

# Statistical Inference Project 2

Isaac G Veras

05/10/2023

## Introduction:

In this project, the exponential distribution in R and the comparison with the Central Limit Theorem will be investigated. The exponential distribution will be simulated in R with `rexp(n, lambda)`, where `lambda` is the rate parameter. The mean of the exponential distribution is  $1/\lambda$  and the standard deviation is also  $1/\lambda$ . Setting `lambda = 0.2` for all simulations. In this way, the distribution of the means of exponential 40 will be investigated. It will be necessary to do a thousand simulations.

## Part 2: Basic Inferential Data Analysis Instructions:

### Questions:

1. Load the `ToothGrowth` data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
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 are other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

### Package installation:

```
if (!require("pacman")) install.packages("pacman")
```

```
## Carregando pacotes exigidos: pacman
```

```
pacman::p_load(pacman,      # Package Manager
               knitr,       # Transform R Markdown documents into various output formats
               plyr,        # Data manipulation
               data.table,  # Manipulate, process and analyze large data sets
               tidyverse    # Data organization
               )
```

### Pre-processing:

1. Load the `ToothGrowth` data and perform some

# basic exploratory data analyses:

```
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 ...
```

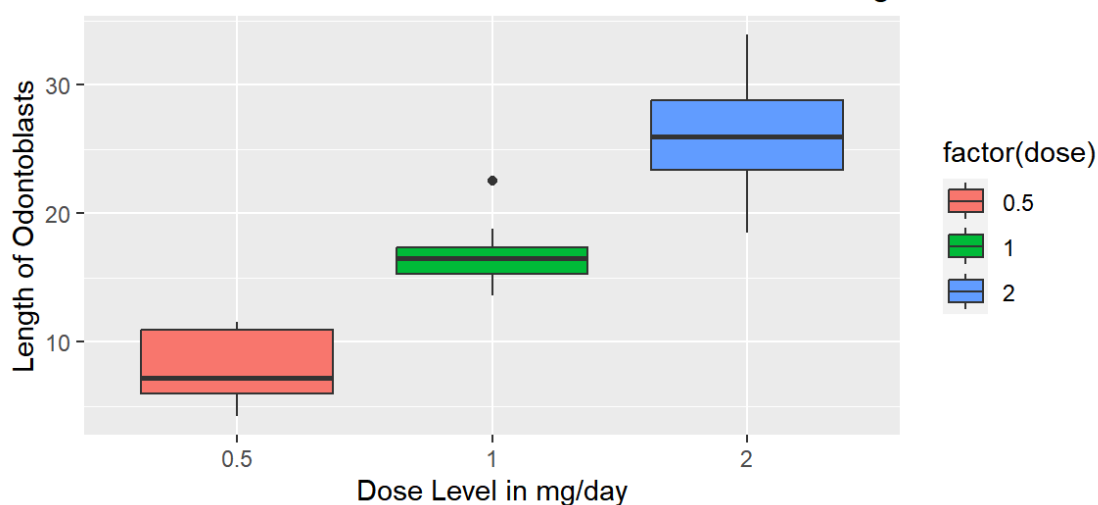
We can see that our data frame ToothGrowth contains 60 observations of 3 variables. Variable supp of type factor has two levels. VC stands for Vitamin C (ascorbic acid) and OJ is Orange Juice.

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
## OJ   10 10 10
## VC   10 10 10
```

We can see that for each supplement type we have 3 dose levels of vitamin C (0.5, 1, and 2 mg/day). Each supplement is tested on 30 guinea pigs and each supplement is tested in different doses on a subgroup of 10 guinea pigs.

Effect of Ascorbic Acid on Tooth Growth in Guinea Pigs



From the boxplot, it's obvious that the cells responsible for tooth growth in guinea pigs are getting longer as we increase the dose of ascorbic acid as a supplement.

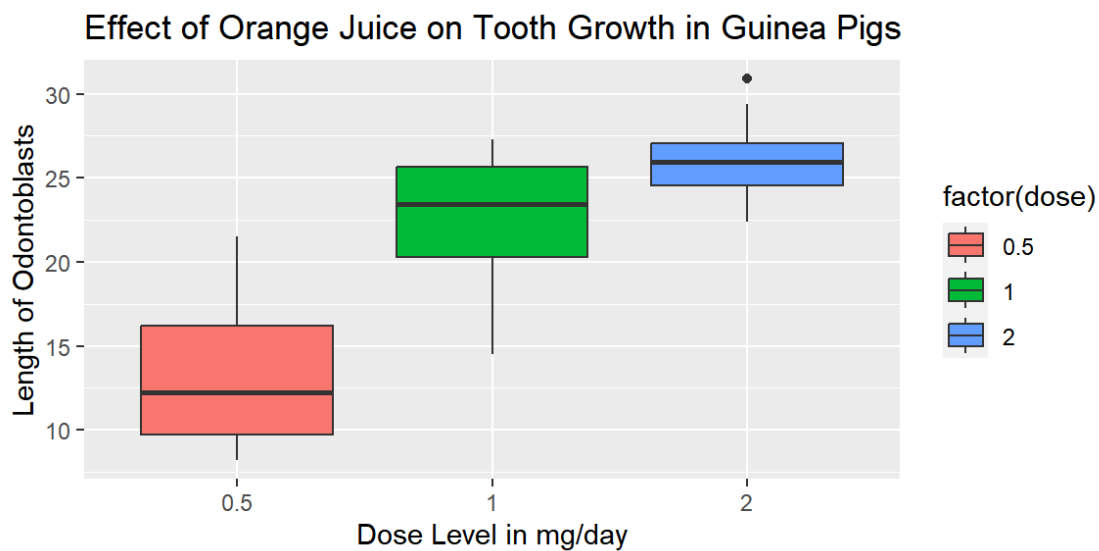
```
tapply(vitaminC$len, vitaminC$dose, mean)
```

```
##      0.5      1      2
##  7.98 16.77 26.14
```

```
tapply(vitaminC$len, vitaminC$dose, var)
```

```
##      0.5      1      2
##  7.544000  6.326778 23.018222
```

We can see that both means and variances of each group differ noticeably.



From these boxplot we can see that as the dose level increases, so does the length of odontoblasts.

```
tapply(orangeJ$len, vitaminC$dose, mean)
```

```
##    0.5    1    2  
## 13.23 22.70 26.06
```

```
tapply(orangeJ$len, vitaminC$dose, var)
```

```
##      0.5      1      2  
## 19.889000 15.295556  7.049333
```

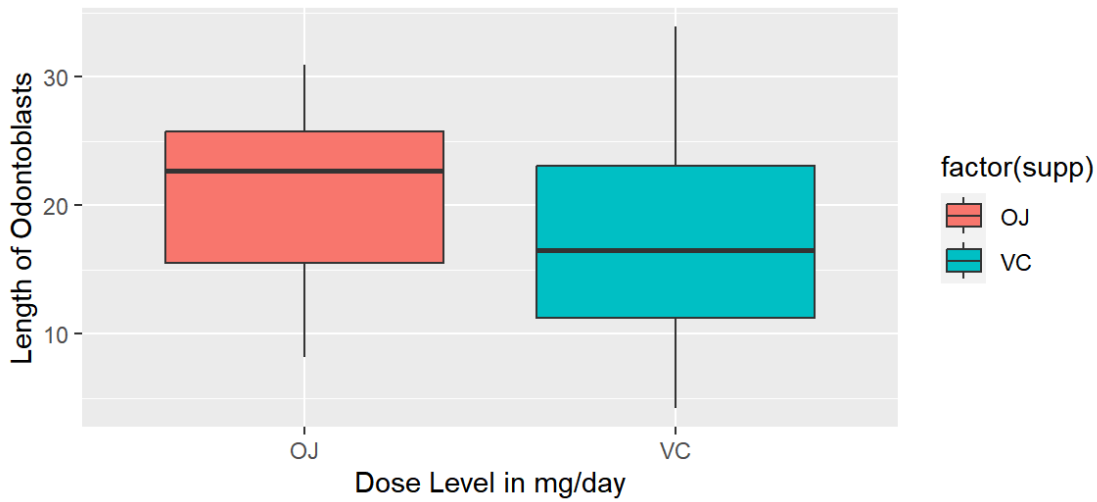
We can also check if there is an overall difference in the effect on tooth growth between two supplements.

```
tapply(ToothGrowth$len, ToothGrowth$supp, mean)
```

```
##      OJ      VC  
## 20.66333 16.96333
```

We can see that orange juice as a supplement is, on average, much more effective in promoting tooth growth in guinea pigs than ascorbic acid.

## Effect of VC and OJ on Tooth Growth in Guinea Pigs



From this graph, we can also see that ascorbic acid as a supplement has more varying effects on tooth growth in guinea pigs.

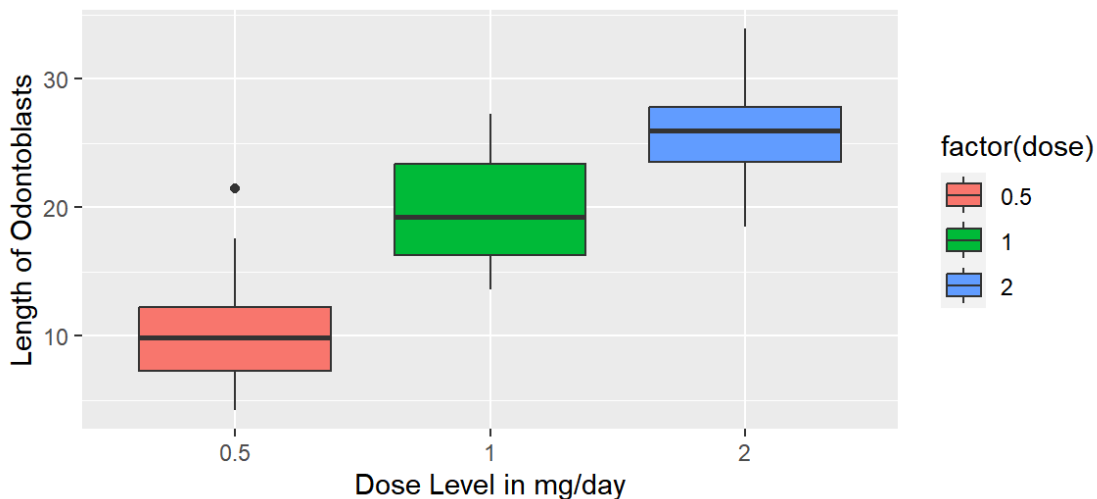
```
tapply(ToothGrowth$len, ToothGrowth$supp, var)
```

```
##      OJ      VC
## 43.63344 68.32723
```

This confirms our initial claim that ascorbic acid causes greater variance in the resulting length of odontoblasts (cells responsible for tooth growth).

To see the overall effect of dose level on tooth growth:

## Effect of Dose Level on Tooth Growth in Guinea Pigs



```
tapply(ToothGrowth$len, ToothGrowth$dose, mean)
```

```
##      0.5      1      2
## 10.605 19.735 26.100
```

```
tapply(ToothGrowth$len, ToothGrowth$dose, var)
```

```
##      0.5      1      2
## 20.24787 19.49608 14.24421
```

As expected, as we increase the dose, the length of odontoblasts increases irrelative of the supplement type used.

## 2. Provide a basic summary of the data:

```
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
```

From this output we can't really see much, except that in each supplement subgroup there are 30 guinea pigs. We also see that the overall mean of length of odontoblasts is 18.81 and the median is 19.25. The Minimum length is 4.20 and the maximum is 33.90.

## 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 are other approaches worth considering)

**First** we'll test the alternative hypothesis that means the length of odontoblasts for ascorbic acid (VC) and orange juice (OJ) differs significantly. We'll assume unequal variance, and we know that these groups are not paired. Even though these settings are default in `t.test()`, I'll assign them `FALSE` value for the sake of clarity.

```
vitaminC <- subset(ToothGrowth, supp == 'VC')
orangeJ <- subset(ToothGrowth, supp == 'OJ')
t.test(vitaminC$len, orangeJ$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vitaminC$len and orangeJ$len
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

As we can see  $p\text{-value} = 0.06063 > 0.05 \Rightarrow$  There is not enough evidence to reject the null hypothesis that the difference in average tooth length (odontoblasts) for VC and OJ group is equal to zero.

In our **second step** we'll do several t-tests to check if there's a statistically significant difference in mean values of length of odontoblasts (tooth growth cells) between groups that were subjected to different dose levels of supplements.

We'll assume again that `paired = FALSE` and `var.equal = FALSE`, but since these are default in `t.test()` this time, we'll skip redundant assignments.

```
dose05 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
dose1 <- ToothGrowth$len[ToothGrowth$dose == 1.0]
dose2 <- ToothGrowth$len[ToothGrowth$dose == 2.0]

t.test(dose05, dose1 )
```

```
##
## Welch Two Sample t-test
##
## data: dose05 and dose1
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.983781 -6.276219
## sample estimates:
## mean of x mean of y
## 10.605 19.735
```

```
t.test(dose05, dose2)
```

```
##
## Welch Two Sample t-test
##
## data: dose05 and dose2
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
## 10.605 26.100
```

```
t.test(dose1, dose2)
```

```
##
## Welch Two Sample t-test
##
## data: dose1 and dose2
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
## 19.735 26.100
```

As we can see, from all 3 t.tests, p value is less than 0.05, and we can reject the null hypothesis that the difference between means is equal to zero. We also see that all the 95% confidence intervals are below zero, which confirms are initial claims that the tooth growth increases as we increase the dose level.

In the **third step** we'll inspect the difference in mean lengths of odontoblastos (cells responsible for tooth growth) between ascorbic acid (VC) group and orange juice (OJ) group for each dose level applied.

```
dose05VC <- vitaminC$len[vitaminC$dose == 0.5]
dose1VC <- vitaminC$len[vitaminC$dose == 1.0]
dose2VC <- vitaminC$len[vitaminC$dose == 2.0]

dose05OJ <- orangeJ$len[orangeJ$dose == 0.5]
dose1OJ <- orangeJ$len[orangeJ$dose == 1.0]
dose2OJ <- orangeJ$len[orangeJ$dose == 2.0]
```

As in previous cases, we'll assume that observations are not paired and that variances are not equal.

```
t.test(dose05VC, dose05OJ)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose05VC and dose05OJ
## 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:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98      13.23
```

```
t.test(dose1VC, dose1OJ)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1VC and dose1OJ
## 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:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

```
t.test(dose2VC, dose2OJ)
```

```
##
## Welch Two Sample t-test
##
## data: dose2VC and dose2OJ
## 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:
## -3.63807 3.79807
## sample estimates:
## mean of x mean of y
## 26.14 26.06
```

We can see from the t-tests that the difference between group means for VC and OJ is statistically significant for dose level 0.5 mg/day, where  $p = 0.006358 < p = 0.05$  and for dose level of 1 mg/day, where the p value  $p = 0.001038 < p = 0.05$ . For dose level 2 mg/day we have that  $p\text{-value} = 0.9639 > 0.05$  and 95 percent confidence interval: -3.63807 3.79807 includes zero. We conclude that there is no statistically significant difference between the group means for VC and OJ when the dose level is 2 mg/day.

## 4. State your conclusions and the assumptions needed for your conclusions.

### Assumptions:

1. Samples used are random iid samples.
2. Each sample is independent of one another, in other words; they are not paired.
3. The population distribution of each sample must be approximately normal or mound shaped and roughly symmetric.

### Conclusions:

1. Supplement type alone does not affect the mean value of length of odontoblasts (cells responsible for tooth growth).
2. Dose level alone, without consideration of the supplement type affects tooth growth significantly. Increasing the dosage will induce better tooth growth.
3. Orange Juice as a supplement, when used in dose levels of 0.5 mg/day and 1mg/day promotes better tooth growth than ascorbic acid. When applied in dose level of 2 mg/day, it has an effect on tooth growth similar to that of an ascorbic acid.