Statistical Inference Project. Part 2

Diego Andres Benitez

24/1/2021

Contents

1 Overview **1**

2 Tasks **1**

3 Analysis **1**
3.1 **Summary of the data** 1
3.2 **Exploratory data analysis** 2
3.3 **Hypothesis Testing** 3

4 Conclusions **4**

5 Appendix **4**

1 Overview

This is the second part of the Statistical Inference Course Project from Coursera. In this part, a sample dataset about Tooth Growth is being investigated with exploratory data analysis. The ToolthGrowth dataset contains observation for the effect of vitamin C on tooth growth in Guinea Pigs, which corresponds to the response is 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).

2 Tasks

Now 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.

3 Analysis

3.1 Summary of the data

Let's load the data and we can peek at the top rows of the data, examine the structure of the datase and provide a quick summary.

```
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 ...
```

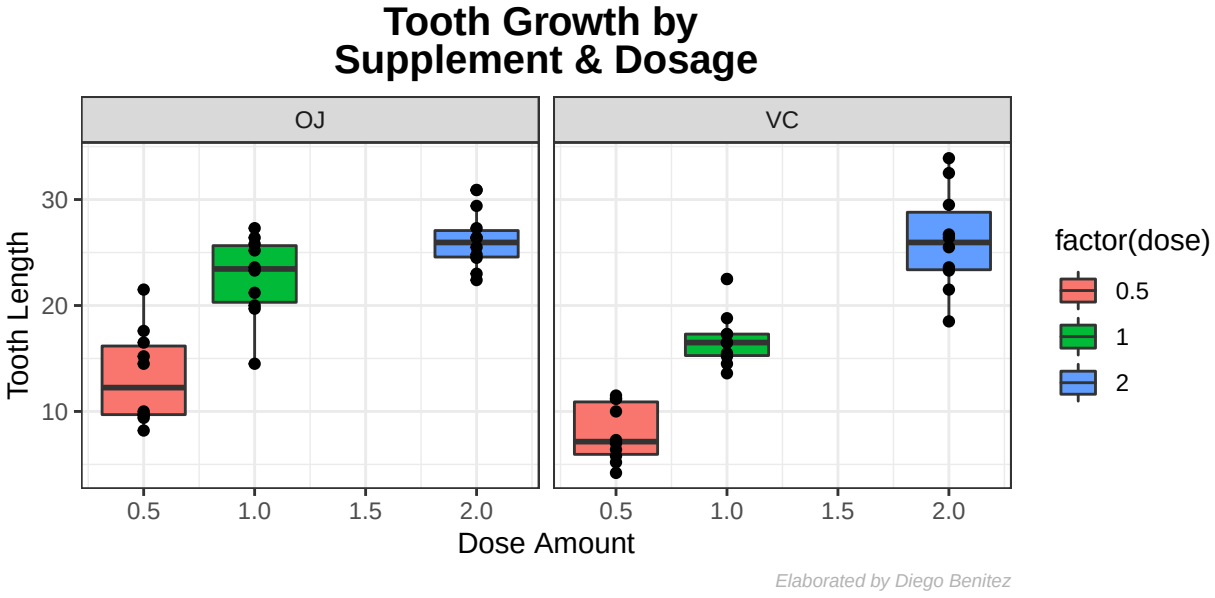
```
summary(ToothGrowth)
```

```
##       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
```

From the summary, we know that 30 observations are related to orange juice (supp = "OJ"), and other 30 are with ascorbic acid (supp = "VC"). Since the dose levels are fixed at 0.5, or 1, or 2 mg level, the median and mean or even quantile values do not carry much meaning in our analysis, we should focus on the length observation in the dataset, which the summary shows the values are between 4.20 and 33.90 with mean value at 28.81 and median at 19.25.

3.2 Exploratory data analysis

Graphically, we can also use boxplot to provide a quick visual on the impact of dosage and supplement on the tooth growth, see below:



Based on the above chart, it appears that when the dosage is high at 2 mg, the mean value of tooth growth appears to be similar between OJ and VC, however, when the dosage is 0.5 mg or 1 mg, the chart definitely shows that OJ has a obvious positive impact on tooth growth compared to VC.

3.3 Hypothesis Testing

Let's look at the below two hypothesis:

- H_0 : "Orange juice does not have an impact on tooth growth compared to ascorbic acid"
- H_1 : "Orange juice has an impact on tooth growth compared to ascorbic acid.

```
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 of this test was 0.0606345. Since the p-value is greater than 0.05 and the confidence interval of the test contains zero we can say that supplement types seems to have no impact on Tooth growth based on this test.

Now we'll compare tooth growth by dose, looking at the different pairs of dose values.

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5))
t.test(len~dose,data=ToothGrowth_sub)
```

```
##
##  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
```

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))
t.test(len~dose,data=ToothGrowth_sub)
```

```
##
##  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
```

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

```
##
##  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
```

As can be seen, the p-value of each test was essentially zero and the confidence interval of each test does not cross over zero (0). Based on this result we can assume that the average tooth length increases with an inceasing dose, and therefore the null hypothesis can be rejected.

4 Conclusions

Given the following assumptions:

1. The sample is representative of the population
2. The distribution of the sample means follows the Central Limit Theorem

In reviewing our t-test analysis from above, we can conclude that supplement delivery method has no effect on tooth growth/length, however increased dosages do result in increased tooth length.

5 Appendix

The Rmd code, generated by `knitr`, are available clicking here.