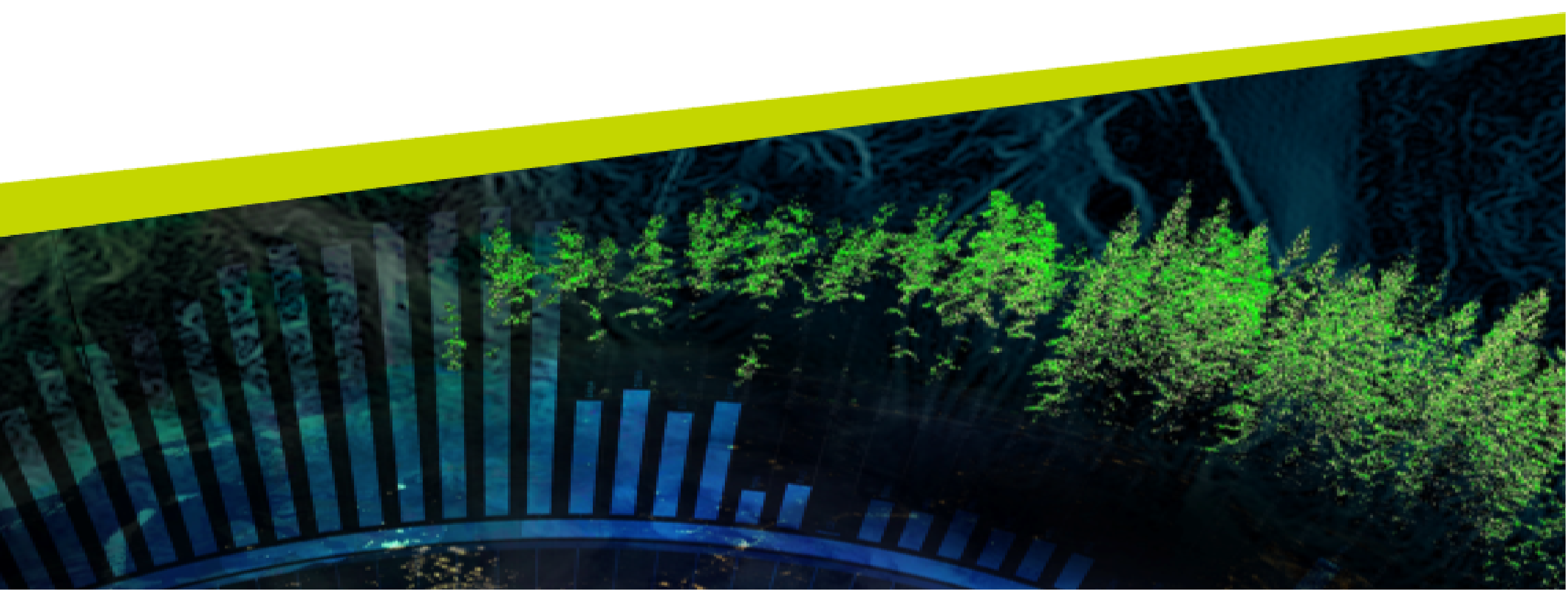
INFERENCIA Y MODELOS ESTADÍSTICOS

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.

En el procedimiento ANOVA para muestras correlacionadas podemos distinguir 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 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 grupos de casos parecidos en bloques). 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 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 10.1[2].

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

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

En este caso, la lógica es muy similar a la que ya conocimos para ANOVA con muestras independientes. Sin

[1] El 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 las y los “ejemplares” o “especímenes” (dos palabras que comienza a ser usada en publicaciones científicas) que están *sujeto/a* a estudio.

[2] Los valores aquí expuestos son ficticios.

embargo, existe una diferencia importante al trabajar con muestras correlacionadas: no toda la variabilidad es pura e inevitable, sino que una parte de ella se debe a diferencias individuales preexistentes 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 el tiempo de ejecución necesitado por cada algoritmo. Si $\mu_{(B\,-\,A)}$ denota la media de las diferencias en tiempos de ejecución necesitado por algoritmos A y B para ordenar todos los arreglos posibles, entonces la hipótesis puede escribirse como: $\mu_{(\text{Bubblesort}\,-\,\text{Quicksort})} = \mu_{(\text{Radixsort}\,-\,\text{Quicksort})} = \mu_{(\text{Mergesort}\,-\,\text{Quicksort})} = \mu_{(\text{Radixsort}\,-\,\text{Bubblesort})} = \mu_{(\text{Mergesort}\,-\,\text{Bubblesort})} = \mu_{(\text{Mergesort}\,-\,\text{Radixsort})} = 0$.

H_A: la media de las diferencias en el tiempo de ejecución necesitado es diferente para al menos un par de algoritmos. Matemáticamente: $\exists A, B \in \{\text{Quicksort}, \text{Bubblesort}, \text{Radixsort}, \text{Mergesort}\}, A \neq B |\ \mu_{(\text{B}\,-\,\text{A})} \neq 0$.

10.1 CONDICIONES PARA USAR ANOVA DE UNA VÍA PARA MUESTRAS CORRELACIONADAS

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 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.

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). A su vez, el enunciado señala que el proceso seguido por el ingeniero garantiza el cumplimiento de la segunda condición.

La figura 10.1 (creada mediante el script 10.1, líneas 19–25) muestra gráficos Q-Q para cada grupo, 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.

La prueba de esfericidad es más compleja, por lo que no se aborda en este texto. Para el ejemplo, se debería revisar que:

$$\sigma^2_{(\text{Bubblesort}\,-\,\text{Quicksort})} = \sigma^2_{(\text{Radixsort}\,-\,\text{Quicksort})} = \sigma^2_{(\text{Mergesort}\,-\,\text{Quicksort})} = \sigma^2_{(\text{Radixsort}\,-\,\text{Bubblesort})} = \sigma^2_{(\text{Mergesort}\,-\,\text{Bubblesort})} = \sigma^2_{(\text{Mergesort}\,-\,\text{Radixsort})}$$

En principio, podemos examinar las diferencias entre las varianzas registradas para cada algoritmo (que se muestran en la tabla 10.2). Podemos ver que las diferencias parecen más bien "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 es más de 3 veces el menor valor. 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.

	Bubblesort	Radixsort	Mergesort
Quicksort	0,57	1,47	1,65
Bubblesort		0,82	1,79
Radixsort			0,91

Tabla 10.2: varianzas muestrales de las diferencias en el tiempo de ejecución entre cada par de algoritmos.

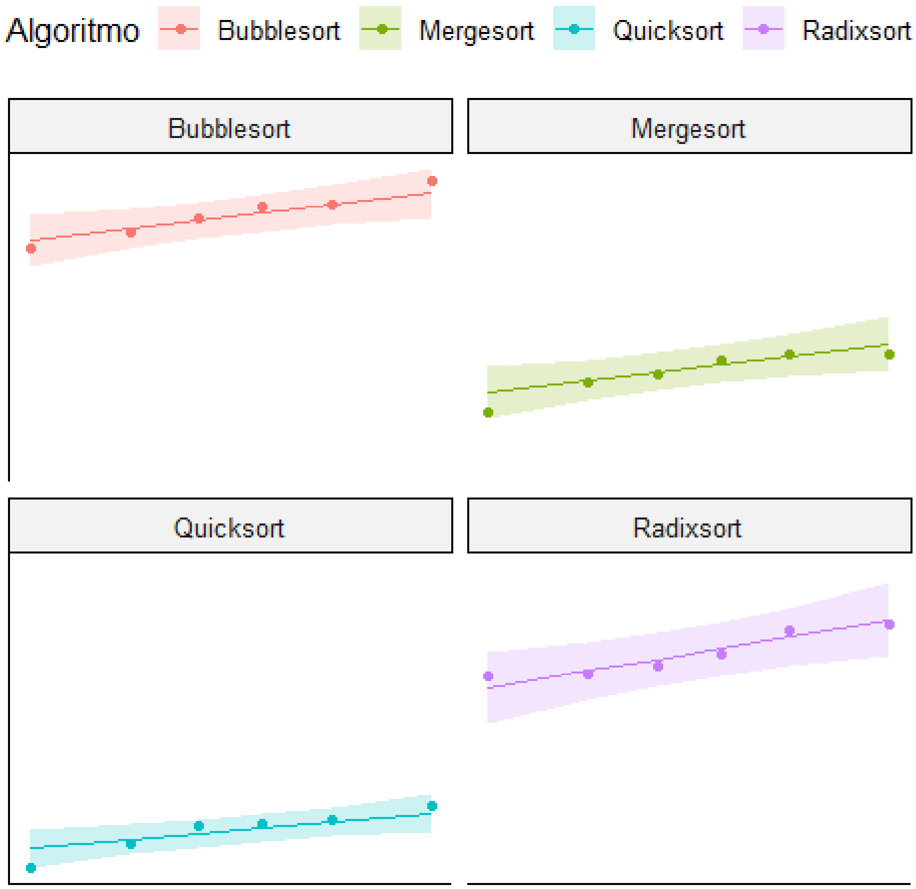


Figura 10.1: gráfico para comprobar el supuesto de normalidad en las muestras del ejemplo.

10.2 PROCEDIMIENTO ANOVA DE UNA VÍA PARA MUESTRAS CORRELACIONADAS

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:

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

10.2.1 Variabilidad total, entre grupos e intragrupos

Los primeros pasos son idénticos a los que ya estudiamos para ANOVA de una vía con muestras independientes, y consisten en calcular la variabilidad total (es decir, para las muestras combinadas), la variabilidad entre grupos y la variabilidad intragrupos, denotadas por SS_T, SS_{bg} y SS_{wg}, respectivamente:

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

10.2.2 Variabilidad debida a los casos

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, 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.

$$SS_{wg} = SS_{\text{casos}} + SS_{\text{error}} \tag{10.1}$$

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

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

$$SS_{\text{casos}} = k \cdot \sum_{i=1}^{n} (\overline{x}_i - \overline{x}_T)^2 \tag{10.2}$$

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

Instancia	Quicksort	Bubblesort	Radixsort	Mergesort	$\overline{x}$
1	23,2	31,6	30,1	25,0	27,475
2	22,6	29,3	28,4	25,7	26,500
3	23,4	30,7	28,7	25,7	27,125
4	23,3	30,8	28,3	23,7	26,525
5	21,8	29,8	29,9	25,5	26,750
6	23,9	30,3	29,1	24,7	27,000

Tabla 10.3: tiempos de ejecución y tiempo medio de ejecución para las diferentes instancias del ejemplo.

Luego, para el ejemplo:

$$\begin{aligned} SS_{\text{casos}} &= 4 \cdot [(27{,}475 - 26{,}896)^2 + (26{,}500 - 26{,}896)^2 + (27{,}125 - 26{,}896)^2 + \\ &= \quad (26{,}525 - 26{,}8967)^2 + (26{,}750 - 26{,}896)^2 + (27{,}000 - 26{,}896)^2] \\ &= 2{,}857 \end{aligned}$$

y

$$\begin{aligned} SS_{\text{error}} &= SS_{wg} - SS_{\text{casos}} \\ &= 11{,}885 - 2{,}857 \\ &= 9{,}028 \end{aligned}$$

10.2.3 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):

$$\nu_T = n_T - 1 = 24 - 1 = 23 \tag{10.3}$$

$$\nu_{bg} = k - 1 = 4 - 1 = 3 \tag{10.4}$$

$$\nu_{wg} = n_T - k = 24 - 4 = 20 \tag{10.5}$$

[3]Tradicionalmente, los textos se refieren a esta variabilidad como "intrasujetos".

Puesto que anteriormente descompusimos la variabilidad intragrupos en variabilidad individual y variabilidad del error, necesitamos también identificar los grados de libertad correspondientes a cada componente, dados por las ecuaciones 10.6 y 10.7, respectivamente.

$$\nu_{\text{casos}} = n_{\text{casos}} - 1 \tag{10.6}$$
$$\nu_{\text{error}} = \nu_{wg} - \nu_{\text{casos}} \tag{10.7}$$

Para el ejemplo, tenemos:

$$\begin{aligned} \nu_{\text{casos}} &= 6 - 1 &= 5 \\ \nu_{\text{error}} &= 20 - 5 &= 15 \end{aligned}$$

Las medias cuadradas del procedimiento ANOVA de una vía para muestras correlacionadas son, respectivamente, la media cuadrada entre grupos para el efecto (igual que para muestras independientes) y la media cuadrada del error, dada por la ecuación 10.8.

$$MS_{\text{error}} = \frac{SS_{\text{error}}}{\nu_{\text{error}}} \tag{10.8}$$

Así, tenemos que:

$$\begin{aligned} MS_{\text{efecto}} = MS_{bg} &= \frac{213{,}045}{3} &= 71{,}015 \\ MS_{\text{error}} = &= \frac{9{,}028}{15} &= 0{,}602 \end{aligned}$$

En consecuencia:

$$\frac{MS_{\text{efecto}}}{MS_{\text{error}}} = F = \frac{71{,}015}{0{,}602} = 117{,}992$$

En R, al hacer la llamada `pf(117.992, 3, 15, lower.tail = FALSE)` para calcular el valor p, obtenemos $p < 0{,}001$ (en rigor matemático, $p \geq 1{,}177 \cdot 10^{-10}$).

10.2.4 Resultado del procedimiento ANOVA

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	71,015	117,991	$1{,}177 \cdot 10^{-10}$
Error	15	9,028	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.

10.2.5 Resumen del procedimiento ANOVA de una vía para muestras correlacionadas

El procedimiento ANOVA para medidas repetidas puede resumirse en los siguientes pasos:

1. Calcular la suma de los cuadrados de las desviaciones para la muestra combinada (SS_T).
2. Para cada grupo g, calcular la suma de los cuadrados de las desviaciones dentro de dicho grupo (SS_g).
3. Calcular la variabilidad entre grupos (SS_{bg}).
4. Calcular la variabilidad al interior de los grupos (SS_{wg}).
5. Calcular la variabilidad intracasos y la variabilidad del error (SS_{casos} y SS_{error}).
6. Calcular los grados de libertad relevantes (ν_T, $\nu_{\text{efecto}} = \nu_{bg}$ y ν_{error}).
7. Calcular las medias cuadradas ($MS_{\text{efecto}} = MS_{bg}$ y MS_{error}).
8. Calcular el estadístico de prueba (F).
9. Obtener el valor p.

10.3 ANOVA DE UNA VÍA PARA MUESTRAS CORRELACIONADAS 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 10.1.

En el caso de `aov()`, podemos apreciar que la fórmula entregada en la llamada (líneas 28–29 del script 10.1) 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 que se requiere descartar la variabilidad entre casos. La figura 10.2 muestra el resultado obtenido, idéntico al presentado en la tabla 10.4 salvo por ligeras diferencias de redondeo.

```
Resultado de la prueba ANOVA para muestras correlacionadas con 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
```

Figura 10.2: resultado del procedimiento ANOVA usando la función `aov()`.

La llamada a `ezANOVA()`, en cambio, es muy similar a la ya conocida, como se puede apreciar en las líneas 34–35 del script 10.1: 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. La figura 10.3 muestra, una vez más, que se obtiene el mismo resultado.

Script 10.1: procedimiento ANOVA de una vía para muestras correlacionadas.

```r
1 library(tidyverse)
2 library(ggpubr)
3 library(ez)
4
5 # Crear el data frame.
6 instancia <- factor(1:6)
7 Quicksort <- c(23.2, 22.6, 23.4, 23.3, 21.8, 23.9)
8 Bubblesort <- c(31.6, 29.3, 30.7, 30.8, 29.8, 30.3)
9 Radixsort <- c(30.1, 28.4, 28.7, 28.3, 29.9, 29.1)
10 Mergesort <- c(25.0, 25.7, 25.7, 23.7, 25.5, 24.7)
11 datos <- data.frame(instancia, Quicksort, Bubblesort, Radixsort, Mergesort)
12
```

```
Resultado de la prueba ANOVA para muestras correlacionadas con ezANOVA

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
```

Figura 10.3: resultado del procedimiento ANOVA usando la función `ezANOVA()`.

```r
# Llevar data frame a formato largo.
datos <- datos %>% pivot_longer(c("Quicksort", "Bubblesort", "Radixsort",
                                  "Mergesort"),
                                names_to = "algoritmo", values_to = "tiempo")
datos[["algoritmo"]] <- factor(datos[["algoritmo"]])

# Comprobación de normalidad.
g <- ggqqplot(datos, x = "tiempo", y = "algoritmo", color = "algoritmo")
g <- g + facet_wrap(~ algoritmo)
g <- g + rremove("x.ticks") + rremove("x.text")
g <- g + rremove("y.ticks") + rremove("y.text")
g <- g + rremove("axis.title")
print(g)

# Procedimiento ANOVA con aov().
prueba <- aov(tiempo ~ algoritmo + Error(instancia/(algoritmo)),
              data = datos)
cat("\nResultado de la prueba ANOVA para muestras correlacionadas con aov\n")
print(summary(prueba))

# Procedimiento ANOVA con ezANOVA().
prueba2 <- ezANOVA(data = datos, dv = tiempo, within = algoritmo,
                   wid = instancia, return_aov = TRUE)
cat("\nResultado de la prueba ANOVA para muestras correlacionadas con ezANOVA\n")
print(summary(prueba2[["aov"]]))

# Pero ezANOVA entrega más información.
cat("\nResultado de la prueba de esfericidad de Mauchly\n")
print(prueba2[["Mauchly's Test for Sphericity"]])

cat("\nFactores de corrección para cuando no se ")
cat("cumple la condición de esfericidad\n")
print(prueba2[["Sphericity Corrections"]])

# Gráfico del tamaño del efecto.
g2 <- ezPlot(data = datos, dv = tiempo, wid = instancia, within = algoritmo,
             y_lab = "Tiempo promedio de ejecución [ms]", x = algoritmo)
g2 <- g2 + theme_pubr()
print(g2)
```

Habíamos mencionado que otra ventaja de `ezANOVA()` es que verifica la condición de esfericidad mediante la prueba de esfericidad de Mauchly, cuyo resultado se muestra en la figura 10.4. Podemos apreciar que el valor p obtenido en esta prueba es muy alto ($p = 0{,}555$), de lo que se desprende que los datos del ejemplo sí cumplen con la condición de esfericidad (hipótesis nula de la prueba de Mauchly).

```
Resultado de la prueba de esfericidad de Mauchly

     Effect         W         p p<.05
2 Algoritmo 0.3367911 0.5545469
```

Figura 10.4: resultado de la prueba de esfericidad de Mauchly realizada por `ezANOVA()`.

Ahora bien, existen dos correcciones que suelen emplearse cuando se producen violaciones a la condición de esfericidad: la de **Greenhouse-Geisser** y la de **Huynd-Feldt**. Ambas estiman la esfericidad, denotada por ϵ, y corrigen el valor p de ANOVA en base a dicha 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 esta es cercana a 1, por lo que se recomienda su uso para $\epsilon < 0{,}75$. Para $\epsilon \geq 0{,}75$ estimada con el método Greenhouse-Geisser, de acuerdo a Karadimitriou y Marshall (2016), suele emplearse la estimación de Huynd-Feldt, 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 10.5, 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 Huynd-Feldt.
- p[HF]: valor p tras la corrección de Huynd-Feldt.

```
Factores de corrección para cuando no se cumple la condición de esfericidad

     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 10.5: correcciones de esfericidad realizadas por `ezANOVA()`.

Si los datos del ejemplo no cumplieran con la esfericidad, deberíamos considerar `p[GG]` como p valor de la prueba, y no el valor (sin corregir) de la tabla entregada por `ezANOVA()` de la figura 10.3. Una vez más, podemos graficar el tamaño del efecto medido (script 10.1, líneas 48–50), obteniéndose como resultado el gráfico de la figura 10.6.

10.4 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 de Bonferroni y Holm, lo único que cambia es que ahora debemos asignar el valor `TRUE` al argumento `paired` de la función `pairwise.t.test()`.

En cuanto a las pruebas HSD de Tukey y de Scheffé, su realización se dificulta al usar la tabla ANOVA resultante de un proceso de una vía para muestras correlacionadas. Esto se debe a que el formato del objeto `aov` resultante difiere al que se obtiene al realizar un procedimiento ANOVA para muestras independientes, por lo que las funciones del paquete `DescTools` arrojan un error. No obstante, existe una alternativa. El primer paso consiste en crear un modelo mixto (concepto que va más allá de los alcances de este documento) mediante la función `lme(formula, data, random)` del paquete `nlme`, donde:

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

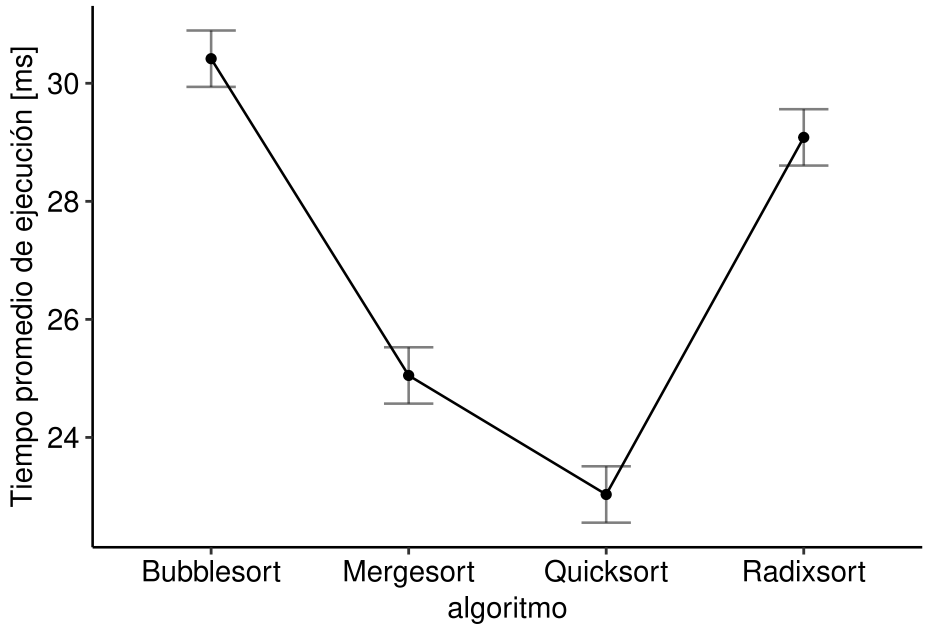


Figura 10.6: Tamaño del efecto medido.

Como segundo paso, estimamos la media de la variable dependiente, con su respectivo intervalo de confianza, para cada nivel de la variable categórica. Para esto usamos la función `emmeans(object, specs)` del paquete homónimo, donde:

- `object`: modelo mixto construido en el paso previo.
- `specs`: nombre del factor en estudio (es decir, la variable independiente), delimitado por comillas.

Por último, estimamos las medias de las diferencias para los contrastes entre pares, con su error estándar y los valores p correspondientes, mediante la función `pairs(x, adjust)`, donde:

- `x`: diferencias obtenidas en el paso precedente.
- `adjust`: método para ajustar los valores p. `"tukey"` para el método HSD de Tukey, `"scheffe"` para el método de Scheffé.

Los mecanismos para estimar las medias marginales ("emmeans") y para construir otros contrastes con el método de Scheffé van más allá de los alcances de este libro, pero pueden encontrarse en la documentación de los paquetes R involucrados.

El script 10.2 efectúa las pruebas post-hoc para el ejemplo, obteniéndose los resultados de la figura 10.7. Considerando el ajuste para múltiples pruebas de Tukey, podemos concluir con 99 % de confianza que todos los algoritmos tienen tiempos de ejecución distintos al resolver las mismas instancias ($t(15) \geq 4{,}531$; $p_{ajustado} \leq 0{,}002$), con la excepción del par Bubblesort/Radixsort, para el que la evidencia no es suficiente para descartar que presentan, en promedio, los mismos tiempos de ejecución ($t(15) = 2{,}996$; $p_{ajustado} = 0{,}040$).

Script 10.2: pruebas post-hoc para el ejemplo.

```
library(tidyverse)
library(nlme)
library(emmeans)
library(ez)

# Crear el data frame.
instancia <- factor(1:6)
Quicksort <- c(23.2, 22.6, 23.4, 23.3, 21.8, 23.9)
Bubblesort <- c(31.6, 29.3, 30.7, 30.8, 29.8, 30.3)
Radixsort <- c(30.1, 28.4, 28.7, 28.3, 29.9, 29.1)
```

```r
Mergesort <- c(25.0, 25.7, 25.7, 23.7, 25.5, 24.7)
datos <- data.frame(instancia, Quicksort, Bubblesort, Radixsort, Mergesort)

# Llevar data frame a formato largo.
datos <- datos %>% pivot_longer(c("Quicksort", "Bubblesort", "Radixsort",
                                  "Mergesort"),
                                names_to = "algoritmo", values_to = "tiempo")

datos[["algoritmo"]] <- factor(datos[["algoritmo"]])

# Nivel de significación.
alfa <- 0.01

# Procedimiento ANOVA.
anova <- ezANOVA(data = datos, dv = tiempo, within = algoritmo,
                 wid = instancia, return_aov = TRUE)

# Procedimiento post-hoc de Bonferroni.
bonferroni <- pairwise.t.test(datos[["tiempo"]], datos[["algoritmo"]],
                              p.adj = "bonferroni", paired = TRUE)

cat("Corrección de Bonferroni\n")
print(bonferroni)

# Procedimiento post-hoc de Holm.
holm <- pairwise.t.test(datos[["tiempo"]], datos[["algoritmo"]],
                        p.adj = "holm", paired = TRUE)

cat("\n\nCorrección de Holm\n")
print(holm)

# Procedimiento post-hoc HSD de Tukey.
mixto <- lme(tiempo ~ algoritmo, data = datos, random = ~1|instancia)
medias <- emmeans(mixto, "algoritmo")
tukey <- pairs(medias, adjust = "tukey")

cat("\n\nPrueba HSD de Tukey\n\n")
print(tukey)

# Procedimiento post-hoc de Scheffé
cat("\n\nComparación de Scheffé\n")
scheffe <- pairs(medias, adjust = "scheffe")
print(scheffe)
```

10.5 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 al momento de nacer, al cabo de 4 días y luego de 8 días, 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, usando el método de Benjamini y Hochberg para la corrección de pruebas múltiples.

(c) Realiza un análisis post-hoc usando la prueba HSD de Tukey. ¿Se llega a la misma conclusión?

```
Corrección de Bonferroni

        Pairwise comparisons using paired t tests

data:  datos[["tiempo"]] and datos[["algoritmo"]]

          Bubblesort Mergesort Quicksort
Mergesort 0.00112    -         -
Quicksort 1.4e-05    0.07196   -
Radixsort 0.09232    0.00088   0.00039

P value adjustment method: bonferroni


Corrección de Holm

Pairwise comparisons using paired t tests

data:  datos[["tiempo"]] and datos[["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


Prueba HSD de Tukey

 contrast               estimate    SE df t.ratio p.value
 Bubblesort - Mergesort     5.37 0.445 15  12.058  <.0001
 Bubblesort - Quicksort     7.38 0.445 15  16.589  <.0001
 Bubblesort - Radixsort     1.33 0.445 15   2.996  0.0403
 Mergesort - Quicksort      2.02 0.445 15   4.531  0.0020
 Mergesort - Radixsort     -4.03 0.445 15  -9.062  <.0001
 Quicksort - Radixsort     -6.05 0.445 15 -13.594  <.0001

Degrees-of-freedom method: containment
P value adjustment: tukey method for comparing a family of 4 estimates


Comparación de Scheffé
 contrast               estimate    SE df t.ratio p.value
 Bubblesort - Mergesort     5.37 0.445 15  12.058  <.0001
 Bubblesort - Quicksort     7.38 0.445 15  16.589  <.0001
 Bubblesort - Radixsort     1.33 0.445 15   2.996  0.0642
 Mergesort - Quicksort      2.02 0.445 15   4.531  0.0040
 Mergesort - Radixsort     -4.03 0.445 15  -9.062  <.0001
 Quicksort - Radixsort     -6.05 0.445 15 -13.594  <.0001

Degrees-of-freedom method: containment
P value adjustment: scheffe method with rank 3
```

Figura 10.7: resultados de las pruebas post-hoc para el ejemplo.

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 Investiga cómo obtener una medida estandarizada del tamaño del efecto para las pruebas ómnibus realizadas en las preguntas 8, 9 y 10.

10.12 Investiga cómo conocer el poder de las pruebas ómnibus realizadas en las preguntas 8, 9 y 10.

10.13 ¿Y qué hay de los tamaños del efecto de las pruebas post-hoc?

10.14 ¿Y qué hay del poder conseguido en las pruebas post-hoc?

10.6 BIBLIOGRAFÍA DEL CAPÍTULO

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

Subject [Accessed through dictionary boxes on Google]. (2023, octubre). En *Oxford Languages*. Oxford University Press.