Part 2: Basic Inferential Data Analysis

David Augusto

1/3/2021

==========

Part 2: Basic Inferential Data Analysis Instructions

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(ggplot2)
library(datasets)
data(ToothGrowth)
```

2. Provide a basic summary of the data.

```
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 ...
```

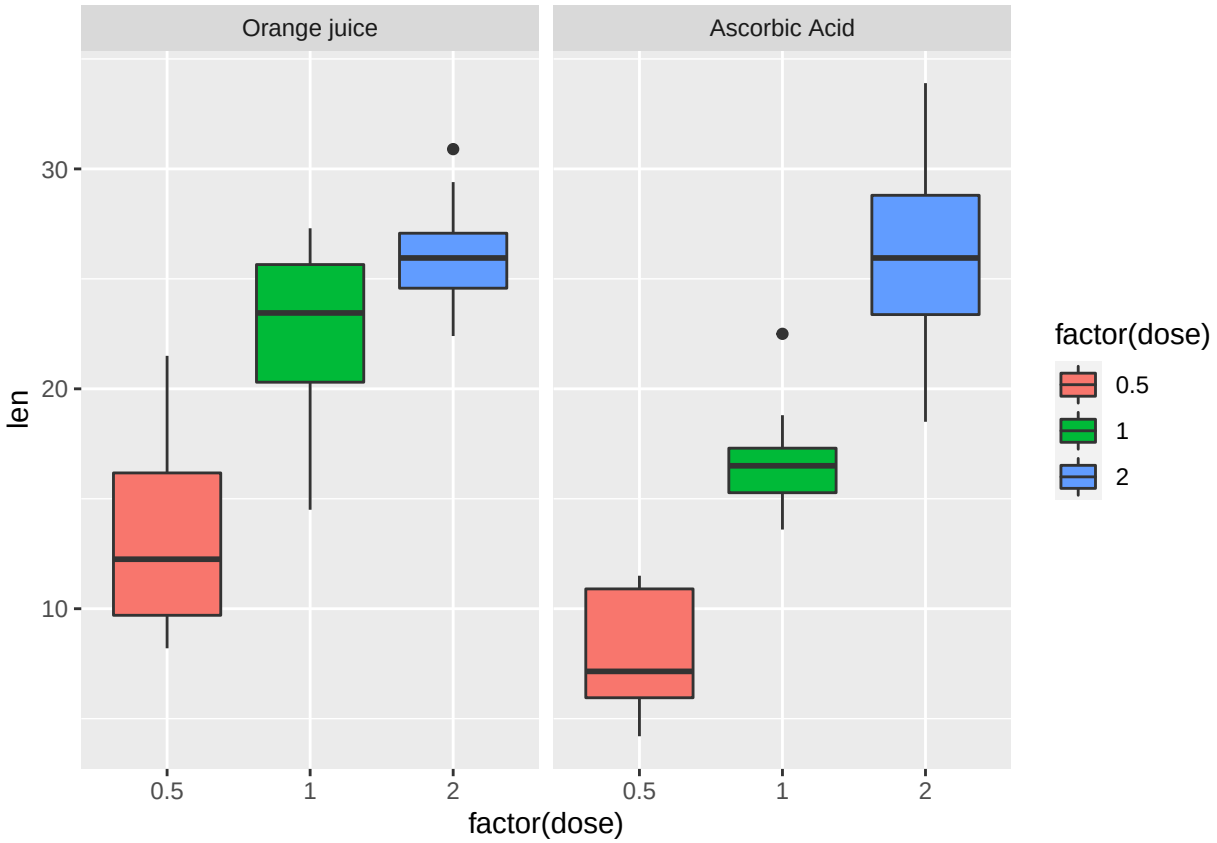
```
summary(ToothGrowth)
```

```
##       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
```

==========

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's other approaches worth considering)

```
ggplot(ToothGrowth, aes(factor(dose), len, fill = factor(dose))) +
      geom_boxplot() +
      facet_grid(.~supp, labeller = as_labeller(
            c("OJ" = "Orange juice",
              "VC" = "Ascorbic Acid")))
```



Next we propose different hypothesis test for the dataset:

==========

1. First hypothesis test

Ho: mean growth of all the population of Orange Juice is == mean growth of all the population of Ascorbic acid

H1: mean growth of all the population of Orange Juice is =! mean growth of all the population of Ascorbic acid

```
hyth1 <- t.test(len ~ supp, data = ToothGrowth)
hyth1$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
hyth1$p.value
```

```
## [1] 0.06063451
```

Since the p-value is 0.06063451 greater than 0.05. The null hypothesis **cannot** be rejected.

We could assume that either one provides the same growth across dosage.

==========

2. Second hypothesis test

Ho: mean growth of Orange Juice when the dose is .5 == mean growth of Ascorbic acid when the dose is .5

H1: mean growth of Orange Juice when the dose is .5 =! mean growth of Ascorbic acid when the dose is .5

```
hyth2<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 0.5))
hyth2$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
hyth2$p.value
```

```
## [1] 0.006358607
```

Since the p-value is 0.006358607 lower than 0.05. The null hypothesis **is** rejected.

We could assume that the mean growth isn´t the same between supplements when the dose is 0.5

==========

3. Third hypothesis test

Ho: mean growth of Orange Juice when the dose is 1 == mean growth of Ascorbic acid when the dose is 1

H1: mean growth of Orange Juice when the dose is 1 =! mean growth of Ascorbic acid when the dose is 1

```
hyth3<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 1))
hyth3$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
hyth3$p.value
```

```
## [1] 0.001038376
```

Since the p-value is 0.001038376 lower than 0.05. The null hypothesis **is** rejected.

We could assume that the mean growth isn´t the same between supplements when the dose is 1

==========

4. Fourth hypothesis test

Ho: mean growth of Orange Juice when the dose is 2 == mean growth of Ascorbic acid when the dose is 2

H1: mean growth of Orange Juice when the dose is 2 =! mean growth of Ascorbic acid when the dose is 2

```r
hyth4<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 2))
hyth4$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```r
hyth4$p.value
```

```
## [1] 0.9638516
```

Since the p-value is 0.9638516 greater than 0.05. The null hypothesis **cannot** be rejected.

We could assume that either one provides the same mean growth when the dose is 2.

==========

Conclusions

From the statistical analysis and the plotted data:

From the boxplot we can observe how the length behaves for each of the dosages:

- .05 and 1 the orange juice shows a greater length in comparison to the Ascorbic Acid.
- but for the dose 2, it appears as both OJ or AA have the approximately the same median, but for the AA bigger variance.

From this we can assume, when the dosage is smaller, the Orange juice tends to provide greater growth.

Now from the hypotesis testing we confirm what we saw in the boxplot

The mean is **not the same** when the dosage is .5 and 1,

And we **cannot reject that** the means are the same when the dosage is 2