Part 2: Basic Inferential Data Analysis

Hasan Shojaei

20 June 2018

Overview

In Part 2 of this project, we analyze the ToothGrowth data and perform a hypothesis test, which indicates the use of two different supplement types (OJ or VC) does not cause a true difference in the average tooth growth.

Loading the Data and Providing a Summary of the Data

We will first need to load the ToothGrowth data into R environment. In this dataset, 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).

The structure of the data frame and a summary of the data are provided here.

```
library(datasets)
data(ToothGrowth)
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 ...
```

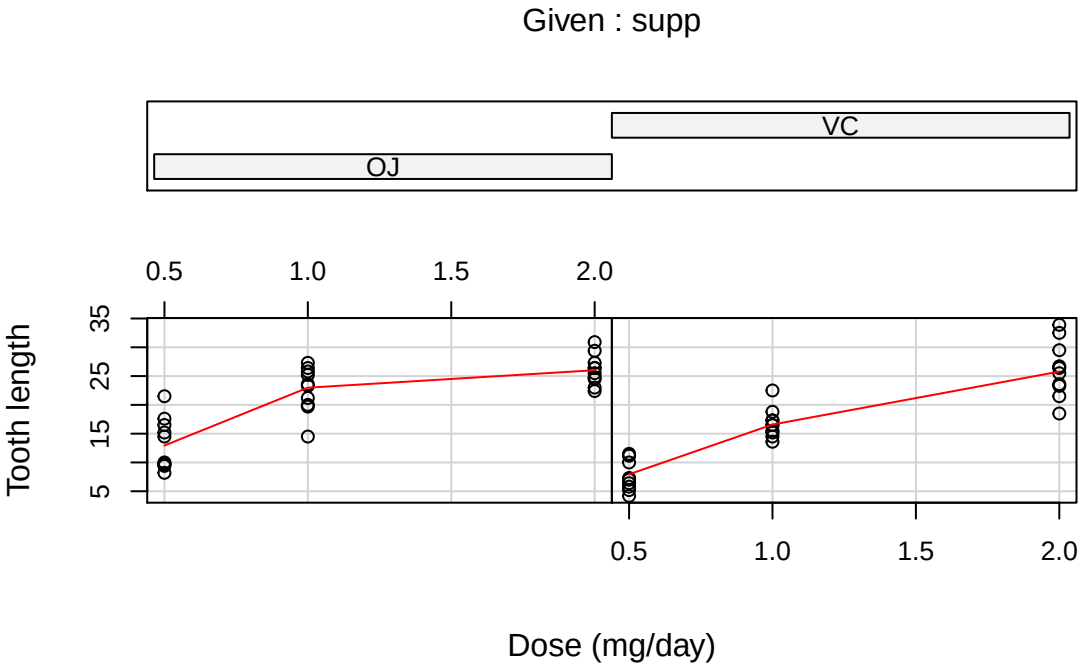
```
summary(ToothGrowth)
```

```
##       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
```

Performing Exploratory Data Analysis

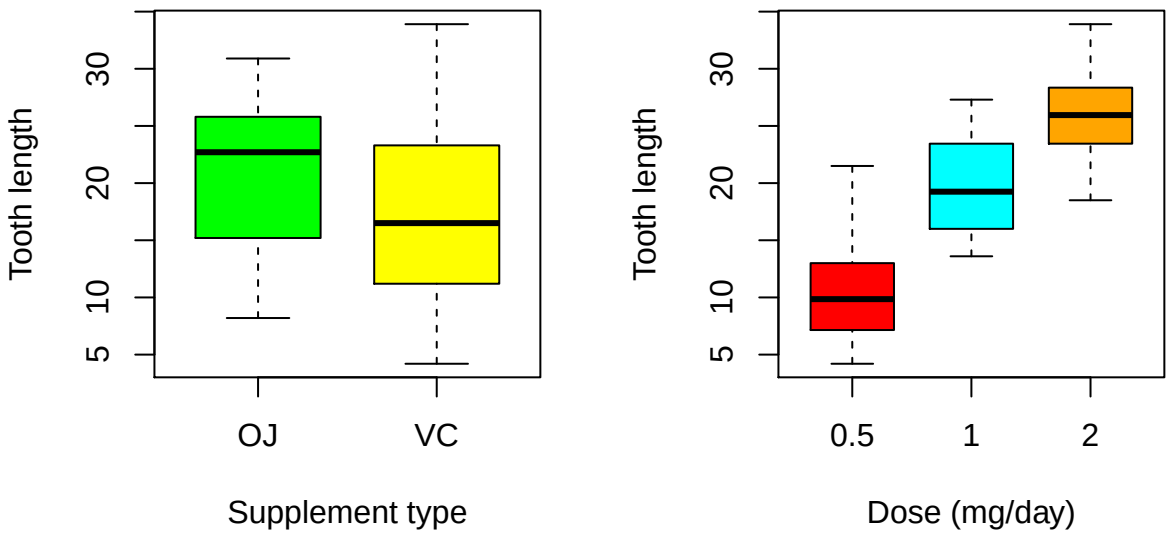
We first use the "graphics" package to create a coplot of length vs dose, given supplement type.

```
library(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "Dose (mg/day)", ylab="Tooth length")
```



We also make boxplots with respect to supplement type and dose to further explore the data.

```
par(mfrow=c(1,2))
boxplot(len~supp,data=ToothGrowth,boxwex=0.7,
        col=c("green", "yellow"), xlab="Supplement type", ylab="Tooth length")
boxplot(len~as.factor(dose),data=ToothGrowth,boxwex=0.7,
        col=c("red", "cyan", "orange"), xlab="Dose (mg/day)", ylab="Tooth length")
```



The main observations from this section is that the tooth length increases with higher doses of both supplement types. It's also observed that the min, Q1, median and Q3 are smaller for VC supplement compared to OJ supplement, while VC supplement results in a larger max value of tooth length.

Performing Hypothesis Test

In this section we perform a t-test to compare tooth growth by supplement type. The null hypothesis is that the true difference in means is equal to 0. The alternative hypothesis is that the true difference in means is NOT equal to 0. Since there are two independent group of pigs in this study, we set paired=FALSE. It's also assumed that the vairances are different.

```
t.test(len ~ supp, data = ToothGrowth,paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Conclusion

The p-value is larger than 0.05 (equivalently 0 is within 95% confidence interval). This means we have weak evidence against the null hypothesis, so we fail to reject the null hypothesis.