Analyze the ToothGrowth data

Tooth Growth data analysis

In this report we will look at the ToothGrowth dataset in R and perform some basic exploratory data analyses. Then we will use hypothesis test to compare tooth growth by different supplement type and dose.

Load data and exploratory data analyses

First we load the dataset from the library. By looking at the basic information of the data we know there are 60 observations of 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).

```
library(datasets)
library(ggplot2)
data(ToothGrowth)
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)
```

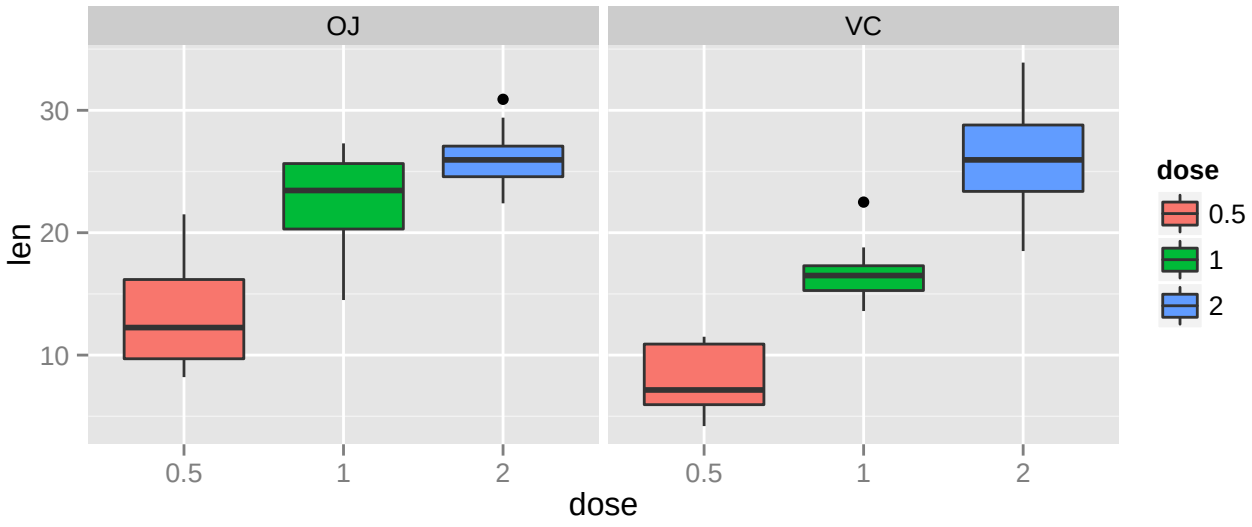
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
str(ToothGrowth)
```

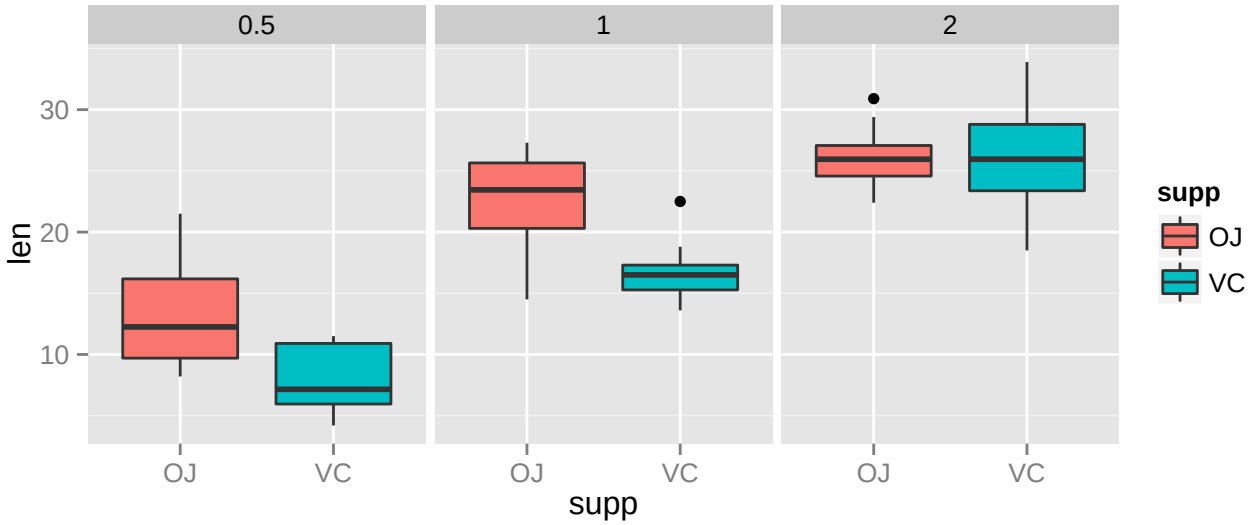
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Then we plot the length of tooth with different dosage level and supplement type.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
ggplot(aes(x=dose,y=len),data=ToothGrowth) + geom_boxplot(aes(fill=dose)) + facet_wrap(~supp)
```




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



From above analysis, we see the supplment type appears to affect the tooth length: orange juice(OJ) is better than VC. But whether the difference is statistically significant remains to be investigated. Also we observe the supplement of Vitamin C has positive influence on the tooth length: the higher the supplement, the longer the tooth length. In the following, we will investigate whether this difference is of statistical importance.

Confidence intervals and Hypothesis Test

Now we will use hypothesis test to find out whether different dosage and supplement type have statistical significance on the tooth lengh. First we want to find whether the supplement type has influence on the tooth length at each dosage level. The null hypothesis is the difference in means of the tooth length with different supplement type is 0(for dosage=0.5, 1 and 2).

```
t.test(len ~ supp, paired = FALSE, var.equal= FALSE, data=ToothGrowth[ToothGrowth$dose==0.5,])
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

With dosage level at 0.5, the p-value of the hypothesis is 0.006, which is less than 0.05. Then we will reject the null hypothesis. Also we can see the confidence interval is [1.719057 8.780943], the fact that it does not contain 0 gives the same conclusion.

```
t.test(len ~ supp, paired = FALSE, var.equal= FALSE, data=ToothGrowth[ToothGrowth$dose==1.0,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

With dosage level at 1.0, the p-value of the hypothesis is 0.001, which is less than 0.05. Again we will reject the null hypothesis. Also we can see the confidence interval is [2.802148 9.057852], the fact that it does not contain 0 gives the same conclusion.

```
t.test(len ~ supp, paired = FALSE, var.equal= FALSE, data=ToothGrowth[ToothGrowth$dose==2.0,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

With dosage level at 2.0, the p-value of the hypothesis is 0.9639, which is greater than 0.05. Then we fail to reject the null hypothesis. Also we can see the confidence interval is [-3.79807 3.63807], the fact that it contains 0 gives the same conclusion.

Next we will investigate wether the dosage level has significant influence on the tooth length with each supplement type. The null hypothesis is the means of the tooth length with different dosage levels is 0 for each supplement type.

```
t.test(len ~ dose, paired = FALSE, var.equal= FALSE,
       data=ToothGrowth[ToothGrowth$supp=="OJ"&
                        (ToothGrowth$dose==0.5|ToothGrowth$dose==1.0),])$p.value
```

```
## [1] 8.784919e-05
```

```
t.test(len ~ dose, paired = FALSE, var.equal= FALSE,
       data=ToothGrowth[ToothGrowth$supp=="OJ"&
                        (ToothGrowth$dose==1.0|ToothGrowth$dose==2.0),])$p.value
```

```
## [1] 0.03919514
```

```
t.test(len ~ dose, paired = FALSE, var.equal= FALSE,
       data=ToothGrowth[ToothGrowth$supp=="OJ"&
                        (ToothGrowth$dose==0.5|ToothGrowth$dose==2.0),])$p.value
```

```
## [1] 1.323784e-06
```

We see all p-values are less than 0.05, thus we will reject the null hypothesis. We say there are significant difference with different dosage level when suppled with OJ.

```
t.test(len ~ dose, paired = FALSE, var.equal= FALSE,
       data=ToothGrowth[ToothGrowth$supp =="VC"&
                        (ToothGrowth$dose==0.5|ToothGrowth$dose==1.0),])$p.value
```

```
## [1] 6.811018e-07
```

```
t.test(len ~ dose, paired = FALSE, var.equal= FALSE,
       data=ToothGrowth[ToothGrowth$supp =="VC"&
                        (ToothGrowth$dose==0.5|ToothGrowth$dose==2.0),])$p.value
```

```
## [1] 4.681577e-08
```

```
t.test(len ~ dose, paired = FALSE, var.equal= FALSE,
       data=ToothGrowth[ToothGrowth$supp =="VC"&
                        (ToothGrowth$dose==1.0|ToothGrowth$dose==2.0),])$p.value
```

```
## [1] 9.155603e-05
```

We see all p-values are less than 0.05, thus we will reject the null hypothesis. We say there are significant difference with different dosage level when suppled with VC.

Conclusion and Assumptions

In general, we find there are significant difference in tooth length with different supplement type and dosage level. But when the dosage is large(dose=2mg), there is no difference between the two supplement types. For above analysis, we assume each observation is independent. We also assume the guinea pigs are randomly selected and their tooth conditions are similar(no other attributes of selected guinea pigs such as age or size would affect the tooth growth).