

Correlaciones

Katheryn Lizet Vázquez Rodríguez

2022-06-27

Utilidad

Esta metodología sirve para identificar la relación entre dos variables de tipo **cuantitativo**. Los datos se distribuyen de forma normal. Area: Estadística Paramétrica y se utiliza la matriz “penguins.xlsx” como ejemplo.

librerías

```
library(readxl)
```

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

#1.- Exploración de variables

1.- Dimensión

```
dim(penguins)
```

```
## [1] 344 9
```

2.- Tipos de variables

```
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 ...
```

3.- Identificar la especie Adelie

```
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"
```

[illegible]

4.- Seleccionar las observaciones de la 1 a la 152 y las variables cuantitativas.

```
colnames(penguins)
```

```
## [1] "ID"           "especie"       "isla"          "largo_pico_mm"  
## [5] "grosor_pico_mm" "largo_aleta_mm" "masa_corporal_g" "genero"  
## [9] "año"
```

```
adelie<-penguins[1:152,4:7]
```

verificación de la matriz

```
str(adelie)
```

```
## tibble [152 x 4] (S3: tbl_df/tbl/data.frame)  
## $ largo_pico_mm : num [1:152] 39.1 39.5 40.3 37.8 36.7 39.3 38.9 39.2 34.1 42 ...  
## $ grosor_pico_mm : num [1:152] 18.7 17.4 18 18.1 19.3 20.6 17.8 19.6 18.1 20.2 ...  
## $ largo_aleta_mm : num [1:152] 181 186 195 190 193 190 181 195 193 190 ...  
## $ masa_corporal_g: num [1:152] 3750 3800 3250 3700 3450 ...
```

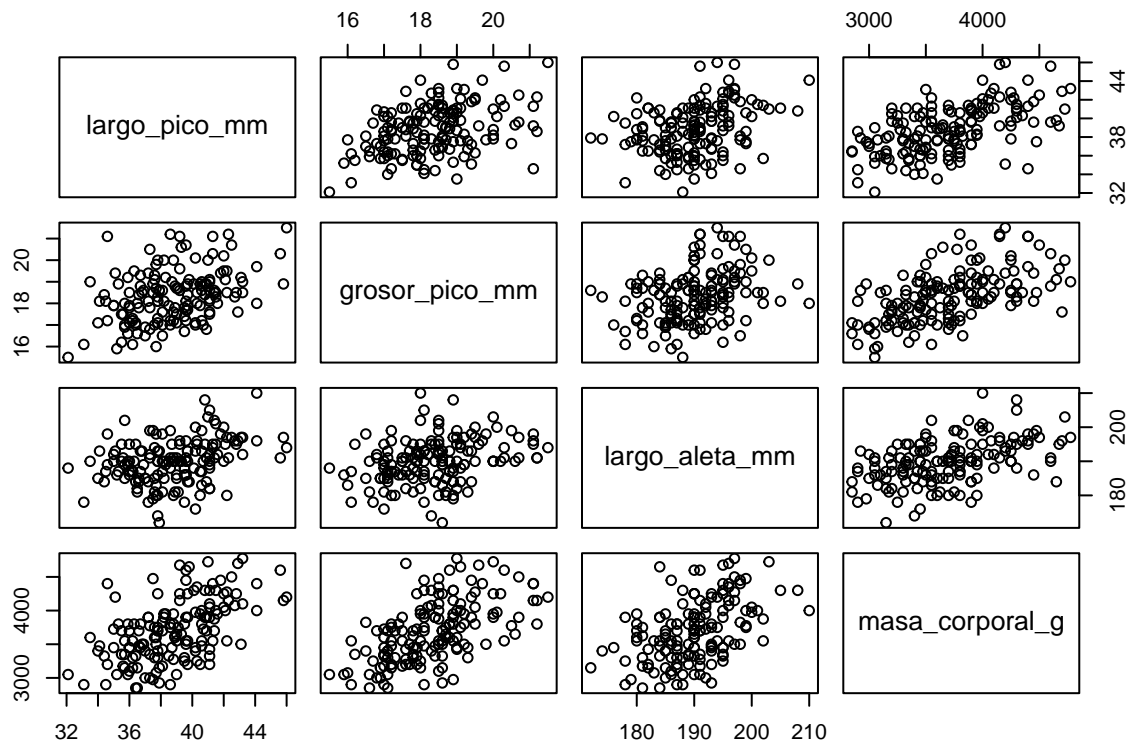
Metodología

1.- Calculo de la correlacion de pearson

```
cor_adelie<-cor(adelie)
```

2.- Realización del plot de correlaciones

```
plot(adelie)
```



3.- Organización de la tabla de correlaciones

librerias

```
library(knitr)
```

Organización

```
kable(cor_adelie)
```

	largo_pico_mm	grosor_pico_mm	largo_aleta_mm	masa_corporal_g
largo_pico_mm	1.0000000	0.3917580	0.3256178	0.5486177
grosor_pico_mm	0.3917580	1.0000000	0.3075689	0.5760619
largo_aleta_mm	0.3256178	0.3075689	1.0000000	0.4682015
masa_corporal_g	0.5486177	0.5760619	0.4682015	1.0000000

Correlación de Spearman

Librerías

```
library(readxl)
```

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

#1.- Exploración de variables

1.- Dimensión

```
dim(marvel)
```

```
## [1] 39 11
```

2.- 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 ...
```

3.- Nombre y posición 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"
```

4.- Seleccionar las variables: rate, metascore, minutes, opening weekend USA, gross USA y gross worldwide

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

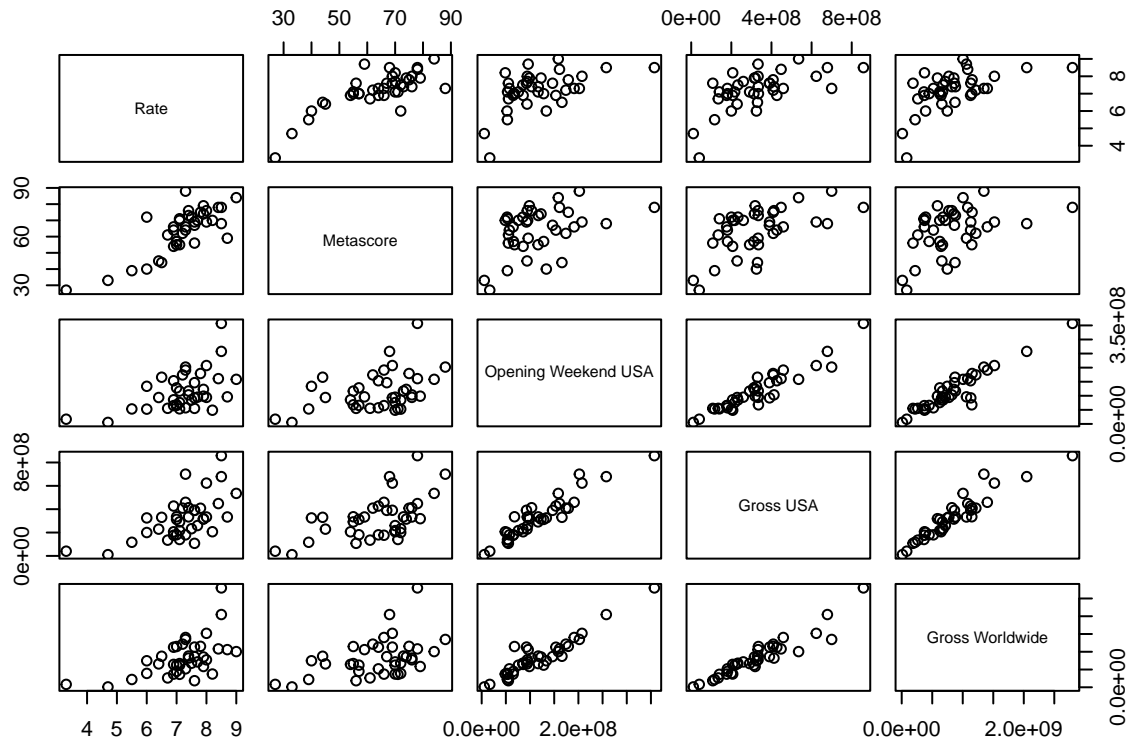
5.- Verificar que el nombre de las variables esten correctas

```
colnames(marvel)
```

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

5.- Realizar un plot de exploracion

```
plot(marvel)
```



6.- Realizar la correlación de spearman

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

6.1.- Visualizar el objeto

```
spearman
```

```
##           Rate Metascore Opening Weekend USA Gross USA
## Rate      1.0000000 0.6938601                0.4711430 0.5830256
## Metascore  0.6938601 1.0000000                0.3733037 0.5201540
## Opening Weekend USA 0.4711430 0.3733037                1.0000000 0.8979757
## Gross USA  0.5830256 0.5201540                0.8979757 1.0000000
## Gross Worldwide 0.5289085 0.3926474                0.8779352 0.9536437
##           Gross Worldwide
## Rate      0.5289085
## Metascore  0.3926474
## Opening Weekend USA 0.8779352
## Gross USA  0.9536437
## Gross Worldwide 1.0000000
```

Librerias

```
library(knitr)
```

Se utiliza la función kable para el formato en markdown

```
kable(spearman)
```

	Rate	Metascore	Opening Weekend USA	Gross USA	Gross Worldwide
Rate	1.0000000	0.6938601	0.4711430	0.5830256	0.5289085
Metascore	0.6938601	1.0000000	0.3733037	0.5201540	0.3926474
Opening Weekend USA	0.4711430	0.3733037	1.0000000	0.8979757	0.8779352
Gross USA	0.5830256	0.5201540	0.8979757	1.0000000	0.9536437
Gross Worldwide	0.5289085	0.3926474	0.8779352	0.9536437	1.0000000