Tooth Growth data analysis

Borye

Friday, November 07, 2014

Introduction

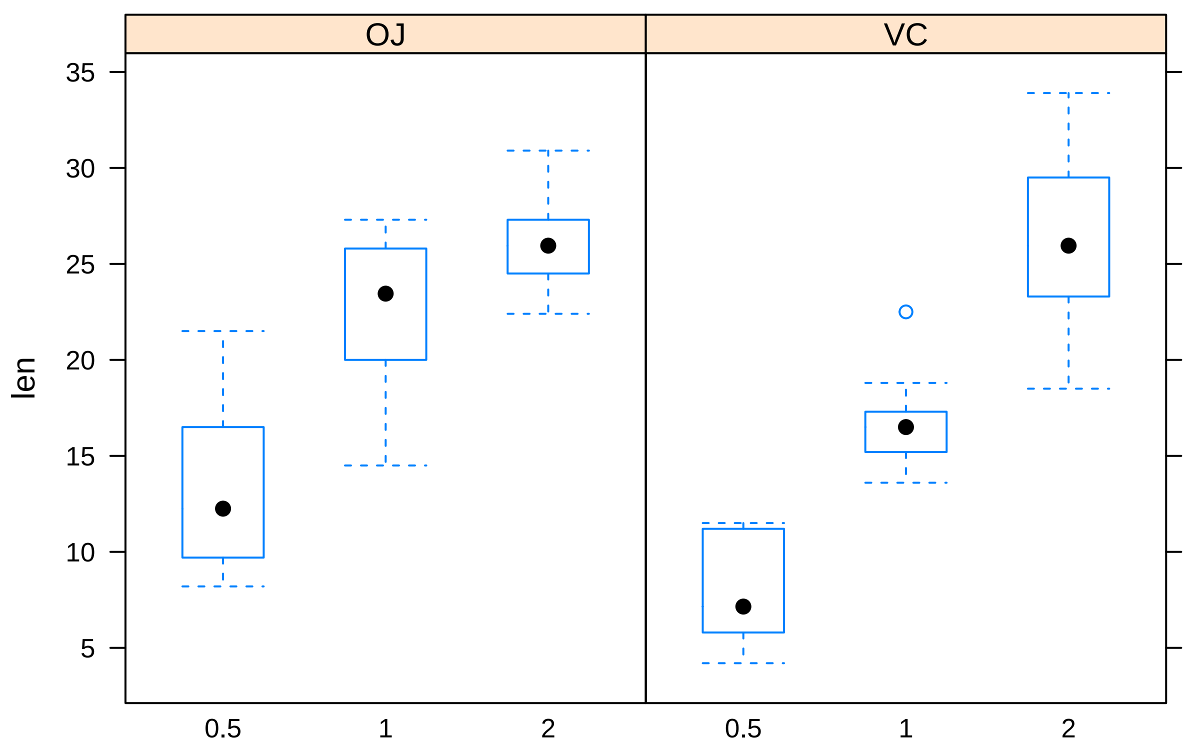
This is a project for the statistical inference class hold by coursera. In this part of the project, I will analyze the ToothGrowth data in the R datasets packages

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```r
library(lattice)
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 ...
```

```r
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
bwplot(len ~ dose | supp, data = ToothGrowth)
```



2. Provide a basic summary of the data

```
data(ToothGrowth)
len <- data.frame(ToothGrowth$len)
colnames(len) <- "len"
mean_len<- aggregate(len, list(supp = ToothGrowth$supp, dose = ToothGrowth$dose), mean)
sd_len <- aggregate(len, list(supp = ToothGrowth$supp, dose = ToothGrowth$dose), sd)
```

The mean of the dataset by supp is shownd below

```
mean_len
```

```
##   supp dose   len
## 1   OJ  0.5 13.23
## 2   VC  0.5  7.98
## 3   OJ  1.0 22.70
## 4   VC  1.0 16.77
## 5   OJ  2.0 26.06
## 6   VC  2.0 26.14
```

The standard deviation of the dataset by supp is shownd below

```
sd_len
```

```
##   supp dose      len
## 1   OJ  0.5 4.459709
## 2   VC  0.5 2.746634
## 3   OJ  1.0 3.910953
## 4   VC  1.0 2.515309
## 5   OJ  2.0 2.655058
## 6   VC  2.0 4.797731
```

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

Calculate the confidence interval of the tooth growth by different supp

```
g1_supp <- mean_len[mean_len$supp == "OJ", ]$len
g2_supp <- mean_len[mean_len$supp == "VC", ]$len
t.test(g1_supp - g2_supp)
```

```
##
##  One Sample t-test
##
## data:  g1_supp - g2_supp
## t = 1.9472, df = 2, p-value = 0.1909
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.475757 11.875757
## sample estimates:
## mean of x
##       3.7
```

Calculate the confidence interval of the tooth growth by different dose

```
g_dose_2 <- mean_len[mean_len$dose == 2.0, ]$len
g_dose_1 <- mean_len[mean_len$dose == 1.0, ]$len
g_dose_0.5 <- mean_len[mean_len$dose == 0.5, ]$len
t.test(g_dose_2 - g_dose_1)
```

```
##
##  One Sample t-test
##
## data:  g_dose_2 - g_dose_1
## t = 2.1181, df = 1, p-value = 0.2808
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -31.81715  44.54715
## sample estimates:
## mean of x
##     6.365
```

```
t.test(g_dose_1 - g_dose_0.5)
```

```
##
##  One Sample t-test
##
## data:  g_dose_1 - g_dose_0.5
## t = 26.8529, df = 1, p-value = 0.0237
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   4.80989 13.45011
## sample estimates:
## mean of x
##      9.13
```

```
t.test(g_dose_2 - g_dose_0.5)
```

```
##
##  One Sample t-test
##
## data:  g_dose_2 - g_dose_0.5
## t = 5.8143, df = 1, p-value = 0.1084
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -18.36704  49.35704
## sample estimates:
## mean of x
##    15.495
```

4. State your conclusions and the assumptions needed for your conclusions

From the results we can observed that the **change in supp is the least efficient one**. The efficient of dose increasing is better then the change of supp. If we set the unit of increasing dose to 0.5. Then the **dose increasing from 0.5 to 1 is the most efficient one**. And the **confidence interval of the dose increasing from dose 0.5 to 1 is the narrowest**.