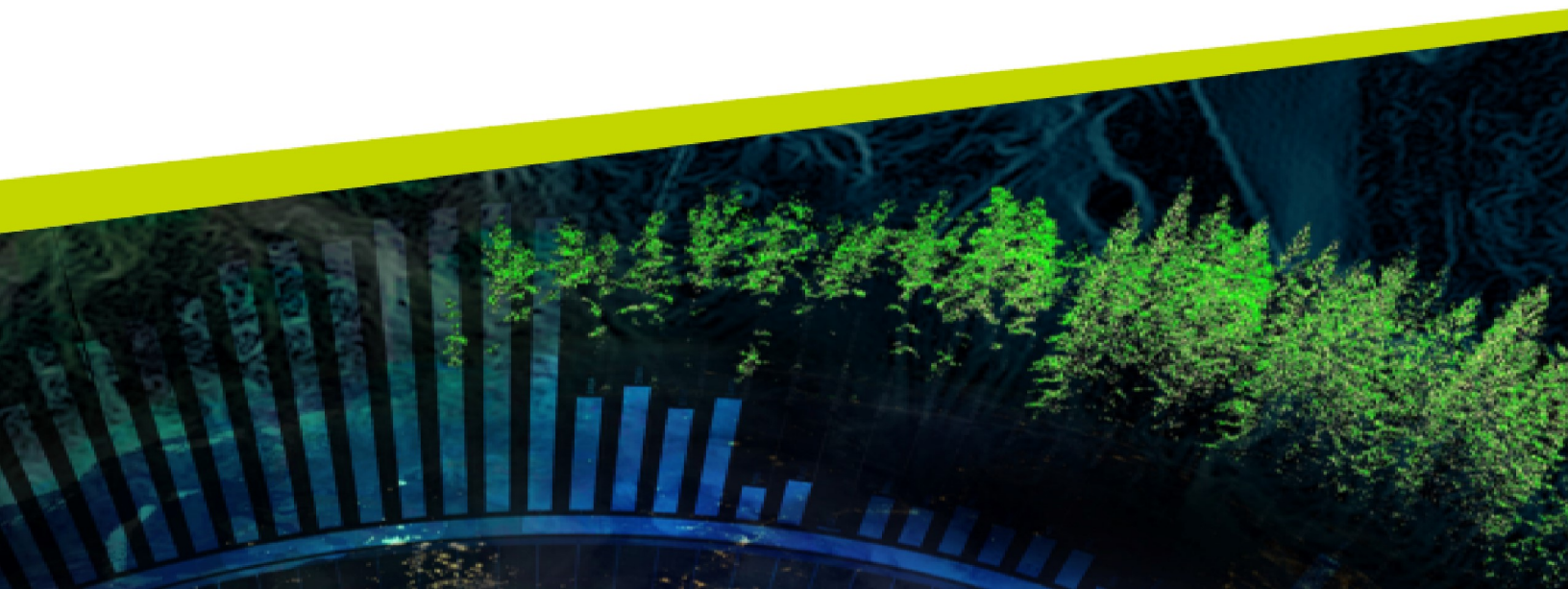




INFERENCIA Y MODELOS ESTADÍSTICOS

Jacqueline Köhler C. y José Luis Jara V.



11.3 PRUEBAS NO PARAMÉTRICAS CON MÁS DE DOS MUESTRAS NUMÉRICAS

Al igual que existen alternativas no paramétricas para inferir con una o dos muestras de una variable numérica, también las hay para cuando se tienen más de dos muestras. Conoceremos ahora alternativas no paramétricas clásicas para el procedimiento ANOVA de una vía, tanto para muestras independientes como para muestras correlacionadas.

11.3.1 Prueba de Kruskal-Wallis

En el capítulo 9 estudiamos el procedimiento ANOVA de una vía para $k > 2$ muestras independientes, el cual requiere el cumplimiento de los siguientes supuestos:

1. La escala con que se mide la variable dependiente tiene las propiedades de una escala de intervalos iguales.
2. Las k muestras son obtenidas de manera aleatoria e independiente desde la(s) población(es) de origen.
3. Se puede suponer razonablemente que la(s) población(es) de origen sigue(n) una distribución normal.
4. Las k muestras tienen varianzas aproximadamente iguales.

Si bien ANOVA es usualmente robusto ante desviaciones leves de las condiciones (excepto la segunda) cuando las muestras son de igual tamaño, no ocurre lo mismo cuando los tamaños de las muestras difieren. En este caso, una alternativa es emplear la **prueba de Kruskal-Wallis**, cuyas condiciones son:

1. La variable independiente debe tener a lo menos dos niveles (aunque, para dos niveles, se suele usar la prueba de Wilcoxon-Mann-Whitney).
2. La escala de la variable dependiente debe ser, a lo menos, ordinal.
3. Las observaciones son independientes entre sí.

Para ilustrar esta prueba, tomemos el ejemplo de un estudiante memorista en informática que para su proyecto de título estudia cuatro criterios diferentes (A , B , C y D) para optimizar una secuencia de operaciones *join* en tablas de bases de datos relacionales con millones de registros. Para sus experimentos, el memorista consiguió armar un conjunto de 48 bases de datos públicas de tamaños similares donde se requiere realizar una secuencia de 8 operaciones *join* para obtener un reporte. Luego, asignó cada base de datos a uno de los cuatro criterios de forma aleatoria y ejecutó las consultas en iguales condiciones de hardware y recursos computacionales, registrando el tiempo de ejecución (en segundos) requerido por cada *query*, obteniendo las siguientes mediciones:

- Criterio A : 21, 22, 22, 23, 23, 23, 23, 24, 24, 24, 25, 26
- Criterio B : 15, 17, 18, 18, 19, 19, 20, 20, 21
- Criterio C : 9, 10, 10, 10, 10, 11, 11, 12, 12, 12, 13, 14, 15
- Criterio D : 15, 15, 16, 16, 16, 18, 18, 18

Las hipótesis a contrastar son, entonces:

H_0 : todos los criterios llevan a consultas igualmente eficientes (o, de manera similar, ningún criterio genera consultas menos ni más eficiente que los demás).

H_A : al menos uno de los criterios genera consultas con eficiencia diferente a las generadas por al menos algún otro criterio.

El procedimiento de la prueba de Kruskal-Wallis tiene elementos similares a los descritos en las pruebas no paramétricas para una y dos medias. El primer paso consiste en combinar las muestras y luego asignar el rango a cada elemento, obteniéndose para el ejemplo el resultado de la tabla 11.7.

A continuación se calcula la suma (S^g) y la media (M^g) de los rangos en cada grupo ($g \in \{A, B, C, D\}$) y en la muestra combinada ($g = T$). La tabla 11.8 presenta los valores obtenidos para el ejemplo, incluyendo además el tamaño muestral de cada caso (n^g).

Como en las pruebas anteriores, usar rangos nos entrega información adicional sobre los datos. Por ejemplo, el promedio de los rangos está siempre dado por la ecuación 11.16.

Observaciones										Ranking de observaciones											
<i>A</i>			<i>B</i>			<i>C</i>			<i>D</i>		<i>A</i>			<i>B</i>			<i>C</i>			<i>D</i>	
21	23	25	15	19	9	11	12	15	18	31,5	36,5	42,0	15,5	27,5	1,0	8,5	8,5	15,5	24,0		
22	23	26	17	20	10	11	13	15	18	33,5	36,5	43,0	21,0	29,5	3,5	8,5	12,0	15,5	24,0		
22	24		18	20	10	12	14	16	18	33,5	40,0		24,0	29,5	3,5	8,5	13,0	19,0	24,0		
23	24		18	21	10	12	15	16		36,5	40,0		24,0	31,5	3,5	8,5	15,5	19,0			
23	24		19		10	12		16		36,5	40,0		27,5		3,5	8,5		19,0			

Tabla 11.7: asignación de rangos a la muestra combinada del ejemplo.

Grupo:	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>T</i>
n^g	12	9	14	8	43
S^g	449,50	230,00	106,50	160,00	946,00
M^g	37,46	25,56	7,61	20,00	22,00

Tabla 11.8: resumen de los rangos para el ejemplo.

$$M^T = \frac{n^T + 1}{2} \quad (11.16)$$

La hipótesis nula, llevada al dominio de los rangos, es que los rangos medios de los distintos grupos son iguales. Bajo este supuesto, el promedio y la varianza de las sumas de los rangos de cada grupo están dados, respectivamente, por las ecuaciones 11.17 y 11.18.

$$S_{H_0}^g = \frac{n^g (n^T + 1)}{2} \quad (11.17)$$

$$\sigma_{H_0}^g = \sqrt{\frac{n^g (n^T + 1)(n^T - n^g)}{12}} \quad (11.18)$$

De manera similar a ANOVA, se requiere determinar la variabilidad que presentan los rangos **entre los grupos**, que denotaremos RSS_{bg} por sus siglas en inglés, que se obtiene con la suma de las desviaciones cuadradas de las medias grupales de los rangos con respecto a la media total de los rangos, como se presenta en la ecuación 11.19.

$$RSS_{bg} = \sum_g n^g (M^g - M^T)^2 \quad (11.19)$$

Con un poco de álgebra:

$$\begin{aligned} RSS_{bg} &= \sum_{g=1}^k n^g \left(\frac{S^g}{n^g} - \frac{n^T + 1}{2} \right)^2 \\ &= \sum_g \left(S^g - \frac{n^g (n^T + 1)}{2} \right)^2 \\ &= \sum_g (S^g - S_{H_0}^g)^2 \end{aligned}$$

Si consideramos que $\frac{S^g - S_{H_0}^g}{\sigma_{H_0}^g} \sim \mathcal{N}(0, 1)$, entonces $\sum_g \frac{(S^g - S_{H_0}^g)^2}{\sigma_{H_0}^g} \sim \chi^2(\nu = k - 1)$, para k grupos.

En base a este análisis, Kruskal y Wallis (1952) definieron el estadístico de prueba H como muestra la ecuación 11.20, que sigue aproximadamente una distribución χ^2 con $k - 1$ grados de libertad cuando **cada grupo tiene a lo menos 5 observaciones**.

$$H = \frac{12}{n^T (n^T + 1)} RSS_{bg} \quad (11.20)$$

Volviendo al ejemplo de las secuencias de operaciones *join*, tendríamos:

$$\begin{aligned} RSS_{bg} &= n^A (M^A - M^T)^2 + n^B (M^B - M^T)^2 + n^C (M^C - M^T)^2 + n^D (M^D - M^T)^2 \\ &= 12 \cdot (37,46 - 22)^2 + 9 \cdot (25,56 - 22)^2 + 14 \cdot (7,61 - 22)^2 + 8 \cdot (20,00 - 22)^2 \\ &= 5.913,211 \end{aligned}$$

y

$$H = \frac{12}{43 \cdot (43 + 1)} \cdot 5.913,211 = 37,505$$

Como todas las muestras tienen más 5 observaciones, el estadístico se distribuye aproximadamente como χ^2 con $\nu = 4 - 1 = 3$ grados de libertad. Así, podemos calcular el valor p para el ejemplo (en R) mediante la llamada `pchisq(37.505, 3, lower.tail = FALSE)`, obteniéndose como resultado $p < 0,001$. Este valor indica que la evidencia es suficientemente fuerte como para rechazar la hipótesis nula en favor de la hipótesis alternativa, incluso para un nivel de significación $\alpha = 0,01$. En consecuencia, podemos concluir con 99 % de confianza que existe al menos un criterio de optimización que lleva a consultas con eficiencias distintas a las generadas por al menos algún otro criterio.

Fijémonos en que, al igual que ANOVA, la prueba de Kruskal-Wallis es de **tipo ómnibus**, por lo que no entrega información en relación a cuáles grupos presentan diferencias. En consecuencia, una vez más es necesario efectuar un análisis post-hoc cuando se detectan diferencias significativas. De manera similar a la estudiada en el capítulo 9, podemos hacer comparaciones entre pares de grupos con la prueba de Wilcoxon-Mann-Whitney (equivalentes a las realizadas con la prueba t de Student para ANOVA de una vía para muestras independientes), usando alguno de los factores de corrección que ya conocimos en el capítulo 8, como los métodos de Holm o de Benjamini y Hochberg (Amat Rodrigo, 2016b).

En R, podemos ejecutar la prueba de Kruskal-Wallis mediante la función `kruskal.test(formula, data)`, donde:

- `formula`: tiene la forma `<variable dependiente> ~ <variable independiente (factor)>`.
- `data`: matriz de datos en formato largo.

La función `pairwise.wilcox.test(x, g, p.adjust.method, paired = FALSE)`, permite realizar los procedimientos post-hoc, cuando corresponda, donde:

- `x`: vector con la variable dependiente.
- `g`: factor o agrupamiento.
- `p.adjust.method`: puede ser “holm” o “BH”, entre otras alternativas.
- `paired`: valor booleano que indica si la prueba es pareada (verdadero) o no. Para la prueba de Kruskal-Wallis debe ser `FALSE`.

El script 11.7 muestra la realización de la prueba de Kruskal-Wallis para el ejemplo e incorpora el procedimiento post-hoc de Benjamini y Hochberg. Los resultados se presentan en la figura 11.14. Observemos que además hemos indicado el argumento `exact = FALSE` en este último, con la intención de evitar la advertencia de que no es posible obtener un valor p exacto en presencia de rangos empatados.

Primero, debemos notar que el estadístico H que reporta la función es difiere ligeramente al obtenido anteriormente ($37,505 \neq 37,714$). Esto se debe a que el método debe introducir ajustes en la ecuación 11.20 por la presencia de empates en los datos analizados. En consecuencia, el valor p que reporta la función también presenta una pequeña diferencia con el obtenido más arriba.

A partir de los resultados del procedimiento post-hoc, considerando un nivel de significación $\alpha = 0,01$, podemos concluir con 99% de confianza que existen diferencias significativas entre la eficiencia conseguida por todos los pares de criterios de optimización, con excepción de los criterios *B* y *D*.

```
Kruskal-Wallis rank sum test

data:  Tiempo by Criterio
Kruskal-Wallis chi-squared = 37.714, df = 3, p-value = 3.249e-08

Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data:  datos[["Tiempo"]] and datos[["Criterio"]]

      A      B      C
B 0.00026 -      -
C 9.3e-05 0.00025 -
D 0.00026 0.02738 0.00026

P value adjustment method: BH
```

Figura 11.14: resultado de la prueba de Kruskal-Wallis y el procedimiento post-hoc de Benjamini y Hochberg para el ejemplo.

Script 11.7: prueba de Kruskal-Wallis y el procedimiento post-hoc de Holm para el ejemplo.

```
1 # Construir la matriz de datos.
2 A <- c(24, 23, 26, 21, 24, 24, 25, 22, 23, 22, 23, 23)
3 B <- c(17, 15, 18, 20, 19, 21, 20, 18, 19)
4 C <- c(10, 11, 14, 11, 15, 12, 12, 10, 9, 13, 12, 12, 10, 10)
5 D <- c(18, 16, 18, 15, 16, 15, 18, 16)
6 Tiempo <- c(A, B, C, D)
7
8 Criterio <- c(rep("A", length(A)), rep("B", length(B)),
9               rep("C", length(C)), rep("D", length(D)))
10 Criterio <- factor(Criterio)
11
12 datos <- data.frame(Tiempo, Criterio)
13
14 # Establecer nivel de significación
15 alfa <- 0.01
16
17 # Hacer la prueba de Kruskal-Wallis.
18 prueba <- kruskal.test(Tiempo ~ Criterio, data = datos)
19 print(prueba)
20
21 # Efectuar procedimiento post-hoc de Benjamini & Hochberg
22 # si se encuentran diferencias significativas.
23 if(prueba[["p.value"]] < alfa) {
24   post_hoc <- pairwise.wilcox.test(datos[["Tiempo"]],
25                                     datos[["Criterio"]],
26                                     p.adjust.method = "BH",
27                                     paired = FALSE,
28                                     exact = FALSE)
29
30   print(post_hoc)
31 }
```

Notemos que `pairwise.wilcox.test()` solo reporta los p valores ajustados. Si queremos conocer el tamaño del efecto de las diferencias detectadas, como se recomienda reportar, debemos realizar las correspondientes pruebas de Wilcoxon-Mann-Whitney para todos los pares de grupos que presenten diferencias significativas.

11.3.2 Prueba de Friedman

Como es natural suponer, podemos considerar la **prueba de Friedman** como una alternativa no paramétrica al procedimiento ANOVA de una vía con muestras correlacionadas descrito en el capítulo 10. Sin embargo, debemos saber que no es exactamente una extensión de esta prueba, puesto que no considera las diferencias relativas entre casos (como lo hace ANOVA y la prueba de rangos con signo de Wilcoxon), y en consecuencia, como señala Baguley (2012), el poder estadístico es bastante menor.

Recordemos las condiciones que se deben verificar para poder aplicar la prueba ANOVA para muestras correlacionadas:

1. La escala con que se mide la variable dependiente tiene las propiedades de una escala de intervalos iguales.
2. Las mediciones son independientes al interior de cada grupo.
3. Se puede suponer razonablemente que la(s) población(es) de origen sigue(n) una distribución normal.
4. La matriz de varianzas-covarianzas es esférica. Como explica Horn (2008, p. 1), esta condición establece que las varianzas entre los diferentes niveles de las medidas repetidas deben ser iguales.

Existen situaciones en las que no podemos comprobar que la escala de medición de la variable dependiente sea de intervalos iguales:

- Cuando las observaciones se miden en una escala logarítmica (por ejemplo, la escala de pH para medir la acidez o la escala de Richter para medir la intensidad de los sismos).
- Cuando las mediciones provienen de una escala ordinal, por ejemplo, un orden de preferencia.
- Cuando las mediciones de base provienen de una escala ordinal. Por ejemplo, cuando se suman o promedian puntajes de diversos elementos evaluados con una escala Likert.

Las condiciones requeridas por la prueba de Friedman son menos exigentes:

1. La variable independiente debe ser categórica y tener a lo menos tres niveles.
2. La escala de la variable dependiente debe ser, a lo menos, ordinal.
3. Las observaciones son una muestra aleatoria e independiente de la población.

Como ejemplo para esta prueba, supongamos ahora que un equipo de desarrolladores desea establecer qué interfaz gráfica (*A*, *B* o *C*) resulta más atractiva para un nuevo sistema, por lo que han seleccionado una muestra aleatoria representativa de los distintos tipos de usuarias y usuarios, y les han solicitado evaluar 6 aspectos de cada interfaz con una escala Likert de 5 puntos, donde el valor 1 corresponde a una valoración muy negativa y 5, a una muy positiva. La valoración que cada participante da a la interfaz evaluada, que llamaremos “índice de usabilidad”, corresponde al promedio simple de las puntuaciones de los 6 aspectos evaluados. El orden en que cada participante evaluó cada una de las interfaces fue asignado aleatoriamente. La tabla 11.9 muestra los índices de usabilidad otorgados por cada participante a cada una de las interfaces.

En consecuencia, las hipótesis a contrastar son:

H_0 : las interfaces obtienen índices de usabilidad similares.

H_A : al menos una interfaz obtiene índices de usabilidad distintos que al menos otra interfaz.

El primer paso del proceso consiste en asignar rangos a las observaciones de cada participante. La interfaz con puntuación más baja recibe un rango de 1 y la más alta, un rango de 3 (generalizando, si se tienen k observaciones pareadas, se asignan rangos con valores 1 a k). En caso de empate, se asigna el promedio de los rangos correspondientes. La tabla 11.10 muestra el resultado de este proceso.

La hipótesis nula para la prueba de Friedman es que los rangos promedio de cada grupo son iguales. Si denotamos el rango promedio de un grupo por M^g , para cada grupo esperamos, entonces, que se cumpla la igualdad de la ecuación 11.21, donde k es la cantidad de grupos.

Participante	Interfaz		
	<i>A</i>	<i>B</i>	<i>C</i>
1	21	6	13
2	10	21	25
3	7	18	18
4	21	7	20
5	24	24	24
6	27	13	8
7	17	13	29

Tabla 11.9: evaluación realizada por las usuarias y los usuarios a cada una de las distintas interfaces.

Participante	Originales			Rangos		
	<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>
1	21	6	13	3	1	2
2	10	21	25	1	2	3
3	7	18	18	1	2,5	2,5
4	21	7	20	3	1	2
5	24	24	24	2	2	2
6	27	13	8	3	2	1
7	17	13	29	2	1	3

Tabla 11.10: rangos de las interfaces por usuaria/o.

$$M^g = \frac{k+1}{2} \quad (11.21)$$

La tabla 11.11 presenta un resumen de estas estadísticas para el ejemplo de las tres interfaces, donde n^g y M^g corresponden, respectivamente, al tamaño de la muestra y al promedio de los rangos de cada interfaz ($g \in \{A, B, C\}$) y en la muestra combinada ($g = T$).

Grupo:	<i>A</i>	<i>B</i>	<i>C</i>	<i>T</i>
n^g	7	7	7	21
M^g	2,14	1,64	2,21	2,00

Tabla 11.11: resumen de los rangos.

Con estos valores, podemos definir una medida para la variabilidad de los grupos agregados, dada por la ecuación 11.22.

$$RSS_{bg} = \sum_g n^g (M^g - M^T)^2 \quad (11.22)$$

Haciendo el cálculo para el ejemplo, tenemos:

$$RSS_{bg} = 7 \cdot (2,14 - 2,00)^2 + 7 \cdot (1,64 - 2,00)^2 + 7 \cdot (2,21 - 2,00)^2 = 1,357$$

Con el resultado anterior, podemos ahora calcular el estadístico de prueba que sigue una distribución χ^2 con $k - 1$ grados de libertad, siendo k el número de mediciones repetidas, como muestra la ecuación 11.23.

$$\chi^2 = \frac{12}{k(k+1)} RSS_{bg} \quad (11.23)$$

Para el ejemplo:

$$\chi^2 = \frac{12}{3 \cdot (3 + 1)} \cdot 1,357 = 1,357$$

Una vez más, calculamos el valor p mediante la llamada `pchisq(1.357, 2, lower.tail = FALSE)`, obteniéndose $p = 0,507$. Considerando un nivel de significación $\alpha = 0,05$, se falla al rechazar la hipótesis nula. En consecuencia, concluimos con 95 % de confianza que no hay evidencia suficiente para descartar que los índices de usabilidad entre las distintas interfaces sean los mismos.

En este caso no es necesario realizar un procedimiento post-hoc, pues la prueba ómnibus no encontró diferencias estadísticamente significativas. No obstante, si fuese necesario, podemos efectuar una prueba de rangos con signo de Wilcoxon por cada par de grupos y aplicar algún factor de corrección.

Para hacer la prueba de Friedman en R, podemos usar la función `friedman.test(formula, data)`, donde:

- `formula`: tiene la forma `<variable dependiente> ~ <variable independiente> | <identificador del caso o bloque>`.
- `data`: matriz de datos en formato largo.

Para los procedimientos post-hoc, podemos aplicar los ajustes ya vistos, como el de Holm, mediante la función `pairwise.wilcox.test()`, del mismo modo descrito para la prueba de Kruskal-Wallis, cuidando en este caso que el argumento `paired` debe tomar forzosamente el valor `TRUE`. Si además queremos conocer el tamaño del efecto detectado para aquellos pares identificados como relevantes, debemos realizar las correspondientes pruebas de rangos con signo de Wilcoxon para todos los pares de grupos que presenten diferencias significativas (Amat Rodrigo, 2016a).

El script 11.8 muestra la realización de la prueba de Friedman para el ejemplo, cuyo resultado se presenta en la figura 11.15, e incorpora el procedimiento post-hoc de Holm por fines académicos, ya que este solamente debería realizarse si la prueba ómnibus encuentra diferencias significativas.

Script 11.8: prueba de Friedman y el procedimiento post-hoc de Holm para el ejemplo.

```

1 # Construir la matriz de datos.
2 A <- c(21, 10, 7, 21, 24, 27, 17)
3 B <- c(6, 21, 18, 7, 24, 13, 13)
4 C <- c(13, 25, 18, 20, 24, 8, 29)
5
6 Puntuacion <- c(A, B, C)
7
8 Interfaz <- c(rep("A", length(A)),
9               rep("B", length(B)),
10              rep("C", length(C)))
11
12 Caso <- rep(1:7, 3)
13
14 Interfaz <- factor(Interfaz)
15
16 datos <- data.frame(Caso, Puntuacion, Interfaz)
17
18 # Establecer nivel de significación
19 alfa <- 0.05
20
21 # Hacer la prueba de Friedman.
22 prueba <- friedman.test(Puntuacion ~ Interfaz | Caso, data = datos)
23 print(prueba)
24
25 # Efectuar procedimiento post-hoc de Holm si se encuentran diferencias
26 # significativas o por fines académicos.
27 fines_académicos <- TRUE
28 if(prueba$p.value < alfa || fines_académicos) {
29   post_hoc <- pairwise.wilcox.test(datos$Puntuacion,

```



```

30         datos$Interfaz,
31         p.adjust.method = "holm",
32         paired = TRUE,
33         exact = FALSE)
34
35     print(post_hoc)
36 }

```

Friedman rank sum test

data: Puntuacion and Interfaz and Caso

Friedman chi-squared = 1.6522, df = 2, p-value = 0.4378

Pairwise comparisons using Wilcoxon signed rank test with continuity correction

data: datos\$Puntuacion and datos\$Interfaz

	A	B
B	0.58	-
C	0.83	0.53

P value adjustment method: holm

Figura 11.15: salida generada por el script 11.8 considerando fines académicos.

Por último, debemos saber que va tomando fuerza la idea de que no se debe usar la prueba de Friedman. En reemplazo, se está recomendando transformar los datos en rangos y luego aplicar directamente el análisis de varianza sobre los datos *rankeados* (Zimmerman & Zumbo, 1993). Es más, esta idea va ganando adeptos incluso para muestras independientes y el análisis de dos muestras.

11.3.3 Ejercicios propuestos (sección 11.3)

- 11.33 Considerando los datos del primer estudio sobre el frenado de automóviles (conjunto `cars` del paquete `datasets`), determina si las distancias requeridas para detenerse ya eran distintas dependiendo si el vehículo viajaba a una velocidad baja, media o alta. En aquella época, década de 1920s, estas categorías estaban definidas, respectivamente, por los siguientes rangos: [8, 13), [13, 20), [21, 25) [millas por hora]. Utiliza una prueba de Kruskal-Wallis para el análisis, y la corrección de Benjamini y Hochberg para comparaciones múltiples. No olvides enunciar las hipótesis y verificar si se cumplen las condiciones de las pruebas que estás aplicando.
- 11.34 El conjunto `airquality` del paquete `datasets` contiene mediciones de la radiación solar (expresada en Langleys) recibida diariamente (con algunos datos perdidos) en la ciudad de New York, EE.UU., desde mayo a septiembre de 1973. Verifica si la radiación recibida en la ciudad es la misma durante los meses de verano (junio a agosto). No olvides enunciar las hipótesis y verificar si se cumplen las condiciones de las pruebas aplicadas. Utiliza la corrección de Benjamini y Hochberg para comparaciones múltiples de ser necesario.
- 11.35 El conjunto `ChickWeight` del paquete `datasets` contiene los resultados de un experimento del efecto de 4 tipos de dietas en el crecimiento temprano de pollitos. Verifica si las dietas producen crecimientos disímiles al sexto día de vida de los pollitos, cuando deben trasladarse a los galpones no calefaccionados. Enuncia las hipótesis y verifica si se cumplen las condiciones de las pruebas aplicadas. Utiliza la corrección de Benjamini y Hochberg para comparaciones múltiples de ser necesario.
- 11.36 Da un ejemplo de una pregunta de investigación sobre las asignaturas comunes en ingeniería que requiera

utilizar una prueba de Kruskal-Wallis. Identifica bien las variables involucradas, justifica por qué no se podría usar una prueba paramétrica, y enuncia las hipótesis a docimar.

- 11.37 Da un ejemplo de una pregunta de investigación sobre los conciertos realizados en Santiago que requiera utilizar una prueba Kruskal-Wallis debido a que la escala de la variable dependiente no permite usar una prueba ANOVA. Identifica bien las variables involucradas y enuncia las hipótesis a docimar.
- 11.38 Da un ejemplo de una pregunta de investigación sobre el estado de la salud mental de estudiantes universitarios que requiera utilizar una prueba de Kruskal-Wallis. Identifica bien las variables involucradas, justifica por qué no se podría usar una prueba paramétrica, y enuncia las hipótesis a docimar.
- 11.39 Investiga qué alternativas existen para estimar el tamaño del efecto para la prueba Kruskal-Wallis. ¿Están implementadas en R?
- 11.40 Investiga qué alternativas conocer el poder o estimar el tamaño de las muestras necesarias para de una prueba de Kruskal-Wallis. ¿Están implementadas en R?
- 11.41 Considera los resultados que tiene la dieta 4 en el crecimiento temprano de pollitos, que se reportan en el conjunto `ChickWeight` del paquete `datasets`. Verifica si el peso de los pollitos difieren entre los días 4, 5, 8 y 10 de vida. Enuncia las hipótesis y verifica si se cumplen las condiciones de las pruebas aplicadas. Utiliza la corrección de Benjamini y Hochberg para comparaciones múltiples de ser necesario.
- 11.42 Una farmacéutica comparó tratamientos para la picazón usando como grupo de estudio 10 hombres de entre 20 y 30 años. Cada sujeto recibió un tratamiento diferente cada día, durante siete días consecutivos, en un orden temporal aleatorio. Para cada tratamiento, se les indujo comezón en los antebrazos mediante un estímulo llamado “cowage”, y los sujetos anotaron la duración del picor, en segundos. Se puede acceder a estos datos usando la función `read_csv()` del paquete `tidyverse` mediante la siguiente llamada: `read_csv("http://statacumen.com/teach/ADA2/notes/ADA2_notes_Ch05_itch.csv")`. Verifica si existen diferencias en la duración de la comezón entre los fármacos Papaverina (`Papv`), Pentobarbitol (`Pento`) y el placebo (`Placebo`). Enuncia las hipótesis y verifica si se cumplen las condiciones de las pruebas aplicadas. Utiliza la corrección de Benjamini y Hochberg para comparaciones múltiples de ser necesario.
- 11.43 Da un ejemplo de una pregunta de investigación sobre las asignaturas comunes en ingeniería que requiera utilizar una prueba de Friedman. Identifica bien las variables involucradas, justifica por qué no se podría usar una prueba paramétrica, y enuncia las hipótesis a docimar.
- 11.44 Da un ejemplo de una pregunta de investigación sobre los conciertos realizados en Santiago que requiera utilizar una prueba Friedman debido a que la escala de la variable dependiente no permite usar una prueba ANOVA. Identifica bien las variables involucradas y enuncia las hipótesis a docimar.
- 11.45 Da un ejemplo de una pregunta de investigación sobre el estado de la salud mental de estudiantes universitarios que requiera utilizar una prueba de Friedman. Identifica bien las variables involucradas, justifica por qué no se podría usar una prueba paramétrica, y enuncia las hipótesis a docimar.
- 11.46 Investiga qué alternativas existen para estimar el tamaño del efecto para la prueba Friedman. ¿Están implementadas en R?
- 11.47 Investiga qué alternativas conocer el poder o estimar el tamaño de las muestras necesarias para de una prueba de Friedman. ¿Están implementadas en R?

11.4 BIBLIOGRAFÍA DEL CAPÍTULO

- Amat Rodrigo, J. (2016a). *Test de Friedman*. Consultado el 29 de mayo de 2021, desde https://www.cienciadedatos.net/documentos/21_friedman_test
- Amat Rodrigo, J. (2016b). *Test Kruskal-Wallis*. Consultado el 29 de mayo de 2021, desde https://www.cienciadedatos.net/documentos/20_kruskal-wallis_test
- Baguley, T. (2012). *Beware the Friedman test!* Consultado el 13 de diciembre de 2021, desde <https://seriousstats.wordpress.com/2012/02/14/friedman/>

- Carchedi, N., De Mesmaeker, D., & Vannoorenberghe, L. (s.f.). RDocumentation.
Consultado el 2 de abril de 2021, desde <https://www.rdocumentation.org/>
- Fay, M. P., & Proschan, M. A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys*, 4, 1.
- Glen, S. (2021a). *Geometric Mean Definition and Formula*.
Consultado el 27 de mayo de 2021, desde <https://www.statisticshowto.com/geometric-mean-2/>
- Glen, S. (2021b). *Kruskal Wallis H Test: Definition, Examples & Assumptions*.
Consultado el 5 de junio de 2021, desde <https://www.statisticshowto.com/kruskal-wallis/>
- Horn, R. A. (2008). *Sphericity in repeated measures analysis*. Consultado el 11 de mayo de 2021, desde <http://oak.ucc.nau.edu/rh232/courses/EPS625/Handouts/RM-ANOVA/Sphericity.pdf>
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), 583-621.
- Lærd Statistics. (2020). *Friedman Test in SPSS Statistics* [Lund Research Ltd.]. Consultado el 5 de junio de 2021, desde <https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php>
- Lane, D. (s.f.). *Online Statistics Education: A Multimedia Course of Study*.
Consultado el 4 de mayo de 2021, desde <https://onlinestatbook.com/>
- Lowry, R. (1999). *Concepts & Applications of Inferential Statistics*.
Consultado el 3 de mayo de 2021, desde <http://vassarstats.net/textbook/>
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. John wiley & sons.
- United States Census Bureau. (2004). *CT1970p2-13: Colonial and Pre-Federal Statistics*. Consultado el 26 de mayo de 2021, desde <https://www2.census.gov/prod2/statcomp/documents/CT1970p2-13.pdf>
- United States Census Bureau. (2021). Decennial Census of Population and Housing. Consultado el 26 de mayo de 2021, desde <https://www.census.gov/programs-surveys/decennial-census/decade.html>
- Zimmerman, D. W., & Zumbo, B. D. (1993). Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. *The Journal of Experimental Education*, 62(1), 75-86.