

# Clase 07 Pruebas no paramétricas

Curso Introducción al Análisis de datos con R para la  
Acuicultura.

Dr. José Gallardo Matus

Pontificia Universidad Católica de Valparaíso

22 July 2022

## 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 prueban 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

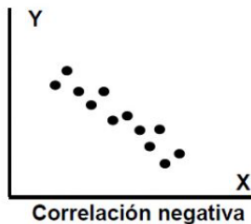
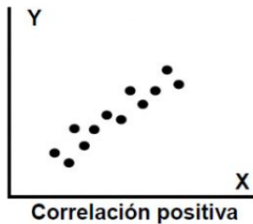
- ▶ Las variables son independientes.
- ▶ Muestras independentes tienen idéntica distribución.
- ▶ No tienen supuestos acerca de la distribución de la variable.
- ▶ La distribución muestral se estima a veces usando permutación.

# 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.



# CORRELACIÓN NO PARAMÉTRICA

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

<b>Fish size (X)</b>	<b>Parásitos (Y)</b>	<b>Ranking X</b>	<b>Ranking Y</b>
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 <sup>2</sup>
4	2	2	4
1	3	-2	4
2	4	-2	4
3	1	2	4

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

$$\sum d^2 = 16$$

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

$$rho = -0,6$$

# PRUEBA DE HIPÓTESIS DE CORRELACIÓN

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



# PRUEBA DE CORRELACIÓN CON R

```
# 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
```

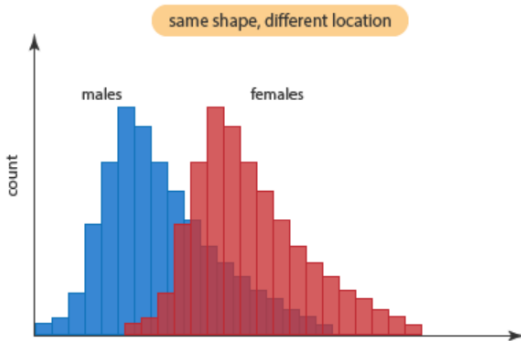
No se rechaza  $H_0$  porque  $p = 0,416$  es mayor a  $0,05$

# 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: Conducta agresiva (minutos de pelea) entre tilapias cultivadas en estanque mosexo (Tratamiento: solo machos) y estanques mixtos (Control: Machos y hembras) **Link**

Tratamiento (T)	Control (C)
9	0
12	4
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	0	4	1
12	4	5	2
13	6	6	3
		$\Sigma = 15$	$\Sigma = 6$

$$W = 15 - 6 = 9$$

Máxima diferencia posible entre T y C.

# PRUEBA DE HIPÓTESIS DE MANN-WHITNEY

---

## Hipótesis de 2 colas

---

**H<sub>0</sub>:** Tratamiento = Control

**H<sub>1</sub>:** Tratamiento  $\neq$  Control

---

---

## Hipótesis 1 cola

---

**H<sub>0</sub>:** Tratamiento = Control

**H<sub>1</sub>:** Tratamiento  $>$  Control

---

---

## Hipótesis 1 cola

---

**H<sub>0</sub>:** Tratamiento = Control

**H<sub>1</sub>:** Tratamiento  $<$  Control

---

# PRUEBA DE MANN-WHITNEY CON R

```
# 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
```

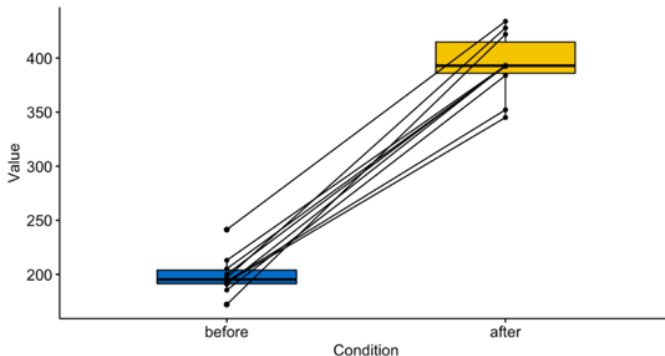
```
## alternative hypothesis: true location shift is greater t
```

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

# 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	<b>2</b>
2	41	50	9	<b>4</b>
3	47	52	5	<b>3</b>
4	52	50	2	<b>-1</b>

$W$  = suma de los ranking = 8

$V$  = suma de casos positivos (aumenta) = 9



# PRUEBA DE HIPÓTESIS DE WILCOXON

---

## Hipótesis 2 colas

---

$$H_0: d = 0$$

$$H_1: d \neq 0$$

---

---

## Hipótesis 1 cola

---

$$H_0: d = 0$$

$$H_1: d < 0$$

---

---

## Hipótesis 1 cola

---

$$H_0: d = 0$$

$$H_1: d > 0$$

---

# PRUEBA DE WILCOXON PAREADAS CON R

```
# 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
```

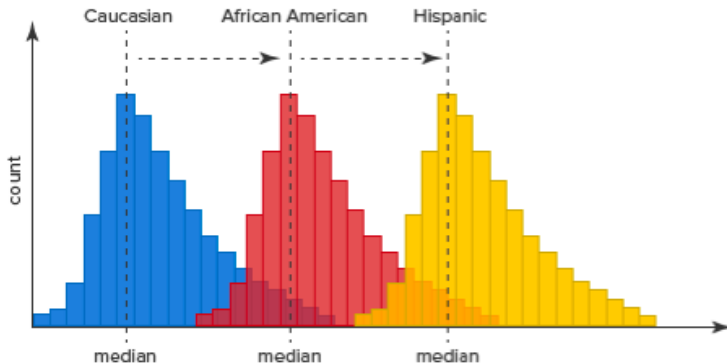
```
# no es necesario indicar muestras pareadas
```

```
# 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: SCORE CALIDAD CAMARÓN

Score de calidad organoléptica (textura) de camarón **link**.

Descripción	Puntaje
Muy compacto y denso	9
Menos elástico, compacto y denso	7
No elástico, no compacto y no denso	5
Ligeramente blando	3
Suave	1

# PRUEBA DE KRUSKAL - WALLIS

Textura luego de 0, 4 y 8 días de almacenamiento de camarón congelado.

0 días	4 días	8 días
9	7	6
8	7	5
9	6	5
8	8	6

---

## Hipótesis

---

**H<sub>0</sub>**: La distribución de los k grupos son iguales.

**H<sub>1</sub>**: Al menos 2 grupos son distintos.

---

# PRUEBA DE KRUSKAL - WALLIS CON R

```
kruskal.test(list(d0, d4, d8))
```

```
##
```

```
##  Kruskal-Wallis rank sum test
```

```
##
```

```
## data:  list(d0, d4, d8)
```

```
## Kruskal-Wallis chi-squared = 9, df = 2, p-value = 0.0111
```

# 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

# PRUEBA DE CHI CUADRADO

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

---

## Hipótesis

---

**H<sub>0</sub>:** La variable predictora y la variable respuesta son independientes (Tratamiento = control)

**H<sub>1</sub>:** 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: SOBREVIVENCIA MANCHA BLANCA CAMARÓN

Sobrevivencia de postlarvas alimentadas con *B* glucanos y desafiadas con WSSP **Chang et al., 1999.**

Tratamiento	Sobrevivientes	Muertos
Con glucanos	20	80
Sin glucanos	5	95

# CÁLCULO DE ESTADÍSTICO CHI CUADRADO

**¿Cómo se calcula el estadístico Chi cuadrado?**

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

Frecuencia esperada

```
##      [,1] [,2]
```

```
## [1,] 12.5 87.5
```

```
## [2,] 12.5 87.5
```

```
## X-squared
```

```
## 10.28571
```

# PRUEBA DE CHI CUADRADO CON R

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

```
##  
## Pearson's Chi-squared test  
##  
## data:  datos  
## X-squared = 10.286, df = 1, p-value = 0.001341
```

- ▶ Guía de trabajo práctico disponible en [Rstudio.cloud](https://Rstudio.cloud).  
**Clase\_07**

# 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.
- ▶ Prueba de Chi-cuadrado.