Statistical Inference - Part 2 Basic inferential data analysis

Andrey Budish

March 31, 2017

Overview

The purpose of this project is to analyze ToothGrowth data in the R datasets package. I will perform basic exploratory data analysis of this data, calculate confidence intervals and perform hypothesis tests to compare tooth growth by supp and dose.

Loading the data

Let's Load ToothGrowth data:

```
library(datasets)
data("ToothGrowth")
```

Basic exploratory data analysis

Let's take a look at the data:

```
head(ToothGrowth)
```

```
##    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
```

```
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 ...
```

From ?ToothGrowth we can read the description and format of the data.

Description
The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Format
A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length
[,2] supp factor Supplement type (VC or OJ).
[,3] dose numeric Dose in milligrams/day

Since we have 3 levels of vitamin C dosage, let's change dose to the factor variable:

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

To count how many observations of each of the unique combinations of supplement/dosage let's make a relevant table

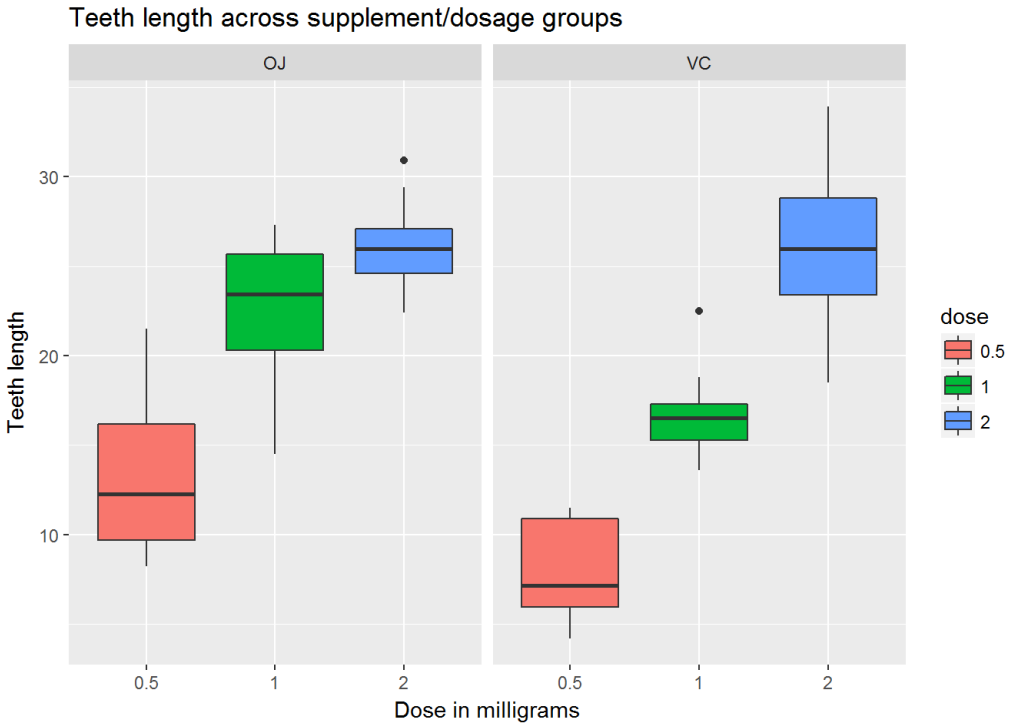
```
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

We can see that we have 10 observations for each of the supplement/dosage group.

Let's built a boxplot which will show changes of the teeth length across supplement/dosage groups:

```r
library(ggplot2)
p <-  ggplot(ToothGrowth,
               aes(x=dose, y=len, fill=dose))
p + geom_boxplot(notch = FALSE) + facet_grid(. ~ supp) +
        scale_x_discrete("Dose in milligrams") +
        scale_y_continuous("Teeth length") +
        ggtitle("Teeth length across supplement/dosage groups")
```



Comparing tooth growth by supp and dose using hypothesis testing

From the plot above we can make a few assumptions:
1. Increasing the dose levels increases the length of teeth.
2. Ascorbic acid and orange juice has different impact on tooth growth at the same dose levels

Let us test those assumptions.

1.1 Testing orange juice (OJ) influence on the teeth length between different dose levels:

```r
# create 3 data frames for each of the dose levels of orange juice that contain relevant tooth length
len_OJ_0.5 <- subset(ToothGrowth, supp == "OJ" & dose == "0.5", select = len)
len_OJ_1 <- subset(ToothGrowth, supp == "OJ" & dose == "1", select = len)
len_OJ_2 <- subset(ToothGrowth, supp == "OJ" & dose == "2", select = len)
```

Now lets perform t-tests to test if increasing the dose levels also increases tooth length for orange juice.
Our t-tests will be independent, since we compare results of 10 different guinea pigs, hence independent groups, and let's assume a different variance.
The null hypothesis will state that there is no difference between tooth length means and alternative hypothesis - that the bigger dose has bigger mean.

```r
# t-test between 1 and 0.5 milligrams/day dose of orange juice
t.test(len_OJ_1 - len_OJ_0.5, paired = FALSE, var.equal = FALSE, alternative = "greater")$p.value
```

```
## [1] 0.00121757
```

```r
# t-test between 2 and 1 milligrams/day dose of orange juice
t.test(len_OJ_2 - len_OJ_1, paired = FALSE, var.equal = FALSE, alternative = "greater")$p.value
```

```
## [1] 0.04191956
```

- **For both tests p-values are less than 0.05 which shows that increasing the dose levels of orange juice increases the length of teeth.**

1.2 Testing ascorbic acid (VC) influence on the teeth length between different dose levels:

```
# create 3 data frames for each of the dose levels of ascorbic acid that contain relevant tooth length
len_VC_0.5 <- subset(ToothGrowth, supp == "VC" & dose == "0.5", select = len)
len_VC_1 <- subset(ToothGrowth, supp == "VC" & dose == "1", select = len)
len_VC_2 <- subset(ToothGrowth, supp == "VC" & dose == "2", select = len)
```

```
# t-test between 1 and 0.5 milligrams/day dose of ascorbic acid
t.test(len_VC_1 - len_VC_0.5, paired = FALSE, var.equal = FALSE, alternative = "greater")$p.value
```

```
## [1] 8.575825e-05
```

```
# t-test between 2 and 1 milligrams/day dose of ascorbic acid
t.test(len_VC_2 - len_VC_1, paired = FALSE, var.equal = FALSE, alternative = "greater")$p.value
```

```
## [1] 0.0002323975
```

- **For both tests p-values are less than 0.05 which shows that increasing the dose levels of ascorbic acid increases the length of teeth**.

2 Testing difference of impact on tooth grows between orange juice(OJ) and ascorbic acid(VC) at the same dose levels

The null hypothesis will state that there is no difference between tooth length means and alternative hypothesis - that orange juice has bigger effect on tooth length than ascorbic acid at the same dose levels.

```
# t-test between orange juice and ascorbic acid, 0.5 milligrams dose
supp_test0.5 <- t.test(len_OJ_0.5 - len_VC_0.5, paired = FALSE, var.equal = FALSE, alternative = "greater")

# t-test between orange juice and ascorbic acid, 1 milligram dose
supp_test1 <- t.test(len_OJ_1 - len_VC_1, paired = FALSE, var.equal = FALSE, alternative = "greater")

# t-test between orange juice and ascorbic acid, 2 milligrams dose
supp_test2 <- t.test(len_OJ_2 - len_VC_2, paired = FALSE, var.equal = FALSE, alternative = "greater")

# table of p-values
data.frame("p-value"= c(supp_test0.5$p.value, supp_test1$p.value, supp_test2$p.value),
           row.names = c("OJ > VC (0.5 milligrams)", "OJ > VC (1 milligram)", "OJ > VC (2 milligrams"))
```

```
##                             p.value
## OJ > VC (0.5 milligrams) 0.007736024
## OJ > VC (1 milligram)    0.004114624
## OJ > VC (2 milligrams    0.516521648
```

- **For 0.5 and 1 milligram, p-value is smaller than 0.05, which shows that orange juice has bigger impact on tooth length than ascorbic acid.**
- **For 2 milligram - p-value is more than 0.05, which shows that there is no difference of impact on tooth grows between orange juice and ascorbic acid.**

Conclusions

Based on the above t-tests we can conclude.
1. Increasing the dose levels increases the length of teeth.
2. At 0.5 and 1 milligram dosages orange juice has greater impact on tooth growth than ascorbic acid.
3. At 2 milligram dosage there is no statistical difference between supplement methods.