

Data Analysis using ToothGrowth dataset

Dinh Tuan Phan

4/29/2021

Synopsis

In the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Introduction

The ToothGrowth Dataset describes the response of the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid). The data contain 60 observations on 3 variables:

- len numeric Tooth length,
- supp factor Supplement type (VC or OJ),
- dose numeric Dose in milligrams.

In this project, we will use the hypothesis tests to compare tooth growth by supplement and dose.

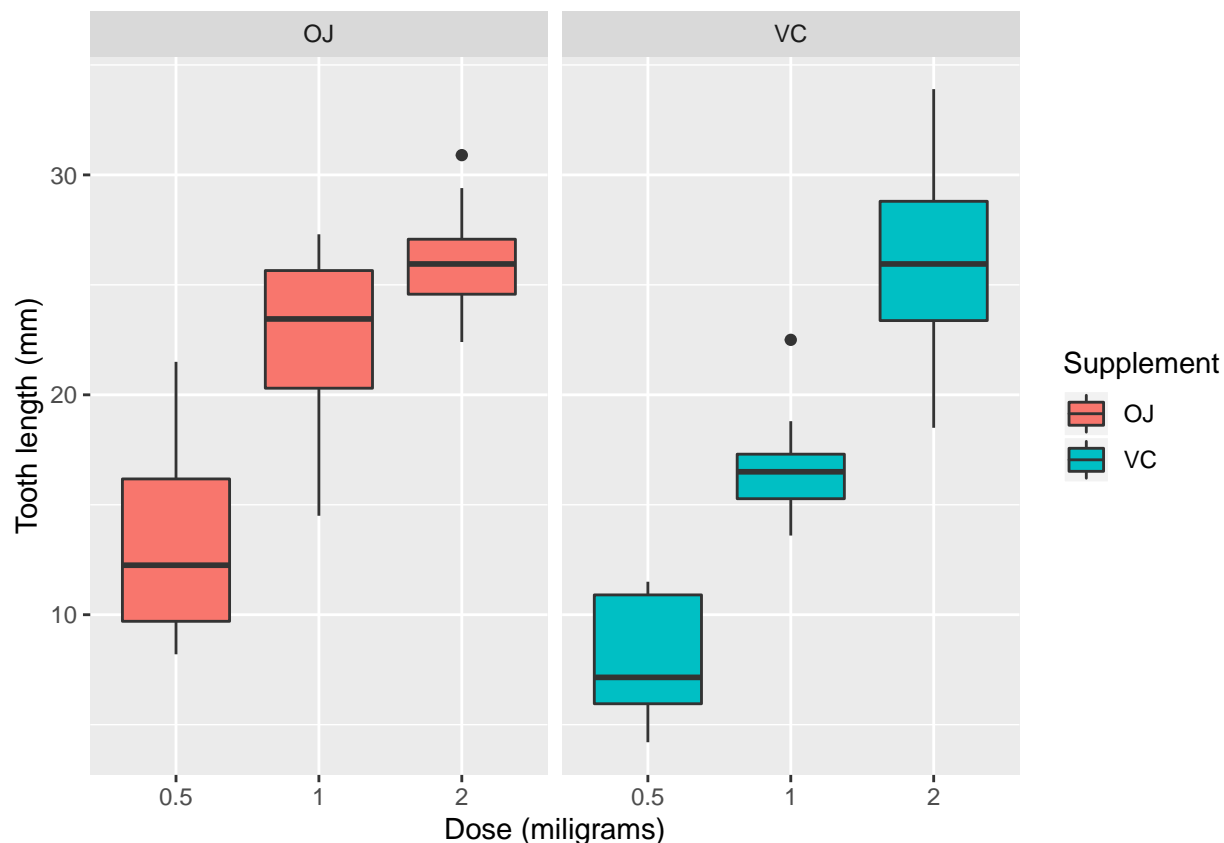
Loading packages

```
library("data.table")
library("ggplot2")
library("datasets")
library("dplyr")
```

Load data

We visualize the data to get the properties of the parameters.

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_boxplot() +
  facet_grid(. ~ supp) +
  xlab("Dose (miligrams)") +
  ylab("Tooth length (mm)") +
  guides(fill=guide_legend(title="Supplement"))
```



The boxplot shows that the tooth length has positive correlation with the dosage of both supplements(VC/OJ).

```

ToothGrowth %>%
  group_by(supp,dose) %>%
  summarise(Q25th_len = quantile(len,0.25),
            Q50th_len = quantile(len,0.5),
            Q75th_len = quantile(len,0.75),
            avg_len = mean(len),
            sd_len = sd(len)) -> data

```

Summary data

```
head(data)
```

```

## # A tibble: 6 x 7
## # Groups:   supp [2]
##   supp  dose Q25th_len Q50th_len Q75th_len avg_len sd_len
##   <fct> <dbl>   <dbl>    <dbl>   <dbl>   <dbl> <dbl>
## 1 OJ    0.5      9.7      12.2     16.2    13.2   4.46
## 2 OJ    1       20.3     23.5     25.6    22.7   3.91
## 3 OJ    2       24.6     26.0     27.1    26.1   2.66
## 4 VC    0.5      5.95     7.15     10.9     7.98   2.75
## 5 VC    1       15.3     16.5     17.3    16.8   2.52
## 6 VC    2       23.4     26.0     28.8    26.1   4.80

```

We can study the relations between len and supp and dose by using t hypothesis test. The following table shows the t-test results combined with p-value and confidence intervals.

```

tests = list()
dose = c(0.5,1,2)
for (d in dose) {
  ojd = ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == "OJ"]
  vcd = ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == "VC"]
  t <- t.test(ojd, vcd)
  id <- paste0("OJ","-", "VC",",",d)
  tests <- rbind(tests, list(id=id, p.value=t$p.value, conf.lo=t$conf.int[1], conf.hi=t$conf.int[2]))
}

```

Summary data

```
head(tests)
```

##	id	p.value	conf.lo	conf.hi
## [1,]	"OJ-VC,0.5"	0.006358607	1.719057	8.780943
## [2,]	"OJ-VC,1"	0.001038376	2.802148	9.057852
## [3,]	"OJ-VC,2"	0.9638516	-3.79807	3.63807

Conclusions

From the t-test results, we need to reject the following hypotheses:

- The difference of mean-value between OJ and VC in dose 0.5 is zero.
- The difference of mean-value between OJ and VC in dose 1.0 is zero.

We cannot distinguish the difference between OJ and VC in dose 2. Therefore, we can conclude that in dose 0.5 and dose 1, the tooth length with OJ supplement is longer than that with VC. However, in dose 2, the tooth length with OJ and VC supplements does not have significant difference.