

# Statistical Inference: Peer Assessment, Part 2

*S C NG*

*20 June, 2015*

## Overview

In the second portion of the assignment, we're going to analyze the ToothGrowth data in the R datasets package. We will use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose, and make the conclusion with assumptions needed.

## Setup

First of all, the following default settings and libraries are loaded

```
#preset default options for Rmd, codes not shown in report  
require(knitr)
```

```
## Loading required package: knitr
```

```
## Warning: package 'knitr' was built under R version 3.1.3
```

```
opts_chunk$set(cache=TRUE, echo=TRUE)
```

```
#load required libraries for data analysis  
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
require(datasets)
```

## Exploratory of Data

According to the help page, the ToothGrowth dataset recorded 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). Below R code shows help page and the data structure of ToothGrowth dataset

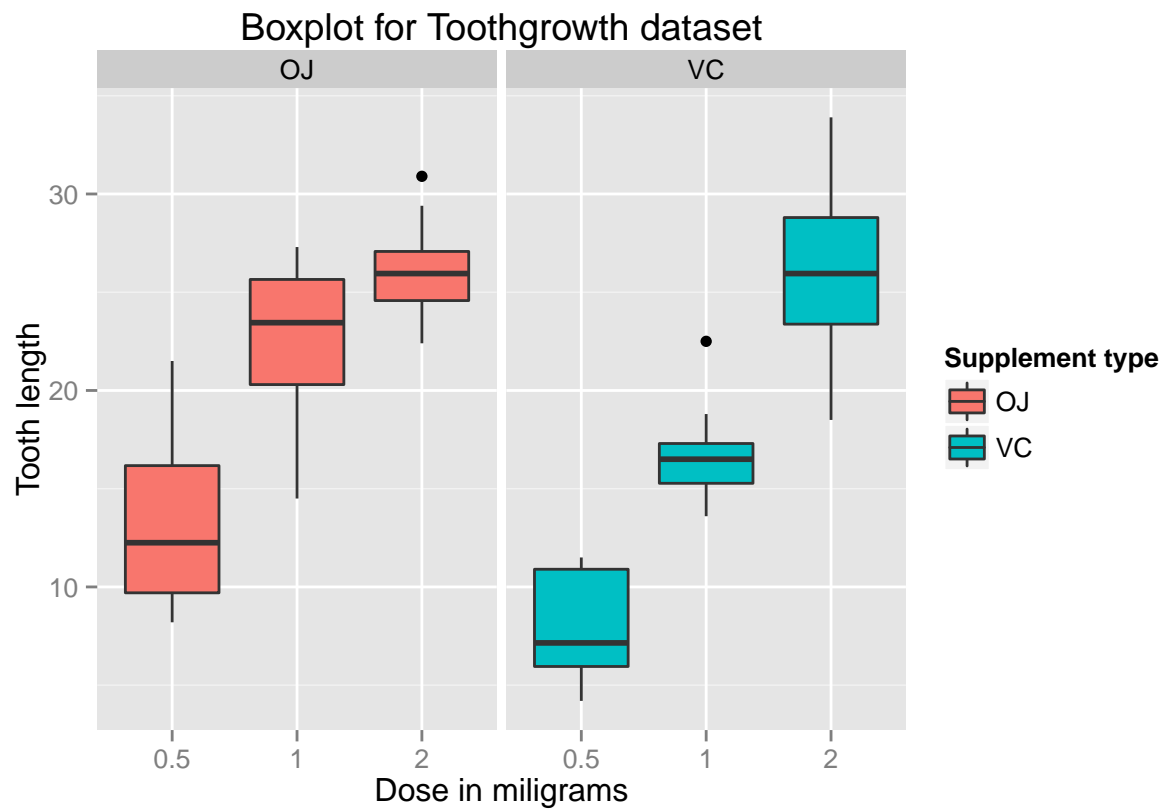
```
?ToothGrowth  
str(ToothGrowth)
```

```
## '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 ...
```

The dataset consists of 60 observations on 3 variables: 1. len (numeric) = Tooth length 2. supp (factor) = Supplement type (VC or OJ). 3. dose (numeric) = Dose in milligrams.

Below boxplot summarize the distribution of Tooth length values per each three dose levels, 1 plot per each Supplement type

```
g2 <- ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp))
g2 <- g2 + geom_boxplot()
g2 <- g2 + facet_grid(. ~ supp)
g2 <- g2 + xlab("Dose in miligrams") + ylab("Tooth length")
g2 <- g2 + guides(fill = guide_legend(title = "Supplement type"))
g2 <- g2 + ggtitle("Boxplot for Toothgrowth dataset")
g2
```



## Data Analysis

The following section tried to use confidence intervals and hypothesis tests to compare the tooth growth by supplement type and dose

We will start with comparing the tooth length by supplement type through the following test in R

```
t1 <- t.test(len~supp, paired = FALSE, var.equal = TRUE, data=ToothGrowth)
t1
```

```
##
## Two Sample t-test
##
```

```
## data: len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1670064 7.5670064
## sample estimates:
## mean in group OJ mean in group VC
## 20.66333 16.96333
```

TODO: Analysis???

Then we will check the tooth length by dose TODO: Analysis between len and dose???

## Assumptions

For the above tests, we assume the len values in the dataset are normally distributed