Tooth Growth Analysis

Luis Talavera

July 26th 2022

Synopsis

The guinea pig, (Cavia porcellus), is a domesticated species of South American rodent belonging to the cavy family (Caviidae). In this analysis we will test if the vitamin C has effect on tooth growth in guinea pigs considering the dose received and delivery method used.

Explore data

For this analysis, we are going to use R included dataset "ToothGrowth", it contains data of 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).

First, the variables of our data are len, supp and dose.

```
'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 ...
```

Although dose is a numerical value, in this case, it will be useful to treat it as a categorical variable.

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

Here is a descriptive summary of data

```
      len          supp       dose
 Min.   : 4.20   OJ:30   0.5:20
 1st Qu.:13.07   VC:30   1  :20
 Median :19.25           2  :20
 Mean   :18.81
 3rd Qu.:25.27
 Max.   :33.90
```

Now, the data will be grouped, first by delivery method (supp), later by dose.

```
library(dplyr)
group.data <- ToothGrowth %>% group_by(supp,dose)
```
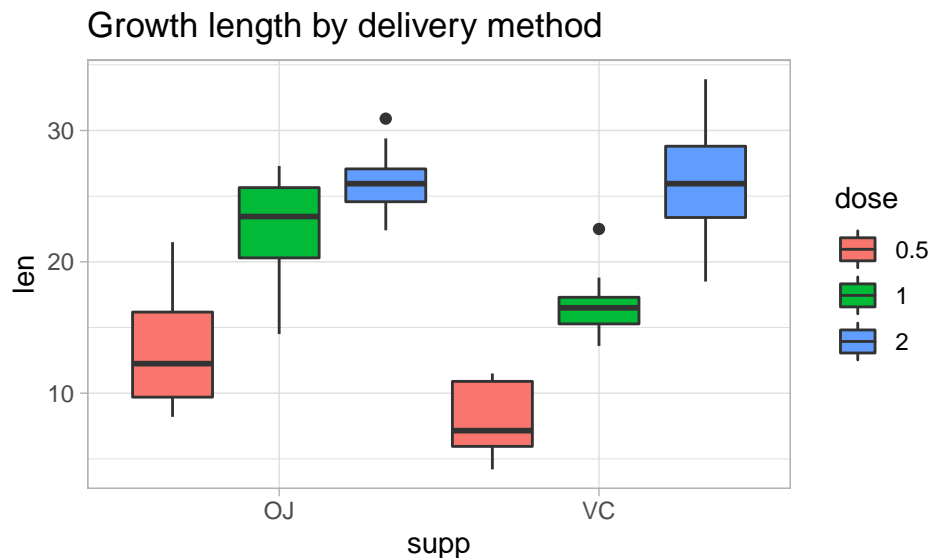
Figure 1: Boxplots of lengths divided by dose and grouped by delivery method

Analysis

From the figure above, we can observe that the doses received by orange juice have a longer growth length mean than those doses received by ascorbic acid. So our question is as follows: is there sufficient evidence at the $\alpha = 0.5$ level to conclude that the growth length mean of doses given by orange juice differs from the growth length mean of doses given by ascorbic acid?

Here is a descriptive summary of grouped data by delivery method.

```
# A tibble: 2 x 4
  supp      n  mean    sd
  <fct> <dbl> <dbl> <dbl>
1 OJ       30  20.7  6.61
2 VC       30  17.0  8.27
```

For this question we will assume the data are **plausibly normal** and that **the obervations are independent of each other** in order to be able to use two sample t test. And because the observed standard deviations of the two samples are of similar magnitude, we'll assume that the population variances are equal.

After those assumptions, we can test the null hypothesis

$$H_0 : \mu_{OJ} = \mu_{VC}$$

against the alternative hypothesis

$$H_a : \mu_{OJ} \neq \mu_{VC}$$

using the test statistic

$$t = \frac{(20.7 - 17.0) - 0}{7.61\sqrt{\frac{1}{30} + \frac{1}{30}}} = 1.88$$

The p-value is

$$P = 2 * P(T_58 > 1.88) = 2 * 0.03235 = 0.0647$$

2

Since our p-value is greater than our $\alpha = 0.5$ then we conclude to fail to reject the null hypothesis.

If we use t.test R function we can see the results are similar.

```
    Two Sample t-test

data:  oj and vc
t = 1.9153, df = 58, p-value = 0.06039
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1670064  7.5670064
sample estimates:
mean of x mean of y
 20.66333  16.96333
```

Appendix

Figure 1

```r
library(ggplot2)
ggplot(group.data, aes(x=supp, y=len, fill=dose)) +
  geom_boxplot(position=position_dodge(1)) +
  theme_light() +
  labs(title = "Growth length by delivery method")
```

t and p-value calculation

```r
temp <- ToothGrowth %>% select(len,supp) %>% group_by(supp)
data <- temp %>% summarise(n = 30, mean = mean(len), sd = sd(len)) %>% distinct()

n <- as.numeric(data[data$supp=="OJ","n"])
m <- as.numeric(data[data$supp=="VC","n"])
sx <- as.numeric(data[data$supp=="OJ","sd"])
sy <- as.numeric(data[data$supp=="VC","sd"])
mx <- as.numeric(data[data$supp=="OJ","mean"])
my <- as.numeric(data[data$supp=="VC","mean"])
sp <- as.numeric(sqrt((n*sx^2 +m*sy^2)/(n+m-2)))
t <- (mx-my)/(sp*sqrt(1/n+1/m))
p <- 2*pt(t,m+n-2,lower.tail = FALSE)
```

t.test

```r
supps.values <- ToothGrowth %>% group_by(supp)
oj <- (supps.values %>% filter(supp=="OJ"))$len
vc <- (supps.values %>% filter(supp=="VC"))$len

t.test
```

```r
function (x, ...)
UseMethod("t.test")
<bytecode: 0x00000000139cb330>
<environment: namespace:stats>
```