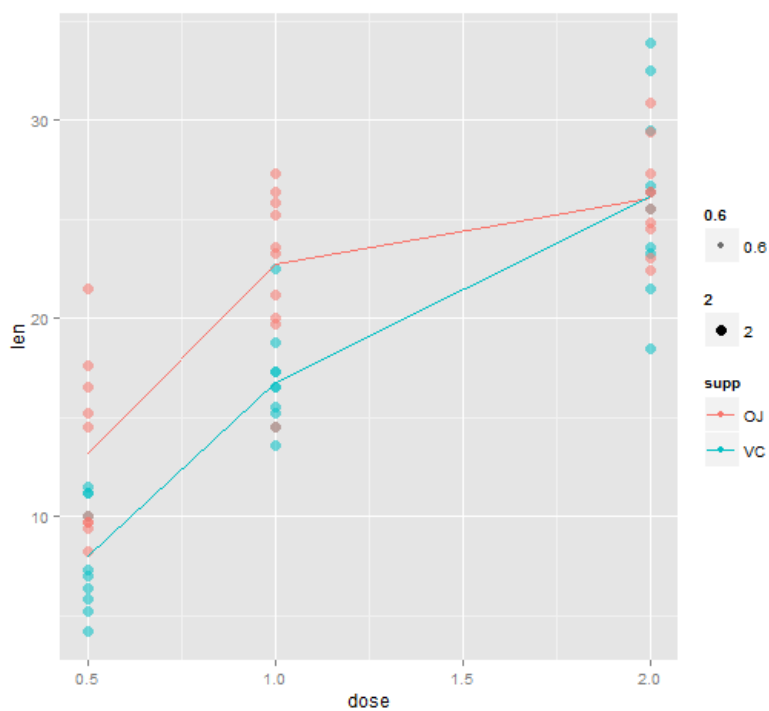


Calculate the length and mean by supplement and dose. Create a scatterplot to evaluate the doses

```
avgs<- aggregate(len~., data=ToothGrowth, mean)
```

plot the scatter points and the averages

```
library(ggplot2)
g <- ggplot(ToothGrowth, aes(x=dose, y=len))
g <- g + geom_point(aes(group=supp, colour=supp, size=2, alpha=0.6))
g <- g + geom_line(data=avgs, aes(group=supp, colour=supp))
print(g)
```



# Compare tooth growth by supp and dose

---

In order to understand how Vitamin C affects tooth growth, the following conclusions below have been made while investigating confidence intervals. Supplement and dosage baselines were taken into consideration.

Let's look at our hypothesis that with increased Vitamin C dosage from 0.5 to 1.0 milligrams, and from 1.0 to 2.0 milligrams the confidence interval does not contain zero. Therefore, we can conclude that dosage does effect the rate of tooth growth in guinea pigs:

To test this hypothesis, create subsets of the data for conducting several T-Tests to better understand the confidence interval. To analyze tooth growth, we carry out two-sample (orange juice vs. ascorbic acid) T-tests at each dosage level.

## T-Test at .5 mg

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ])  
  
> 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 OJ mean in group VC  
>13.23 7.98
```

## T-Test at 1 mg

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1, ])  
>  
> 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 OJ mean in group VC  
>22.7016.77
```

## T-Test at 2 mg

```
> t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 2, ])  
>  
> 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 OJ mean in group VC  
>26.0626.14
```

## Summary and Assumptions:

- The confidence intervals are assumed to not be paired, comparing two different supplement types from individual guinea pig did not take place.
- The dataset contains six measurements for ten guinea pigs each, three for each dose for each of the two supplement types. As it is not the same ten guinea pigs for each of the six measurements, we cannot effectively analyze the differences of the three different dose measurements (for each supplement type).

- We cannot assume equal variances within each group. As we only have  $n=10$  observations, T-tests were performed with  $\alpha=0.05$
- The samples are independent.
- The distribution approximately is normal.

## **Conclusions:**

- As Vitamin C dose size alone increases, the tooth length increases as well
- We have analyzed the influence of both the dosage and the supplement type of Vitamin C on the length of the guinea pigs' teeth.
- We have found that the dosage very likely has an impact on the length of the teeth, whereas the supplement type has no significant impact.
- The supplement type of Orange Juice affects tooth length greater than Ascorbic Acid, or Vitamin C with a 0.5 and 1.0 dose size.
- When the dose size reached 2.0 milligrams, the effect between Orange Juice and Ascorbic Acid is very similar.