Statistical Inference Assignment: Part 2

Ferran Martí

Monday, May 18, 2015

Overview

This report conducts an analysis on the ToothGrowth dataset in R. According to the official description, *this set contains data about the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).* The report is divided in three parts: First, an introduction to get the dataset and set up the environment in R. Second, a basic summary of the data. Finally, we will use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

Setting up the environment and the data

First of all, we will set up the working directory:

```
setwd("C:/Users/ftorrent/Desktop/Data Science Track1/Coursera/Statistical Inference")
```

Now, we load the required packages and get the data to perform the analysis:

```
library(knitr)
library(ggplot2)
library(graphics)
library(datasets)
tg<-ToothGrowth
```

Basic Summary of the Data

Let's have a quick view on the data we have:

```
summary(tg)
```

```
##       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
```

As we can see, n the dataset we have three variables: 1. Length of the teeth 2. Supplement type (VC or OJ). This refers to the delivery method (Ascorbic Acid or Orange Juice). VC corresponds to Vitamin C. They have 30 observations each. 3. Dose in mg (0.5, 1 or 2). 10 observations each.

```
table(ToothGrowth$supp, ToothGrowth$dose)
```
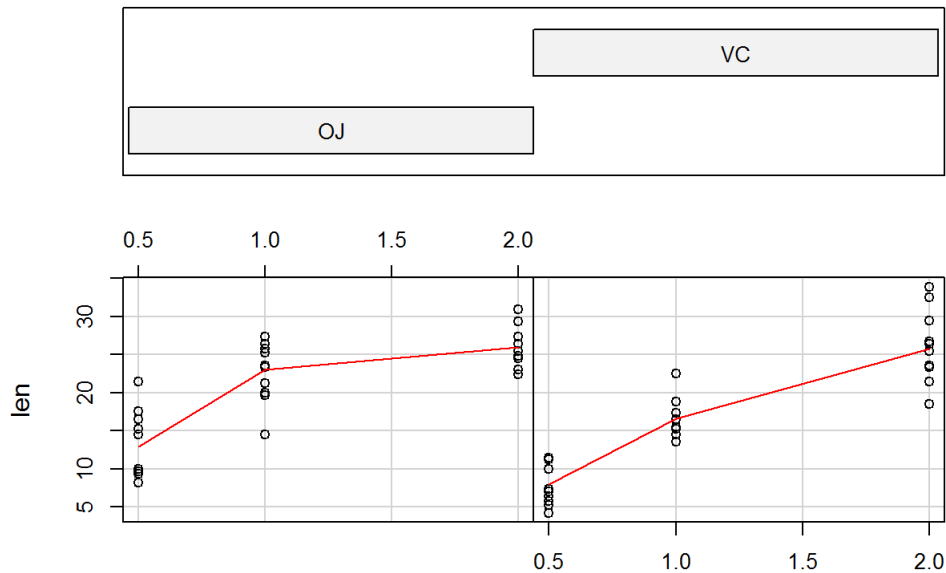
```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

We can observe that there are 10 observations for each combination of dose and supplement.

Let's have a look at the data using a graph to see the values of length and dose, given the type of supplement:

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

Given : supp

ToothGrowth data: length vs dose, given type of supplement

Therefore, as we can see there is a tendency to have larger teeth when being given a 2.0 dose of vitamin, no matter which supplement type.

We can also do a plot box to see this tendency more neat visually:

```
oj <- ToothGrowth[ToothGrowth$supp == "OJ",]
plot.oj <- ggplot(oj, aes(x=factor(dose),y=len,fill=factor(dose))) +
    geom_boxplot() +
    scale_x_discrete("Dose (mg)") +
    scale_y_continuous("Tooth Length") +
    ggtitle("Tooth Length by Dosage of Orange Juice")

# Load and prepare Vitamin C data for plot
vc <- ToothGrowth[ToothGrowth$supp == "VC",]
plot.vc <- ggplot(vc, aes(x=factor(dose),y=len,fill=factor(dose))) +
    geom_boxplot() +
    scale_x_discrete("Dose (mg)") +
    scale_y_continuous("Tooth Length") +
    ggtitle("Tooth Length by Dosage of Vitamin C")
```

```
plot.oj
```

Tooth Length by Dosage of Orange Juice

```
plot.vc
```
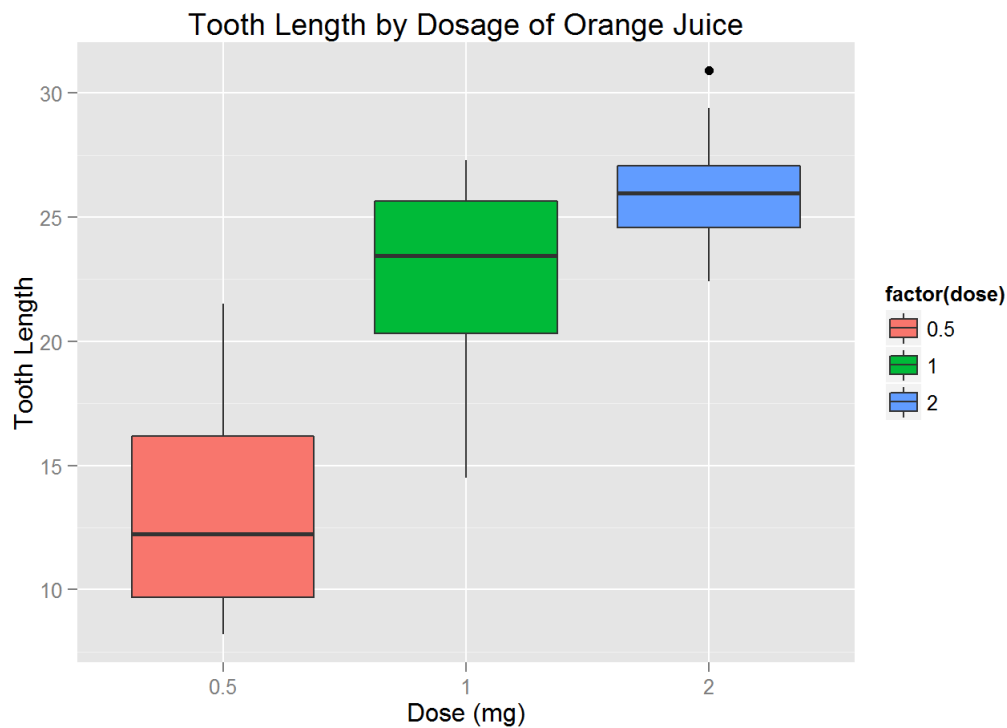

Tooth Length by Dosage of Vitamin C
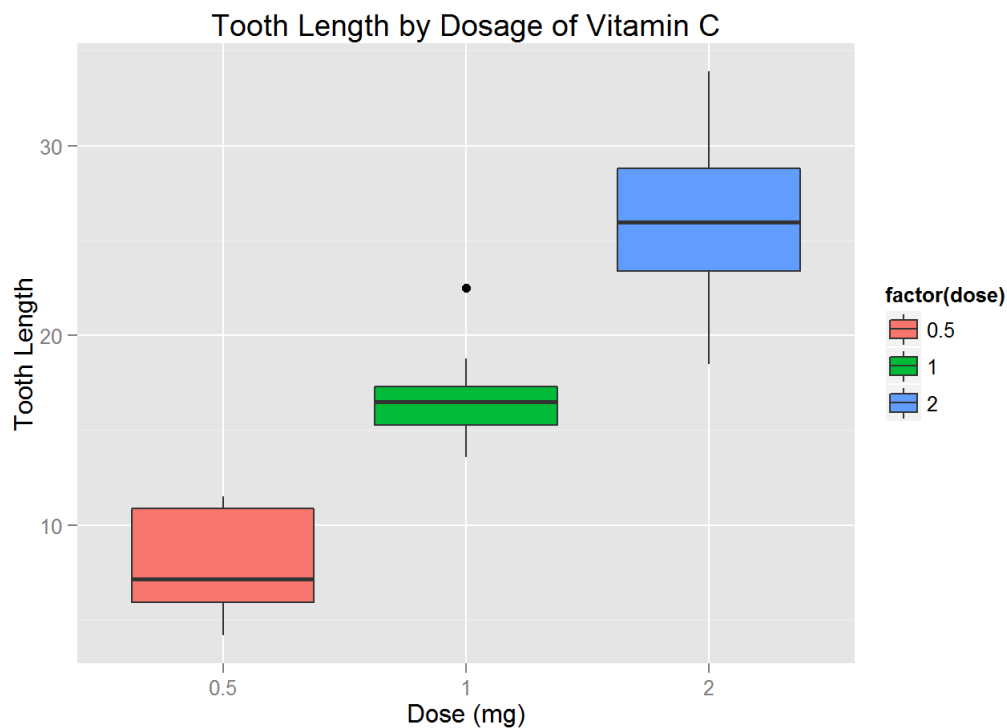
However, to get more accurate results, we have to do some statistical analysis by using confidence intervals and hypothesis tests.

Compare tooth growth by supp and dose.

To compare tooth growth by supplement, we will use a t-test study on the data, with a confidence interval of 95%. The null hypothesis here is that the means of the two groups being tested are equal, independently on how the vitamin is supplied. The alternative hypothesis is that teeth lengh changes depending on which type of supplement type is given.

First, we want to break the data in groups, by dosage and by supplement type, so we have 6 groups of 10 observations each:

```
vc05 <- vc[vc$dose==0.5,]
vc10 <- vc[vc$dose==1.0,]
vc20 <- vc[vc$dose==2.0,]
oj05 <- oj[oj$dose==0.5,]
oj10 <- oj[oj$dose==1.0,]
oj20 <- oj[oj$dose==2.0,]
```

Comparison by supplement type:

Now we can test the effect of supplement type on tooth length, while holding every other variable (in this case, dosage) constant at each level (ceteris paribus).

```
# Perform t test on vc vs oj at 0.5mg dose
t.05.supps.vc.and.oj <- t.test(len ~ supp, data=rbind(vc05,oj05), var.equal=FALSE)

# Perform t test on vc vs oj at 1.0mg dose
t.10.supps.vc.and.oj <- t.test(len ~ supp, data=rbind(vc10,oj10), var.equal=FALSE)

# Perform t test on vc vs oj at 2.0mg dose
t.20.supps.vc.and.oj <- t.test(len ~ supp, data=rbind(vc20,oj20), var.equal=FALSE)
```

Comparisons:

```
t.05.supps.vc.and.oj
```

```
##
##  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
```

```
t.10.supps.vc.and.oj
```

```
##
##  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
```

```
t.20.supps.vc.and.oj
```

```
## 
##  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
```

Let's clarify this a bit with this table:

DOSAGE / P-VALUES / 95% CONFIDENCE INTERVALS

*0.5 dose / 0.006359 / 1.719057 8.780943

*1 dose / 0.001038 / 2.802148 9.057852

*2 dose / 0.9639 /-3.79807 3.63807

Therefore, we can observe that at 0.5 and 1 doses, the null hypothesis is rejected (p-values below 0.05 and confidence intervals without including 0), whereas for a 2 mg dose, the null is not rejected. So this means that for low level doses, it actually matters the supplement type given, as when Orange Juice is given, tooth grow more than the other kind of supplement (VC), with a 95% of confidence. However, for high dosages (2 mg), this difference disappears and it does not matter whereas you apply the dosage through VC or OJ.

Comparison by dosage

```
# Perform t test on 0.5mg vs 1.0mg, within each supplement
t.vc.doses.05.and.10 <- t.test(len ~ dose, data=rbind(vc05,vc10), var.equal=TRUE)
t.oj.doses.05.and.10 <- t.test(len ~ dose, data=rbind(oj05,oj10), var.equal=TRUE)

# Perform t test on 1.0mg vs 2.0mg, within each supplement
t.vc.doses.10.and.20 <- t.test(len ~ dose, data=rbind(vc10,vc20), var.equal=TRUE)
t.oj.doses.10.and.20 <- t.test(len ~ dose, data=rbind(oj10,oj20), var.equal=TRUE)

# Perform t test on 0.5mg vs 2.0mg, within each supplement
t.vc.doses.05.and.20 <- t.test(len ~ dose, data=rbind(vc05,vc20), var.equal=TRUE)
t.oj.doses.05.and.20 <- t.test(len ~ dose, data=rbind(oj05,oj20), var.equal=TRUE)
```

```
t.vc.doses.05.and.10
```

```
## 
##  Two Sample t-test
## 
## data:  len by dose
## t = -7.4634, df = 18, p-value = 6.492e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.264346  -6.315654
## sample estimates:
## mean in group 0.5   mean in group 1
##              7.98             16.77
```

```
t.oj.doses.05.and.10
```

```
## 
##  Two Sample t-test
## 
## data:  len by dose
## t = -5.0486, df = 18, p-value = 8.358e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.410814  -5.529186
## sample estimates:
## mean in group 0.5   mean in group 1
##             13.23             22.70
```

DOSAGE COMPARED / SUPPLEMENT / P-VALUES / 95% CONFIDENCE INTERVALS

*0.5-1 dose / VC /0.000000 / -11.264346 -6.315654

*0.5-1 dose / OJ /0.000083 / -13.410814 -5.529186

So by now we can state that with a 95% of confidence, there is a difference between applying a dosage of 0.5 or 1 mg of vitamin, independently of the supplement given.

```
t.vc.doses.10.and.20
```

```
## 
##  Two Sample t-test
## 
## data:  len by dose
## t = -5.4698, df = 18, p-value = 3.398e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12.96896  -5.77104
## sample estimates:
## mean in group 1 mean in group 2
##           16.77           26.14
```

```
t.oj.doses.10.and.20
```

```
## 
##  Two Sample t-test
## 
## data:  len by dose
## t = -2.2478, df = 18, p-value = 0.03736
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5005017 -0.2194983
## sample estimates:
## mean in group 1 mean in group 2
##           22.70           26.06
```

DOSAGE COMPARED / SUPPLEMENT / P-VALUES / 95% CONFIDENCE INTERVALS

- 1-2 dose / VC /0.000034 / -12.96896 -5.77104

- 1-2 dose / OJ /0.03736 /-6.5005017 -0.2194983

We can state that with a 95% of confidence, there is a difference between applying a dosage of 1 or 2 mg of vitamin, independently of the supplement given, as our p-values are smaller than 0.05, and the confidence intervals do not contemplate 0 within the interval.

```
t.vc.doses.05.and.20
```

```
##
##   Two Sample t-test
##
## data:  len by dose
## t = -10.3878, df = 18, p-value = 4.957e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.83284 -14.48716
## sample estimates:
## mean in group 0.5   mean in group 2
##               7.98             26.14
```

```
t.oj.doses.05.and.20
```

```
##
##   Two Sample t-test
##
## data:  len by dose
## t = -7.817, df = 18, p-value = 3.402e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.278223  -9.381777
## sample estimates:
## mean in group 0.5   mean in group 2
##              13.23             26.06
```

DOSAGE COMPARED / SUPPLEMENT / P-VALUES / 95% CONFIDENCE INTERVALS

- 0.5-2 dose / VC /0 / -21.83284 -14.48716

- 0.5-2 dose / OJ /0 /-16.278223 -9.381777

Obviously, by comparing a 0.5 to a 2 dosage we also find a significant effect on the teeth growth, as applying a 2 mg dose increases more the length of the teeth than applying a 0.5 mg dosage.

Conclusions

We have found that an increase of the dosage from 0.5 to 1 mg and from 1 mg to 2 mg have a significant effect, increasing teeth length with a 95% confidence. This results are robust and independent of the supplement type of administration.

We have also found that giving Orange Juice to the Guinea Pigs increase their tooth lenght more than giving them Vitamin C when the dosage is either 0.5 mg or 1 mg. However, in high doses, the supplement type is indistinct, as we can't reject the null hypothesis of both means being equal.

So let's summarize again all this amount of information with a table:

DOSAGE COMPARED / SUPPLEMENT / P-VALUES / 95% CONFIDENCE INTERVALS *0.5-1 dose / VC /0.006359 / 1.719057 8.780943* 0.5-1 dose / OJ /0.001038 / 2.802148 9.057852 *1-2 dose / VC /0.9639 /-3.79807 3.63807* 1-2 dose / OJ *0.5-2 dose / VC* 0.5-2 dose / OJ