# ToothGrow Analysis

## Anandu R

## 9/20/2020

## ToothGrowth data analysis

Loading the data
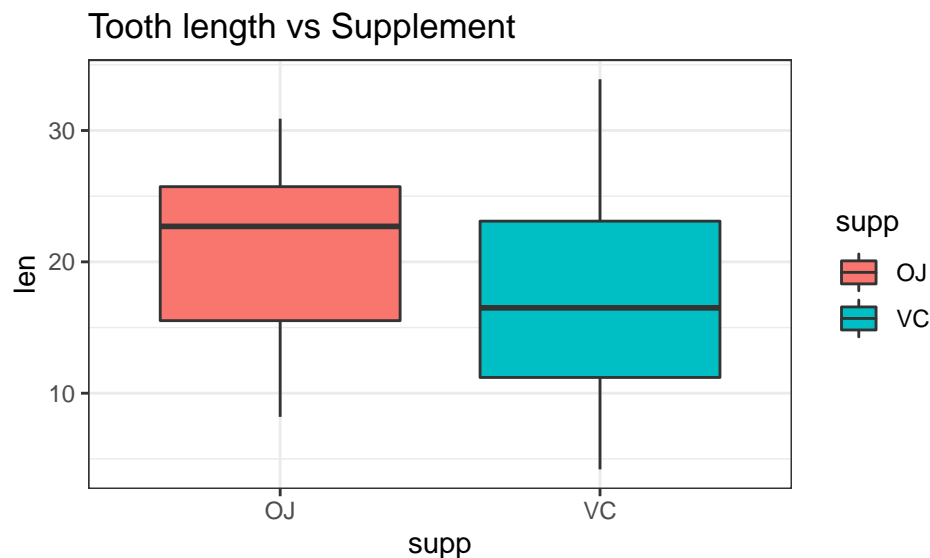
```
data("ToothGrowth")
```

Summarizing the data

```
ToothGrowth$dose = factor(ToothGrowth$dose)
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```
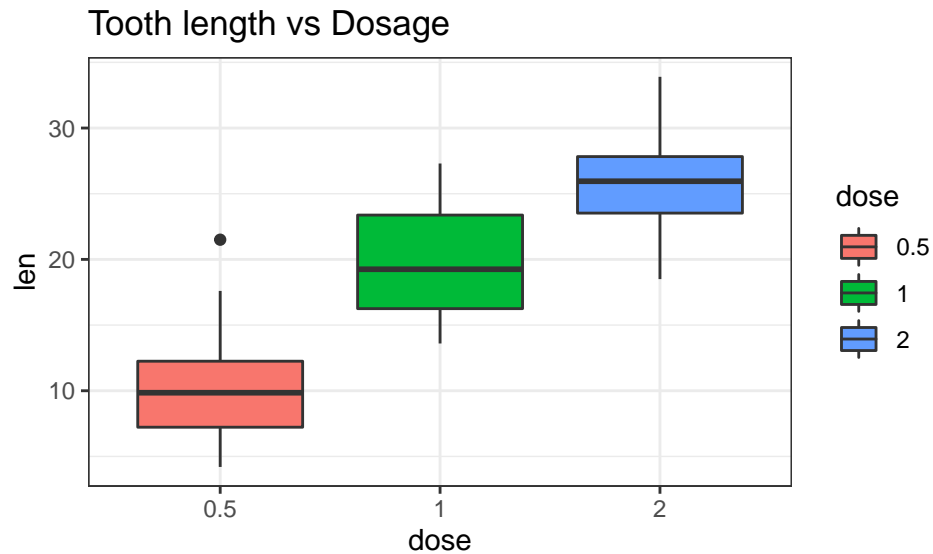
## Exploratory analysis using boxplots

**Mean change in tooth length vs supplement delivery mode**



There is considerable difference in the mean difference in tooth growth observed in the patients for the different delivery methods of supplement provided

**Difference in the tooth length for various dosages**



Tooth length vs Dosage

As dosage is increased the average tooth length increases at each step increase in dosage.

**Mean change in tooth length analyzed across each supplement and dosage provided**



Supplement~Dosage effect on Toothlength
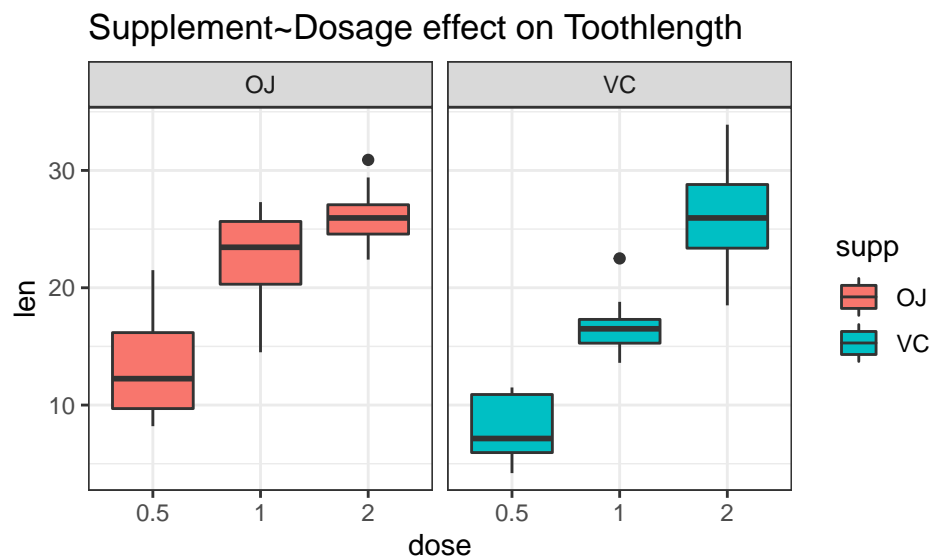
We observe that there is marked difference mean distribution of tooth length for patients that consumed 0.5 or 1 dosage of supplement OJ and VC, but the difference is almost non-existent for the 2mg dosage.

## Hypothesis testing

**Testing by supplement delivery mode**

Consider the supplement OJ being better performing than supplement VS

```
## Subsetting the data
x = ToothGrowth$len
group = ToothGrowth$supp
## Perfomring t-test
t.test(x~group, paired = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  x by group
## 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
```

The p value is greater than hence we reject the null hypothesis

Calculating the power with which we can reject the null hypothesis

```
n = length(x)
mu0 = mean(x[group=="OJ"])
mua = mean(x[group=="VC"])
sigma = sd(x[group=="VC"])
delta = (mua - mu0)/sigma
power.t.test(n = n, sd = sigma, delta = delta, type = "one.sample", alt = "one.sided")$power
```

```
## [1] 0.01972157
```

Our power is pretty low but regardless, the null hypothesis has been rejected but we can also bring out due
to the low power that there isn't much significant difference between the supplement delivery modes.



As we can see from the visualization that the two distributions are highly overlapping.

**Testing by dosage**

Another thing we noticed from the supplement~dosage visualization was that there is considerable difference in output for the various dosage levels 0.5, 1 and 2.
Consider the null hypothesis being dosage level has no effect on the tooth length.

Let us now check to see if we get a significant p-value for each pairing of the dosage.

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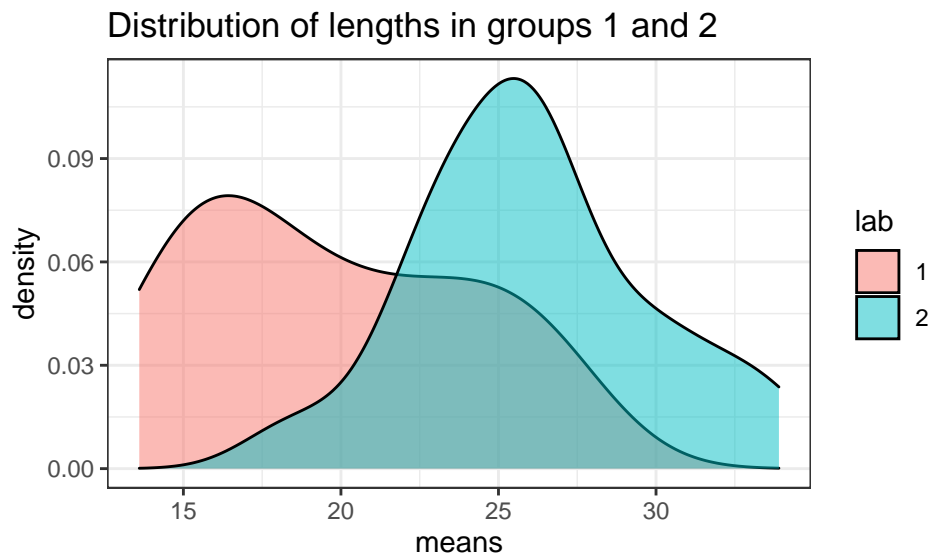
```
##
##  Welch Two Sample t-test
##
## data:  x by group
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##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



Distribution of lengths in groups 0.5 and 1
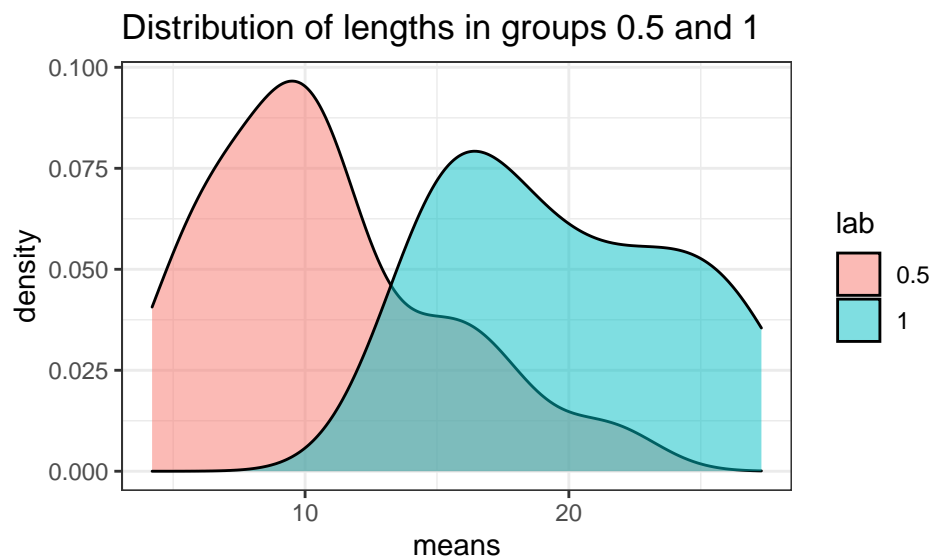
```
x = subset(ToothGrowth ,dose %in% c(1,2))$len
group = subset(ToothGrowth, dose %in% c(1,2))$dose
t.test(x~group, paired = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  x by group
## 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
```

## Distribution of lengths in groups 1 and 2



```
x = subset(ToothGrowth ,dose %in% c(0.5,2))$len
group = subset(ToothGrowth, dose %in% c(0.5,2))$dose
t.test(x~group, paired = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  x by group
## 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
```

## Distribution of lengths in groups 0.5 and 2
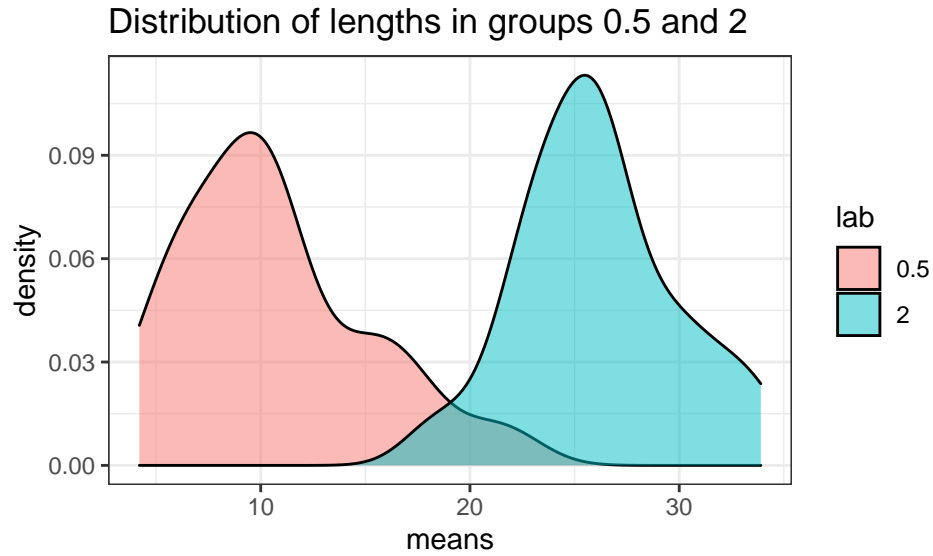


In each case the p-value is very low hence we can conclude that the hypothesis can be easily rejected.

## Conclusions

There appears to be statistical significance on tooth growth by varying the dosage levels, as dosage increases the mean tooth length increases.
And, There seems to be negligible role played by the delivery methods - VC, OJ although it is to be noted that at dosage levels 0.5 and 1 OJ has higher overall outcome in toothgrowth, whereas negligible for dosage of level 2