P2

Wang Luyuan

2019/8/24

Overview:

This report analyzes the `ToothGrowth` data in the R data sets package.The full code is pulished on (https://github.com/Wang-CODEPROJECT/R-statstical-inference (https://github.com/Wang-CODEPROJECT/R-statstical-inference))

```
setwd('D:/R-Studio/Coursera/Stats/Statistical-Inference-master/part 2')
suppressWarnings(library(ggplot2)); suppressMessages(library(data.table));
library(grid);
source('multiplot.R');
```

Data Cleansing:

From the source, we will load the `ToothGrowth` data into a `data.table` object, change the column names to something more meaningful, and declare a join key.

```
# load data and make column names meaningful
dt<-data.table(ToothGrowth)
setnames(dt,c('len','supp','dose'),c('Length','Supplement','Dose'))

# add 'Dosage'and set the join key
dt<-dt[,Dosage:=sapply(as.character(dt$Dose),function(x) as.factor(switch(x,'0.5'='SM','1'='M
D','2'='LG')))]
setkey(dt,Supplement,Dosage)
head(dt,1)
```

```
##    Length Supplement Dose Dosage
## 1:   15.2         OJ  0.5     SM
```

Exploratory Analysis:

The following result sets are two a simple exploratory methods to understand the content and the structure of the `data.table`

```
summary(dt)
```

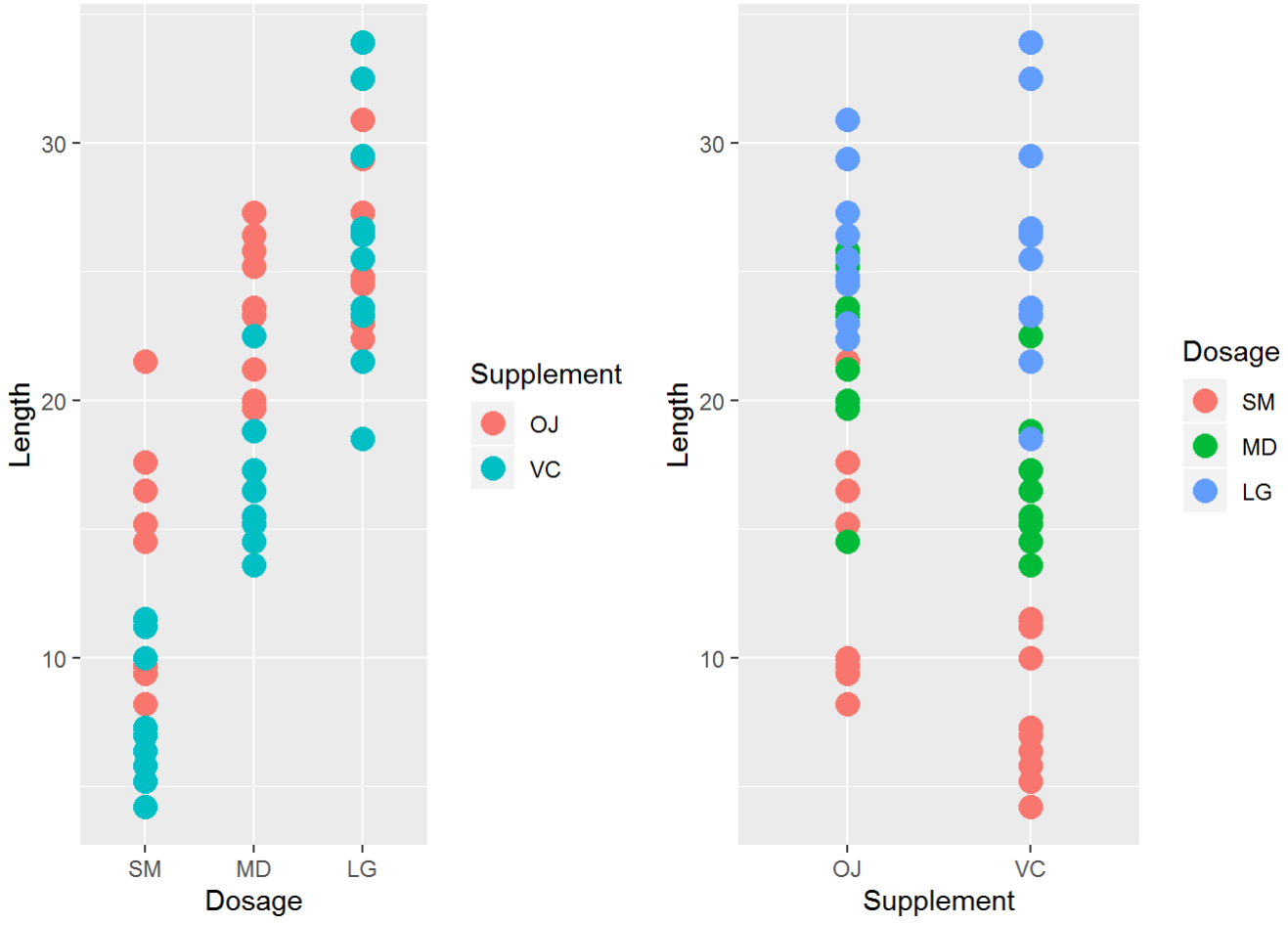
```
##      Length      Supplement      Dose       Dosage
##  Min.   : 4.20   OJ:30      Min.   :0.500   SM:20
##  1st Qu.:13.07   VC:30      1st Qu.:0.500   MD:20
##  Median :19.25              Median :1.000   LG:20
##  Mean   :18.81              Mean   :1.167
##  3rd Qu.:25.27              3rd Qu.:2.000
##  Max.   :33.90              Max.   :2.000
```

```
str(dt)
```

```
## Classes 'data.table' and 'data.frame':   60 obs. of  4 variables:
##  $ Length    : num  15.2 21.5 17.6 9.7 14.5 10 8.2 9.4 16.5 9.7 ...
##  $ Supplement: Factor w/ 2 levels "OJ","VC": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Dose      : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
##  $ Dosage    : Factor w/ 3 levels "SM","MD","LG": 1 1 1 1 1 1 1 1 1 1 ...
##  - attr(*, ".internal.selfref")=<externalptr>
##  - attr(*, "sorted")= chr  "Supplement" "Dosage"
```

To further conduct the exploratory analysis, we can plot `Length` against both `Dosage` and `Supplement`. When we do this we see that the larger the `Dosage`, the longer the tooth `Length`. However, it is slightly unclear as to which supplement is more effective, Orange Juice `OJ` or Ascorbic Acid `VC`.

```
# plot 1
g1<-ggplot(dt,aes(x=Dosage,y=Length))
g1<-g1+geom_point(aes(color=Supplement),size=4)
# plot 2
g2<-ggplot(dt,aes(x=Supplement,y=Length))
g2<-g2+geom_point(aes(color=Dosage),size=4)
# plot together
multiplot(g1,g2,cols=2)
```



Compare Dosage Alone

```
t1<-subset(dt,Dosage=='SM')$Length
t2<-subset(dt,Dosage=='MD')$Length
t<-t.test(t1,t2,paired=FALSE,var.equal=FALSE)
t$conf.int[1:2]
```

```
## [1] -11.983781  -6.276219
```

We increase the Vitamin C dose from 0.5 to 1.0 milligrams, the confidence interval does not contain zero, so we can reject the null hypothesis that this dose increase does not increase tooth length.

```
t1<-subset(dt,Dosage=='MD')$Length
t2<-subset(dt,Dosage=='LG')$Length
t<-t.test(t1,t2,paired=FALSE,var.equal=FALSE)
t$conf.int[1:2]
```

```
## [1] -8.996481 -3.733519
```

Next, if we increase the Vitamin C dose from 1.0 to 2.0 milligrams, the confidence interval againg does not contain zero, so we can reject the null hypothesis that this dose increase does not increase tooth length.

Compare Supplement Alone

```
t1<-subset(dt,Supplement=='VC')$Length
t2<-subset(dt,Supplement=='OJ')$Length
t<-t.test(t1,t2,paired=FALSE,var.equal=FALSE)
t$p.value
```

```
## [1] 0.06063451
```

```
t$conf.int[1:2]
```

```
## [1] -7.5710156  0.1710156
```

Compare Supplement by Each Dosage

```
t1<-subset(dt,Supplement=='VC' & Dosage=='SM')$Length
t2<-subset(dt,Supplement=='OJ' & Dosage=='SM')$Length
t<-t.test(t1,t2,paired=FALSE,var.equal=FALSE)
t$conf.int[1:2]
```

```
## [1] -8.780943 -1.719057
```

When we continue the analysis, and compare a ‘SM’ dosage of Ascorbic Acid to a’SM’ dosage of Orange Juice, we see the confidence interval does not contain zero, so we can reject the null hypothesis that supplement type with a ‘SM’ dosage does not affect tooth growth.

```
t1<-subset(dt,Supplement=='VC' & Dosage=='MD')$Length
t2<-subset(dt,Supplement=='OJ' & Dosage=='MD')$Length
t<-t.test(t1,t2,paired=FALSE,var.equal=FALSE)
t$conf.int[1:2]
```

```
## [1] -9.057852 -2.802148
```

Next, we compare a ‘MD’ dosage of Ascorbic Acid to a ‘MD’ dosage of Orange Juice, and, again, we see the confidence interval does not contain zero; so, we can reject the null hypothesis that supplement type with a ‘MD’ dosage does not affect tooth growth.

```
t1<-subset(dt,Supplement=='VC' & Dosage=='LG')$Length
t2<-subset(dt,Supplement=='OJ' & Dosage=='LG')$Length
t<-t.test(t1,t2,paired=FALSE,var.equal=FALSE)
t$p.value
```

```
## [1] 0.9638516
```

```
t$conf.int[1:2]
```

```
## [1] -3.63807  3.79807
```

Lastly, we compare a ‘LG’ dosage of Ascorbic Acid to a ‘LG’ dosage of Orange Juice; this time, however, we observer the confidence interval contains zero and there is a p-value of almost 1.0. In turn, we do not reject the null hypothesis that supplement type with a ‘LG’ dosage does not affect tooth growth. Meaning, with a ‘LG’ Dosage, we cannot conclude which supplement type has a greater affect on tooth growth.

Conclusions:

1. As Vitamin C dose size alone increases, the tooth length increases as well, and
2. Irrespective of dose size, supplement type alone does not affect tooth growth; however,
3. The supplement type of Orange Juice, or ‘OJ’, affects tooth length greater then Ascorbic Acid, or ‘VC’, with a 0.5 and 1.0 dose size, in turn,
4. When the dose size reached 2.0 milligrams, there is no difference between Orange Juice and Ascorbic Acid.