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

First glimpse of the ToothGrowth dataset:

```
## '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 ...
```

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

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

```
a <- ToothGrowth %>% group_by(dose, supp) %>% summarise(n = n())
knitr::kable((a), caption = "Nr. of observations in each group", align = c("l") )
```

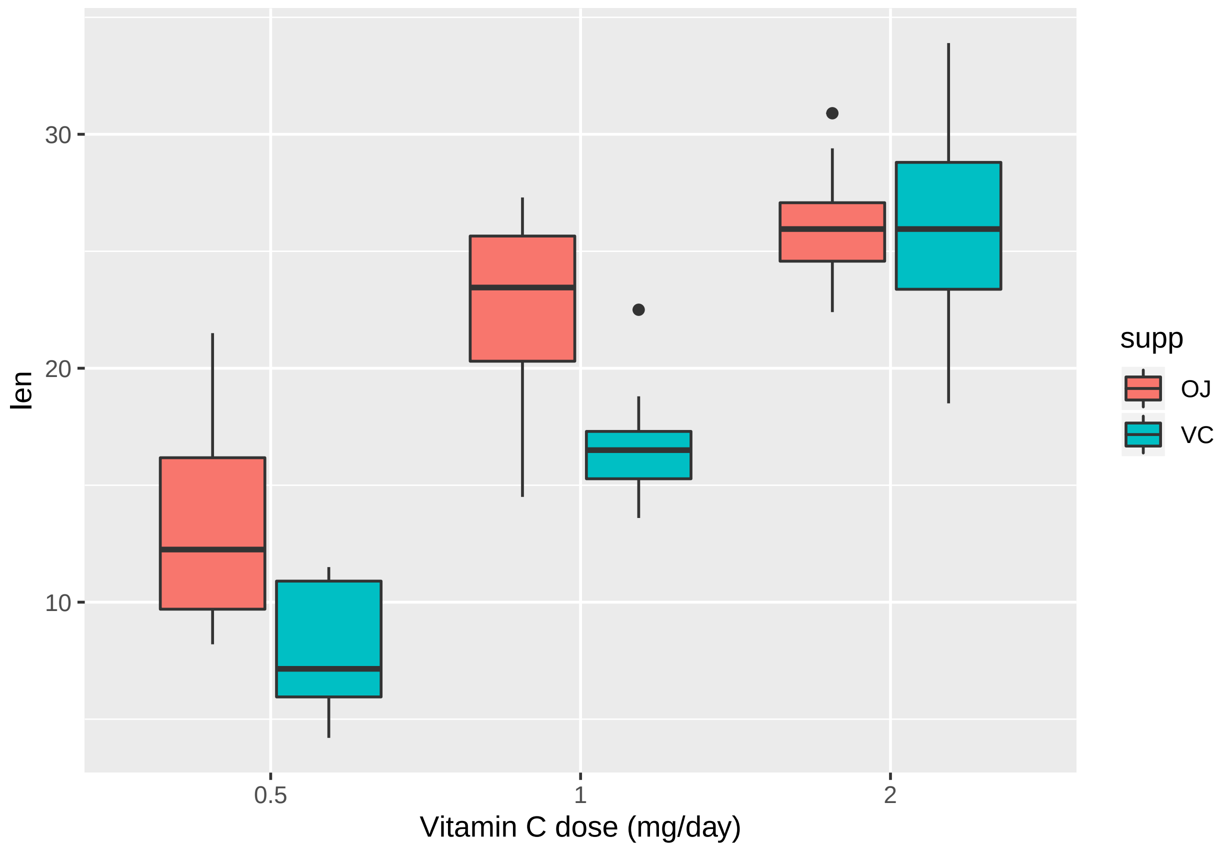
Table 1: Nr. of observations in each group

dose	supp	n
0.5	OJ	10
0.5	VC	10
1.0	OJ	10
1.0	VC	10
2.0	OJ	10
2.0	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:

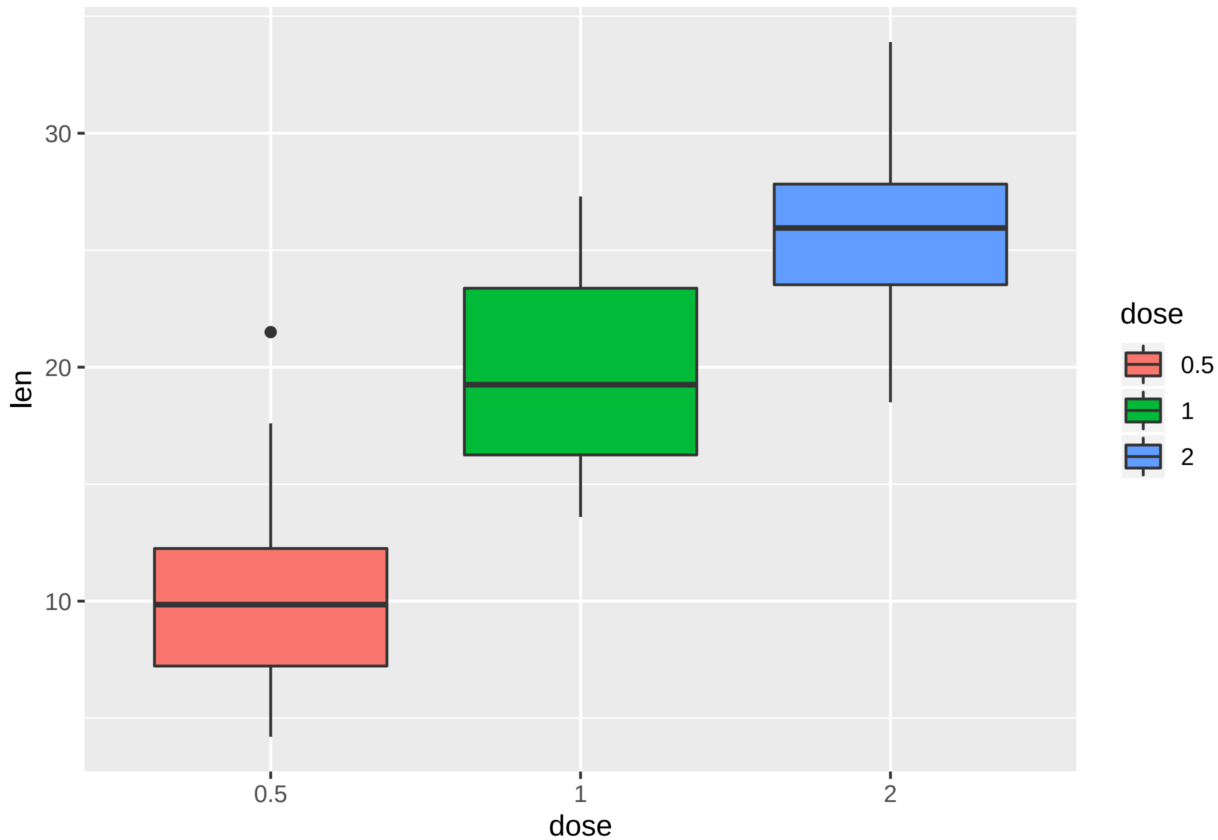
```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

dose_supp_plot <- ggplot(ToothGrowth, aes(x = dose, y = len, fill = supp))
dose_supp_plot + geom_boxplot() + xlab("Vitamin C dose (mg/day)")
```



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 Vitamin C via **VC** vs **OJ** for the **0.5 & 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, fill = dose))
dose_plot + geom_boxplot()
```




```
b <- ToothGrowth %>% group_by(dose) %>% summarise(Mean_Tooth_Growth = mean(len))
knitr::kable( (b), align = c("l") )
```

dose	Mean_Tooth_Growth
0.5	10.605
1	19.735
2	26.100

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

HYPOTHESIS TESTING - RESULTS - CONCLUSIONS

```
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 ($p = 9.53e\text{-}16$, thus $p < 0.05$).

But ANOVA does not allow to know which of the pairwise DOSE comparisons are significant –we need to perform a 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 the **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 delivery method (**supp**, VC vs. OC), we must perform a t-test between the 30 pigs that received Vitamin C via VC versus the 30 that received it via OJ:

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The results show that **there is no statistically significant difference in Tooth Growth depending on the Vitamin C delivery method used** ($p = 0.06$). We can also confirm this result by noting that the 95% confidence interval for the *difference in means* includes "0" as one of the possible values (so, one of the statistically possible values would be that the Mean(VC) - Mean(OJ) = 0)

If we were to perform paired t-tests for VC/OJ, within each of the dose groups, the result would be similar, with the difference that we would loose statistical confidence (we would obtain a confidence interval of 85% at a maximum, which is less certainty, thus is not desired)