Analysis of the ToothGrowth data

Andrey Alferov

Sunday, April 26, 2015

This report provides the results of analysis of the ToothGrowth data. It contains:

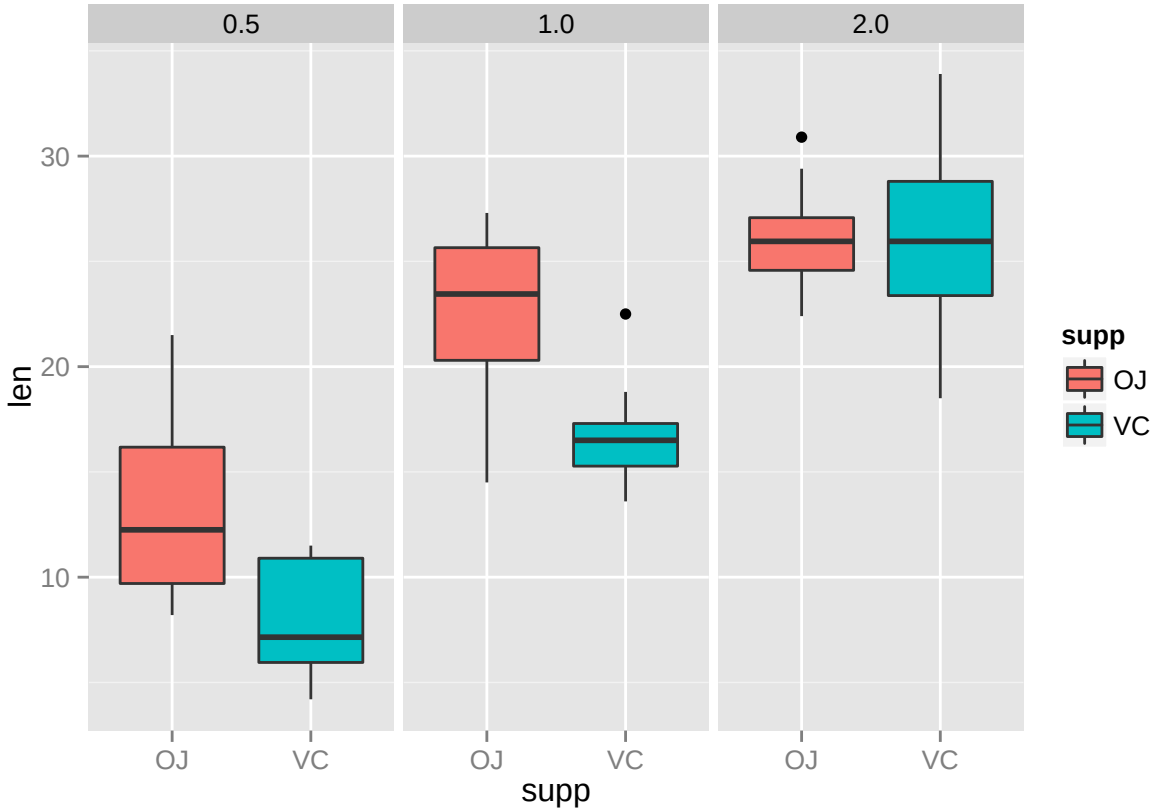
1. basic exploratory data analyses of the data;
2. tooth growth comparative analysis by supp and dose.

```
library(ggplot2)

data(ToothGrowth) #Load dataset
```

ToothGrowth exploratory analysis Let's plot tooth growth for different supplements and different doses.

```
ggplot(aes(x = supp, y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = supp)) + facet_wrap(~ dose)
```



As one can see the OJ supplement is more effective at lower and intermediate dose levels but at higher dose levels both supplements have similar effects.

The summary statistics of ToothGrowth dataset:

```
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
```

Analysis of tooth growth by supp and dose Analysis was performed assuming there are independent groups of observations in the dataset as well as unequal variance for different groups.

Split original dataset by dose level.

```
Tooth.dose05 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose10 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose20 <- subset(ToothGrowth, dose == 2.0)
```

The t-test for the lowest dose level (0.5 mg) for two different supplements

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose05)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

Confidence interval of [1.72, 8,78] for mean(OJ)-mean(VC) shows that we can reject the null hypothesis and that orange juice (OJ) is more effective for this dose level.

The t-test for the lowest dose level (1.0 mg) for two different supplements

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose10)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

Confidence interval of [2.80, 9.06] for mean(OJ)-mean(VC) shows that we can reject the null hypothesis and that orange juice (OJ) is more effective for this dose level.

The t-test for the lowest dose level (2.0 mg) for two different supplements

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose20)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Confidence interval of [-3.80, 3.64] for mean(OJ)-mean(VC) shows that we cannot reject the null hypothesis that there is not a significant difference in tooth length between the two supplement types at this dose level.

Conclusion Analysis via confidence intervals prooved the exploratory analysis conclusion that OJ supplement is more effective at lower and intermediate dose levels but at higher dose levels both supplements have similar effects.