Statistical Inference Course - ToothGrowth Data Analysis

Vijay Vaidya

Thursday, July 23, 2015

Overview

We will analyze the "ToothGrowth" data from R package and compare the impact of the toothgrowh with the supplement and dose of supplement

Exploratory data analysis

We will load "ToothGrowth" data and do some basic analysis

```
rm(list = ls()) # clear the environment
data(ToothGrowth) # load the data
class(ToothGrowth) # check the class/data type
```

```
## [1] "data.frame"
```

Now that we know that we have frame, we can do basic analysis

```
#Total no.of rows
nrow(ToothGrowth)
```

```
## [1] 60
```

```
#Column names
colnames(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

```
#Supplement types
levels(ToothGrowth$supp)
```

```
## [1] "OJ" "VC"
```

```
#How many unique doses
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
#Max length
ToothGrowth[ which.max(ToothGrowth$len), ]
```

```
##     len supp dose
## 23 33.9   VC    2
```

```r
#Min Length
ToothGrowth[ which.min(ToothGrowth$len), ]
```

```
##   len supp dose
## 1 4.2   VC  0.5
```

```r
#Avg Length
mean(ToothGrowth$len)
```

```
## [1] 18.81333
```

```r
#Avg Length of tooth from the group taking OJ
mean(ToothGrowth[ ToothGrowth$supp == "OJ", ]$len)
```

```
## [1] 20.66333
```

```r
#Avg Length of tooth from the group taking VC
mean(ToothGrowth[ ToothGrowth$supp == "VC", ]$len)
```

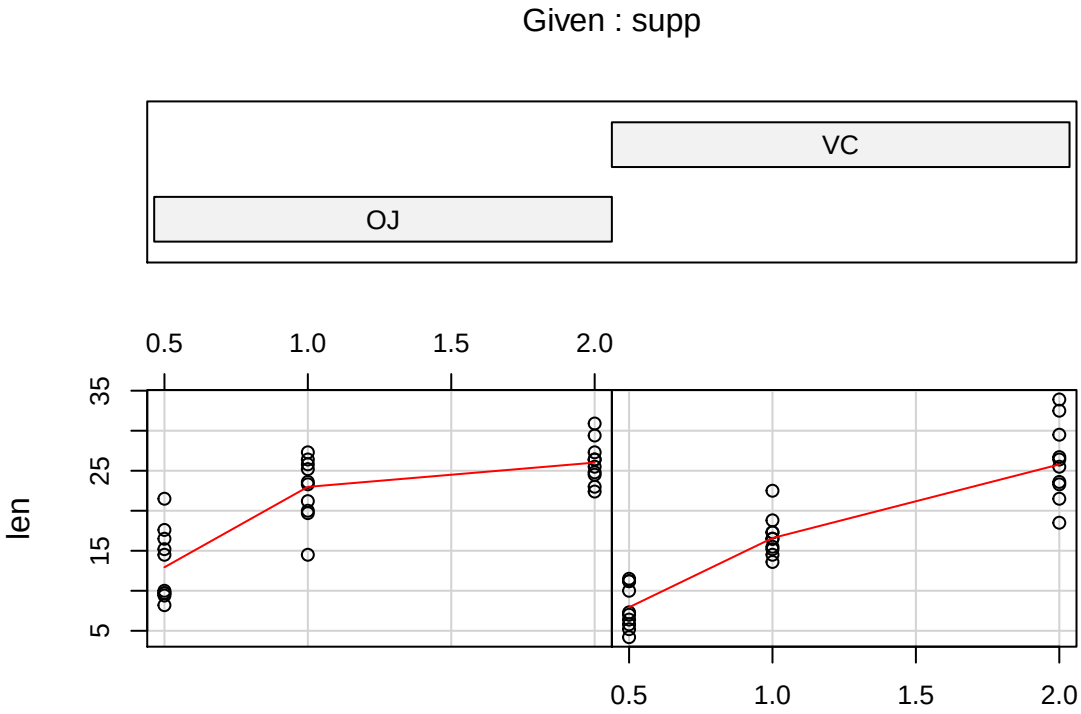
```
## [1] 16.96333
```

```r
#Any NA values?
which(is.na(ToothGrowth$len))
```

```
## integer(0)
```

we can now chart a graph to show consumption of orange juice or vitamin C and the doses impacts teeth growth:

```r
library(ggplot2)
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,xlab = "ToothGrowth data: length vs
```



ToothGrowth data: length vs dose, given type of supplement

Comparison of tooth growth by supp and dose

So we want to compare whether length of the tooth dependins upon the dose of OJ or VC. So we split *ToothGrowth* by doses and calculate the respective confidence intervals:

```
doses <- unique(ToothGrowth$dose)
```

** Dose 0.5 mg**

Avg tooth length with OJ - 13.23 Std deviation 4.4597085

Avg tooth length with VC - 7.98 Std deviation 2.7466343

```
d <- ToothGrowth[ ToothGrowth$dose == doses[ 1 ], ]
t.test(len ~ supp, data = d, paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.770262 8.729738
## sample estimates:
```

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

** Dose 1mg**

Avg tooth length with OJ - 22.7 Std deviation 3.9109533

Avg tooth length with VC - 16.77 Std deviation 2.5153087

```
d <- ToothGrowth[ ToothGrowth$dose == doses[ 2 ], ]
t.test(len ~ supp, data = d, paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.840692 9.019308
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

** Dose 2mg**

Avg tooth length with OJ - 26.06 Std deviation 2.6550581

Avg tooth length with VC - 26.14 Std deviation 4.7977309

```
d <- ToothGrowth[ ToothGrowth$dose == doses[ 3 ], ]
t.test(len ~ supp, data = d, paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.722999  3.562999
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Conclusion

1. From the results we can conclude that when dose size s of 0.5mg, average tooth growth is higher when suppeent of OJ is provied than the supplement of VC. We can say that with 95% confidence because p-value (0.005304) is < 0.01.
2. Same thing (as described above) is true for the dose size of 1mg. We have the strong evidence for thes dose size of 1mg because p-value (0.0007807) is too small.
3. However we can't cconclude the same for the dose of 2mg because p-value (0.9637) is way higher that 0.01