

Clase 19 Pruebas no paramétricas

Diplomado en Análisis de Datos con R e Investigación reproducible para Biociencias.

Dr. José Gallardo Matus & Dra. Angélica Rueda Calderón

Pontificia Universidad Católica de Valparaíso

10 November 2022

PLAN DE LA CLASE

1.- Introducción

- ▶ ¿Qué son las pruebas no paramétricas?.
- ▶ Test de Correlación no paramétrico.
- ▶ Pruebas de contraste no paramétrico.
- ▶ Prueba de asociación Chi cuadrado.

2.- Práctica con R y Rstudio cloud

- ▶ Realizar pruebas no paramétricas.
- ▶ Realizar gráficas avanzadas con ggplot2.

MÉTODOS NO PARAMÉTRICOS

- ▶ Conjunto diverso de pruebas estadísticas.
- ▶ El concepto de “no paramétrico” a veces es confuso, pues los métodos no paramétricos si estiman y someten a prueba hipótesis usando parámetros, pero no los de distribución normal.
- ▶ Se aplican usualmente para variables cuantitativas que no cumplen con el supuesto de normalidad y para variables cualitativas.
- ▶ Alternativamente se conocen como métodos de distribución libre.
- ▶ El concepto matemático de permutación está subyacente a muchos métodos no paramétricos y se utiliza para someter a prueba las hipótesis.

SUPUESTOS DE LOS MÉTODOS NO PARAMÉTRICOS

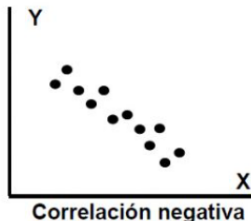
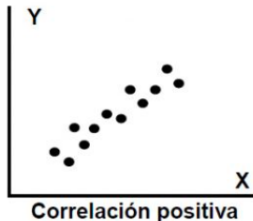
- ▶ Las variables son independientes.
- ▶ Muestras independientes con idéntica distribución.
- ▶ No tienen supuestos acerca de la distribución de la variable (algunas asumen chi-cuadrado).
- ▶ La distribución del estadístico se estima muy a menudo por permutación.

PRUEBA DE CORRELACIÓN NO PARAMÉTRICA

¿Para que sirve?

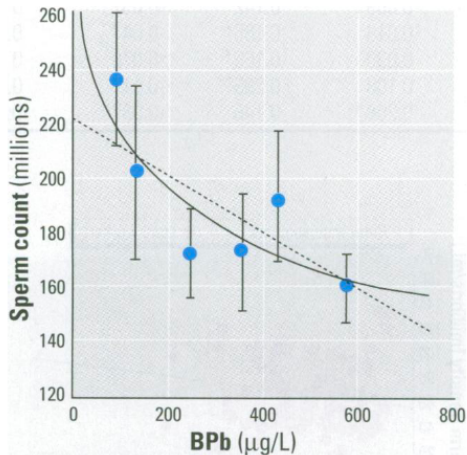
Para estudiar asociación de dos variables, cuando no se cumple uno o varios supuestos de la correlación paramétrica:

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



CASO 1: Nº ESPERMIOS - PLOMO SANGUÍNEO

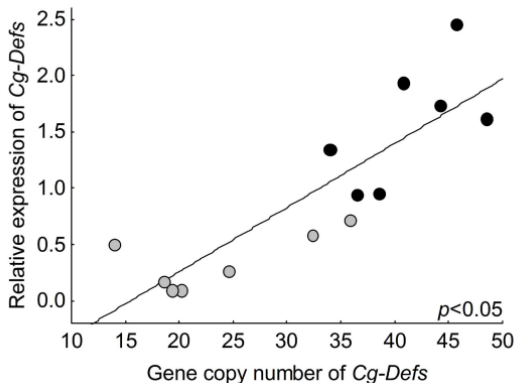
¿Cuáles son los supuestos que no se cumplen?



Fuente: Telisman et al. 2000

CASO 2: COPIAS GEN DEFENSINA - EXPRESIÓN

¿Cuál es el supuesto que no se cumple?



Fuente: Schmitt et al. 2013

CORRELACIÓN NO PARAMÉTRICA

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

Plomo sangre (X)	Nº espermios (Y)	Ranking X	Ranking Y
742	170	4	2
101	180	1	3
313	210	2	4
600	160	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

- Recuerde que el muestreo aleatorio podría generar diferentes resultados.

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 el número de permutaciones/correlaciones para 4 elementos.

```
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

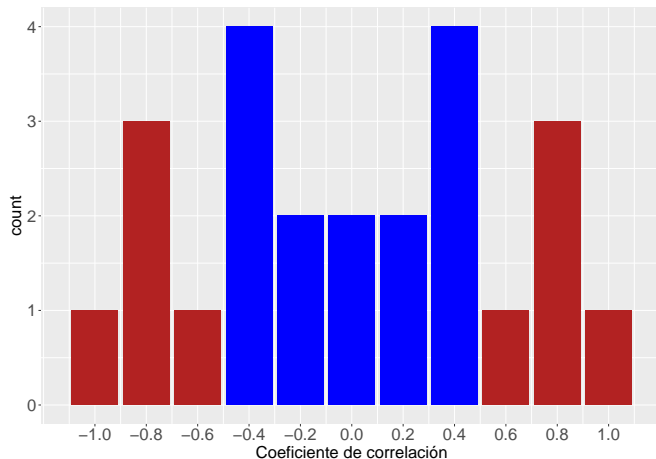
$$\rho = -0.6$$

- ▶ 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

DISTRIBUCIÓN MUESTRAL DE CORRELACIÓN

¿Cuántas correlaciones son ≥ 0.6 y ≤ -0.6 ?



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(742,101,313,600)
```

```
Y <- c(170,180,210,160)
```

```
# 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
```

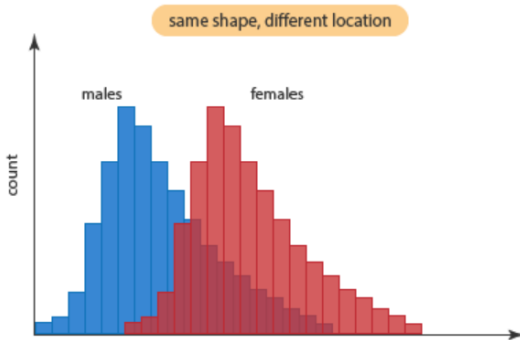
```
## -0.6
```

COMPARACIÓN DE MUESTRAS INDEPENDIENTES

¿Para qué sirve?

Para comparar dos muestras con idéntica distribución, con diferentes medianas y sin normalidad.

Usualmente para variables discretas.



PRUEBA DE MANN-WHITNEY (W)

Estudio de caso: Formación de biofilm (μm^2) de *Staphylococcus epidermidis* en presencia de plasma humano. Fuente: Skovdal et al. 2021

Tratamiento con plasma (T)	Control sin plasma (C)
9	4
12	5
13	6

CÁLCULO ESTADÍSTICO MANN-WHITNEY (W)

¿Cómo se calcula el estadístico W?

Como la diferencia de los ranking entre tratamiento y control

Tratamiento (T)	Control (C)	Ranking T	Ranking C
9	4	4	1
12	5	5	2
13	6	6	3
		$\Sigma = 15$	$\Sigma = 6$

$$W = 15 - 6 = 9$$

Máxima diferencia posible entre T y C.

¿CUÁNTAS COMBINACIONES SON POSIBLES?

¿Cuántas combinaciones son posibles?

$$6! / 3! \times 3! = 720 / 36 = 20$$

2 resultados posibles de 20

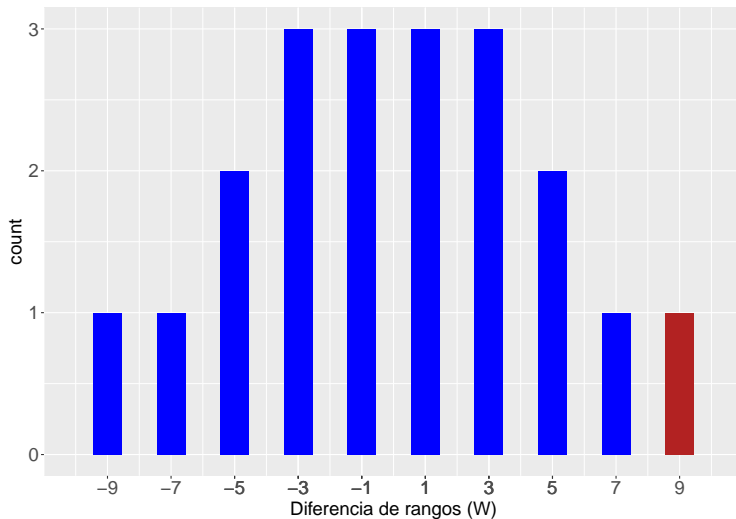
Control mayor que tratamiento.

T	C
1	4
2	5
3	6
6	15
$W =$	- 9

Tratamiento mayor que Control.

T	C
2	1
5	3
6	4
13	8
$W =$	5

DISTRIBUCIÓN MUESTRAL DE W



PRUEBA DE HIPÓTESIS DE MANN-WHITNEY

Hipótesis

H_0 : Tratamiento = Control

H_1 : Tratamiento > Control

Resultado obtenido $W=9$.

$$p = 1/20$$

$$p = 0.05$$

No se rechaza H_0 porque $p = 0,05$

PRUEBA DE MANN-WHITNEY CON R

```
# Crea objetos tratamiento y control
```

```
t <- c(9, 12, 13)
```

```
c <- c(0, 4, 6)
```

```
# Realiza prueba de Mann-Whitney
```

```
wilcox.test(t, c, alternative = "g",  
            paired = FALSE)
```

```
##
```

```
## Wilcoxon rank sum exact test
```

```
##
```

```
## data: t and c
```

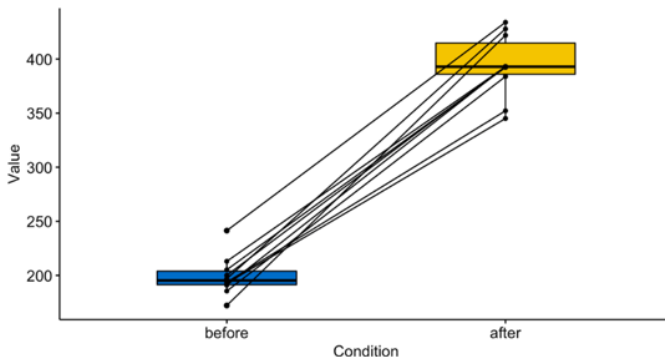
```
## W = 9, p-value = 0.05
```

```
## alternative hypothesis: true location shift is greater t
```

COMPARACIÓN DE MUESTRAS PAREADAS

¿Para que sirve?

Para comparar dos muestras *pareadas* con idéntica distribución, con diferentes medianas y sin normalidad.



PRUEBA DE WILCOXON MUESTRAS PAREADAS

Estudio de caso: Gonadotrofina en trucha 7 y 14 días **post ovulación**.

¿Aumenta la gonadotrofina post ovulación?

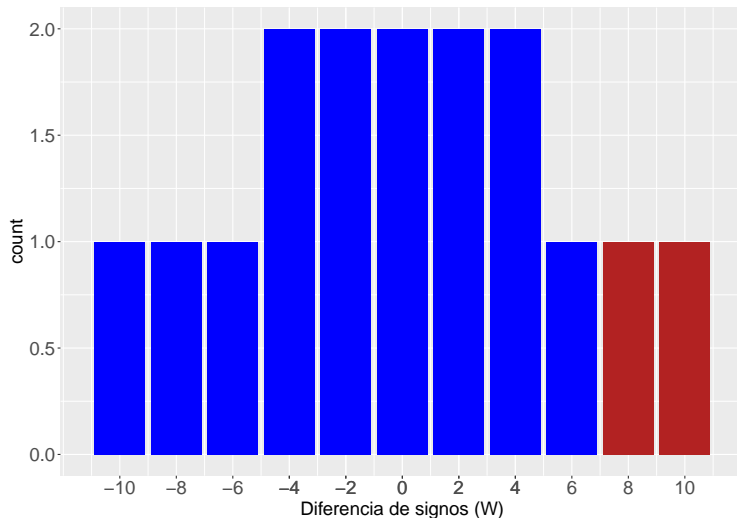
Trucha	7 días	14 días	d	Ranking con signo
1	45	49	4	2
2	41	50	9	4
3	47	52	5	3
4	52	50	2	-1

W = suma de los ranking = 8

V = suma de casos positivos (aumenta) = 9

DISTRIBUCIÓN MUESTRAL DE W

¿Cuántas combinaciones de signos (+ o -) son posibles? $2^4 = 16$



PRUEBA DE HIPÓTESIS DE WILCOXON

Hipótesis

$$H_0: d = 0$$

$$H_1: d > 0$$

$$p = 2/16$$

$$p = 0,125$$

No se rechaza H_0 porque $p = 0,125$ es mayor a $0,05$

PRUEBA DE WILCOXON PAREADAS CON R

```
# Crea objetos pre y post  
pre <- c(45, 41, 47, 52)  
post <- c(49, 50, 52, 50)  
# Realiza prueba de Wilcoxon  
wilcox.test(post - pre, alternative = "greater")
```

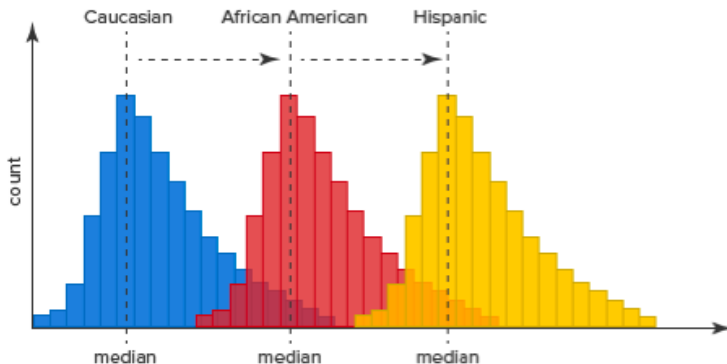
```
##  
## Wilcoxon signed rank exact test  
##  
## data: post - pre  
## V = 9, p-value = 0.125  
## alternative hypothesis: true location is greater than 0
```

```
# no es necesario indicar muestras pareadas  
# pues estamos haciendo la resta en la función.
```

COMPARACIÓN DE MÚLTIPLES MUESTRAS INDEPENDIENTES

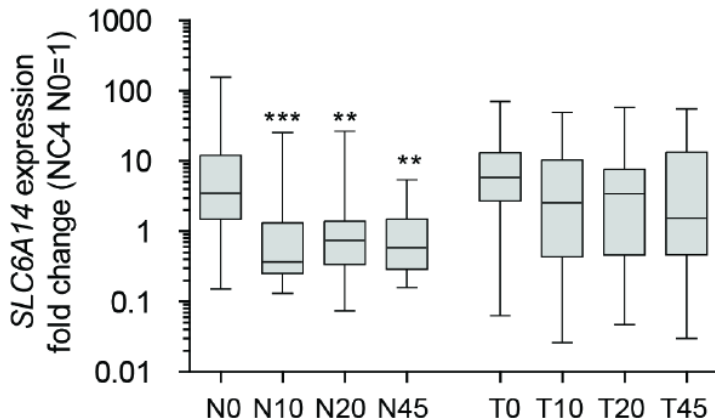
¿Para que sirve?

Para comparar múltiples muestras con idéntica distribución, con diferentes medianas y sin normalidad.



ESTUDIO DE CASO: EXPRESIÓN SLC6A14 EN RESPUESTA A ISQUEMIA.

Lange et al. 2015. Respuesta a isquemia en tejido normal (N) y tumoral (T) sometido a isquemia por 0, 10, 20 y 45 min.



PRUEBA DE KRUSKAL - WALLIS CON R

Hipótesis

H_0 : La distribución de los k grupos son iguales.

H_1 : Al menos 2 grupos son distintos.

```
# Realiza prueba de kruskal
```

```
kruskal.test(list(N0, N10, N20, N45)) %>% pander()
```

Table 15: Kruskal-Wallis rank sum test: list(N0, N10, N20, N45)

Test statistic	df	P value
39.54	3	1.331e-08 * * *

PRUEBA DE DUNN PARA COMPARACIONES MULTIPLES

```
# Realice prueba de dunn  
dunnTest(FC ~ Time,  
         data=tidy,  
         method="bonferroni")
```

##	Comparison	Z	P.unadj	P.adj
## 1	N0 - N10	5.4970805	3.861305e-08	2.316783e-07
## 2	N0 - N20	4.6903882	2.726872e-06	1.636123e-05
## 3	N10 - N20	-0.8066923	4.198438e-01	1.000000e+00
## 4	N0 - N45	5.0886287	3.606620e-07	2.163972e-06
## 5	N10 - N45	-0.4084518	6.829420e-01	1.000000e+00
## 6	N20 - N45	0.3982405	6.904529e-01	1.000000e+00

PRUEBA DE ASOCIACIÓN VARIABLES CATEGÓRICAS

¿Para que sirve?

Se utilizan para investigar la asociación de dos o más variables categóricas una de las cuales es una variable respuesta y la otra es una variable predictora.

Tratamiento	Respuesta +	Respuesta -
Si	a	c
No	b	d

¿Cómo se calcula el estadístico Chi cuadrado?

$$\chi^2 = \sum \frac{(\text{freq.obs.} - \text{freq.esp.})^2}{(\text{freq.esperada})} = \sum \frac{(O - E)^2}{(E)}$$

PRUEBA DE CHI CUADRADO

Esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula.

Hipótesis

H₀: La variable predictora y la variable respuesta son independientes
(Tratamiento = control)

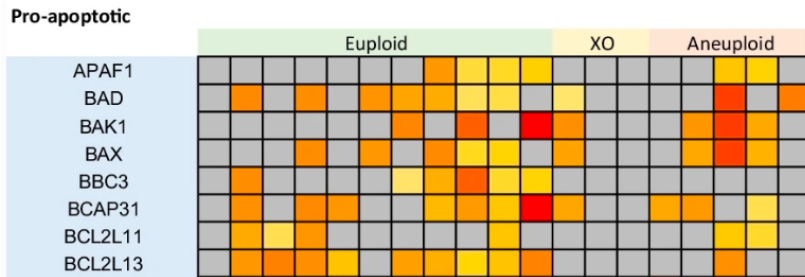
H₁: La variable predictora y la variable respuesta NO son independientes

Supuestos:

- Los datos provienen de una muestra aleatoria de la población de interés.
- El tamaño de muestra es lo suficientemente grande para que el número esperado en las categorías sea mayor 5 y que ninguna frecuencia sea menor que 1.

ESTUDIO DE CASO: PREDICCIÓN MOLECULAR DE EMBRIONES ANEUPLOIDES

Lal et al. 2022. Expresión del gen BCL2L13 en embriones euploides y aneuploides.



PRUEBA CHI CUADRADO

```
datos
```

```
##           Expresado No expresado
## Euploides           9           2
## Aneuploides          1           7
```

```
# Test de Chi-squared en R (chisq.test)
test<-chisq.test(datos, correct = FALSE)

test %>% pander()
```

Table 18: Pearson's Chi-squared test: datos

Test statistic	df	P value
8.927	1	0.00281 * *

PRÁCTICA ANÁLISIS DE DATOS

- ▶ Guía de trabajo práctico disponible en Rstudio.cloud.
Clase_19

RESUMEN DE LA CLASE

Revisión de conceptos de estadística no paramétrica.

- ▶ Correlación de Spearman.
- ▶ Prueba de Man-Whitney.
- ▶ Prueba de Wilcoxon.
- ▶ Prueba de Kruskal Wallis + DUNN test.
- ▶ Prueba de Chi-cuadrado.