

G4.P-1

David Emanuel Craciusecu y Laura Pérez Medeiro

October 12, 2019

The test laboratory will be divided into two parts:

1. An exercise mentored by the teacher, in which all the contents explained in the first unit will be revised. The files used in this exercise are:

1.1 "Satelites.txt", which contains information about the different ratios of Uranus' moons. The aim of the exercise is obtained the same result as we get into the theoretical class.

```
> while (!"satelites.txt" %in% list.files(getwd()))
+ {
+   print("Data file not found. Add \"satelites.txt\" to the current directory.")
+   invisible(readline(prompt="Press [enter] to continue"))
+ }
>     satelites <- read.table("satelites.txt")
>     satelites

      nombre radio
1      CORDELIA   13
2       OFELIA   16
3       BIANCA   22
4      CRESIDA   33
5    LESDEMONA   39
6      JULIETA   42
7    ROSALINDA   27
8      BELINDA   34
9 LUNA-1986U1020   20
10     CALIBANO   30
11     LUNA-119   20
12   LUNA_119U2   15

>     radius <- satelites $radio
>     radius

[1] 13 16 22 33 39 42 27 34 20 30 20 15
```

It is important to take into account that the working directory must be the same to the file's directory when read.table is going to be used. Otherwise, we must indicate the absolute route where the file is.

After reading the data, different analyzes will be applied to those Uranus' moon radius:

a) Calculate absolute and relative satellite radius frequencies:

ABSOLUTE FREQUENCY:

```
> absoluteFreq <- function(set) {table(set)}
> absoluteFreq(radius)
```

```
set
13 15 16 20 22 27 30 33 34 39 42
 1  1  1  2  1  1  1  1  1  1  1
```

ACUMULATIVE ABSOLUTE FREQUENCY:

```
> cumAbsoluteFreq <- function(set) {cumsum(absoluteFreq(set))}
> cumAbsoluteFreq(radius)
```

```
13 15 16 20 22 27 30 33 34 39 42
 1  2  3  5  6  7  8  9 10 11 12
```

RELATIVE FREQUENCY

```
> relativeFreq <- function(set) {table(set) / length(set)}
```

ACUMULATIVE RELATIVE FREQUENCY

```
> cumRelativeFreq <- function(set) {cumsum(relativeFreq(set))}
```

b) Arithmetic mean

```
> arithmeticMean <- function(set, usrTrim = 0) (mean(set, trim = usrTrim))
> arithmeticMean(radius)
```

```
[1] 25.91667
```

c) Measures of dispersion, where the following page was used as a reference for this section: <http://iridl.ldeo.columbia.edu/doehelp/StatTutorial/Dispersion/index.htmlIntro>
- RANGE:

```
> range <- function(set) {max(set) - min(set)}
> range(radius)
```

```
[1] 29
```

- STANDARD DEVIATION

```
> stdDeviation <- function(set)
+ {
+   sd(set) * (sqrt((length(set) - 1) / length(set)))
+ }
> stdDeviation(radius)
```

```
[1] 9.277736
```

- VARIANCE:

```
> variance <- function(set) {var(set) * (length(set) - 1 / length(set))}
> variance(radius)
```

```
[1] 1118.993
```

- ROOT MEAN SQUARE:

```
> rootMeanSqr <- function(set) {sqrt(mean(set ^ 2))}
> rootMeanSqr(radius)

[1] 27.52726
```

- ROOT MEAN SQUARE ANOMALY:

```
> rootMeanSqrAn <- function(set) {sqrt(sum(set - mean(set)) ^ 2) / length(set)}
> rootMeanSqrAn(radius)

[1] 1.184238e-15
```

- INTERQUARTILE RANGE:

```
> interQuartRange <- function(set) {IQR(set)}
> interQuartRange(radius)

[1] 14.25
```

- MEDIAN ABSOLUTE DEVIATION

```
> medAbsDeviation <- function(set) {mad(set)}
> medAbsDeviation(radius)

[1] 12.6021
```

d) Finally, measures of order:

-MEDIAN:

```
> getMedian <- function(set) {median(set)}
> getMedian(radius)

[1] 24.5
```

-MODE:

```
> getMode <- function(set)
+ {
+   uniqueVal <- unique(set)
+   uniqueVal[which.max(tabulate(match(set, uniqueVal)))]
+ }
> getMode(radius)

[1] 20
```

-QUARTILES:

```
> getQuartiles <- function(set) {quantile(set)}
> getQuartiles(radius)

 0%   25%   50%   75%  100%
13.00 19.00 24.50 33.25 42.00
```

-54th QUANTILE:

```
> getQuantiles <- function(set, range = 0) {quantile(set, probs = range)}
> getQuantiles(radius)
```

```
0%
13
```

1.2 Now, the file will be "cardata.sav" where the same analyzes of data will be applied to mpg_agrupada. *But this time, it is not necessary to calculate the 54th and the frequencies.*

It is necessary to import the foreign's library to read data from SPSS files in R.

```
> library(foreign)
```

Then, the file is read, only data related to mpg_agrupada is loaded into the variable mpg_agrupada and we use a filter to

```
> dataset = read.spss("cardata.sav", to.data.frame=TRUE)
> mpg_agrupada = dataset$mpg_agrupada
> mpg_agrupada = mpg_agrupada[!is.na(mpg_agrupada)]
> mpg_agrupada
\end{Sinput}
\begin{Soutput}
 [1] _duplicated_7 _duplicated_3 _duplicated_3 _duplicated_3 _duplicated_3 _duplicated_3
 [7] _duplicated_3 _duplicated_4 _duplicated_3 _duplicated_3 _duplicated_3 _duplicated_3
[13] _duplicated_2 _duplicated_2 _duplicated_3 _duplicated_2 _duplicated_2 _duplicated_2
[19] _duplicated_6 _duplicated_6 _duplicated_4 _duplicated_4 _duplicated_3 _duplicated_3
[25] _duplicated_3 _duplicated_3 _duplicated_3 _duplicated_2 _duplicated_2 _duplicated_2
[31] _duplicated_2 _duplicated_2 _duplicated_3 _duplicated_2 _duplicated_2 _duplicated_2
[37] _duplicated_5 _duplicated_4 _duplicated_4 _duplicated_7 _duplicated_7 _duplicated_7
[43] _duplicated_5 _duplicated_5 _duplicated_7 _duplicated_6 _duplicated_5 _duplicated_5
[49] _duplicated_4 _duplicated_3 _duplicated_5 _duplicated_4 _duplicated_5 _duplicated_5
[55] _duplicated_4 _duplicated_4 _duplicated_6 _duplicated_8 _duplicated_7 _duplicated_7
[61] _duplicated_6 _duplicated_3 _duplicated_5 _duplicated_3 _duplicated_2 _duplicated_2
[67] _duplicated_5 _duplicated_7 _duplicated_6 _duplicated_5 _duplicated_5 _duplicated_5
[73] _duplicated_4 _duplicated_8 _duplicated_7 _duplicated_4 _duplicated_8 _duplicated_8
[79] _duplicated_3 _duplicated_7 _duplicated_5 _duplicated_5 _duplicated_6 _duplicated_6
[85] _duplicated_6 _duplicated_9 _duplicated_3 _duplicated_2 _duplicated_3 _duplicated_3
[91] _duplicated_6 _duplicated_6 _duplicated_4 _duplicated_5 _duplicated_8 _duplicated_8
[97] _duplicated_7 _duplicated_10 _duplicated_10 _duplicated_8 _duplicated_6 _duplicated_6
[103] _duplicated_6 _duplicated_7 _duplicated_7 _duplicated_7 _duplicated_5 _duplicated_5
[109] _duplicated_7 _duplicated_10 _duplicated_7 _duplicated_8 _duplicated_7 _duplicated_7
[115] _duplicated_5 _duplicated_3 _duplicated_4 _duplicated_6 _duplicated_7 _duplicated_7
[121] _duplicated_8 _duplicated_8 _duplicated_6 _duplicated_6 _duplicated_8 _duplicated_8
[127] _duplicated_10 _duplicated_9 _duplicated_10 _duplicated_7 _duplicated_6 _duplicated_6
[133] _duplicated_6 _duplicated_8 _duplicated_7 _duplicated_6 _duplicated_8 _duplicated_8
[139] _duplicated_7 _duplicated_7 _duplicated_6 _duplicated_7 _duplicated_6 _duplicated_6
[145] _duplicated_4 _duplicated_8 _duplicated_6 _duplicated_7 _duplicated_7 _duplicated_7
[151] _duplicated_8 _duplicated_6 _duplicated_8 _duplicated_6
10 Levels: _duplicated_2 _duplicated_3 _duplicated_4 _duplicated_5 _duplicated_6 ... _duplicated_10
\end{Soutput}
\end{Schunk}
```

Now, data is ready to be analyzes so we can use the same functions as it has been in the previous section.