

Actividad. Correlación lineal

Daniel Orellana Pérez

1) Define brevemente el concepto de correlación lineal.

La correlación lineal se trata de una medida estadística que indica la relación entre dos variables cuantitativas, donde un cambio en una de estas se relaciona en el cambio de la segunda. Esta correlación lineal se mediría en una escala de -1 (correlación negativa) a 1 (correlación positiva). Entre ambos parámetros encontraríamos el 0, que viene a significar la total ausencia de correlación.

2) ¿Por qué decimos que la correlación lineal es una prueba de correlación paramétrica? ¿En qué se diferencian las pruebas paramétricas de las no paramétricas?

Decimos que la correlación lineal es una prueba de correlación paramétrica puesto que la relación entre los datos que provienen de una distribución normal y sus variables se pueden expresar mediante las matemáticas y gráficos.

La diferencia entre las pruebas paramétricas y las no paramétricas es que estas no necesitas, al contrario de las paramétricas, aunque es posible que en ocasiones puedan tener menor poder estadístico que las primeras, no requiere que haya relación entre los diferentes datos.

3) Calcula la correlación entre las variables almacenadas en la tabla 'data'.

```
getwd()
nuevo_dir <- "C:/practicacorreacion"
setwd(nuevo_dir)
if(file.exists(nuevo_dir)) {
  cat("Directorio creado correctamente: ", nuevo_dir, "\n")
} else {
  cat("Fallo al crear directorio: ", nuevo_dir, "\n")
}

library(readxl)
View(data)
print(data)
```

```
#Actividad 3
correlacion_variables <- cor(data)
print(correlacion_variables)
```

```
> correlacion_variables <- cor(data)
> print(correlacion_variables)
```

	id	peso	altura
id	1.000000000	0.08321171	0.003792534
peso	0.083211709	1.000000000	0.147768260
altura	0.003792534	0.14776826	1.000000000

4) Calcula los coeficientes de correlación de las variables junto con el nivel de significancia (p-value) en 1 solo gráfico. Interpreta los resultados.

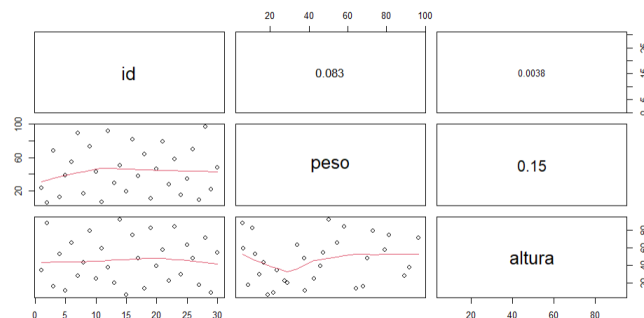
```
#Actividad 4
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...) {
  usr <- par("usr")
  on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
```

```

Cor <- abs(cor(x, y))
txt <- paste0(prefix, format(c(Cor, 0.123456789), digits = digits)[1])
if(missing(cex.cor)) {
  cex.cor <- 0.4 / strwidth(txt)
}
text(0.5, 0.5, txt,
     cex = 1 + cex.cor*Cor)
}

pairs(data,
      upper.panel = panel.cor,
      lower.panel = panel.smooth)

```



- Es importante mencionar que en la consola cuando le doy a correr al programa, me sale el siguiente mensaje que, a mi parecer, hace referencia a lo marcado en amarillo. Por tanto, no se si este ejercicio es correcto del todo.

5) Emplea una función para obtener en una matriz de correlación lineal, IC 95% y p-value de todas las variables en el data frame 'data'.

#Actividad 5

```

library(correlation)
matriz <- correlation(data)
matriz
print(matriz)

```

p-value adjustment method: Holm (1979)
Observations: 30
> print(matriz)
Correlation Matrix (pearson-method)

Parameter1	Parameter2	r	95% CI	t(28)	p
id	peso	0.08	[-0.29, 0.43]	0.44	> .999
id	altura	3.79e-03	[-0.36, 0.36]	0.02	> .999
peso	altura	0.15	[-0.22, 0.48]	0.79	> .999

p-value adjustment method: Holm (1979)
Observations: 30

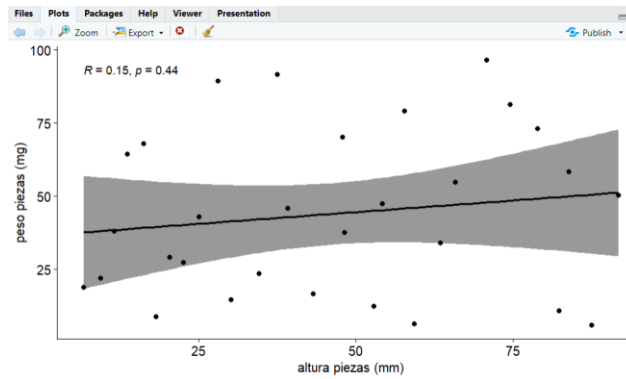
6) Visualiza gráficamente la correlación lineal existente entre las variables 'longitud' y 'peso'.

#Actividad 6

```

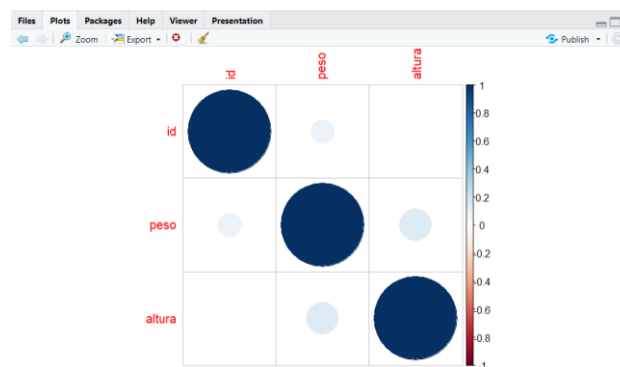
library(ggpubr)
library(ggplot2)
ggscatter(data, x = "longitud", y = "peso",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "altura piezas (mm)", ylab = "peso piezas (mg)")

```



7) Emplea la librería ``corrplot()`` para visualizar la correlación entre variables.

```
library(corrplot)
corrplot(cor(data))
```



8) A partir de la siguiente secuencia de valores numéricos:

- Distancia (km): 1.1,100.2,90.3,5.4,57.5,6.6,34.7,65.8,57.9,86.1

- Número de cuentas (valor absoluto): 110,2,6,98,40,94,31,5,8,10

a. Crea 2 vectores: 'distancia' y 'n_piezas' para almacenarlos en un data frame

#A)

```
distancia <- c( 1.1,100.2,90.3,5.4,57.5,6.6,34.7,65.8,57.9,86.1)
```

```
n_piezas <- c(110,2,6,98,40,94,31,5,8,10)
```

```
datos_act8 <- data.frame(distancia, n_piezas)
```

```
print(datos_act8)
```

```
> distancia <- c( 1.1,100.2,90.3,5.4,57.5,6.6,34.7,65.8,57.9,86.1)
> n_piezas <- c(110,2,6,98,40,94,31,5,8,10)
> datos_act8 <- data.frame(distancia, n_piezas)
> print(datos_act8)
  distancia n_piezas
1      1.1      110
2    100.2         2
3     90.3         6
4      5.4        98
5     57.5        40
6      6.6        94
7     34.7        31
8     65.8         5
9     57.9         8
10    86.1        10
```

b. Calcula el coeficiente de correlación

#B

```
correlacion_datos_act8 <- cor(datos_act8)
```

```
print(correlacion_datos_act8)
```

```
> correlacion_datos_act8 <- cor(datos_act8)
> print(correlacion_datos_act8)
```

	distancia	n_piezas
distancia	1.0000000	-0.9249824
n_piezas	-0.9249824	1.0000000

c. Calcula el nivel de significancia

```
significancia_datos_act8 <- cor.test(datos_act8$distancia, datos_act8$n_piezas)$p.value
print(significancia_datos_act8)
> significancia_datos_act8 <- cor.test(datos_act8$distancia, datos_act8$n_piezas)$p.value
> print(significancia_datos_act8)
[1] 0.0001264706
```

d. Calcula el Intervalo de confianza al 95% en relación con el coeficiente de correlación

```
intervaloconfianza_datos_act8 <- cor.test(datos_act8$distancia, datos_act8$n_piezas)$conf.int
print(intervaloconfianza_datos_act8)
```

```
> intervaloconfianza_datos_act8 <- cor.test(datos_act8$distancia, datos_act8$n_piezas)$conf.int
> print(intervaloconfianza_datos_act8)
[1] -0.9824414 -0.7072588
attr(,"conf.level")
[1] 0.95
```

e. ¿Qué intensidad y dirección presentan ambas variables?

Presenta una baja intensidad en una gráfica lineal ascendente.

f. ¿Es significativa esta relación?

Esto nos podría indicar un patrón y una clara tendencia al alza según la gráfica lineal.

g. Resulta apropiado afirmar la correlación (o no) entre variables con un tamaño muestral tan reducido (n=10).

Con un tamaño muestral tan reducido (n=10) no sería del todo apropiado afirmar una correlación puesto que dicho tamaño no presenta una gran variable estadística puesto que faltaría más representación. Si bien, sí que se podría afirmar una correlación, hay que ser muy cautelosos a la hora de realizar una interpretación de los resultados y así hay que dejarlo constatado en nuestro informe.

9) Explicame con un ejemplo en R la diferencia entre una relación lineal y monótona entre 2 variables.

La diferencia entre una relación lineal y una relación monótona en R es que, tras poner los diferentes parámetros encontramos es que en la primera presenta una línea de regresión lineal que se ajusta bien de los parámetros de los datos, mientras que la segunda relación, se nos presentaría de una forma no lineal.

#Actividad 9

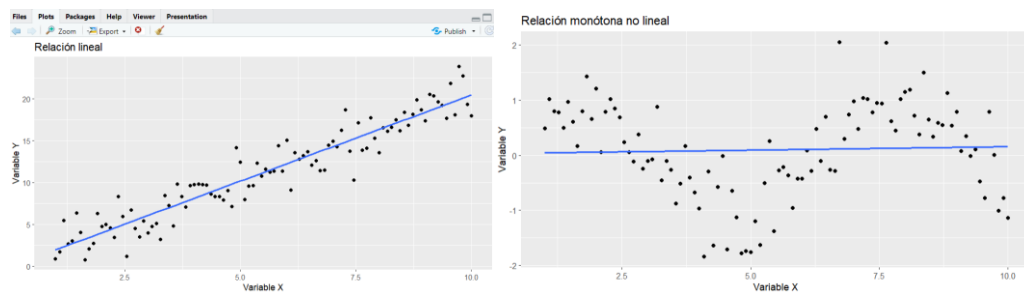
```
library(ggplot2)
set.seed(123)
x_linear <- seq(1, 10, length.out = 100)
y_linear <- 2 * x_linear + rnorm(100, sd = 2)
```

```
x_monotonic <- seq(1, 10, length.out = 100)
y_monotonic <- sin(x_monotonic) + rnorm(100, sd = 0.5)

data_linear <- data.frame(x = x_linear, y = y_linear)
data_monotonic <- data.frame(x = x_monotonic, y = y_monotonic)
```

```
ggplot(data_linear, aes(x = x, y = y)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Relación lineal", x = "Variable X", y = "Variable Y")
```

```
ggplot(data_monotonic, aes(x = x, y = y)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Relación monótona no lineal", x = "Variable X", y = "Variable Y")
```



10) ¿Qué tipo de prueba de correlación se aplica a las variables que experimentan una relación monótona? Expón un ejemplo en R

La prueba que se aplica a las variables que experimentan una relación monótona suele ser la correlación no paramétrica, en concreto el coeficiente de correlación de Spearman (ρ), la cual se encarga de la relación de los rangos entre dos variables no paramétricas.

Para calcular esta correlación se ordenan los diferentes datos y se le asignan rangos calculándose posteriormente rangos en vez de los datos originales. Como resultado nos debe dar un coeficiente entre -1 y 1. Para R se utiliza la función “cor.test”.

```
set.seed(123)
x_monotonic <- seq(1, 10, length.out = 100)
y_monotonic <- sin(x_monotonic) + rnorm(100, sd = 0.5)
data_monotonic <- data.frame(x = x_monotonic, y = y_monotonic)
correlation <- cor.test(data_monotonic$x, data_monotonic$y, method = "spearman")
print(correlation)
```

```
Console Terminal Background Jobs
R 4.3.2 · C:/practicacorrelacion/
> set.seed(123)
> x_monotonic <- seq(1, 10, length.out = 100)
> y_monotonic <- sin(x_monotonic) + rnorm(100, sd = 0.5)
> data_monotonic <- data.frame(x = x_monotonic, y = y_monotonic)
> correlation <- cor.test(data_monotonic$x, data_monotonic$y, method = "spearman")
> print(correlation)

Spearman's rank correlation rho

data: data_monotonic$x and data_monotonic$y
S = 161390, p-value = 0.7549
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.03156316
```