Exploring the tooth growth

Overview

In this report we will explore the response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

(We can get this info with the command ?ToothGrowth)

Loading data

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

Exploratory analysis

First, let's get some summary info.

```
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 ...
```

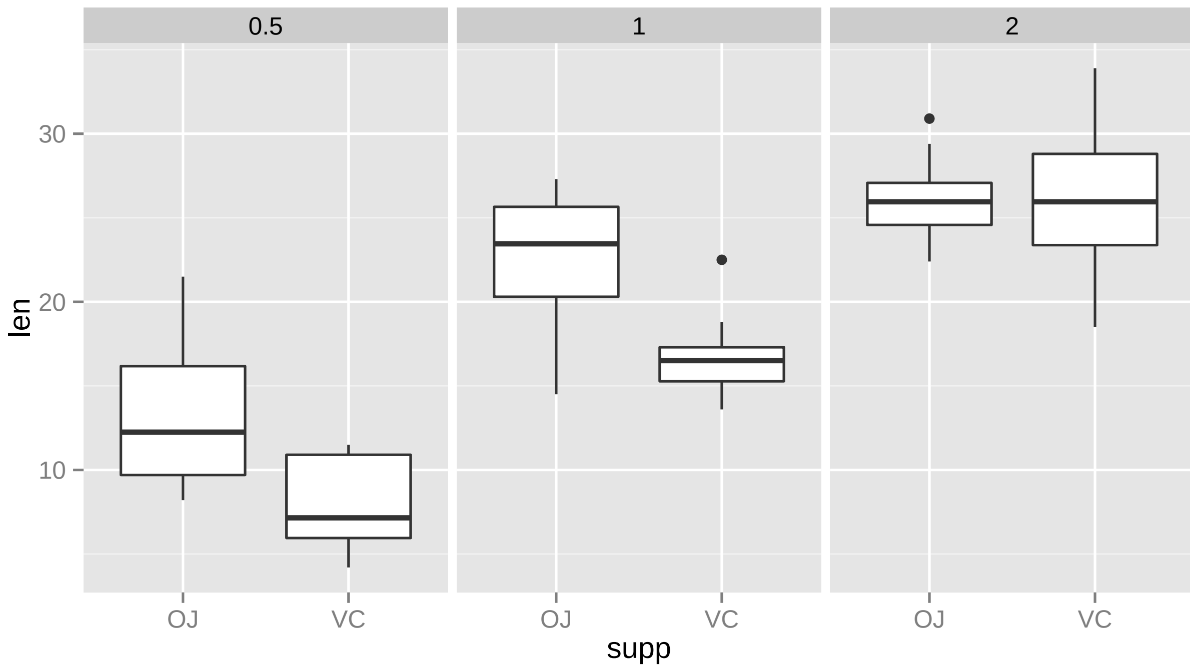
```
summary(ToothGrowth)
```

```
##       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  
```

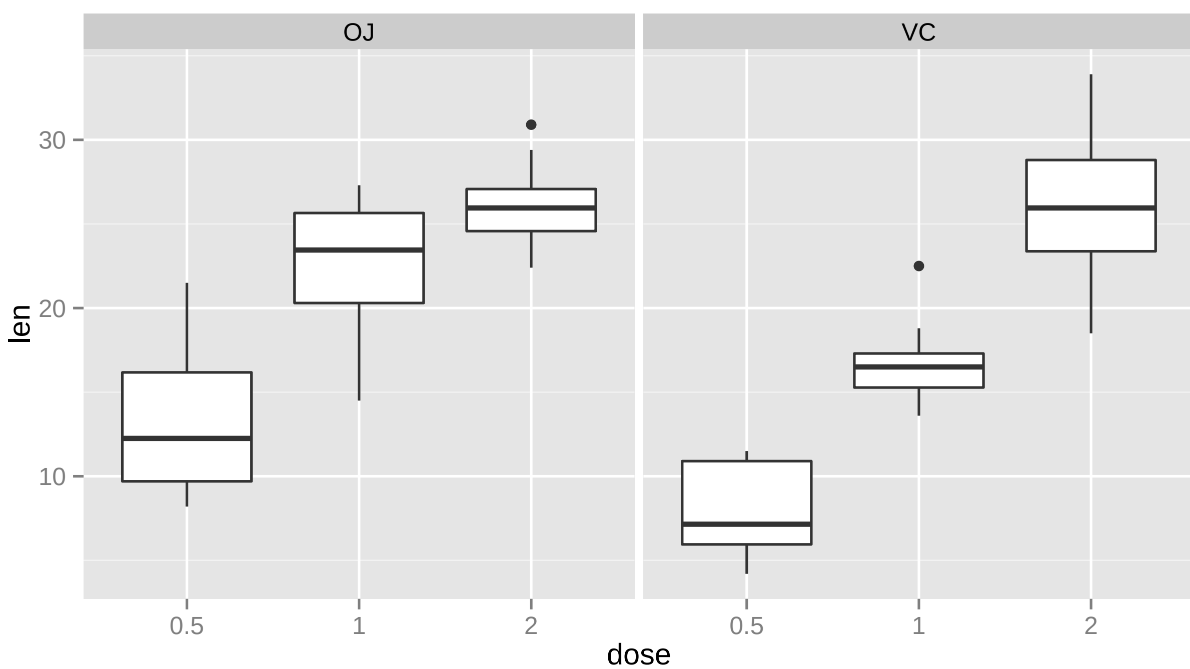
Now we build some conditional graphs, using ggplot2.

```
library(ggplot2)
ToothGrowth$dose<-factor(ToothGrowth$dose)

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



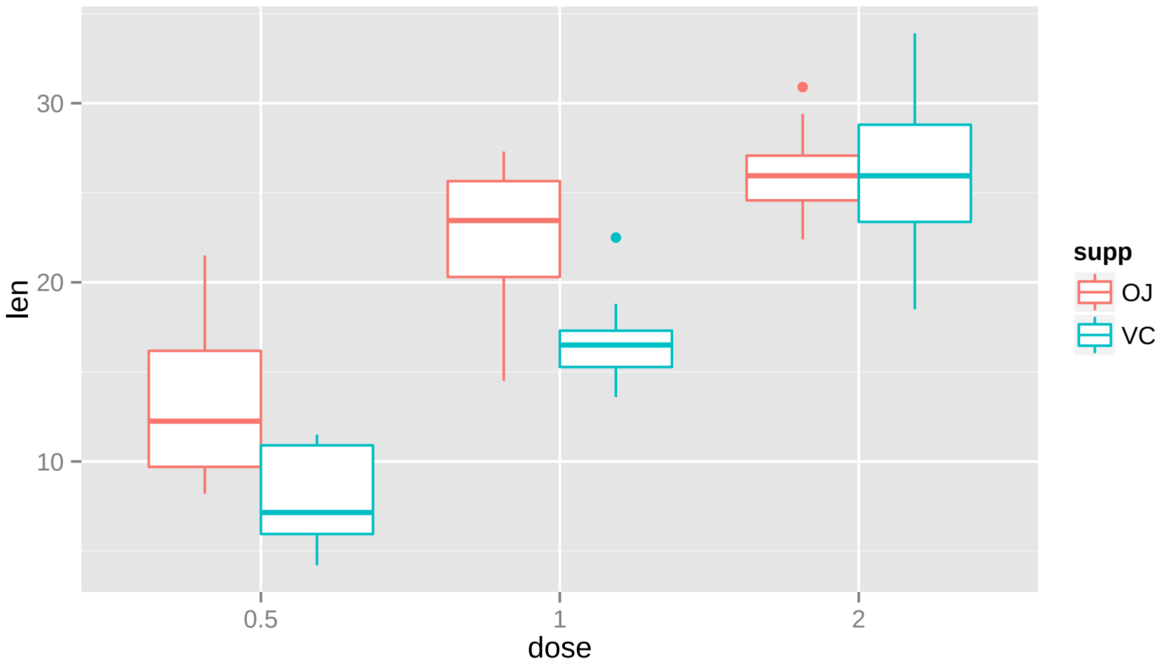

```
qplot(dose,len, data=ToothGrowth, facets=.~supp, geom="boxplot")
```



From this figures we see that increasing dose leads to greater tooth growth (which makes sense if vitamin works) and orange juice (OJ) seems to be more effective delivery method at least at two dose levels (0.5 and 1).

We can see the whole picture on single figure by combining these graphs with the help of colors.

```r
qplot(dose,len, data=ToothGrowth, col=supp, geom="boxplot")
```



Hypothesis testing

We'll use t-tests for our hypothesis, assuming that sample means have approximately normal distributions, so with small sample sizes t-distribution will be a good approximation. Since we don't know object's id for each record in dataset we can't perform dependent (paired) t-test, so we'll use independent one.

Considering our previous exploratory analysis we'll test two hypothesis:

1) Increasing dose increases tooth growth

MUo = difference in means is equal to 0

We have 3 types of dose, so we should do 3 t-tests.

Consider variances to be not equal.

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
t.test(filter(ToothGrowth, dose=="0.5")$len,filter(ToothGrowth, dose=="1")$len, paired=F, var.equal=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  filter(ToothGrowth, dose == "0.5")$len and filter(ToothGrowth, dose == "1")$len
## 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 of x mean of y
##    10.605    19.735
```

p-value is very small, so difference is significant at 0.05 level and even at 0.01. Also we see that confidence interval doesn't contain 0, but is far below it.

We can make the same inference for other two tests.

```r
t.test(filter(ToothGrowth, dose=="1")$len,filter(ToothGrowth, dose=="2")$len, paired=F, var.equal=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  filter(ToothGrowth, dose == "1")$len and filter(ToothGrowth, dose == "2")$len
## 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 of x mean of y
##    19.735    26.100
```

```r
t.test(filter(ToothGrowth, dose=="0.5")$len,filter(ToothGrowth, dose=="2")$len, paired=F, var.equal=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  filter(ToothGrowth, dose == "0.5")$len and filter(ToothGrowth, dose == "2")$len
## 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 of x mean of y
##    10.605    26.100
```

2) Delivery method influences the tooth growth

MUo = difference in means is equal to 0.

Consider variances to be not equal.

```r
t.test(filter(ToothGrowth, supp=="VC")$len,filter(ToothGrowth, supp=="OJ")$len, paired=F, var.equal=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  filter(ToothGrowth, supp == "VC")$len and filter(ToothGrowth, supp == "OJ")$len
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

p-value>0.05, so difference is not significant at 0.05 level. Looking at confidence interval, we see that it contains 0.

But considering our exploratory analysis above we may want to exclude records with dose=2 and test differences for other doses.

```r
t.test(filter(ToothGrowth, supp=="VC" & dose !=2)$len,filter(ToothGrowth, supp=="OJ"& dose!=2)$len, pai
```

```
##
##  Welch Two Sample t-test
##
## data:  filter(ToothGrowth, supp == "VC" & dose != 2)$len and filter(ToothGrowth, supp == "OJ" & dose
## t = -3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.304766 -1.875234
## sample estimates:
## mean of x mean of y
##    12.375    17.965
```

Now we see significance at 0.05 and 0.01 levels, and confidence interval doesn't contain 0.

Summary

We assume that sample means have approximately normal distributions, so with small sample sizes t-distribution will be a good approximation.

Having done independent t-tests, we can conclude that increasing dose increases tooth growth (at 0.01 significance level) and orange juice is more effective delivery method for tooth growth for doses=0.5 or 1 (at 0.01 significance level).