Part 2: Basic Inferential Data Analysis

Mukesh Saxena

30/06/2018

Overview

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

Loading and basic exploratory analysis

The first step is to load the dataset package and the ToothGrowth data.

```
#  loading the datasets package
library(datasets)
#  loading th ToothGrowth data
df <- ToothGrowth
```

From the help command ?ToothGrowth, it is shown that the data are about The Effect of Vitamin C on Tooth Growth in Guinea Pigs. It report the response is 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).

Taking a look at the dataframe :

```
#  some informations about the data
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 are going to perform a serie of t.test analysis in order to compare the two delivery methods and then the three dose levels. For each test we are going to assume a null hypothesis and an alternative one.

For all tests, we are going to use some common assumptions which are:

- One sided test.
- Samples are not paired
- Non equal variances
- Confidence level of 95%.

Delivery method

In the first step, we are going to examen the diffrence between tooth growth associated to the delivery methods: orange juice and ascorbic acid. Considering as a null hypothesis that the ascorbic acid is more peroformant or there is no diffrence between the two methods mean(VC) >= mean(OJ). The alternative hypothesis is then that the orange juce performant in terms of tooth growth mean(VC) < mean(OJ).

```r
# performing a t-test and looking at confidence interval
t.test(df[df$supp == "VC", "len"], df[df$supp == "OJ", "len"], alternative="less")$conf
```

```
## [1]         -Inf -0.4682687
#    attr(,"conf.level")
#    [1] 0.95
```

```r
# looking at the p value
t.test(df[df$supp == "VC", "len"], df[df$supp == "OJ", "len"], alternative="less")$p.value
```

```
## [1] 0.03031725
```

The confidence interval did not include 0 and the p-value is equal to 0.03 so less than 0.05. We can reject the null hypothesis in favor to the alternative one.

Dose levels.

We are going to compare now the diffrence between teeth growth associated to each dose levels in order to have an idea which level is the most performant in terms of tooth growth. Since we are going to perform a multiple comparaison (3 test), we have to use the Bonferroni Correction. Then ower new alpha is alpha = 0.05/3 ~ 0.017 which is used to decide for the significant p-values.

- 0.5 mg/day

Considering a null hypothesis assuming that a 0.5 mg/day dose is less performant than 1 mg/day dose and 2 mg/day dose or they are equivalent mean(0.5 mg/day) <= mean(1 mg/day or 2 mg/day). So, the alternative hypotesis suppose that mean(0.5 mg/day) > mean(1 mg/day or 2 mg/day).

```r
# performing a t-test and looking at confidence interval
t.test(df[df$dose != 0.5, "len"], df[df$dose == 0.5, "len"], alternative="less")$conf
```

```
#    [1]-Inf 14.49309
#    attr(,"conf.level")
#    [1] 0.95
```

```r
# looking at the p-value
t.test(df[df$dose != 0.5, "len"], df[df$dose == 0.5, "len"], alternative="less")$p.value
```

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

The confidence interval include 0 and the p-value is equal to 1. We can't reject the null hypothesis in favor of the alternative one.

- 1 mg/day

Considering a null hypothesis assuming that a 1 mg/day dose is less performant than 0.5 mg/day dose and 2 mg/day dose or they are equivalent mean(1 mg/day) <= mean(0.5 mg/day or 2 mg/day). So, the alternative hypotesis suppose that mean(1 mg/day) > mean(0.5 mg/day or 2 mg/day).

```r
# performing a t-test and looking at confidence interval
t.test(df[df$dose != 1, "len"], df[df$dose == 1, "len"], alternative="less")$conf
```

```
#    [1]-Inf 1.48068
#    attr(,"conf.level")
#    [1] 0.95
```

```r
# looking at the p-value
t.test(df[df$dose != 1, "len"], df[df$dose == 1, "len"], alternative="less")$p.value
```

```
## [1] 0.2114481
```

The confidence interval include 0 and the p-value is equal to 0.211. We can't reject the null hypothesis in favor of the alternative one.

- 2 mg/day

Considering a null hypothesis assuming that a 2 mg/day dose is less performant than 0.5 mg/day dose and 1 mg/day dose or they are equivalent mean(2 mg/day) <= mean(0.5 mg/day or 1 mg/day). So, the alternative hypotesis suppose that mean(2 mg/day) > mean(0.5 mg/day or 1 mg/day).

```
# performing a t-test and looking at confidence interval
t.test(df[df$dose != 2, "len"], df[df$dose == 2, "len"], alternative="less")$conf
```

```
## [1]           -Inf -8.729885
#   attr(,"conf.level")
#   [1] 0.95
```

```
# looking at the p-value
t.test(df[df$dose != 2, "len"], df[df$dose == 2, "len"], alternative="less")$p.value
```

```
## [1] 1.173268e-11
```

The confidence interval did not include 0 and the p-value is equal to 1.17e-11 < alpha=0.017. We can reject the null hypothesis in favor of the alternative one.

Conclusion.

From the serie of conducted text, we have mainly two conclusions :

- The orange juce is more efficient in terms of tooth growth compared to the ascorbic acid.
- The most efficient dose level is 2 mg/day compared to 0.5 mg/day and 1 mg/day. We did 3 tests to compare the diffrent dose levels. For each one, we compare the p value to the alpha from using the Bonferroni Correction which is very concervative. We can so can assume that ower result is significant.