

Statistical Inference Course Project: Part 2

Yiyang Pei

Friday, January 23, 2015

This report is for Part 2 of the course project of the Coursera course “Statistical Inference”. In this part, we will analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

First, we load the ToothGrowth data into R.

```
data(ToothGrowth)
```

Then, we have a quick look at the internal structure of the dataset.

```
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 ...
```

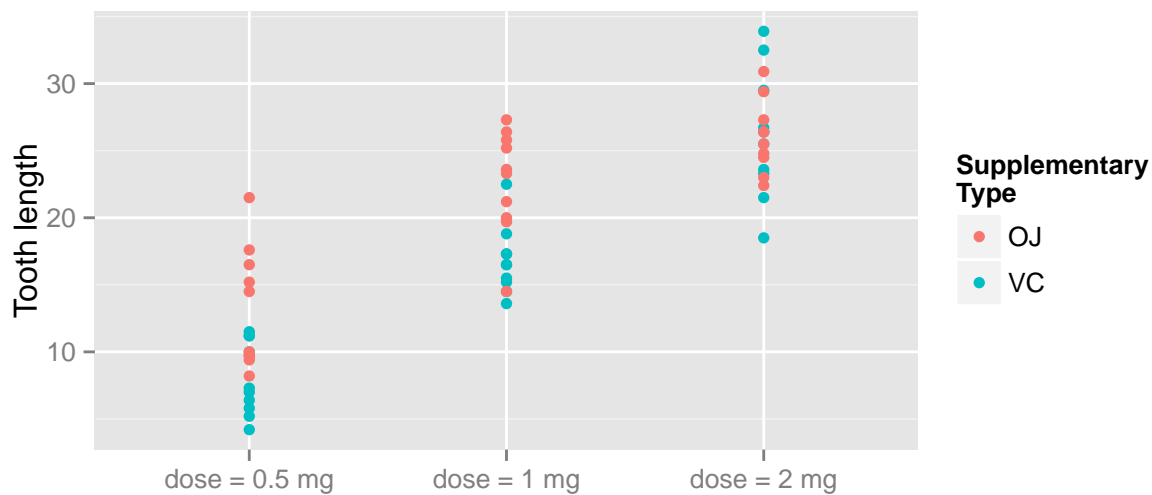
Next, we look at the number of observations for each supplementary type and dose level.

```
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##
##      OJ VC
## 0.5 10 10
## 1   10 10
## 2   10 10
```

In addition, We plot the original data using scatterplot.

```
library(ggplot2)
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
levels(ToothGrowth$dose) <- paste("dose =", levels(ToothGrowth$dose), "mg")
ggplot(ToothGrowth, aes(dose, len, color=supp)) + geom_point() +
  scale_color_discrete(name="Supplementary \nType") +
  labs(x="") + labs(y="Tooth length")
```



2. Provide a basic summary of the data.

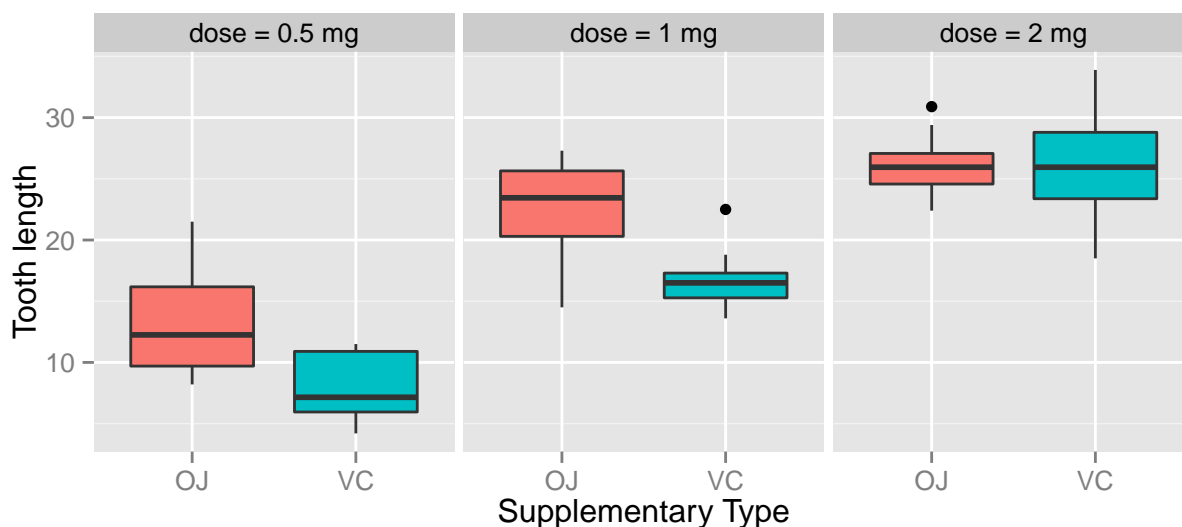
Let's look at the summary by supplementary type and dose levels.

```
library(dplyr)
ToothGrowth %>%
  group_by(supp, dose) %>%
  summarise(ave_len=mean(len), var_len=var(len))
```

```
## Source: local data frame [6 x 4]
## Groups: supp
##
##   supp      dose ave_len  var_len
## 1   OJ dose = 0.5 mg  13.23 19.889000
## 2   OJ dose = 1 mg   22.70 15.295556
## 3   OJ dose = 2 mg   26.06  7.049333
## 4   VC dose = 0.5 mg   7.98  7.544000
## 5   VC dose = 1 mg   16.77  6.326778
## 6   VC dose = 2 mg   26.14 23.018222
```

Graphically, we can use box plot to have a summary of the original data.

```
ggplot(ToothGrowth, aes(x=supp, y=len, fill=supp)) + geom_boxplot() +
  facet_grid(. ~ dose) + guides(fill=FALSE) +
  labs(x="Supplementary Type") + labs(y="Tooth length")
```



3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

For each of three dose levels dose = 0.5 mg, dose = 1 mg, dose = 2 mg, We perform t test to compare the tooth growth for the two supplementary types OJ, VC.

- For dose = 0.5 mg,

```
t.test(len ~ supp, data = subset(ToothGrowth, dose=="dose = 0.5 mg"),
      paired = FALSE, var.equal = FALSE)
```

```
##
##  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
```

Since the confidence interval does not contain 0, we can reject the null hypothesis. The test indicates that OJ is more effective for tooth growth than VC when dose equals 0.5 mg.

- For dose = 1 mg,

```
t.test(len ~ supp, data = subset(ToothGrowth, dose=="dose = 1 mg"),
      paired = FALSE, var.equal = FALSE)
```

```
##
##  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
```

Since the confidence interval does not contain 0, we can reject the null hypothesis. The test indicates that OJ is more effective for tooth growth than VC when dose equals 1 mg.

- For dose = 2 mg,

```
t.test(len ~ supp, data = subset(ToothGrowth, dose=="dose = 2 mg"),
       paired = FALSE, var.equal = FALSE)
```

```
##
## 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
```

Since the confidence interval contains 0, we cannot reject the null hypothesis. The test indicates that there is no significant difference between the two supplementary types when dose equals 2 mg.

4. State your conclusions and the assumptions needed for your conclusions.

- From the above analysis, we have drawn the following conclusions:
 1. When dose is equal to 0.5 mg and 1 mg, OJ is more effective for tooth growth than VC.
 2. When dose is equal to 2 mg, there is no significant difference between two supplementary types.
- The conclusions drawn are based on the following assumptions:
 1. The two groups of subjects receiving the supplementary types VC and OJ are randomly drawn and hence are independent.
 2. The variances of the two groups are to be different.