

# Analyzing Tooth Growth Data

by Robin Chauhan

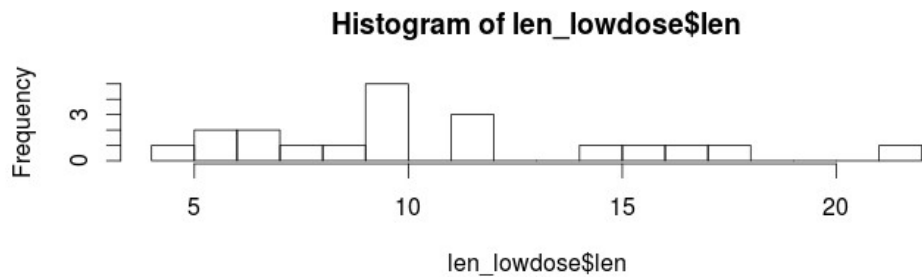
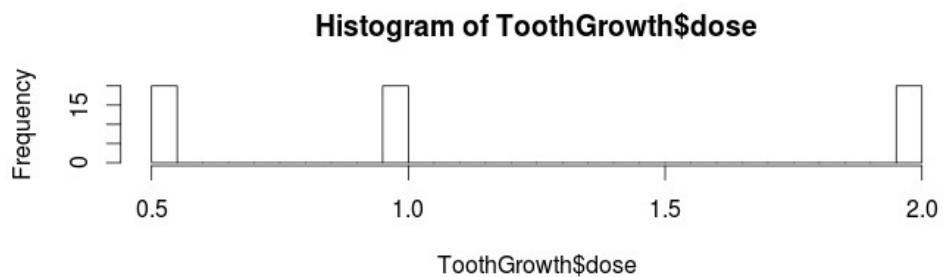
## Exploration

```
library(datasets)
```

```
# basic info
head(ToothGrowth)
summary(ToothGrowth)
unique(ToothGrowth$dose )
unique(ToothGrowth$supp )
```

```
> # basic info
> head(ToothGrowth)
   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(ToothGrowth)
      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
> unique(ToothGrowth$dose )
[1] 0.5 1.0 2.0
> unique(ToothGrowth$supp )
[1] VC OJ
Levels: OJ VC
```

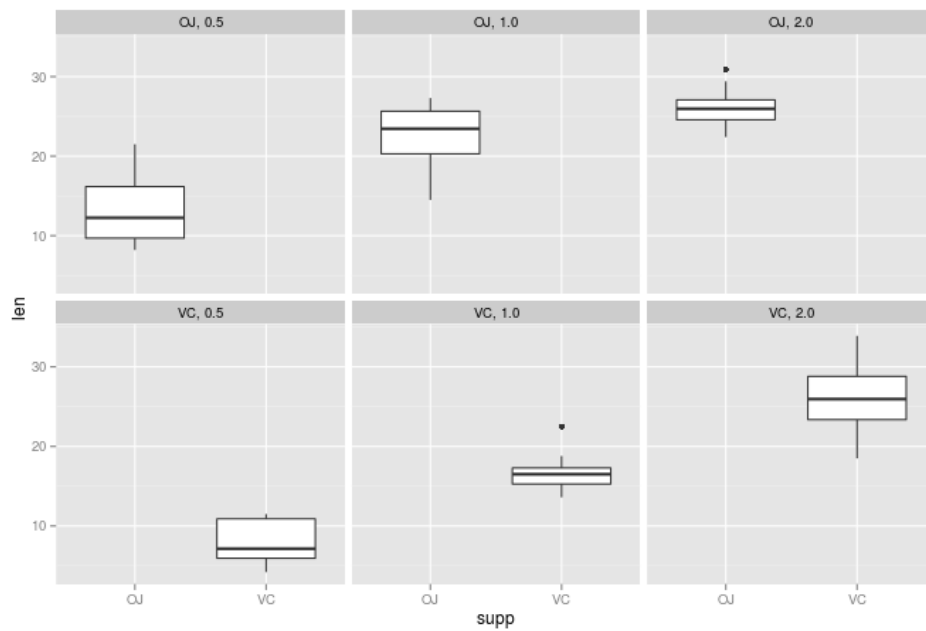
```
# basic plots
par(mfrow=c(2,1))
hist(ToothGrowth$dose, breaks=40)
hist(len_lowdose$len, breaks=20)
```

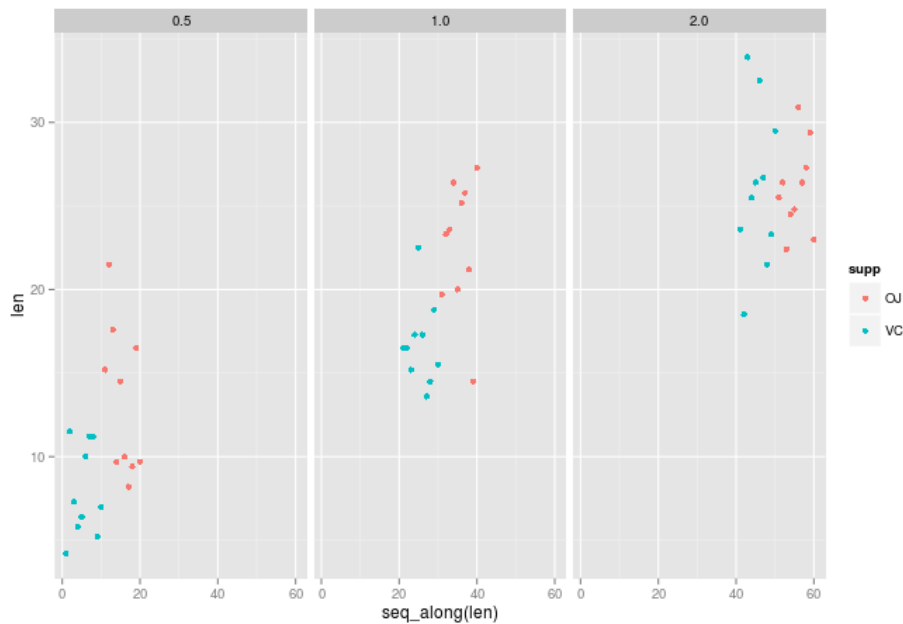


# faceted plots

```
qplot(data=ToothGrowth,x=supp,y=len,facets=~supp+dose,geom="boxplot")
```

```
qplot(data=ToothGrowth,y=len, color=supp, facets=~dose)
```





## Comparing Length based on Supp and Dose

First we take subsets. Then we t-test between various subsets, and find the 95% confidence intervals:

```
# subsetting
lowdose_oj<-subset( ToothGrowth, dose==0.5 & supp=='OJ' )
lowdose_vc<-subset( ToothGrowth, dose==0.5 & supp=='VC' )
meddose_oj<-subset( ToothGrowth, dose==1.0 & supp=='OJ' )
meddose_vc<-subset( ToothGrowth, dose==1.0 & supp=='VC' )
hidose_oj<-subset( ToothGrowth, dose==2.0 & supp=='OJ' )
hidose_vc<-subset( ToothGrowth, dose==2.0 & supp=='VC' )

# t-testing for hypothesis of length change
t.test(lowdose_oj$len,lowdose_vc$len,var.equal=FALSE,conf.level=0.95)$conf.int
t.test(meddose_oj$len,meddose_vc$len,var.equal=FALSE,conf.level=0.95)$conf.int
t.test(hidose_oj$len,hidose_vc$len,var.equal=FALSE,conf.level=0.95)$conf.int

t.test(lowdose_oj$len,meddose_oj$len,var.equal=FALSE,conf.level=0.95)$conf.int
t.test(meddose_oj$len,hidose_oj$len,var.equal=FALSE,conf.level=0.95)$conf.int

t.test(lowdose_vc$len,meddose_vc$len,var.equal=FALSE,conf.level=0.95)$conf.int
t.test(meddose_vc$len,hidose_vc$len,var.equal=FALSE,conf.level=0.95)$conf.int
```

## COMPARING OJ TO VC:

- Null Hypothesis: Mean length does not change between OJ and VC, at same dosage level

### RESULTS

```
> t.test(lowdose_oj$len, lowdose_vc$len, var.equal=FALSE, conf.level=0.95)$conf.int
[1] 1.719057 8.780943
attr(,"conf.level")
[1] 0.95
> t.test(meddose_oj$len, meddose_vc$len, var.equal=FALSE, conf.level=0.95)$conf.int
[1] 2.802148 9.057852
attr(,"conf.level")
[1] 0.95
> t.test(hidose_oj$len, hidose_vc$len, var.equal=FALSE, conf.level=0.95)$conf.int
[1] -3.79807 3.63807
attr(,"conf.level")
[1] 0.95
```

### CONCLUSIONS

- Comparing OJ to VC at different dosage levels, low and med dose result in positive 95% confidence intervals that do NOT contain zero.
- Therefore we can reject the null hypothesis for low and med dosage at 95% confidence
- At high dosage, 95% confidence intervals DO contain zero, thus we CANNOT reject null hypothesis in this case.

## COMPARING LOWER DOSES TO HIGHER DOSES:

- Null hypothesis: dosage does not change mean, for same supplement

### RESULTS

```
> t.test(lowdose_oj$len, meddose_oj$len, var.equal=FALSE, conf.level=0.95)$conf.int
[1] -13.415634 -5.524366
attr(,"conf.level")
[1] 0.95
> t.test(meddose_oj$len, hidose_oj$len, var.equal=FALSE, conf.level=0.95)$conf.int
[1] -6.5314425 -0.1885575
attr(,"conf.level")
[1] 0.95
> t.test(lowdose_vc$len, meddose_vc$len, var.equal=FALSE, conf.level=0.95)$conf.int
[1] -11.265712 -6.314288
attr(,"conf.level")
[1] 0.95
> t.test(meddose_vc$len, hidose_vc$len, var.equal=FALSE, conf.level=0.95)$conf.int
[1] -13.054267 -5.685733
attr(,"conf.level")
[1] 0.95
```

## CONCLUSIONS

- In every case, 95% confidence intervals did NOT contain zero.
- Thus we can reject the null hypothesis at 95% confidence
- We can conclude, dosage does affect length
- Confidence intervals were entirely negative, this we can conclude that increasing dose leads to increase length

## ASSUMPTIONS

- Data is not paired
- t statistic used due to small sample size
- We assume variance is not the same between groups
- This is a 2-tailed test (testing for SIGNIFICANT DIFFERENCE in means in EITHER DIRECTION)
- Lengths are normally distributed