Inferential Analysis on Tooth Growth by Supplement Types and Dosages

Jiachang (Ernest) Xu

6/12/2017

Overview

This project is a breif graphical and numericla exploration of the dataset, by performing the techniques of confidence interval and hypothesis testing.

Data Processing and Cleaning

```
library(datasets)
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(RColorBrewer)
```

```
## Loading required package: RColorBrewer
```

```
require(grDevices)

data(ToothGrowth)
attach(ToothGrowth)

# first look: 3 variables and 60 observations
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 ...
```

```
summary(ToothGrowth)
```

```
##       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
```

```
# convert $dose to factors
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

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

Exploratory Analysis

This section examines the relationship between dose size, dose type, and tooth length

```
require(ggplot2)
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
theme <- theme(
    panel.background = element_rect(fill = "lightgrey", colour = "lightgrey", size = 0.5, linetype = "so
    panel.grid.major = element_line(size = 0.5, linetype = 'solid', colour = "white"),
    panel.grid.minor = element_line(size = 0.25, linetype = 'solid', colour = "white")
)

plot1 <- ggplot(aes(x = factor(dose), y = len), data = ToothGrowth) +
                geom_boxplot(aes(fill = factor(dose))) +
                theme + scale_fill_brewer(palette="GnBu") +
                labs(title = "Figure 1")

plot2 <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
                geom_boxplot(aes(fill = supp)) +
                theme + scale_fill_brewer(palette="PuOr") +
                labs(title = "Figure 2")

grid.arrange(plot1, plot2, ncol=2)
```

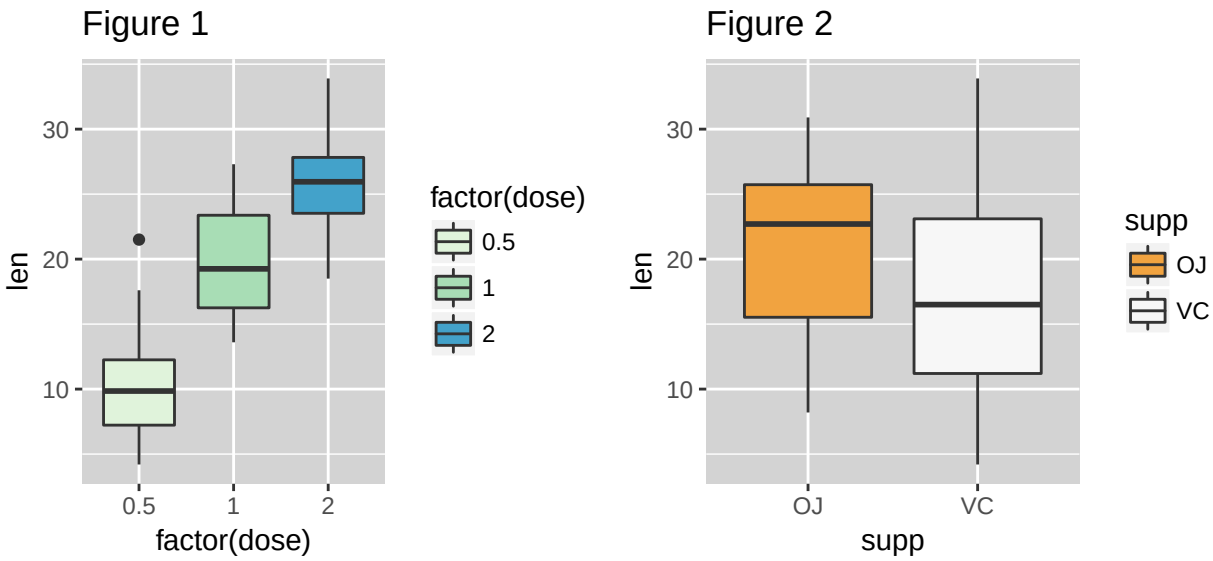

Figure 1 shows that tooth length increases as the doseage increases.

Figure 2 shows that Orange Juice is more effective than Vitamin C, including all doseage levels.

```
require(ggplot2)
require(gridExtra)

plot3 <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
                geom_boxplot(aes(fill = supp)) +
                facet_wrap(~ dose) + theme +
                scale_fill_brewer(palette="YlOrRd") +
```

```
                labs(title = "Figure 3: Tooth Growth due to Two Supplements by Incremental Dosages") +
                annotate("text", x = 1.5, y = 2, label = c("p = .03","p = .0005","p = .5"))
plot3
```

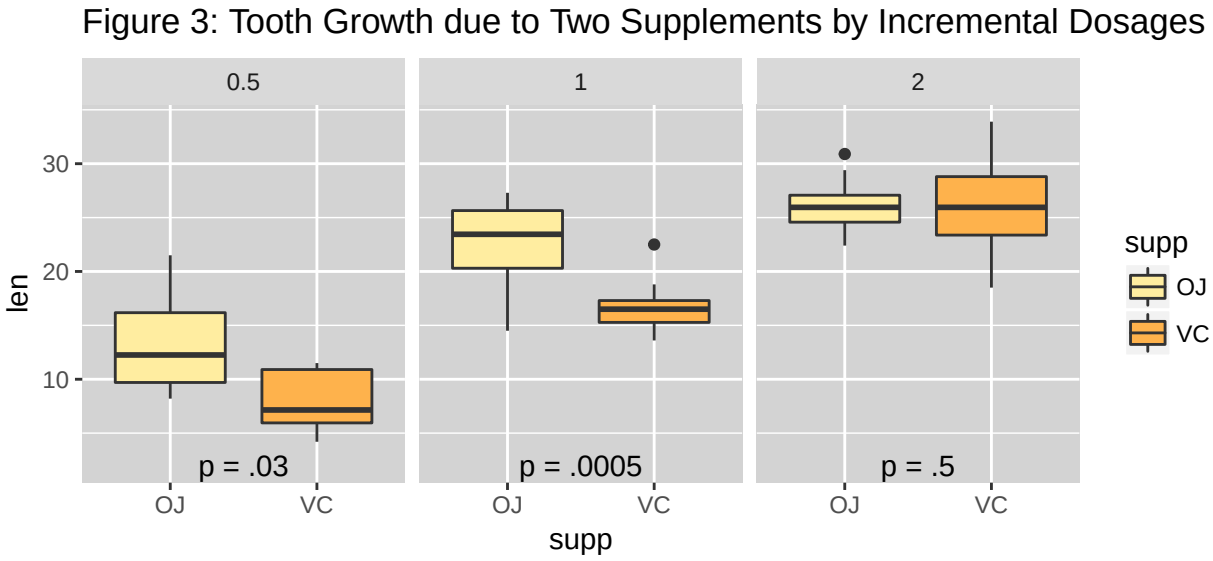

Figure 3 shows that Orange Juice is more effective than Vitamin C only in the first two doseage, but they are equally effective in the third doseage.

Hypothesis Testing

First, we need to reformat the data by using the split() function. Therefore, we can do hypothesis testing more conviently from the reformatted dataset.

```
# split the data frame by dose and supplement type
split_tooth <- split(ToothGrowth, f = list(ToothGrowth$dose, ToothGrowth$supp))
```

Second, aggregations show the sample mean and standard deviation.

```
aggregate(len, list(supp, dose), mean)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 13.23
## 2      VC     0.5  7.98
## 3      OJ     1.0 22.70
## 4      VC     1.0 16.77
## 5      OJ     2.0 26.06
## 6      VC     2.0 26.14
```

```
aggregate(len, list(supp, dose), sd)
```

```
##   Group.1 Group.2        x
## 1      OJ     0.5 4.459709
## 2      VC     0.5 2.746634
## 3      OJ     1.0 3.910953
## 4      VC     1.0 2.515309
## 5      OJ     2.0 2.655058
## 6      VC     2.0 4.797731
```

Third, we perform hypothesis testing at the **5%** significance level. Each p-value correponds to the question immediately above it

Test 1: Is Orange Juice more effective than vitamin C across doses? YES

```
t.test(c(split_tooth[[1]]$len, split_tooth[[2]]$len, split_tooth[[3]]$len),
       c(split_tooth[[4]]$len, split_tooth[[5]]$len, split_tooth[[6]]$len),
       alternative = "greater")$p.value
```

```
## [1] 0.03031725
```

Test 2: Is Orange Juice more effective than Vitamin C for a .5 dose? YES

```
t.test(split_tooth[[1]]$len, split_tooth[[4]]$len,
       alternative = "greater")$p.value
```

```
## [1] 0.003179303
```

Test 3: Is Orange Juice more effective than Vitamin C for a 1 dose? YES

```
t.test(split_tooth[[2]]$len, split_tooth[[5]]$len,
       alternative = "greater")$p.value
```

```
## [1] 0.0005191879
```

Test 4: Is Orange Juice more effective than Vitamin C for a 2 dose? INCONCLUSIVE

```
t.test(split_tooth[[3]]$len, split_tooth[[6]]$len,
       alternative = "greater")$p.value
```

```
## [1] 0.5180742
```

Test 5: Is a 2 dose (any supplement type) more effective than a .5 dose? YES

```
t.test(c(split_tooth[[1]]$len, split_tooth[[4]]$len),
       c(split_tooth[[2]]$len, split_tooth[[5]]$len),
       alternative = "less")$p.value
```

```
## [1] 6.341504e-08
```

Test 6: Is a 2 dose (any supplement type) more effective than a .5 dose? YES

```
t.test(c(split_tooth[[1]]$len, split_tooth[[4]]$len),
       c(split_tooth[[3]]$len, split_tooth[[6]]$len),
       alternative = "less")$p.value
```

```
## [1] 2.198762e-14
```

Test 7: Is a 2 dose (any supplement type) more effective than a .5 dose? YES

```
t.test(c(split_tooth[[2]]$len, split_tooth[[5]]$len),
       c(split_tooth[[3]]$len, split_tooth[[6]]$len),
       alternative = "less")$p.value
```

```
## [1] 9.532148e-06
```