Inferential Data Analysis

Donato Scarano

03 luglio 2018

Basic Inferential Data Analysis

In this project, we're going to analyze the ToothGrowth data in the R datasets package.

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

```
library(datasets)
data("ToothGrowth")
```

Let's start looking at the structure of the dataset and let us consult the help section with the dataset details.

```
str(ToothGrowth)
```

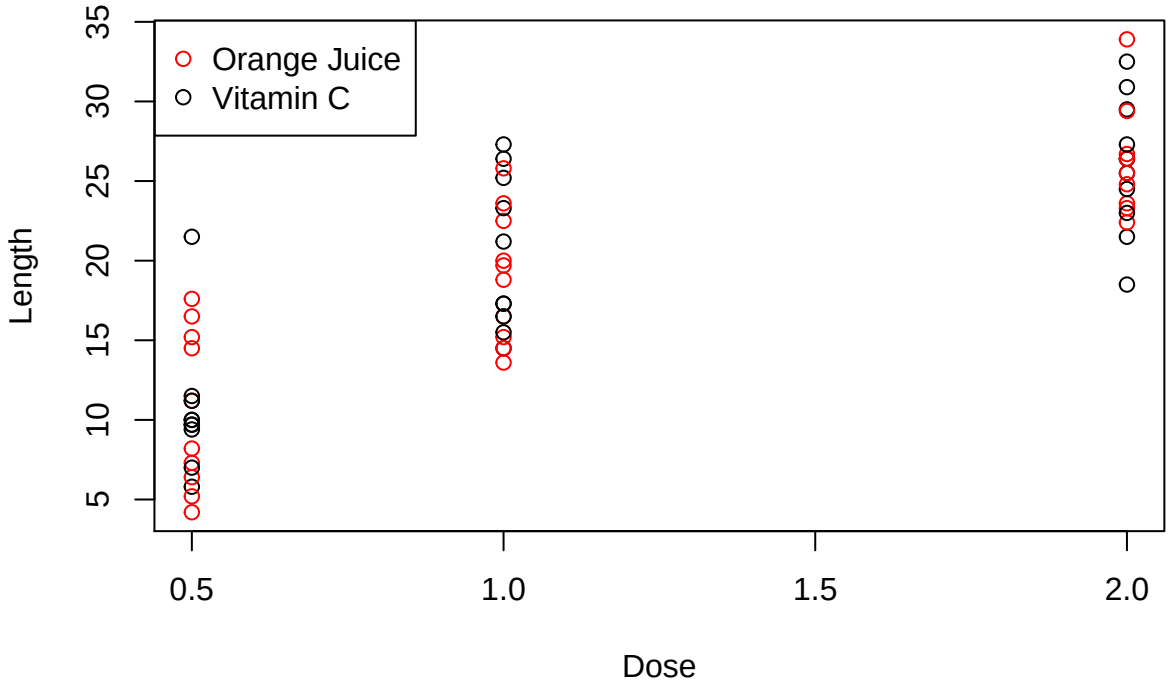
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
?ToothGrowth
```

```
## starting httpd help server ... done
```

We want to see first how the dose received either via Vitamin C or Orange Juice is related to the tooth length. Let us plot the relation above to get a first look.

```
plot(ToothGrowth$len ~ ToothGrowth$dose, xlab = "Dose", ylab = "Length", col = c("red", "black"))
legend("topleft",c("Orange Juice", "Vitamin C"),pch = 1, col = c("red", "black"))
```



There seems to be a positive correlation between the dose and the tooth growth.

Provide a basic summary of the data.

Let us use the summary function to obtain a basic outline

```
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
```

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)

Confidence intervals, we fit a linear model and we apply the confint function to obtain them:

```
fitd <- lm(ToothGrowth$len ~ ToothGrowth$dose)
confint(fitd)
```

```
##                       2.5 %    97.5 %
## (Intercept)      4.900171  9.944829
## ToothGrowth$dose 7.856870 11.670273
```

Same process can be apllied for supp variable as well:

```
fits <- lm(ToothGrowth$len ~ ToothGrowth$supp)
confint(fits)
```

```
##                          2.5 %     97.5 %
## (Intercept)         17.928947 23.3977198
## ToothGrowth$suppVC -7.567006  0.1670064
```

or for both togheter:

```
fittotal <- lm(ToothGrowth$len~ToothGrowth$supp + ToothGrowth$dose)
confint(fittotal)
```

```
##                          2.5 %    97.5 %
## (Intercept)          6.704608 11.840392
## ToothGrowth$suppVC -5.889905 -1.510095
## ToothGrowth$dose    8.007741 11.519402
```

State your conclusions and the assumptions needed for your conclusions.

There is certainly a positive correlation between tooth length and dose but we see a negative one between tooth length and supp VC as shown in confidence interval above. OJ seems to have a greater impact than VC.