

# Clase 10 Pruebas no paramétricas

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

Dr. José A. Gallardo | jose.gallardo@pucv.cl | Pontificia  
Universidad Católica de Valparaíso

20 July 2021

# 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.
- ▶ Elaborar un reporte dinámico en formato pdf.

# 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.

# SUPUESTOS: MÉTODOS NO PARAMÉTRICOS

El principal supuesto de los métodos no paramétricos es que las variables aleatorias son independientes y con idéntica distribución.

Usualmente no tienen supuestos acerca de la distribución de la variable, 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 una hipótesis.

# 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:

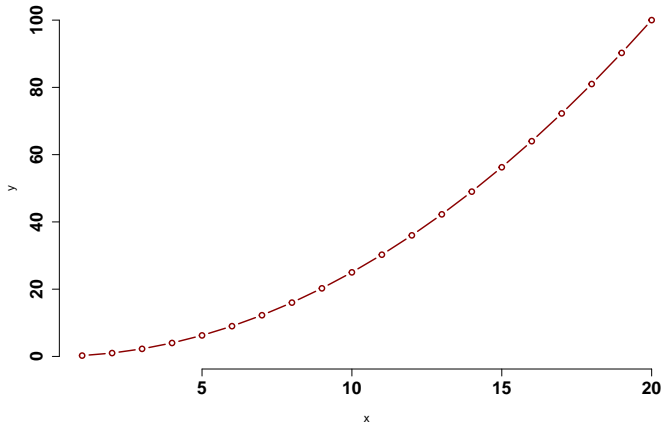
- Las variables  $X$  e  $Y$  no son continuas.
- No existe relación lineal.
- La distribución conjunta de  $(X, Y)$  no es una distribución Bivariable normal.

# EJEMPLO FUNCIÓN MONÓTONA

**¿Cuál es el supuesto que no se cumple?**

No existe una relación lineal

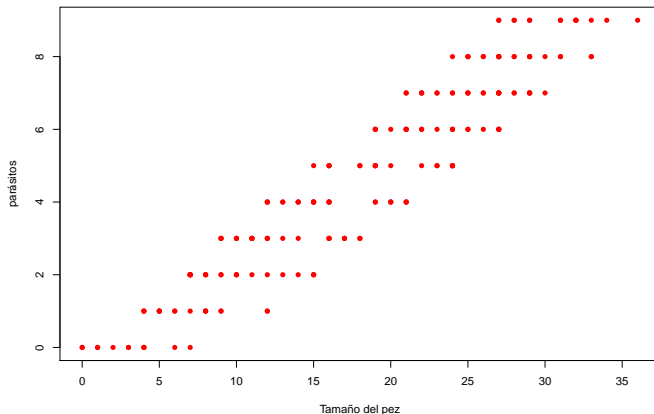
Relación no lineal



# EJEMPLO VARIABLES DISCRETAS U ORDINALES

**¿Cuál es el supuesto que no se cumple?**

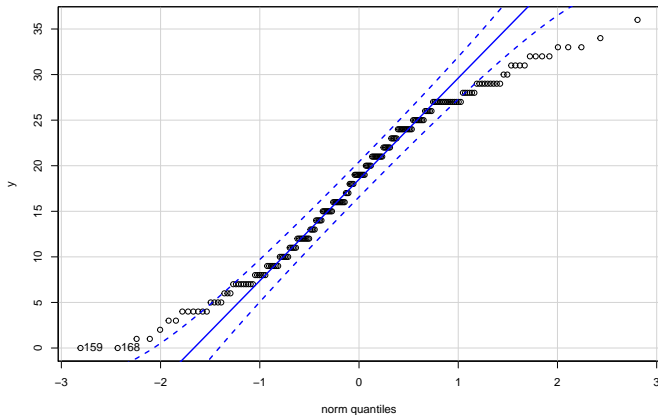
Parásitos es variable discreta.



# EJEMPLO VARIABLES DISCRETAS U ORDINALES 2

**¿Cuál es el supuesto que no se cumple?**

Parásitos no tiene distribución normal



## [1] 159 168



# COEFICIENTE DE CORRELACIÓN DE SPEARMAN

## ¿Cómo se calcula?

Coeficiente de correlación de Spearman ( $\rho = \text{rho}$ )

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = -0,6$$

Fish size (X)	Parásitos (Y)	Ranking X	Ranking Y	d	d <sup>2</sup>
942	13	4	2	2	4
101	14	1	3	-2	4
313	18	2	4	-2	4
800	10	3	1	2	4

$$\sum d^2 = 16$$

# ¿CUÁNTAS CORRELACIONES SON POSIBLES?

Ejemplo correlación -1 Y 1

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

# DISTRIBUCIÓN MUESTRAL DE CORRELACIÓN

Permutación: combinación ordenada de elementos.

$P_4 = 4! = 4 \times 3 \times 2 \times 1 = 24$  permutaciones posibles.

¿Cuántas combinaciones son mayores o igual a nuestros datos?

**-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	$\rho \neq 0$

$$p = 10 / 24 \quad p = 0.4167$$

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
```

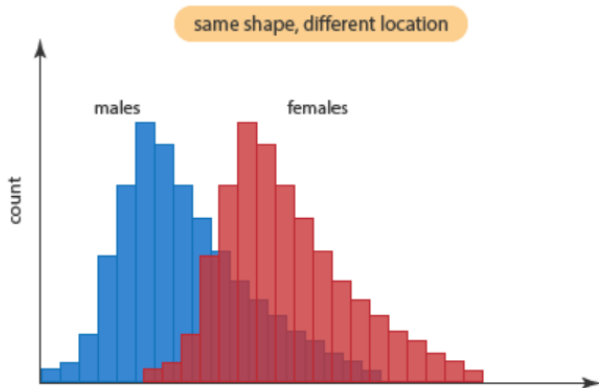
```
## -0.6
```

# 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 DE 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.



# DISTRIBUCIÓN MUESTRAL DE W

¿Cuántas combinaciones son posibles?

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

3 resultados posibles de 20

<b>T</b>	<b>C</b>	<b>T</b>	<b>C</b>	<b>T</b>	<b>C</b>
1	4	1	2	2	1
2	5	4	3	5	3
3	6	5	6	6	4
<i>6</i>	<i>15</i>	<i>10</i>	<i>11</i>	<i>13</i>	<i>8</i>
<i>W =</i>	<b>- 9</b>	<i>W =</i>	<b>- 1</b>	<i>W =</i>	<b>5</b>

# PRUEBA DE HIPÓTESIS DE MANN-WHITNEY

---

**Hipótesis**

---

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

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

---

Resultado obtenido  $W=9$ .

$$p = 1/20$$

$$p = 0.05$$

No se rechaza **H<sub>0</sub>** 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 test
```

```
##
```

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

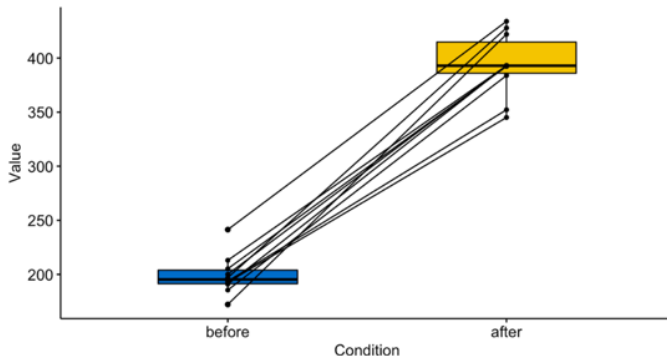
```
## W = 9, p-value = 0.05
```

```
## 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**

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>

**Hipótesis**

**$H_0$ :  $d = 0$**

**$H_1$ :  $d > 0$**

# DISTRIBUCIÓN MUESTRAL DE W

¿Cuántas combinaciones de signos (+ o -) son posibles?

$$2^4 = 16$$

¿Cuántas combinaciones son mayores o igual a nuestros datos?

-10, -8, -6, -4, -4, -2, -2, 0, 0, 2, 2, 4, 4, 6, **8, 10**

$$p = 2/16$$

$$p = 0,125$$

No se rechaza **H<sub>0</sub>** 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")
```

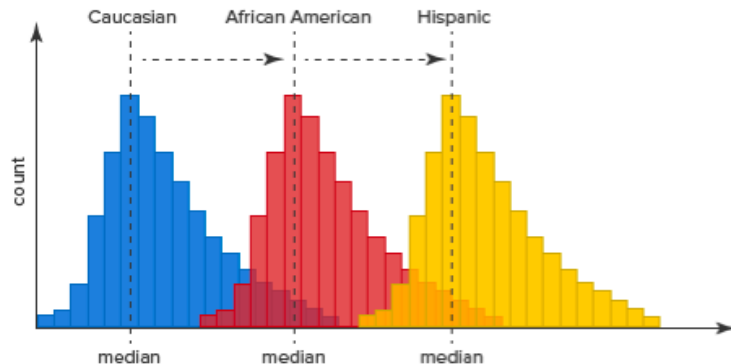
```
##  
## Wilcoxon signed rank test  
##  
## data: post - pre  
## V = 9, p-value = 0.125  
## alternative hypothesis: true location is greater than 0
```

```
# 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**.

<b>Descripción</b>	<b>Puntaje</b>
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

```
d0 <- c(9,8,9,8) # day0  
d4 <- c(7,7,6,8) # day4  
d8 <- c(6,5,5,6) # day8  
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 **link**.

Tratamiento	Sobrevivientes	Muertos
Con glucanos	20	80
Sin glucanos	5	95

# CÁLCULO DE ESTADÍSTICO CHI CUADRADO

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

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

Frecuencia esperada

```
##      [,1] [,2]  
## [1,] 12.5 87.5  
## [2,] 12.5 87.5
```

```
## X-squared  
## 10.28571
```

# PRUEBA DE CHI CUADRADO CON R

```
# Crea matriz de datos  
datos <- c(20, 5, 80, 95)  
dim(datos) <- c(2,2)  
# 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
```



# PRÁCTICA ANÁLISIS DE DATOS

- ▶ Guía de trabajo práctico disponible en drive y Rstudio.cloud.

## **Clase\_10**

- ▶ El trabajo práctico se realiza en Rstudio.cloud.

## **Clase 10 - Estadística no paramétrica**

# 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.