Statistical Inference Project. Part 2.

Sergey Rakhmatullin

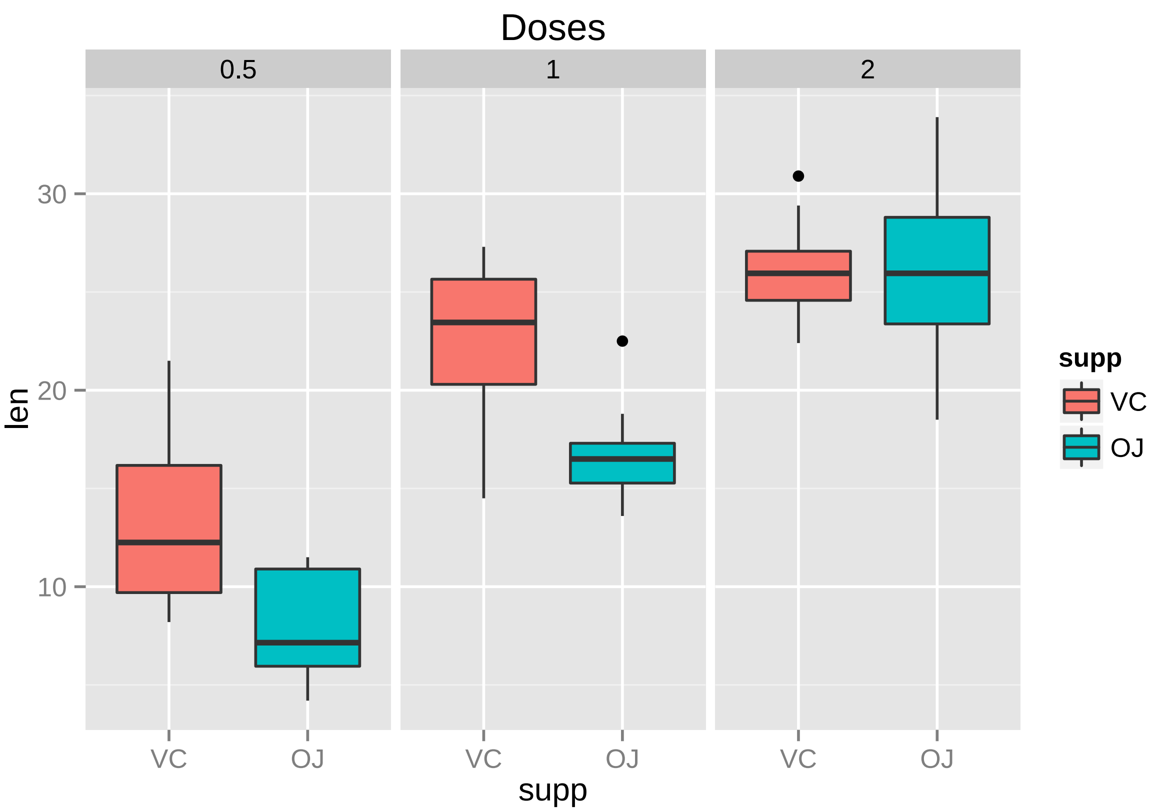
Overview

This project studies the ToothGrowth dataset and estimates effects of different supplements(OJ, VC) on tooth growth, data is given for three dose volumes: 0.5, 1.0, 2.0 ml.

Exploratory Data Analysis

Let's see how dose and supplement affect the tooth growth.

```
library(ggplot2)
require(datasets)
levels(ToothGrowth$supp) <-c('VC','OJ')
levels(ToothGrowth$dose) <-c(0.5, 1.0, 2.0)
d1<-ggplot(ToothGrowth, aes(y=len, x=supp, group=supp))
d1<-d1+geom_boxplot(aes(fill=supp))+facet_grid(.~dose)
d1 + ggtitle('Doses')
```



We can see 6 groups divided by two supplement factors (OJ, VC) and three dose factors (0.5, 1.0, 2.0).
For doses 0.5 and 1.0 we can clearly observe difference of results length caused by supplement (VC gives better results).
For dose 2.0 we can't see difference of results length caused by supplement.

Calculating 95% T-confidence intervals

Calculating 95% T-confidence interval for each of 6 groups:

```
n<-10
tCI<-aggregate(len~dose+supp, data=ToothGrowth,
               FUN=function(x){mean(x)+c(-1,1)*qt(0.975,n-1)*sd(x)/sqrt(n)})
colnames(tCI[,3])<-c('tCI_low', 'tCI_high')
tCI[tCI$dose==0.5,]
```

```
##   dose supp len.tCI_low len.tCI_high
## 1  0.5   VC   10.039717    16.420283
## 4  0.5   OJ    6.015176     9.944824
```

```
tCI[tCI$dose==1.0,]
```

```
##   dose supp len.tCI_low len.tCI_high
## 2    1   VC    19.90227     25.49773
## 5    1   OJ    14.97066     18.56934
```

```
tCI[tCI$dose==2.0,]
```

```
##   dose supp len.tCI_low len.tCI_high
## 3    2   VC    24.16069     27.95931
## 6    2   OJ    22.70791     29.57209
```

For doses 0.5, 1.0 95% T-confidence intervals have no intersection: 95% T-conf interval for VC is above 95% T-conf interval for OJ.
For dose 2.0 95% T-confidence intervals have significant overlay and means are pretty close - we can't say which supplement has better results.

Hypothesys testing

Let's estimate same data using hypothesys testing.
Testing the null hypothesys that means for different supplements (VC, OJ) for each dose (0.5, 1.0, 2.0) are equal, i.e. $\mu_{vc,0.5} = \mu_{oj,0.5}$, $\mu_{vc,1.0} = \mu_{oj,1.0}$, $\mu_{vc,2.0} = \mu_{oj,2.0}$:

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

```
##
##  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 VC mean in group OJ
##            13.23             7.98
```

```r
t.test(len~supp, data=ToothGrowth, subset=(dose==1.0), paired=FALSE, var.equal=FALSE)
```

```
##
##  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 VC mean in group OJ
##            22.70            16.77
```

For doses 0.5 and 1.0 difference for supplements (VC, OJ) has positive value with 95% T-confidence interval above zero. The null hypothesis is not false with probability more than 95%, so we can accept alternative hypothesis that results for VC is better that the results for OJ.
$\mu_{vc,0.5} > \mu_{oj,0.5}$, $\mu_{vc,1.0} > \mu_{oj,1.0}$

```r
t.test(len~supp, data=ToothGrowth, subset=(dose==2.0), paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, 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 VC mean in group OJ
##            26.06            26.14
```

For dose 2.0 difference using supplements (VC, OJ) has close to zero value with 95% T-confidence interval that includes zero. The null hypothesis is prove to be correct on the given data.
$\mu_{vc,2.0} = \mu_{oj,2.0}$

Conclusions

For our conclusions we assumed that we can clearly estimate effectiveness of supplements depending on the dose according to its 95% T-confidence intervals of resulting growth: if the mean growth of one supplement is larger than other and 95% T-confidence intervals have no intersection - this supplement is better for our experiment.
Or, in other words, if one supplement causes growth difference than other with 95% T-conf interval above zero - we are to conclude that it gives better result than other.

Both T-confidence interval estimations and hypothesis testing illustrates that:
for doses 0.5 and 1.0 using VC supplement gives better results than OJ
there is no difference between using supplements (VC, OJ) for dose 2.0

So we can conclude that for the small amount of supplement more efficient to use VC supplement.