

Basic Inferential Data Analysis Exercise

Statistical Inference

Linda Angulo Lopez

01/01/2021

key words: ToothGrowth data, Data Exploration, T Tests, Box Plots, Data Analysis, Statistical Inference.

Summary

The project consists of two parts (i) a simulation exercise and a (ii) basic inferential data analysis, the former is presented here. The hypothesis tested, was that the dosage and supplement do not affect tooth length, the alternative is that it does. The population was assumed to be near normally distributed. It was also found that at p-value is 0.03032, when comparing orange juice to vitamin C, so we can reject the null hypothesis. Further investigation showed that orange juice is linked to higher tooth growth length at dose = .5mg and dose = 1.0mg. But that there is no significant difference of tooth length at dose = 2.0mg, the p-value was almost 1 at 0.9639.

#R4.0 Environmental Set for:

```
library(knitr) # creating a pdf document ;
library(ggplot2) # making plots
library(dplyr) #exploring data
library(DataExplorer) # creating reports
library(UsingR); data("ToothGrowth") # Read in the ToothGrowth data
```

Plot & explore data

Here is a glimpse of the data: ToothGrowth, and a basic data summary:

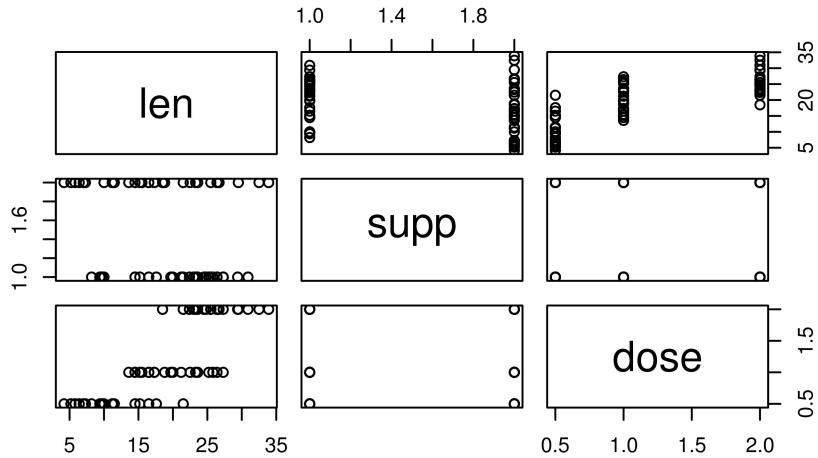
```
glimpse(ToothGrowth); summary(ToothGrowth)
```

```

## Rows: 60
## Columns: 3
## $ len <dbl> 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16.5, 16...
## $ supp <fct> VC, VC...
## $ dose <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1.0, 1....
```

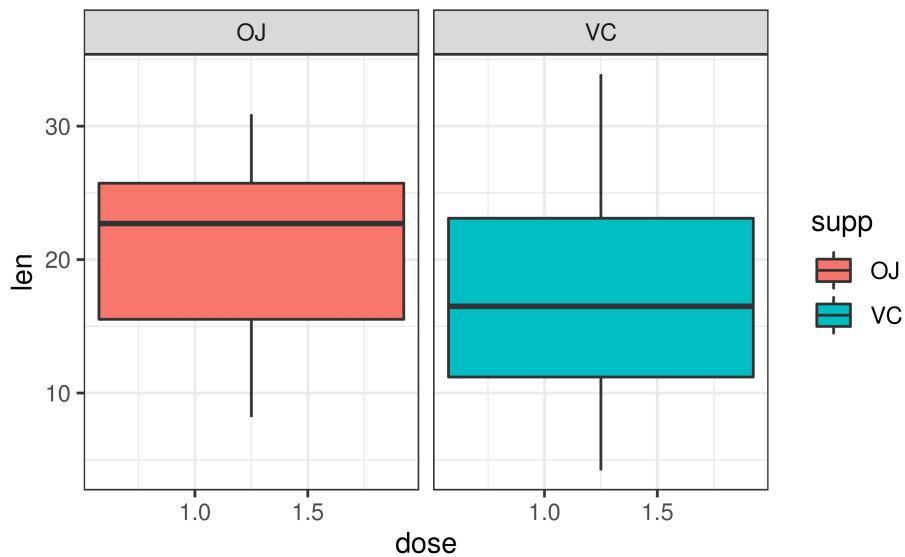
| | len | supp | dose |
|------------|--------|-------|---------------|
| ## Min. | : 4.20 | 0J:30 | Min. :0.500 |
| ## 1st Qu. | :13.07 | VC:30 | 1st Qu.:0.500 |
| ## Median | :19.25 | | Median :1.000 |
| ## Mean | :18.81 | | Mean :1.167 |
| ## 3rd Qu. | :25.27 | | 3rd Qu.:2.000 |
| ## Max. | :33.90 | | Max. :2.000 |

and a quick plot of the data: *ToothGrowth*



From the quick data plots we see that the tooth length varies by dose & supplements, the following box plots affirm this.

Closer Inspection: Dose



It seems that orange juice is linked to better tooth grown results, we will use a *t test* to evaluate the statistical significance.

```
##      [,1]      [,2]
## [1,] 1.719057 8.780943
## [2,] 2.802148 9.057852
## [3,] -3.798070 3.638070
```

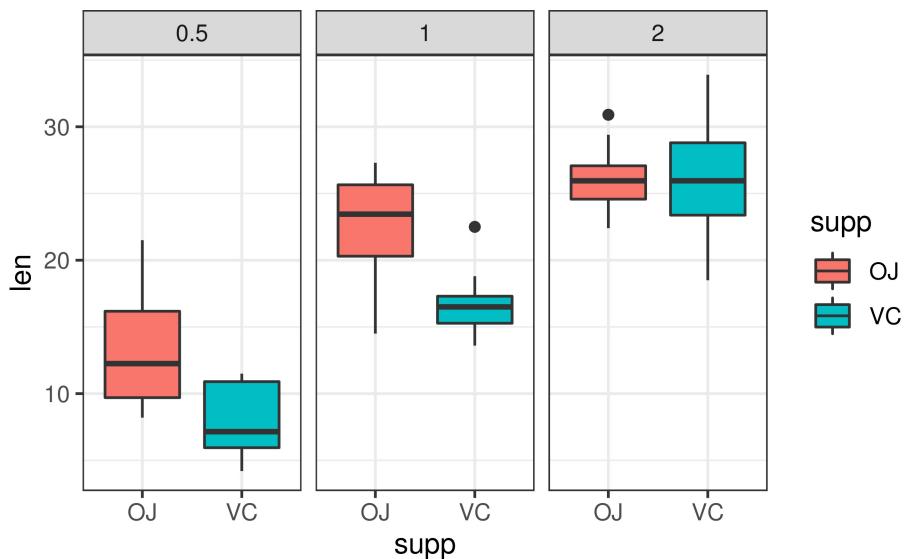
So, we see that there is a greater tooth length at doses .5 and 1.0 but that it is split at zero at dose 2.0. So we can accept the hypothesis at dose .5 and 1.0 as at these doses orange juice is linked to higher tooth

growth length.

Let's split the data to get more insights with a Welch t-test:

```
##  
## Welch Two Sample t-test  
##  
## data: OJ and VC  
## t = 1.9153, df = 55.309, p-value = 0.03032  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.4682687      Inf  
## sample estimates:  
## mean of x mean of y  
## 20.66333 16.96333
```

If we set alpha to .05 we can reject the null hypothesis, as the p -value, .03, is lower than alpha.



Let's plot that:

From the box plot we see that at dose = 2mg both vitamin C, VC, and Orange Juice, OJ, produce tooth length of around 26. This can be affirmed with a two sided Welch t-test.

```
##  
## Welch Two Sample t-test  
##  
## data: OJ2 and VC2  
## 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 of x mean of y  
## 26.06 26.14
```

The means are the same, so there is no difference, the p-value = 0.9639 almost 1.

After-note: - This is an R Markdown document, created as a submission for the Statistical Inference Course by Johns Hopkins University on Coursera. - The exercise can be replicated hidden code is available here.