

# Statistical Inference Project:ToothGrowthdata analysis

*Reshu*

*1 June 2017*

## Overview

In this assignment we will analyze the ToothGrowth data in R datasets package. We will Load ToothGrowth data & perform some basic exploratory analysis. Then we will use hypothesis tests to compare tooth growth by supp and dose.

## 1.Data Exploration

```
data("ToothGrowth") #Loading ToothGrowth dataset
dim(ToothGrowth) #dimensions of ToothGrowth dataset
```

```
## [1] 60 3
```

```
str(ToothGrowth) #Structure of ToothGrowth dataset
```

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

As shown above, Toothgrowth dataset is a dataframe, which has 60 obs of 3 variables - len(numeric), supp(Factor with 2 levels "OJ", "VC"), dose (numeric).

## 2.Summary of the dataset is shown as follows

```
summary(ToothGrowth) #Summary of ToothGrowth dataset
```

```
##      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
```

## 3.Comparison of Tooth growth by supp and dose

*#Comparison by dose*

```
d1<-subset(ToothGrowth,dose %in% c(0.5,1.0))
d2<-subset(ToothGrowth,dose %in% c(0.5,2.0))
d3<-subset(ToothGrowth,dose %in% c(1.0,2.0))
t.test(len~dose,paired=F,var.equal=F,data=d1)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.983781 -6.276219
## sample estimates:
## mean in group 0.5 mean in group 1
## 10.605 19.735
```

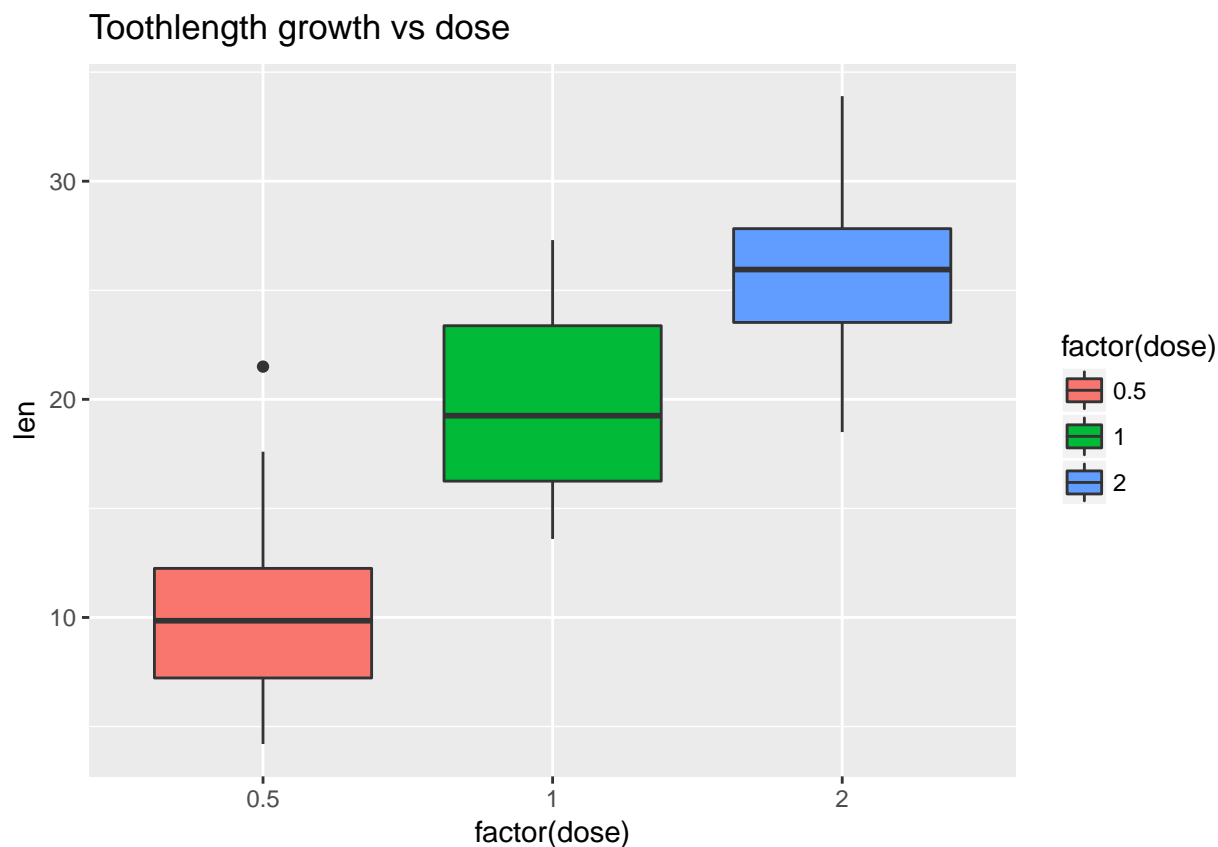
```
t.test(len~dose,paired=F,var.equal=F,data=d2)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -18.15617 -12.83383
## sample estimates:
## mean in group 0.5 mean in group 2
## 10.605 26.100
```

```
t.test(len~dose,paired=F,var.equal=F,data=d3)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
## 19.735 26.100
```

```
ggplot(aes(x=factor(dose),y=len),data=ToothGrowth) + geom_boxplot(aes(fill=factor(dose))) + ggtitle ("Tooth Growth by Dose")
```



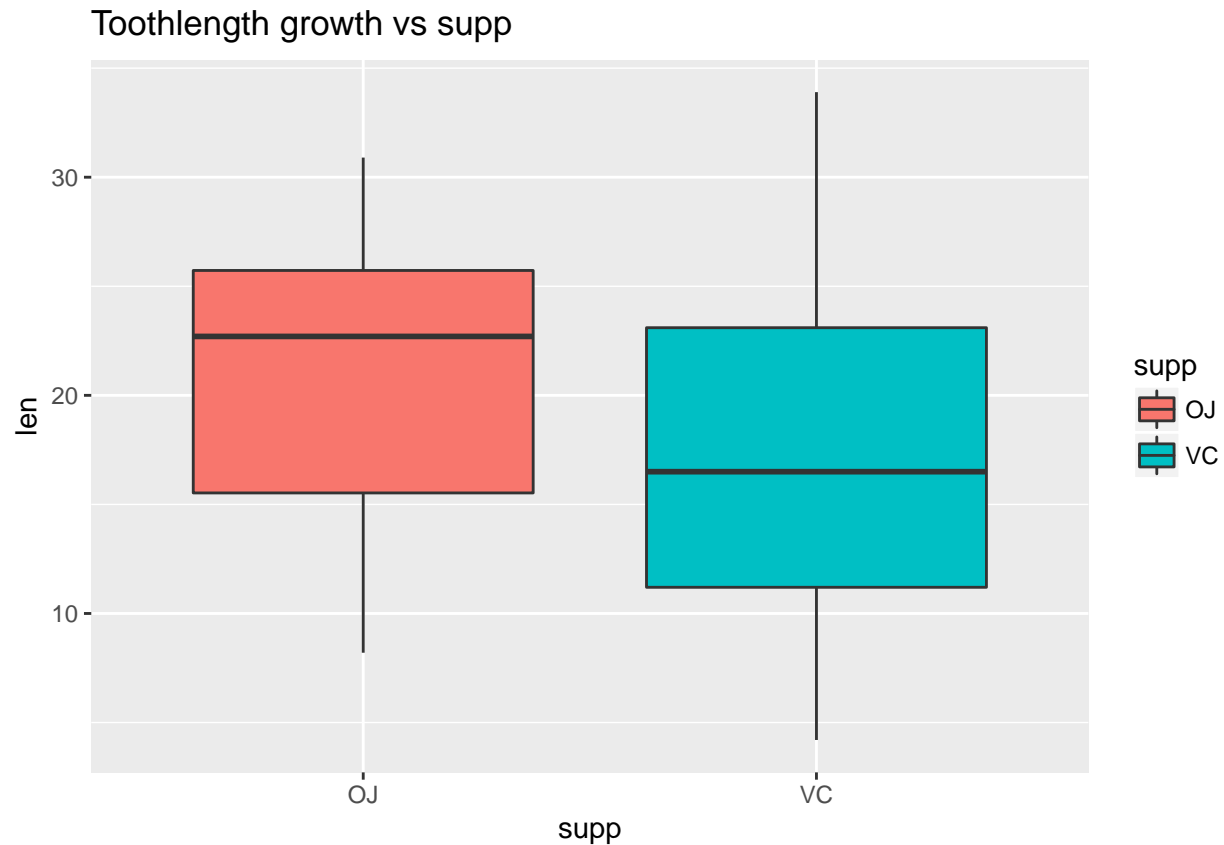
The confidence intervals and p-values are as follows -for dosage 0.5 and 1.0  $[-11.98, -6.276]$  ,p-value= $1.268e-07$  -for dosage 0.5 and 2.0  $[-18.16, -12.83]$  ,p-value= $4.398e-14$  -for dosage 1.0 and 2.0  $[-8.99, -3.73]$  ,p-value= $1.906e-05$  p-value is very small as compared to 0.05(default value for error tolerance alpha ) so, We reject null hypothesis and deduce that there is correlation between dosage and tooth length

*#Comparison by supp*

```
t.test(len~supp,paired=F,var.equal=F,data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##      20.66333      16.96333
```

```
ggplot(aes(x=supp,y=len),data=ToothGrowth) + geom_boxplot (aes(fill=supp)) +ggtitle("Toothlength growth
```



The confidence interval is  $[-0.171, 7.571]$  for dosage vs sup t-test. Also, p-value is 0.06063, which is greater than 0.05 (alpha default value). So, we accept null hypothesis and deduce that there is no correlation between Supp and toothlength growth.

#### 4. Conclusion

1. Supplement type has no effect on tooth growth. 2. Increasing dose level leads to increased tooth growth.

#### Assumptions

1. Populations are independent, variances between groups are different, and random sample is used.

#### Appendix

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp)) + facet_wrap(~dose) + ggtitle("Plot")
```

Plot showing Toothlength growth vs supp for every dosage

