

Unidad III

PRUEBAS ESTADÍSTICAS PARA EVALUAR UNA MUESTRA PAREADA Y DOS MUESTRAS INDEPENDIENTES

"Learning is not compulsory...neither is survival"

W. Edwards Deming

Introducción

En muchas investigaciones se tiene como principal objetivo realizar comparaciones de procedimientos o métodos. Por ejemplo, se desea comparar: si un grupo de estudiantes han incrementado su rendimiento en la segunda práctica calificada en comparación con la primera práctica calificada, si un grupo de empleados al utilizar un método A son más eficientes que al utilizar un método B, si los hombres tienen un CI diferente al de las mujeres, si los empleados de un país A ganan menos que los empleados de un país B.

Para poder analizar los datos de los dos primeros ejemplos, se debe seleccionar una muestra aleatoria de tamaño n y realizar mediciones en dos momentos diferentes o con dos procedimientos distintos. A este tipo de muestra se le denomina muestra pareada o relaciona y para la recolección de los datos se debe hacer un estudio longitudinal en dos etapas.

Para poder analizar los datos de los dos últimos ejemplos, se debe seleccionar dos muestras que sean aleatorias e independientes de tamaños n_1 y n_2 . Es decir, se debe hacer un estudio transversal. Esto se puede lograr siguiendo un muestreo estratificado con dos estratos.

Para el caso de pruebas de una muestra relacionada se puede deducir que un requisito para poder realizar estas pruebas es que se requiere de una muestra aleatoria simple de tamaño n . Algunas de las pruebas que se presentarán en este capítulo son extensiones del caso de una muestra aleatoria. Por ejemplo, uno de los objetivos de estas pruebas es básicamente comparar un parámetro de posición como la mediana.

Para el caso de pruebas de dos muestras independientes un supuesto es que estas muestras sean aleatorias de tamaños n_1 y n_2 .

Por otro lado, desde el punto de vista no paramétrico, han surgido pruebas que tienen como principal característica el uso intensivo de computadoras. Muchos autores están incluyendo este tema en los textos actuales de Estadística No Paramétrica. No queremos ser ajenos a este desarrollo es por eso que en este capítulo se implementará la idea básica de la Prueba de Permutación.

Al igual que en el capítulo anterior donde existen pruebas que tienen el mismo objetivo se desarrollará un procedimiento para comparar dichas pruebas, que permita determinar cuál de ellas es mejor para ciertos escenarios. Para esto, se elaborará funciones en R que facilitarán su entendimiento.

Este capítulo tiene como objetivo desarrollar de manera teórica-práctica algunas de las muchas pruebas no paramétricas que se pueden utilizar para analizar los datos de una variable proveniente de una muestra relacionada o de dos muestras independientes.

PRUEBAS ESTADÍSTICAS PARA EVALUAR UNA MUESTRA RELACIONADA (O PAREADA)

1. Prueba para evaluar una variable dicotómica

A. Prueba de Mc Nemar

➤ Aspectos Generales

Es una prueba aplicable a los diseños “antes-después”, en los cuales cada sujeto se utiliza como su propio control. En estas condiciones puede emplearse para probar la efectividad de un tratamiento particular.

Los datos están compuestos por n pares de sujetos los cuales para análisis se puede representar en una tabla de contingencia 2×2 .

Esta prueba es un caso particular de la prueba Q de Cochran que será desarrollada posteriormente.

➤ Supuesto

- Los datos deben encontrarse en una escala nominal u ordinal y si se trabaja con variables de tipo intervalo o razón estas previamente se deben dicotomizar.

➤ Inferencia Estadística

Hipótesis

H_0 : No existe variación antes y después de aplicado el tratamiento (el efecto del tratamiento no es significativo)

H_1 : Existe variación antes y después de aplicado el tratamiento (el efecto del tratamiento es significativo)

Para llevar a cabo la prueba se debe realizar lo siguiente:

- Clasificar la muestra en una tabla de contingencia 2×2 de la siguiente manera:

		Después	
		+	-
Antes	+	A	B
	-	C	D

+ y – se usan para denotar diferentes respuestas. Note que todos aquellos cambios entre la primera y segunda respuesta aparecen en las celdas superior derecha (de + a -) e inferior izquierda (de – a +) de la tabla. Las entradas en la tabla corresponden a las frecuencias de los resultados asociados. Así, B denota el número de individuos cuyas respuestas fueron + en la primera medición y – en la segunda medición. De manera similar, C es el número de individuos quienes cambiaron de – a +. A es la frecuencia de individuos que respondieron + en ambas ocasiones y D es el número de personas que respondieron – en ambas evaluaciones.

Así $B+C$ es el total de personas cuyas respuestas cambiaron., así que si H_0 es verdadera se espera que en cada una de las celdas de cambios (B y C) se espera una frecuencia de $(B+C)/2$.

- Aplicar el estadístico de prueba

$$\chi_c^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \sim \chi_{[1-\alpha, 1]}^2$$

El cual en términos de la tabla 2x2 propuesta se puede simplificar a:

$$\chi_c^2 = \frac{(B - C)^2}{B + C} \sim \chi_{[1-\alpha, 1]}^2$$

La aproximación por medio de la distribución Chi Cuadrado de la distribución muestral de χ^2 llega a ser más precisa si se hace una corrección por continuidad. La corrección por continuidad es necesaria porque una distribución continua (Chi cuadrado) se utiliza para aproximarse a una distribución discreta χ^2 . Cuando las frecuencias esperadas son pequeñas, la aproximación puede ser muy pobre. El propósito de la corrección es eliminar esta fuente de imprecisión. Con la corrección por continuidad incluida se tiene:

$$\chi_c^2 = \frac{(|B - C| - 1)^2}{B + C} \sim \chi_{[1-\alpha, 1]}^2$$

También el p-valor se puede calcular utilizando la distribución binomial

$$pvalor = 2 \sum_{i=0}^r \binom{B+C}{i} 0.5^{B+C} 0.5^{n-(B+C)} \quad r = \min(B, C)$$

➤ **Aplicación**

En un estudio de la depresión se tomaron los datos antes y después del año de seguimiento con el fin de determinar si la frecuencia de depresión ha variado luego de una terapia realizado por un médico. Los resultados del estudio de una muestra de 26 pacientes elegidos al azar son:

ANTES	DESPUÉS	ANTES	DESPUÉS
depresión	depresión	no depresión	depresión
depresión	depresión	no depresión	depresión
depresión	depresión	no depresión	depresión
depresión	depresión	no depresión	depresión
no depresión	depresión	no depresión	depresión
no depresión	depresión	depresión	no depresión
depresión	depresión	depresión	no depresión
no depresión	depresión	no depresión	no depresión
depresión	depresión	no depresión	no depresión

no depresión	depresión		no depresión	no depresión
depresión	depresión		no depresión	no depresión
depresión	depresión		no depresión	no depresión
no depresión	depresión		no depresión	no depresión

		Después	
		Depresión	No Depresión
Antes	Depresión	8	2
	No Depresión	10	6

Realice la prueba de Mc Nemar, de sus conclusiones a un nivel de significación de 0.05.

H_0 : La frecuencia de depresión no ha variado

H_1 : La frecuencia de depresión ha variado

$\alpha = 0.05$

$p\text{-valor} = 0.039 < \alpha$ se rechaza H_0 .

Conclusión

Existe suficiente evidencia estadística a un nivel de significación de 0.05, para rechazar H_0 .

Por lo tanto, se puede afirmar que la frecuencia de depresión ha variado.

➤ **Funciones en R**

Existe la función `mcnemar.test`

`mcnemar.test(x,y)` o `mcnemar.test(tabla)`

Existe la función `mcnemar.exact` del paquete `exact2x2` que utiliza la distribución binomial.

➤ **Resultados con R**

```
tabla<-matrix(c(8,10,2,6),2,2)
```

```
mcnemar.test(tabla)
```

```
McNemar's Chi-squared test with continuity
correction
data:  datos1[, 1] and datos1[, 2]
McNemar's chi-squared = 4.0833, df = 1, p-value = 0.04331
```

```
library(exact2x2)
mcnemar.exact(tabla)
```

```
Exact McNemar test (with central confidence
intervals)
data:  tabla
```

```
b = 2, c = 10, p-value = 0.03857422
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.02130704421 0.93850198857
sample estimates:
odds ratio
      0.2
```

El odds ratio se obtiene de dividir b/c y se puede interpretar como la ventaja de pasar de un estado de depresión a no depresión. En otras palabras 2 de cada 10 pacientes pasan de un estado de depresión a no depresión luego de la terapia.

➤ **Algunas consideraciones en R**

- Se puede realizar la prueba con los datos sin agrupar o agrupados en una tabla de contingencia 2x2.
- La función `mcnemar` utiliza la distribución Chi Cuadrado para calcular el pvalor.
- La función `mcnemar.exact` utiliza la distribución Binomial para calcular el pvalor.

2. Pruebas para evaluar un parámetro de locación (o posición)

Hipótesis

Bilateral	Unilateral	
Caso A	Caso B	Caso C
$H_0 : Me_1 - Me_2 = Me_0$	$H_0 : Me_1 - Me_2 \leq Me_0$	$H_0 : Me_1 - Me_2 \leq Me_0$
$H_1 : Me_1 - Me_2 \neq Me_0$	$H_1 : Me_1 - Me_2 > Me_0$	$H_1 : Me_1 - Me_2 < Me_0$
O	O	O
$H_0 : Me_d = Me_0$	$H_0 : Me_d \leq Me_0$	$H_0 : Me_d \geq Me_0$
$H_1 : Me_d \neq Me_0$	$H_1 : Me_d > Me_0$	$H_1 : Me_d < Me_0$

A. Prueba de Signos para datos pareados

➤ Aspectos Generales

Esta prueba es el equivalente no paramétrico a la prueba de comparación de medias con observaciones pareadas. Un caso especial es para probar la hipótesis nula de que las dos medianas de la población son iguales.

Si la hipótesis de que las dos poblaciones son de igual nivel de magnitud es cierta, el número de signos (+) debería ser aproximadamente igual al número de signos (-).

➤ Supuestos

- Los valores muestrales deben estar en al menos en la escala intervalo, es decir, la variable debe ser cuantitativa.
- Las mediciones de la variable de interés hechas a los n elementos de la muestra se deben realizar en dos momentos diferentes o bajo dos procedimientos distintos.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Se aplica un signo (+) a cada par de valores cuya medida en la primera muestra es mayor que la medida en la segunda muestra, y un signo (-) cuando ocurre lo contrario. Si un par de medidas tiene el mismo valor, estos valores relacionados se excluyen del análisis, con lo que el tamaño de muestra efectivo se reduce.
- Definir el tamaño de la muestra (n) definitivo excluyendo a los valores iguales en las dos mediciones.
- Contabilizar el número de signos (+).

Las hipótesis pueden tomar alguna de las siguientes formas:

El cálculo de los p-valores para cada caso es como si se tratara de una muestra.

➤ Aplicación

Ejemplo 1

A un grupo de consumidores compuesto por 14 individuos elegidos al azar se le pide calificar dos marcas de refresco de cola de acuerdo con un sistema de

evaluación por puntos basado en varios criterios. Los resultados son presentados en la tabla siguiente:

Miembro del Grupo	Calificación (puntos asignados)		Signo de la Diferencia
	Marca 1	Marca 2	
1	20	16	+
2	24	26	-
3	28	18	+
4	24	17	+
5	20	20	0
6	29	21	+
7	19	23	-
8	27	22	+
9	20	23	-
10	30	20	+
11	18	18	0
12	28	21	+
13	26	17	+
14	24	26	-

Pruebe la hipótesis de que existe diferencia en el nivel mediano de calificaciones de las dos marcas de refresco de cola al nivel de significación de 5%.

Solución

$$H_0: Me_1 = Me_2 \quad H_0: \pi = 0.50$$

$$H_1: Me_1 \neq Me_2 \quad H_1: \pi \neq 0.50$$

$$\alpha=0.05$$

Cálculo de p-valor

$$p - valor = 2P(X \leq c') = 2 \sum_{x=0}^{c'} \binom{n}{x} 2^{-n}$$

Criterio de Decisión

Si $p\text{-valor} \geq \alpha$ no se rechaza la H_0 .

Si $p\text{-valor} < \alpha$ se rechaza la H_0 .

Desarrollo de la Prueba

El valor de n inicial es 14 pero debido a la existencia de dos valores iguales entre las medianas, n ahora sería igual a 12, el menor valor entre signos (+) y (-) es 4

$$p - valor = 2P(X \leq c') = 2 \sum_{x=0}^4 \binom{12}{x} 2^{-12}$$

```
2*pbinom(4,12,0.5)
```

```
[1] 0.3876953
```

Conclusión

A un nivel de significación de 0.05 existe suficiente evidencia estadística para no rechazar la hipótesis nula.

Por lo tanto, no podemos afirmar que las dos marcas de refresco de cola tienen diferente preferencia del consumidor.

Ejemplo 2

A un grupo de 28 pacientes elegidos al azar que asistía a una clínica de fisioterapia se le sometió a una determinada prueba diseñada para medir su nivel de motivación, antes de que participaran en un curso experimental de remotivación.

Al término del programa, los pacientes fueron sometidos a una nueva prueba. Las calificaciones antes y después fueron las siguientes:

Paciente	Calificación		Dif. D-A	Signo
	Antes	Después		
1	10	15	5	+
2	8	10	2	+
3	5	10	5	+
4	14	14	0	0
5	15	25	10	+
6	22	20	-2	-
7	17	20	3	+
8	10	22	12	+
9	8	16	8	+
10	20	18	-2	-
11	11	16	5	+
12	16	24	8	+
13	10	23	13	+
14	15	25	10	+
15	5	15	10	+
16	24	20	-4	-

Paciente	Calificación		Dif. D-A	Signo
	Antes	Después		
17	20	24	4	+
18	14	23	9	+
19	10	25	15	+
20	15	21	6	+
21	14	15	1	+
22	17	17	0	0
23	22	21	-1	-
24	14	13	-1	-
25	16	15	-1	-
26	9	10	1	+
27	16	13	-3	-
28	15	19	4	+
29	14	15	1	+
30	13	12	-1	-
31	9	9	0	0
32	10	12	2	+

Se sabe que una calificación alta indica un alto nivel de motivación.

¿Se concluiría que el programa de remotivación es eficaz? Use $\alpha=0.05$.

Solución:

$H_0: Me_D \leq Me_A$ $H_0: \pi \leq 0.50$

$H_1: Me_D > Me_A$ $H_1: \pi > 0.50$

$\alpha=0.05$

Prueba Estadística

$$Z_c = \frac{x - 0.5n}{0.50\sqrt{n}} \sim N(0,1)$$

$$p\text{-value} = P(Z > Z_c) = 1 - P(Z < Z_c)$$

Criterio de Decisión

No se rechaza H_0 si $Z_c \leq 1.65$.

Se rechaza la H_0 si $Z_c > 1.65$.

ó

Si $p\text{-valor} \geq \alpha$ no se rechaza la H_0 .

Si $p\text{-valor} < \alpha$ se rechaza la H_0 .

Desarrollo de la Prueba

$x = 21$; $n = 29$

$$Z_c = \frac{21 - 0.5(29)}{0.5\sqrt{29}} = \frac{6}{2.693} = 2.227$$

$$p\text{-valor} = 1 - P(Z < Z_c) = 1 - P(Z < 2.227) = 1 - 0.9871 = 0.0129$$

Conclusión

A un nivel de significación de 0.05 existe suficiente evidencia estadística para rechazar la H_0 .

Por lo que se puede afirmar que el programa de remotivación fue eficaz.

➤ Funciones en R

Existe la función `SIGN.test` del paquete `BSDA`.

En esta última función se debe ingresar los dos vectores de datos.

```
SIGN.test(vector de datos1, vector de datos2, valor  
hipotético, alternativa, conf.level)
```

Otra función que se puede utilizar es `signTest` del paquete `EnvStats`

```
signTest(vector de datos1, vector de datos2, valor  
hipotético, alternativa, conf.level, paired=TRUE)
```

También se puede utilizar la función `sign_test` del paquete `coin`

```
sign_test(y~g, alternativa, valor hipotético)
```

➤ Resultados con R

```
m1<-c(20,24,28,24,20,29,19,27,20,30,18,28,26,24)  
m2<-c(16,26,18,17,20,21,23,22,23,20,18,21,17,26)  
library(BSDA)  
SIGN.test(m1,m2)
```

```
library(EnvStats)  
signTest(m1,m2,paired=TRUE)
```

```
Results of Hypothesis Test  
-----  
Null Hypothesis:                median of differences = 0  
Alternative Hypothesis:         True median of differences  
is not equal to 0  
Test Name:                     Paired Sign test  
Estimated Parameter(s):        median of differences = 6  
Data:                          x = m1  
                               y = m2  
Test Statistic:                # Diffs > median of  
differences = 8  
P-value:                       0.3876953  
Confidence Interval for:       median of differences  
Confidence Interval Method:    exact  
Confidence Interval Type:      two-sided  
Confidence Level:              90.77148%  
Confidence Limit Rank(s):      4 10  
Confidence Interval:           LCL = -2  
                               UCL = 9
```

```
library(coin)  
y<-c(m1,m2)  
g<-as.factor(rep(1:2,c(14,14)))
```

```
sign_test(y~g, alternative="t")
```

Dependent-samples Sign-Test

```
data:  m1 and m2
S = 8, p-value = 0.3877
alternative hypothesis: true median difference is not equal
to 0
95 percent confidence interval:
 -2.000000  8.165934
sample estimates:
median of x-y          4.5
```

	Conf.Level	L.E.pt	U.E.pt
Lower Achieved CI	0.9426	-2	8.0000
Interpolated CI	0.9500	-2	8.1659
Upper Achieved CI	0.9871	-2	9.0000

Asymptotic Sign Test

```
data:  y by
      x (pos, neg)
      stratified by block
Z = 5.2915, p-value = 1.213e-07
alternative hypothesis: true mu is not equal to 0
```

➤ **Algunas consideraciones con R**

La función SIGN.test (del paquete BSDA) y la función signTest (del paquete EnvStats):

- Permite analizar una hipótesis bilateral o unilateral para un valor hipotético específico.
- La función sign_test del paquete coin no brinda los mismos resultados que la función SIGN.test

B. Prueba de Wilcoxon para datos pareados

➤ **Aspectos Generales**

Al igual que la prueba de Signos para observaciones pareadas, esta prueba es el equivalente no paramétrico a la prueba de comparación de medias con observaciones pareadas.

Dado que la prueba de Wilcoxon considera la magnitud de la diferencia, es una prueba más sensible que la Prueba de Signos.

➤ **Supuestos**

- Los valores muestrales deben estar en al menos en la escala intervalo es decir la variable debe ser cuantitativa.
- Las mediciones de la variable de interés hechas a los n elementos de la muestra se deben realizar en dos momentos diferentes o bajo dos procedimientos distintos.
- La distribución (forma) de la diferencia de las observaciones con respecto al valor hipotético debe ser simétrica.

➤ **Inferencia Estadística**

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Calcular la diferencia entre el primer y segundo grupo de observaciones (o viceversa dependiendo de la prueba alternativa de interés).
- Obtener los rangos de las diferencias (con respecto al valor hipotético) sin tomar en cuenta el signo de las mismas (en valor absoluto). En caso de haber empate se asigna un rango promedio a todas las diferencias empatadas.
- Todo el procedimiento y criterio de decisión posterior es similar al de la Prueba de Wilcoxon de una muestra debido a que en este caso se sigue trabajando con una columna de diferencias.

➤ **Aplicación**

Ejemplo

A un grupo de 17 pacientes elegidos al azar que asistía a una clínica de fisioterapia se les sometió a una determinada prueba diseñada para medir su nivel de motivación, antes de que participaran en un curso experimental de remotivación. Al término del programa, los pacientes fueron sometidos a una nueva prueba. Las calificaciones antes y después fueron las siguientes:

Paciente	Calificación		Diferen. $X_D - X_A$	Rango de $ d $	Rango con Signo	
	Antes	Después			(+)	(-)
1	10	15	5	9.0	9	
2	8	10	2	3.0	3	
3	5	10	5	9.0	9	
4	14	15	1	1.0	1	
5	15	25	10	14.0	14	
6	22	20	-2	3.0		3
7	17	20	3	5.0	5	
8	10	22	12	16.0	16	
9	8	16	8	11.5	11.5	

10	20	18	-2	3.0		3
11	11	16	5	9.0	9	
12	16	24	8	11.5	11.5	
13	10	23	13	17.0	17	
14	15	25	10	14.0	14	
15	5	15	10	14.0	14	
16	24	20	-4	6.5		6.5
17	20	24	4	6.5	6.5	
Total	14	23	9		140.5	12.5

Primero se debe verificar la simetría de la diferencia de los datos antes y después del programa de remotivación.

$H_0: A_s = 0$

$H_1: A_s \neq 0$

$\alpha=0.05$

Pvalor=0.901

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que la diferencia de las puntuaciones antes y después del programa de remotivación presenta una distribución asimétrica.

Pruebe a un nivel de significación de 0.05 si el programa fue efectivo.

Solución:

$H_0: Me_D \leq Me_A$

$H_1: Me_D > Me_A$

$\alpha=0.05$

Prueba Estadística

$$Z_c = \frac{(w \pm 1/2) - \mu_w}{\sigma_w} \sim N(0,1) \text{ ó}$$

$$p\text{-valor} = P(Z > Z_c) = 1 - P(Z < Z_c)$$

Criterio de Decisión

No se rechaza H_0 si $Z_c \leq 1.65$.

Se rechaza la H_0 si $Z_c > 1.65$.

ó

Si $p\text{-valor} \geq \alpha$ no se rechaza la H_0 .

Si $p\text{-valor} < \alpha$ se rechaza la H_0 .

Desarrollo de la Prueba

$$\mu_w = \frac{17(17+1)}{4} = 76.5$$

$$\sigma_w = \sqrt{\frac{17(17+1)[2(17)+1]}{24} - \left[\frac{3^3-3}{2} + \frac{2^3-2}{2} + \frac{3^3-3}{2} + \frac{2^3-2}{2} + \frac{3^3-3}{2} \right]} = 20.11$$

$$Z_c = \frac{(140.5 - 1/2) - 76.5}{20.11} = 3.16$$

$$p\text{-valor} = P(Z > -3.16) = 1 - P(Z < 3.16) = 1 - 0.9992 = 0.0008$$

Conclusión

A un nivel de significación de 0.05 existe suficiente evidencia estadística para rechazar la H_0 .

Por lo se puede afirmar que el programa de remotivación fue eficaz.

➤ **Funciones en R**

En esta función se debe ingresar los dos vectores de datos.

```
wilcox.test(vector de datos1, vector de datos2, alternativa,  
valor hipotético, pareados=VERDADERO)
```

La función `wilcox.exact` del paquete `exactRankTests` permite obtener los resultados con el ajuste para empates

```
wilcox.exact(vector de datos1, vector de datos2,  
alternativa, valor hipotético, paired=TRUE)
```

También se puede utilizar la función `wilcoxsign_test` del paquete `coin`

```
wilcoxsign_test(y~g, alternativa, valor hipotético,método)
```

➤ **Resultados con R**

```
antes<-c(10,8,5,14,15,22,17,10,8,20,11,16,10,15,5,24,20)  
desp<-c(15,10,10,14,25,20,20,22,16,18,16,24,23,25,15,20,24)
```

```
symmetry.test(antes-desp,option="MGG",boot=F)
```

Symmetry test by Miao, Gel, and Gastwirth (2006)

```
data: antes - desp  
Test statistic = -0.12441, p-value = 0.901  
alternative hypothesis: the distribution is asymmetric.
```

```
wilcox.test(desp,antes,alternative="g",paired=TRUE)
```

```
library(exactRankTests)  
wilcox.exact(desp,antes,alternative="g",paired=TRUE)
```

```
library(coin)  
y<-c(antes,desp)  
g<-as.factor(rep(1:2,c(17,17)))  
wilcoxsign_test(y~g,alternative="g",zero.method="Wilcoxon")
```

Wilcoxon signed rank test with continuity correction

```
data: desp and antes  
V = 126.5, p-value = 0.001323  
alternative hypothesis: true location shift is greater than  
0
```

Exact Wilcoxon signed rank test

```
data: desp and antes  
V = 126.5, p-value = 0.0005188  
alternative hypothesis: true mu is greater than 0
```

```
Asymptotic Wilcoxon Signed-Rank Test
```

```
data:  y by x (pos, neg)
       stratified by block
Z = 5.0893, p-value = 1.797e-07
alternative hypothesis: true mu is greater than 0
```

➤ **Algunas consideraciones en R**

Se usa la función `wilcox.test` o la función `wilcox.exact`:

- Permite analizar una hipótesis bilateral o unilateral para un valor hipotético específico.
- La función `wilcoxsign_test` del paquete `coin` no brinda los mismos resultados que la función `wilcox.test` o `wilcox.exact`.

3. Pruebas para evaluar un parámetro de escala

A. Prueba de Grambsch

➤ Aspectos Generales

La prueba robusta de Grambsch permite evaluar la diferencia entre el parámetro de escala en datos pareados. Es decir, la dispersión de las mediciones antes y después de una muestra de n elementos.

➤ Supuestos

- Los valores muestrales del primer y segundo grupo de observaciones deben estar en al menos en la escala intervalo es decir la variable debe ser cuantitativa.
- Las mediciones de la variable de interés hechas a los 'n' elementos de la muestra se debe realizar en dos momentos diferentes, o bajo dos procedimientos distintos.
- La diferencia de los datos proviene de una distribución simétrica.

➤ Inferencia Estadística

Procedimientos para el desarrollo de la prueba.

- Sea $X_1, X_2, X_3, \dots, X_n$ y $Y_1, Y_2, Y_3, \dots, Y_n$ la muestra pareada para una misma variable de interés.
- Calcular $s_i = X_i + Y_i$ y $d_i = X_i - Y_i$ basado en el hecho que:

$$\text{var}(x) - \text{var}(y) = \text{cov}(x + y, x - y) = \text{cov}(s, d)$$
- Evaluar:

$$u_i = (s_i - \bar{s})(d_i - \bar{d})$$

Siendo:

$$\bar{s} = \sum_{i=1}^n s_i / n, \quad \bar{d} = \sum_{i=1}^n d_i / n$$

- El cálculo de la prueba estadística se puede tratar como una prueba t para una muestra.

$$z = \frac{\bar{u}}{\hat{\sigma}_u / \sqrt{n}} \sim N(0, 1)$$

Siendo:

$$\bar{u} = \sum_{i=1}^n u_i / n \quad \text{y} \quad \hat{\sigma}_u = \sqrt{\sum_{i=1}^n (u_i - \bar{u})^2 / (n-1)}$$

Prueba de hipótesis:

Las hipótesis pueden tomar las siguientes formas, donde θ_i es un parámetro de escala:

Bilateral	Unilateral	
$H_0 : \theta_x = \theta_y$	$H_0 : \theta_x \leq \theta_y$	$H_0 : \theta_x \geq \theta_y$
$H_1 : \theta_x \neq \theta_y$	$H_1 : \theta_x > \theta_y$	$H_1 : \theta_x < \theta_y$

➤ **Aplicación**

Se midió el tiempo de ejecución de un procedimiento (en comparación con un estándar) mediante dos equipos electrónicos, un valor positivo indica que se encuentra sobre el valor del estándar y valor negativo indica que se encuentra debajo del estándar. Se evaluaron 10 procesos elegidos de forma aleatoria, obteniéndose los siguientes resultados:

Equipo A	Equipo B
6.0	5.2
0.8	0.2
3.2	2.9
6.2	6.3
3.8	2.7
-1.6	-1.4
1.8	1.5
3.3	2.8
1.3	0.8
5.6	5.3

Pruebe a un nivel de significación de 0.05, si existe diferencia en la dispersión del tiempo de ejecución del procedimiento de los dos equipos electrónicos.

X_i	Y_i	$s_i = X_i + Y_i$	$d_i = X_i - Y_i$	$u_i = (s_i - \bar{s})(d_i - \bar{d})$
6.0	5.2	11.2	0.8	2.1567
0.8	0.2	1.0	0.6	-0.8873
3.2	2.9	6.1	0.3	-0.0473
6.2	6.3	12.5	-0.1	-3.4833
3.8	2.7	6.5	1.1	0.5727
-1.6	-1.4	-3.0	-0.2	5.2887
1.8	1.5	3.3	0.3	0.2607
3.3	2.8	6.1	0.5	0.0387
1.3	0.8	2.1	0.5	-0.3213
5.6	5.3	10.9	0.3	-0.5753

$$\bar{d} = \sum_{i=1}^n d_i / n = \frac{0.8+0.6+0.3-0.1+1.1+-0.2+0.3+0.5+0.5+0.3}{10} = 0.41$$

$$\bar{s} = \sum_{i=1}^n s_i / n = \frac{11.2+1+6.1+12.5+6.5-3+3.3+6.1+2.1+10.9}{10} = 5.67$$

$$\bar{u} = \sum_{i=1}^n u_i / n = \frac{2.1567-0.8873-0.0473-3.4833+0.5727+5.2887+0.2607+0.0387-0.3213-0.5753}{10} = 0.3003$$

$$\hat{\sigma}_u = \sqrt{\sum_{i=1}^n (u_i - \bar{u})^2 / (n-1)} = \sqrt{\frac{(2.1567-0.3003)^2 + (-0.8873-0.3003)^2 + \dots + (-0.3213-0.3003)^2 + (-0.5753-0.3003)^2}{10-1}} = 0.7108248$$

$$H_0 : \theta_x = \theta_y$$

$$H_a : \theta_x \neq \theta_y$$

$$\alpha = 0.05$$

Prueba Estadística:

$$z = \frac{\bar{u}}{\hat{\sigma}_u / \sqrt{n}} \sim N(0,1)$$

$$z = \frac{\bar{u}}{\hat{\sigma}_u / \sqrt{n}} = \frac{0.3003}{0.7108248 / \sqrt{10}} = 0.42246$$

$$pvalor = 1 - 2*(0.42246) = 0.6726842$$

Conclusión

A un nivel de significación del 5%, existe suficiente evidencia estadística para no rechazar la hipótesis nula. Por lo tanto, no se puede afirmar que existe diferencia en la dispersión del tiempo de ejecución del procedimiento de los dos equipos electrónicos.

➤ Funciones con R

Existe la función `grambsch.Var.test` del paquete `PairedData`

`grambsch.Var.test(vector de datos 1 , vector de datos 2, alternative)`

También existe la función `grambsch.Var.test.paired` del paquete mencionado.

En este caso se debe ingresar los vectores de datos como tipo `paired`.

`grambsch.Var.test.paired(objeto clase pareado, alternative)`

➤ Resultados con R

```
x<-c(6, 0.8, 3.2 ,6.2, 3.8 ,-1.6, 1.8, 3.3, 1.3, 5.6)
```

```
y<-c(5.2, 0.2, 2.9 ,6.3, 2.7, -1.4 ,1.5, 2.8 ,0.8, 5.3)
```

```
grambsch.Var.test(x,y,alternative = "t")
```

Paired Grambsch test

data: x and y

z = 0.42247, p-value = 0.6727

alternative hypothesis: true ratio of variances is not equal to 1

Se almacenan ambos vectores `x` e `y` en un objeto de clase `pareada`

```
o<-paired(x,y)
```

```
grambsch.Var.test.paired(o)
```

<pre>Paired Grambsch test data: x and y z = 0.42247, p-value = 0.6727 alternative hypothesis: true ratio of variances is not equal to 1</pre>
--

En el paquete PairedData. existen otras pruebas para la comparación del parámetro de escala en una muestra pareada:

Prueba de Bonett-Seier con la función `bonettseier.Var.test()`,

Prueba de Imam con la función `imam.Var.test()`,

Prueba de Levene con la función `levener.Var.test.paired()`,

Prueba de McCulloch test con la función `mcculloch.Var.test.paired()` y

Prueba de Sandvik-Olsson con la función `sandvikolsson.Var.test()`

Por otro lado, en ese mismo paquete hay pruebas para otros parámetros de datos relacionados como

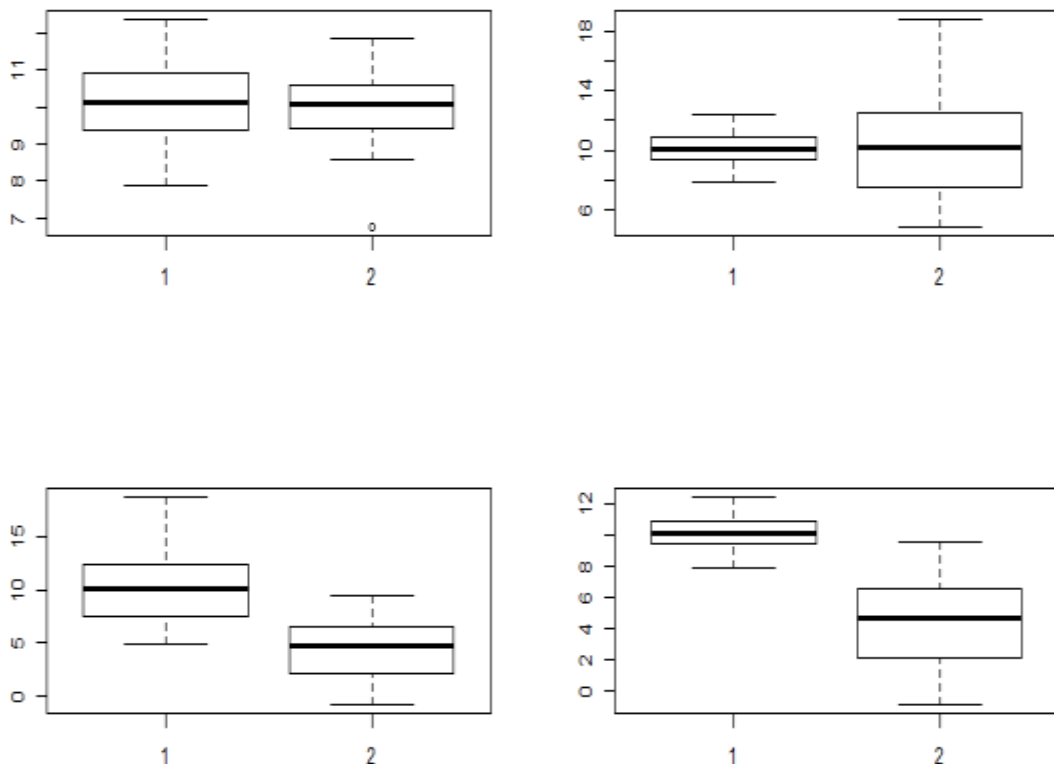
La prueba de Yuen para evaluar la media podada con la función `yuen.t.test.paired`.

La prueba Winsorizada para evaluar la correlación con la función `winsor.cor.test.paired`.

PRUEBAS PARA DOS MUESTRAS INDEPENDIENTES

Cuando se tienen datos provenientes de dos muestras aleatorias e independientes se puede estar interesado en comparar parámetros de: locación (posición), escala (dispersión) o la forma que tienen.

La idea inicial de comparación de dos grupos se puede establecer analizando los siguientes diagramas:



En el primer gráfico de cajas (superior izquierda) se puede observar que no hay mucha diferencia en la mediana, los tamaños de las cajas son muy parecidos y la mediana en ambas cajas está ubicada en la parte central. Con lo que se podría suponer que no hay diferencia en la locación, dispersión en el 50% central de los datos y forma en el 50% central de los datos.

En el segundo gráfico de cajas (superior derecha) se puede apreciar que no hay diferencia en el valor de la mediana ni en su ubicación dentro de la caja, pero los tamaños de las cajas si presentan diferencias. Esto podría hacer sospechar que no hay diferencia en la locación y forma en el 50% central de los datos, pero si hay diferencia en la dispersión en el 50% central de los datos.

En el tercer gráfico de cajas (inferior izquierda) hay una diferencia en el valor de la mediana, pero no en el tamaño de las cajas y en la ubicación de la mediana dentro de la caja. Por lo que se podría sospechar que hay diferencia en la locación, pero no en el de dispersión y forma en el 50% central de los datos.

En el cuarto gráfico de cajas (inferior derecha) hay diferencia en el valor de las medianas y en tamaño de las cajas, pero no hay diferencia en la ubicación de la mediana dentro de la caja. Por lo que se supondría que hay diferencia en la locación, dispersión en el 50% central de los datos, pero no hay diferencia en la forma en el 50% central de los datos.

1. Para comparar la distribución (forma)

Como se discutirá más adelante existen pruebas que exigen que dos conjuntos de datos deben presentar la misma forma, es decir que no exista diferencia en la distribución de los datos de donde provienen.

Hipótesis

$$H_0: F_X(x) = G_X(x)$$

$$H_1: F_X(x) \neq G_X(x)$$

A. Prueba de Kolmogorov-Smirnov

➤ Aspectos Generales

Smirnov extendió la prueba KS para comparar dos distribuciones basadas en muestras independientes.

Esta prueba puede ser considerada como un análisis previo a diversas pruebas paramétricas, dado que muchas de ellas requieren que las muestras aleatorias e independientes provengan de la misma distribución.

➤ Supuesto

- La variable en estudio esta medida en una escala por lo menos intervalo.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Si X_1, X_2, \dots, X_m y Y_1, Y_2, \dots, Y_n son dos muestras independientes de poblaciones con funciones de distribución acumulada F_X y G_Y . Si $F_m(x)$ y $G_n(x)$ son las correspondientes funciones de distribución empírica.
- Obtener $Z = Z_1, \dots, Z_{m+n}$ es la muestra combinada X_1, X_2, \dots, X_m y Y_1, Y_2, \dots, Y_n
- Calcular

$$D_{m,n} = \sup_x |F_m(x) - G_n(x)|$$

Donde $D_{m,n}$ puede ser simplificada (para una conveniente programación) en

$$D_{m,n} = \max_x \{|F_m(Z_i) - G_n(Z_i)|\}$$

$D_{m,n}$ podría ser grande si hay conglomerados de valores cuando las muestras son combinadas.

Por lo tanto, el valor de las observaciones no es directamente relevantes a excepción de cómo están ordenadas luego de ser combinadas.

➤ Aplicación

Ejemplo 1

En un experimento diseñado para estimar los efectos de la inhalación prolongada de óxido de cadmio, 15 animales de laboratorio elegidos al azar sirvieron de sujetos para el experimento, mientras que 10 animales similares elegidos al azar sirvieron de controles. La variable de interés fue el nivel de hemoglobina después del experimento. Los resultados se muestran a continuación:

Animales Expuestos	Rango	Animales no Expuestos	Rango
14.4	7.0	17.4	24.0
14.2	6.0	16.2	16.0
13.8	2.0	17.1	23.0
16.5	18.0	17.5	25.0
14.1	4.5	15.0	8.5
16.6	19.0	16.0	15.0
15.9	14.0	16.9	22.0
15.6	12.0	15.0	8.5
14.1	4.5	16.3	17.0
15.3	10.5	16.8	21.0
15.7	13.0		
16.7	20.0		
13.7	1.0		
15.3	10.5		
14.0	3.0		
	145		180

Verifique si los datos correspondientes al nivel de hemoglobina de los animales expuestos y no expuestos no provienen de la misma distribución teórica. Use $\alpha=0.05$.

Primero realizaremos un análisis gráfico

Diagrama de Cajas

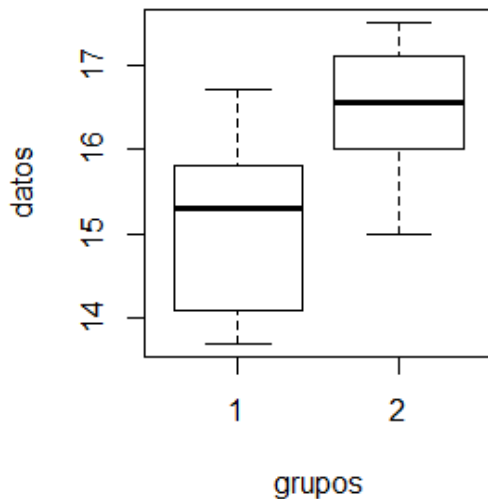
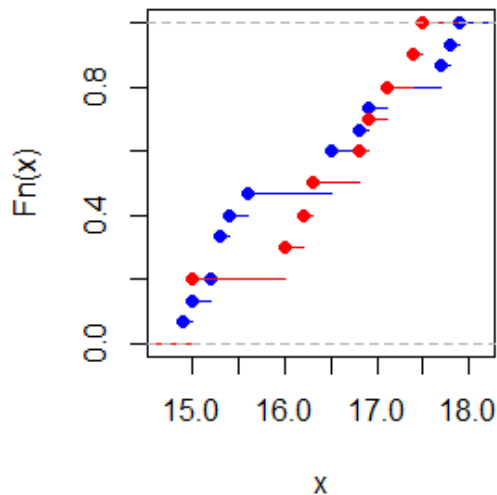


Gráfico de Dis. Acum. Emp.



$$H_0 : F_X(x) = G_Y(x)$$

$$H_1 : F_X(x) \neq G_Y(x)$$

$$\alpha=0.05$$

$$p\text{-valor}= 0.0266$$

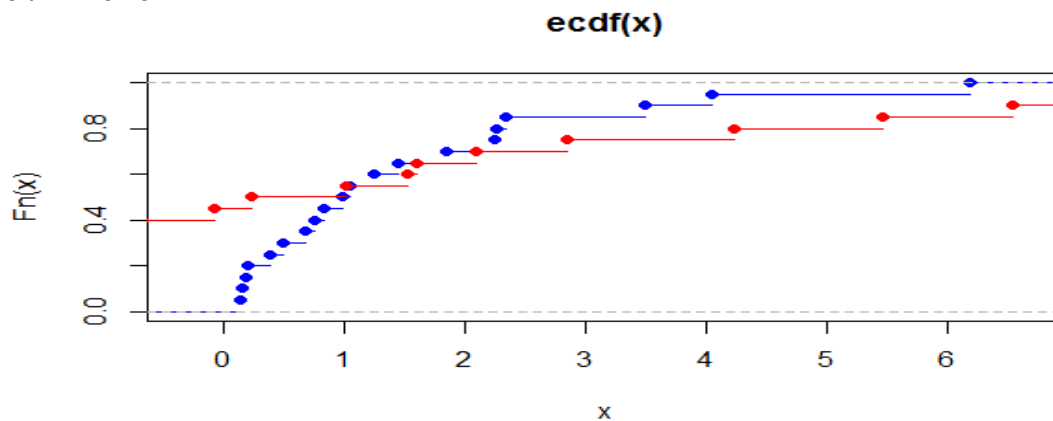
Conclusión

A un nivel de significación de 0.05, se puede afirmar que las distribuciones teóricas del nivel de hemoglobina de los animales no expuestos y expuestos al óxido de cadmio son diferentes.

Ejemplo 2

Esta prueba se ve fácilmente afectada por la locación de los datos como se puede apreciar en el siguiente ejemplo:

```
set.seed(100)
d1<-rexp(20,1/2)
d2<-rexp(20,1/6)
mean(d1)
1.55242
mean(d2)
5.272529
```



```
ks.test(d1,d2)
```

```
Two-sample Kolmogorov-Smirnov test
data: d1 and d2
D = 0.55, p-value = 0.003967
alternative hypothesis: two-sided
```

Como se puede apreciar a pesar que los datos provienen de la misma distribución la prueba rechaza eso.

Para evitar este problema se puede reescalar los datos con respecto al centro, antes de la aplicación de la prueba.

```
ks.test(d1+3.7,d2)
```

```
Two-sample Kolmogorov-Smirnov test
data: d1 + 3.7 and d2
D = 0.4, p-value = 0.08106
alternative hypothesis: two-sided
```

```
ks.test(scale(d1,center=T,scale=F),scale(d2,center=T,scale=
F))
```

Two-sample Kolmogorov-Smirnov test

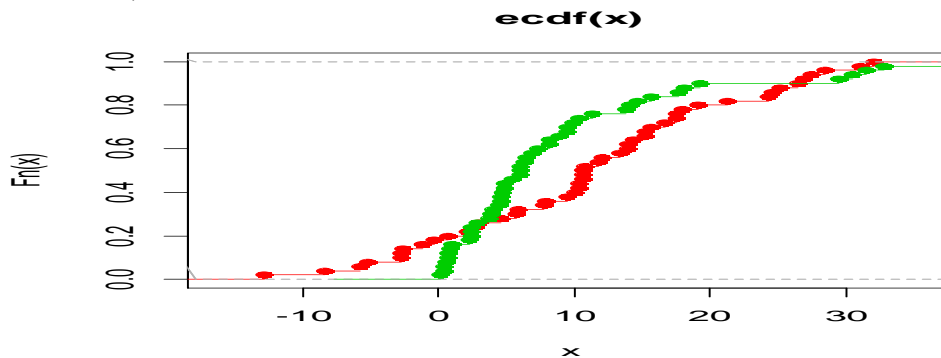
```
data: scale(d1, center = T, scale = F) and scale(d2, center
= T, scale = F)
D = 0.4, p-value = 0.08106
alternative hypothesis: two-sided
```

Esto quiere decir que en el ejemplo 1 también se tendría que reescalar los datos antes de aplicar la prueba de Kolmogorov-Smirnov. Con lo cual resultaría un pvalor=0.787 y no se rechazaría H_0 .

Ejemplo 3

Simule dos muestras de tamaño 50, una de ellas proveniente de una distribución normal con media 10 y desviación estándar 10 y la otra con distribución exponencial con media 10. Realice un análisis estadístico completo.

```
set.seed(21)
m1<-rnorm(50,10,10)
m2<-rexp(50,1/10)
plot.ecdf(m1, col="red")
plot.ecdf(m2,ad=T,col=3)
ks.test(m1,m2)
t.test(m1,m2)
```



```
Two-sample Kolmogorov-Smirnov test
data: m1 and m2
D = 0.34, p-value = 0.005842
alternative hypothesis: two-sided
```

```
Welch Two Sample t-test
data: m1 and m2
t = 0.9015, df = 97.566, p-value = 0.3696
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
-2.250675  5.997099
sample estimates:
mean of x mean of y
11.244565  9.371353
```

➤ **Secuencia en R**

```
ks.test(vector de datos1, vector de datos2)
```

➤ **Resultados con R**

```
ex<-c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,  
15.7, 16.7,13.7,15.3,14.0)  
ne<-c(17.4,16.2,17.1,17.5,15.0,16.0,16.9,15.0,16.3,16.8)  
n1<-length(ex)  
n2<-length(ne)  
grupos<-c(rep(1,n1),rep(2,n2))  
datos<-c(ex,ne)  
par(mfrow=c(1,2))  
boxplot(datos~grupos, main="Diagrama de Cajas")  
plot.ecdf(ex+1.2, col="blue", main="Gráfico de Dis. Acum.  
Emp.")  
plot.ecdf(ne,ad=T,col="red")  
ks.test(ex,ne)
```

```
Two-sample Kolmogorov-Smirnov test  
data: ex and ne  
D = 0.6, p-value = 0.0266  
alternative hypothesis: two-sided  
Mensajes de aviso perdidos  
In ks.test(ex, ne) : cannot compute exact p-value with ties
```

```
mean(ex); mean(ne)  
15.06  
16.42
```

```
ks.test(scale(ex,center=T,scale=F),scale(ne,center=T,scale=  
F))
```

```
Two-sample Kolmogorov-Smirnov test  
data: scale(ex, center = T, scale = F) and scale(ne, cente  
r = T, scale = F)  
D = 0.26667, p-value = 0.787  
alternative hypothesis: two-sided
```

```
ks.test(ex+1.4,ne)
```

```
Two-sample Kolmogorov-Smirnov test  
  
data: ex + 1.4 and ne  
D = 0.26667, p-value = 0.787  
alternative hypothesis: two-sided
```


B. La prueba de Rachas o Aleatoriedad de Wald-Wolfowitz

➤ Aspectos Generales

Esta prueba tiene la misma finalidad que la prueba de Kolmogorov-Smirnov, es decir permite verificar si dos muestras independientes provienen de la misma distribución teórica. La diferencia con la prueba de Kolmogorov-Smirnov radica en que en esta prueba se contabiliza el número de rachas existentes entre las dos muestras luego de haberlas combinado como una sola.

➤ Supuesto

- La variable de interés presenta una escala de medición por lo menos intervalo.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Si X_1, X_2, \dots, X_{n_1} y Y_1, Y_2, \dots, Y_{n_2} son dos muestras independientes se deben combinar los dos conjuntos de datos en uno solo ($n=n_1+n_2$) y ordenarlos de menor a mayor.
- Analizar el número de corridas (o rachas) considerando si los datos provienen de una u otra muestra. Por ejemplo si la primera muestra está formada por 4 elementos y la segunda por 5 elementos un arreglo puede ser XYYXXYXY se puede observar que hay 6 rachas.

Un patrón con muy pocas corridas podría sugerir que este grupo de n observaciones no sea una sola muestra aleatoria sino que compone de dos muestras de dos poblaciones distintas.

Si el arreglo es XXXXYYYYY hay dos corridas, los elementos de la muestra de X son más pequeños que los elementos de los elementos de la muestra de Y. Esta particular configuración podría indicar no solo que las poblaciones no son idénticas sino que las X's son estocásticamente más pequeñas que las Y's. Sin embargo, el orden reverso también contiene solo dos corridas y por lo tanto un criterio basado únicamente en el número total de corridas no permite distinguir esos dos casos.

- Si se define la variable aleatoria R como el número total de corridas de la muestra combinada de X e Y. Si hay pocas corridas se rechazaría H_0 .
- Los criterios para rechazar H_0 y la aproximación asintótica son los mismos que para prueba de rachas de una muestra.

En el caso de muestras grandes se puede hacer uso a la aproximación a la Normal

$$Z = \frac{R - \mu_R}{\sigma_R}$$

$$\mu_R = \frac{2n_1n_2}{(n_1 + n_2 + 1)}$$

$$\sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

➤ Aplicación

Ejemplo 1

Con los datos del Ejemplo 1 utilizados en la prueba de Kolmogorov-Smirnov verifique si los datos presentan distinta forma. Use $\alpha=0.05$.

$$H_0 : F_X(x) = G_Y(x)$$

$$H_1 : F_X(x) \neq G_Y(x)$$

$$\alpha=0.05$$

$$p\text{-valor} = 0.004$$

Conclusión

A un nivel de significación de 0.05, se puede afirmar que las distribuciones teóricas del nivel de hemoglobina de los animales no expuestos y expuestos al óxido de cadmio son diferentes.

Si los datos se centran el pvalor es igual a 0.2871, con lo cual no se rechaza H_0 .

Ejemplo 2

Al igual que la prueba de Kolmogorov-Smirnov, la prueba de Rachas de Wald-Wolfowitz podrían brindar resultados errados cuando se analizan muestras con parámetros de locación muy distintos.

Esto es fácil de deducir en la prueba de Wald-Wolfowitz debido a que si un grupo presenta mayor locación que otro, este tenderá a ubicar sus observaciones al final lo que implicaría pocas rachas así los datos provengan de distribuciones similares.

```
set.seed(20)
d1<-rnorm(20)
d2<-rnorm(15,8)
ks.test(d1,d2)
      Two-sample Kolmogorov-Smirnov test
```

```
data:  d1 and d2
D = 1, p-value = 6.158e-10
alternative hypothesis: two-sided
```

```
RunsTest(d1,d2)
Wald-Wolfowitz Runs Test
```

```
data:  d1 and d2
z = -5.4831, runs = 2, m = 20, n = 15, p-value = 4.18e-08
alternative hypothesis: true number of runs is not equal th
e expected number
```

```
RunsTest(scale(d1,center=T,scale=F),scale(d2,center=T,scale
=F))
```

```
RunsTest(scale(d1,center=T,scale=F),scale(d2,center=T,scale
=F))
```

```
Wald-Wolfowitz Runs Test
```

```
data:  scale(d1, center = T, scale = F) and scale(d2, center = T, scale = F)
z = 0.82622, runs = 21, m = 20, n = 15, p-value = 0.4087
alternative hypothesis: true number of runs is not equal the expected number
```

➤ **Funciones con R**

Existe la función `pruns.exact` del paquete `randomizeBE`, pero el defecto es que se debe conocer el número de rachas existentes.

`pruns.exact(número de rachas, n1, n2)`

También existe la función `RunsTest` del paquete `DescTools`

`RunsTest(vector1, vector2)`

➤ **Resultados con R**

Ejemplo1

```
x<-c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,
15.7, 16.7,13.7,15.3,14.0)
y<-c(17.4,16.2,17.1,17.5,15.0,16.0,16.9,15.0,16.3,16.8)
n1<-length(x)
n2<-length(y)
todo<-c(x,y)
grupos<-c(rep(1,n1),rep(2,n2))
data<-cbind(todo,grupos)
ordenados<-data[order(data[,1]),]
pruns.exact(6,15,10)
[1] 0.004025074
```

`RunsTest(x,y)`

Wald-Wolfowitz Runs Test

```
data:  x and y
runs = 6, m = 15, n = 10, p-value = 0.004025
alternative hypothesis: true number of runs is not equal the expected number
```

`RunsTest(scale(x,center=T,scale=F),scale(y,center=T,scale=F))`

Wald-Wolfowitz Runs Test

```
data:  scale(x, center = T, scale = F) and scale(y, center = T, scale = F)
runs = 10, m = 15, n = 10, p-value = 0.2871
alternative hypothesis: true number of runs is not equal the expected number
```

2. Para evaluar un parámetro de locación

Las pruebas que se discutirán a continuación son consideradas alternativas no paramétricas de las pruebas Z o t para la diferencia de medias de dos muestras independientes.

Hipótesis

Bilateral	Unilateral	
Caso A	Caso B	Caso C
$H_0 : Me_1 - Me_2 = Me_0$	$H_0 : Me_1 - Me_2 \leq Me_0$	$H_0 : Me_1 - Me_2 \geq Me_0$
$H_1 : Me_1 - Me_2 \neq Me_0$	$H_1 : Me_1 - Me_2 > Me_0$	$H_1 : Me_1 - Me_2 < Me_0$

A. Prueba de Mann Whitney

➤ Aspectos Generales

Es una prueba que se basa en los rangos de las observaciones.

La prueba de Mann-Whitney permite comparar si la mediana de una población es diferente, mayor o menor a la mediana de la otra población.

Si las dos poblaciones son simétricas, de modo que dentro de cada población la media y la mediana son las mismas, las conclusiones a las que se llegan respecto a las medias de las dos poblaciones se aplicarán también a las medianas de ambas poblaciones.

➤ Supuestos

- La variable de interés se encuentra medida en una escala por lo menos intervalo.
- Las muestras provienen de distribuciones similares.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Combinar las dos muestras y se obtienen el rango de todas las observaciones (de menor a mayor), teniendo presente a cual muestra pertenece cada observación. Si se tienen observaciones similares se asigna un rango promedio (obtenido de las posiciones del rango para el cual se establecieron).
- Desarrollar la prueba estadística correspondiente según el caso de tamaño de muestra

Casos

Muestras pequeñas

La prueba estadística es:

$$U = S - \frac{n_1(n_1 + 1)}{2}$$

donde: n_1 es el número de observaciones de la muestra de interés y S es la suma de los rangos asignados a las observaciones de la muestra de la población de valores correspondiente al grupo de interés.

Este valor se compara con otro obtenido de la Tabla de Mann Whitney que se busca con los valores n_1 (número de observaciones del primer grupo), n_2 (número de observaciones del segundo grupo) y con el nivel de significación α . Los criterios de decisión para muestras pequeñas son:

- Si $H_1: Me_X \neq Me_Y$ se rechaza H_0 si $U < W_{\alpha/2}$ ó $U > (W_{1-\alpha/2} = n_1 n_2 - W_{\alpha/2})$.
- Si $H_1: Me_X < Me_Y$ se rechaza H_0 si $U < W_{\alpha}$.
- Si $H_1: Me_X > Me_Y$ se rechaza H_0 si $U > (W_{1-\alpha} = n_1 n_2 - W_{\alpha})$.

Muestras Grandes

Si las muestras son grandes puede utilizarse la siguiente aproximación a la distribución normal en el caso que no haya empates:

$$Z = \frac{U - \frac{n_1(n_2)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0,1)$$

y si hubiera empates la expresión sería:

$$Z = \frac{U - \frac{n_1(n_2)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \left[\sum_{i=1}^g \frac{t_i^3 - t_i}{(n_1 + n_2)(n_1 + n_2 - 1)} \right]}} \sim N(0,1)$$

donde: U puede ser calculado mediante las siguientes expresiones:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \text{ y } U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

R_1 y R_2 son las sumas de los rangos de la primera y segunda muestra respectivamente.

Se elige el menor valor entre U_1 o U_2 para reemplazarlo en la prueba de la aproximación a la normal.

➤ **Aplicación**

Ejemplo 1

En un experimento diseñado para estimar los efectos de la inhalación prolongada de óxido de cadmio, 15 animales de laboratorio sirvieron de sujetos para el experimento, mientras que 10 animales similares sirvieron de controles. La variable de interés fue el nivel de hemoglobina después del experimento. Los resultados se muestran a continuación:

Animales Expuestos	Rango	Animales no Expuestos	Rango
14.4	7.0	17.4	24.0
14.2	6.0	16.2	16.0
13.8	2.0	17.1	23.0
16.5	18.0	17.5	25.0
14.1	4.5	15.0	8.5
16.6	19.0	16.0	15.0
15.9	14.0	16.9	22.0
15.6	12.0	15.0	8.5

14.1	4.5	16.3	17.0
15.3	10.5	16.8	21.0
15.7	13.0		
16.7	20.0		
13.7	1.0		
15.3	10.5		
14.0	3.0		
	145		180

Se desea saber si puede concluirse que la inhalación prolongada de óxido de cadmio disminuye el nivel de hemoglobina. Use $\alpha=0.05$.

Primero se debe probar si ambas muestras provienen de distribuciones similares. Esto ya se probó con ayuda de las dos pruebas anteriores con Wald-Wolfowitz (pvalor=0.2871) y con Kolmogorov-Smirnov (pvalor=0.787).

Solución

$H_0: Me_1 \geq Me_2$

$H_1: Me_1 < Me_2$

$\alpha=0.05$

Prueba Estadística

$$U = S - \frac{n_1(n_1 + 1)}{2}$$

Criterio de Decisión

No se rechaza H_0 si $U \geq 45$.

Se rechaza H_0 si $U < 45$.

Desarrollo de la Prueba

$$U = 145 - \frac{15(15 + 1)}{2} = 25$$

Conclusiones

Existe suficiente evidencia estadística para rechazar la H_0 a un nivel de significación de 0.05.

Por lo tanto, se puede afirmar que la inhalación prolongada de óxido de cadmio si disminuye el nivel de hemoglobina.

Ejemplo 2

Los salarios mensuales (en cientos de soles) de 21 empleados elegidos al azar de la empresa "Primer S.A." y los salarios de 19 empleados elegidos al azar de la empresa "The Second S.A." se muestran a continuación:

Primer S.A.	Rango (X)	The Second	Rango (Y)
14.3	36.0	13.7	26.0
14.0	30.0	11.4	12.5
14.2	33.5	13.2	24.0
14.2	33.5	13.1	22.5
13.0	21.0	12.4	17.5
9.2	2.0	10.3	5.0
14.7	38.0	10.7	8.0

9.1	1.0	13.6	25.0
14.8	39.5	11.4	12.5
14.0	30.0	10.9	11.0
10.0	4.0	13.9	27.5
12.0	15.0	11.7	14.0
14.2	33.5	10.8	9.5
14.8	39.5	12.4	17.5
14.2	33.5	13.1	22.5
14.4	37.0	10.8	9.5
12.8	20.0	10.6	7.0
9.3	3.0	12.3	16.0
13.9	27.5	14.0	30.0
10.5	6.0		
12.5	19.0		
	502.5		317.5

Pruebe si los salarios medianos de la empresa Primer S.A. son mayores a los salarios medianos de la empresa The Second S.A. Use $\alpha=0.05$

Primero se debe probar si ambos conjuntos provienen de la misma distribución

$H_0: F_x = F_y$

$H_1: F_x \neq F_y$

$\alpha=0.05$

Pvalor=0.1487

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que los salarios de los empleados de las empresas Primer S.A y The Second S.A provienen de distribuciones diferentes.

Se puede realizar la prueba de Mann Whitney

Solución

$H_0: Me_x \leq Me_y$

$H_1: Me_x > Me_y$

$\alpha=0.05$

Prueba Estadística

Como hay empates se utilizará la siguiente prueba:

$$Z = \frac{U - \frac{n_1(n_2)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \left[\sum_{i=1}^g \frac{t_i^3 - t_i}{(n_1 + n_2)(n_1 + n_2 - 1)} \right]}} \sim N(0,1) \text{ ó } p\text{-value} = P(Z < Z_c)$$

Criterio de Decisión

No se rechaza H_0 si $Z_c \leq 1.65$.

Se rechaza la H_0 si $Z_c > 1.65$.

ó

Si p-valor $\geq \alpha$ no se rechaza la H_0 .

Si p-valor $< \alpha$ se rechaza la H_0 .

Desarrollo de la Prueba

$$U_1 = (21)(19) + \frac{21(21+1)}{2} - 502.5 = 127.5$$

$$U_2 = (21)(19) + \frac{19(19+1)}{2} - 317.5 = 271.5$$

$$Z = \frac{127.5 - \frac{21(19)}{2}}{\sqrt{\frac{21(19)(21+19+1)}{12} - \left[\frac{6(2^3 - 2) + (3^3 - 3) + (4^3 - 4)}{(21+19)(21+19-1)} \right]}} = -1.95 \text{ ó}$$

$$p\text{-valor} = P(Z < -1.95) = 0.0256$$

Conclusiones

Existe suficiente evidencia estadística para rechazar la H_0 a un nivel de significación de 0.05.

Por lo tanto, se puede afirmar que los salarios medianos de las dos empresas en estudio difieren significativamente.

➤ Funciones en R

En esta función se debe ingresar los dos vectores de datos.

```
wilcox.test(vector de datos1, vector de datos2, alternativa,
valor hipotético, pareados=FALSE)
```

La función `wilcox.exact` del paquete `exactRankTests` permite obtener los resultados con el ajuste para empates

```
wilcox.exact(vector de datos1, vector de datos2,
alternativa, valor hipotético, paired=FALSE)
```

En el paquete `coin` también existe la función `wilcox_test`

```
wilcox_test(y~x)
```

➤ Resultados con R

```
ex<-c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,
15.7, 16.7,13.7,15.3,14.0)
ne<-c(17.4,16.2,17.1,17.5,15.0,16.0,16.9,15.0,16.3,16.8)
wilcox.test(ex,ne,alternative="l", paired=FALSE)
library(exactRankTests)
```

```
wilcox.exact(ex,ne,alternative="l", paired=FALSE)
library(coin)
dat<-c(ex,ne)
g<-as.factor(c(rep(1,length(ex)),rep(2,length(ne))))
library(coin)
wilcox_test(dat~g, alternative="l")
```

```
Wilcoxon rank sum test with continuity correction
data: ex and ne
W = 25, p-value = 0.003004
alternative hypothesis: true location shift is less than 0
```



```
Exact Wilcoxon rank sum test
data:  ex and ne
W = 25, p-value = 0.002093
alternative hypothesis: true mu is less than 0
```

```
Asymptotic Wilcoxon Mann-Whitney Rank Sum Test

data:  dat by g (1, 2)
Z = -2.7751, p-value = 0.002759
alternative hypothesis: true mu is less than 0
```

```
pri<-c(14.3, 14.0, 14.2, 14.2, 13.0, 9.2, 14.7,
9.1,14.8,14.0,10.0,12.0,
14.2,14.8,14.2,14.4,12.8,9.3,13.9,10.5,12.5)
sec<-c(13.7, 11.4, 13.2,
13.1,12.4,10.3,10.7,13.6,11.4,10.9,13.9,11.7,10.8,
12.4,13.1,10.8,10.6,12.3,14.0)
mean(pri)
12.8619
mean(sec)
12.12105
ks.test(pri,sec+.7)
```

```
Two-sample Kolmogorov-Smirnov test

data:  pri and sec + 0.7
D = 0.3609, p-value = 0.1487
alternative hypothesis: two-sided
```

```
library(exactRankTests)
wilcox.exact(pri,sec,alternative="g")
```

```
Exact Wilcoxon rank sum test

data:  pri and sec
W = 271.5, p-value = 0.02552
alternative hypothesis: true mu is greater than 0
```

➤ **Algunas consideraciones en R**

La función `wilcox.test`:

- Permite analizar una hipótesis bilateral o unilateral para un valor hipotético específico.
Para un ajuste para empates se debe utilizar la función `wilcox.exact`

La función `wilcox_test`

- Permite solo realizar hipótesis cuando el valor hipotético es igual a cero.

B. Prueba de Fligner-Policello

➤ Aspectos Generales

La prueba de Mann Whitney para comparar la mediana de dos muestras independientes exige que los datos de las muestras independientes provengan de distribuciones con similar forma.

En muchas situaciones esta condición no se cumple por lo que es necesario recurrir a otra prueba como la de Fligner-Policello.

➤ Supuestos

- La variable de interés se encuentra medida en una escala por lo menos intervalo.
- Los datos de cada una de las muestras provienen de distribuciones simétricas.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

El estadístico está basado en una cantidad denominada “colocación”. Una colocación es el número de valores en una muestra estrictamente menor que un valor dado de una segunda muestra.

Consideremos muestras X_1, X_2, \dots, X_{n_1} y Y_1, Y_2, \dots, Y_{n_2} . Para un valor X_i , $i = 1, 2, \dots, n_1$, su correspondiente colocación P_i es el número de valores de Y_1, Y_2, \dots, Y_{n_2} menores que X_i . Similarmente para un valor Y_j , $j = 1, 2, \dots, n_2$, su correspondiente colocación Q_j es el número de valores de la primera muestra menores que Y_j .

Los siguientes pasos involucran los cálculos similares a los requeridos para la prueba t de Student para dos muestras.

Se debe calcular los promedios y las varianzas de las colocaciones

$$\bar{P} = \frac{1}{n_1} \sum_{i=1}^{n_1} P_i \quad y \quad \bar{Q} = \frac{1}{n_2} \sum_{j=1}^{n_2} Q_j$$
$$V_1 = \sum_{i=1}^{n_1} (P_i - \bar{P})^2 \quad y \quad V_2 = \sum_{j=1}^{n_2} (Q_j - \bar{Q})^2$$

La prueba estadística de Fligner –Policello es:

$$F = \frac{n_1 \bar{P} - n_2 \bar{Q}}{2\sqrt{V_1 + V_2 + \bar{P}\bar{Q}}}$$

Se rechaza H_0 a favor de H_1 si $F > \varepsilon_\alpha$, donde ε_α es el valor crítico de la distribución de Fligner-Policello. Por el Teorema Central de Límite, $F \sim N(0,1)$, ε_α puede ser convenientemente reemplazado con el cuantil de la distribución normal estándar cuando n_1 y n_2 son grandes.

Si hay empates se necesita ajustar el estadístico de Fligner Policello, definido arriba.

El ajuste se debe realizar en las colocaciones. Cuando se calcula un conjunto de colocaciones, por ejemplo para calcular P_i , contamos el número de Y 's que son menores que X_i y adicionamos la mitad de valores igual a X_i . Lo mismo se realiza para el otro conjunto de colocaciones. El balance del método permanece sin cambios.

➤ **Aplicación**

Considere los datos del experimento donde se expusieron 15 animales a óxido de cadmio y otros 10 que no fueron expuestos para analizar el nivel de hemoglobina.

Se desea saber si puede concluirse que la inhalación prolongada de óxido de cadmio disminuye el nivel de hemoglobina. Use $\alpha=0.05$.

Primero se debe evaluar la simetría de los dos conjuntos de datos

$H_0: A_{s1}=0$ $H_1: A_{s1} \neq 0$ $\alpha=0.05$ $P_{valor}=0.2929$ Conclusión A un nivel de significación de 0.05, no se puede afirmar que los datos del nivel de hemoglobina provienen de una distribución asimétrica	$H_0: A_{s2}=0$ $H_1: A_{s2} \neq 0$ $\alpha=0.05$ $P_{valor}=0.5465$ Conclusión A un nivel de significación de 0.05, no se puede afirmar que los datos del nivel de hemoglobina provienen de una distribución asimétrica
--	--

Solución

$H_0: Me_1 \geq Me_2$

$H_1: Me_1 < Me_2$

$\alpha=0.05$

Prueba Estadística

$$U = 3.873 \quad p_{valor} = 0.0012$$

Conclusiones

Existe suficiente evidencia estadística para rechazar la H_0 a un nivel de significación de 0.05.

Por lo tanto, se puede afirmar que la inhalación prolongada de óxido de cadmio si disminuye el nivel de hemoglobina.

➤ **Funciones con R**

Se puede hacer uso de las funciones:

`fp.test` del paquete `RVAideMemoire`

`pFligPoli` del paquete `NSM3`

`fp.test` del paquete `npsm`

`mod.wmw.test` del paquete `robustrank`

➤ **Resultados con R**

```
library(lawstat)
```

```
symmetry.test(ex,option="MGG",boot=F)
```

```
Symmetry test by Miao, Gel, and Gastwirth (2006)
data:  ex
Test statistic = -1.0518, p-value = 0.2929
alternative hypothesis: the distribution is asymmetric.
```

```
symmetry.test(ne,option="MGG",boot=F)
```

```
Symmetry test by Miao, Gel, and Gastwirth (2006)
data:  ne
Test statistic = -0.60299, p-value = 0.5465
alternative hypothesis: the distribution is asymmetric.
```

```
library(RVAideMemoire)
fp.test(ex,ne,delta=0,alternative="less")
```

```
Fligner-Policello test
data:  ex and ne
U* = 3.873, p-value = 1
```

```
library(NSM3)
pFligPoli(ex,ne,method="Asymptotic")
```

```
Number of X values: 15 Number of Y values: 10
Fligner-Policello U Statistic: 3.873
Asymptotic upper-tail probability: 1e-04
Asymptotic two-sided p-value: 1e-04
```

```
library(npsm)
fp.test(ex,ne,delta=0,alternative="less")
```

```
statistic = 3.872983 , p-value = 0.9999542
```

```
library(robustrank)
mod.wmw.test(ex,ne,alternative="l",method="fp")
[1] 0.0012
```

➤ **Algunas consideraciones con R**

Se debe tener cuidado con esta prueba pues como se ha visto presenta resultados muy diferentes con respecto a la prueba de Mann Whitney cuando se usa la función `fp.test` de los paquetes `RVAideMemoire` y `npsm`.

La función `pFligPoli` solo permite realizar el caso bilateral.

Si se quiere probar desigualdad para el test de Fligner-Policello usaría el paquete `NSM3` y si se quiere probar “great” y “less” el paquete `robustrank`. Lo malo de esas funciones es que asumen que el valor hipotético es igual a cero.

3. Para evaluar un parámetro de escala (dispersión)

Las pruebas que se desarrollarán a continuación son consideradas alternativas no paramétricas de la prueba F de razón de varianzas.

Se usará θ_i para denotar al parámetro de dispersión de la población i (se refiere en general a cualquier medida de dispersión)

Hipótesis

Bilateral		Unilateral
Caso A	Caso B	Caso C
$H_0 : \theta_1 = \theta_2$	$H_0 : \theta_1 \leq \theta_2$	$H_0 : \theta_1 \geq \theta_2$
$H_1 : \theta_1 \neq \theta_2$	$H_1 : \theta_1 > \theta_2$	$H_1 : \theta_1 < \theta_2$

A. Prueba de Homogeneidad de Varianzas de Conover

➤ Aspectos Generales

La prueba F es utilizada para comparar las varianzas de dos poblaciones independientes asumiendo que las muestras provienen de una distribución normal. Sin embargo, esta prueba paramétrica tiene un pobre rendimiento cuando no se cumple el supuesto de normalidad.

Surgen diferentes alternativas y una de ellas es la prueba de homogeneidad de varianzas propuesta por Conover.

Esta prueba solo permite realizar el caso bilateral.

➤ Supuesto

- La variable de interés se debe encontrar en una escala de medición por lo menos intervalo.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Si X_1, X_2, \dots, X_{n1} y Y_1, Y_2, \dots, Y_{n2} son dos muestras independientes de poblaciones con funciones de distribución acumulada F_X y G_Y . Se deben calcular las medias muestrales \bar{x} y \bar{y}
- Calcular $(x_i - \bar{x})^2$ y $(y_i - \bar{y})^2$
- Obtener $\tilde{R}(x_i)$ y $\tilde{R}(y_i)$ los rangos de todas las $n = n_1 + n_2$ diferencias al cuadrado previamente calculadas.
- Calcular el estadístico

$$T = \sum_{i=1}^{n_1} \tilde{R}(x_i)$$

Si hay empates es mejor usar:

$$T^* = \frac{T - n_1 V_R}{\sqrt{\frac{n_1 n_2}{n(n-1)} W_R - \frac{n_1 n_2}{(n-1)} V_R^2}}$$

$$V_R = n^{-1} \left[\sum_{i=1}^{n_1} \tilde{R}(x_i)^2 + \sum_{i=1}^{n_2} \tilde{R}(y_i)^2 \right] \quad W_R = \sum_{i=1}^{n_1} \tilde{R}(x_i)^4 + \sum_{i=1}^{n_2} \tilde{R}(y_i)^4$$

Si se utilizan muestras n_1 y n_2 pequeñas se deben recurrir a tablas estadísticas; si las muestras son grandes se puede hacer uso de la siguiente aproximación a la distribución normal

$$T \sim N(\mu_T, \sigma_T^2) \text{ con } \mu_T = \frac{n_1(n+1)(2n+1)}{6} \quad \sigma_T^2 = \frac{n_1 n_2 (n+1)(2n+1)(8n+11)}{180}$$

Por lo tanto

$$Z_c = \frac{T - \mu_T}{\sigma_T}$$

➤ **Aplicación**

Ejemplo

Con los datos del Ejemplo 1 de la prueba de Mann Whitney verifique si los datos presentan varianza distinta. Use $\alpha=0.05$.

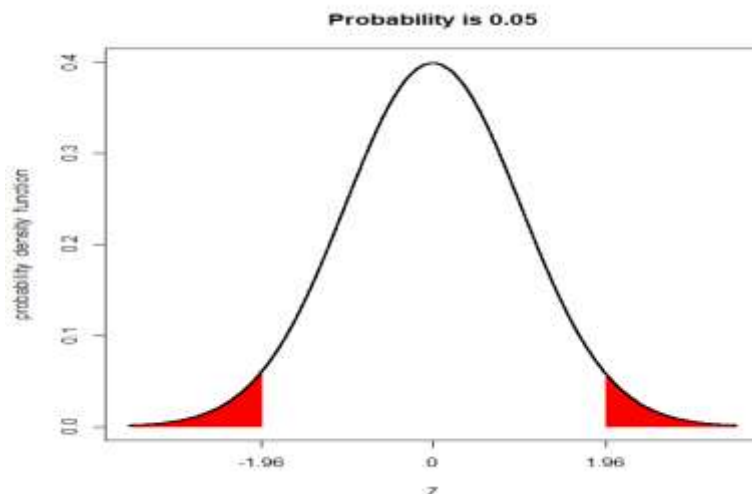
$$H_0 : \theta_1 = \theta_2$$

$$H_1 : \theta_1 \neq \theta_2$$

$$\alpha=0.05$$

$$Z_{cal} = -6.41$$

```
library(fastGraph)
shadeDist(qnorm(c(0.05/2, 1-0.05/2), lower.tail=F), "dnorm")
```



Conclusión

A un nivel de significación de 0.05, se puede afirmar que las varianzas del nivel de hemoglobina de los animales no expuestos y expuestos al óxido de cadmio son diferentes.

➤ **Funciones con R**

```
n1<-length(x)
n2<-length(y)
n<-n1+n2
xc<-(x-mean(x))^2
yc<-(y-mean(y))^2
dac<-c(xc,yc)
daco<-rank(dac)
rx<-daco[1:n1]
ry<-daco[(n1+1):n]
T<-sum(rx)
V<-(sum(rx^2)+sum(ry^2))/n
W<-sum(rx^4)+sum(ry^4)
Tc<-(T-n1*V)/sqrt((n1*n2*W)/(n*(n-1))-((n1*n2*V^2)/(n-1)))
mut<-(n1*(n+1)*(2*n+1))/6
s2t<-(n1*n2*(n+1)*(2*n+1)*(8*n+11))/180
z<-(T-mut)/sqrt(s2t)
```

Esta la función `conover_test` del paquete `coin`

```
library(coin)
ex<-c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,
15.7, 16.7,13.7,15.3,14.0)
ne<-c(17.4,16.2,17.1,17.5,15.0,16.0,16.9,15.0,16.3,16.8)
dat<-c(ex,ne)
g<-as.factor(c(rep(1,length(ex)),rep(2,length(ne))))
library(coin)
conover_test(dat~g)
```

Asymptotic Two-Sample Conover-Iman Test

```
data: dat by g (1, 2)
Z = 0.9718, p-value = 0.3311
alternative hypothesis: true ratio of scales is not equal to
1
```

B. Prueba de Reacciones Extremas de Moses

➤ Aspectos Generales

Esta prueba tiene el mismo objetivo que las pruebas de Conover y solo es utilizada para el caso bilateral. Moses

➤ Supuestos

- La variable en estudio se encuentra medida en una escala por lo menos ordinal.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

Sean X_1, X_2, \dots, X_{n_1} y Y_1, Y_2, \dots, Y_{n_2} dos muestras independientes

- Rankear las observaciones como si fuera una única muestra.
- Calcular la amplitud del grupo 1 ($X_{\max} - X_{\min}$) y sumando 1 a esta diferencia (redondear el resultado al entero más próximo).

Dado que la amplitud es una medida de dispersión muy inestable, Moses sugiere utilizar la amplitud recortada (A_r). Para ello se fija un valor pequeño (r) y se calcula la amplitud después de descartar r valores por arriba y por debajo ($2r$ valores en total)

La amplitud recortada se obtiene restando los rangos correspondientes al valor más grande y más pequeño del primer grupo tras eliminar del cómputo los r valores mas grandes y los r valores más pequeños y sumando 1 a esa diferencia y redondeando al entero más próximo.

Es evidente que A_r no puede ser menor a $n_1 - 2r$ (ni mayor que $n_1 + n_2 - 2r$). Además, si en el grupo 2 se han producido reacciones extremas, la amplitud del grupo 1 tenderá a su valor mínimo pues habrá pocas observaciones del grupo 2 entremezcladas con las del grupo 1. Por lo tanto, podría resultar muy informativo conocer la probabilidad asociada a los valores A_r que superen en alguna cantidad el valor $n_1 - 2r$. Si llamamos s a la cantidad en que un determinado valor de observado de A_r supera a $n_1 - 2r$, podemos obtener la probabilidad de encontrar valores A_r como el observado o menores mediante:

$$P(A_s \leq n_1 - 2r + s) = \frac{\sum_{i=0}^s \left[\binom{i + n_1 - 2r - 2}{i} \binom{n_2 + 2r + 1 - i}{n_2 - i} \right]}{\binom{n_1 + n_2}{n_1}}$$

Si esta probabilidad es pequeña se rechaza la hipótesis de que ambas muestras provienen de poblaciones con la misma amplitud.

➤ Aplicación

Ejemplo

Con los datos del Ejemplo 1 de la prueba de Mann Whitney verifique si los datos presentan distinta dispersión. Use $\alpha=0.05$.

$$H_0 : \theta_1 = \theta_2$$

$$H_1 : \theta_1 \neq \theta_2$$

$\alpha=0.05$

p-valor=0.1281 > α no se rechaza H_0 .

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que las varianzas del nivel de hemoglobina de los animales no expuestos y expuestos al óxido de cadmio son diferentes.

➤ **Funciones con R**

Esta prueba se encuentra dentro del paquete DescTools y se debe utilizar la función `MosesTest`

`MosesTest(vector de datos1, vector de datos2)`

➤ **Resultados con R**

```
ex<-c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,  
15.7, 16.7,13.7,15.3,14.0)  
ne<-c(17.4,16.2,17.1,17.5,15.0,16.0,16.9,15.0,16.3,16.8)  
MosesTest(ex,ne)
```

Moses Test of Extreme Reactions

```
data: ex and ne  
S = 18, p-value = 0.1281  
alternative hypothesis: extreme values are more likely in x  
than in y
```

C. Prueba de Mood

➤ Aspectos Generales

La prueba de Mood permite probar si dos parámetros de dispersión (amplitud, escala, variabilidad) son diferentes, uno es mayor o menor que el otro.

Si se obtienen los rangos de n observaciones, el promedio es $(n+1)/2$. La desviación del rango de la observación i con respecto a su media es $i-(n+1)/2$ y se puede considerar como un indicador de dispersión. Sin embargo, como en el caso de la clásica estadística descriptiva las desviaciones pueden ser positivos o negativos. La solución es usar el valor absoluto o el cuadrado de las desviaciones. Un valor grande de las dispersiones podría implicar que los datos se encuentran muy dispersos y un valor pequeño implica lo contrario.

Esta prueba permite realizar los tres tipos de hipótesis (bilateral, unilateral derecha e izquierda).

➤ Supuestos

- La variable de interés se encuentra medida en una escala por lo menos intervalo.
- Las dos poblaciones son idénticas (forma y medianas iguales) excepto por una posible diferencia en dispersión.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Los datos consisten en dos muestras aleatorias X_1, \dots, X_{n_1} y Y_1, \dots, Y_{n_2} , de dos poblaciones 1 y 2 respectivamente.
- Combinar las dos muestras ($n=n_1+n_2$) y se obtienen el rango de todas las observaciones (de menor a mayor), teniendo presente a cuál muestra pertenece cada observación.
- Determinar r_i que es el rango del primer grupo X de todos los datos combinados.
- Desarrollar la siguiente prueba estadística

$$M = \sum_{i=1}^{n_1} \left(r_i - \frac{n+1}{2} \right)^2$$

Si las muestras son pequeñas se debe contrastar el valor calculado con el valor crítico obtenido de tablas.

Sin embargo, si las muestras son grandes se puede hacer uso de la siguiente aproximación a la distribución Normal propuesta por Laubscher:

$$Z_c = \frac{M - E(M)}{\sqrt{V(M)}}$$

Donde:

$$E(M) = \frac{n_1(n^2-1)}{12} \text{ y } V(M) = \frac{n_1 n_2 (n+1)(n^2-4)}{180}$$

➤ **Aplicación**

Ejemplo

Se tiene datos sobre puntajes en la PC1 de alumnos seleccionados al azar de dos secciones. Verifique si presentan las notas de los alumnos de la primera sección presentan mayor dispersión que las de la segunda sección. Use $\alpha=0.05$.

```
S1<-c(14,14,13,15,14,16,15,15,14,15,15,16,13,15,14)
```

```
S2<-c(15,16,15,15,15,14,16,15,16,14)
```

Primero se debe probar si ambos conjuntos provienen de la misma distribución

$H_0: F_x = F_y$

$H_1: F_x \neq F_y$

$\alpha=0.05$

Pvalor=0.787

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que las distribuciones de las notas de los alumnos de las dos secciones en estudio provienen de distribuciones diferentes.

Segundo se debe probar si ambos conjuntos no presentan diferente mediana

$H_0: Me_1 = Me_2$

$H_1: Me_1 \neq Me_2$

$\alpha=0.05$

Pvalor=0.1545

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que las notas de los estudiantes de las secciones en estudio son diferentes.

$H_0: \theta_1 \leq \theta_2$

$H_1: \theta_1 > \theta_2$

$\alpha=0.05$

Pvalor= 0.6989

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que la dispersión de las notas de los alumnos de la primera sección es mayor que las de la segunda sección.

➤ **Funciones en R**

```
mood.test(vector de datos1, vector de datos2, alternativa)
```

➤ **Resultados con R**

```
S1<-c(14,14,13,15,14,16,15,15,14,15,15,16,13,15,14)
```

```
S2<-c(15,16,15,15,15,14,16,15,16,14)
```

```
mood.test(ex,ne,alternative="g")
```

```
Mood two-sample test of scale
```

```
data: ex and ne
```

```
Z = -0.52112, p-value = 0.6989
```

```
alternative hypothesis: greater
```

D. Prueba de Ansari-Bradley

➤ **Aspectos Generales**

Contrasta la hipótesis sobre la dispersión de dos poblaciones continuas, medida por un parámetro al que denotaremos como θ , pudiéndose dar los siguientes casos respecto a las medianas: iguales conocidas, distintas conocidas y desconocidas.

Si dos poblaciones A y B son idénticas en todos sus aspectos, incluida la dispersión y de cada una se extrae una muestra aleatoria simple, mutuamente independientes de tamaños n_1 y n_2 , es de esperar que las dispersiones muestrales sean parecidas, por lo que las observaciones en la muestra conjunta ordenada se repartirán al azar. Por el contrario, al diferir las poblaciones exclusivamente en la dispersión, los elementos de la muestra menos dispersa formarán un enclave dentro de las observaciones de la muestra más dispersa, en otras palabras tenderán a ser interiores con respecto a los de la muestra más dispersa.

➤ **Supuestos**

- Las poblaciones de donde se extraen las muestras tienen la misma mediana.
- La variable de interés se encuentra medida en una escala por lo menos intervalo.

➤ **Inferencia Estadística**

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Combinar las dos muestras ($n=n_1+n_2$) y ordenarlas de menor a mayor
- Asignar a las observaciones coeficientes de la siguiente forma: al primero y al último el coeficiente 1, al segundo y penúltimo el coeficiente 2 y así sucesivamente resultando:

Si n es par

i	1	2	3	4	...	n/2	...	n-3	n-2	n-1	n
R _i	1	2	3	4	...	n/2	...	4	3	2	1

Si n es impar

i	1	2	3	4	...	(n+1)/2	...	n-3	n-2	n-1	n
R _i	1	2	3	4	...	(n+1)/2	...	4	3	2	1

Cuando las dispersiones sean significativamente distintas los coeficientes R_i , correspondientes a la muestra con mayor dispersión tenderán a ser menores (por ser exteriores) que los de la otra muestra (interiores) o lo equivalente la suma de los primeros tenderá a ser menor que la de los segundos.

- Desarrollar la siguiente prueba estadística asociada a los coeficientes asignados a la muestra A. Si las medianas poblacionales (Me_A y Me_B) son conocidas y distintas se calculan nuevos valores de las observaciones $x'_i = x_i - Me_A$ e $y'_i = y_i - Me_B$ y se aplica la prueba a estos datos. Si no se conocen las medianas poblacionales se pueden sustituir por las medianas muestrales.

$$W = \sum_{i=1}^{n_1} R_i$$

Si W toma un valor reducido indica que la variabilidad de la población A es mayor que la de la B y si W es grande es mayor la dispersión de B.
 Si las muestras son pequeñas se debe comparar el valor de W con un valor tabular de Ansari-Bradley.

Cuando las muestras n_1 y n_2 son grandes se puede hacer uso de la aproximación asintótica a la Normal, dependiendo de la cantidad n de observaciones

Si n es par	Si no existen empates	Si existen empates
$E(W) = \frac{n_1(n+2)}{4}$	$V(W) = \frac{n_1 n_2 (n^2 - 4)}{48(n-1)}$	$V_c(W) = \frac{n_1 n_2 \sum_{h=1}^g t_h r_h^2}{n(n-1)} \frac{n_1 n_2 (n+2)^2}{16(n-1)}$
Si n es impar	Si no existen empates	Si existen empates
$E(W) = \frac{n_1(n+1)^2}{4n}$	$V(W) = \frac{n_1 n_2 (n+1)(n^2 + 3)}{48n^2}$	$V_c(W) = \frac{n_1 n_2 \sum_{h=1}^g t_h r_h^2}{n(n-1)} \frac{n_1 n_2 (n+1)^4}{16n^2(n-1)}$

En las expresiones de empates, g indica el número de rangos diferentes cuando aparecen observaciones con valores repetidos, t_h , es el tamaño del h -ésimo grupo y r_h el rango medio del h -ésimo grupo.

Por lo tanto

$$Z_c = \frac{W - E(W)}{\sqrt{V(W)}} \quad \text{o} \quad Z_c = \frac{W - E(W)}{\sqrt{V_c(W)}}$$

➤ **Aplicación**

Ejemplo

Con los datos del Ejemplo de la prueba de Mood verifique si las notas de la sección 1 son más dispersas que los de la sección 2. Use $\alpha=0.05$.

Primero se debe probar si ambos conjuntos no presentan diferente mediana

$H_0: Me_1 = Me_2$

$H_1: Me_1 \neq Me_2$

$\alpha=0.05$

Pvalor=0.1545

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que las notas de los estudiantes de las secciones en estudio sean diferentes.

$H_0: \theta_1 \leq \theta_2$

$H_1: \theta_1 > \theta_2$

$\alpha=0.05$

Pvalor= 0.6705

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que la dispersión de las notas de los alumnos de la primera sección es mayor que las de la segunda sección.

➤ **Funciones en R**

```
ansari.test(vector de datos1, vector de datos2, alternativa)
```

También se encuentra la función `ansari_test` del paquete `coin`

```
ansari_test(formula, alternativa)
```

Existe la función `ansari.exact` del paquete `ExactRankTests` que permite obtener los resultados corregidos por empates. Se recomienda utilizar los resultados de esta función.

➤ **Resultados con R**

```
S1<-c(14,14,13,15,14,16,15,15,14,15,15,16,13,15,14)
```

```
S2<-c(15,16,15,15,15,14,16,15,16,14)
```

```
ansari.test(S1,S2, alternative="g")
```

```
Ansari-Bradley test
data:  ex and ne
AB = 107, p-value = 0.5345
alternative hypothesis: true ratio of scales is not equal to
1
Mensajes de aviso perdidos
In ansari.test.default(ex, ne) : cannot compute exact p-
value with ties
```

```
n1<-length(S1)
n2<-length(S2)
g<-as.factor(c(rep(1,n1),rep(2,n2)))
dat<-c(S1,S2)
library(coin)
ansari_test(dat~g, alternative="g")
```

```
Asymptotic Two-Sample Ansari-Bradley Test

data:  dat by g (1, 2)
Z = -0.4012, p-value = 0.3441
alternative hypothesis: true ratio of scales is greater than
1
```

```
library(exactRankTests)
ansari.exact(S1,S2, alternative="g")
```

```
Ansari-Bradley test

data:  S1 and S2
AB = 105, p-value = 0.6705
alternative hypothesis: true ratio of scales is greater than
1
```

E. Prueba de Siegel-Tukey

➤ **Aspectos Generales**

La prueba de Siegel-Tukey, fue publicada en 1960 por Sidney Siegel y John Wilder Tukey en una revista de la Asociación Americana de Estadística en el artículo "Un procedimiento no paramétrico de suma de rangos para una distribución relativa en muestras no pareadas".

Esta prueba únicamente tiene valor cuando existen razones a priori para creer que una condición experimental nos conducirá a puntuaciones extremas en cualquier dirección manteniendo la misma mediana.

Esta prueba no paramétrica se utiliza para determinar si uno de los dos grupos de datos tiende a tener valores más dispersos que los otros. En otras palabras, la prueba determina si uno de los dos grupos tiende a moverse, a veces a la derecha, a veces a la izquierda, pero lejos del centro.

➤ **Supuestos**

- La escala de medición que representan los datos es al menos ordinal.
- Las muestras provienen de poblaciones tienen medianas similares.

➤ **Inferencia Estadística**

La prueba consiste en que, una vez ordenados todos los elementos de ambas muestras, combinadas, se asignan rangos comenzando desde el menor y el mayor, hacia el centro: al menor valor se le asocia el rango 1; al valor más elevado y al que le precede se asignan los rangos 2 y 3; al segundo y tercer valores más bajos se asignan los rangos 4 y 5, y así sucesivamente. Si el número total de observaciones en ambas muestras es par, una de ellas se quedará sin rango. Las expresiones anteriores se utilizan para calcular el estadístico R_m , que es la suma de rangos de la muestra de menor tamaño. La interpretación del contraste estriba en que, si una de las dos muestras procede de una población con mayor dispersión, recibirá los rangos menores, mientras que la que procede de una muestra de menor variabilidad recibirá los rangos mayores. Puede apreciarse que el contraste tiene interés cuando condicionamos en que ambas distribuciones tienen una media de posición central similar.

Estos son los pasos que hay que seguir en la aplicación de la prueba de Siegel-Tukey:

- Se combinan las puntuaciones de los grupos n_1 y n_2 y se arreglan en una sola serie ordenada, cuidando de identificar cada puntuación correspondiente a cada grupo.
- Se asigna un rango a cada puntuación ordenada en la secuencia; cada rango se asigna alternando los extremos de la secuencia ordenada. Se asignan rangos comenzando desde el menor al mayor, hacia el centro: al menor valor se le asocia el rango 1; al valor más elevado y al que le precede se asignan los rangos 2 y 3; al segundo y tercer valores más bajos se asignan los rangos 4 y 5, y así sucesivamente. Si el número total de observaciones en ambas muestras es par, una de ellas se quedará sin rango.

$X_{(1)} \rightarrow$ rango 1

$X_{(n)} \rightarrow$ rango 2

$X_{(n-1)} \rightarrow$ rango 3

$X_{(2)} \rightarrow$ rango 4

$X_{(3)} \rightarrow$ rango 5

Tabla de resumen:

$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$...	$X_{(n-3)}$	$X_{(n-2)}$	$X_{(n-1)}$	$X_{(n)}$
1	4	5	8	9	...	7	6	3	2

- Se calcula la sumatoria de rangos para los grupos n_1 y n_2 . Se obtiene el estadístico R_m que es la suma de rangos de la muestra de menor tamaño. Si uno de los dos grupos está más disperso, su suma será menor, debido a que recibirá más de los rangos bajos reservados para las colas extremas, mientras que el otro grupo recibirá más puntuaciones altas asignadas al centro (esto es análogo al Prueba de suma de rangos de Wilcoxon). Sin embargo, resulta evidente que la validez de la prueba queda en entredicho si los promedios de las series son sustancialmente distintos (en el caso extremo de que todas las medidas de una serie sean inferiores a todas las medidas de la otra, las sumas de sus rangos serán muy parecidas sea cual sea el nivel de dispersión de las dos muestras). Para corregir esta disfunción se le suma a cada una de las medidas del conjunto inferior la diferencia entre los promedios de las dos series, por lo que se eliminará el efecto producido por las medias diferentes, mientras que no se altera la dispersión de los datos. También deben corregirse los estadísticos en función del número de datos y tomándose como estadístico de prueba el menor valor de ellos.

Estadístico de prueba:

$$R_m = \sum_{i=1}^{n_m} R_i$$

Donde:

$n_m = \min(n_1, n_2)$ y R_i es el i – ésimo rango asignando

Estadístico de prueba corregido:

$$R_m = \sum_{i=1}^{n_m} R_i - \frac{n_m(n_m + 1)}{2}$$

Donde:

$n_m = \min(n_1, n_2)$ y R_i es el i – ésimo rango asignado

- Para muestras pequeñas, determine el valor-p o el estadístico tabular utilizando la tabla de la suma de rangos de Wilcoxon-Mann-Whitney. El valor calculado R_m o el nivel de significancia α se compara con el obtenido en la Tabla de Wilcoxon-Mann-Whitney, que se busca con los valores (n_1) como el número de observaciones del primer grupo, (n_2) como el número de observaciones del segundo grupo y/o con el nivel de significación α . Los criterios de decisión para muestras pequeñas son:

Si el valor tabular es comparable al estadístico R_m :

Si $H_1: \theta_x \neq \theta_y$, se rechaza H_0 si, $R_m < W_{\frac{\alpha}{2}}$ ó $R_m > W_{1-\frac{\alpha}{2}}$

Si $H_1: \theta_x < \theta_y$, se rechaza H_0 si, $R_m < W_{\frac{\alpha}{2}}$

Si $H_1: \theta_x > \theta_y$, se rechaza H_0 si, $R_m > W_{1-\frac{\alpha}{2}}$

- Para muestras mayores que 20 el estadístico R_m puede aproximarse a una distribución normal.

Si, $n_1 + n_2 > 20$

Donde: $n_1 + n_2 = n$

$$z^* = \frac{R_m - \frac{n_m(n+1)}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}} \sim N(0,1)$$

➤ Aplicación

Se quiere evaluar la dispersión del puntaje de un examen obtenido en dos secciones. Los resultados son:

Sección A: 33 62 84 85 88 93 97

Sección B: 4 16 48 51 66 98

Prueba de hipótesis:

$H_0: \theta_A = \theta_B$

$H_1: \theta_A \neq \theta_B$

$\alpha = 0.05$

Al combinar los grupos, se obtiene un grupo de 13 entradas. La clasificación se realiza por extremos alternos (el rango 1 es el más bajo, 2 y 3 son los dos más altos, 4 y 5 son los dos siguientes más bajos, etc.).

Sección	B	B	A	B	B	A	B	A	A	A	A	A	B
Valor	4	16	33	48	51	62	66	84	85	88	93	97	98
Rango	1	4	5	8	9	12	13	11	10	7	6	3	2



Data fuente: grupos



Data ordenada



Extremos alternos

La suma de los rangos dentro de cada grupo W :

$$W_A = 5 + 12 + 11 + 10 + 7 + 6 + 3 = 54$$

$$W_B = 1 + 4 + 8 + 9 + 13 + 2 = 37$$

Si la hipótesis nula es verdadera, se espera que la suma de los rangos (teniendo en cuenta el tamaño de los dos grupos) será más o menos la misma.

Como ya se mencionó si uno de los dos grupos está más disperso, su suma será menor, debido a que recibirá más de los rangos bajos reservados para las colas extremas, mientras que el otro grupo recibirá más puntuaciones altas asignadas al centro.

A partir de las sumas de rangos, las estadísticas R_m corregidas se calculan restando el puntaje mínimo posible, $n_m(n_m + 1) / 2$ para cada grupo:

$$Rm_A = 54 - 7(8) / 2 = 26$$

$$Rm_B = 37 - 6(7) / 2 = 16$$

Ya que:

$$\Pr[X \leq \min(Rm_A, Rm_B)]$$

$$\Pr[X \leq \min(26, 16)]$$

Dónde:

$$X \sim \text{Wilcoxon}(m, n)$$

Es el valor p deseado, se puede usar una tabla de la distribución de suma de rangos de Wilcoxon para encontrar la significación estadística de los resultados. Usando los datos del ejemplo nuevamente, con grupos de tamaños 6 y 7:

$$\Pr[X \leq 16] = 0.2669$$

Por lo tanto, no se rechaza H_0 .

Conclusión:

A un nivel de significación de 0.05, no se puede afirmar que la dispersión del puntaje es diferente en las secciones en evaluación.

➤ **Funciones con R**

Esta prueba se encuentra dentro del paquete DescTools y se debe utilizar la función SiegelTukeyTest

SiegelTukeyTest (vector de datos1, vector de datos2, alternativa)

➤ **Resultados con R**

```
A<-c(33, 62, 84, 85, 88, 93, 97)
```

```
B<-c(4, 16, 48, 51, 66, 98)
```

```
library(DescTools)
```

```
SiegelTukeyTest(A,B,alternative="t")
```

Siegel-Tukey-test for equal variability

data: A and B

ST = 34, p-value = 0.4848

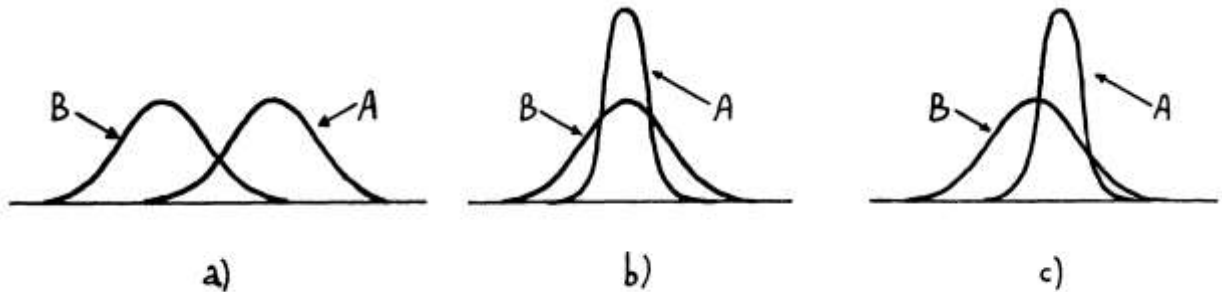
alternative hypothesis: true ratio of scales is not equal to 1

Existe también la función siegelTukeyTest del paquete PMCMRplus pero su pvalor es calculado con la aproximación a la normal.

4. Para evaluar un parámetro de locación y/o escala

Las pruebas que se presentan a continuación permiten evaluar un parámetro de locación y escala de manera simultánea.

Los casos en los que se rechazaría la hipótesis nula son los siguientes:



Cucconi, O. (1968). Ilustración de situaciones de rechazo de la hipótesis nula. [Figura]. Recuperado de: https://www.jstor.org/stable/23241361?seq=1#page_scan_tab_contents

- Las poblaciones A y B difieren en su parámetro de locación, pero no en el de escala.
- Las poblaciones A y B difieren en su parámetro de escala, pero no de locación.
- Las poblaciones A y B difieren tanto en su parámetro de escala, como el de locación.

Prueba de Hipótesis:

$$H_0: \Delta = 0 \text{ y } \tau^2 = 1$$
$$H_1: \Delta \neq 0 \text{ y/o } \tau^2 \neq 1$$

Donde: Δ es la diferencia de las medianas ($\Delta = Me_1 - Me_2$), y τ^2 es el ratio de los parámetros de escala ($\tau^2 = \frac{\theta_1}{\theta_2}$).

A. Prueba de Lepage

➤ Aspectos Generales

La prueba desarrollada por Lepage (1971) es la prueba no paramétrica comúnmente usada en la práctica en los campos de toxicología, medicina y epidemiología; cuando no solo el parámetro de posición difiere en las distribuciones de dos muestras, sino también el parámetro de dispersión. El estadístico de prueba es la suma estandarizada cuadrática de los estadísticos de prueba de Mann Whitney y Ansari-Bradley.

➤ Supuestos

- Las muestras provienen de poblaciones con distribuciones similares.
- La variable de interés en las dos muestras esta medida en una escala al menos intervalo.

➤ Inferencia Estadística

Dadas las fórmulas de Mann Whitney y Ansari Bradley (del presente capítulo), el estadístico posición-escala es denotado de la siguiente forma:

$$L = \frac{(W - E_0(W))^2}{\sqrt{Var_0(W)}} + \frac{(AB - E_0(AB))^2}{\sqrt{Var_0(AB)}}.$$

En caso de la prueba asintótica, se lleva a cabo a pesar de los pequeños tamaños de muestra; el p-valor basado en la distribución χ^2 con dos grados de libertad es igual $1 - P_0(\chi^2_2 \geq L)$.

➤ **Aplicación**

Baer and Schmid-Hempel (1999) desarrollaron un experimento para evaluar el tamaño de las colonias de abejorros *Bombus terrestris* (especie que habita en Europa). Reinas jóvenes fueron inseminadas con espermatozoides de alta (grupo 1) o baja (grupo 2) diversidad genética. Los datos son los siguientes:

Grupo	Tamaño de muestra	Observaciones
Alta diversidad	7	14,24,26,98,12,105,85
Baja diversidad	12	40,14,18,28,11,39,17,37,52,30,65,35

Se observa que el tamaño de colonia igual a 14 se presenta dos veces, lo cual sería un empate. Primero, debemos ordenar las observaciones de menor a mayor y asignar los respectivos rangos, según Mann Whitney tenemos que asignar un rango medio al empate y asignar los rangos según Ansari-Bradley para un total impar. En la siguiente tabla se muestra el desarrollo:

Grupo	Observaciones	Rango para AB	Rango para MW
2	11	1	1
1	12	2	2
1	14	3.5	3.5
2	14	3.5	3.5
2	17	5	5
2	18	6	6
1	24	7	7
1	26	8	8
2	28	9	9
2	30	10	10
2	35	9	11
2	37	8	12
2	39	7	13
2	40	6	14
2	52	5	15
2	65	4	16
1	85	3	17
1	98	2	18
1	105	1	19

Las filas resaltadas en amarillo representan nuestro grupo de interés, el cual nos servirá para hallar los resultados que se requieren para obtener L . Segundo, debemos empezar calculando W , su esperado y su variancia cuando hay empates; así:

$$W1 = 2 + 3.5 + 7 + 8 + 17 + 18 + 19 = 74.5$$

$$W2 = 1 + 3.5 + 5 + 6 + 9 + 10 + 11 + 12 + 13 + 14 + 15 + 16 = 115.5$$

$$U_1(W) = 7(12) + \frac{7(7+1)}{2} - 74.5 = 37.5$$

$$U_2(W) = 7(12) + \frac{12(12+1)}{2} - 115.5 = 46.5$$

$$E_0(W) = \frac{7(12)}{2} = 42 \text{ y } Var_0(W) = \frac{7(12)(19+1)}{12} - \frac{(2-1)2(2+1)}{19(19-1)} = 139.98$$

Tercero, calculamos el estadístico de Ansari-Bradley, su media y su variancia con empates, de esta manera:

$$AB = 2 + 3.5 + 7 + 8 + 3 + 2 + 1 = 26.5$$

$$E_0(AB) = \frac{\frac{1}{4}7(19+1)^2}{19} = 36.842 \text{ y } Var_0(AB) = \frac{7(12)(16*19*2*3.5^2 - 18^3)}{16*19^2*18} = 5.93$$

Cuarto, reemplazamos los valores obtenidos arriba para obtener L .

$$L = \left(\frac{74.5 - 70}{\sqrt{139.98}} \right)^2 + \left(\frac{26.5 - 36.842}{5.93} \right)^2 = 3.186$$

Por último, calculamos la prueba asintótica donde el p-valor se basa en la distribución χ^2 con dos grados de libertad; entonces obtenemos lo siguiente:

$$p - \text{valor} = 1 - P_0(\chi^2_2 \geq 3.186) = 1 - 0.7966852 = 0.2033$$

Dada la hipótesis siguiente:

$$H_0: \Delta = 0 \text{ y } \tau^2 = 1$$

$$H_1: \Delta \neq 0 \text{ y/o } \tau^2 \neq 1$$

Concluimos:

A un $\alpha=0.05$, no se puede afirmar que las medianas y/o varianzas de los grupos de alta y baja diversidad genética son diferentes.

➤ Funciones con R

```
alta_diversidad=c(14,24,26,98,12,105,85)
baja_diversidad=c(40,14,18,28,11,39,17,37,52,30,65,35)
```

Se puede realizar la prueba con `pLepage(x,y, method=...)` que está dentro del paquete NSM3. Se obtiene el siguiente resultado:

```
pLepage(alta_diversidad,      baja_diversidad,      method      =  
"Asymptotic")
```

<pre>Ties are present, so p-values are based on conditional null distribution. Number of X values: 7 Number of Y values: 12 Lepage D Statistic: 3.1754 Asymptotic upper-tail probability: 0.2044</pre>
--

Como podemos apreciar el estadístico de Lepage y el p-valor difieren de lo calculado manualmente, sin embargo, los valores son cercanos y en ambos procedimientos no se rechaza la hipótesis planteada, afirmando que no hay diferencias significativas en los parámetros de posición y escala en el grupo de alta diversidad genética y en el grupo de baja diversidad genética.

B. Prueba de Cucconi

➤ Aspectos Generales

Cucconi elaboró una prueba para evaluar los cambios simultáneos de locación y escala en su artículo llamado *“Un novo test non parametrico per il confronto fra due gruppi di valori campionari”*, publicado en la revista italiana *“Giornale degli Economisti e Annali di Economia”* en 1968. Se puede usar en campos muy variados, desde meteorológicos y médicos, hasta análisis de precios de acciones en el mercado bursátil. Su estadístico de prueba se basa en los cuadrados de los rangos de las muestras.

A pesar de que la prueba de Lepage es la más conocida para abordar este problema, Cucconi fue propuesta años antes y tiene varias ventajas particulares sobre esta, como lo señala Duarte (2018):

- No es una combinación de una prueba de locación y una de escala, como otras.
- Como lo muestra Marozzi (2009), se compara favorablemente con los contrastes de tipo Lepage en términos de potencia y probabilidad de error tipo I.
- Su cálculo es más sencillo.

Hay que tener en cuenta que, si bien la prueba fue propuesta por Cucconi en 1968, este solo proporcionó los valores críticos asintóticos, mientras que Marozzi calculó por primera vez una tabla de valores críticos exactos y estudió su potencia y probabilidad de error tipo I después de más de cuarenta años.

➤ Supuestos

- Las muestras son mutuamente independientes.
- La variable de interés en las dos muestras esta medida en una escala al menos intervalo.

➤ Inferencia Estadística

Se tiene $X_1 = (X_{11}, \dots, X_{1n_1})$ y $X_2 = (X_{21}, \dots, X_{2n_2})$, que son dos muestras aleatorias independientes de tamaño n_1 y n_2 , respectivamente.

La prueba de Cucconi, para evaluar el problema de locación-escala, propone:

$$CUC = \frac{1}{2} \sum_{k=1}^2 \frac{U_k^2 + V_k^2 - 2\rho U_k V_k}{2(1 - \rho^2)}$$

Donde:

$$U_k = \frac{6 \sum_{i=1}^{n_1} R_{ki}^2 - n_1(n+1)(2n+1)}{\sqrt{n_1 n_2 (n+1)(2n+1)(8n+11)/5}}$$

$$V_k = \frac{6 \sum_{i=1}^{n_1} (n+1 - R_{ki})^2 - n_1(n+1)(2n+1)}{\sqrt{n_1 n_2 (n+1)(2n+1)(8n+11)/5}}$$

$$n = n_1 + n_2; k = 1, 2$$

$$\rho = Cor(U_k, V_k) = \frac{2(n^2 - 4)}{(2n+1)(8n+11)} - 1$$

Además, R_{ki} denota los rangos de X_{ki} en la muestra $X = (X_1, X_2)$. Es decir, para realizar el ranqueo se toma la muestra de tamaño $n = n_1 + n_2$.

Como menciona Duarte (2018), podemos observar que U_k se basa en los cuadrados de los rangos R_{ki} , mientras que V_k se basa en los cuadrados de los rangos contrarios $(n + 1 - R_{ki})$.

En ausencia de empates, vemos que:

$$U_2 = -U_1 \text{ y } V_2 = -V_1$$

Por lo tanto, para calcular el estadístico CUC , solo será necesario calcular U y V en una de las muestras, y la nueva fórmula será:

$$CUC = \frac{U^2 + V^2 - 2\rho UV}{2(1 - \rho^2)}$$

Una vez establecido un nivel de significación α :

- No se rechazará la hipótesis nula si el valor del estadístico CUC no supera al $\ln(\alpha)$:

$$CUC < \ln(\alpha)$$

- Se rechazará la hipótesis nula en caso contrario:

$$CUC \geq \ln(\alpha)$$

(Cucconi, 1968)

➤ Aplicación

Los ingresos de empresas de elegidas al azar de dos zonas diferenciadas de una ciudad son:

Ingreso Zona D	Rango	Ingreso Zona E	Rango
490	1	1040	18
1090	19	610	9.5
590	7	590	7
1911	21	790	13
550	5	590	7
930	17	644	11
810	15	890	16
710	12	610	9.5
800	14	1220	20
529	4		
520	3		
510	2		
n1=12		n2=9	

$$\begin{array}{ll} H_0: \delta = 0 & \text{y} \quad \gamma^2 = 1 \\ H_1: \delta \neq 0 & \text{y/o} \quad \gamma^2 \neq 1 \end{array}$$

Se puede observar que en el conjunto de datos existen empates, es decir, observaciones repetidas.

1. Hallar:

$$U_1 = \frac{6 \sum_{i=1}^{n_1} R_{1i}^2 - n_1(n+1)(2n+1)}{\sqrt{n_1 n_2 (n+1)(2n+1)(8n+11)/5}}$$

$$= \frac{6 * (1760) - 12 * (21+1) * (42+1)}{\sqrt{12 * 9 * (21+1) * (42+1) * (168+11)/5}} = -0.41412$$

$$U_2 = \frac{6 \sum_{i=1}^{n_1} R_{2i}^2 - n_1(n+1)(2n+1)}{\sqrt{n_1 n_2 (n+1)(2n+1)(8n+11)/5}}$$

$$= \frac{6 * (1548.5) - 12 * (21+1) * (42+1)}{\sqrt{12 * 9 * (21+1) * (42+1) * (168+11)/5}} = -1.077653$$

$$V_1 = \frac{6 \sum_{i=1}^{n_1} (n+1 - R_{1i})^2 - n_1(n+1)(2n+1)}{\sqrt{n_1 n_2 (n+1)(2n+1)(8n+11)/5}}$$

$$= \frac{6 * 2288 - 12 * (21+1) * (42+1)}{\sqrt{12 * 9 * (21+1) * (42+1) * (168+11)/5}} = 1.24236$$

$$V_2 = \frac{6 \sum_{i=1}^{n_1} (n+1 - R_{2i})^2 - n_1(n+1)(2n+1)}{\sqrt{n_1 n_2 (n+1)(2n+1)(8n+11)/5}}$$

$$= \frac{6 * 1020.5 - 12 * (21+1) * (42+1)}{\sqrt{12 * 9 * (21+1) * (42+1) * (168+11)/5}} = -2.734133$$

$$\rho = Cor(U_k, V_k) = \frac{2(n^2 - 4)}{(2n+1)(8n+11)} - 1 = \frac{2(21^2 - 4)}{(42+1)(168+11)} - 1$$

$$= -0.8864493$$

$$CUC = \frac{1}{2} \sum_{k=1}^2 \frac{U_k^2 + V_k^2 - 2\rho U_k V_k}{2(1 - \rho^2)}$$

$$CUC = \frac{1}{2} \left(\frac{-0.41412^2 + 1.24236^2 - 2 * -0.8864 * -0.41412 * 1.24236}{2(1 - (-0.8864)^2)} \right.$$

$$\left. + \frac{-1.077653^2 + -2.734133^2 - 2 * -0.8864 * -1.077653 * -2.734133}{2(1 - (-0.8864)^2)} \right)$$

$$CUC = 17.11353$$

$$\ln(\alpha) = -2.995732 \rightarrow CUC > \ln(\alpha) \rightarrow \text{Se rechaza } H_0.$$

Conclusión:

A un nivel de significación de 0.05, se concluye que existe diferencia entre el parámetro de locación y/o escala del monto total de ganancias de la zona D y la zona E.

➤ Funciones con R

```
CucconiTestStat<-function(datos1,datos2) {
  #Se ingresan los VECTORES de datos
  datos<-c(datos1,datos2)
  n1<-length(datos1)
  n2<-length(datos2)
  n<-n1+n2

  #Calculando factores en común
  den<-sqrt(n1*n2*(n+1)*(2*n+1)*(8*n+11)/5)
  num<-n1*(n+1)*(2*n+1)
  rho<-((2*(n^2-4))/((2*n+1)*(8*n+11)))-1

  R<-rank(datos)
  R1<-R[1:n1]
  R2<-R[(n1+1):n]

  #Para muestra 1
  Uk1<-((6*sum(R1^2)-num)/den)
  Vk1<-((6*sum((n+1-R1)^2)-num)/den)

  #Para muestra 2
  Uk2<-((6*sum(R2^2)-num)/den)
  Vk2<-((6*sum((n+1-R2)^2)-num)/den)

  CUC<-((1/2)*((Uk1^2+Vk1^2-2*rho*Uk1*Vk1) + (Uk2^2+Vk2^2-
  2*rho*Uk2*Vk2))/(2*(1-rho^2)))

  return(list(Uk1=Uk1,Uk2=Uk2,Vk1=Vk1,Vk2=Vk2,R1=R1,R2=R2,
  CUC=CUC))
}

#APLICACIÓN DE LA FUNCIÓN CREADA
```

➤ **CucconiTestStat**(D\$Income,E\$Income)

```
## Uk1
## [1] -0.41412
##
## Uk2
## [1] -1.077653
##
```

```
## Vk1
## [1] 1.24236
##
## Vk2
## [1] -2.734133
##
## R1
## [1] 1 19 7 21 5 17 15 12 14 4 3 2
##
## R2
## [1] 18.0 9.5 7.0 13.0 7.0 11.0 16.0 9.5 20.0
##
## CUC
## [1] 17.11353
```

➤ **Algunas consideraciones con R**

Se debe tener en cuenta que existe, en R, la función “cuconi.teststat” dentro del paquete “nonpar” para calcular el estadístico de prueba de Cucconi. Esta no trabaja con la fórmula general, por lo que para conjuntos de datos que presenten empates en el ranqueo, el cálculo del estadístico no será correcto. Por esto, se recomienda utilizar la función programada en el presente documento. Lamentablemente esta función ya no está disponible en el paquete

5. Prueba de Permutación para comparar parámetro de locación o escala

➤ Aspectos Generales

La idea de permutar datos entre tratamientos es una forma de estadística inferencial que data de la época del fundador del padre de la estadística moderna Ronald Fisher (1935). Sin embargo, los teóricos han desarrollado esta idea desde entonces. Sin embargo, la disponibilidad de los sistemas de cómputo más veloces ha hecho que los estadísticos hagan uso de esta metodología.

La típica prueba para comparar la media de dos poblaciones requiere el cumplimiento de los supuestos de que las observaciones de las muestras a analizar sean: independientes y que provengan de distribuciones normales. Si esto no se cumple se corre el riesgo de declarar un resultado como estadísticamente significativo cuando realmente no lo es.

La asignación aleatoria de los sujetos experimentales a dos métodos que se desean comparar provee una idea básica de inferencia sin necesidad de asunciones asociadas a la prueba t. El argumento es el siguiente: “Si no hay diferencia entre los dos métodos, cuando se unen los dos grupos de datos en uno solo conjunto, entonces no debería existir diferencia en la mayoría de los conjuntos de datos cuando nuevamente se generen dos grupos de datos por asignación aleatoria”.

Por otro lado, si uno de los grupos tiende a ser mayor que el otro, la diferencia de medias de los puntajes de los grupos podría ser grande.

➤ Supuestos

- La variable de interés se encuentra medida en una escala por lo menos intervalo.

➤ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Si X_1, X_2, \dots, X_{n_1} y Y_1, Y_2, \dots, Y_{n_2} son dos muestras independientes, calcular la diferencia entre las medias de los dos grupos, a este valor se le denotará como D_{obs}
- Permutar las, $n_1 + n_2$ observaciones entre los dos grupos. El número de posibles permutaciones es:

$$\binom{n_1 + n_2}{n_1} = \frac{(n_1 + n_2)!}{n_1! n_2!}$$

- Para cada permutación calcular la diferencia, D , entre las dos medias de los dos grupos.
- Si se cree que el primer grupo produce observaciones que son en promedio más grandes que las del grupo 2 (prueba de cola superior), el cálculo del p-valor es la proporción de D 's mayores o iguales a D_{obs} .

$$P_{cola superior} = \frac{\# D's \geq D_{obs}}{\binom{n_1 + n_2}{n_1}}$$

Similarmente se calcula la proporción de D's menores o iguales que D_{obs} para pruebas cola inferior. Para pruebas de dos colas, se realiza un procedimiento similar considerando el valor absoluto de las diferencias de las medias estos es:

$$P_{dos\ colas} = \frac{\#|D's| \geq |D_{obs}|}{\binom{n_1 + n_2}{n_1}}$$

Sea θ_i es un parámetro de interés

Hipótesis

Bilateral

$$H_0 : \theta_1 = \theta_2$$

$$H_1 : \theta_1 \neq \theta_2$$

Unilateral

$$H_0 : \theta_1 \leq \theta_2$$

$$H_1 : \theta_1 > \theta_2$$

$$H_0 : \theta_1 \geq \theta_2$$

$$H_1 : \theta_1 < \theta_2$$

➤ **Aplicación**

Ejemplo

Con los datos del Ejemplo 1 de la prueba de Mann Whitney verifique si los datos presentan medias distintas. Use $\alpha=0.05$.

$$H_0 : \theta_1 = \theta_2$$

$$H_1 : \theta_1 \neq \theta_2$$

$$\alpha=0.05$$

$$p\text{-valor}= 0.002$$

Conclusión

A un nivel de significación de 0.05, se puede afirmar que las distribuciones teóricas del nivel de hemoglobina de los animales no expuestos y expuestos al óxido de cadmio son diferentes.

➤ **Funciones con en R**

Se presentará una serie de comandos para realizar la prueba para el caso bilateral

```
x<-c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,
15.7, 16.7,13.7,15.3,14.0)
y<-c(17.4,16.2,17.1,17.5,15.0,16.0,16.9,15.0,16.3,16.8)
n1<-length(x)
n2<-length(y)
n<-n1+n2
todo<-c(x,y)
umbral<-mean(x)-mean(y)
diferencias<-rep(0,1000)
set.seed(20)
for (i in 1:1000) {
mezcla<-sample(todo,n,F)
diferencias[i]<-mean(mezcla[1:n1])-mean(mezcla[(n1+1):n])
}
pvalor<-sum(abs(diferencias)>=abs(umbral))/1000
print(pvalor)
```

➤ **Resultados con R**

```
[1] 0.004
```

Existen varios paquetes en R que realizan la prueba de permutación

- La función “*perm.test*” del paquete *exactRankTests* realiza pruebas de permutación de una muestra (y pareada) y dos muestras independientes.
- La prueba de permutación de una muestra se basa en *stat (x-mu)*.
- La prueba de permutación pareada de dos muestras se basa en *stat (x-y-mu)*.
- Y la prueba de permutación no apareada de dos muestras se basa en *stat (x-mu) -stat (y)*.

```
library(exactRankTests)
x<-c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,
15.7, 16.7,13.7,15.3,14.0)
y<-c(17.4,16.2,17.1,17.5,15.0,16.0,16.9,15.0,16.3,16.8)
perm.test( x, y, alternative = "two.sided", stat= mean)
```

- La función “*twotPermutation*” del paquete *DAAG* calcula el valor p de la prueba t de dos muestras usando una prueba de permutación. La densidad de permutación también se pueden trazar.

```
library(DAAG)
twot.permutation(x,y,nsim=2000)
```

- La función *permTS* del paquete *perm* realiza prueba de permutación de 2 muestras, donde el estadístico de prueba es equivalente a la media de un grupo menos la media del otro. Existen 2 algoritmos para realizar la prueba de permutación: Exactos (network algorithm o Monte Carlo) o asintóticos ([*pclt*])

```
library(perm)
permTS(x,y)
```

- La función *perm.t.s* del paquete *RVAideMemoire* realiza la prueba t de Student de permutación.

```
library(RVAideMemoire)
perm.t.test(x, y)
```

Otras pruebas no paramétricas para comparan un parámetro de locación con datos provenientes de dos muestras independientes son: Normal y Savage.

Otras pruebas no paramétricas para comparar un parámetro de escala (dispersión) con datos provenientes de dos muestras independientes son: Taha, Klotz, Fligner.

Estas pruebas se pueden encontrar en el paquete *coin*.