

Peer-graded Assignment: Statistical Inference Course Project

Irina Z

9/22/2020

Part 2: Basic inferential data analysis

Synopsis

In this report, I provide the analysis of the **ToothGrowth data** in the R datasets package. ToothGroth data shows **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).

Question 1 : Load the ToothGrowth data and provide a basic summary of the data

```
# Reading the data
ToothGrowth <- data.table(ToothGrowth)
# Looking at the structure of the data
str(ToothGrowth)
```

```
## Classes 'data.table' and '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 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

Dataset has **60 observations** and **3 variables**:

- len - Tooth length
- supp - Supplement type (ascorbic acid or orange juice)
- dose - Dose in milligrams/day

```
# Summaries of the variables
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
```

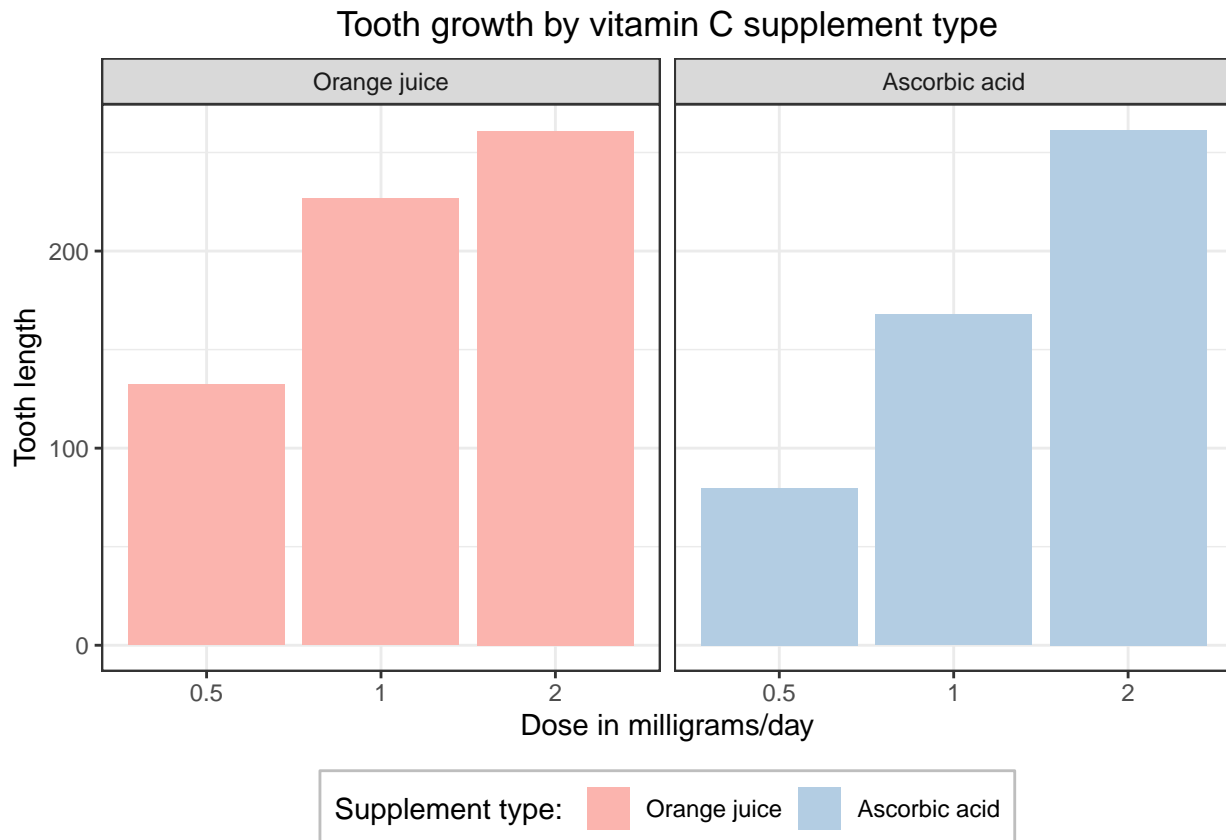
We can make the following statements:

- *Tooth length* varies from 4.2 to 33.9
- Half of the guinea pigs received vitamin C in the form of orange juice (OJ), while another half of the guinea pigs received in the form of ascorbic acid (VC)
- *Dosage* varies from 0.5 mg/day to 2 mg/day

Question 2 : Perform some basic exploratory data analyses

First, I show how tooth growth varies by the dose of vitamin C consumed. The boxplot is shown in the Appendix (Picture 1). We can see that **the higher dose of vitamin C is associated with longer teeth**.

However, we know there are two types of vitamin C supplements, and different types of supplements may have different trends. To see how teeth length varies by vitamin C dose for different supplement types, I create the following graph (code for this graph is presented in the Appendix, Code 1):



We can see that for both supplement types **tooth growth increases with the increase of dosage**. However, **lower dosages (0.5 and 1 mg/day) of vitamin C via orange juice is associated with longer teeth than same dosage of vitamin C via ascorbic acid**.

Question 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)

First, I test the hypothesis that higher dose of vitamin C results in longer teeth.

To start, I **compare 0.5 mg/day and 1 mg/day doses** of vitamin C. Null hypothesis suggests that there is no difference between the two groups of guinea pigs, and the alternative hypothesis is that guinea pigs that received 1 mg/day of vitamin C had longer teeth compared to those who received 0.5 mg/day of vitamin C.

```
t1 <- t.test(ToothGrowth[dose == 1]$len, ToothGrowth[dose == 0.5]$len,  
            paired = FALSE,  
            alternative = "greater")$p.value
```

P-value is equal to 6.3415036×10^{-8} , and if we compare it to 0.05, we get $6.3415036 \times 10^{-8} < 0.05 = \text{TRUE}$. Which means that the result is statistically significant, and teeth were longer for guinea pigs who received 1 mg/day of vitamin C compared to those who received 0.5 mg/day of vitamin C.

Now I **compare 1 mg/day and 2 mg/day doses** of vitamin C. Similarly, null hypothesis suggests that there is no difference between the two groups of guinea pigs, and the alternative hypothesis is that guinea pigs that received 2 mg/day of vitamin C had longer teeth compared to those who received 1 mg/day of vitamin C. The code for this part is presented in the Appendix (Code 2).

P-value is equal to 9.5321476×10^{-6} , and if we compare it to 0.05, we get $9.5321476 \times 10^{-6} < 0.05 = \text{TRUE}$. Which means that the result is statistically significant, and teeth were longer for guinea pigs who received 2 mg/day of vitamin C compared to those who received 1 mg/day of vitamin C.

Therefore, the hypothesis that **higher dose of vitamin C results in longer teeth** is accepted.

Second, I test the hypothesis that taking vitamin C supplements in form of orange juice gave better results compared to ascorbic acid.

To start, I **compare 0.5 mg/day dose for orange juice and for ascorbic acid**. Null hypothesis suggests that there is no difference between the two groups of guinea pigs, and the alternative hypothesis is that guinea pigs that received orange juice had longer teeth compared to those who received ascorbic acid.

```
t3 <- t.test(ToothGrowth[dose == 0.5 & supp == "OJ"]$len,  
            ToothGrowth[dose == 0.5 & supp == "VC"]$len,  
            paired = FALSE,  
            alternative = "greater")$p.value
```

P-value is equal to 0.0031793, and if we compare it to 0.05, we get $0.0031793 < 0.05 = \text{TRUE}$. Which means that the result is statistically significant, and teeth were longer for guinea pigs who received 0.5 mg/day of vitamin C via orange juice compared to those who received it via ascorbic acid.

Now I **compare 1 mg/day dose for orange juice and for ascorbic acid**. Again, null hypothesis suggests that there is no difference between the two groups of guinea pigs, and the alternative hypothesis is that guinea pigs that received orange juice had longer teeth compared to those who received ascorbic acid. The code for this part is presented in the Appendix (Code 3).

P-value is equal to 5.1918794×10^{-4} , and if we compare it to 0.05, we get $5.1918794 \times 10^{-4} < 0.05 = \text{TRUE}$. Which means that the result is statistically significant again, and teeth were longer for guinea pigs who received 1 mg/day of vitamin C via orange juice compared to those who received it via ascorbic acid.

Finally, I **compare 2 mg/day dose for orange juice and for ascorbic acid**. Again, null hypothesis suggests that there is no difference between the two groups of guinea pigs, and the alternative hypothesis is that guinea pigs that received orange juice had longer teeth compared to those who received ascorbic acid. The code for this part is presented in the Appendix (Code 4).

P-value is equal to 0.5180742, and if we compare it to 0.05, we get $0.5180742 > 0.05 = \text{FALSE}$. Which means that the result is not statistically significant, and there was no difference in the teeth length between the guinea pigs who received 2 mg/day of vitamin C via orange juice and those who received it via ascorbic acid.

Therefore, the hypothesis that **orange juice gives better results for the teeth growth compared to ascorbic acid** is not accepted.

Question 4 : State your conclusions and the assumptions needed for your conclusions

Based on the performed T-tests and the exploratory analysis of the data, we can conclude that the **dosage of the vitamin C supplements has a positive impact on tooth growth**, while the **kind of supplement (orange juice vs. ascorbic acid) does not have a clear impact on the teeth growth**.

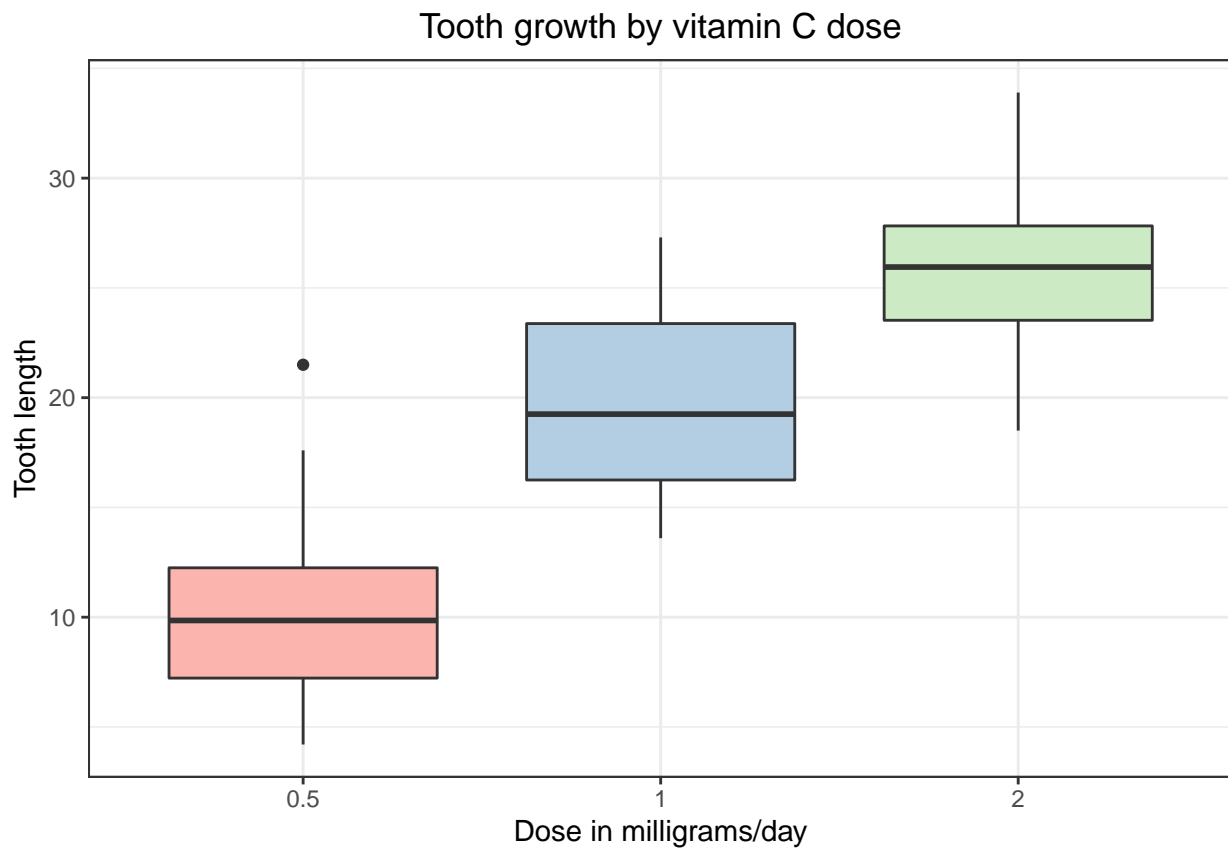
To derive these conclusions I had to impose the following **assumptions**:

- Guinea pigs tooth lengths are *normally distributed*
- Guinea pigs from the sample are *representative* of the entire population of guinea pigs

Appendix to Part 2

Picture 1. Tooth growth by vitamin C dose

```
ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = factor(dose))) +  
  geom_boxplot() +  
  theme_bw() +  
  labs(x = "Dose in milligrams/day", y = "Tooth length",  
        title = "Tooth growth by vitamin C dose") +  
  scale_fill_brewer(palette = "Pastel1") +  
  theme(legend.position = "none",  
        plot.title = element_text(hjust = 0.5))
```



Code 1. Comparing tooth growth by vitamin C supplement type

```
ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = supp)) +  
  geom_bar(stat = "identity") +  
  theme_bw() +  
  facet_grid(. ~ supp,  
             labeller = labeller(supp = c(`OJ` = "Orange juice",  
                                           `VC` = "Ascorbic acid")))) +  
  labs(x = "Dose in milligrams/day", y = "Tooth length",  
       title = "Tooth growth by vitamin C supplement type") +  
  scale_fill_brewer(palette = "Pastel1",  
                   name = "Supplement type: ",  
                   labels = c("Orange juice", "Ascorbic acid")) +  
  theme(legend.position = "bottom",  
        legend.background = element_rect(colour = 'grey'),  
        plot.title = element_text(hjust = 0.5))
```

Code 2. Compare 1 mg/day and 2 mg/day doses

```
t2 <- t.test(ToothGrowth[dose == 2]$len, ToothGrowth[dose == 1]$len,  
            paired = FALSE,  
            alternative = "greater")$p.value
```

Code 3. Compare 1 mg/day dose for orange juice and ascorbic acid

```
t4 <- t.test(ToothGrowth[dose == 1 & supp == "OJ"]$len,  
            ToothGrowth[dose == 1 & supp == "VC"]$len,  
            paired = FALSE,  
            alternative = "greater")$p.value
```

Code 4. Compare 2 mg/day dose for orange juice and ascorbic acid

```
t5 <- t.test(ToothGrowth[dose == 2 & supp == "OJ"]$len,  
            ToothGrowth[dose == 2 & supp == "VC"]$len,  
            paired = FALSE,  
            alternative = "greater")$p.value
```