ToothGrowth Data

ToothGrowth data consists of three columns: len, supp, dose. len represents tooth length, supp represents the method of how vitamin C was delivered (ascorbic acid or orange juice) and dose represents the dose (obviously :)).

```
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 ...
```

```
head(ToothGrowth)
```

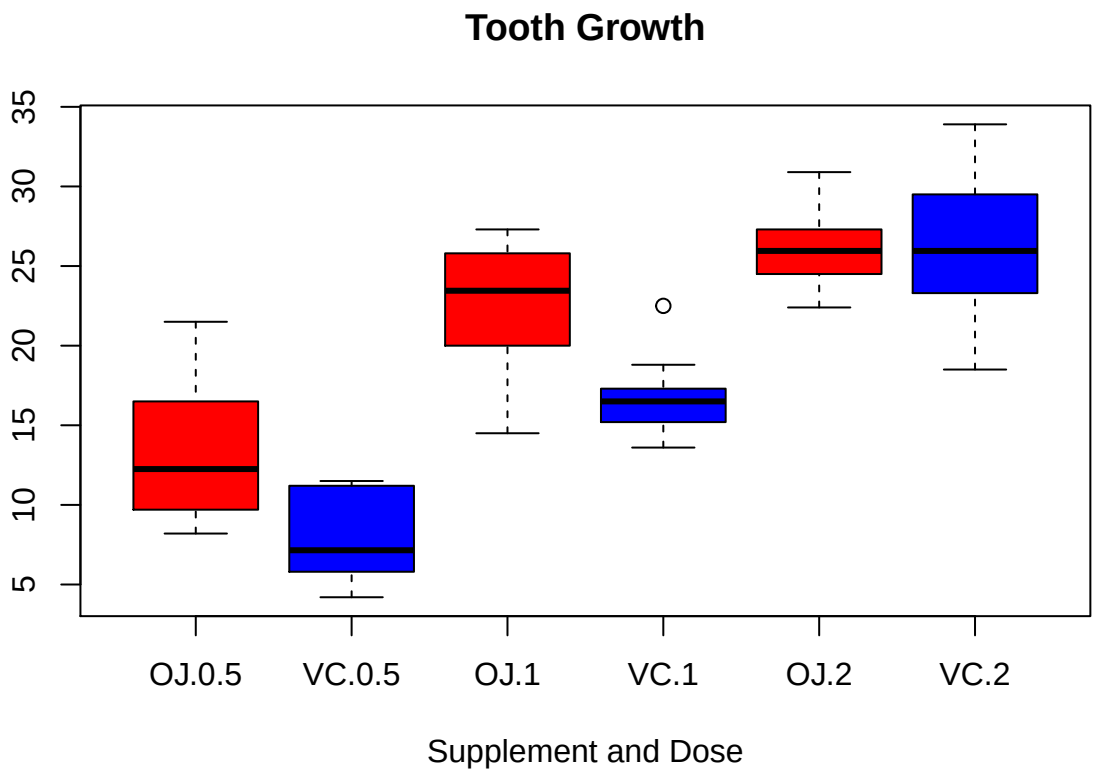
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
tail(ToothGrowth)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

Probably, the best way to explore raw data would be the boxplot:

```
boxplot(len~supp*dose, data=ToothGrowth, notch=FALSE,
  col=(c("red","blue")),
  main="Tooth Growth", xlab="Supplement and Dose")
```



As can be seen from the graph, there is some difference between delivery methods when the dose is either 0.5 mg or 1 mg and virtually no difference in case of 2 mg dose. In order to check, if there is any significant difference between two methods, I will use t-test. Prior to that, for convenience, I will add index column so that each index represents specific dose and method:

```
ToothGrowth$index <- rep(1:6, each = 10)
head(ToothGrowth, n = 12)
```

```
##     len supp dose index
## 1   4.2   VC  0.5     1
## 2  11.5   VC  0.5     1
## 3   7.3   VC  0.5     1
## 4   5.8   VC  0.5     1
## 5   6.4   VC  0.5     1
## 6  10.0   VC  0.5     1
## 7  11.2   VC  0.5     1
## 8  11.2   VC  0.5     1
## 9   5.2   VC  0.5     1
## 10  7.0   VC  0.5     1
## 11 16.5   VC  1.0     2
## 12 16.5   VC  1.0     2
```

Now I can refer to any combination of dose and supp using only index, for example, if I need entries with dose == 2 and supp == 'VC' I can do

```
ToothGrowth[ToothGrowth$index %in% 1, ]
```

```
##     len supp dose index
## 1   4.2   VC  0.5     1
## 2  11.5   VC  0.5     1
## 3   7.3   VC  0.5     1
## 4   5.8   VC  0.5     1
## 5   6.4   VC  0.5     1
## 6  10.0   VC  0.5     1
## 7  11.2   VC  0.5     1
## 8  11.2   VC  0.5     1
## 9   5.2   VC  0.5     1
## 10  7.0   VC  0.5     1
```

We can now apply t-test to test if means of each group are equal or not:

- for 0.5 mg

```
t.test(ToothGrowth[ToothGrowth$index %in% 4, ]$len - ToothGrowth[ToothGrowth$index %in% 1, ]$len)
```

```
##
##  One Sample t-test
##
## data:  ToothGrowth[ToothGrowth$index %in% 4, ]$len - ToothGrowth[ToothGrowth$index %in%     1, ]$len
## t = 2.9791, df = 9, p-value = 0.01547
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.263458 9.236542
## sample estimates:
## mean of x
##      5.25
```

If we compare t-statistic to the 95th percentile of the T distribution 1.8331129, it is obvious, that t statistic is too big, so we need to reject the null hypothesis in this case.

- for 1 mg

```
t.test(ToothGrowth[ToothGrowth$index %in% 5, ]$len - ToothGrowth[ToothGrowth$index %in% 2, ]$len)
```

```
##
##  One Sample t-test
##
## data:  ToothGrowth[ToothGrowth$index %in% 5, ]$len - ToothGrowth[ToothGrowth$index %in%     2, ]$len
## t = 3.3721, df = 9, p-value = 0.008229
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.951911 9.908089
## sample estimates:
## mean of x
##      5.93
```

Same happens with 1 mg dose, its t-statistic is too big.

- for 2 mg

```
t.test(ToothGrowth[ToothGrowth$index %in% 6, ]$len - ToothGrowth[ToothGrowth$index %in% 3, ]$len)
```

```
##
##  One Sample t-test
##
## data:  ToothGrowth[ToothGrowth$index %in% 6, ]$len - ToothGrowth[ToothGrowth$index %in%     3, ]$len
## t = -0.0426, df = 9, p-value = 0.967
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.328976  4.168976
## sample estimates:
## mean of x
##     -0.08
```

In case of 2mg doses there is no significant difference in the means of the samples.

We can also compare overall effectiveness of two delivery methods without accounting for the dose, but based on the plot I would say that orange juice is more effective way of administering vitamin C:

```
t.test(ToothGrowth[ToothGrowth$index %in% c(4, 5, 6), ]$len - ToothGrowth[ToothGrowth$index %in% c(1, 2
```

```
##
##  One Sample t-test
##
## data:  ToothGrowth[ToothGrowth$index %in% c(4, 5, 6), ]$len - ToothGrowth[ToothGrowth$index %in%
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.408659 5.991341
## sample estimates:
## mean of x
##       3.7
```

As expected, t-statistics is too big, so we need to reject null hypothesis and based on the confidence interval, which is completely above zero, it is obvious that orange juice is prefered delivery method. However when we talk about big doses (like 2 mg) there is no significant difference in both methods.