

# Statistical Inference (ToothGrowth)

Prateek Sarangi

## Introduction

In the second part of Assignment, 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 hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there are other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

## Data analysis

### 1. Load data

In this project, we will analyze the effect of Vitamin C on tooth growth in guinea pigs. Dataset is come from the ToothGrowth data in the R package.

First we need to load the datasets

```
library(datasets)
data(ToothGrowth)
str(ToothGrowth)
```

### 2. Show basic data of dataset (Question 1)

Lets get some data from our dataset:

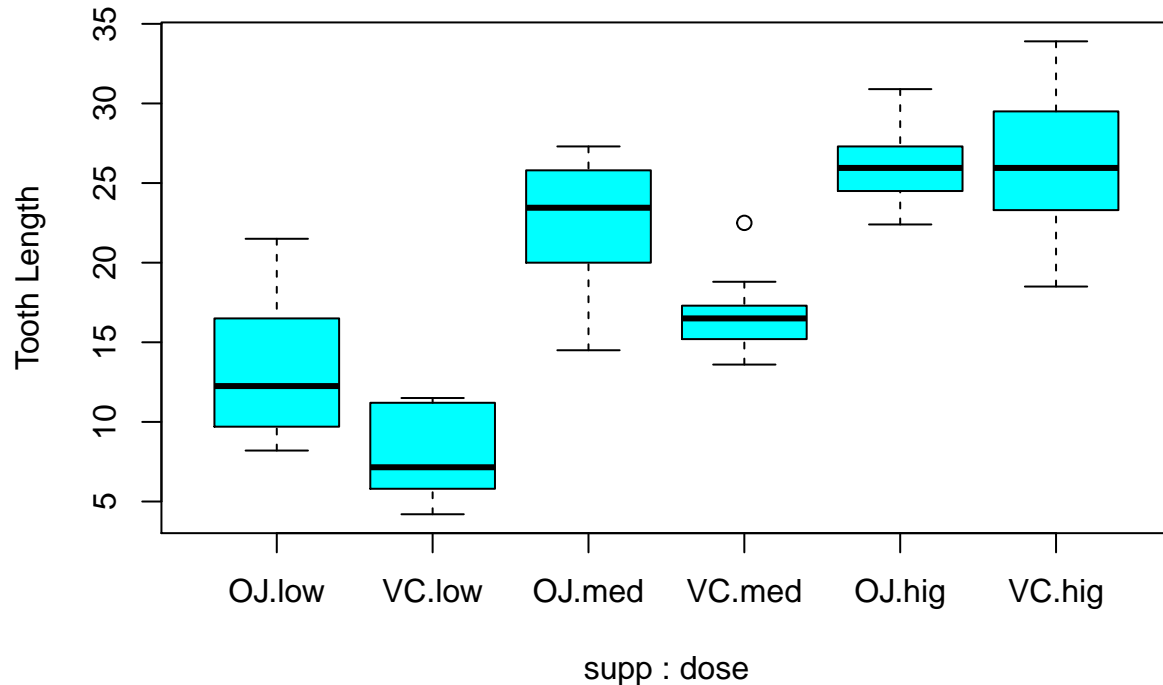
```
ToothGrowth$dose = factor(ToothGrowth$dose, levels=c(0.5,1.0,2.0),
                           labels=c("low","med","hig"))
replications(len ~ supp * dose, data=ToothGrowth)
```

```
##      supp      dose supp:dose
##      30      20      10
```

We can also plot some data to understand the data:

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

## Boxplots of Tooth Growth Data



Here above the results from dataset. First code shows us number of data for our set. We created a three levels of doses to show how much data are in each category we would like to use.

Our plot(second code) shows us the length of teeth in each of 10 guinea pigs at each of three doses of Vitamin C (levels: 0.5, 1, and 2 mg). Each dose is presented with two delivery methods (orange juice marked as “OJ” and ascorbic acid marked as “VC”). Totally we have 60 observations.

### 3. Basic summary (Question 2)

We can get a summary of the above plot using this code:

```
with(ToothGrowth, tapply(len, list(supp,dose), var))
```

```
##      low      med      hig
## OJ 19.889 15.295556 7.049333
## VC  7.544  6.326778 23.018222
```

or using a mean of each values:

```
with(ToothGrowth, tapply(len, list(supp,dose), mean))
```

```
##      low  med  hig
## OJ 13.23 22.70 26.06
## VC  7.98 16.77 26.14
```

### 4. Confidence/Hypothesis for tests (Question 3)

Because of only 60 observations we can use only a t-confidence intervals. We would like to show that dose of Orange Juice is higher than Ascorbic Acid (VC).

```
TG_OJ_low <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose=="low",1]
TG_VC_low <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose=="low",1]
t.test(TG_OJ_low, TG_VC_low, alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: TG_OJ_low and TG_VC_low
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 2.34604      Inf
## sample estimates:
## mean of x mean of y
##      13.23      7.98
```

As we can see - the Orange Juice (0.5mg) dose is greater has a higher value than basic Ascorbic Acid (Vitamin C)

```
TG_OJ_low <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose=="med",1]
TG_VC_low <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose=="med",1]
t.test(TG_OJ_low, TG_VC_low, alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: TG_OJ_low and TG_VC_low
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 3.356158      Inf
## sample estimates:
## mean of x mean of y
##      22.70      16.77
```

For the medium value of dose (1mg) also shows us, that Orange Juice dose is greater than basic Ascorbic Acid (Vitamin C)

```
TG_OJ_low <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose=="hig",1]
TG_VC_low <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose=="hig",1]
t.test(TG_OJ_low, TG_VC_low, alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: TG_OJ_low and TG_VC_low
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -3.1335      Inf
## sample estimates:
## mean of x mean of y
##      26.06      26.14
```

The third test also shows us, that high dose (2mg) of Orange Juice is greater to the value from Vitamin C. But - values are very similar and we can not say that it is definitely higher.

## 5. Conclusions (Question 4)

We have 95% confidence, that dosage levels for low and medium values (0.5 and 1 mg) for Orange Juice causes larger tooth growth than Vitamin C dosage. However, when the dosage level is high (2 mg), there isn't statistical difference in tooth growth between two dosage types.