Statistical Inference - Course Project Part 2

Eduardo Alves

August 09, 2019

Overview

This project analyze the ToothGrowth data in the R datasets package. Per the course project instructions, the following items should occur:

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- 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)
- State your conclusions and the assumptions needed for your conclusions.

Loading necessary R packages

```
if(!require(ggplot2)){install.packages('ggplot2')}; library(ggplot2)
```

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

```
if(!require(datasets)){install.packages('datasets')}; library(datasets)
```

Load the ToothGrowth data and perform some basic exploratory data analyses

The dataset ToothGrowth is about 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).

loading dataset...

```
data(ToothGrowth)
```

We can see below that the dataset consists of 60 observations of 3 types of variables:

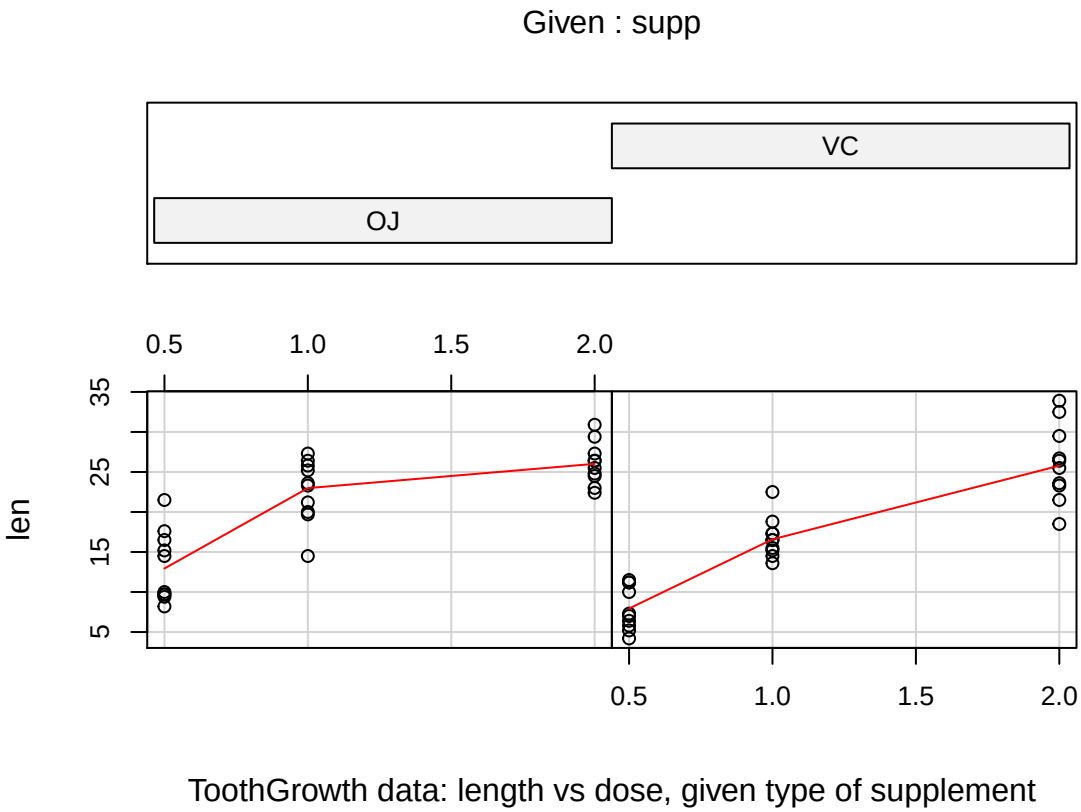
1. len -> Tooth length
2. supp -> Supplement type (VC or OJ).
3. dose -> Dose in milligrams/day

```
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 ...
```

Below is an exploratory analysis crossing information on tooth size versus dosage applied by type of supplement.

```
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



Provide a basic summary of the data.

As requested, below we have a basic summary of ToothGrowth dataset.

```
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
```

As we can see that in each supplement subgroup there are 30 guinea pigs, the mean length of the odontoblasts is 18.81 µm, also presenting an mean dose of 1,167 mg/day.

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

1st Hypothesis

The both supplements, Vitamin C and Orange Juice, deliver the same tooth growth across the dataset.

```
hypo1 <- t.test(len ~ supp, data = ToothGrowth)
hypo1$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
hypo1$p.value
```

```
## [1] 0.06063451
```

The confidence interval includes 0 and p-value is greater than 0.05 threshold.

The null hypothesis cannot be rejected.

2nd Hypothesis

For the dose of 0.5 mg/day, the 2 supplements deliver the same tooth growth:

```
hypo2 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 0.5))
hypo2$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
hypo2$p.value
```

```
## [1] 0.006358607
```

The confidence interval doesn't include 0 and p-value is smaller than 0.05 threshold.

The null hypothesis can be rejected.

3nd Hypothesis

For the dose of 1 mg/day, the 2 supplements deliver the same tooth growth:

```r
hypo3 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 1))
hypo3$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```r
hypo3$p.value
```

```
## [1] 0.001038376
```

The confidence interval doesn't include 0 and p-value is smaller than 0.05 threshold.

The null hypothesis can be rejected.

4th Hypothesis

For the dose of 2 mg/day, the 2 supplements deliver the same tooth growth:

```r
hypo4 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 2))
hypo4$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```r
hypo4$p.value
```

```
## [1] 0.9638516
```

The confidence interval doesn't include 0 and p-value is greater than 0.05 threshold.

The null hypothesis cannot be rejected.

Assumptions and Conclusions

Assumptions

- Null hypotheses assumed equal length means. Whereas alternative hypotheses assumed not equal length means.
- The experiment was carried out on 60 guinea pigs divided in 6 groups of 10's.
  - 0.5 mg of VC to 10 pigs and 0.5 mg of OJ to another group of 10 pigs
  - 1.0 mg of VC to 10 pigs and 1.0 mg of OJ to another group of 10 pigs
  - 2.0 mg of VC to 10 pigs and 2.0 mg of OJ to another group of 10 pigs
- Doses are given to pigs for equal periods of time enough to compare their odontoblasts lengths fairly.

Conclusions

- Dosages 0.5 mg/day and 1.0 mg/day of OJ delivers more tooth growth than VC.
- Dosage 2.0 mg/day of VC delivers more tooth growth than OJ.
- We can conclude that the higher of VC dose, the greater the tooth growth.