Basic Inferential Data Analysis

Mariia Danilenko

Introduction

We're going to analyze the ToothGrowth data in the R datasets package. For this project we only use the techniques from class, even if there's other approaches worth considering.

Analysis

Given data

ToothGrowth data:

The response is the length (**len** column) of odontoblasts (cells responsible for tooth growth) in **60** guinea pigs. Each animal received one of three dose (**dose** column) levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods: orange juice (OJ) or ascorbic acid (VC) (**supp** column). Take a look to observing data:

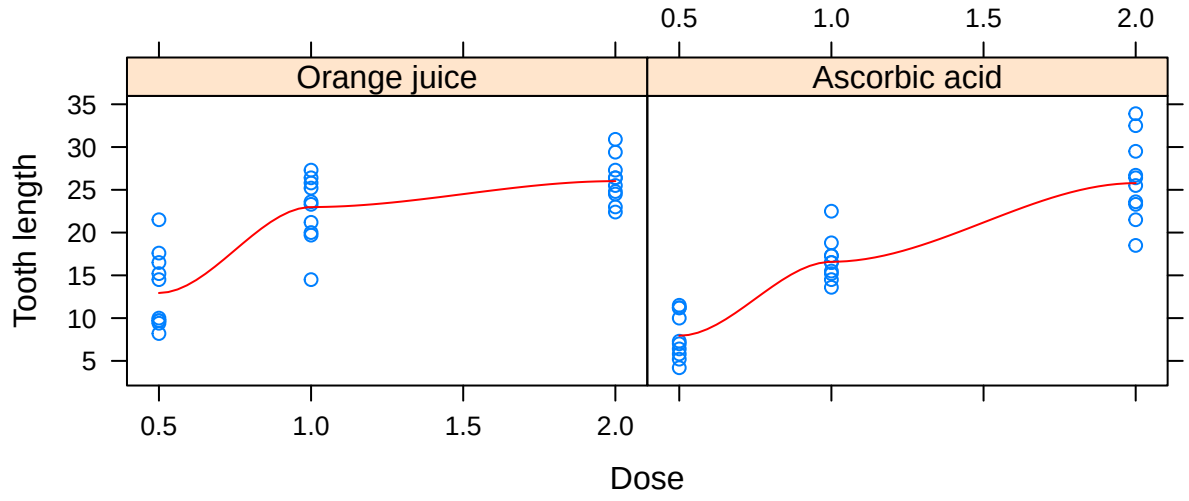
```
str(ToothGrowth)
```

```
## '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 ...
```

Visual analysis

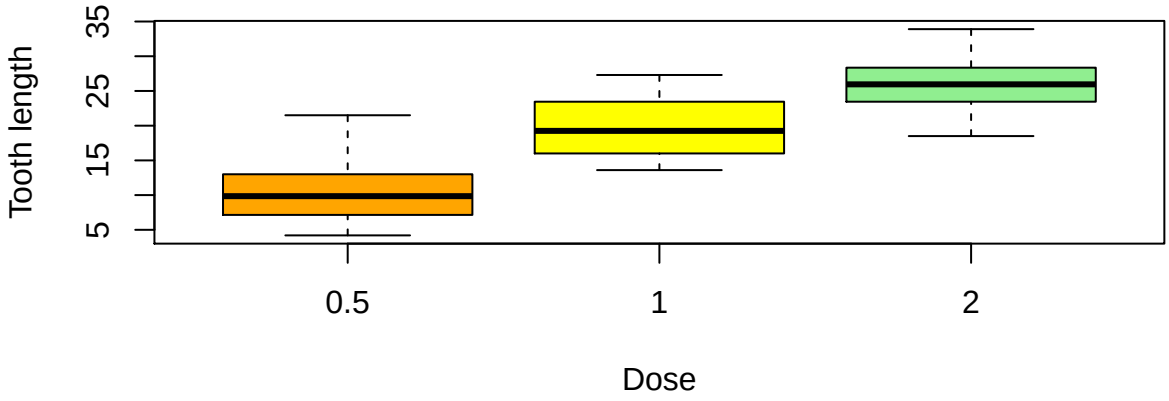
Make an introduction plot of length factored by supplement type and depends on dose:

```
library(lattice)
par(mfcol=c(2,1))
xyplot(len ~ dose | factor(supp,labels=c("Orange juice","Ascorbic acid")),
       data = ToothGrowth, type=c("p","smooth"),
       xlab = "Dose", col.line="red",
       ylab="Tooth length")
```



Notice correlation between dose and length. Before we begin analysis, let's plot boxplot factored by dose:

```
with(ToothGrowth,
     boxplot(len~dose, xlab="Dose", ylab="Tooth length",
             col=c("orange","yellow", "lightgreen")))
```



Calculations

Cut given data into six groups: 2 sup types and 3 dose types:

```
TGOJ0.5<-subset(ToothGrowth, ToothGrowth$supp=="OJ" & ToothGrowth$dose==0.5)
TGVC0.5<-subset(ToothGrowth, ToothGrowth$supp=="VC" & ToothGrowth$dose==0.5)
TGVC1<-subset(ToothGrowth, ToothGrowth$supp=="VC" & ToothGrowth$dose==1)
TGOJ1<-subset(ToothGrowth, ToothGrowth$supp=="OJ" & ToothGrowth$dose==1)
TGOJ2<-subset(ToothGrowth, ToothGrowth$supp=="OJ" & ToothGrowth$dose==2)
TGVC2<-subset(ToothGrowth, ToothGrowth$supp=="VC" & ToothGrowth$dose==2)
```

And also select two subsets with orange juice and ascorbic acid:

```
TGOJ<-subset(ToothGrowth, ToothGrowth$supp=="OJ")
TGVC<-subset(ToothGrowth, ToothGrowth$supp=="VC")
```

Comparisons of doses

Comparison between 0.5 and 1 doses

```
t.test(cbind(TGOJ0.5$len,TGVC0.5$len), cbind(TGOJ1$len,TGVC1$len),
       alternative = "less")$conf.int
```

```
## [1]      -Inf -6.753323
## attr(,"conf.level")
## [1] 0.95
```

We reject H_0 hypothesis and claim that length mean for 0.5 dose is less than length mean for 1 dose. Comparison between 1 and 2 doses

```
t.test(cbind(TGOJ1$len,TGVC1$len), cbind(TGOJ2$len,TGVC2$len),
       alternative = "less")$conf.int
```

```
## [1]     -Inf -4.17387
## attr(,"conf.level")
## [1] 0.95
```

We reject H_0 hypothesis and claim that length mean for 1 dose is less than length mean for 2 dose.

We don't need compare 0.5 and 2 doses because of transitivity.

Also, we checked same hypothesis for different sups separately and got same results. R code for these calculations you can find in the appendix. Finally we claim that dose changes are significant for tooth growing.

Comparisons of sups

We compared two sup types for every dose:

```
t.test(TGOJ0.5$len,TGVC0.5$len, alternative = "greater")$conf.int
```

```
## [1] 2.34604     Inf
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(TGOJ1$len,TGVC1$len, alternative = "greater")$conf.int
```

```
## [1] 3.356158      Inf
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(TGOJ2$len,TGVC2$len, alternative = "greater")$conf.int
```

```
## [1] -3.1335     Inf
## attr(,"conf.level")
## [1] 0.95
```

We reject H_0 hypothesis in first and second models. It means that orange juice is more effective than ascorbic acid on 0.5 and 1 doses.

We fail to reject H_0 hypothesis in fhird model, however we can explain it:
Exist maximum level of vitamin C. which pigs can absorb. That's why in 2 dose sup type doesn't matter.

Results

We made following conclusions:
* dose growing is significant for tooth growing;
* sup type impact on tooth length until 2 dose.

Appendix

Doses hypothesis separately for sups:

```
# for orange juice
t.test(TGOJ0.5$len,TGOJ1$len, alternative = "less")$conf.int
```

```
## [1]      -Inf -6.214316
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(TGOJ1$len,TGOJ2$len, alternative = "less")$conf.int
```

```
## [1]       -Inf -0.7486236
## attr(,"conf.level")
## [1] 0.95
```

```
# for ascorbic acid
t.test(TGVC0.5$len,TGVC1$len, alternative = "less")$conf.int
```

```
## [1]      -Inf -6.746867
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(TGVC1$len,TGVC2$len, alternative = "less")$conf.int
```

```
## [1]      -Inf -6.346525
## attr(,"conf.level")
## [1] 0.95
```

We reject H_0 hypothesis for all 4 models.

Sup whole hypothesis:

```
t.test(TGOJ$len,TGVC$len, alternative = "greater")$conf.int
```

```
## [1] 0.4682687       Inf
## attr(,"conf.level")
## [1] 0.95
```

We reject H_0 hypothesis.