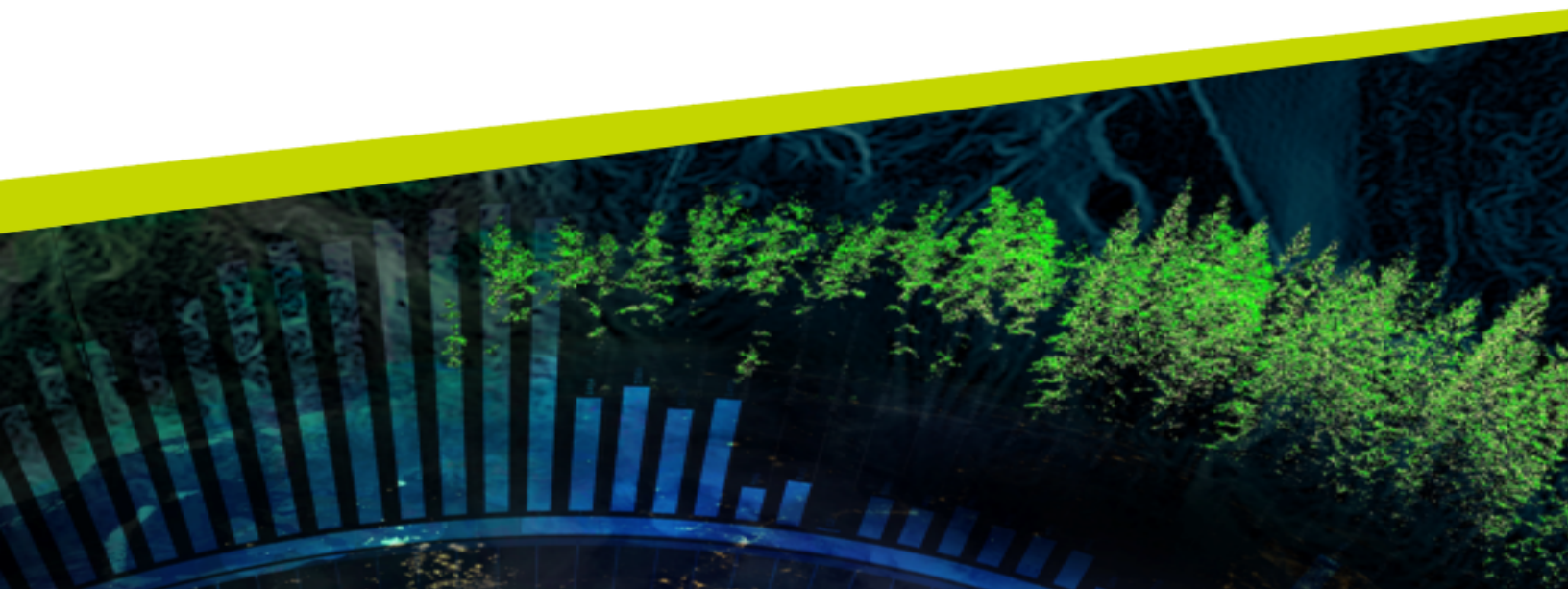




INFERENCIA Y MODELOS ESTADÍSTICOS

Jacqueline Köhler C. y José Luis Jara V.



CAPÍTULO 10. ANOVA DE UNA VÍA PARA MUESTRAS CORRELACIONADAS

En el capítulo 9 conocimos el procedimiento ANOVA de una vía para muestras independientes, que podemos entender como una extensión de la prueba *t* de Student para muestras independientes. De manera similar, ahora abordaremos el **procedimiento ANOVA de una vía para muestras correlacionadas** (también llamado **ANOVA para medidas repetidas** o **ANOVA intracasos**¹) que puede asociarse a la prueba *t* con muestras apareadas, pero ahora considerando un factor con tres o más niveles.

El procedimiento ómnibus y los procedimientos post-hoc que estudiamos en este capítulo se basan en las ideas presentadas por Lowry (1999, cap. 15).

10.1 HIPÓTESIS

En el procedimiento ANOVA para muestras correlacionadas podemos distinguir, como con la prueba *Q* de Cochran vista en el capítulo 8, entre dos escenarios:

- **Diseño con medidas repetidas:** a cada caso se le toman medidas en diferentes momentos o condiciones. Por ejemplo, medir la estatura de estudiantes de un colegio al comenzar 7º básico, 8º básico y 1º medio. Otro ejemplo sería registrar los tiempos de ejecución para una misma instancia de un problema con k algoritmos diferentes.
- **Diseño con bloques aleatorios:** cada bloque contiene diferentes casos agrupados según determinadas características. Por ejemplo, considerando que puede haber una alta migración de estudiantes durante la investigación, podríamos separar el grupo en 8 bloques: niños muy activos, niños con actividad moderada, niños sedentarios, niños muy sedentarios, y análogamente para las niñas. De esta forma podríamos asegurar medidas en 7º básico, 8º básico y 1º medio, aún cuando no siempre se trate de las mismas personas. Similarmente, podemos comparar los tiempos de ejecución de k algoritmos si tenemos registros para instancias de prueba distintas pero con dificultad similar (como grafos que tengan el mismo número de vértices y aristas).

El método es el mismo en ambos casos e intenta controlar estadísticamente la variación introducida por factores distintos al que se desea estudiar, usando para ello varias mediciones de un caso o un bloque. Si bien el diseño con bloques aleatorios es común, especialmente en medicina, este apunte usa las medidas repetidas en su discusión, ya que son más comunes en el área de la ingeniería y la informática.

Como es habitual, usemos un ejemplo para ver cómo se lleva a cabo el procedimiento ANOVA de una vía para muestras correlacionadas. Supongamos que una estudiante de un curso de programación debe comparar la eficiencia de cuatro algoritmos de ordenamiento: Quicksort, Bubblesort, Radixsort y Mergesort. Para ello, ha generado aleatoriamente 6 arreglos de tamaño y dificultad diversa, registrando para cada uno de ellos el tiempo de ejecución utilizado por cada algoritmo (en milisegundos), como muestra la tabla 10.1².

En este caso, la lógica es muy similar a la que ya conocimos para ANOVA con muestras independientes. Sin embargo, existe una diferencia importante al trabajar con muestras correlacionadas: no toda la variabilidad intragrupos es pura e inevitable, sino que una parte de ella se debe a **diferencias individuales** preexistentes

¹El uso de la palabra “casos” en este contexto es nuestro intento por utilizar un término sin sesgo de género para reemplazar la palabra “sujetos”, que es muy usada en los textos de estadística inferencial. Debemos recordar que los conceptos que estamos estudiando fueron desarrollados por angloparlantes, quienes usan el sustantivo *subjects*, que en inglés tiene el significado “*a person or thing that is being discussed, described, or dealt with*” («Subject», 2023). Es decir, se refiere, sin importar su sexo o género, a “personas”, “cosas” o “animales”, aunque esto último no lo dice explícitamente la definición anterior. Así, cuando los textos usan el término “sujetos”, no se hace referencia a un grupo de hombres, sino que a los “ejemplares” o “espécimenen” (dos palabras que comienza a ser usada en publicaciones científicas) que están *sujetos* a estudio. Incluso, podrían ser algoritmos, vehículos, máquinas, etc., y no seres humanos.

²Los valores aquí expuestos son ficticios.

Instancia	Bubblesort	Mergesort	Quicksort	Radixsort
1	31,6	25,0	23,2	30,1
2	29,3	25,7	22,6	28,4
3	30,7	25,7	23,4	28,7
4	30,8	23,7	23,3	28,3
5	29,8	25,5	21,8	29,9
6	30,3	24,7	23,9	29,1

Tabla 10.1: tiempos de ejecución para las diferentes instancias con cada algoritmo del ejemplo.

entre los casos (por ejemplo, un arreglo puede estar ordenado desde el inicio, mientras otro podría estar en orden inverso).

Recordando las hipótesis de la prueba t de Student con muestras apareada, podemos extenderlas para nuestro ejemplo:

H_0 : en promedio, no hay diferencias en los tiempos requeridos por los algoritmos estudiados para ordenar los mismos arreglos numéricos.

H_A : la media de las diferencias en los tiempos de ejecución necesarios para ordenar los mismos arreglos numéricos es diferente para al menos un par de los algoritmos estudiados.

Matemáticamente:

Sean $t_B^{(i)}$, $t_M^{(i)}$, $t_Q^{(i)}$ y $t_R^{(i)}$ los tiempos de ejecución requeridos por los algoritmos Bubblesort, Mergesort, Quicksort y Radixsort respectivamente, para ordenar el i -ésimo arreglo numérico en el espacio de todos los posibles arreglos numéricos. Denotaremos $d_{X-Y}^{(i)}$ a la diferencia en los tiempos de ejecución necesarios para ordenar el i -ésimo arreglo numérico por los algoritmos X e Y (es decir, $d_{X-Y}^{(i)} = t_X^{(i)} - t_Y^{(i)}$) con $X, Y \in \{B, M, Q, R\}$. Entonces: Denotando como $\mu_{d_{X-Y}}$ a la media de las diferencias en los tiempos de ejecución necesarios por los algoritmos X e Y en ordenar todos los posibles arreglos numéricos:

H_0 : $\mu_{d_{B-M}} = \mu_{d_{B-Q}} = \mu_{d_{B-R}} = \mu_{d_{M-Q}} = \mu_{d_{M-R}} = \mu_{d_{Q-R}} = 0$.

H_A : $\exists X, Y \in \{B, M, Q, R\}, X \neq Y \mid \mu_{d_{X-Y}} \neq 0$.

Vemos que, como la versión para muestras independientes, la hipótesis nula de la prueba ANOVA con muestras correlacionadas es de **tipo ómnibus**.

10.2 CONDICIONES

Al igual que otras pruebas que hemos conocido en capítulos anteriores, este procedimiento requiere que se cumplan algunas condiciones:

1. La escala con que se mide la variable dependiente tiene las propiedades de una escala de intervalos iguales.
2. Las observaciones son independientes al interior de cada muestra.
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.

Veamos si nuestro ejemplo cumple con las condiciones. La primera se verifica, puesto que el tiempo, como toda magnitud física, tiene una escala de intervalos iguales (de hecho tiene escala de razón, que es más exigente). A su vez, el enunciado señala que la estudiante generó las instancias de prueba de forma aleatoria, y los 6 casos evaluados son infinitamente menor al 10 % del espacio de todos los arreglos numéricos. Por otro lado, cuánto se tarde un algoritmo en ordenar un determinado arreglo no influye en cuánto se tarda en otras instancias. Luego, la condición de independencia al interior de cada muestra está garantizada. La figura 10.1 (creada mediante el script 10.1) muestra gráficos Q-Q para cada medición, donde se puede apreciar que no se observan valores que pudieran ser considerados atípicos y se puede suponer razonablemente que las distribuciones se asemejan a la normal.

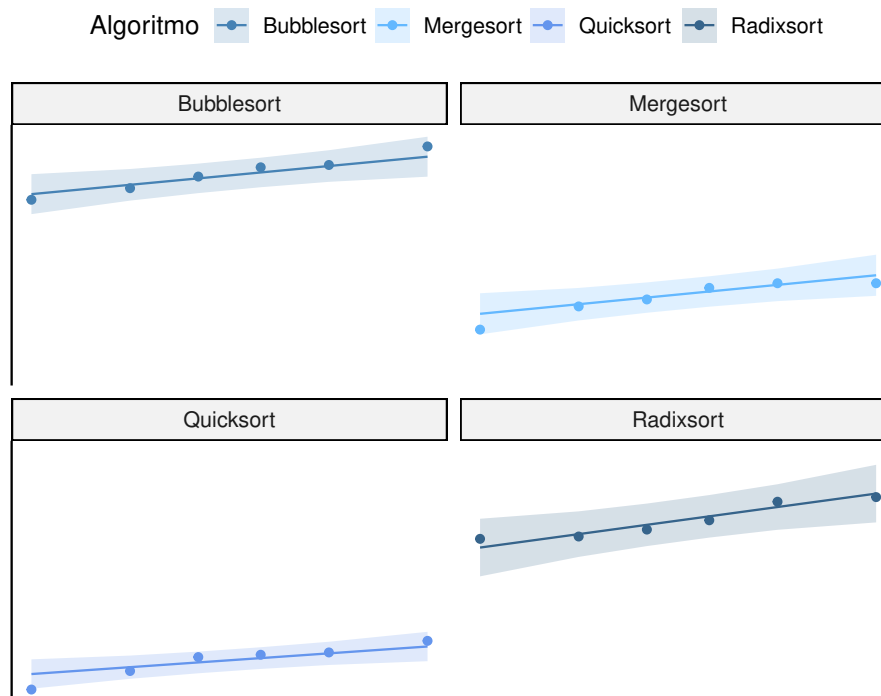


Figura 10.1: gráfico para comprobar el supuesto de normalidad en las muestras del ejemplo.

Script 10.1: verificación gráfica de la condición de normalidad para un procedimiento ANOVA de una vía para muestras correlacionadas.

```

1 library(emmeans)
2 library(ez)
3 library(ggpubr)
4 library(nlme)
5 library(tidyr)
6
7 # Crear la matriz de datos
8 Instancia <- factor(1:6)
9 Bubblesort <- c(31.6, 29.3, 30.7, 30.8, 29.8, 30.3)
10 Mergesort <- c(25.0, 25.7, 25.7, 23.7, 25.5, 24.7)
11 Quicksort <- c(23.2, 22.6, 23.4, 23.3, 21.8, 23.9)
12 Radixsort <- c(30.1, 28.4, 28.7, 28.3, 29.9, 29.1)
13 datos_anchos <- data.frame(Instancia, Bubblesort, Mergesort, Quicksort, Radixsort)
14
15 # Llevar los datos a formato largo
16 datos_largos <- datos_anchos |> pivot_longer(-Instancia,
17                                           names_to = "Algoritmo",
18                                           values_to = "Tiempo") |>
19   mutate(Algoritmo = factor(Algoritmo))
20
21 # Verificar normalidad por medio de gráficos Q-Q
22 tonos_azules = c("steelblue", "steelblue1", "cornflowerblue", "steelblue4")
23 g <- ggqqplot(datos_largos, x = "Tiempo", y = "Algoritmo",
24              color = "Algoritmo", palette = tonos_azules)
25 g <- g + facet_wrap(~ Algoritmo)
26 g <- g + rremove("x.ticks") + rremove("x.text")
27 g <- g + rremove("y.ticks") + rremove("y.text")
28 g <- g + rremove("axis.title")
29 print(g)

```

La prueba de esfericidad es más compleja, por lo que no se estudia en detalle en este texto. En palabras simples, se podría describir como que se requiere que las varianzas (o desviaciones estándar) de las **diferencias entre todos los pares** de niveles dentro del factor de medidas repetidas deben ser iguales. Para el ejemplo, se debería cumplir:

$$\sigma_{d_{B-M}} = \sigma_{d_{B-Q}} = \sigma_{d_{B-R}} = \sigma_{d_{M-Q}} = \sigma_{d_{M-R}} = \sigma_{d_{Q-R}}$$

Las diferencias en los tiempos de ejecución requeridos para cada una de las instancias entre cada par de algoritmos se muestran en la tabla 10.2, junto a sus medias y varianzas muestrales. Podemos ver que las varianzas son “pequeñas” si se considera que los tiempos promedio están en el rango de 23 a 30 [ms] aproximadamente, por lo que podríamos asumir que son iguales; aunque también podríamos considerar lo contrario, puesto que el mayor valor observado de las varianzas muestrales (1,795) es más de 3 veces el menor valor (0,574). Afortunadamente, la función de R `ezANOVA()` incluye una prueba auxiliar para verificar esta condición: la **prueba de esfericidad de Mauchly**, e incluso proporciona un método para controlar posibles violaciones, como veremos más adelante en este capítulo.

Instancia	B - M	B - Q	B - R	M - Q	M - R	Q - R
1	6,6	8,4	1,5	1,8	-5,1	-6,9
2	3,6	6,7	0,9	3,1	-2,7	-5,8
3	5,0	7,3	2,0	2,3	-3,0	-5,3
4	7,1	7,5	2,5	0,4	-4,6	-5,0
5	4,3	8,0	-0,1	3,7	-4,4	-8,1
6	5,6	6,4	1,2	0,8	-4,4	-5,2
Promedio	5,367	7,383	1,333	2,017	-4,033	-6,050
Varianza	1,795	0,574	0,819	1,646	0,915	1,475

B: Bubblesort, M: Mergesort, Q: Quicksort, R: Radixsort.

Tabla 10.2: diferencias en los tiempos de ejecución necesitado por los algoritmos para cada instancia, sus respectivos promedios y varianzas observadas.

10.3 PROCEDIMIENTO

Al igual que para el caso de muestras independientes, el procedimiento ANOVA para muestras correlacionadas opera en base a la variabilidad, calculada en base a la suma de los cuadrados de las desviaciones. Recordemos la forma general de este cálculo:

$$SS = \sum_{i=1}^n (x_i - \bar{x})^2$$

Los primeros pasos son análogos a los que ya estudiamos para ANOVA de una vía con muestras independientes, y consisten en calcular la variabilidad total (con la muestra combinada), la variabilidad entre mediciones (*between groups*) y la variabilidad intramediciones (*within groups*), denotadas por SS_T , SS_{bg} y SS_{wg} , respectivamente. Para el ejemplo:

$$\begin{aligned} SS_T &\approx 224,930 \\ SS_{bg} &\approx 213,045 \\ SS_{wg} &\approx 11,885 \end{aligned}$$

Como mencionamos en la introducción de este capítulo, al trabajar con muestras correlacionadas es posible **separar la variabilidad debida a diferencias preexistentes entre los casos** seleccionados en las muestras (SS_{casos})³, expresada por las variaciones observadas en las mediciones repetidas de un mismo caso,

³Tradicionalmente, los textos se refieren a esta variabilidad como “intrasujetos”.

pues estas son ajenas al factor en estudio, y solo nos interesa conservar la variabilidad pura (SS_{error}). Así, en ANOVA para muestras correlacionadas aparece una nueva identidad, dada por la ecuación 10.1:

$$SS_{wg} = SS_{casos} + SS_{error} \quad (10.1)$$

De manera similar a la que ya empleamos en cálculos previos, la **variabilidad individual** o **intracasos** está dada por la ecuación 10.2:

$$SS_{casos} = k \sum_{i=1}^n (\bar{x}_i - \bar{x}_T)^2 \quad (10.2)$$

donde:

- k corresponde a la cantidad de observaciones (medidas) por cada caso.
- \bar{x}_i es la media de las observaciones del i -ésimo caso.
- \bar{x}_T es la media combinada de las mediciones.

La tabla 10.3 muestra una vez más las observaciones del ejemplo, incluyendo el tiempo promedio de ejecución para cada instancia.

Instancia	Promedio	Desviación	Desviación ²
1	27,475	0,579	0,335
2	26,500	-0,396	0,157
3	27,125	0,229	0,052
4	26,525	-0,371	0,138
5	26,750	-0,146	0,021
6	27,000	0,104	0,011
			$\Sigma = 0,714$
			$SS_{casos} = 9,029$

Tabla 10.3: tiempos de ejecución medios y desviaciones observadas para las diferentes instancias del ejemplo, usadas para calcular la variabilidad atribuible a los casos.

10.4 EL ESTADÍSTICO DE PRUEBA F

Al igual que en el capítulo 9, calculamos ahora los grados de libertad ya conocidos (recordemos que ahora k corresponde a la cantidad de mediciones por caso):

$$\nu_T = n_T - 1 \quad (10.3)$$

$$\nu_{bg} = k - 1 \quad (10.4)$$

$$\nu_{wg} = n_T - k \quad (10.5)$$

Para el ejemplo tenemos $\nu_T = 24 - 1 = 23$, $\nu_{bg} = 4 - 1 = 3$ y $\nu_{wg} = 24 - 4 = 20$.

Puesto que anteriormente descompusimos la variabilidad intramediciones en variabilidad individual y variabilidad del error, necesitamos también identificar los grados de libertad correspondientes a cada componente, utilizando las ecuaciones 10.6 y 10.7:

$$\nu_{casos} = n_{casos} - 1 \quad (10.6)$$

$$\nu_{error} = \nu_{wg} - \nu_{casos} \quad (10.7)$$

Para el ejemplo: $\nu_{casos} = 6 - 1 = 5$ y $\nu_{error} = 20 - 5 = 15$.

Los cuadrados medios del procedimiento ANOVA de una vía para muestras correlacionadas son, respectivamente, el cuadrado medio del efecto, que corresponde al cuadrado medio entre mediciones, y el cuadrado medio del error, dada por la ecuación 10.8:

$$MS_{error} = \frac{SS_{error}}{\nu_{error}} \quad (10.8)$$

Así, tenemos que para el ejemplo:

$$MS_{efecto} = MS_{bg} = \frac{213,065}{3} \approx 71,022$$

$$MS_{error} = \frac{9,029}{15} \approx 0,602$$

y en consecuencia:

$$\frac{MS_{efecto}}{MS_{error}} = F = \frac{71,022}{0,602} \approx 117,989$$

En R, al hacer la llamada `pf(117.989, 3, 15, lower.tail = FALSE)` para calcular el valor p, obtenemos $p < 0,001$ (en rigor matemático, $p \approx 1,178 \cdot 10^{-10}$).

10.5 RESULTADO

Una vez más, el resultado del procedimiento se representa en la forma tabular, como muestra la tabla 10.4.

Fuente	ν	SS	MS	F	p
Efecto	3	213,045	71,022	117,989	$1,178 \cdot 10^{-10}$
Error	15	9,029	0,602		
TOTAL	23	224,930			

Tabla 10.4: resultado del procedimiento ANOVA.

El valor p obtenido es muy menor a cualquier nivel de significación típico que podamos considerar, por lo que rechazamos la hipótesis nula en favor de la hipótesis alternativa. Así, la estudiante del ejemplo concluye con más de 99 % de confianza existen diferencias significativas en los tiempos de ejecución requeridos entre al menos dos de los algoritmos de ordenamiento comparados al resolver los mismos arreglos.

10.6 PROCEDIMIENTO ÓMNIBUS EN R

Para efectuar el procedimiento ANOVA de una vía para muestras correlacionadas en R, podemos usar las mismas funciones ya estudiadas en el capítulo 9: `aov()` y `ezANOVA()`, como ilustra el script 10.2.

En el caso de `aov()`, podemos apreciar que la fórmula entregada en la llamada (líneas 35–36 del script) es bastante diferente a la del capítulo 9. Esto se debe a que esta función realiza por defecto un procedimiento ANOVA para muestras independientes, por lo que se debe explicitar en la fórmula la expresión `Error(Instancia/Algoritmo)` que le indica que la variable independiente `Algoritmo` está **anidada** dentro de los niveles de la variable `Instancia`, es decir que al calcular la variabilidad del error debe considerar que los tiempos requeridos por los algoritmos varían (o podrían variar) para de cada instancia, por lo que se requiere descartar esta variabilidad del error aleatorio. La figura 10.2 muestra el resultado obtenido, que es el mismo presentado en la tabla 10.4 salvo por ligeras diferencias de redondeo.

La llamada a `ezANOVA()`, en cambio, es muy similar a la ya conocida, como se puede apreciar en las líneas 43–44 del script 10.2: la única diferencia con respecto a la llamada realizada en el capítulo 9 es que ahora el argumento `between` ha sido **reemplazado** por `within` para la variable independiente. Esta diferencia indica a `ezANOVA()` que se trata de un procedimiento ANOVA con muestras correlacionadas. El resultado de esta llamada también es mostrado en la figura 10.2, debajo de `$ANOVA`, donde podemos apreciar que, una vez más, se obtiene el mismo resultado, con algunas diferencias por el redondeo, de la tabla 10.4.

Habíamos mencionado que otra ventaja de `ezANOVA()` es que verifica la condición de esfericidad mediante la prueba de esfericidad de Mauchly. En la figura 10.2, puede verse que el resultado de esta prueba se muestra a continuación del efecto encontrado, bajo el título `$‘Mauchly’s Test for Sphericity’`.

```

Procedimiento ómnibus
=====

Procedimiento ANOVA para muestras repetidas usando aov
-----

Error: Instancia
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  5  2.857  0.5714

Error: Instancia:Algoritmo
      Df Sum Sq Mean Sq F value    Pr(>F)
Algoritmo  3 213.04   71.01    118 1.18e-10 ***
Residuals 15   9.03    0.60
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Procedimiento ANOVA para muestras repetidas usando ezANOVA
-----
$ANOVA
      Effect DFn DFd      F      p p<.05      ges
2 Algoritmo   3  15 117.9921 1.177725e-10 * 0.9471612

$'Mauchly's Test for Sphericity'
      Effect      W      p p<.05
2 Algoritmo 0.3367911 0.5545469

$'Sphericity Corrections'
      Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF] p[HF]<.05
2 Algoritmo 0.6803135 8.377723e-08 * 1.154155 1.177725e-10 *

```

Figura 10.2: resultado ómnibus del procedimiento ANOVA para muestras repetidas entregadas por las funciones `aov()` y `ezANOVA()`.

Script 10.2: (continuación del script 10.1) procedimiento ANOVA de una vía para muestras correlacionadas.

```

31 cat("Procedimiento ómnibus\n")
32 cat("=====\n\n")
33
34 # Realizar y mostrar el procedimiento ANOVA para muestras repetidas con aov()
35 prueba <- aov(Tiempo ~ Algoritmo + Error(Instancia/Algoritmo),
36              data = datos_largos)
37
38 cat("Procedimiento ANOVA para muestras repetidas usando aov\n")
39 cat("-----\n")
40 print(summary(prueba))
41
42 # Realizar y mostrar el procedimiento ANOVA para muestras repetidas con ezANOVA()
43 prueba2 <- ezANOVA(data = datos_largos, dv = Tiempo, wid = Instancia,
44                   within = Algoritmo)
45
46 cat("\nProcedimiento ANOVA para muestras repetidas usando ezANOVA\n")
47 cat("-----\n")
48 print(prueba2)

```

Podemos apreciar que la prueba no resulta significativa para el ejemplo ($W = 0,337$; $p = 0,555$), por lo que no podemos sospechar que los datos del ejemplo no cumplen con la condición de esfericidad (hipótesis nula de la prueba de Mauchly).

Ahora bien, la prueba ANOVA de una vía para muestras correlacionadas no es muy robusta a violaciones de la condición de esfericidad. Por fortuna, cuando existen problemas con esta condición, no todo está perdido

puesto que existen dos correcciones que podemos emplear para corregir las estimaciones del procedimiento que suponen la esfericidad: la de **Greenhouse-Geisser** y la de **Huynh-Feldt**. Ambas estiman la esfericidad, denotada por ϵ , y corrigen el valor p de la prueba ANOVA en base a esta estimación ajustando los grados de libertad de la distribución F usada en el cálculo. La corrección de Greenhouse-Geisser es más conservadora y tiende a subestimar ϵ cuando este es cercano a 1, por lo que se recomienda su uso para $\epsilon < 0,75$. Para $\epsilon \geq 0,75$ estimado con el método Greenhouse-Geisser (Karadimitriou & Marshall, 2016), suele emplearse la estimación de Huynh-Feldt que es algo más liberal (Lærd Statistics, 2020). `ezANOVA()` lleva a cabo ambas correcciones y reporta para cada una de ellas tanto la estimación de la esfericidad como el valor p corregido, como se aprecia en la figura 10.2 bajo el título `$‘Sphericity Corrections’`, donde:

- `GGe`: estimación de ϵ con el método de Greenhouse-Geisser.
- `p[GG]`: valor p tras la corrección de Greenhouse-Geisser.
- `HFe`: estimación de ϵ con el método de Huynh-Feldt.
- `p[HF]`: valor p tras la corrección de Huynh-Feldt.

Cuando los datos no cumplen con la condición de esfericidad, se debe reportar el valor estimado de ϵ , `GGe` o `HFe` dependiendo si el primero es o no menor a 0,75, y considerar para todo efecto `p[GG]` o `p[HF]`, respectivamente, como el valor p de la prueba ANOVA.

En nuestro ejemplo, se da una situación **muy particular**: las estimaciones `GGe` $\approx 0,68$ y `HFe` $\approx 1,15 = 1,00$ son muy contradictorias.

En general, ϵ varía en el intervalo $\left[\frac{1}{k-1}, 1\right]$ donde k es el número de mediciones para cada caso. En general, valores mayores a 0,9 indican leves desviaciones de esfericidad, entre 0,7 y 0,9 sugieren una desviación mediana, y valores bajo 0,7 son indicativos de desviaciones importantes.

De esta forma, para el ejemplo, `GGe` está indicando una desviación importante mientras que `HFe` sugiere que la condición de esfericidad se cumple completamente⁴.

¿Qué hacemos entonces? Una consideración importante es que, a pesar de su popularidad en alternativas de software estadísticos, la prueba de Mauchly ha sido fuertemente criticada porque con frecuencia falla en detectar desviaciones de la esfericidad en muestras pequeñas y sobre estimarlas en muestras grandes (Lærd Statistics, 2020). La recomendación es siempre tomar en cuenta las estimaciones de ϵ y analizar si la prueba es consistente con ellas. Si no hay consistencia, lo aconsejable es irse con la conclusión más prudente (conservadora).

Esto es lo que podría estar pasando en nuestro ejemplo, después de todo solo se eligieron seis casos. Luego, deberíamos seguir la opción más conservadora y reportar: existe una desviación importante de la condición de esfericidad $\epsilon = 0,68$, estimada con el método de Greenhouse-Geisser, por lo que la prueba resulta significativa con un estadístico de prueba muy alto ($F(3, 15) = 117,992$) y valor p corregido muy bajo ($p \approx 8,378 \cdot 10^{-8} < 0,001$).

10.7 TAMAÑO DEL EFECTO

Tal y como hicimos al trabajar con muestras independientes, podemos obtener **una aproximación gráfica** del tamaño del efecto observado en los datos utilizando la función `ezPlot(data, dv, wid, within, x)`, notando que ahora debemos especificar la variable independiente por medio del argumento `within` en vez de `between`. Para el ejemplo, esta llamada aparece en las líneas 51–52 del script 10.3, obteniéndose como resultado el gráfico de la figura 10.3.

Recordemos que la función `ezANOVA()` nos reporta el tamaño del efecto observado usando la medida estandarizada eta cuadrado generalizada, que se encuentra bajo el título `ges` en la figura 10.2. Vemos que para el ejemplo $\eta_c^2 \approx 0,95$, que es ¡un efecto gigante! si consideramos los umbrales 0,01, 0,06 y 0,14 recomendados por Cohen (1992) para efectos pequeños, medianos y grandes, respectivamente, aunque estos fueron hechos para η^2 y no la versión generalizada. Sin embargo, como vimos para muestras independientes, en procedimientos ANOVA de una vía $\eta_c^2 = \eta^2$, por lo que los umbrales son válidos en este caso. De todas formas, los valores

⁴Notemos que si bien un valor mayor que 1 es posible en el cálculo que emplea el método de Huynh-Feldt, en la práctica, cuando esto ocurre, se trunca a este valor máximo.

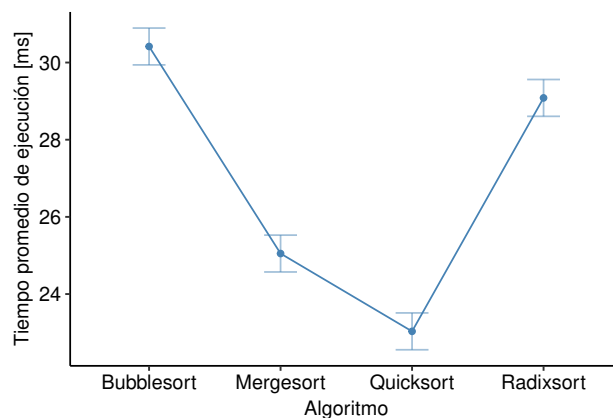


Figura 10.3: Tamaño del efecto observado en el ejemplo.

de estas medidas en general no suelen ser muy diferentes, a menos que se estén usando varios factores para definir bloques o incluyendo variables sin manipular (covariables) en el diseño de la prueba.

Script 10.3: (continuación del script 10.2) procedimiento ANOVA de una vía para muestras correlacionadas.

```

50 # Obtener el gráfico del tamaño del efecto
51 g2 <- ezPlot(data = datos_largos, dv = Tiempo, wid = Instancia, within = Algoritmo,
52             y_lab = "Tiempo promedio de ejecución [ms]", x = Algoritmo)
53 # Cambiar sus colores
54 g2 <- g2 + theme_pubr()
55 g2[["layers"]][[1]][["aes_params"]][["colour"]] <- "steelblue"
56 g2[["layers"]][[2]][["aes_params"]][["colour"]] <- "steelblue"
57 g2[["layers"]][[3]][["aes_params"]][["colour"]] <- "steelblue"
58 # Y mostrar el gráfico
59 print(g2)

```

10.8 POTENCIA ESTADÍSTICA

A diferencia de lo estudiado hasta ahora, la función `pwr.anova.test()` del paquete R `pwr` que usamos con el procedimiento para muestras independientes **no sirve** al trabajar con muestras correlacionadas.

Sin embargo, podemos encontrar una alternativa válida en el paquete `WebPower` que recoge un conjunto de herramientas básicas y avanzadas para realizar análisis de poder estadístico (Zhang & Yuan, 2018) que se encuentran disponibles en línea de manera gratuita en <https://webpower.psychstat.org>.

Esta alternativa corresponde a la función `wp.rmanova(n, ng = 1, nm, f, nscor, alpha, power, type = 1)`, donde:

- `n`: número de casos.
- `ng`: número de grupos, que debe tener valor 1 para ANOVA de medidas repetidas.
- `nm`: número de medidas repetidas para cada caso.
- `f`: tamaño del efecto (f de Cohen).
- `nscor`: estimación de la esfericidad en las poblaciones. Por omisión toma valor 1 que asume el cumplimiento de la condición esfericidad y que no se requieren correcciones, como se considerarían con valores menores.
- `alpha`: nivel de significación de la prueba. Por omisión toma valor 0.05.
- `power`: poder de la prueba.
- `type`: tipo del efecto de interés. Se usan los valores 0 para indicar el efecto entre grupos, 1 para el efecto entre medidas repetidas y 2 para la interacción entre grupos y mediciones. Es decir, siempre

debemos usar el valor 1 para las pruebas ANOVA de una vía para muestras correlacionadas que estamos estudiando.

La función `wp.rmanova()` sigue la idea de las funciones del paquete `pwr` en el sentido que permite obtener estimaciones de uno de los factores que influyen en la calidad de la prueba (nivel de significación, potencia estadística, tamaño del efecto, y tamaño de la muestra) otorgándole valor `NULL` al argumento correspondiente.

En nuestro ejemplo tenemos seis instancias de prueba (`n = 6`), cuatro algoritmos de ordenamiento (`nm = 4`) y estamos usando 99 % confianza (`alpha = 0.01`).

La estimación de la esfericidad, como vimos en la sección 10.6, puede obtenerse de las correcciones calculadas por la función `ezANOVA()`. Siguiendo la discusión hecha allí, optamos la alternativa más prudente: `nscor = 0.68`.

Como vimos, la medida estandarizada del efecto f de Cohen está relacionada a la medida η^2 de acuerdo a la ecuación 9.14. Además, considerando que la función `ezANOVA()` nos reporta $\eta_c^2 \approx 0,947$ como medida del tamaño del efecto observado en los datos del ejemplo, y que, como se dijo en la sección 10.7, para la prueba ANOVA de una vía para muestras correlacionadas se cumple que $\eta_c^2 = \eta^2$, podemos calcular la f de Cohen:

$$f = \sqrt{\frac{\eta_c^2}{1 - \eta_c^2}} = \sqrt{\frac{0,947}{1 - 0,947}} \approx 4,227$$

Entonces, el siguiente código permite obtener una estimación de la potencia de la prueba realizada para el ejemplo:

```
> wp.rmanova(n = 6, ng = 1, nm = 4, f = 4.227, nscor = 0.68, alpha = 0.01, type = 1)
Repeated-measures ANOVA analysis
```

n	f	ng	nm	nscor	alpha	power
6	4.227	1	4	0.68	0.01	0.9998076

NOTE: Power analysis for within-effect test

URL: <http://psychstat.org/rmanova>

>

donde apreciamos que el poder estadístico fue aproximadamente 1, lo que no es extraño para el enorme efecto encontrado en los datos. Debemos recordar que se considera que los valores $f = 0,10$, $f = 0,25$ y $f = 0,40$ son característicos de un efecto pequeño, mediano y grande, respectivamente, en ausencia de información específica del estudio (por lo que tienen limitaciones y deberían usarse con precaución) (Cohen, 1992). Luego, $f = 4,227$ es ¡más de 10 veces! el umbral considerado para efectos grandes.

Una gracia de la función `wp.rmanova()`, que no tienen las funciones del paquete `pwr`, es que si se especifican múltiples valores a uno de los factores, con ciertas restricciones, la función devuelve múltiples valores para el factor que se está estimando. Por ejemplo, el siguiente código permite obtener una curva de poder para una prueba ANOVA de una vía para tres medidas repetidas para diferentes número de casos en las muestras considerando que se cumple la condición de esfericidad, que el efecto debería ser grande (pero no gigante) y un nivel de significación $\alpha = 0,05$:

```
> wp.rmanova(n = seq(20, 40, 5), ng = 1, nm = 3, f = 0.40, nscor = 1.00, alpha = 0.05,
+           power = NULL, type = 1)
Repeated-measures ANOVA analysis
```

n	f	ng	nm	nscor	alpha	power
20	0.4	1	3	1	0.05	0.3174641
25	0.4	1	3	1	0.05	0.3930828
30	0.4	1	3	1	0.05	0.4655163
35	0.4	1	3	1	0.05	0.5334160
40	0.4	1	3	1	0.05	0.5959355

NOTE: Power analysis for within-effect test

URL: <http://psychstat.org/rmanova>

>

10.9 PROCEDIMIENTOS POST-HOC

Podemos ocupar los mismos procedimientos post-hoc estudiados en el capítulo 9 tras realizar un procedimiento ANOVA de una vía con muestras correlacionadas. En el caso de las correcciones genéricas, lo único que cambia es que ahora debemos asignar el valor `TRUE` al argumento `paired` de la función `pairwise.t.test()`.

Para el ejemplo, el script 10.4 muestra este código para el ejemplo, cuya salida es mostrada en la figura 10.4.

```
Procedimiento post-hoc
=====

Procedimiento post-hoc de Holm
-----

Pairwise comparisons using paired t tests

data:  datos_largos[["Tiempo"]] and datos_largos[["Algoritmo"]]

      Bubblesort Mergesort Quicksort
Mergesort 0.00059      -      -
Quicksort 1.4e-05     0.02399  -
Radixsort 0.02399     0.00059  0.00033

P value adjustment method: holm

Procedimiento post-hoc de Benjamini y Hochberg
-----

Pairwise comparisons using paired t tests

data:  datos_largos[["Tiempo"]] and datos_largos[["Algoritmo"]]

      Bubblesort Mergesort Quicksort
Mergesort 0.00028      -      -
Quicksort 1.4e-05     0.01439  -
Radixsort 0.01539     0.00028  0.00020

P value adjustment method: fdr
```

Figura 10.4: resultado del procedimiento post-hoc usando la función `pairwise.t.test()` y correcciones genéricas para pruebas múltiples.

Script 10.4: (continuación del script 10.3) análisis post-hoc con los métodos genéricos

```
61 cat("\nProcedimiento post-hoc\n")
62 cat("=====\n\n")
63
64 # Definir el nivel de significación
65 alfa <- 0.01
66
67 # Realizar y mostrar un procedimiento post-hoc de Holm
68 holm <- pairwise.t.test(datos_largos[["Tiempo"]], datos_largos[["Algoritmo"]],
69                          p.adj = "holm", paired = TRUE, conf.level = 1 - alfa)
70 cat("Procedimiento post-hoc de Holm\n")
71 cat("-----\n")
72 print(holm)
73
74 # Realizar y mostrar un procedimiento post-hoc de Benjamini y Hochberg
75 bh <- pairwise.t.test(datos_largos[["Tiempo"]], datos_largos[["Algoritmo"]],
76                       p.adj = "fdr", paired = TRUE, conf.level = 1 - alfa)
77 cat("\nProcedimiento post-hoc de Benjamini y Hochberg\n")
78 cat("-----\n")
79 print(bh)
```

En cuanto a las pruebas HSD de Tukey y de Scheffé, su realización se dificulta puesto que el modelo subyacente a un procedimiento ANOVA de una vía para muestras correlacionadas es distinto al requerido para muestras independientes, y si bien usamos las mismas funciones para aplicar la prueba, estas diferencias impiden obtener un objeto con el formato "aov" requerido por las funciones `TukeyHSD()` y `ScheffeTest()` que utilizamos anteriormente.

No obstante, existe alternativas para conseguir estos análisis. Primero necesitamos crear un **modelo lineal (de efectos) mixto** (LME, o también LMM, del inglés *linear mixed-effects model*). Como se ha mencionado, a pesar que el estudio de este tipo de métodos va más allá de los alcances de este texto, corresponde a modelos que consideran tanto factores de efecto fijo como de efecto aleatorio para poder analizar estos efectos (lineales) sobre una variable numérica dependiente. En una prueba ANOVA de una vía con muestras correlacionadas, la variable independiente puede verse como un factor de efecto fijo, mientras que los casos incluidos en la muestra corresponden a un factor de efecto aleatorio (una selección aleatoria de todos los posibles casos).

La función `lme(formula, data, random)` del paquete `nlme` (acrónimo de *Linear and Nonlinear Mixed Effects Models*) permite construir modelos lineales mixtos, donde:

- `formula`: fórmula de la forma `<variable_dependiente>~<variable_independiente>`.
- `data`: matriz de datos en formato largo.
- `random`: fórmula de la forma `~1|<identificador del caso>`.

La fórmula del argumento `random` indica que la línea base de la variable dependiente (el intercepto, denotado como `1`) varía aleatoriamente (denotado `|`) entre los diferentes casos, capturando así la dependencia de las múltiples mediciones para el mismo caso.

Como segundo paso, debemos **estimar medias marginales** (EMM del inglés *estimated marginal mean*) para la variable dependiente para cada nivel de la variable independiente. A diferencia de las medias muestrales, las EMM estiman el valor medio que tendría la variable dependiente en la población si los otros factores del modelo mantuvieran un valor específico (usualmente su media). Esto permite realizar comparaciones más justas y significativas entre los niveles de un factor de interés, eliminando la influencia de cualquier otro factor en el modelo. Por supuesto, en el análisis de varianza de una vía no hay otros factores más que la variable independiente, por lo que las estimaciones son más simples en este caso.

La función `emmeans(object, specs)` del paquete homónimo, permite obtener las EMM, donde:

- `object`: modelo lineal mixto (construido previamente con la función `lme()`).
- `specs`: nombre del factor en estudio (es decir, la variable independiente), delimitado por comillas.

Con esto realizado, podemos analizar diferentes contrastes mediante la función `contrast(x, method, adjust)`, donde:

- `x`: resultado de la estimación de las medias marginales obtenidas (con la función `emmeans()`) en el paso precedente.
- `method`: string que indica el nombre de un método de contraste predefinido. Alternativamente, una lista en que cada elemento es un vector con los coeficientes de un contraste (para cada nivel del factor y que deben sumar cero), asociado a un nombre para el contraste que define.
- `adjust`: método para ajustar los valores p de las comparaciones múltiples.

Para realizar un procedimiento HSD de Tukey, basta con indicar el método predefinido `method = "pairwise"`, que compara todos los pares de mediciones, y elegir el ajuste `adjust = "tukey"`. Para aplicar la corrección de Scheffé, basta cambiar este último parámetro a `adjust = "scheffe"`. Si no se quiere comparar todos los pares de niveles del factor, como indica la descripción anterior, se deben indicar un nombre y un vector de coeficiente para cada contraste solicitado.

Veamos cómo utilizar esta alternativa en el ejemplo de los algoritmos de ordenamiento, que es mostrado en el script 10.5. En la línea 82 se construye el modelo lineal mixto, que es usando en la línea 83 para estimar las medias marginales para cada algoritmo. Luego, la línea 86 aplican la comparaciones de a pares de Tukey, que es mostrada en pantalla en la línea 89. Notemos aquí, que el método `print()` para los contrastes incluyen, entre varios otros, los argumentos `infer` y `level`. El primer argumento es un vector de uno o dos valores lógicos, donde el primero indica si se muestran intervalos de confianza, y el segundo determina si se muestran los resultados de las pruebas t múltiples, incluyendo los valores p obtenidos. Si el vector solo tiene un valor lógico, este se utiliza para ambas decisiones. El segundo argumento permite especificar el nivel de confianza a utilizar al calcular intervalos de confianza, si es que se ha solicitado. Así, la línea 89 del script solicita mostrar intervalos con 99 % confianza.

Script 10.5: (continuación del script 10.4) análisis post-hoc utilizando funciones para la creación y análisis de modelos lineales de efectos mixtos

```

81 # Construir un modelo lineal mixto y estimar medias marginales
82 mixto <- lme(Tiempo ~ Algoritmo, data = datos_largos, random = ~1|Instancia)
83 medias <- emmeans(mixto, "Algoritmo")
84
85 # Realizar, mostrar y graficar un procedimiento post-hoc HSD de Tukey
86 hsd <- contrast(medias, method = "pairwise", adjust = "tukey")
87 cat("\nProcedimiento HSD de Tukey\n")
88 cat("-----\n")
89 print(hsd, infer = TRUE, level = 1 - alfa)
90 print(plot(hsd, level = 1 - alfa, colors = "steelblue") +
91       ggtitle("Post-hoc HSD de Tukey") +
92       xlab("Diferencias en tiempos de ejecución\npara las mismas instancias") +
93       ylab("Pares de algoritmos") + theme_pubr())
94
95 # Crear la lista de contrastes a analizar con la prueba de Scheffé
96 contrastes <- list("Q vs B" = c(-1, 0, 1, 0), "Q vs M" = c(0, -1, 1, 0),
97                  "Q vs R" = c(0, 0, 1, -1), "M vs B" = c(-1, 1, 0, 0),
98                  "M vs R" = c(0, 1, 0, -1))
99
100 # Realizar y mostrar un procedimiento post-hoc de Scheffé
101 scheffe <- contrast(medias, method = contrastes, adjust = "scheffe")
102 cat("\nProcedimiento post-hoc de Scheffé\n")
103 cat("-----\n")
104 print(scheffe, infer = TRUE, level = 1 - alfa)
105 print(plot(scheffe, level = 1 - alfa, colors = "steelblue") +
106       ggtitle("Post-hoc de Scheffé") +
107       xlab("Diferencias en tiempos de ejecución\npara las mismas instancias") +
108       ylab("Pares de algoritmos") + theme_pubr())
109
110 # Mostrar y graficar tamaños del efecto
111 cat("\nIntervalos de confianza para las medias marginales\n")
112 cat("-----\n")
113 print(summary(medias, level = 1 - alfa, adjust = "none"))
114 print(plot(medias, level = 1 - alfa, adjust = "none", colors = "steelblue") +
115       ggtitle(sprintf("Tamaño del efecto (%d% IC)", round((1 - alfa) * 100))) +
116       xlab("EMM del tiempos de ejecución\nrequerido para las mismas instancias") +
117       theme_pubr())

```

Mas adelante, en las línea 96–98, se definen contrastes específicos para el procedimiento post-hoc de Scheffé. Estos se definieron considerando la aproximación gráfica de la figura 10.3, donde se aprecia que el algoritmo Quicksort (Q) podría ser más rápido que todos los otros, mientras que el algoritmo Mergesort (M) parece superar en eficiencia a los algoritmos Bubblesort (B) y Radixsort (R). Los efectos definidos en la lista construida en estas líneas especifican exactamente eso: comparar las diferencias en el tiempo de ejecución necesitado por Q contra los de los demás, y los del algoritmo M contra los requeridos por B y R. Estos contrastes se calculan en la línea 101 del script, y sus resultados se despliegan en pantalla en la línea 104, junto a los intervalos de confianza e información de las pruebas t realizadas.

Los resultados de ejecutar esta parte del script puede encontrarse en la figura 10.5, donde podemos ver que las comparaciones de a pares con el ajuste de Tukey detecta diferencias significativas, con un nivel de significación $\alpha = 0,99$ entre todos los pares de algoritmos ($t(15) \geq 4,531$; $p_{\text{adj.}} \leq 0,002$), con la excepción del par Bubblesort-Radixsort ($t(15) = 2,996$; $p_{\text{adj.}} = 0,040$).

También observamos que estos resultados son consistentes con los entregados por el procedimiento post-hoc de Scheffé, que ratifica que existen diferencias significativas entre Quicksort y los otros algoritmos de ordenamiento ($t(15) \geq 4,531$; $p_{\text{adj.}} \leq 0,004$), y entre Mergesort y las otras dos alternativas ($t(15) \geq 9,062$; $p_{\text{adj.}} < 0,001$).

Así, podemos concluir con 99 % confianza que Quicksort tarda menos que todos los otros algoritmos en resolver las mismas instancias de prueba, y que le sigue el algoritmo Mergesort que es más rápido que Bubblesort y Radixsort. Entre estos últimos, no parece haber diferencias significativas en el tiempo de ejecución que requieren.

Notemos dos ventajas interesantes de utilizar esta alternativa con modelos lineales mixtos. La primera, es que podemos obtener más detalles de las pruebas múltiples aplicadas en los procedimientos post-hoc de lo que podíamos conseguir con las funciones tradicionales. De hecho, existe un buen número de contrastes diferentes,

Procedimiento HSD de Tukey

contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Bubblesort - Mergesort	5.37	0.445	15	3.714	7.02	12.058	<.0001
Bubblesort - Quicksort	7.38	0.445	15	5.731	9.04	16.589	<.0001
Bubblesort - Radixsort	1.33	0.445	15	-0.319	2.99	2.996	0.0403
Mergesort - Quicksort	2.02	0.445	15	0.364	3.67	4.531	0.0020
Mergesort - Radixsort	-4.03	0.445	15	-5.686	-2.38	-9.062	<.0001
Quicksort - Radixsort	-6.05	0.445	15	-7.703	-4.40	-13.594	<.0001

Degrees-of-freedom method: containment

Confidence level used: 0.99

Conf-level adjustment: tukey method for comparing a family of 4 estimates

P value adjustment: tukey method for comparing a family of 4 estimates

Procedimiento post-hoc de Scheffé

contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Q vs B	-7.38	0.445	15	-9.18	-5.589	-16.589	<.0001
Q vs M	-2.02	0.445	15	-3.81	-0.223	-4.531	0.0040
Q vs R	-6.05	0.445	15	-7.84	-4.256	-13.594	<.0001
M vs B	-5.37	0.445	15	-7.16	-3.573	-12.058	<.0001
M vs R	-4.03	0.445	15	-5.83	-2.239	-9.062	<.0001

Degrees-of-freedom method: containment

Confidence level used: 0.99

Conf-level adjustment: scheffe method with rank 3

P value adjustment: scheffe method with rank 3

Figura 10.5: resultado del procedimiento post-hoc usando modelos lineales de efectos mixtos.

además de los específicos que se puedan definir, que pueden ser estudiados. Más aún, estos contrastes no están amarrados al tipo de ajuste aplicado a los valores p múltiples. Incluso, si nos fijamos bien en los reportes de la figura 10.5, podemos ver que ni siquiera es obligación aplicar la misma corrección a los valores p que a los intervalos de confianza⁵. Así, podríamos usar los contrastes específicos definidos en la línea 96-98 del script 10.5, pero ajustar los valores p con el método de Benjamini y Hochberg (obteniendo información de las pruebas t múltiples) y los intervalos de confianza con la corrección de Tukey.

La segunda ventaja, es que ahora sí podemos tener una representación gráfica más exacta del tamaño del efecto, tanto para los diferentes niveles de la variable independiente, como para las diferencias (o contrastes). La línea 113 del script muestra en pantalla los siguientes intervalos de confianza:

Intervalos de confianza para las medias marginales

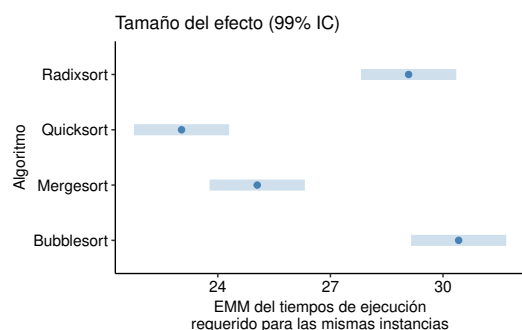
Algoritmo	emmean	SE	df	lower.CL	upper.CL
Bubblesort	30.4	0.315	5	29.1	31.7
Mergesort	25.1	0.315	5	23.8	26.3
Quicksort	23.0	0.315	5	21.8	24.3
Radixsort	29.1	0.315	5	27.8	30.4

Degrees-of-freedom method: containment

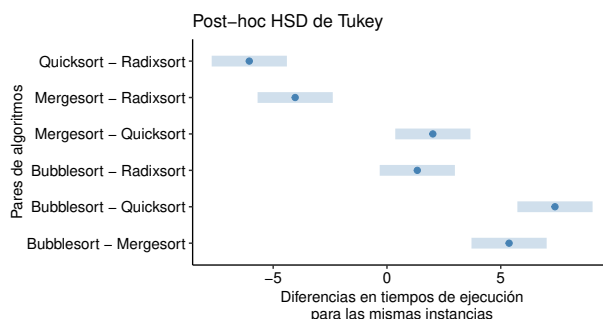
Confidence level used: 0.99

que fácilmente podemos graficar como muestran la línea 114 del script 10.5 (con algunos ajustes estéticos en las líneas siguientes). El gráfico resultante es mostrado en la figura 10.6a. Análogamente, se obtienen representaciones gráficas de los resultados de los procedimientos post-hoc HSD de Tukey (líneas 90-93) y de Scheffé (líneas 105-108), que son mostradas, respectivamente, por las figuras 10.6b. y 10.6c.

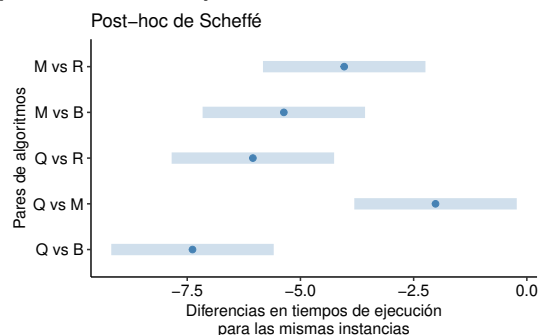
⁵En realidad, no todos los métodos funcionan para ambos tipos de ajuste.



(a) Tamaño del efecto estimado para cada algoritmo.



(b) Tamaño de las diferencias con ajuste HSD de Tukey.



(c) Tamaño de las diferencias con ajuste de Scheffé.

Figura 10.6: Gráficos de los tamaños de los efectos usando estimaciones de medias marginales.

10.10 RESUMEN DEL PROCEDIMIENTO

Como hicimos para las muestras independientes, una descripción muy mecánica el procedimiento ANOVA de una vía para muestras correlacionadas podría resumirse en los siguientes pasos:

1. Determinar las hipótesis ómnibus.
2. Verificar el cumplimiento de las condiciones, y no seguir si hay dudas de ello.
3. Definir un nivel de significación.
4. Calcular la suma de cuadrados para la muestra combinada (SS_T).
5. Para cada medición j , calcular la suma de cuadrados de la muestra (SS_j).
6. Calcular la variabilidad entre mediciones (SS_{bg}).
7. Calcular la variabilidad intramediciones (SS_{wg}).
8. Calcular la variabilidad intracasos y la variabilidad del error (SS_{casos} y SS_{error}).
9. Calcular los grados de libertad (ν_T , ν_{bg} y ν_{error}).
10. Calcular los cuadrados medios (MS_{bg} y MS_{error}).
11. Calcular el estadístico de prueba (F).
12. Obtener el valor p.

13. Concluir sobre las hipótesis ómnibus.
14. Si la prueba resulta significativa:
 - a) Seleccionar un procedimiento post-hoc.
 - b) Verificar que se cumplen las condiciones de la prueba post-hoc.
 - c) Aplicar el procedimiento post-hoc seleccionado, ya sea genérico o usando un modelo lineal de efectos mixtos equivalente.
 - d) Concluir sobre los resultados post-hoc.

Como siempre, no es recomendable seguir ciegamente estos pasos sin considerar que los resultados que se vayan obteniendo tienen sentido.

10.11 EJERCICIOS PROPUESTOS

- 10.1 El conjunto de datos `ChickWeight` registra el peso (en gramos) de 50 pollitos al momento de nacer y al cabo de varios días después de nacidos. Una empresaria agrícola está construyendo un nuevo criadero y necesita saber si existen diferencias significativas en el peso de los pollitos a los 8, 14 y 20 días de nacidos, para poder planificar el número y tamaño de los galpones de crianza.
 - (a) Enuncia las hipótesis a ser contrastadas en este caso y verifica si se cumplen las condiciones para aplicar una prueba ANOVA para muestras correlacionadas.
 - (b) Independiente del resultado anterior, realiza la prueba y entrega una conclusión para la empresaria.
 - (c) Realiza un análisis post-hoc usando la prueba HSD de Tukey. ¿El resultado post-hoc es consistente con el resultado ómnibus?
- 10.2 Da un ejemplo de una pregunta de investigación sobre las asignaturas comunes en ingeniería que requiera utilizar una prueba ANOVA para medidas repetidas. Identifica bien las variables involucradas y enuncia las hipótesis a docimar.
- 10.3 Da un ejemplo de una pregunta de investigación sobre los conciertos realizados en Santiago que requiera utilizar una prueba ANOVA para medidas repetidas. Identifica bien las variables involucradas y enuncia las hipótesis a docimar.
- 10.4 Da un ejemplo de una pregunta de investigación sobre el estado de la salud mental de estudiantes universitarios que requiera utilizar una prueba ANOVA para medidas repetidas. Identifica bien las variables involucradas y enuncia las hipótesis a docimar.
- 10.5 Justificando tus suposiciones, inventa muestras de los datos que se podrían encontrar en el estudio propuesto en la pregunta 10.2 que tengan entre 15 y 18 observaciones y que cumplan las condiciones requeridas por la prueba ANOVA para medidas repetidas.
- 10.6 Justificando tus suposiciones, inventa muestras de los datos que se podrían encontrar en el estudio propuesto en la pregunta 10.3 que tengan entre 15 y 18 observaciones y que cumplan las condiciones requeridas por la prueba ANOVA para medidas repetidas.
- 10.7 Justificando tus suposiciones, inventa muestras de los datos que se podrían encontrar en el estudio propuesto en la pregunta 10.4 que tengan entre 15 y 18 observaciones y que cumplan las condiciones requeridas por la prueba ANOVA para medidas repetidas.
- 10.8 Usando la función `ezANOVA()` realiza el análisis de las muestras definidas en la pregunta 10.5. Aplica la prueba HSD de Tukey para el análisis post-hoc (aún cuando este sea innecesario en estricto rigor).
- 10.9 Usando la función `ezANOVA()` realiza el análisis de las muestras definidas en la pregunta 10.6. Aplica la prueba de Scheffé para el análisis post-hoc (aún cuando este sea innecesario en estricto rigor).
- 10.10 Usando la función `ezANOVA()` realiza el análisis de las muestras definidas en la pregunta 10.7. Aplica el método de Benjamini y Hochberg para el análisis post-hoc (aún cuando este sea innecesario en estricto rigor).
- 10.11 Calcula el poder obtenido en alguna de las tres pruebas realizadas en 10.8, 10.9 y 10.10.

- 10.12 Usando modelos lineales mixtos, rehace el procedimiento post-hoc HSD de Tukey de la pregunta 10.8, obteniendo información detallada del tamaño del efecto y las pruebas t múltiples.
- 10.13 Usando modelos lineales mixtos, rehace el procedimiento post-hoc de Scheffé de la pregunta 10.9, obteniendo información detallada del tamaño del efecto y las pruebas t múltiples.
- 10.14 Usando modelos lineales mixtos, rehace el procedimiento post-hoc de Benjamini y Hochberg de la pregunta 10.10, obteniendo información detallada del tamaño del efecto y las pruebas t múltiples.

10.12 BIBLIOGRAFÍA DEL CAPÍTULO

- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- 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>
- Karadimitriou, S. M., & Marshall, E. (2016). *Repeated measures ANOVA in R* [statstutor community project]. Consultado el 12 de mayo de 2021, desde https://www.sheffield.ac.uk/polopoly_fs/1.885219!/file/105_RepeatedANOVA.pdf
- Lærd Statistics. (2020). *Sphericity* [Lund Research Ltd.]. Consultado el 11 de mayo de 2021, desde <https://statistics.laerd.com/statistical-guides/sphericity-statistical-guide.php>
- Lowry, R. (1999). *Concepts & Applications of Inferential Statistics*. Consultado el 3 de mayo de 2021, desde <http://vassarstats.net/textbook/>
- Subject [Accessed through dictionary boxes on Google]. (2023, octubre). En *Oxford Languages*. Oxford University Press.
- Zhang, Z., & Yuan, K.-H. (2018). *Practical statistical power analysis using Webpower and R*. Isdsa Press. <https://doi.org/10.35566/power>