Basic inferential data analysis

Georgy Makarov

April 24, 2020

Overview

This is the part two of the course project of Statistical Inference course. Here we use confidence intervals and hypothesis tests to compare tooth growth by supplement and dose from the *ToothGrowth* data. There are supporting plots in appendix to this project.

Summary of the data

ToothGrowth dataset shows the length of cells responsible for tooth growth in 60 observations. There are three features in the dataset: *len* stands for the tooth length; *supp* is a factor with supplement *VC* or *OJ*; *dose* shows the dose of the supplement received.

```
data("ToothGrowth")
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
```

Dose feature has three levels: 0.5, 1.0, 2.0. Therefore we consider it a factor variable.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1   2
## Levels: 0.5 1 2
```

The distribution of the tooth length is slightly skewed to the left, as the mean is lower than the median. The histogram of the tooth length is in appendix in *Figure 1*.

While the dose is 2.0 mg there is no difference in the impact of delivery method on the tooth growth. When the dose is either 0.5 or 1.0 mg the orange juice is more effective for the tooth growth. Visualization of this idea is given in appendix in *Figure 2*.

Hypothesis testing

We test three hypothesis. First hypothesis: tooth growth does not depend on application method. Second hypothesis: tooth growth depends on application method when the dose is small. Third hypothesis: tooth growth does not depend on the dose of vitamin C. These hypothesis appear through H_0 notation:

1. H_0: OJ and VC means are the same;
2. H_0: OJ and VC means are the same when dose is lower than 2;
3. H_0: 0.5 and 2.0 doses means are the same.

Hypothesis 1

```
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
```

P-value is **0.06** and it is greater than 0.05. Confidence interval includes 0. There is no evidence to reject H_0 - application methods do not impact tooth growth.

Hypothesis 2

```
t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose %in% c(0.5, 1.0),])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.875234 9.304766
## sample estimates:
## mean in group OJ mean in group VC
##           17.965           12.375
```

P-value is **0.004**, which is less than 0.05. Confidence interval does not contain 0. There is evidence enough to reject H_0 - application methods impact tooth growth when the dose is less than 2.0 mg.

Hypothesis 3

```
t.test(len ~ dose, data = ToothGrowth[ToothGrowth$dose %in% c(0.5, 2),])
```

```
##
##  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
```

P-value is **4.4e-14**. which is less than 0.05. Confidence interval does not contain 0. There is evidence enough to reject H_0 - the dose impacts on tooth growth.

Conclusions

This analysis assumed that each subject was randomly assigned and it was a representative of the population. It is also based on the assumption that all observations are independent. The analysis shows that the tooth growth depends on the dose of vitamin C. While the method of application is critical when the dose is below 2.0 mg, it does not imply any difference on tooth growth when the dose icreases.

Appendix

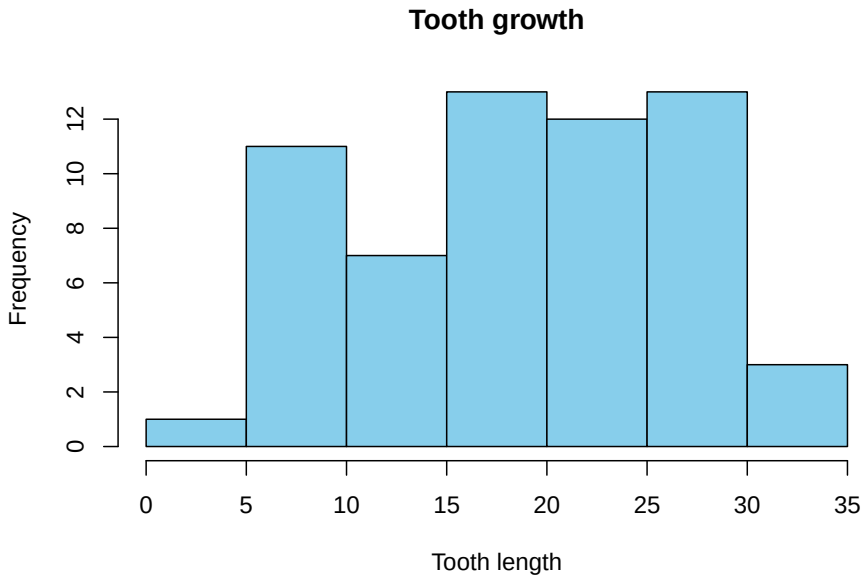

Figure 1. Distribution of tooth length

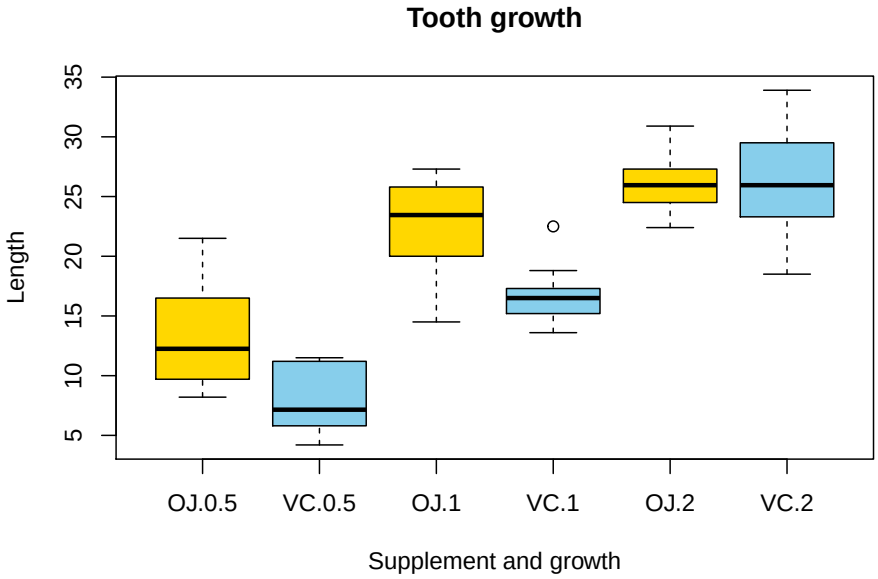

Figure 2. Tooth growth vs supplement & dose