ToothGrowth Analysis with Confidence Intervals

Jesus A. Gonzalez

Wednesday, May 20, 2015

Introduction

The ToothGrowth dataset contains data about the effect of vitamin C on the tooth growth of guinea pigs. A description of this dataset can be found here. The dataset has three attributes:

- len, a numeric value for the tooth length
- supp, a factor with the supplement type (VC or OJ)
- dose, a numeric value for the dose in milligrams

Basic Exploratory Analyses

Initially, we load the ToothGrowth dataset.

```
data(ToothGrowth)
```

Then we use the **str** command in order to know the data types of the attributes and the number of observations.

The number of observations is equal to **60**. We then see the first 10 samples of the ToothGrowth dataset using the **head** command.

```
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 ...

head(ToothGrowth, 10)

##     len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 3   7.3   VC  0.5
## 4   5.8   VC  0.5
## 5   6.4   VC  0.5
## 6  10.0   VC  0.5
## 7  11.2   VC  0.5
## 8  11.2   VC  0.5
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
```

In figure 1 we can appreciate how the tooth length is corelated to the dose in a positive way.

```
library(ggplot2)
ggplot() + geom_jitter(data=ToothGrowth, mapping=aes(x=len, y=dose),
size=1)
```
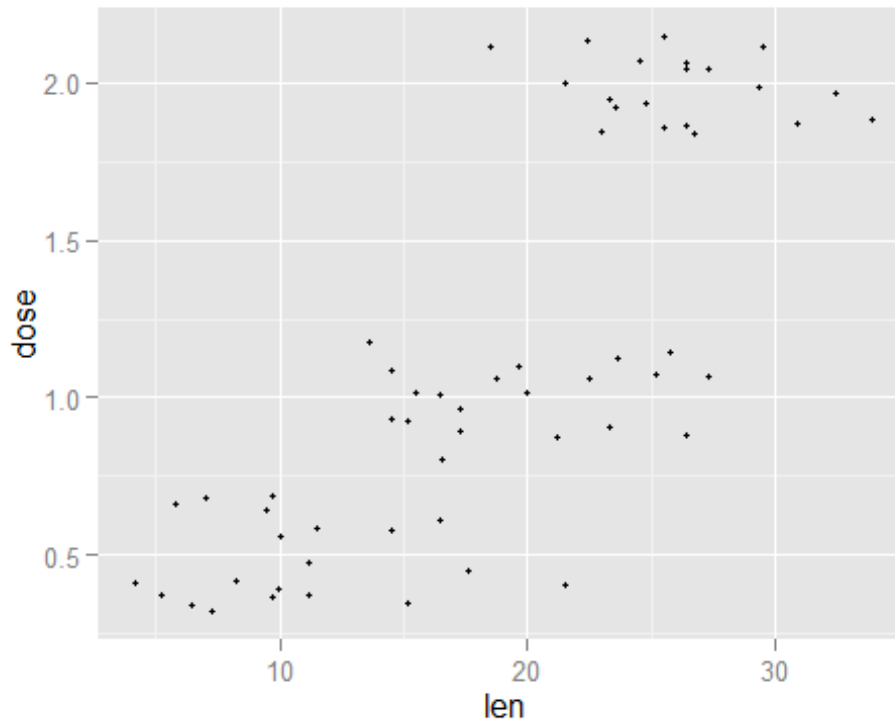

Figure 1. Jitter Plot that shows how the **dose** attribute is positively correlated to the **length** attribute.

Now, in figure 2, we use a coplot that help us see how each type of supplement (independently) is related to the tooth length. In this plot we can see that the tooth length increases with higher doses of supplement. Then, we think that there is a difference between the means of the lengths of tooth according to the dose of supplement and to the type of supplement.

```
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,  xlab
= "ToothGrowth data: length vs dose, given type of supplement")
```

Given : supp

VC

OJ
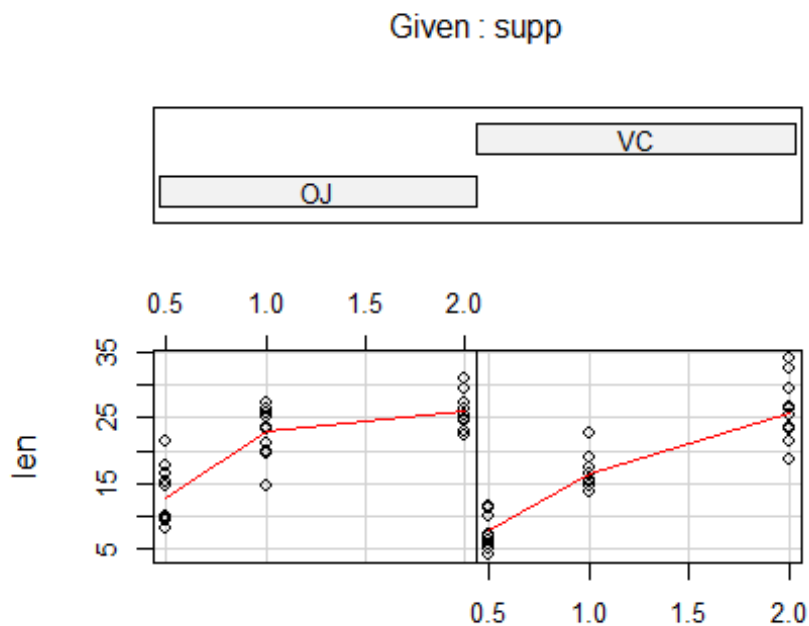
ToothGrowth data: length vs dose, given type of supplement

Figure 2. Plot showing how each type of supplement relates to the tooth length.

Summary of the Data

We summarized the data in order to identify the number of samples per each type of vitamin C supplement and also to know the mean tooth length and the mean dose for the samples of each type of supplement. This can be observed in figure 4. This is the code used to obtain the summary of the data:

```
library(dplyr)
by_supp <- group_by(ToothGrowth, supp)
supp_len <- summarise(by_supp, num_samples = n(), mean_len = mean(len,
na.rm=TRUE), mean_dose = mean(dose, na.rm=TRUE))
```

	supp	num_samples	mean_len	mean_dose
1	OJ	30	20.66333	1.166667
2	VC	30	16.96333	1.166667

Table 1. Summary of the ToothGrowth dataset. We summarized the data to calculate the number of samples, the mean tooth length and the mean dose per each type of supplement.

Confidence Intervals to Compare Tooth Growth by supp and dose

If we compute the differences of tooth length by type of supplement we can obtain the confidence interval for the mean of the difference. We can then perform a hypothesis test to verify if the difference between the means is significant or not. The NULL hipothesis is "the means are equal and there is not effect on tooth growth caused by supplement type". In this case we reject the null hypothesis because the **p-value = 0.00255** is smaller than the confidence level of **0.05**. The confidence interval is **[1.41 - 5.99]**, it is important to note that this confidence interval does not contain the value of **0**.

Experiment to compare the tooth length with respect to the supplement type:

```
g1 <- ToothGrowth$len[1:30]
g2 <- ToothGrowth$len[31:60]
difference <- g2 - g1
mn <- mean(difference)
s <- sd(difference)
n <- 30
mn + c(-1, 1) * qt(.975, n-1) * s / sqrt(n)
```

```
## [1] 1.408659 5.991341
```

```
t.test(difference)
```

```
##
##   One Sample t-test
##
## data:  difference
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   1.408659 5.991341
## sample estimates:
## mean of x
##      3.7
```

Now we compute the difference among the tooth lenghts with respect to the supplement dose and also perform a t-test over the scenario (0.5 vs 2 of dose). In these case we also find that the difference in tooth length is also significative with respect to the supplement dose. In this case the p-value is smaller than the confidence level associated to the t-test with a value of **0.5**, then we reject the null hipothesis that states that the mean of the difference is equal to cero. It is important to note that the confidence interval does not contain the value of **0**.

Experiment comparing a dose of 2 vs a dose of 0.5:

```
g1 <- ToothGrowth$len[ToothGrowth$dose==2]
g2 <- ToothGrowth$len[ToothGrowth$dose==0.5]
difference <- g2 - g1
mn <- mean(difference)
s <- sd(difference)
n <- 30
mn + c(-1, 1) * qt(.975, n-1) * s / sqrt(n)
```

```
## [1] -17.78659 -13.20341
```

```
t.test(difference)
```

```
##
##   One Sample t-test
##
## data:  difference
## t = -11.2915, df = 19, p-value = 7.19e-10
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -18.3672 -12.6228
## sample estimates:
## mean of x
##    -15.495
```

Conclusions and Assumptions

From our experiments we can conclude that the difference in tooth growth with respect to the supplement type and dose is significant at a confidence level of **95%**. For these experiments we assumed that our samples were obtained randomly from an identical and independent distribution.