Statistical interference project2

MS

January 14, 2016

ToothGrowth analysis

Overview

This document performs a simple statistic analysis of the "toothgrowth" data package. The Data is first analysed by comparing two observables at a time and combined to an analysis of all three observables.

Data description

Taken from R Help 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).

Analysis

Data basics

Load the data and show a brief description of the contents.

```
#Load the data set
data("ToothGrowth")
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 ...
```

```
#Seperate out the columns for faster analysis
supp<-ToothGrowth$supp
len<-ToothGrowth$len
dose<-ToothGrowth$dose
```

The data has 60 observations with 3 variables: "len", "supp" and "dose". "len" is the length of the tooth, "supp" is the supplementory type as a factor (Orange Juice or Ascorbic Acid), and "dose" is the dosage in milligrams/day.

Lets have a close look at the data

```
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
```

The tooth length varies from 4.2 to 33.9 with a mean of 18.8. The observations are equally distributed among OJ and VC. Dosage ranges from 0.5 to 2 mg vitmin C, however we can't see how many are in each group.

```
dfDose<-as.data.frame(table(dose))
dfDose
```

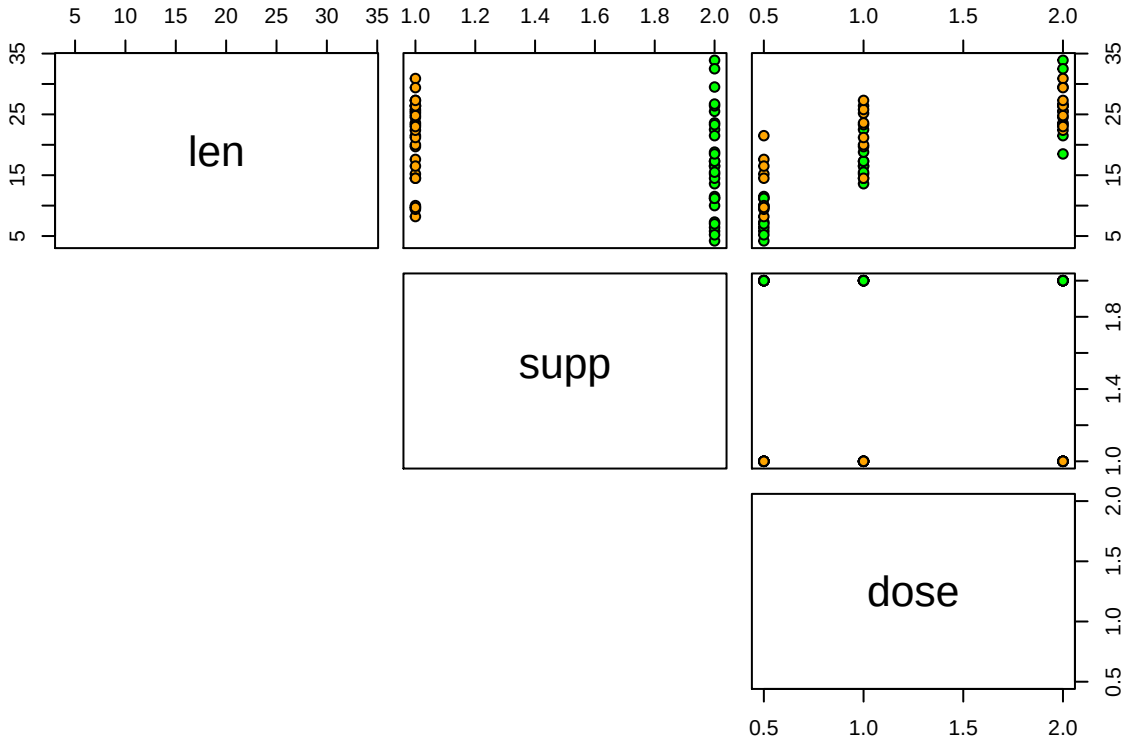
```
##   dose Freq
## 1  0.5   20
## 2    1   20
## 3    2   20
```

Now we know that there are three different dosages: 0.5, 1 and 2 mg. Each with 20 observations.

Exploring the dataset using two observations

First we plot all the data using the pairs function with OJ in orange and VC in green.

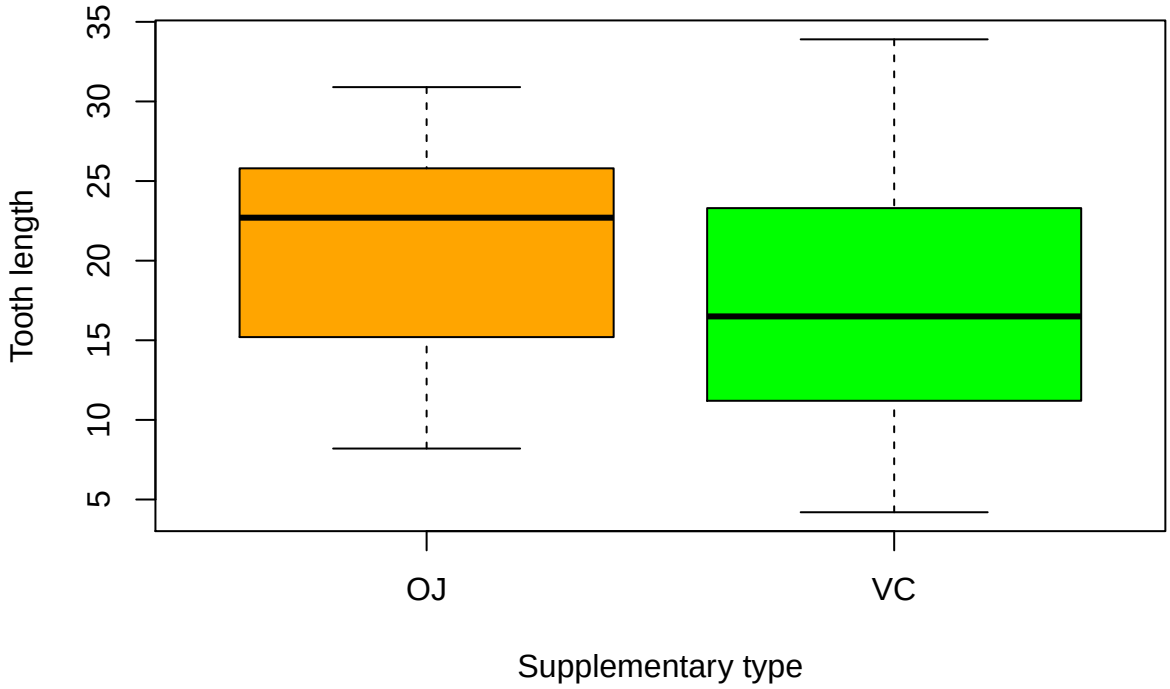
```
##Call pairs and color according to supp
pairs(ToothGrowth, pch = 21, bg = c("orange","green")[supp], lower.panel=NULL)
```



The plotting of “dose” vs. “supp” does not tell us anything we dont already know. In “supp” vs “len” it seems that OJ has a smaller variance than VC but the averages are not clear. “dose” vs. “len” show a general positive correlation between the dose and the tooth length.

Lets start by exploring the relationship between “supp” and “len” without looking at the dose.

```
#Plot supp vs len
boxplot(len ~ supp, type = "p", xlab ="Supplementary type", ylab = "Tooth length", col = c("orange","gr
```



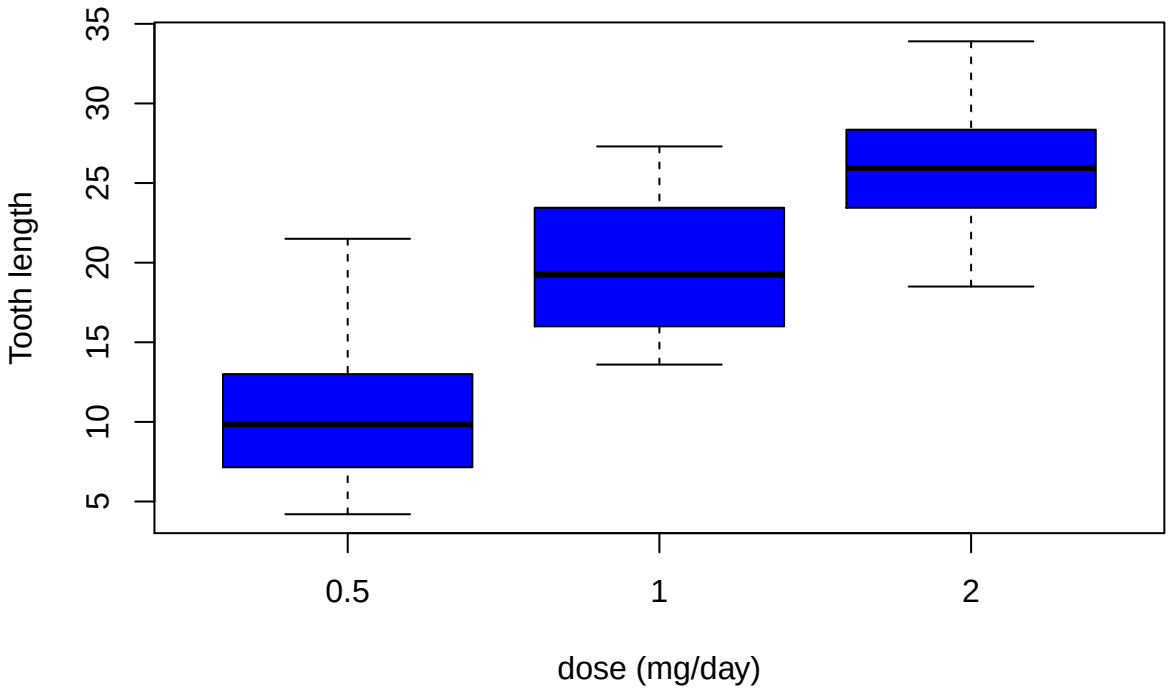
The mean of VC is lower than OJ. But is it significant? The p-value and 95% confidence interval is calculated for a two-sided T-distribution with unequal variables with the null hypothesis that the means are equal.

```
#Calculating the p value and confidence interval for supp vs len
lensuppP <- t.test(len ~ supp, alt ="two.sided", paired = F, var.equal = F)$p.value
lensuppC <- t.test(len ~ supp, alt ="two.sided", paired = F, var.equal = F)$conf
```

The p value is 0.0606345 and the confidence interval is [-0.1710156, 7.5710156]. As p value > 0.05 we fail to reject the null hypothesis (that the means are equal).

Lets have a look at the "dose" vs "len"":

```
#Plot dose vs len
boxplot(len ~ dose, xlab ="dose (mg/day)", ylab = "Tooth length", col = c("blue"))
```



The tooth length means seem to be positive correlated with the dosage. Lets have a look at the p value using the same parameters as above for all three posibilities

```
tResults <- NULL
for (i in 1:2) {
    for (j in (i+1):3) {
        doseIndex <- c(i,j)
        dosesTest <- dose %in% unique(dose)[doseIndex]
        pval <- t.test(len[dosesTest] ~ dose[dosesTest], alt ="two.sided", paired = F, var.equal = F)$p
        conf <- t.test(len[dosesTest] ~ dose[dosesTest], alt ="two.sided", paired = F, var.equal = F)$c
        tResults<-as.data.frame(rbind(tResults,cbind(pval,conf[1],conf[2])))
    }
}
rownames(tResults) <- c("0.5 & 1.0|","0.5 & 2.0|","1.0 & 2.0|")
colnames(tResults)<-c("p.value","conf.low","conf.high")
tResults
```

```
##                  p.value   conf.low  conf.high
## 0.5 & 1.0| 1.268301e-07 -11.983781  -6.276219
## 0.5 & 2.0| 4.397525e-14 -18.156167 -12.833833
## 1.0 & 2.0| 1.906430e-05  -8.996481  -3.733519
```

All p values are <0.05 and we reject the null hypothesis for all comparisons. Thus we conclude, that the mean tooth length increases when the dosage is increased from 0 to 0.5 mg/day and increased from 0.5 to 2.0 mg/day.

Does this conclusion still valid when we adjust for multitesting using the Bonferroni method?

```r
p.adjust(tResults$p.value, method = "BH")
```

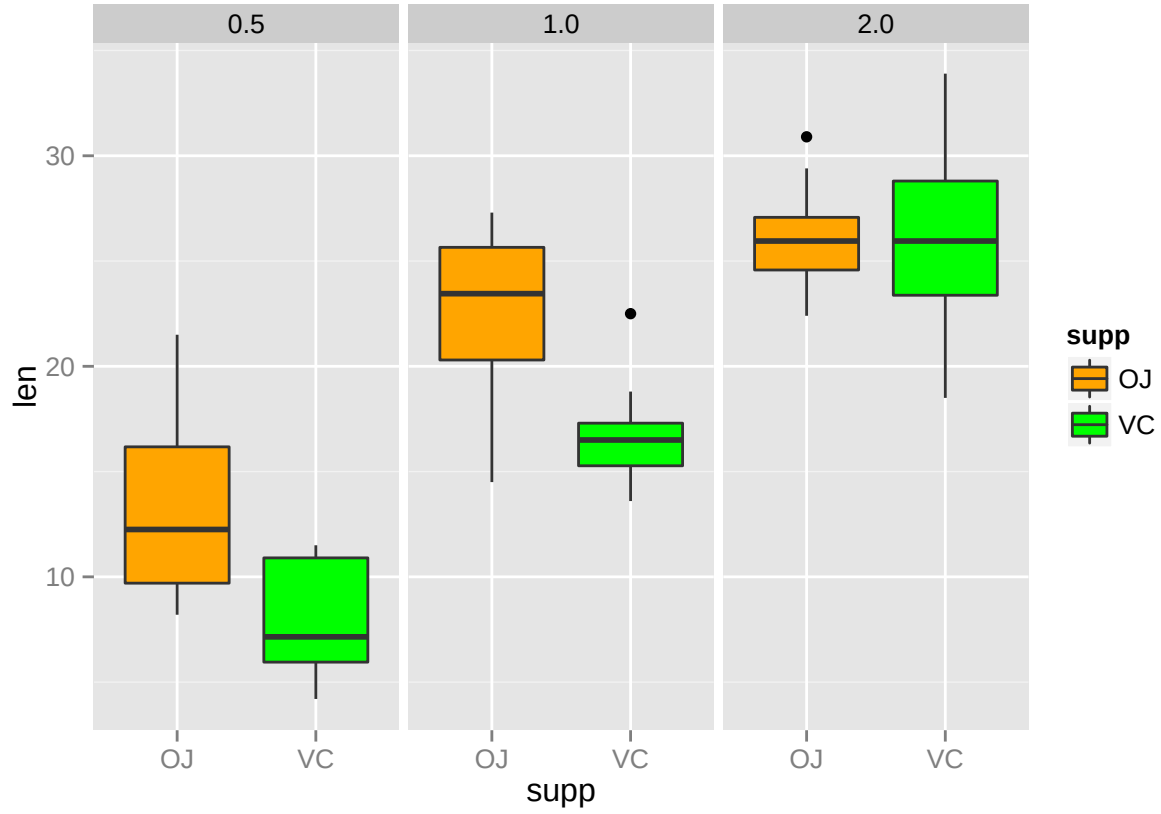
```
## [1] 1.902451e-07 1.319257e-13 1.906430e-05
```

Yes they do.

Exploring the dataset using two observations

Lets plot the length according to dose and supplementary.

```r
library(ggplot2)
ggplot(aes(x = supp, y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = supp)) + facet_wrap(~ dose) +
    scale_fill_manual(values=c("Orange","Green"))
```



The mean seems different for the delivery method at concentration 0.5 and 1.0 mg/day, but equal for 2.0 mg/day. Lets calculate the p values and confidence intervals as above.

```r
tResults <- NULL

for (i in (1:3)) {
    dosesTest <- dose %in% unique(dose)[i]
    pval <- t.test(len[dosesTest] ~ supp[dosesTest], alt ="two.sided", paired = F, var.equal = F)$p.valu
    conf <- t.test(len[dosesTest] ~ supp[dosesTest], alt ="two.sided", paired = F, var.equal = F)$conf
```

```
    tResults<-as.data.frame(rbind(tResults,cbind(pval,conf[1],conf[2])))
}
colnames(tResults)<-c("p.value","conf.low","conf.high")
rownames(tResults) <- unique(dose)
tResults
```

```
##         p.value  conf.low conf.high
## 0.5 0.006358607  1.719057  8.780943
## 1   0.001038376  2.802148  9.057852
## 2   0.963851589 -3.798070  3.638070
```

For both the 0.5 and 1.0 mg/day concentration the p-value>0.05 and hence we reject the null hypothesis and conclude: For a concentration of 0.5 and 1.0 mg/day the effect on tooth length is greater for orange juice delivery than for Ascorbic acid. At the highest concentration we fail to reject the null hypoethesis and accept it. Hence, for a concentration of 2.0 mg/day the effect on tooth length is not different for the two delivery methods.

Does this conclusion still valid when we adjust for multitesting using the Bonferroni method?

```
p.adjust(tResults$p.value, method = "BH")
```

```
## [1] 0.009537910 0.003115128 0.963851589
```

Yes they do.

Conclusions

1. The mean tooth length increases when the dosage is increased from 0 to 0.5 mg/day and increased from 0.5 to 2.0 mg/day.
2. For a concentration of 0.5 and 1.0 mg/day the effect on tooth length is greater for orange juice delivery than for Ascorbic acid.

This leads to other potential working hypothesis' that would require more data:

There is a maximum effect of Vitamin C on tooth length that is independent of delivery method

```
In our current data this is achieved at 2.0 mg/mL for both delivery methods and is supported by the fac
```

The effect of Vitamin C on tooth length depends on the absorption by the gut and hence explains the differences observed for the delivery methods.

```
Needs a completely different data set of Vitamin C absorption by delivery method and should then be cor
```

Assumptions

The following assumptions have been made:

1. The poplulations are independent
2. Divided into groups randomly
3. The delivery method does not have other effects on tooth length then Vitamin C (i.e. other ingredients in Orange Juice, which would require a control group of Guinea Pigs receiving Orange Juice without Vitamin C)