Effect of Vitamin C on Tooth Growth of Guinea pigs

Tomás A. Maccor

22/3/2020

Introduction

This assignment explores whether 3 different doses & 2 different delivery methods of Vitamin C have an influence on the tooth lengths of Guinea pigs. The dataset used is **ToothGrowth**, which comes with R as one of its practice/sample datasets.

Each animal received one of 3 doses of vitamin C, by one of 2 delivery methods: ascorbic acid (VC) versus orange juice (OJ).

This report comprises:

- A basic/descritive statistics summary of our sample data, plus an exploratory analysis.
- Use of statistical inference methods so that if any effects/conclusions are obtained from the SAMPLE data (contained in the ToothGrowth dataset), we can infer that these effects also apply to the entire population of Guinea pigs.

Summary of data & exploratory analysis

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------- tidyverse 1.2.1

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------ tidyverse_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(reshape2)
```

```
##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(BSDA)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'BSDA'

## The following object is masked from 'package:datasets':
##
##     Orange
```

```
## load Tooth Growth data

data("ToothGrowth")
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
```

LEN = length of odontoblast. NUMERIC variable SUPP = delivery method (VC or OJ). FACTOR DOSE = Vitamin C dose: 0.5, 1 & 2 mg/day. NUMERIC variable

How many Guinea pigs received which dose, and by which method?:

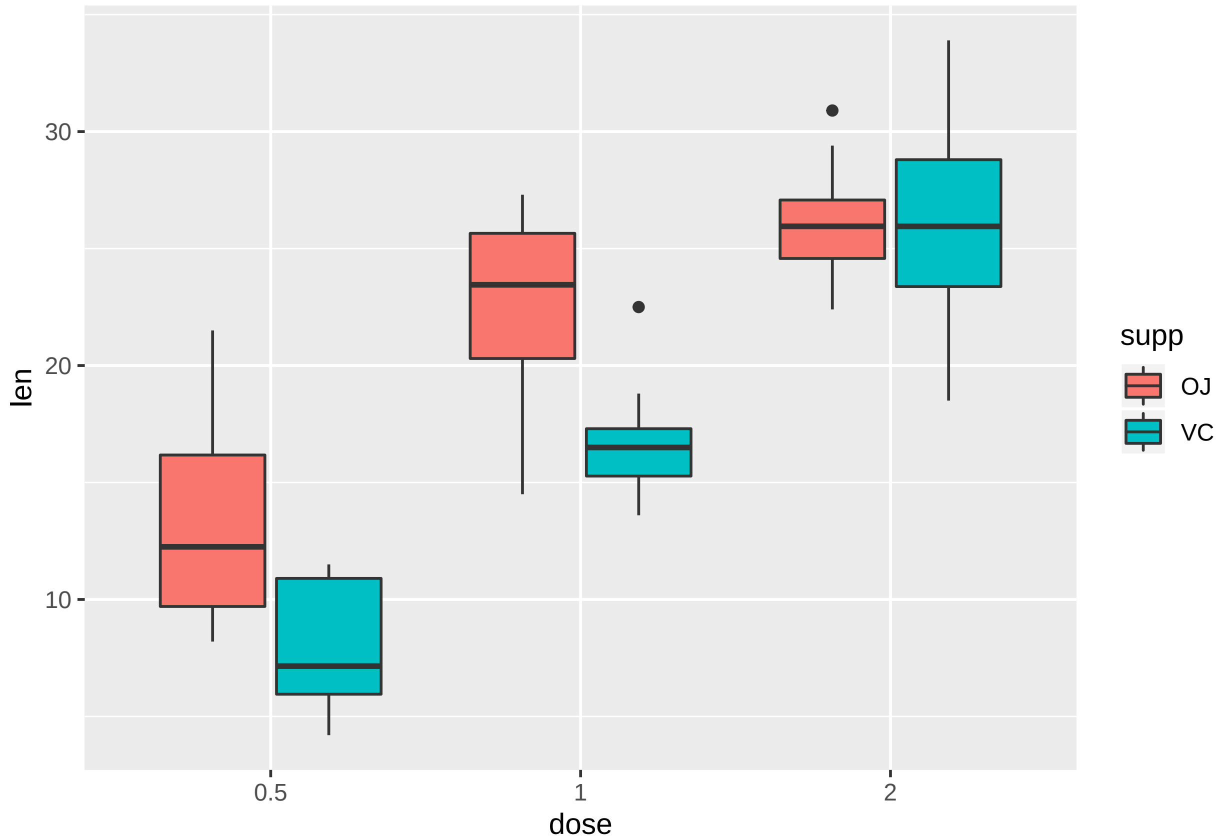
```
ToothGrowth %>% group_by(dose, supp) %>% summarise(n = n())
```

```
## # A tibble: 6 x 3
## # Groups:   dose [3]
##    dose supp      n
##   <dbl> <fct> <int>
## 1   0.5 OJ       10
## 2   0.5 VC       10
## 3   1   OJ       10
## 4   1   VC       10
## 5   2   OJ       10
## 6   2   VC       10
```

So then -20 pigs received each of the 3 doses —> 10 via VC, 10 via OJ.

Let's have a first visualization of the data:

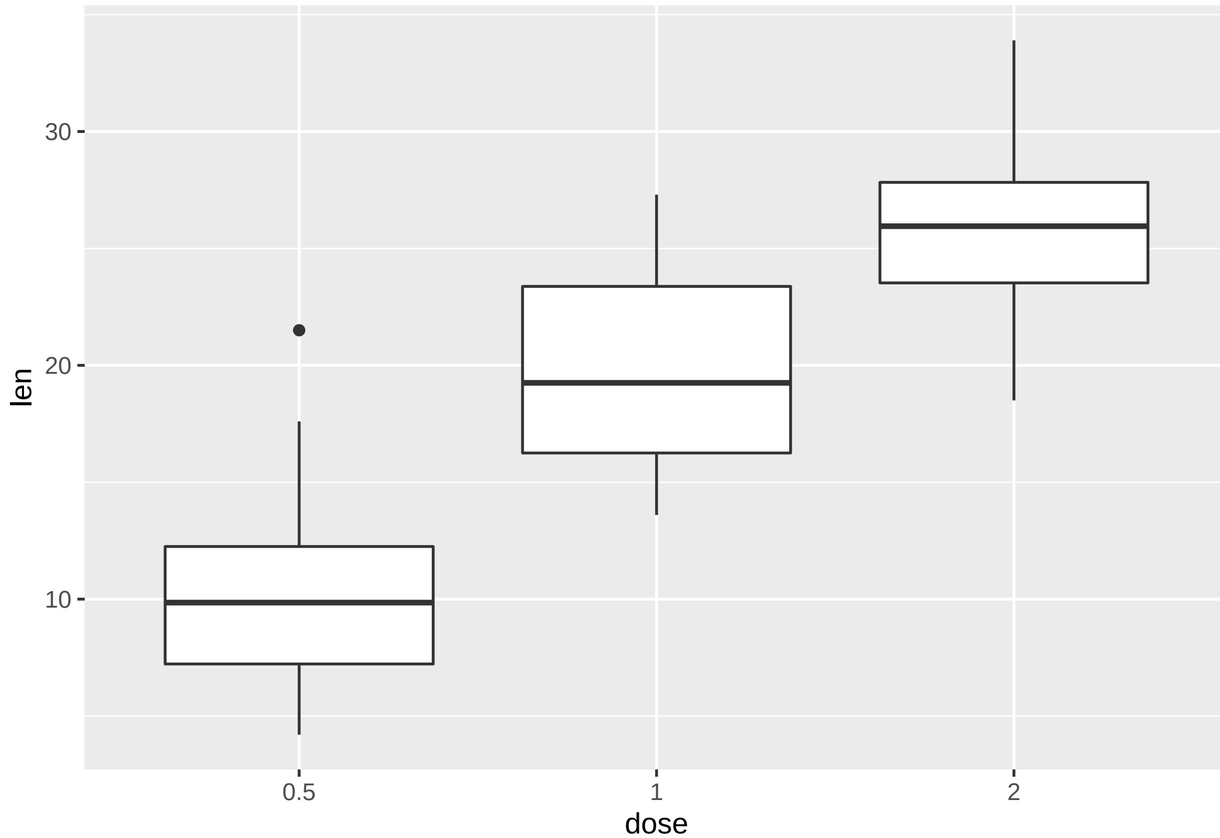
```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

dose_supp_plot <- ggplot(ToothGrowth, aes(x = dose, y = len, fill = supp))
dose_supp_plot + geom_boxplot()
```



We can observe what could be a significant difference in Tooth Growth between the 0.5 mg/day dose versus the 1 & 2 mg/day doses. We need to test for statistical significance to ascertain this. It would also appear there is a difference between delivering Vit_C via *VC* vs *OJ* for the *0.5* and *1mg/day* doses.

If we plot the tooth growth differences by Vitamin C Dose only, and obtain the means for the 3 different Vit C doses:

```
dose_plot <- ggplot(ToothGrowth, aes(x = dose, y = len))
dose_plot + geom_boxplot()
```




```
ToothGrowth %>% group_by(dose) %>% summarise(Mean_Tooth_Growth = mean(len))
```

```
## # A tibble: 3 x 2
##   dose  Mean_Tooth_Growth
##   <fct>             <dbl>
## 1 0.5                10.6
## 2 1                  19.7
## 3 2                  26.1
```

This time, it could well be that there's statistical difference between all 3 doses.

As we have > 2 groups with what appear similar intergroup variance, we use ANOVA for testing significance (equal variance).

So far, our assumptions are: - That tooth growth in Guinea pigs is normally distributed - That the 3 different DOSE groups have an equal variance

```
ANOVA_dose <- aov(len ~ dose, data = ToothGrowth)
summary.aov(ANOVA_dose)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## dose         2   2426    1213   67.42 9.53e-16 ***
## Residuals   57   1026      18                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So, per ANOVA, we have a statistically significant difference in Mean Tooth Growth between the 3 DOSES ($p < 0.05$ for the 3 doses).

But, ANOVA does not allow to know which of the pairwise DOSE comparisons are significant –so now we perform TUKEY TEST to determine this:

```
TukeyHSD(ANOVA_dose)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = len ~ dose, data = ToothGrowth)
##
## $dose
##          diff       lwr       upr    p adj
## 1-0.5   9.130  5.901805 12.358195 0.00e+00
## 2-0.5  15.495 12.266805 18.723195 0.00e+00
## 2-1     6.365  3.136805  9.593195 4.25e-05
```

Which also results in statistically significant differences between all 3 doses ($p < 0.05$). The *differences between the mean DOSES* and their *confidence intervals for those mean differences* are listed(provided) in the Tukey Test.

So for example, we can state that if the entire population of Guinea pigs was given Vitamin C at 3 doses, and we took random samples of these pigs, 95% of the times we would obtain a Mean difference in Tooth Growth that would be between 5.90 to 12.36 (in the 0.5 mg vs. 1.0 mg/day Vitamin C groups).

In relation to the TYPE of Vitamin C supplement, it is not advisable to test solely between VC vs OJ, as DOSE could a confounder in this relation.

So we do pairwise t-tests for each dose

```
ToothGrowth <- ToothGrowth %>% arrange(dose)
Dose0.5 <- ToothGrowth[1:20, ]
Dose1 <- ToothGrowth[21:40, ]
Dose2 <- ToothGrowth[41:60, ]

t.test(len ~ supp, data = Dose0.5)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, data = Dose1)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC 
##            22.70            16.77
```

```
t.test(len ~ supp, data = Dose2)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC 
##            26.06            26.14
```