coeficiente de correlaciones

Pamela Michel De La Cruz Vargas

2022-09-30

```
library (readxl)
```

```
penguins <-read_excel("penguins.xlsx")
```

```
BD<-read_excel("penguins.xlsx")
```

#_____ COEFICIENTE DE CORRELACIÓN _____

Para datos con distribucion normal

Area: Estadistica Parametrica

Utilizamos la matriz "penguins.xlsx"

#_____ # Coeficiente de correlacion de Pearson #_____

#1.- Seleccionamos las variables que vayamos a correlacionar

```
str(penguins)
```

```
## tibble [344 x 9] (S3: tbl_df/tbl/data.frame)
##  $ ID             : chr [1:344] "i1" "i2" "i3" "i4" ...
##  $ especie        : chr [1:344] "Adelie" "Adelie" "Adelie" "Adelie" ...
##  $ isla           : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
##  $ largo_pico_mm  : num [1:344] 39.1 39.5 40.3 37.8 36.7 39.3 38.9 39.2 34.1 42 ...
##  $ grosor_pico_mm : num [1:344] 18.7 17.4 18 18.1 19.3 20.6 17.8 19.6 18.1 20.2 ...
##  $ largo_aleta_mm : num [1:344] 181 186 195 190 193 190 181 195 193 190 ...
##  $ masa_corporal_g: num [1:344] 3750 3800 3250 3700 3450 ...
##  $ genero         : chr [1:344] "male" "female" "female" "female" ...
##  $ año            : num [1:344] 2007 2007 2007 2007 2007 ...
```

```
penguins$especie
```

```
##   [1] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##   [7] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [13] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [19] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [25] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [31] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [37] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [43] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [49] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [55] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
```

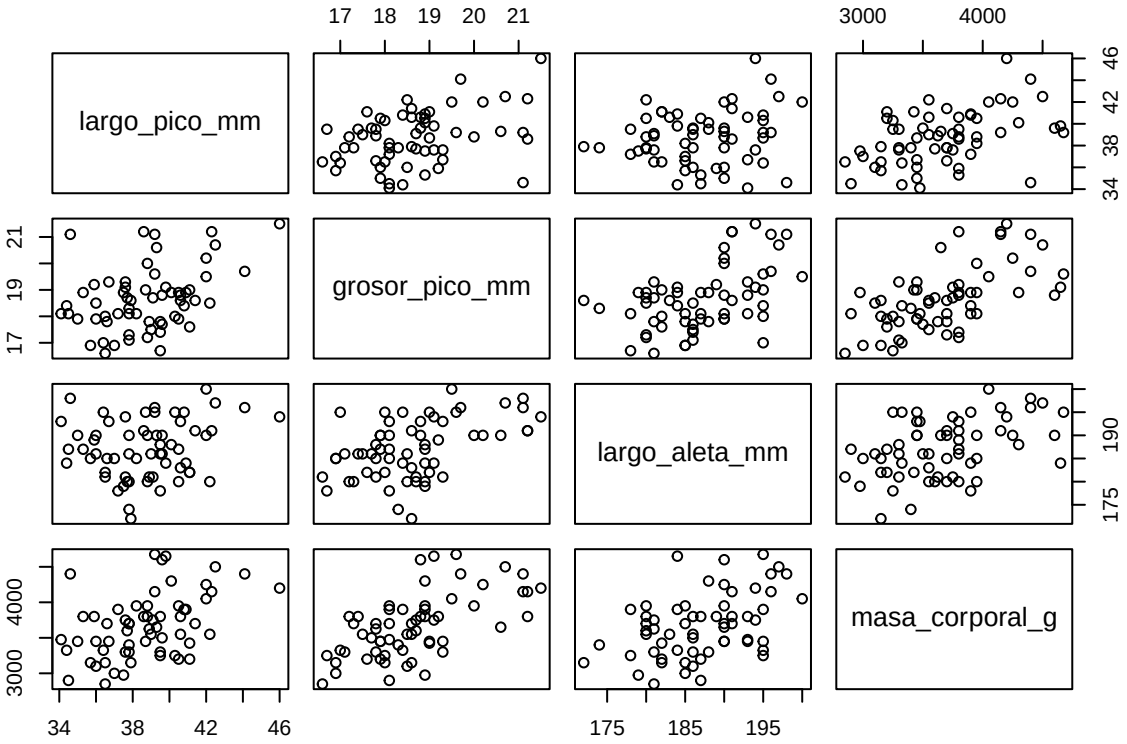
```
##  [61] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [67] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [73] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [79] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [85] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [91] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
##  [97] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [103] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [109] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [115] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [121] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [127] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [133] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [139] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [145] "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"    "Adelie"
## [151] "Adelie"    "Adelie"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [157] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [163] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [169] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [175] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [181] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [187] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [193] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [199] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [205] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [211] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [217] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [223] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [229] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [235] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [241] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [247] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [253] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [259] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [265] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [271] "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"    "Gentoo"
## [277] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [283] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [289] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [295] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [301] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [307] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [313] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [319] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [325] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [331] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [337] "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap" "Chinstrap"
## [343] "Chinstrap" "Chinstrap"
```

2.- Se seleccionan las filas 1 a la 61 que corresponden a la especie Adeli y las variables cuantitativas.

```
adeli<-penguins[1:61,4:7]
```

#4.- Generacion del grafico de correlacion

```
plot(adeli)
```



5.- Calculo de la correlacion de pearson

```
cor(adeli)
```

```
##                 largo_pico_mm grosor_pico_mm largo_aleta_mm masa_corporal_g
## largo_pico_mm       1.0000000      0.3778875      0.1766987       0.4535845
## grosor_pico_mm      0.3778875      1.0000000      0.4760336       0.6144894
## largo_aleta_mm      0.1766987      0.4760336      1.0000000       0.4458517
## masa_corporal_g     0.4535845      0.6144894      0.4458517       1.0000000
```

6.- Organizacion visual de la tabla de correlaciones

#6.1.- Se genera un nuevo objeto con el nombre de pearson

```
pearson<-cor(adeli)
```

#6.2.- Se abre la libreria knitr

```
library(knitr)
```

#6.3.- Se utiliza la funcion kable

```
kable(pearson)
```

	largo_pico_mm	grosor_pico_mm	largo_aleta_mm	masa_corporal_g
largo_pico_mm	1.0000000	0.3778875	0.1766987	0.4535845
grosor_pico_mm	0.3778875	1.0000000	0.4760336	0.6144894
largo_aleta_mm	0.1766987	0.4760336	1.0000000	0.4458517
masa_corporal_g	0.4535845	0.6144894	0.4458517	1.0000000

#———————————— # Correlacion de Spearman #————————————

Para datos con distribucion NO Normal

Area: Estadistica NO Parametrica

Se utiliza la matriz marvel_dc.csv

```
library(readxl)
marvel <- read_excel("marvel.xlsx")
```

1.- Identificar las variables cuantitativas

```
str(marvel)
```

```
## tibble [39 x 11] (S3: tbl_df/tbl/data.frame)
##  $ ID                 : num [1:39] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Original Title     : chr [1:39] "Iron Man" "The Incredible Hulk" "Iron Man 2" "Thor" ...
##  $ Company            : chr [1:39] "Marvel" "Marvel" "Marvel" "Marvel" ...
##  $ Rate               : num [1:39] 7.9 6.7 7 7 6.9 8 7.2 6.9 7.7 8 ...
##  $ Metascore          : num [1:39] 79 61 57 57 66 69 62 54 70 76 ...
##  $ Minutes            : chr [1:39] "126" "112 " "124 " "115" ...
##  $ Release            : num [1:39] 2008 2008 2010 2011 2011 ...
##  $ Budget             : chr [1:39] "140000000" "150000000" "200000000" "150000000 " ...
##  $ Opening Weekend USA: num [1:39] 9.86e+07 5.54e+07 1.28e+08 6.57e+07 6.51e+07 ...
##  $ Gross USA          : num [1:39] 3.19e+08 1.35e+08 3.12e+08 1.81e+08 1.77e+08 ...
##  $ Gross Worldwide    : num [1:39] 5.85e+08 2.63e+08 6.24e+08 4.49e+08 3.71e+08 ...
```

2.- Nombre y posicion de la variable

```
colnames(marvel)
```

```
##  [1] "ID"                 "Original Title"     "Company"
##  [4] "Rate"               "Metascore"          "Minutes"
##  [7] "Release"            "Budget"             "Opening Weekend USA"
## [10] "Gross USA"          "Gross Worldwide"
```

3.- Seleccionar las variables:

rate, minutos, budget y gross.worldwide

```
marvel<-marvel[,c(1,4,5,7,9,10,11)]
```

4.- Verificar que el nombre de las variables

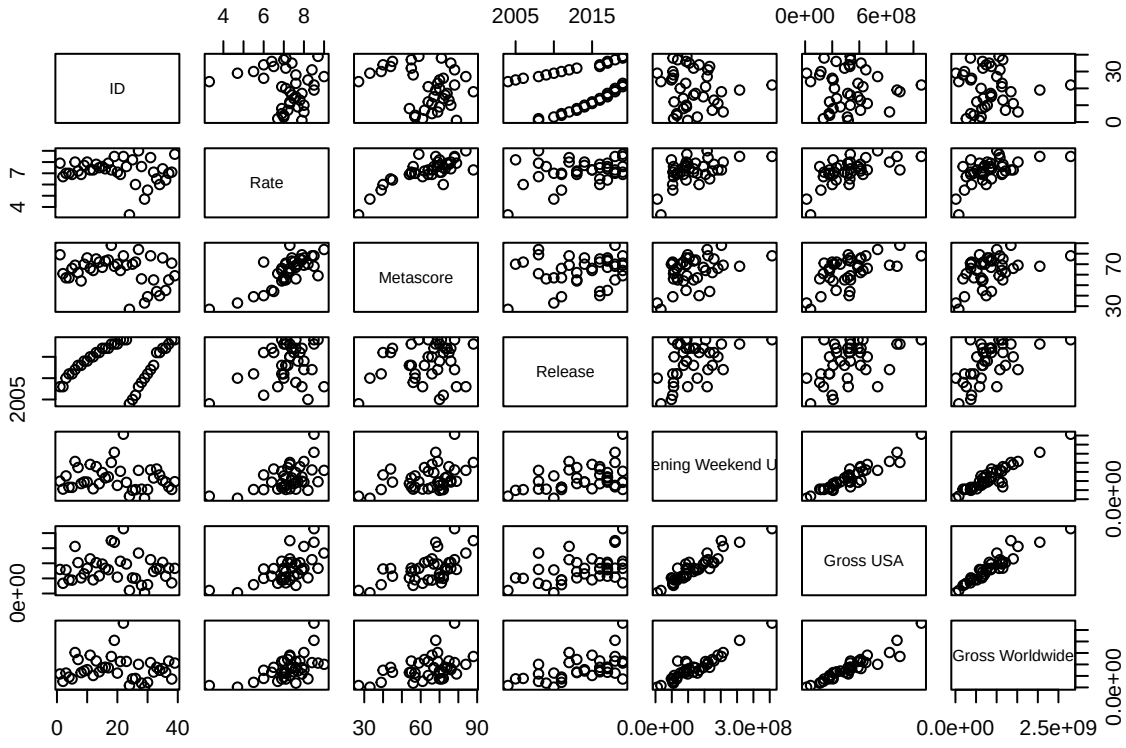
esten correctas

```
colnames(marvel)
```

```
## [1] "ID"                   "Rate"                 "Metascore"
## [4] "Release"              "Opening Weekend USA" "Gross USA"
## [7] "Gross Worldwide"
```

5.- Realizar un plot de exploracion

```
plot(marvel)
```



6.- Realizar la correlacio n de spearman

```
spearman<-cor(marvel, method = "spearman")
```

6.1.- Visualizar el objeto

```
spearman
```

```
##                                ID       Rate  Metascore   Release
## ID                    1.000000000 -0.1080315 -0.1991088 0.3171127
## Rate                 -0.108031503  1.0000000  0.6938601 0.2044872
## Metascore            -0.199108807  0.6938601  1.0000000 0.1794438
## Release               0.317112707  0.2044872  0.1794438 1.0000000
## Opening Weekend USA -0.146761134  0.4711430  0.3733037 0.4003281
## Gross USA            -0.008704453  0.5830256  0.5201540 0.4818162
## Gross Worldwide      -0.013967611  0.5289085  0.3926474 0.5795613
##                      Opening Weekend USA    Gross USA Gross Worldwide
## ID                            -0.1467611 -0.008704453     -0.01396761
## Rate                           0.4711430  0.583025550      0.52890846
## Metascore                      0.3733037  0.520154036      0.39264743
## Release                        0.4003281  0.481816231      0.57956132
## Opening Weekend USA            1.0000000  0.897975709      0.87793522
## Gross USA                      0.8979757  1.000000000      0.95364372
## Gross Worldwide                0.8779352  0.953643725      1.00000000
```

#7.- Se utiliza la funcion kable para tabla en # formato markdown.

```
kable(spearman)
```

	ID	Rate	Metascore	Release	Opening Weekend USA	Gross USA	Gross Worldwide
ID	1.0000000	-0.1080315	-0.1991088	0.3171127	-0.1467611	-0.0087045	-0.0139676
Rate	-0.1080315	1.0000000	0.6938601	0.2044872	0.4711430	0.5830256	0.5289085
Metascore	-0.1991088	0.6938601	1.0000000	0.1794438	0.3733037	0.5201540	0.3926474
Release	0.3171127	0.2044872	0.1794438	1.0000000	0.4003281	0.4818162	0.5795613
Opening Weekend USA	-0.1467611	0.4711430	0.3733037	0.4003281	1.0000000	0.8979757	0.8779352
Gross USA	-0.0087045	0.5830256	0.5201540	0.4818162	0.8979757	1.0000000	0.9536437
Gross Worldwide	-0.0139676	0.5289085	0.3926474	0.5795613	0.8779352	0.9536437	1.0000000