Part 2: Basic Inferential Data Analysis [Peer-graded Assignment: Statistical Inference Course Project]

Dr.Chamika Senanayake

9/26/2020

the part 2 of the projects are executed according to the insturctions given by

Load the ToothGrowth data and perform some basic exploratory data analyses

All required metadata were loaded as following

```
set.seed(1)
library(ggplot2, warn.conflicts = FALSE)
library(reshape2, warn.conflicts = FALSE)
```

Provide a basic summary of the data.

then started by loading the dataset, followed by a basic exploration of how the data is structured.

```
data(ToothGrowth)
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 ...
```

```
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

finally the summary can be given as follows

```
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
```

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)

Data Processing

In this dataset, rows 1 to 10 correspond to the observations from Guinea pigs 1 to 10. The eleventh row contains another observation of Guinea pig number 1, and so on. In order to tidy this dataset, we are going to fit it into 6 columns and 10 rows. Each row corresponds to one specific Guinea pig, and each column represents a supplement (VC or OJ) and its associated dose (0.5, 1.0 or 2.0).

```
ToothGrowth$guinea.pig <- rep(c(1:10), 6)
print(tidy <- dcast(ToothGrowth, guinea.pig ~ supp + dose, value.var = "len"))
```

```
##    guinea.pig OJ_0.5 OJ_1 OJ_2 VC_0.5 VC_1 VC_2
## 1           1   15.2 19.7 25.5    4.2 16.5 23.6
## 2           2   21.5 23.3 26.4   11.5 16.5 18.5
## 3           3   17.6 23.6 22.4    7.3 15.2 33.9
## 4           4    9.7 26.4 24.5    5.8 17.3 25.5
## 5           5   14.5 20.0 24.8    6.4 22.5 26.4
## 6           6   10.0 25.2 30.9   10.0 17.3 32.5
## 7           7    8.2 25.8 26.4   11.2 13.6 26.7
## 8           8    9.4 21.2 27.3   11.2 14.5 21.5
## 9           9   16.5 14.5 29.4    5.2 18.8 23.3
## 10         10    9.7 27.3 23.0    7.0 15.5 29.5
```

Data Analysis

Data Analysis

In this section, we are going to perform several tests to compare tooth growth of Guinea pigs by supplement (OJ, VC) and dose (0.5, 1.0, and 2.0 milligrams). For every comparation, we will consider the following hypotheses:

- H_0: For a given dose level of Vitamin C, **there is no difference in average tooth growth** when we compare the delivery method (Orange Juice or Ascorbic Acid).
- H_a: For a given dose level of Vitamin C, **there is an actual difference in average tooth growth** when we compare the delivery method (Orange Juice or Ascorbic Acid).

The `t.test` function requires two other parameters: `var.equal` and `paired`. The first parameter indicates whether to treat the two variances as being equal. If TRUE then the pooled variance is used to estimate the variance otherwise the Welch (or Satterthwaite) approximation to the degrees of freedom is used. The dataset provides no information on this subject, hence we consider the safer choice and take the equality as FALSE. The second parameter is set as FALSE, due to the description of the dataset, which states that there were 60 distinct Guinea pigs in the experiment.

1) 0.5 milligrams

```
tres_0.5 <- t.test(tidy$OJ_0.5, tidy$VC_0.5, paired = FALSE, var.equal = FALSE)
tres_0.5$conf.int
```

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

Based on `t.test` results, we decide to **reject**, with a 95% confidence interval, the null hypothesis H_0. The p-value of 0.0063586 indicates a low probability that OJ and VC equally promote the same average tooth growth when the dosage is 0.5 mg.

2) 1.0 milligrams

```
tres_1 <- t.test(tidy$OJ_1, tidy$VC_1, paired = FALSE, var.equal = FALSE)
tres_1$conf.int
```

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

Similarly to the previous result, we decide to **reject**, with a 95% confidence interval, the null hypothesis H_0. The p-value of 0.0010384 indicates a low probability that OJ and VC equally promote the same average tooth growth when the dosage is 1.0 mg.

3) 2.0 milligrams

```
tres_2 <- t.test(tidy$OJ_2, tidy$VC_2, paired = FALSE, var.equal = FALSE)
tres_2$conf.int
```

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

This time, we decide to **accept**, with a 95% confidence interval, the null hypothesis H_0. The p-value of 0.9638516 indicates a high probability that OJ and VC equally promote the same average tooth growth when the dosage is 2.0 mg.

State your conclusions and the assumptions needed for your conclusions.

In this report, we compared the effectiveness of tooth growth in Guinea pigs for each combination of dose level of Vitamin C (0.5, 1, and 2 mg) and delivery method (OJ or VC). We executed three `t.tests`, one for each dose, with a 95% confidence interval. As the dataset contains data of 60 distinct Guinea pigs, the sample is not paired. Additionally, we considered the two variances as not being equals. Summing up, lower and intermediate doses of Orange Juice are more effective than Vitamin C. With a 2.0 mg dose, there is a high probability that these delivery methods are equally effective.