

Coursera Statistical Inference Project

Seo-yeon Lim

2018/03/03

Part 2: Basic Inferential Data Analysis - ToothGrowth data

Overview

In this second part of the project, our goal is to conduct a basic inferential data analysis using ToothGrowth data in R datasets package. In this dataset, there are total 60 observations (60 guinea pigs as sample). Each guinea pig received either 0.5, 1 or 2mg/day of vitamin c. Vitamin C was delivered by one of two methods, as orange juice or ascorbic acid. The purpose of this dataset was to see the effect of vitamin c on tooth growth in guinea pigs. ToothGrowth dataset has 3 variables: 1. len: tooth length 2. supp: supplement type (VC or OJ) 3. dose: dose in milligrams/day

Getting ready for the analysis

Load dataset:

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

Basic summary of the data

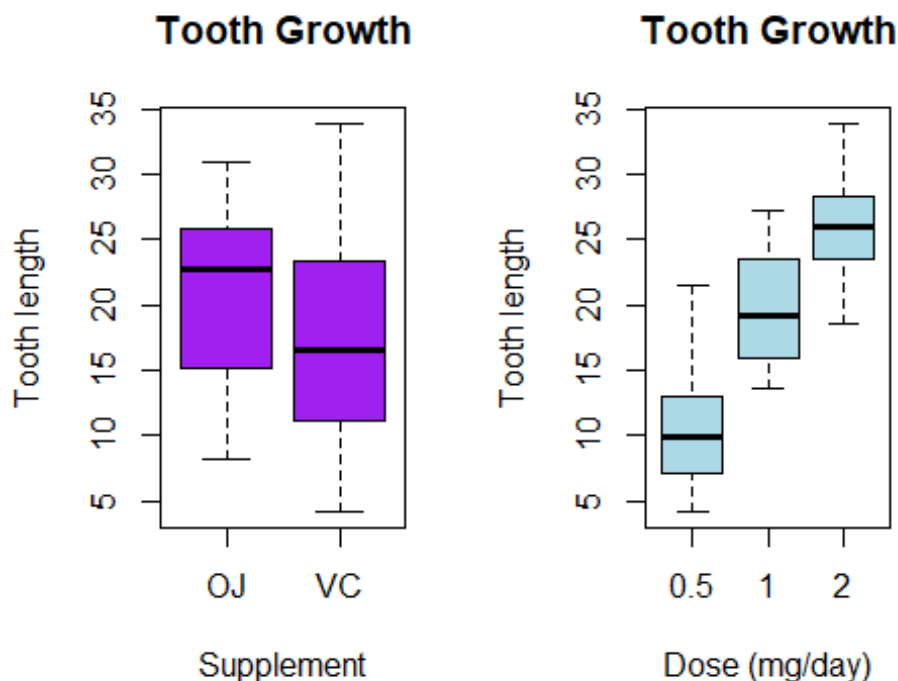
```
summary(ToothGrowth) #Summary statistics of each variable.
```

```
##      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
```

Basic exploratory data analyses

I used boxplots to visualize the summary statistics I got above.

```
par(mfrow=c(1,2))
boxplot(len~supp, data=ToothGrowth, col="purple", main= "Tooth Growth", xlab= "Supplement", ylab= "Tooth length") ##Length based on supplement type
boxplot(len~dose, data=ToothGrowth, col="lightblue", main= "Tooth Growth", xlab= "Dose (mg/day)", ylab= "Tooth length") ##Length based on dose level
```



Confidence intervals and hypothesis testing using t-test

Compare tooth growth by supplement

Null hypothesis: There is no difference between the two types of supplement (orange juice (OJ) and ascorbic acid(VC))

```
t.test(len~supp, data=ToothGrowth)

##
##  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
```

The p-value from the t-test is 0.06063, which is GREATER than our alpha value of 0.05. Also, the 95% confidence interval contains zero. We fail to reject the null hypothesis, and we can say that there is no significant difference on tooth growth based on supplement types.

Compare tooth growth by different doses

Null hypothesis 1: There is no difference between 0.5mg/day and 1.0mg/day.

```
dose_0.5_1.0 <- subset(ToothGrowth, dose %in% c(0.5,1.0))
t.test(len~dose, data=dose_0.5_1.0)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.983781 -6.276219
## sample estimates:
## mean in group 0.5 mean in group 1
## 10.605 19.735
```

Null hypothesis 2: There is no difference between 0.5mg/day and 2.0mg/day.

```
dose_0.5_2.0 <- subset(ToothGrowth, dose %in% c(0.5,2.0))
t.test(len~dose, data=dose_0.5_2.0)

##
## Welch Two Sample t-test
##
## data: len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -18.15617 -12.83383
## sample estimates:
## mean in group 0.5 mean in group 2
## 10.605 26.100
```

Null hypothesis 3: There is no difference between 1.0mg/day and 2.0mg/day.

```
dose_1.0_2.0 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len~dose, data=dose_1.0_2.0)

##
## Welch Two Sample t-test
##
## data: len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
## 19.735 26.100
```

For all 3 hypothesis tests, the p-value from the t-test is way smaller than the alpha value of 0.05. The confidence interval does not include zero. Hence, we can reject the null hypothesis and state that there is a significant difference on dose level of 0.5mg/day and 1.0mg/day, 0.5mg/day and 2.0mg/day, 1.0mg/day and 2.0mg/day on tooth growth.

Conclusions and assumptions

We assume that samples are randomly and independently sampled from the population, and the distribution is normal. Based on my data analysis, I can conclude that the supplement type (either OJ or AC) does not affect the tooth growth. However, the dose level (0.5, 1.0 and 2.0mg/day) did show significant effect on tooth growth.