

CLASE 05 - PERMUTACIÓN Y CORRELACIÓN.

OCE 313 - Técnicas de análisis no paramétricos

Dr. José Gallardo Matus | <https://genomics.pucv.cl/>

11 April 2022

PLAN DE LA CLASE

1.- Introducción

- ▶ Permutación: ¿Qué es? ¿Por qué es importante?.
- ▶ Permutación: caso 1, 2 y 3.
- ▶ Correlación no paramétrica.
- ▶ Calculo de correlación no paramétrica.
- ▶ Distribución muestral por permutación.
- ▶ Prueba de hipótesis de correlación no paramétrica.

2). Práctica con R y Rstudio cloud.

- ▶ Realizar cálculo de permutación usando R.
- ▶ Realizar prueba de correlación.
- ▶ Realizar gráficas avanzadas con ggplot2.

PERMUTACIÓN

¿Qué es?

Una permutación es una combinación ordenada de elementos.

¿Por qué es importante?

El concepto matemático de permutación está subyacente a muchos métodos de análisis no paramétricos.

¿Qué haremos?

Calcularemos las posibles permutaciones de los elementos de un conjunto de datos de un experimento aleatorio y evaluaremos si se acepta o rechaza la hipótesis de las pruebas paramétricas más comunes.

PRÁCTICA PERMUTACIÓN

1.- Si para el conjunto $\{a,b\}$ existen 2 permutaciones a-b y b-a

¿Cuántas permutaciones existen para conjunto $\{a,b,c\}$?

$$P_3 = 3! = 3 \times 2 \times 1 = 6$$

```
factorial(3)
```

```
## [1] 6
```

```
permutations(3,3,letters[1:3])
```

```
##      [,1] [,2] [,3]
## [1,] "a"  "b"  "c"
## [2,] "a"  "c"  "b"
## [3,] "b"  "a"  "c"
## [4,] "b"  "c"  "a"
## [5,] "c"  "a"  "b"
## [6,] "c"  "b"  "a"
```

CASO 2 - PERMUTACIÓN CON REPETICIÓN

- ▶ ¿Cuántas claves diferentes existen en el candado de 3 filas?

Para 3 filas con 10 números {0 al 9} existen $10 \times 10 \times 10$ permutaciones = 1000.

$$10^3$$

[1] 1000



CASO 3 - PERMUTACIÓN CON REPETICIÓN

- ▶ ¿Cuántas permutaciones/palabras se forman la palabra GATA? Ej. {TAGA, ...}
- ▶ Si, $G = 1$ vez; $T = 1$ vez y $A = 2$ veces , entonces

```
factorial(4) / (factorial(1)*factorial(1)*factorial(2))
```

```
## [1] 12
```

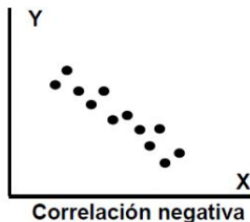
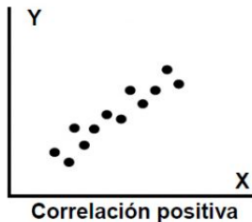


PRUEBA DE CORRELACIÓN NO PARAMÉTRICA

¿Para que sirve?

Para estudiar asociación de dos variables, cuando no se cumple uno o varios supuestos de la correlación paramétrica:

- ▶ Las variables X e Y no son continuas.
- ▶ No existe relación lineal.
- ▶ La distribución conjunta de (X, Y) no es una distribución Bivariable normal.

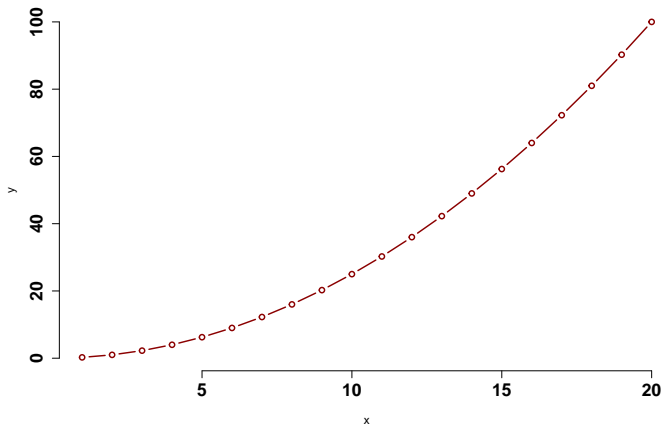


EJEMPLO FUNCIÓN MONÓTONA

¿Cuál es el supuesto que no se cumple?

No existe una relación lineal

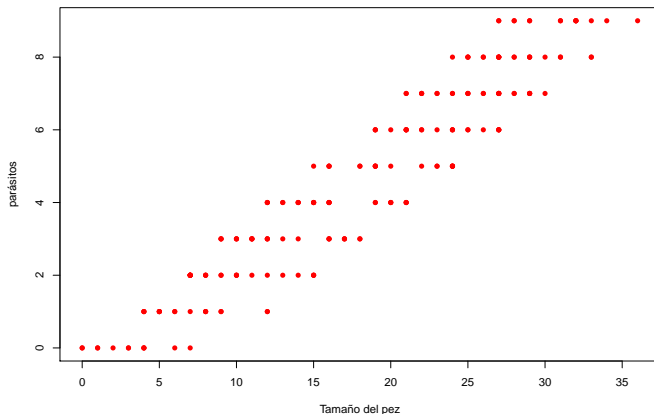
Relación no lineal



EJEMPLO VARIABLES DISCRETAS U ORDINALES

¿Cuál es el supuesto que no se cumple?

Parásitos es variable discreta.



CORRELACIÓN NO PARAMÉTRICA

- ▶ Se basa en calcular el ranking de las variables.
- ▶ Calculamos ranking para cada variable.

Fish size (X)	Parásitos (Y)	Ranking X	Ranking Y
942	13	4	2
101	14	1	3
313	18	2	4
800	10	3	1

- ▶ Si la correlación es +, valores ordenados.
- ▶ Si la correlación es -, valores en orden inverso.
- ▶ Si la correlación es 0, valores desordenados.

COEFICIENTE DE CORRELACIÓN DE SPEARMAN

¿Cómo se calcula?

Ranking X	Ranking Y	d	d^2
4	2	2	4
1	3	-2	4
2	4	-2	4
3	1	2	4

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} =$$

$$\sum d^2 = 16$$

$$\rho = 1 - \frac{6 * 16}{4(4^2 - 1)} =$$

$$rho = -0,6$$

OTRAS CORRELACIONES POSIBLES

Opción 1: Correlación negativa.

Ranking X	Ranking Y
4	1
1	4
2	3
3	2
$\rho = -1$	

Opción 2: Correlación positiva.

Ranking X	Ranking Y
4	4
1	1
2	2
3	3
$\rho = 1$	

¿CUÁNTAS CORRELACIONES SON POSIBLES?

- ▶ Calculamos número de permutaciones/correlaciones.

```
factorial(4)
```

```
## [1] 24
```

- ▶ Las 24 permutaciones/correlaciones corresponden a nuestro espacio muestral para 4 pares de variables.
- ▶ Esto es independiente de las variables utilizadas.

ESPACIO MUESTRAL

- ▶ En nuestro experimento

$$\rho = -0.6$$

- ▶ 1 de 24 correlaciones posibles.

-1.0	-0.8	-0.8	-0.8	-0.6	-0.4	-0.4	-0.4
-0.4	-0.2	-0.2	0.0	0.0	0.2	0.2	0.4
0.4	0.4	0.4	0.6	0.8	0.8	0.8	1.0

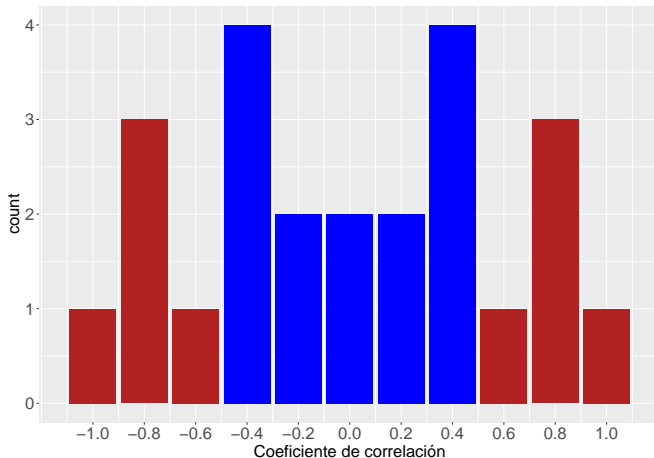
PRUEBA DE HIPÓTESIS DE CORRELACIÓN

Hipótesis	Verdadera cuando
H_0 : X e Y mutuamente independientes	$\rho = 0$
H_1 : X e Y no son mutuamente independientes	$\rho \neq 0$

- ▶ ¿Cuántas correlaciones son ≥ 0.6 y ≤ -0.6 ?

DISTRIBUCIÓN MUESTRAL

- ▶ Solo por azar 10 correlaciones podrían tomar estos valores.



PRUEBA DE HIPÓTESIS DE CORRELACIÓN

Hipótesis	Verdadera cuando
H_0 : X e Y mutuamente independientes	$\rho = 0$
H_1 : X e Y no son mutuamente independientes	$\rho \neq 0$

$$p = 10 / 24$$

$$p = 0.4167$$

No se rechaza H_0 porque $p = 0,416$ es mayor a $0,05$

PRUEBA DE CORRELACIÓN CON R

```
# Crea objetos X e Y  
X <- c(942,101,313,800)  
Y <- c(13,14,18,10)
```

```
# Realiza test de correlación  
cor.test(X,Y, method = "spearman",  
          alternative = "two.sided")
```

```
##  
## Spearman's rank correlation rho  
##  
## data: X and Y  
## S = 16, p-value = 0.4167  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho  
## -0.6
```

RESUMEN DE LA CLASE

- ▶ Revisión de conceptos de permutación.
- ▶ Recordatorio correlación Pearson.
- ▶ Funciones monótonas.
- ▶ Uso de permutación para calcular espacio y distribución muestral.
- ▶ Aplicación interpretación prueba de correlación de Spearman con R.