Statistical Inference Course Project - Part 2

Mounika Gopinath

April 11, 2020

Overview

In this project, we will be analyzing the ToothGrowth data from the R datasets package, through the following steps:

1. Load the ToothGrowh data and do some basic exploratory data analyses
2. Summarize the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
4. State conclusions and assumptions needed for the conclusions.

Description of Data

The ToothGrowh data table contains 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). (decription from https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html)

Analysis

Let�s load the ToothGrowth data and take a high-level look at it.

```
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
```

We can see that ToothGrowth has 3 columns: len, supp (either OJ or VC), and dose (either 0.5, 1, or 2), and there are 60 observations. We can also plot the data to get an overview of the relationship between tooth length and dose, split up by the two delivery methods.
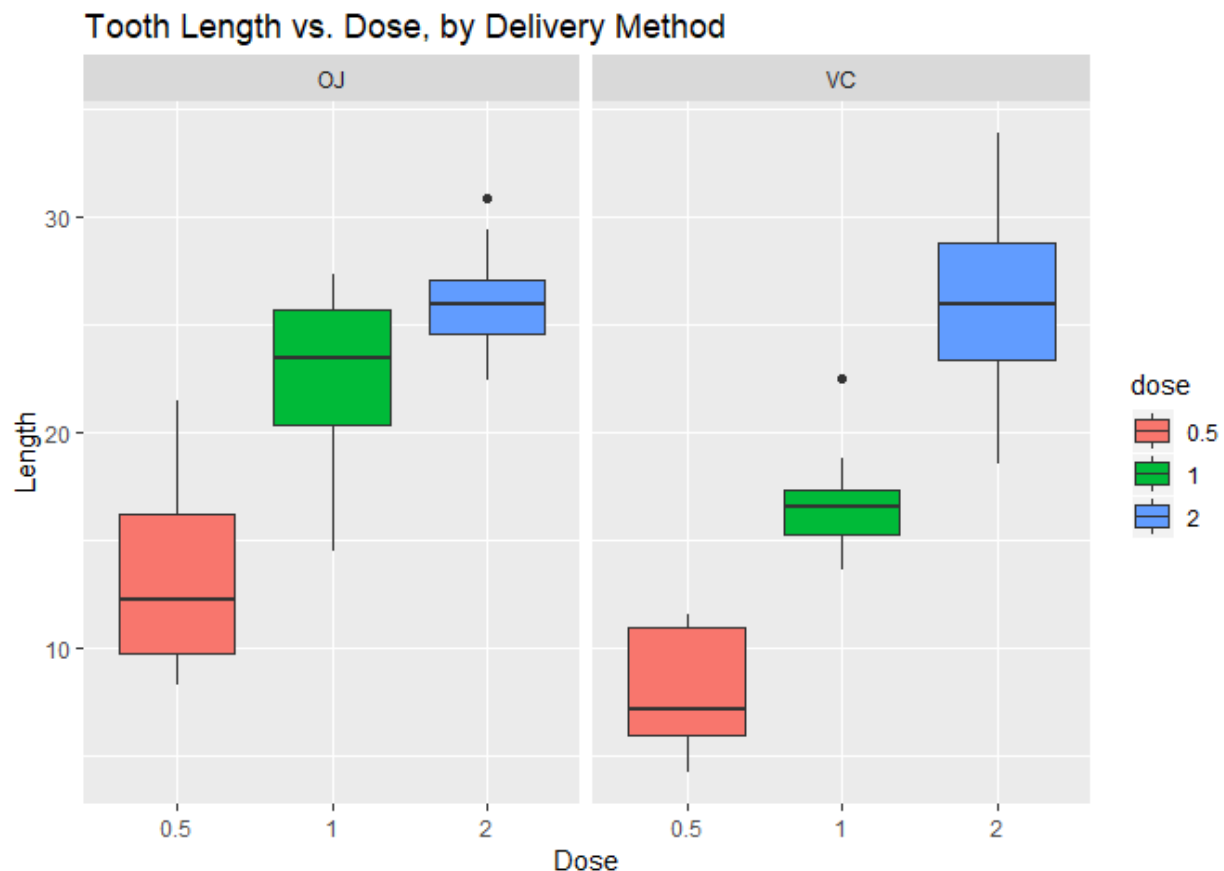
```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.2

ToothGrowth$dose = as.factor(ToothGrowth$dose)

g <- ggplot(aes(x=dose, y=len), data=ToothGrowth)

g <- g + geom_boxplot(aes(fill=dose)) + xlab("Dose") + ylab("Length") + facet
_grid(~ supp) + ggtitle("Tooth Length vs. Dose, by Delivery Method")

g
```


Basic Data Summary

Based on the box plot, it appears that higher doses result in higher levels of tooth growth, regardless of the delivery type. If the dose is 0.5 or 1, the OJ is more effective than VC (moreso for dose = 1). When dose = 2, the reulsting growth is very similar, on average.

Comparison of Tooth Growth by Supp and Dose (Using Confidence Intervals and/or Hypothesis Tests)

We will test a few hypotheses.

Hypothesis 1 - Tooth Growth is independent of supp (i.e., OJ and VC are equally effective)

```
t = ToothGrowth
hyp1 <- t.test(len ~ supp, data = t)
hyp1$conf.int
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
hyp1$p.value
## [1] 0.06063451
```

Since the interval includes zero and the p-value is greater than our threshold of 0.05, the null hypothesis cannot be rejected.

Hypothesis 2 - For dose = 2, OJ and VC result in the same tooth growth.

```
hyp2 <- t.test(len ~ supp, data = subset(t, dose == 2))
hyp2$conf.int
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
hyp2$p.value
## [1] 0.9638516
```

Again, we can see that the interval includes 0 and the p-value is greater than 0.05, and so the null hypothesis cannot be rejected. This is consistent with the boxplot above.

Conclusions

We can see that OJ results in higher tooth growth for dose = 0.5 and dose = 1. For dose = 2, OJ and VC result in about the same amount of tooth growth. Thus for the data as a whole, we can conclude that supp alone does NOT affect the mean length. Also, dose level has a strong effect and length goes up by a significant amount as dose increases. The assumptions here are that the data is normally distributed and the sample data is representative of the total popuation.