

# Análisis de la varianza

Víctor Aceña - Isaac Martín - Carmen Lancho

DSLAB

2025-09-25



- En temas previos hemos comparado las medias de dos poblaciones

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

- ¿Y si queremos comprobar si hay diferencia entre 3 o más medias muestrales? Es decir, comprobar la hipótesis de que todas las medias (3 o más) son iguales → **ANOVA**

- **Análisis de la varianza**, conocido como **ANOVA** (del inglés *Analysis of Variance*)
- Técnica estadística utilizada para comparar las medias de dos o más grupos y determinar si existen diferencias significativas entre ellos
- Desarrollado por Fisher en las primeras décadas del siglo XX.
- La idea central es analizar la variabilidad de los datos y dividirla en componentes atribuibles a diferentes fuentes de variación

- En su forma más simple, el ANOVA se utiliza para probar hipótesis sobre las diferencias entre las medias de grupos
- Por ejemplo, si quisiéramos comparar el rendimiento de tres tipos diferentes de métodos educativos, podríamos usar ANOVA para determinar si el método educativo tiene un efecto significativo en el rendimiento académico:

Tipo de método	Rendimiento académico
Método 1	85, 78, 90, 82
Método 2	88, 79, 91, 85
Método 3	80, 75, 88, 83

1. **Formulación de hipótesis:** Se establece una hipótesis nula que indica que no hay diferencias entre las medias de los grupos y una hipótesis alternativa que sugiere que al menos una media es diferente

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$H_1$  : al menos una de las medias es distinta

2. **Cálculo de la varianza:** Se calculan dos tipos de varianza: la varianza dentro de los grupos (variabilidad debido a diferencias dentro de los mismos grupos) y la varianza entre los grupos (variabilidad debido a diferencias entre los grupos)

3. **F-test:** Se realiza una prueba  $F$  de Fisher para evaluar la relación entre las varianzas. Si la varianza entre los grupos es significativamente mayor que la varianza dentro de los grupos, esto sugiere que hay diferencias significativas entre las medias de los grupos
4. **Análisis de resultados:** Si la prueba  $F$  indica que hay diferencias significativas, se pueden realizar pruebas adicionales para identificar entre qué grupos existen estas diferencias

- Es una herramienta poderosa porque permite comparaciones múltiples mientras controla la tasa de error tipo I
- Es ampliamente utilizado en experimentos donde se comparan tratamientos o condiciones en diferentes grupos o en diferentes momentos

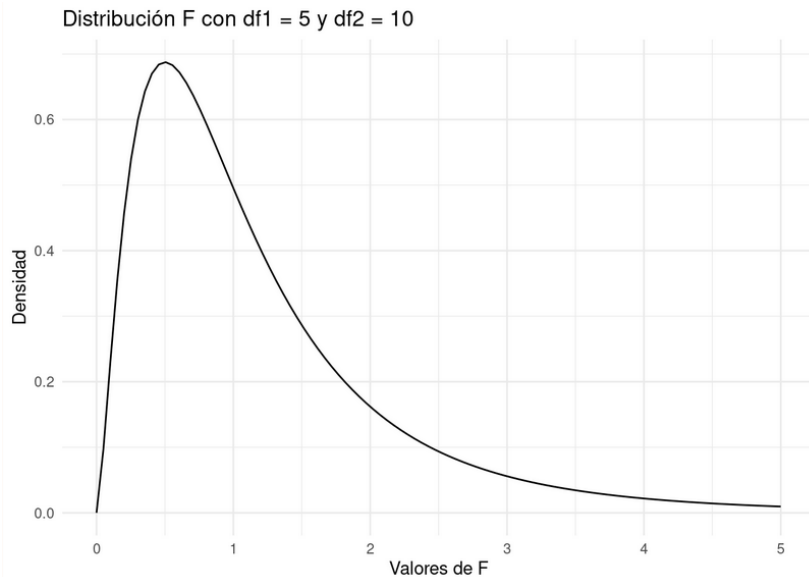
- Necesitamos una distribución muestral para comparar varianzas
- Se toma el estadístico cociente  $S_X^2/S_Y^2$ . Cuanto más próximo a 1, más parecidas las varianzas
- La distribución muestral de  $S_X^2/S_Y^2$  es la distribución  $F$
- Dadas dos muestras  $\mathbf{X} = (X_1, \dots, X_{n+1})$ ,  $\mathbf{Y} = (Y_1, \dots, Y_{m+1})$  de tamaño  $n$  y  $m$ , procedentes de dos poblaciones Normales con varianzas  $\sigma_X^2$  y  $\sigma_Y^2$ . Entonces

$$\frac{nS_X^2}{\sigma_X^2} \sim \chi_n^2 \quad \frac{mS_Y^2}{\sigma_Y^2} \sim \chi_m^2$$

y

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{m,n}$$





- Se utiliza cuando se estudia el efecto de un solo factor (variable independiente) en una variable dependiente continua
- Permite comparar las medias de varios grupos para determinar si existen diferencias significativas entre ellos
- **Hipótesis:**
  - **Hipótesis Nula ( $H_0$ ):** Todas las medias de los grupos son iguales ( $\mu_1 = \mu_2 = \dots = \mu_k$ ).
  - **Hipótesis Alternativa ( $H_1$ ):** Al menos una de las medias de los grupos es diferente.

Tenemos una variable aleatoria  $Y$  que toma valores reales y una variable cualitativa o factor  $X$  con  $k$  niveles  $1, 2, \dots, i, \dots, k$ . La variable  $Y$  toma valores  $Y_{ij}, j = 1, \dots, n_i$  en el nivel  $i$  del factor  $X$ , siendo  $n_i$  el número de observaciones en el nivel  $i$  del factor  $X$

$X_1$	$Y_{11}, Y_{12}, \dots, Y_{1n_1}$
$\vdots$	$\vdots$
$X_i$	$Y_{i1}, Y_{i2}, \dots, Y_{in_i}$
$\vdots$	$\vdots$
$X_k$	$Y_{k1}, Y_{k2}, \dots, Y_{kn_k}$

Tenemos los siguientes supuestos:

- Normalidad: Las distribuciones de las poblaciones de las que provienen las muestras son normales. Se supone que los errores  $\epsilon_{ij}$  están distribuidos como una Normal de media 0 y varianza  $\sigma^2$
- Homogeneidad de varianzas: Las varianzas de las poblaciones son iguales
- Independencia: Las observaciones son independientes entre sí

El modelo teórico es como sigue:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Donde:

- $Y_{ij}$  es la observación  $j$ -ésima del grupo  $i$ -ésimo
- $\mu$  es la media general
- $\epsilon_{ij}$  es el término de error aleatorio,  $\epsilon_{ij} \sim N(0, \sigma^2)$  e independientes
- $\tau_i$  es el efecto del grupo  $i$ -ésimo en la media de la variable respuesta  $Y$ . Esto es, cuánto aumenta o disminuye la media de  $Y$  por pertenecer la observación a la categoría  $i$ . De modo que podemos llamar

$$Y_i = \mu + \tau_i$$

al efecto medio del grupo  $i$ -ésimo.

La suma de las diferencias al cuadrado de cada dato respecto a la media general se calcula como sigue:

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$$

donde  $\bar{Y}_{..}$  es la media general de todas las observaciones.

Teniendo en cuenta que:  $Y_{ij} - \bar{Y}_{..} = Y_i + \epsilon_{ij} - \bar{Y}_{..}$

Podemos descomponer la suma de cuadrados, como sigue:

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \underbrace{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2}_{SSB} + \underbrace{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}_{SSW}$$

¿Qué son SSB y SSW? ¿Qué interpretación le dais?

- SSB: La varianza entre grupos se calcula como la suma de las diferencias al cuadrado de las medias de los grupos respecto a la media general, ponderada por el tamaño de los grupos:

$$SSB = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2$$

donde  $\bar{Y}_i$  es la media del grupo  $i$ .

- SSW: La varianza dentro de los grupos es la suma de las diferencias al cuadrado de cada dato respecto a la media de su grupo

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$



Esto es, se descompone la variabilidad total de los datos en dos componentes, SSB que refleja la diferencia de cada grupo respecto a la media global y SSW que refleja la variabilidad intrínseca dentro de cada grupo:

$$SST = SSB + SSW$$

Nota:

- SST: Sum of squares total
- SSB: Sum of squares between
- SSW: Sum of squares within

## Cálculo del Estadístico F:

$$F = \frac{\text{Varianza Entre Grupos (MSB)}}{\text{Varianza Dentro de los Grupos (MSW)}}$$

Donde:

- MSB (Mean Square Between): Media cuadrática entre grupos.
- MSW (Mean Square Within): Media cuadrática dentro de los grupos.

Esto es:

$$F = \frac{SSB/df_B}{SSW/df_W}$$

siendo  $df_B = k - 1$  son los grados de libertad entre los grupos y  $df_W = N - k$  son los grados de libertad dentro de los grupos y  $N$  es el número total de observaciones.

Una vez se dispone de toda esta información, es común representarla en forma de tabla, en la llamada *Tabla ANOVA*:

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado Medio
Disferencias entre grupos	SSB	$k-1$	MSB
Diferencias dentro de los grupos, Residual o Error	SSW	$N-k$	MSW
Total	SST	$N-1$	

El estadístico de prueba  $F \sim F_{df_B, df_W}$  bajo la hipótesis nula de igualdad de medias.

El  $p - valor$  se obtiene a partir de la distribución  $F$ , considerando los grados de libertad de los numeradores y denominadores. Esto es:

$$p - valor = P(F_{df_b, df_W} > F_{muestral})$$

Como en otros contrastes, si el  $p - valor$  es menor que el nivel de significancia  $\alpha$ , se rechaza la hipótesis nula, concluyendo que al menos una de las medias de los grupos es diferente.

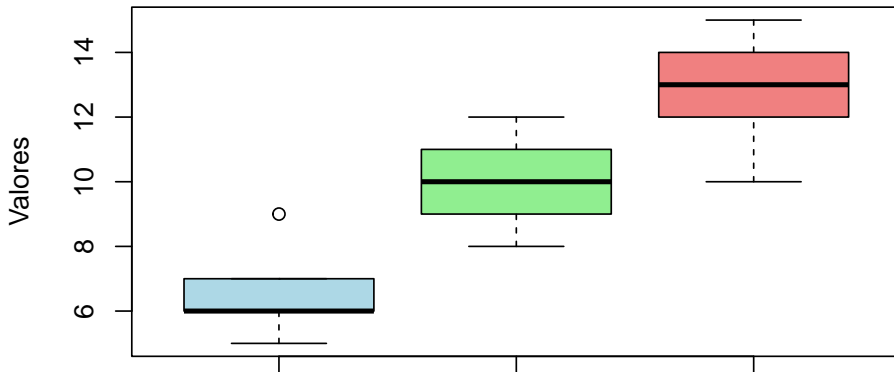
Supongamos que tenemos tres tratamientos (A, B y C) y sus correspondientes muestras de datos son:

- Grupo A: [5, 7, 6, 9, 6]
- Grupo B: [8, 12, 9, 11, 10]
- Grupo C: [14, 10, 13, 15, 12]

Nuestro objetivo es determinar si existe una diferencia significativa entre las medias de estos tres grupos.

Hagamos un boxplot con los tres tratamientos:

**Boxplot de los Grupos A, B y C**



Comenzamos calculando la media de cada grupo y la media general:

- Media de Grupo A  $\bar{Y}_A = \frac{5+7+6+9+6}{5} = \frac{33}{5} = 6.6$
- Media de Grupo B  $\bar{Y}_B = \frac{8+12+9+11+10}{5} = \frac{50}{5} = 10$
- Media de Grupo C  $\bar{Y}_C = \frac{14+10+13+15+12}{5} = \frac{64}{5} = 12.8$
- Media General  $\bar{Y}$ :

$$\bar{Y} = \frac{6.6 + 10.0 + 12.8}{3} = \frac{29.4}{3} = 9.8$$

Calculemos ahora cada uno de los componentes de  $SST = SSB + SSW$

Comencemos por SSB:

$$SSB = n_A(\bar{Y}_A - \bar{Y})^2 + n_B(\bar{Y}_B - \bar{Y})^2 + n_C(\bar{Y}_C - \bar{Y})^2$$

siendo  $n_A = n_B = n_C = 5$  (número de observaciones en cada grupo).

$$SSB = 5(6.6 - 9.8)^2 + 5(10.0 - 9.8)^2 + 5(12.8 - 9.8)^2 = 96.4$$



Calculemos a continuación la suma de los cuadrados dentro de los grupos:

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

Para cada grupo, calculamos la suma de las diferencias al cuadrado entre cada dato y la media del grupo:

- Grupo A:  $(5 - 6.6)^2 + (7 - 6.6)^2 + (6 - 6.6)^2 + (9 - 6.6)^2 + (6 - 6.6)^2 = 9.2$
- Grupo B:  $(8 - 10)^2 + (12 - 10)^2 + (9 - 10)^2 + (11 - 10)^2 + (10 - 10)^2 = 10$
- Grupo C:  
 $(14 - 12.8)^2 + (10 - 12.8)^2 + (13 - 12.8)^2 + (15 - 12.8)^2 + (12 - 12.8)^2 = 14.8$

Con ello,

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = 9.2 + 10.0 + 14.8 = 34.0$$

Por tanto que la Suma Total de Cuadrados (SST) es:

$$SST = SSB + SSW = 96.4 + 34.0 = 130.4$$

Para realizar el contraste, calculamos el estadístico  $F$

$$F = \frac{MSB}{MSW} = \frac{SSB/df_B}{SSW/df_W}$$

Los grados de libertad son:

- Grados de libertad entre los grupos:  $df_B = k - 1 = 3 - 1 = 2$
- Grados de libertad dentro de los grupos:  $df_W = N - k = 15 - 3 = 12$

Luego, el valor del estadístico del contraste es

$$F = \frac{MSB}{MSW} = \frac{SSB/df_B}{SSW/df_W} = \frac{96.4/2}{34/12} = \frac{48.2}{2.83} = 17.01$$

El p-valor se obtiene utilizando la distribución  $F$ -Snedecor con  $df_B = 2$  y  $df_W = 12$ .

El comando en R es `pf(17.01, df1 = 2, df2 = 12, lower.tail = FALSE)` que devuelve un p-valor de 0.00031. Como es menor que el grado de significancia  $\alpha = 0.05$ , indica una diferencia significativa entre los grupos.

```
grupo_a <- c(5, 7, 6, 9, 6)
grupo_b <- c(8, 12, 9, 11, 10)
grupo_c <- c(14, 10, 13, 15, 12)

datos <- data.frame(
  valores = c(grupo_a, grupo_b, grupo_c),
  grupo = factor(rep(c("Grupo A", "Grupo B", "Grupo C"), each = 5)))

anova = aov(datos$valores ~ datos$grupo)
summary(anova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
datos\$grupo	2	96.4	48.20	17.01	0.000314 ***
Residuals	12	34.0	2.83		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

- Se utiliza cuando se estudian dos factores simultáneamente para evaluar su efecto individual y conjunto en una variable dependiente
- Puede verse como una generalización del caso de ANOVA con un único factor
- Este modelo es más complejo y permite entender no solo los efectos principales de cada factor, sino también si hay una interacción entre ellos

Sean  $A$  y  $B$  dos factores que se desean estudiar, con  $m_A$  y  $m_B$  niveles. Trabajaremos con las siguientes hipótesis nulas:

## Opciones de hipótesis:

- Hipótesis nula para los efectos principales  $H_0$ :
  - No hay efecto del primer factor
  - No hay efecto del segundo factor
- Hipótesis nula para la interacción  $H_0$ : No hay interacción entre los dos factores

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

Donde:

- $Y_{ijk}$  es la observación  $k$ -ésima del nivel  $j$ -ésimo del factor  $B$  y nivel  $i$ -ésimo del factor  $A$
- $\mu$  es la media general
- $\alpha_i$  es el efecto del nivel  $i$ -ésimo del factor  $A$
- $\beta_j$  es el efecto del nivel  $j$ -ésimo del factor  $B$
- $\epsilon_{ijk}$  es el término de error aleatorio,  $\epsilon_{ijk} \sim N(0, \sigma^2)$  e independientes



En este caso, la tabla ANOVA queda como sigue:

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado Medio
Diferencias entre niveles del factor A	$SSB_A$	$m_A - 1$	$MSB_A$
Diferencias entre niveles del factor B	$SSB_B$	$m_B - 1$	$MSB_B$
Error	$SSW$	$N - m_A - m_B + 1$	$MSW$
Total	$SST$	$N - 1$	

Para estudiar la importancia de cada factor se calcula el estadístico  $F$  particular para cada uno de ellos como sigue:

$$F_A = \frac{MSB_A}{MSW} \sim F_{m_A-1, N-m_A-m_B+1}$$

y

$$F_B = \frac{MSB_B}{MSW} \sim F_{m_B-1, N-m_A-m_B+1}$$

A partir de estos estadísticos de prueba podemos contrastar las hipótesis nulas de no existencia de efectos asociados a los factores  $A$  y  $B$  respectivamente.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

Donde  $(\alpha\beta)_{ij}$  representa el efecto de interacción entre el nivel  $i$ -ésimo del factor  $A$  y el nivel  $j$ -ésimo del factor  $B$

En este caso, la tabla ANOVA añade el factor de interacción:

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado Medio
Diferencias entre niveles del factor A	$SSB_A$	$m_A - 1$	$MSB_A$
Diferencias entre niveles del factor B	$SSB_B$	$m_B - 1$	$MSB_B$
Diferencias debidas la interacción	$SSB_{AB}$	$(m_A - 1) * (m_B - 1)$	$MSB_{AB}$
Error	$SSW$	$N - m_A * m_B$	$MSW$
Total	$SST$	$N - 1$	

Para estudiar la importancia de cada de la interacción se calculan el estadístico  $F$  correspondiente:

$$F_{AB} = \frac{MSB_{AB}}{MSW} \sim F_{(m_A-1)*(m_B-1), N-m_A*m_B}$$

Con este estadístico, contrastamos la hipótesis nula de no existencia de interacción entre los dos factores  $A$ ,  $B$ :

- Si podemos rechazar esa hipótesis, es decir, si existe interacción entre los factores, entonces hemos terminado. Es decir, no podemos eliminar ningún factor del modelo.
- En cambio, si no rechazamos la hipótesis nula, es decir, si no existe interacción entre los factores, podemos eliminar dicho efecto (la interacción) del modelo y pasar a un modelo sin interacción (como el previo)

Supongamos que estamos estudiando el efecto de dos factores sobre el rendimiento de los estudiantes:

1. El método de enseñanza: tradicional y experimental
2. El tipo de material de estudio: Libro, vídeo y online

Queremos saber si estos factores, y su posible interacción, tienen un efecto significativo en el rendimiento.

- Con ANOVA podemos rechazar la siguiente hipótesis nula:

$H_0 : \mu_1 = \mu_2 \dots = \mu_k$ , siendo  $k$  el número de niveles en el factor

- ¿En qué niveles del factor se encuentran las principales diferencias? Es decir, qué hipótesis (una o varias) de las siguientes son rechazadas:

$$H_0 : \mu_1 = \mu_2$$

$$H_0 : \mu_1 = \mu_3$$

...

$$H_0 : \mu_{k-1} = \mu_k$$

- En un caso con  $k$  niveles en el factor, hay  $k * (k - 1)/2$  posibles contrastes de igualdad de medias
- Si realizamos todos esos contrastes, aumenta la probabilidad de cometer errores de tipo I (rechazar incorrectamente la hipótesis nula)
- Este fenómeno se conoce como el problema de las **comparaciones múltiples**



- Cuando realizamos una sola prueba de hipótesis, establecemos un nivel de significancia predeterminado (ejemplo  $\alpha = 0.05$ )
- Esto es, aceptamos una probabilidad de error de tipo I del 5%, es decir, hay un 5% de probabilidad de rechazar incorrectamente la hipótesis nula cuando es verdadera
- Cuando se realizan múltiples test de hipótesis, la probabilidad acumulada de cometer al menos un error de tipo I aumenta con cada prueba adicional
  - Ejemplo: Con 10 tests de hipótesis independientes, cada uno con un nivel de significancia de  $\alpha = 0.05$ , la probabilidad de cometer al menos un error de tipo I aumenta a más del 40% ( $1 - (1 - 0.05)^{10} \approx 0.40$ )

- Solución: Aplicar correcciones cuando se realizan comparaciones múltiples para controlar este aumento en el riesgo de error
- Existen varios métodos para controlar el problema de las pruebas múltiples:
  - Bonferroni, Holm-Bonferroni, LSD (Least Significant Differences), Tukey HSD (Honestly-significant-difference), entre otros
- Estos métodos controlan la tasa global de error de tipo I para todas las comparaciones realizadas, manteniendo un nivel de significancia general específico

- Este método es relativamente simple y conservador
- Idea: ajustar el nivel de significancia individual para cada prueba de hipótesis realizada. En lugar de utilizar un nivel de significancia estándar (por ejemplo,  $\alpha = 0.05$ ), se divide el nivel de significancia global deseado por el número total de pruebas realizadas ( $k$ ):

$$\alpha' = \frac{\alpha}{k}$$

Esta división produce un nivel de significancia más estricto ( $\alpha'$ ) para cada prueba individual, lo que ayuda a controlar el riesgo global de error de tipo I

- Se utiliza el nivel de significancia individual ajustado para cada prueba de hipótesis. Si el  $p$  – *valor* de una prueba es menor que el nivel de significancia ajustado, se rechaza la hipótesis nula de la prueba

- Fácil de entender e implementar
- Proporciona un control conservador sobre el error de tipo I en comparaciones múltiples
- Puede ser un método demasiado conservador en situaciones donde se realizan muchas comparaciones, lo que puede resultar en una pérdida de potencia estadística

Controlar el **error de tipo I** en comparaciones múltiples con un enfoque jerárquico (step-down procedure)

## Procedimiento:

1. Se calculan los p-valores de todas las comparaciones
2. Se ordenan de menor a mayor  $p_{(1)} \leq \dots \leq p_{(j)} \leq \dots \leq p_{(k)}$  asignando a cada uno un índice  $j$  que indica su posición en la lista empezando por el mayor
3. Cada  $p_{(j)}$  se compara con un umbral ajustado  $\alpha_{(j)} = \frac{\alpha}{k-j+1}$  siendo  $\alpha$  el nivel de significación deseado y  $m$  el número total de pruebas
4. Regla de rechazo: Comenzando por el p-valor más pequeño ( $p_{(1)}$  al estar ordenados) se compara, en orden, cada  $p_{(j)}$  con su  $\alpha_{(j)}$  correspondiente. Cuando se encuentre la primera comparación no significativa, es decir  $p_{(j)} > \alpha_{(j)}$ , se para y se asume que todas las comparaciones restantes son también no significativas

- En cada comparación consecutiva, se va corrigiendo el nivel de significación por el número de comparaciones restantes
- La idea clave es que el método es escalonado:
  - Mientras que las comparaciones sean significativas, se prosigue a la siguiente comparación
  - En cuanto una comparación resulte no significativa, se para y se asume que las siguientes son también no significativa

Controla la tasa de descubrimientos falsos (FDR, False Discovery Rate):

$$FDR = \frac{\text{nº of falsely rejected null hypotheses}}{\text{total nº of rejected null hypothesis}}$$

Se trata de un procedimiento step-up:

1. Se calculan los p-valores de todas las comparaciones
2. Se ordenan de menor a mayor  $p_{(1)} \leq \dots \leq p_{(j)} \leq \dots \leq p_{(k)}$  asignando a cada uno un índice  $j$  que indica su posición en la lista empezando por el menor
3. Cada  $p_{(j)}$  se compara con un umbral ajustado  $\alpha_{(j)} = \frac{j}{k}\alpha$  siendo  $\alpha$  el nivel de significación deseado y  $m$  el número total de pruebas
4. Regla de rechazo: Comenzando por el p-valor mayor ( $p_{(k)}$  al estar ordenados) se compara, en orden, cada  $p_{(j)}$  con su  $\alpha_{(j)}$  correspondiente. Cuando se encuentre el mayor  $j$  tal que  $p_{(j)} \leq \alpha_{(j)}$ , es decir, la comparación es significativa, se para y se asume que todas las comparaciones restantes son también significativas

**Table 10.7** Critical values for  $p$  based on variations on Bonferroni (\* indicates that a comparison is significant)

		<i>Bonferroni</i>			<i>Holm</i>		<i>Benjamini-Hochberg</i>		
	$p$	$p_{\text{crit}} = \frac{\alpha}{k}$		$j$	$p_{\text{crit}} = \frac{\alpha}{j}$		$j$		$p_{\text{crit}} = \left(\frac{j}{k}\right)\alpha$
NT-Super	.0000	.0083	*	6	.0083	*	1	.0083	*
Super-Hulk	.0014	.0083	*	5	.0100	*	2	.0167	*
Spider-Super	.0127	.0083		4	.0125		3	.0250	*
NT-Spider	.0252	.0083		3	.0167		4	.0333	*
NT-Hulk	.1704	.0083		2	.0250		5	.0417	
Spider-Hulk	.3431	.0083		1	.0500		6	.0500	

**Figure 1:** Ejemplo del libro Field, A., Field, Z., & Miles, J. (2012). Discovering statistics using R.



- Prueba post-hoc para comparar las medias entre los pares de grupos tras un ANOVA significativo
- Mismas suposiciones que la ANOVA
- Contraste:
  - $H_0 : \mu_i = \mu_j$
  - $H_1 : \mu_i \neq \mu_j$

- **Idea principal:** calcular la diferencia honestamente significativa (*honestly significant difference HSD*) entre dos medias utilizando una distribución estadística definida por Student y denominada distribución  $Q$
- Esta distribución proporciona la distribución de muestreo exacta de la mayor diferencia entre un conjunto de medias procedentes de la misma población
- Todas las diferencias entre pares de medias se evalúan utilizando dicha distribución basada en la mayor diferencia

Bajo la hipótesis nula  $H_0$  el valor del estadístico  $Q$  que evalúa la diferencia entre los grupos  $i$  y  $j$  es

$$Q = \frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{\frac{MSW}{2} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

donde:

- $\bar{y}_i$  y  $\bar{y}_j$  son las medias muestrales de los grupos  $i$  y  $j$
- $MSW$  es el error cuadrático medio del ANOVA (también se denota  $MSE$ )
- $n_i$  y  $n_j$  es el tamaño muestral del grupo  $i$  y del  $j$

Se compara  $Q$  con un valor crítico de la distribución de Tukey:

$$Q_{cri} = q_{\alpha, k, df} \sqrt{\frac{MSE}{2} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

siendo  $q_{\alpha, k, df}$  el cuantil de la distribución del rango estudentiado para  $k$  grupos con  $df = N - k$  grados de libertad (siendo  $N$  el tamaño total de la muestra)

## Procedimiento:

- **Realizar un ANOVA:** Verifica si existe al menos una diferencia significativa entre las medias de los grupos
- **Calcular  $Q$**  para cada par de grupos
- Obtener  $Q_{cri}$  mediante tablas o con R
- Comparar  $Q$  con  $Q_{cri}$ :
  - Si  $Q > Q_{cri}$ , se rechaza  $H_0$  para ese par de medias
  - En caso contrario, no hay evidencia para rechazar  $H_0$

Gomez Villegas, M. A. (2005). *Inferencia estadística*. Ediciones Díaz de Santos.

Spiegel, M., & Stephens, L. (2009). Estadística–Serie Schaum. *Mc Graw-Hill*.

Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical statistics for data scientists: 50+ essential concepts using R and Python*. O'Reilly Media.

Canavos, G. C., & Medal, E. G. U. (1987). *Probabilidad y estadística* (p. 651). México: McGraw Hill.

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. *Encyclopedia of research design*, 3(1), 1-5.

Field, A., Field, Z., & Miles, J. (2012). *Discovering statistics using R*.