Statistical Inference on the ToothGrow data in R

Cleitus

Saturday, January 17, 2015

We are going to analyze the ToothGrowth data in the R datasets package.
Loading required packages.

```
# Load necessary library
library(ggplot2)
library(datasets)
```

1. Exploratory Data Analysis

We're gonna store the ToothGrowth data in the dataframe `df`.

```
df <- data.frame(ToothGrowth)
str(df)
```

```
## '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 ...
```

60 total Observations with 3 columns :

- Tooth Length (len)
- Supplementary (supp) : OJ(orange juice) or VC(vitamin C)
- Dosage (dose) : 0.5, 1 or 2

```
summary(df)
```

```
##       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
```

Things are starting to unfold, basically we now know that we need to compare tooth length for both supplements as well as the effect of the dose of each supplement.

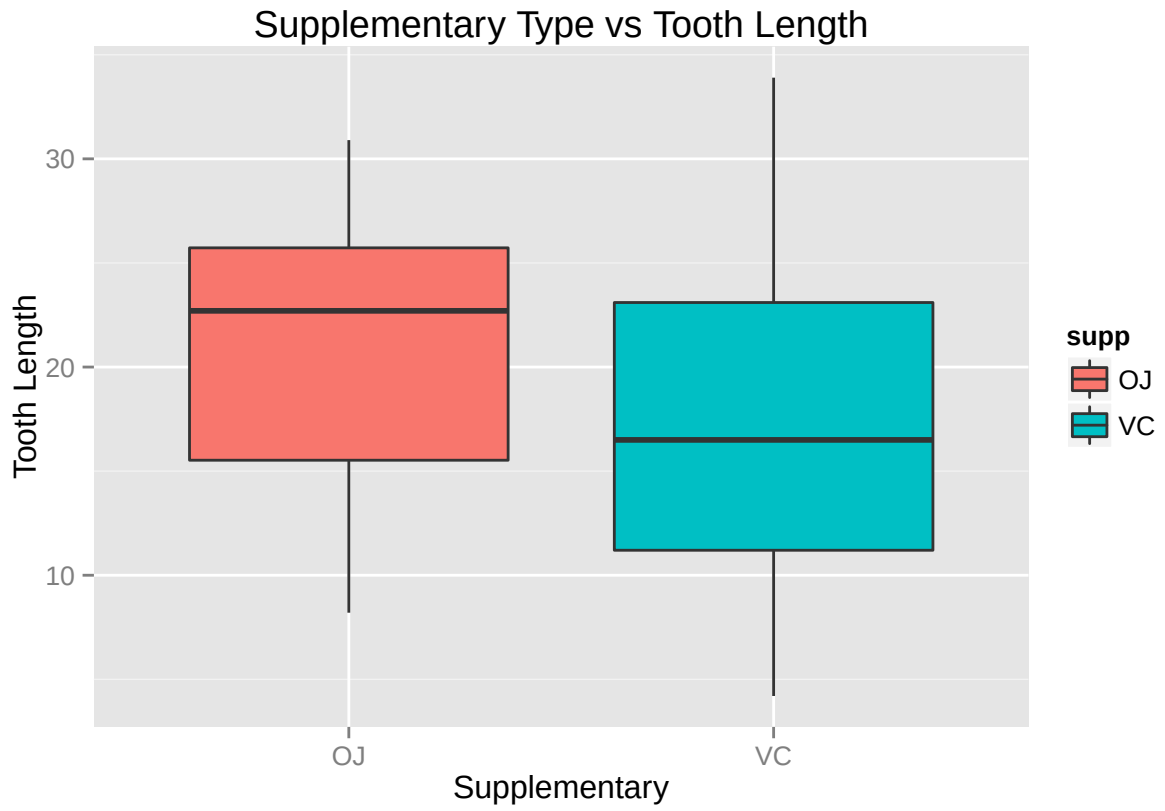
We could start by splitting the data according to what supplement is being used.

```
df.len = data.frame(split(df, df$supp))
head(df.len)
```

```
##    OJ.len OJ.supp OJ.dose VC.len VC.supp VC.dose
## 31   15.2      OJ     0.5    4.2      VC     0.5
## 32   21.5      OJ     0.5   11.5      VC     0.5
## 33   17.6      OJ     0.5    7.3      VC     0.5
## 34    9.7      OJ     0.5    5.8      VC     0.5
## 35   14.5      OJ     0.5    6.4      VC     0.5
## 36   10.0      OJ     0.5   10.0      VC     0.5
```

Plot of Supplement Type VS Tooth Length

```
ggplot(data=df, aes(x=supp, y=len)) +
  geom_boxplot(aes(fill=supp), outlier.colour = "green", outlier.size = 3) +
  labs(title="Supplementary Type vs Tooth Length", x="Supplementary", y="Tooth Length")
```



Plot of Dose VS Tooth Length

```
p <- ggplot(data=df, aes(x=dose, y=len)) +
  geom_boxplot(aes(fill=factor(dose)), outlier.colour = "green", outlier.size = 3) +
  labs(title="Dose Quantity vs Tooth Length", x="Dose Quantity", y="Tooth Length")
p + facet_grid(. ~ supp)
```

Preliminary conclusions :

- OJ seems to have higher affect on tooth length than VC
- It also seems that, for both supplements, the higher the dose the longer the tooth

2. Statistical Inference

2.1 Effect of Supplements

The null hypothesis (H0) is such as : Supplements of Orange Juice and Vitamin C do not have differential affects on the tooth length.
From our exploratory analysis we might be tempted to reject H0 in favor of OJ having higher effect than VC.

Let's settle this with a t-test. First find the variance of the 2 sets of supplements.

```
sapply(split(df$len, df$supp), var)
```

```
##       OJ       VC
## 43.63344 68.32723
```

Variances look pretty far apart therefore we wont assume equality of variances in the following t-test.

```
t.test(df.len$VC.len, df.len$OJ.len, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  df.len$VC.len and df.len$OJ.len
## t = -1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

Notice how the 95% confidence interval contains ZERO, thus we can't reject H0 - that means we can not say conclusively that the supplements have any differential impact on the Tooth Length (OJ and VC are the same 'tooth length' wise).

On the other hand, intuitively, we could be asking the alternative question : Is OJ more influential than VC? Therefore, we might think of conducting a one-sided test. In this case :

```
t.test(df.len$OJ.len, df.len$VC.len, paired=FALSE, var.equal=FALSE, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  df.len$OJ.len and df.len$VC.len
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Here we see that the interval doesn't contain ZERO and is positive, so we could infer that the effect of OJ is larger than that of VC on tooth length.

2.2 Effect of the Dose

Now to see the effect of the dose amount on tooth length. The null hypothesis (H0) is such as : Different doses do not have differential affects on the tooth length. Here we have three levels to study :

- [0.5 ; 1.0]
- [1.0 ; 2.0]
- [0.5 ; 2.0] *this one would turn out to be redundant if we reject H0 in either one of the first two*

First level :

```
t.test(df$len[df$dose==1], df$len[df$dose==0.5], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
```

```
##
## data:  df$len[df$dose == 1] and df$len[df$dose == 0.5]
## t = 6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276252 11.983748
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Since in a sense we tested $1mg - 0.5mg$, then an all positive confidence interval suggests that $1mg$ dose has greater impact of tooth growth than a $0.5mg$ dose. Reject H0.

Second level :

```
t.test(df$len[df$dose==2], df$len[df$dose==1], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  df$len[df$dose == 2] and df$len[df$dose == 1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Here too we reject H0. We can then infer that a (2mg) dose has greater impact of tooth growth than a $1mg$ dose.

This obviously entails that we reject H0 for the third level as well.

2.3 Conclusion

- We could not prove with certainty that there is a difference between the supplements on tooth length
- There is a clear increasing effect on the tooth length following an increase in the dose of a supplement