FACULTAD DE
PSICOLOGÍA Y LOGOPEDIA
Universidad de Málaga

Universidad de Valladolid

MÁSTER OFICIAL INVESTIGACIÓN EN ACTIVIDAD FÍSICA Y DEPORTE

Análisis de la Varianza I

Jairo Rodríguez Medina

Universidad de Valladolid

jairo.rodriguez.medina@uva.es

Málaga 4 y 5 de noviembre de 2026

Estructura de la sesión

I. Introducción

- 1.1 Algo de historia
- 1.2 Conceptos Clave

II. Fundamentos del análisis de la varianza

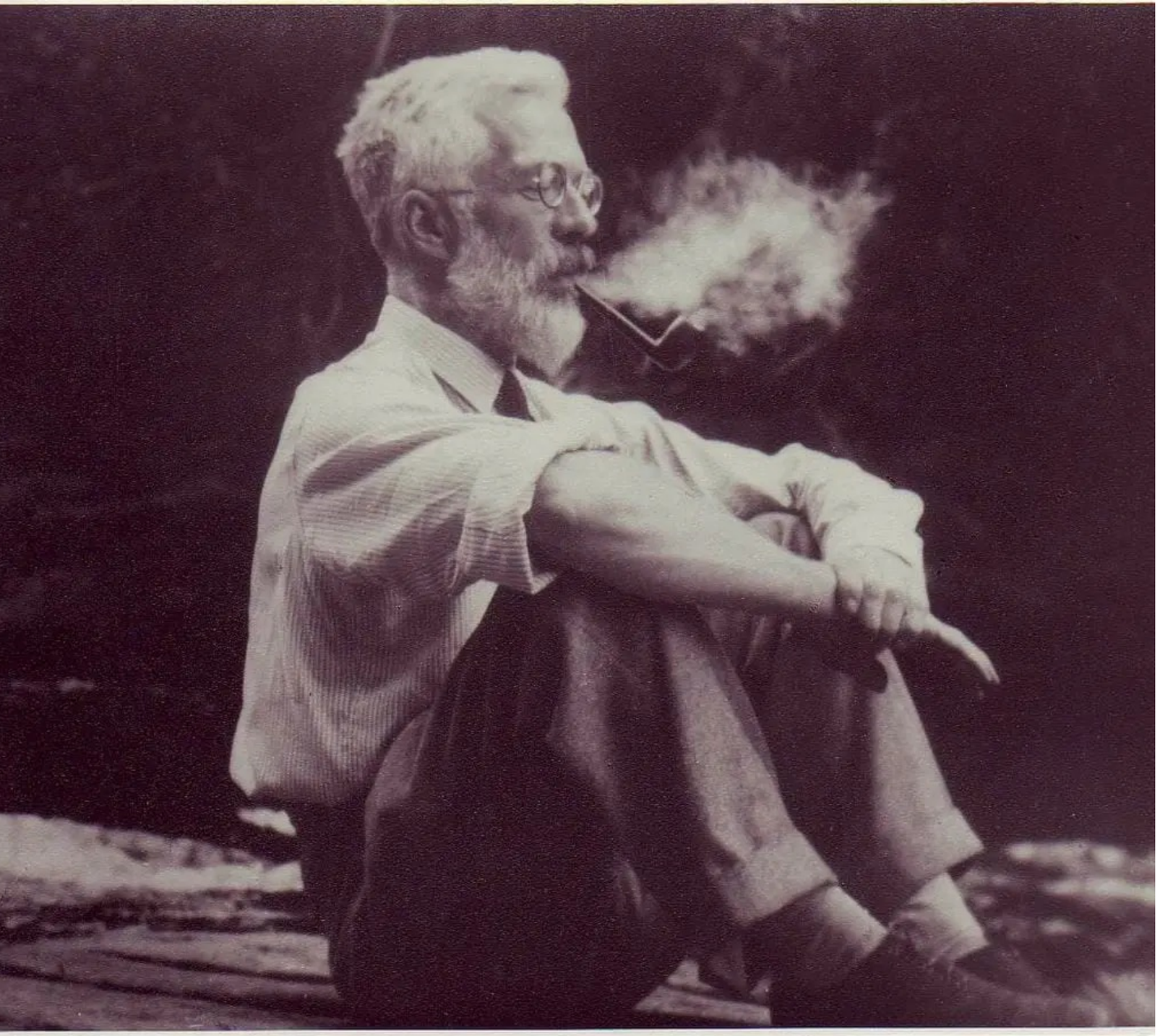
III. F ratio y Tamaño del Efecto

IV. Supuestos ANOVA

V. Pruebas post hoc

VI. Cálculo con Excel, R y SPSS

VII. Informe de Resultados



I. Introducción

- La estadística se preocupa por la variabilidad

- Inicialmente la técnica del análisis de la varianza se desarrolló para analizar las diferencias entre medias

El análisis de varianza es un método simple de ordenar datos aritméticos para aislar y mostrar las características esenciales de un conjunto de datos
Ronald Fisher (1934)





1. Algo de historia

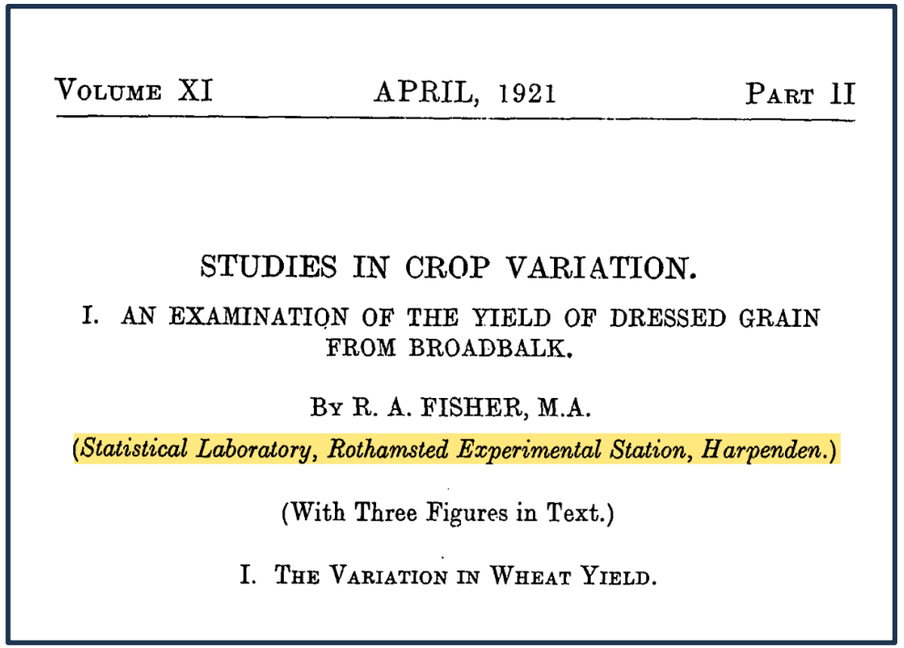
- El análisis de la varianza se remonta a los años 20

Sir Ronald Aylmer Fisher desarrolló el análisis de la varianza en la década de 1920

Utilizó el análisis de la varianza para estudiar el rendimiento de los cultivos en diferentes condiciones de cultivo (Prefirió este trabajo a colaborar con Pearson :)

Durante 90 años en este centro habían estado experimentando con fertilizantes usando un solo producto durante un año en toda la tierra

Rothamsted Agricultural Experimental Station



VOLUME XI APRIL, 1921 PART II

STUDIES IN CROP VARIATION.

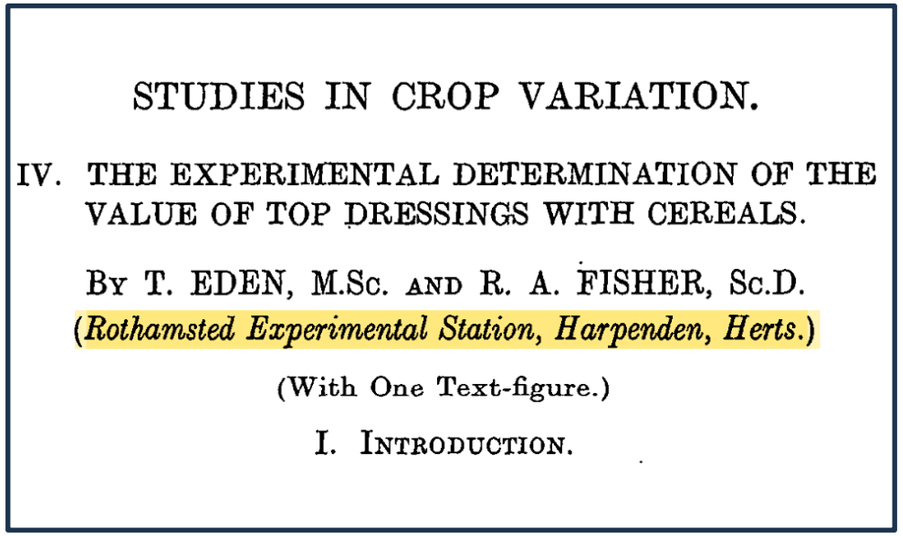
I. AN EXAMINATION OF THE YIELD OF DRESSED GRAIN FROM BROADBALK.

BY R. A. FISHER, M.A.

(*Statistical Laboratory, Rothamsted Experimental Station, Harpenden.*)

(With Three Figures in Text.)

I. THE VARIATION IN WHEAT YIELD.



STUDIES IN CROP VARIATION.

IV. THE EXPERIMENTAL DETERMINATION OF THE VALUE OF TOP DRESSINGS WITH CEREALS.

BY T. EDEN, M.Sc. AND R. A. FISHER, Sc.D.

(*Rothamsted Experimental Station, Harpenden, Herts.*)

(With One Text-figure.)

I. INTRODUCTION.

1. Algo de historia

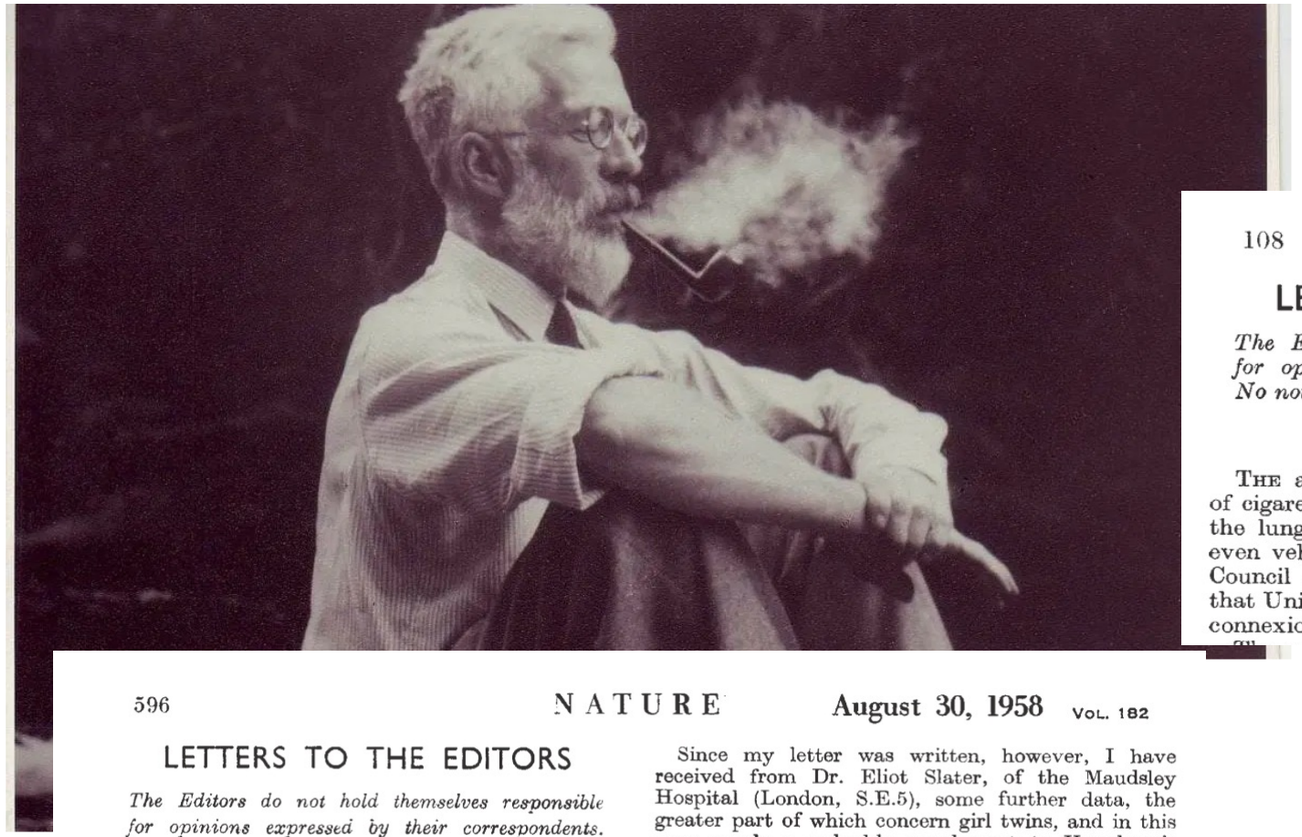
- El análisis de la varianza se remonta a los años 20

Lo que hizo Fisher fue revolucionar el modo en el que se experimentaba en agricultura

Comparó los efectos de varios fertilizantes simultáneamente

Fisher no solo inventó este tipo de diseño de investigación experimental, sino que también desarrolló la técnica estadística para analizar los datos obtenidos

el Análisis de la Varianza



LETTERS TO THE EDITORS

The Editors do not hold themselves responsible for opinions expressed by their correspondents. No notice is taken of anonymous communications.

Lung Cancer and Cigarettes?

THE association observable between the practice of cigarette smoking and the incidence of cancer of the lung, to which attention has been actively, or even vehemently, directed by the Medical Research Council Statistical Unit, has been interpreted, by that Unit, almost as though it demonstrated a causal connexion between these variables.

to minor variations in the smoking habit. In all cases, however, the monozygotic twins show closer similarity and fewer divergencies than the dizygotic.

There can therefore be little doubt that the genotype exercises a considerable influence on smoking, and on the particular habit of smoking adopted, and that a study of twins on a comparatively small scale is competent to demonstrate the rather considerable differences which must exist between the different groups who classify themselves as non-smokers, or the different classes of smokers. Such genotypically different groups would be expected to differ in cancer incidence; and their existence helps to explain such oddities as that pipe and cigar smokers should show much less lung cancer than

LETTERS TO THE EDITORS

The Editors do not hold themselves responsible for opinions expressed by their correspondents. No notice is taken of anonymous communications.

Cancer and Smoking

THE curious associations with lung cancer found in relation to smoking habits do not, in the minds of some of us, lend themselves easily to the simple conclusion that the products of combustion reaching the surface of the bronchus induce, though after a long interval, the development of a cancer. If, for example, it were possible to infer that smoking cigarettes is a cause of this disease, it would equally be possible to infer on exactly similar grounds that inhaling cigarette smoke was a practice of considerable prophylactic value in preventing the disease, for the practice of inhaling is rarer among patients with cancer of the lung than with others.

Such results suggest that an error has been made, of an old kind, in arguing from correlation to causation, and that the possibility should be explored that the different smoking classes, non-smokers, cigarette smokers, cigar smokers, pipe smokers, etc., have adopted their habits partly by reason of their personal temperaments and dispositions, and are not lightly

Since my letter was written, however, I have received from Dr. Eliot Slater, of the Maudsley Hospital (London, S.E.5), some further data, the greater part of which concern girl twins, and in this way supply a valuable supplement to Verschuer's data, and in which, moreover, a considerable number of pairs were separated at or shortly after birth.

For the resemblance in smoking habit, these female pairs give:

	Alike	Unlike	
Monozygotic	44	9	53
Dizygotic	9	9	18

So far, there is only a clear confirmation of the conclusion from the German data that the monozygotics are much more alike than the dizygotics in their smoking habits. The peculiar value of these data, however, lies in the subdivision of the monozygotic pairs into those separated at birth and those brought up together. These are:

	Alike	Unlike	
Separated	23	4	27
Not separated	21	5	26

Of the nine cases of unlike smoking habit, only four occur among the twenty-seven separated at birth. It would appear that the small proportion unlike among these 53 monozygotic pairs is not to be ascribed to mutual influence.

There is nothing to stop those who greatly desire



Las **curiosas asociaciones** con el **cáncer de pulmón** encontradas en relación con el hábito de fumar **no** se prestan fácilmente (…) a la conclusión de que los productos de la combustión (…) inducen el desarrollo de un cáncer

¿?¿¿?¿?¿?¿

Hay tres tipos de mentiras: mentiras, malditas mentiras y estadísticas

2. Conceptos Clave

Los Modelos de Análisis de la Varianza (modelos ANOVA) nos permiten comparar las medias de tres o más grupos

Por ejemplo:

Queremos comparar el efecto de diferentes tipos de entrenamiento en la fuerza de los atletas:

ANOVA DE UN FACTOR

Variable independiente (Tipo de entrenamiento) (es un Factor)

Grupo 1: Entrenamiento de fuerza tradicional

Grupo 2: Entrenamiento pliométrico

Grupo 3: Entrenamiento combinado (entrenamiento de fuerza y pliométrico)

Variable dependiente (Fuerza)

2. Conceptos Clave

Variable dependiente (Fuerza)

Refleja el resultado del estudio. A veces se denominan **variables resultado**

Variable independiente (Tipo de entrenamiento) (es un Factor)

Son las variables que suponemos que influyen en la variable dependiente

A veces se denominan:

Efectos de tratamiento, factores de tratamiento

Efectos de clasificación, factores de clasificación

Por tanto, esperamos que las variables independientes “causen” diferencias en las variables dependientes

2. Conceptos Clave

El elemento clave que debemos tener en cuenta cuando nos referimos a variables dependientes e independientes es la variación

Una variable que no varía es una constante :)

Ejemplo:

Si todos los participantes realizan el mismo tipo de entrenamiento, la variable tipo de entrenamiento pasa a ser una constante

En ese caso no podría ser una variable independiente en nuestro estudio

Lo mismo ocurre con la variable dependiente

¿Algún ejemplo en el que esto pueda ocurrir con la variable dependiente?

2. Conceptos Clave

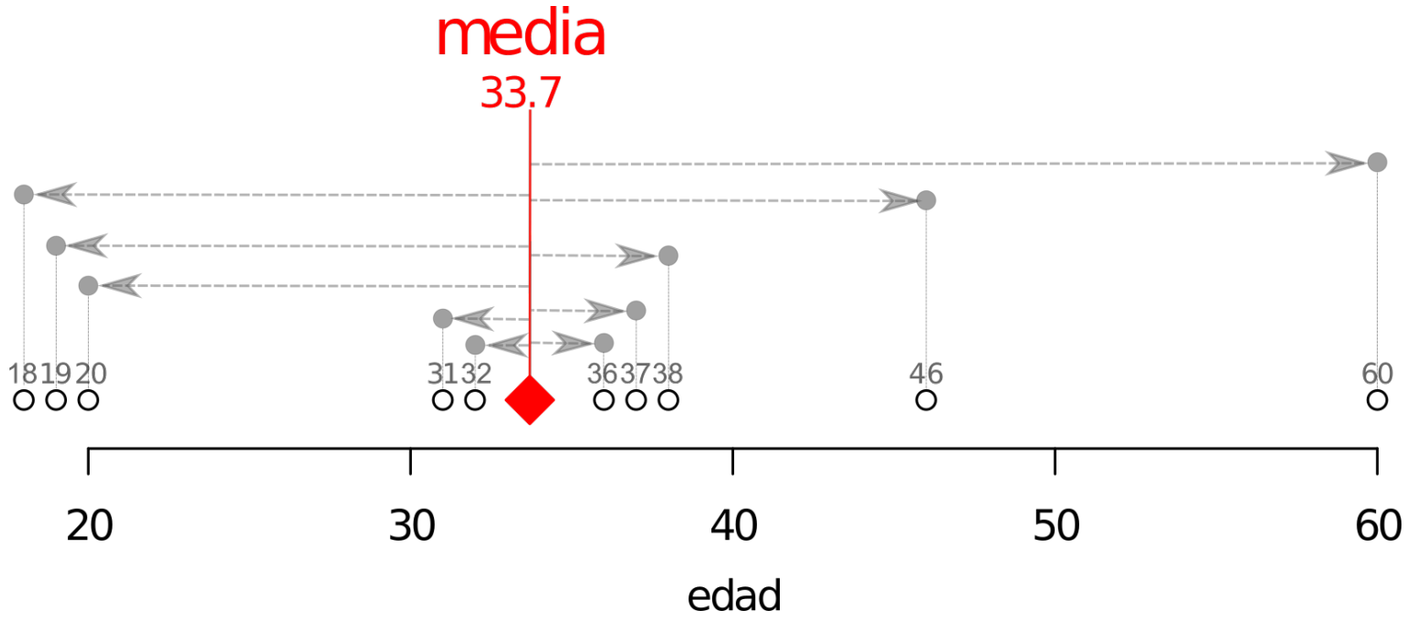
¿Qué es la varianza?

Es una medida de dispersión

Es la media de los cuadrados de las desviaciones a la media ☺

Responde a la pregunta:

¿en qué medida los datos están separados o agrupados en torno a la media?



Media	**32.7**
Mediana	**34**
Rango	**42**
Desviación Típica	**13.07**
Varianza	**170.9**

2. Conceptos Clave

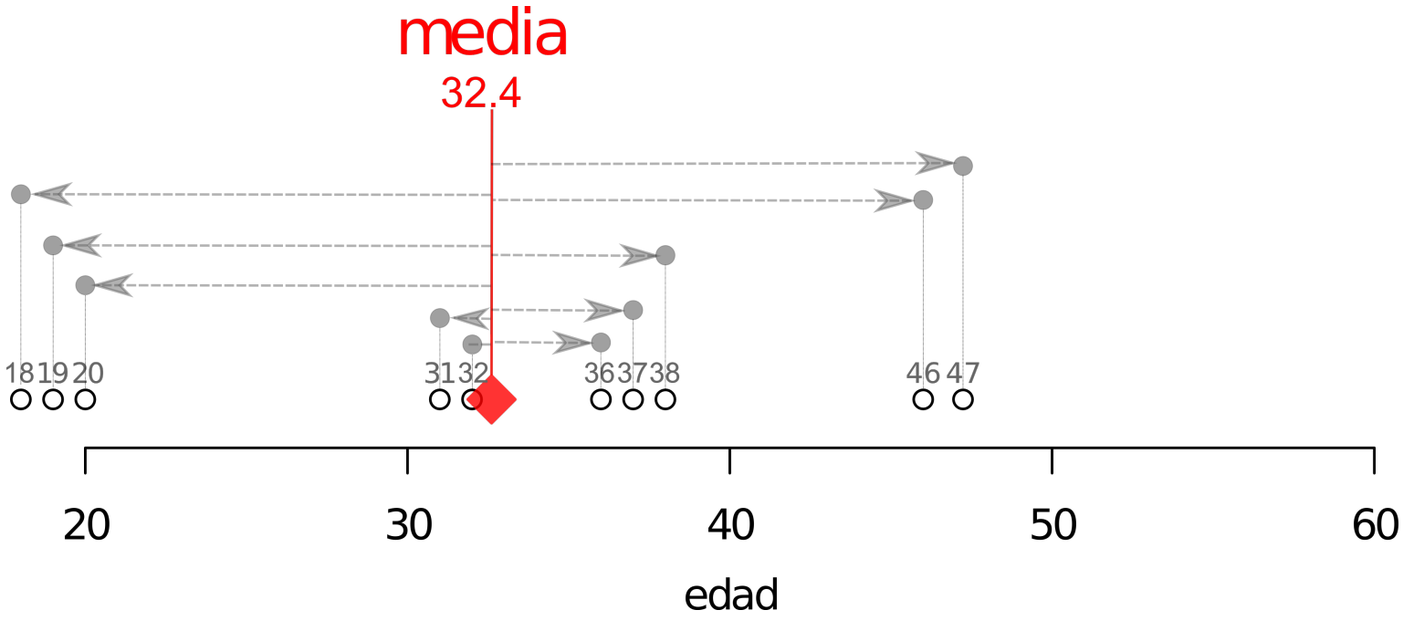
¿Qué es la varianza?

Es una medida de dispersión

Es la media de los cuadrados de las desviaciones a la media ☺

Responde a la pregunta:

¿en qué medida los datos están separados o agrupados en torno a la media?



Media	**32.4**
Mediana	**34**
Rango	**29**
Desviación Típica	**10.58**
Varianza	**111.82**

2. Conceptos Clave

Ejercicio 1.

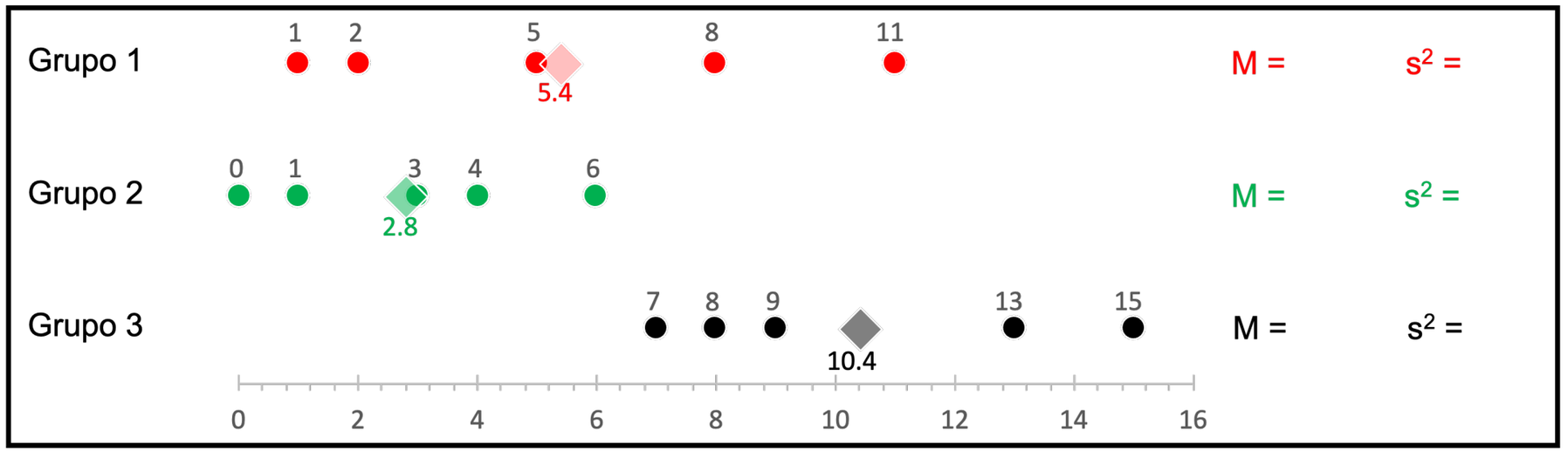
Calcular las varianzas de las siguientes variables: Y_G1; Y_G2; Y_G2

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

$$s^2 = \frac{\Sigma(Y_i - \overline{Y})^2}{n - 1}$$

→ Suma de Cuadrados

→ Grados de Libertad



ejercicios_ANOVA_I_Málaga

2. Conceptos Clave

Ejercicio 1.

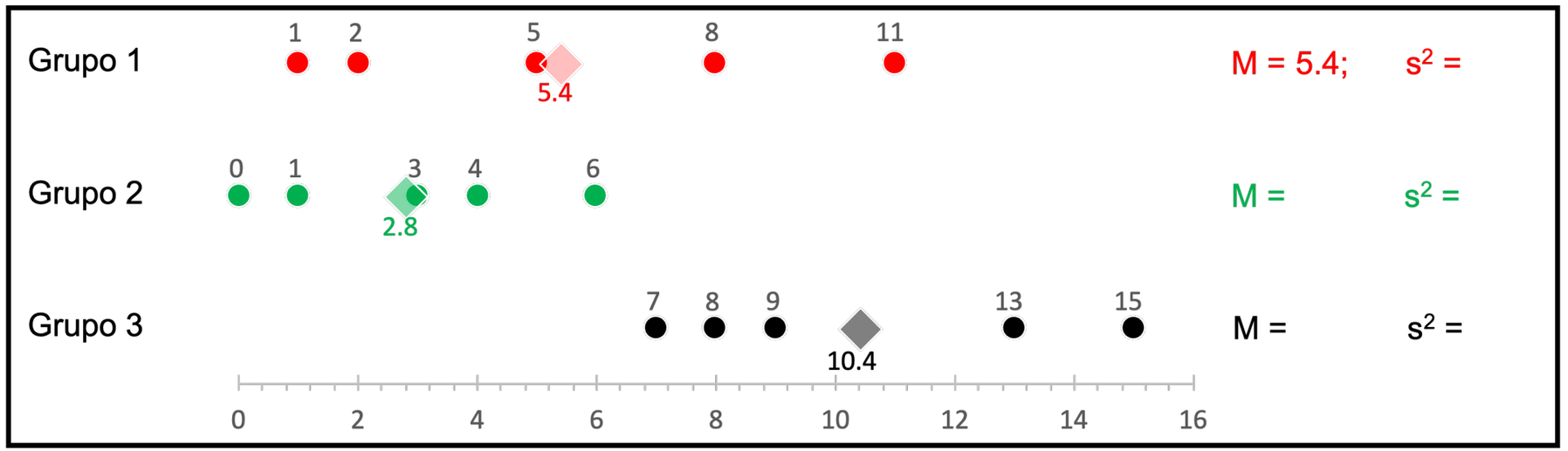
Calcular las varianzas de las siguientes variables: Y_G1; Y_G2; Y_G2

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

$$s^2 = \frac{\Sigma(Y_i - \overline{Y})^2}{n - 1}$$

→ Suma de Cuadrados

→ Grados de Libertad



ejercicios_ANOVA_I_Málaga

2. Conceptos Clave

Ejercicio 1.

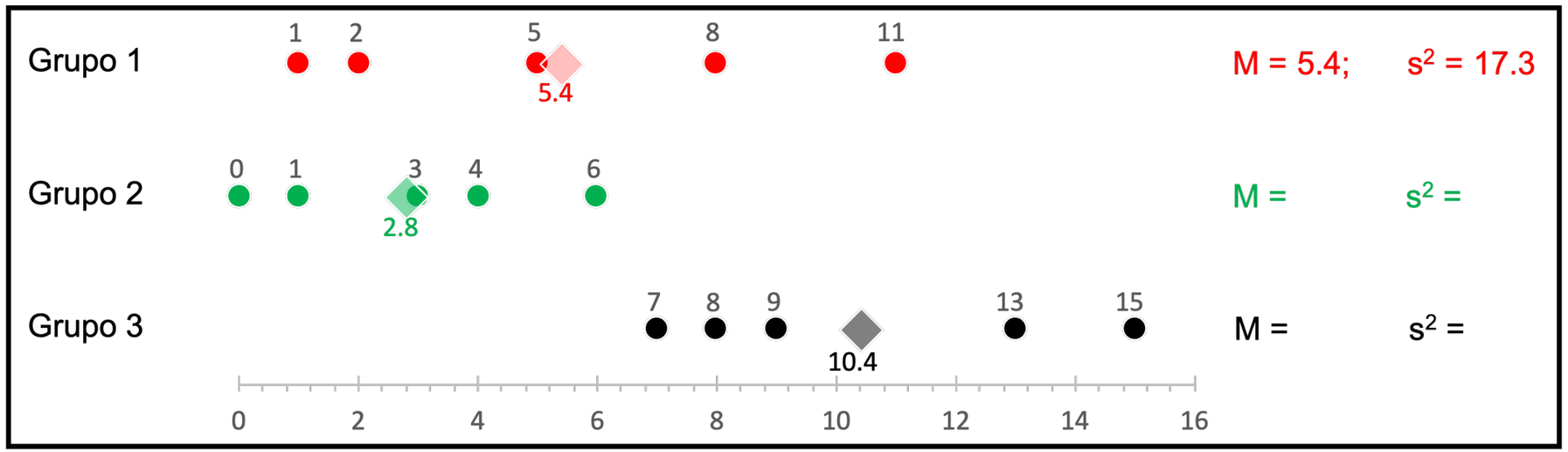
Calcular las varianzas de las siguientes variables: Y_G1; Y_G2; Y_G2

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

$$s^2 = \frac{\Sigma(Y_i - \overline{Y})^2}{n - 1}$$

→ Suma de Cuadrados

→ Grados de Libertad



ejercicios_ANOVA_I_Málaga

2. Conceptos Clave

Ejercicio 1.

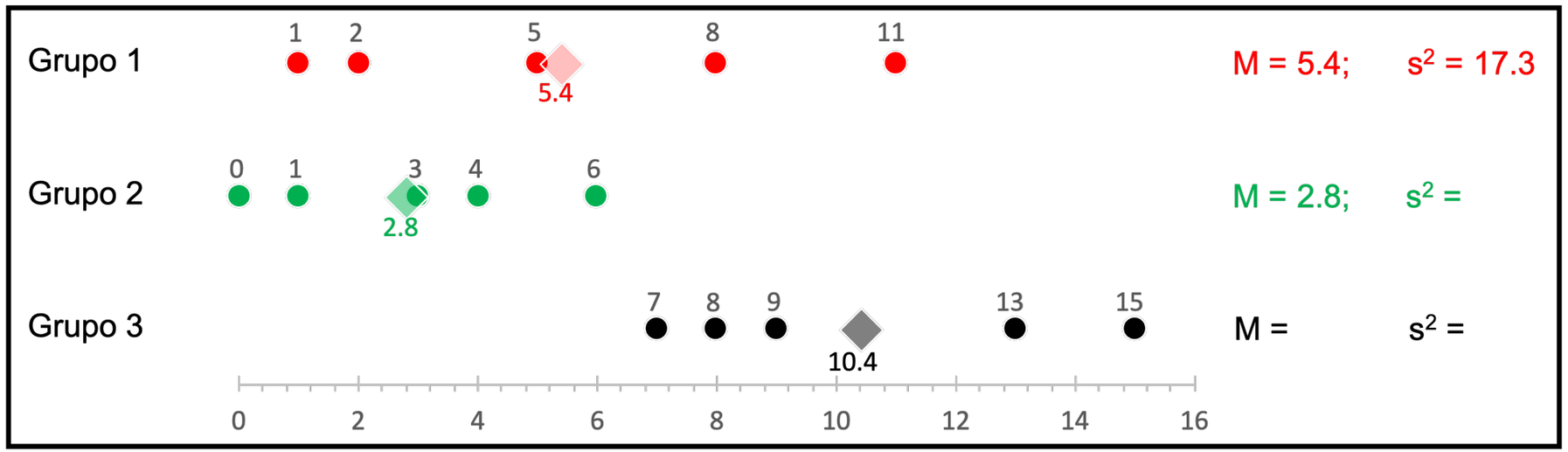
Calcular las varianzas de las siguientes variables: Y_G1; Y_G2; Y_G2

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

$$s^2 = \frac{\Sigma(Y_i - \overline{Y})^2}{n - 1}$$

→ Suma de Cuadrados

→ Grados de Libertad



ejercicios_ANOVA_I_Málaga

2. Conceptos Clave

Ejercicio 1.

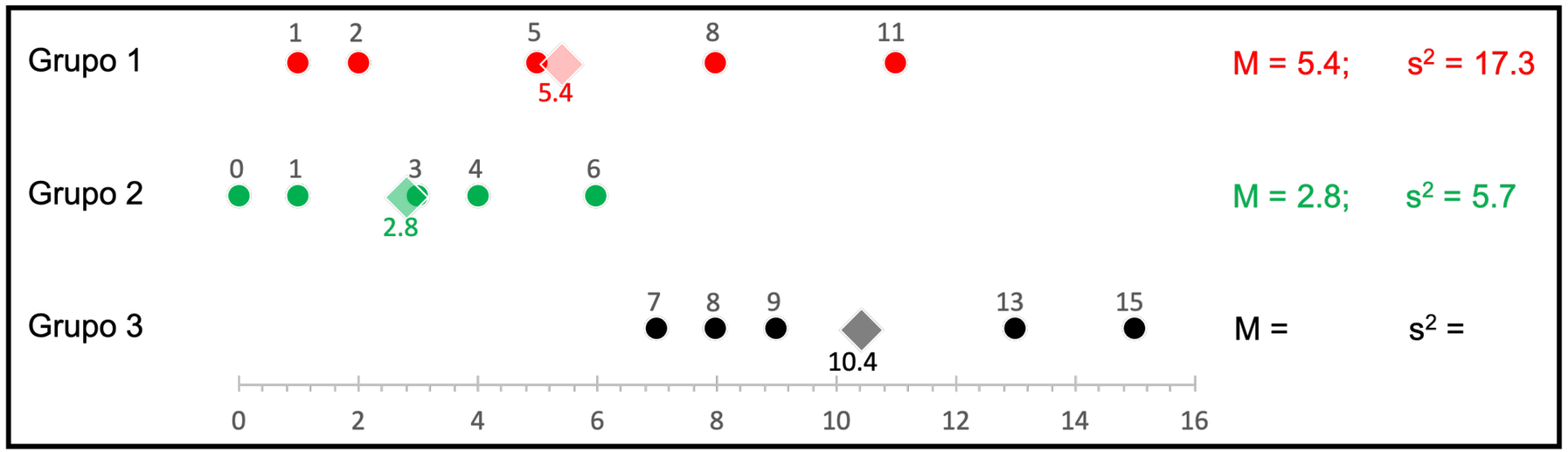
Calcular las varianzas de las siguientes variables: Y_G1; Y_G2; Y_G2

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

$$s^2 = \frac{\Sigma(Y_i - \overline{Y})^2}{n - 1}$$

→ Suma de Cuadrados

→ Grados de Libertad



ejercicios_ANOVA_I_Málaga

2. Conceptos Clave

Ejercicio 1.

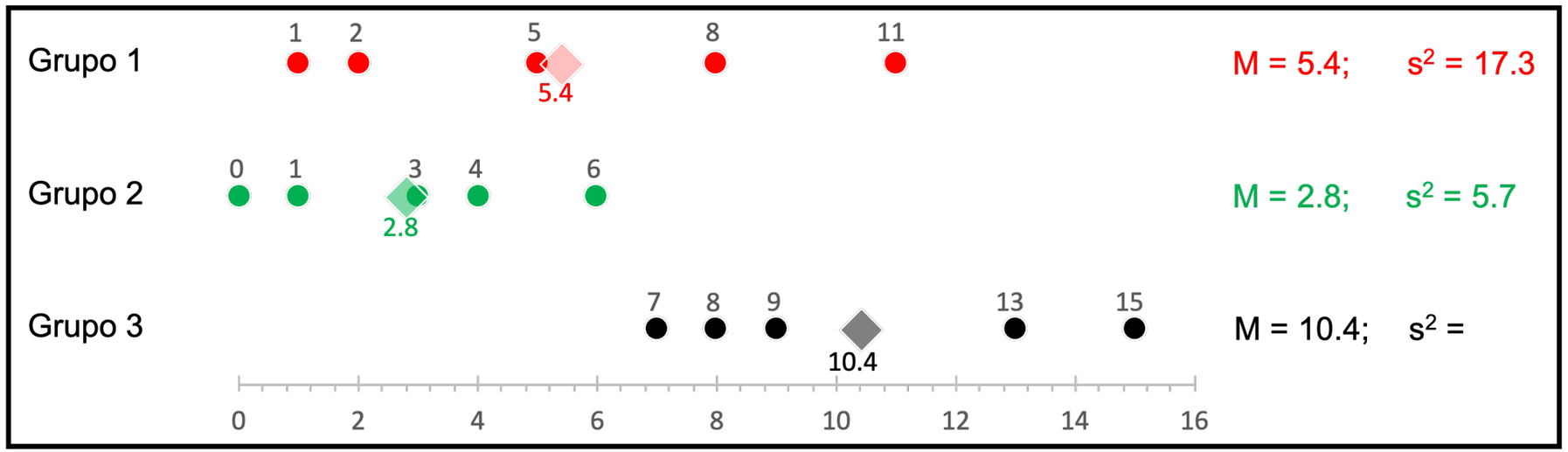
Calcular las varianzas de las siguientes variables: Y_G1; Y_G2; Y_G2

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

$$s^2 = \frac{\Sigma(Y_i - \overline{Y})^2}{n - 1}$$

→ Suma de Cuadrados

→ Grados de Libertad



ejercicios_ANOVA_I_Málaga

2. Conceptos Clave

Ejercicio 1.

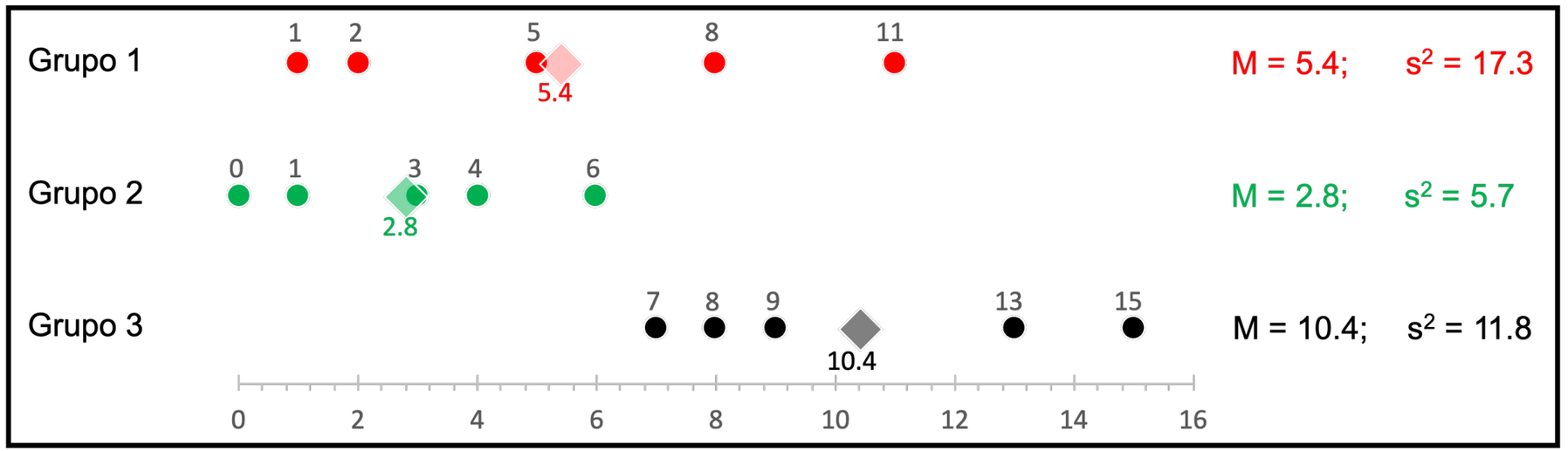
Calcular las varianzas de las siguientes variables: Y_G1; Y_G2; Y_G2

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

$$s^2 = \frac{\Sigma(Y_i - \overline{Y})^2}{n - 1}$$

→ Suma de Cuadrados

→ Grados de Libertad



ejercicios_ANOVA_I_Málaga

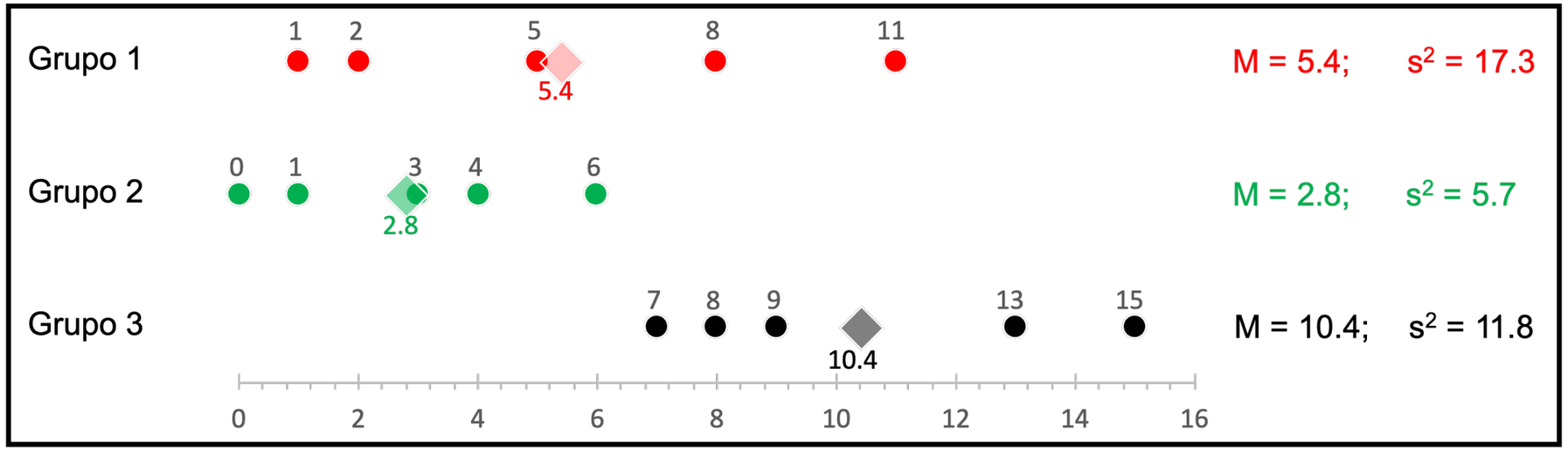
2. Conceptos Clave

Ejercicio 1.

Otra posible fórmula

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

$$s^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n-1}$$



ejercicios_ANOVA_I_Málaga

2. Conceptos Clave

- Son una simplificación de los modelos de regresión
- Son también una generalización a más de dos grupos de la prueba t de Student

Si queremos comparar las medias de tres grupos …

¿Por qué no hacer una prueba *t* para cada par de grupos?

Si usamos un nivel de confianza del 95 % (α = .05) para cada prueba de *t*

- La probabilidad de un error Tipo I (rechazar la hipótesis nula erróneamente) es del 5 %
- La probabilidad de no cometer este error en las tres pruebas sería:

.95 x .95 x .95 = 85.73 %

Por tanto, la probabilidad de cometer este error sería de:

100 – 85.73 = 14.3 %

2. Conceptos Clave

Factor

Un factor es una variable categórica medida en escala nominal u ordinal

Sus categorías se denominan niveles, condiciones o tratamientos

Factor

	Deporte					
	Atletismo			Ciclismo		
Factor →	Complemento Alimenticio			Complemento Alimenticio		
Factor → **Sexo**	A	B	C	A	B	C
Fem						
Mas						

En este caso tenemos **tres factores**: Sexo (con 2 niveles), Deporte (con 2 niveles) y Complemento Alimenticio (con 3 niveles)
2 x 2 x 3 = 12 celdas

2. Conceptos Clave

¿En qué medida los diferentes niveles de cada factor afectan a la variable dependiente (VD)?

VD: RENDIMIENTO

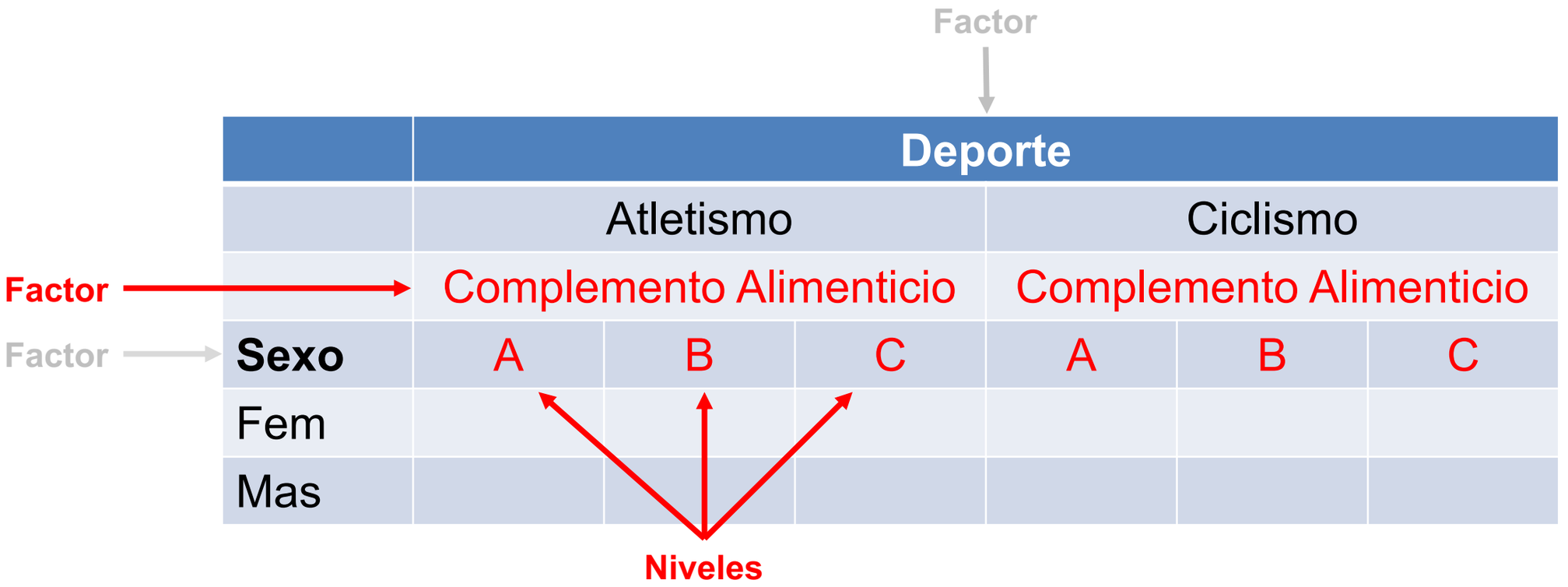
Factor

	Deporte					
	Atletismo			Ciclismo		
Factor →	Complemento Alimenticio			Complemento Alimenticio		
Factor → **Sexo**	A	B	C	A	B	C
Fem						
Mas						

2. Conceptos Clave

Podemos distinguir dos clases de factores

Factores de Tratamiento



2. Conceptos Clave

Podemos distinguir dos clases de factores

Factores de Clasificación

Factor

	Deporte					
	Atletismo ← Niveles →			Ciclismo		
Factor →	Complemento Alimenticio			Complemento Alimenticio		
Factor → **Sexo**	A	B	C	A	B	C
Niveles → Fem						
Mas						

2. Conceptos Clave

Como se ha señalado las variables toman diferentes valores

En los diseños ANOVA que veremos en este seminario las variables dependientes se miden en escalas cuantitativas

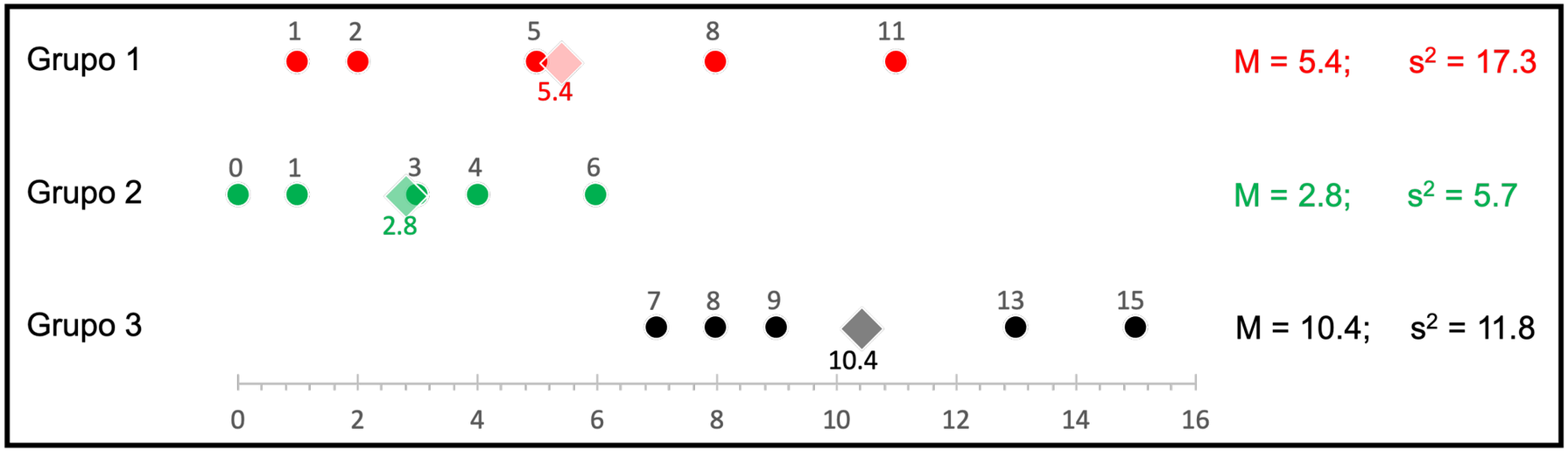
Las medidas de tendencia central proporcionan un indicador de la puntuación más típica de una distribución

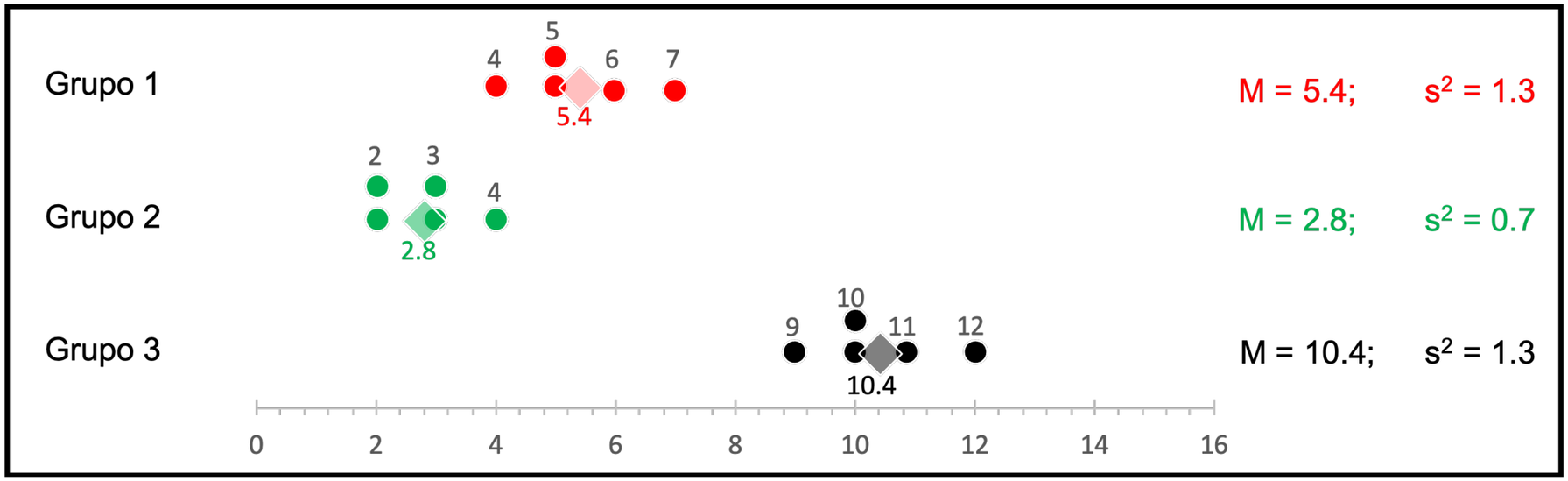
Usaremos la media como medida del impacto de la variable independiente en la variable dependiente

Hipotetizamos que la media de un grupo será diferente a la media de otros grupos

La variabilidad nos indica cómo varían las medidas o puntuaciones dentro de un grupo

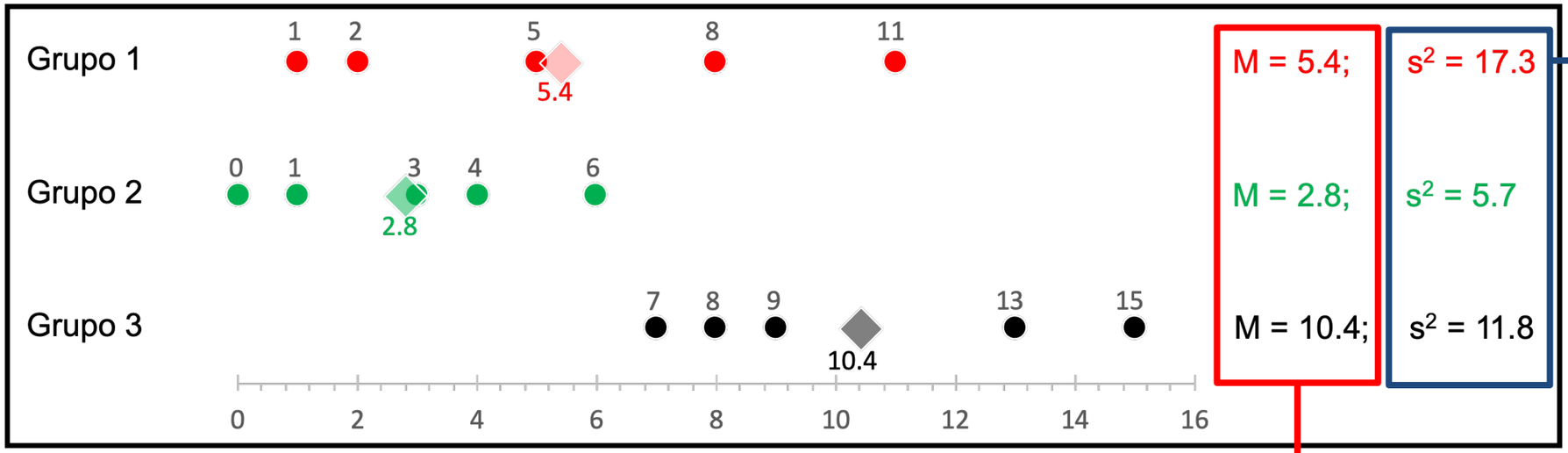
2. Conceptos Clave

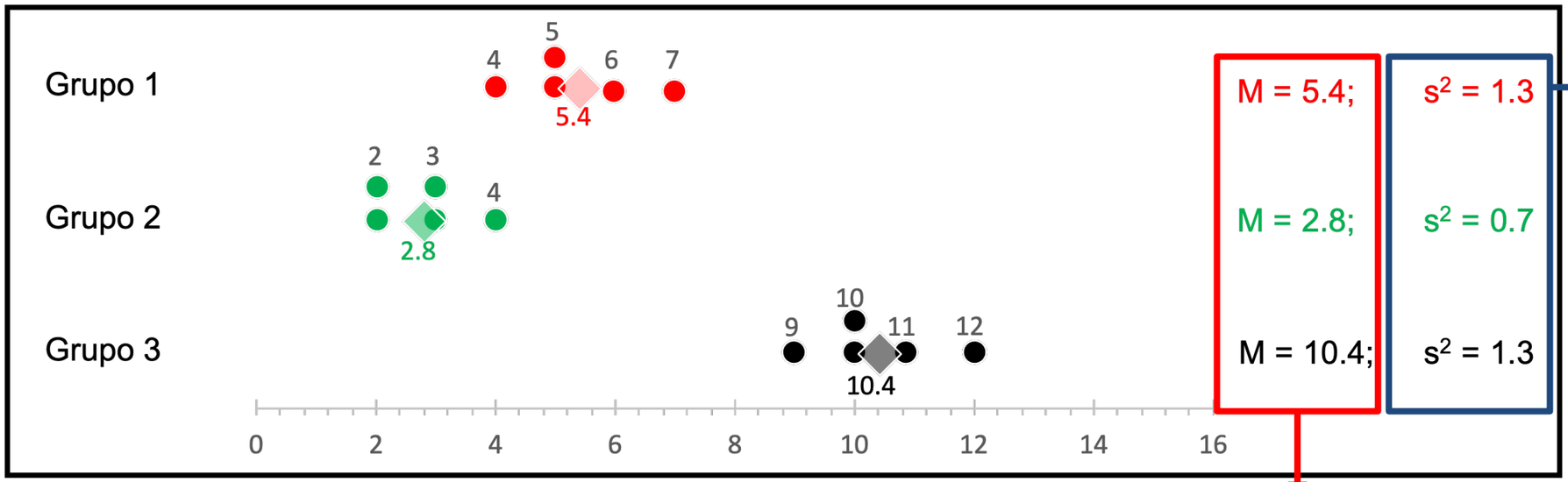




Analizamos las diferencias en las medias con respecto a la variabilidad dentro de cada grupo

2. Conceptos Clave





II. Fundamentos del Análisis de la varianza

- Analizar es entender los elementos individuales de un todo
- El ANOVA divide los componentes de la varianza de la variable dependiente
- Permite comparar las puntuaciones en diferentes grupos
- Podemos distinguir tres clases de diseños ANOVA:
  - Entre sujetos
  - Intra Sujetos o de Medidas Repetidas
  - Diseños Mixtos



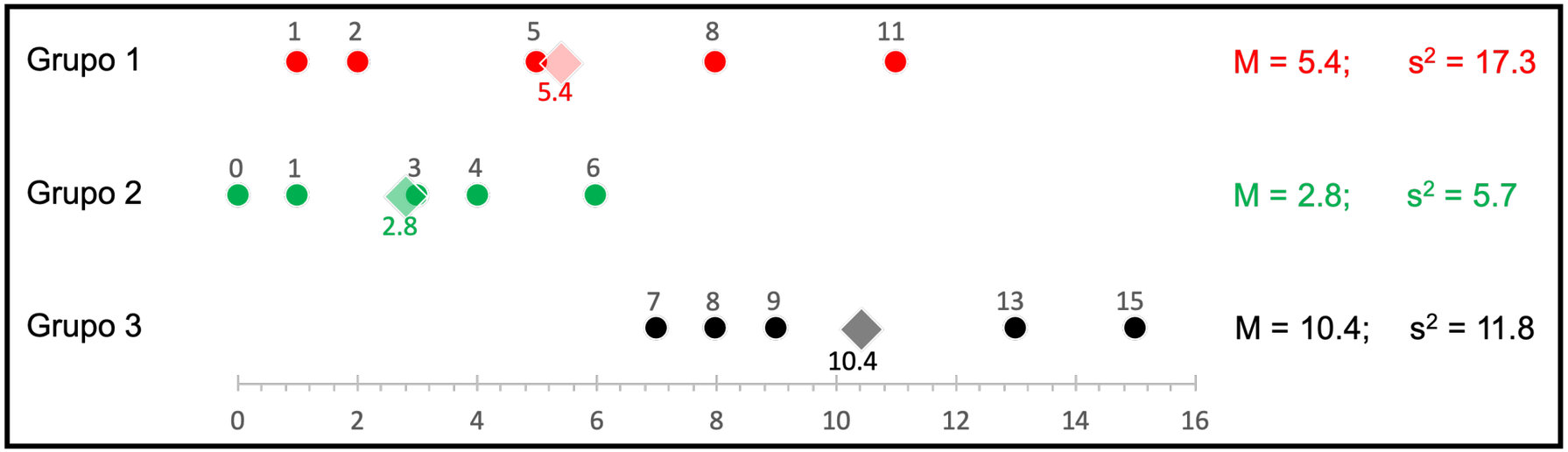
II. Fundamentos del Análisis de la Varianza

Un ejemplo básico

	CÁLCULO DE LA VARIANZA		
Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

- Disponemos de las puntuaciones obtenidas por 15 participantes en una determinada prueba
- Los participantes están divididos en 3 grupos en función del tipo de entrenamiento realizado:
  - Grupo 1: Entrenamiento de fuerza tradicional
  - Grupo 2: Entrenamiento pliométrico
  - Grupo 3: Entrenamiento combinado

¿Las diferencias entre las puntuaciones son significativas?



II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Suma de cuadrados de las desviaciones entre la media de cada grupo y la media general

Tabla ANOVA

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

Varianza entre grupos

Varianza intra grupos

Varianza Total

Suma de cuadrados de las desviaciones entre cada dato y la media general de su grupo

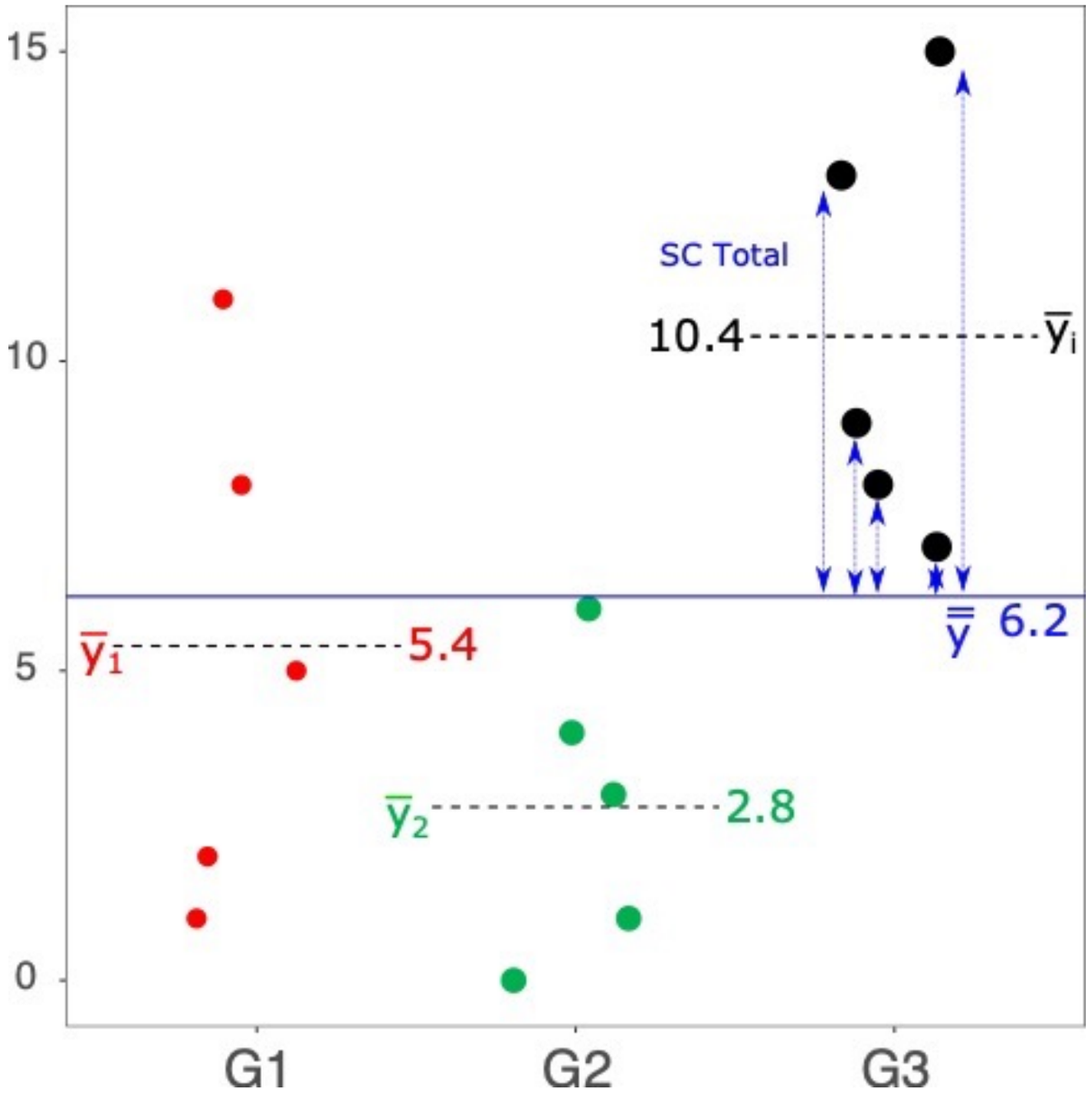
Lo veremos con un ejemplo gráfico

II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Suma de Cuadrados Total

La suma de los cuadrados de las desviaciones de cada punto respecto a la gran general

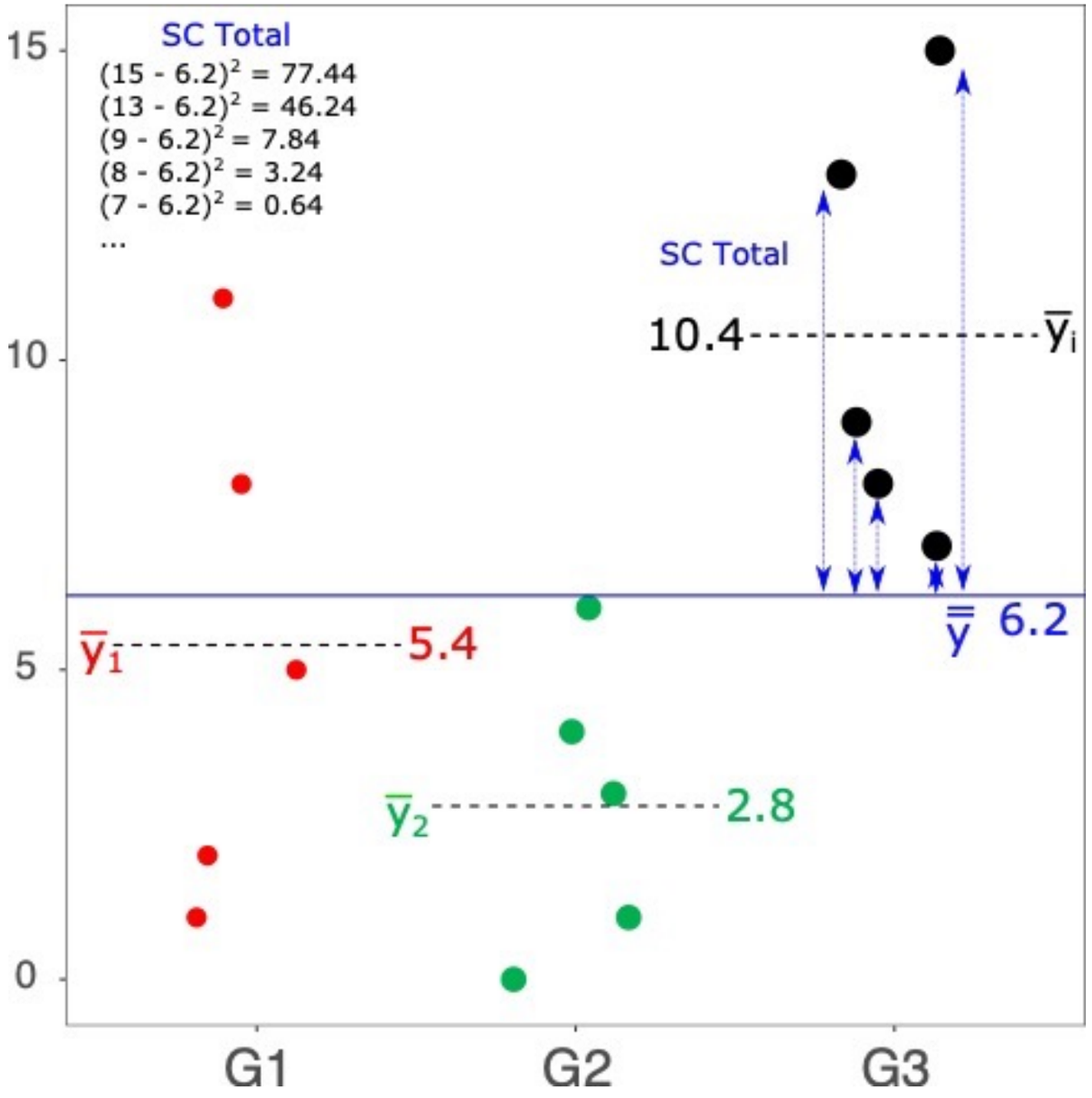


II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Suma de Cuadrados Total

La suma de los cuadrados de las desviaciones de cada punto respecto a la media general



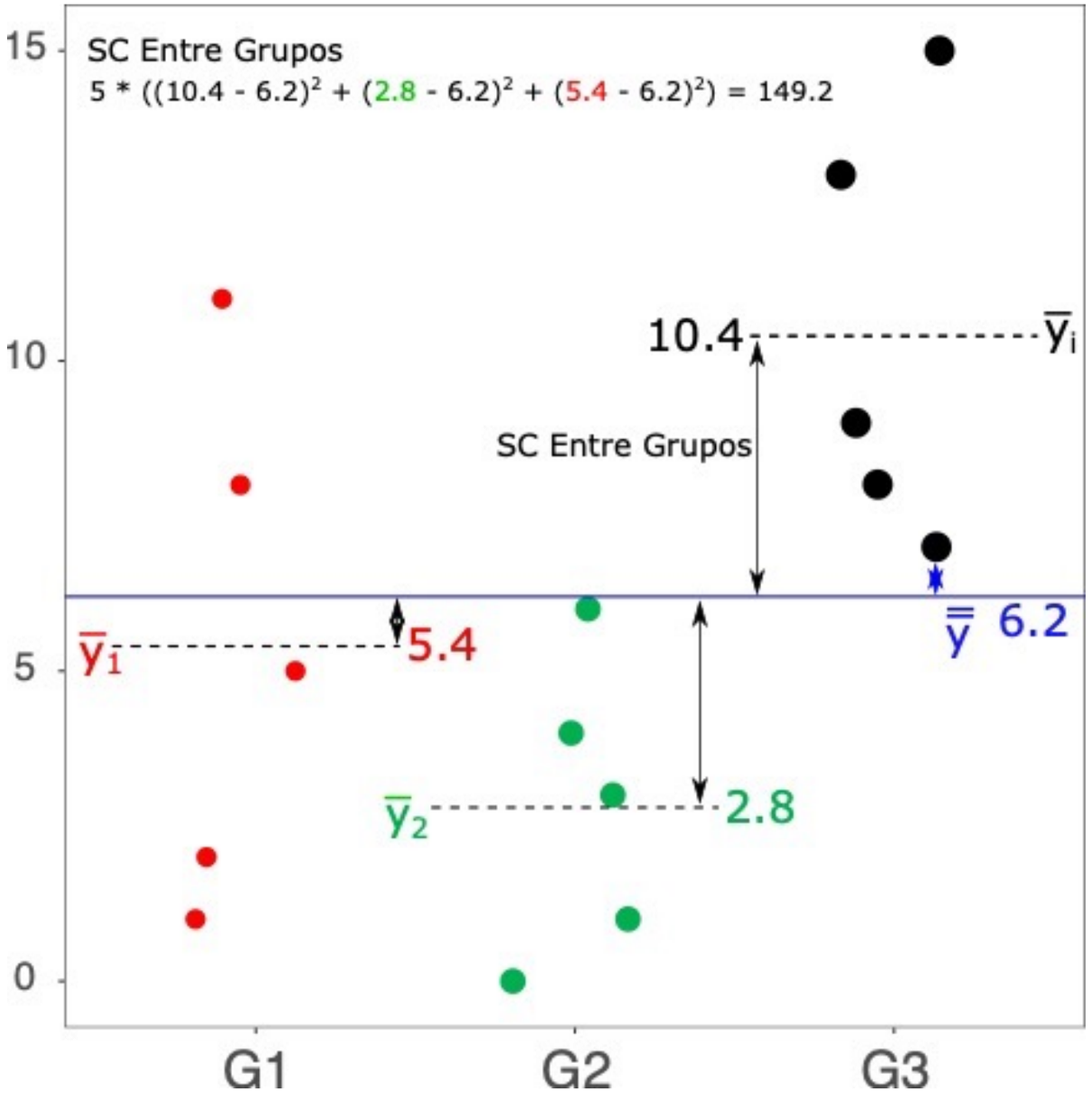
II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Suma de Cuadrados Entre Grupos

- **Suma de Cuadrados del Factor (SS_F)**
- **Suma de Cuadrados del Modelo (SS_M)**
- **Suma de Cuadrados del Tratamiento**
- **Variabilidad Explicada**

La suma de los cuadrados de las desviaciones entre la media de cada grupo y la media general



II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

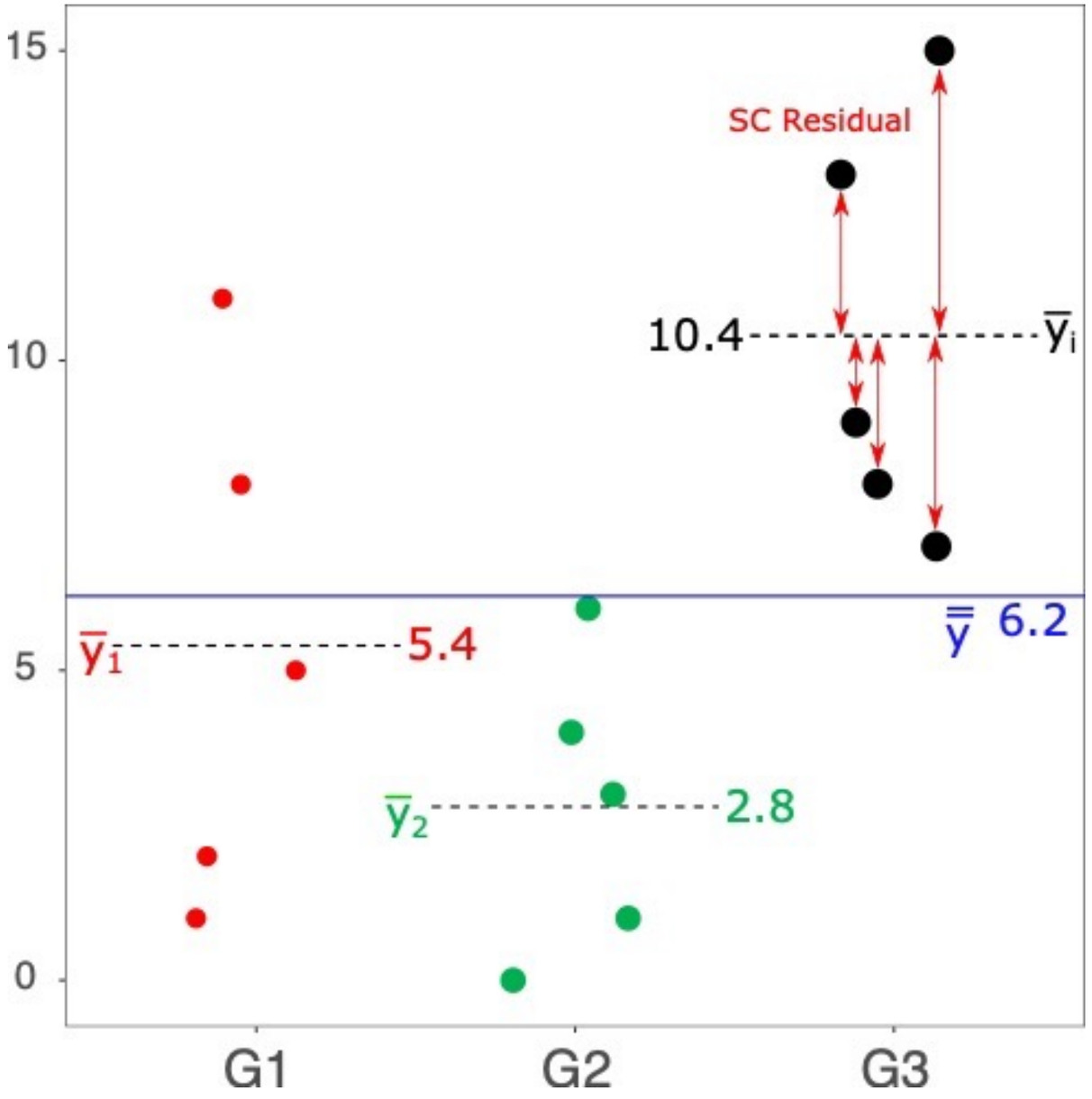
Suma de Cuadrados Intra Grupos

o Suma de Cuadrados Residual

o Suma de Cuadrados del Error

o Variabilidad Residual

La suma de los cuadrados de las desviaciones entre cada observación y la media de su grupo

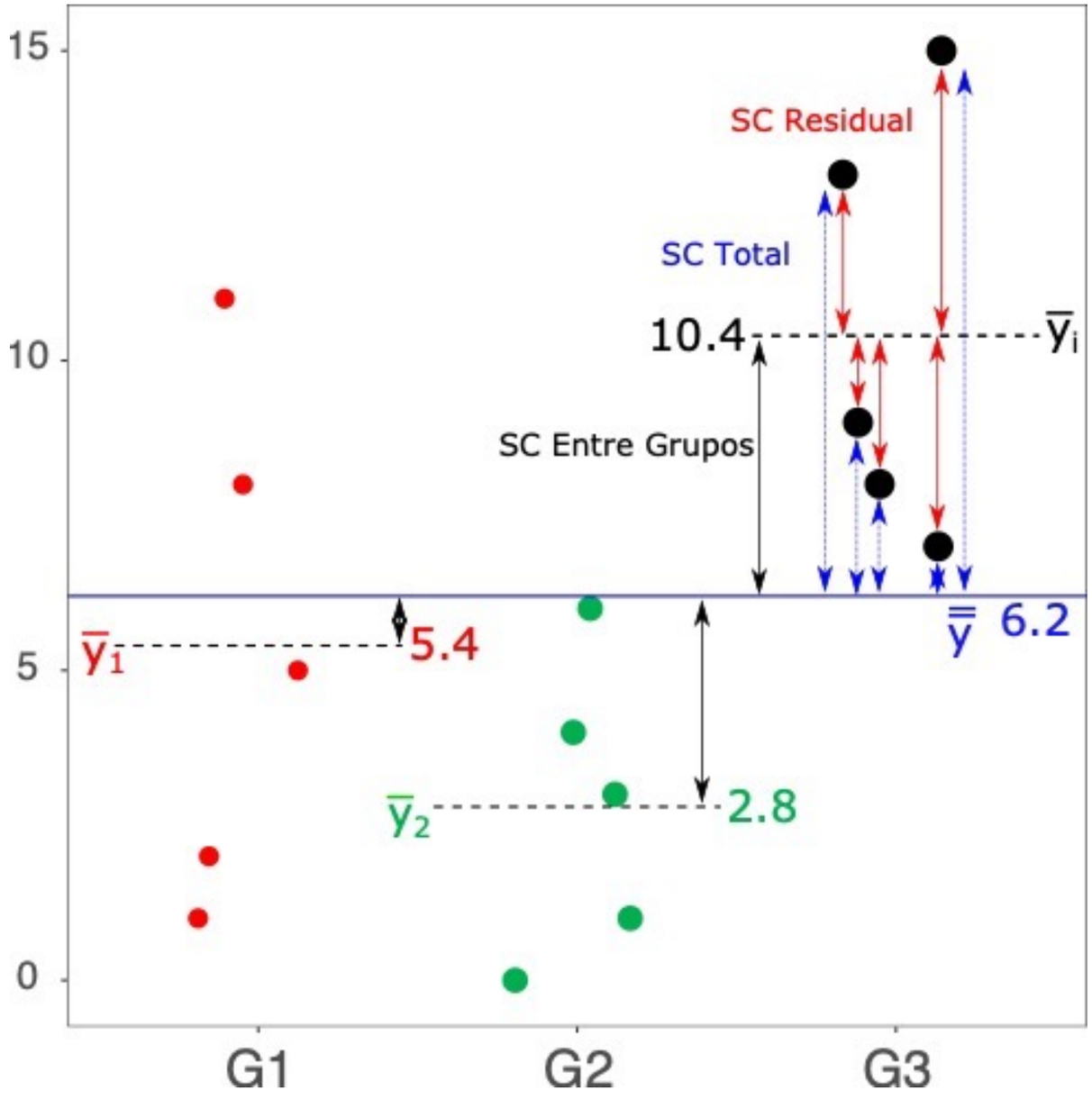


II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Descomposición de la variabilidad

Desviación = $\sum(observado\ -modelo)^2$



II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Ejercicio 2.

Hallar las sumas de cuadrados del siguiente ejemplo

Yi	**Y_G1**	**Y_G2**	**Y_G3**
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

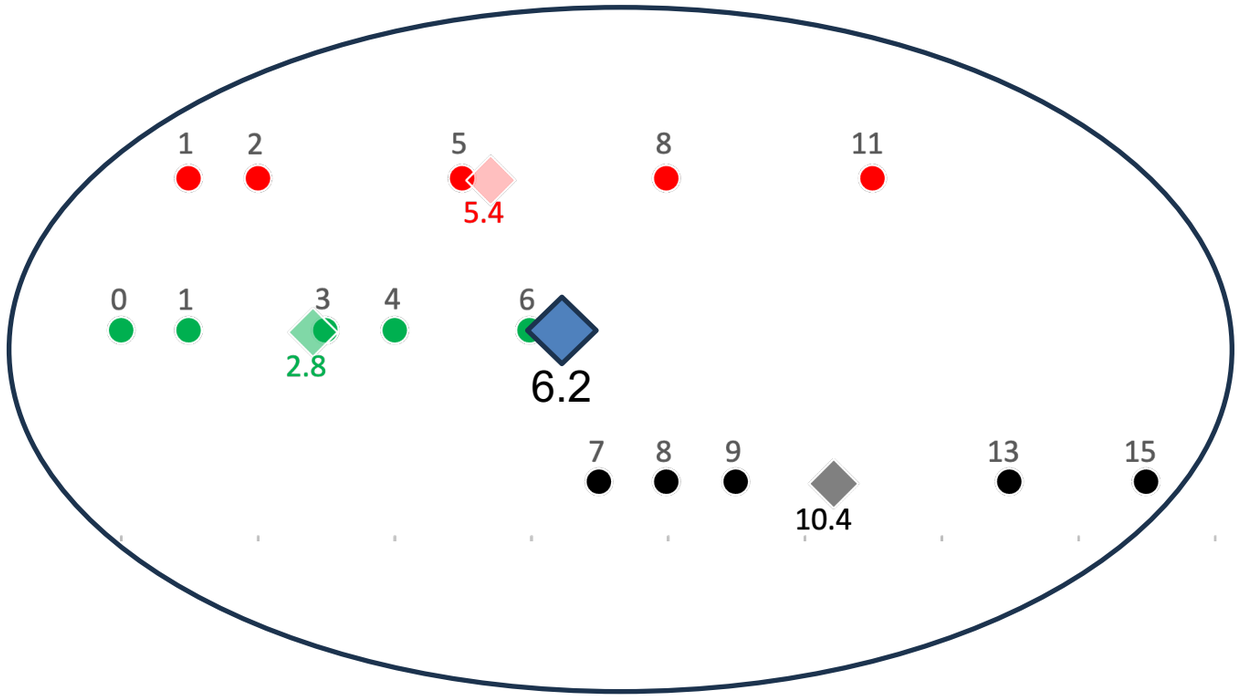
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Ejercicio 2.1.

Suma de cuadrados total



Varianza Total de la Variable Dependiente

1º Suma de Cuadrados Total

$$SS_{\text{Total}} = \sum (Y_i - \overline{Y}_T)^2$$

Suma de Cuadrados Total

La suma de los cuadrados de las desviaciones de cada punto respecto a la gran general

II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Ejercicio 2.

1º Suma de Cuadrados Total

Hallar las sumas de cuadrados del siguiente ejemplo

Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7
ΣY	27	14	52

^2

Y_G1^2	Y_G2^2	Y_G3^2
25	0	169
1	1	81
121	16	64
4	36	225
64	9	49

ΣY^2	865

$\sum Y^2$

$\frac{(\sum Y)^2}{n}$

$(\Sigma Y)^2$	8649
$(\Sigma Y)^2/n$	576,6

$(\Sigma Y)^2/n$	576,6

Suma de Cuadrados Total	288,4

II. Fundamentos del Análisis de la Varianza

2.1 Fuentes de Variación

Ejercicio 2.

1º Suma de Cuadrados Total

Hallar las sumas de cuadrados del siguiente ejemplo

Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7
ΣY	27	14	52

^2

Y_G1^2	Y_G2^2	Y_G3^2
25	0	169
1	1	81
121	16	64
4	36	225
64	9	49

ΣY^2	865

$\sum Y^2$

$\frac{(\sum Y)^2}{n}$

$(\Sigma Y)^2$	8649
$(\Sigma Y)^2/n$	576,6

$(\Sigma Y)^2/n$	576,6

Suma de Cuadrados Total	288,4

II. Fundamentos del Análisis de la Varianza

$$Y_{gi} = \mu + \alpha_g + e_{gi}$$

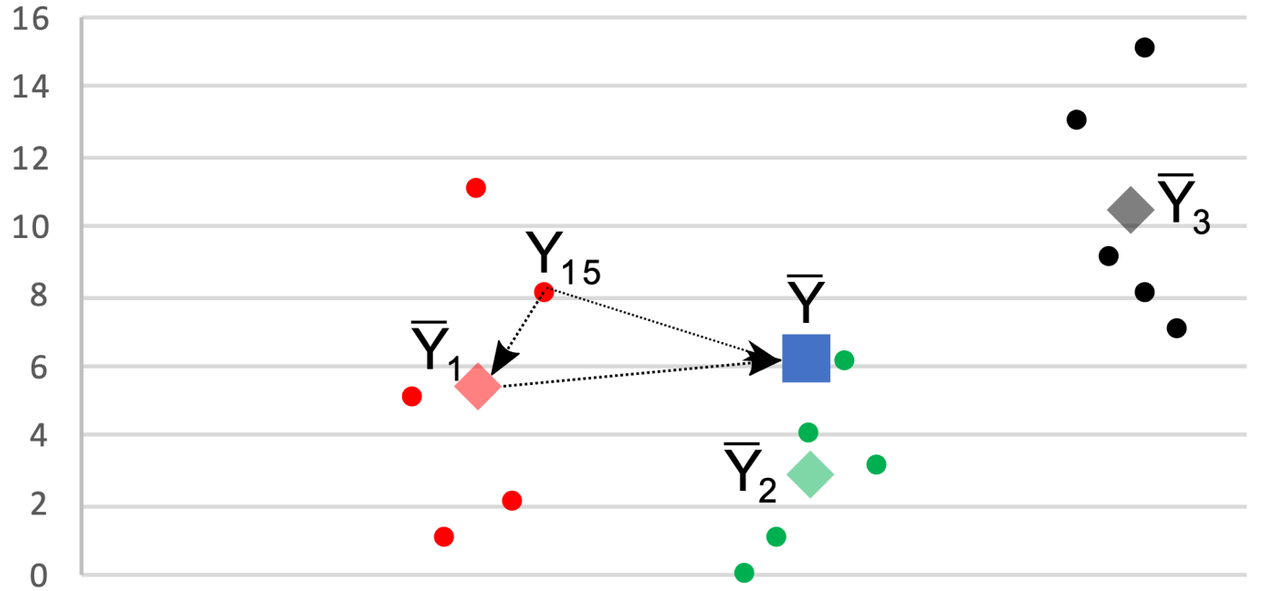
$\mathbf{Y_{gi}}$ es el valor observado de la variable respuesta

$\boldsymbol{\mu}$ es la respuesta esperada global (la media global)

$\boldsymbol{\mu_g}$ son los valores medios de la respuesta esperada para el tratamiento g

$\boldsymbol{\alpha_g}$ es el efecto del tratamiento g que es igual a $\boldsymbol{\mu_g - \mu}$

$\mathbf{e_{gi}}$ es el residuo o error experimental que es igual a $\mathbf{Y_{gi} - \mu_g}$



De modo que cualquier valor observado Y_{gi}

$$Y_{gi} = \overline{Y} + (\overline{Y}_g - \overline{Y}) + (Y_{gi} - \overline{Y}_g)$$

O lo que es lo mismo:

$$(Y_{gi} - \overline{Y}) = (Y_{gi} - \overline{Y}_g) + (\overline{Y}_g - \overline{Y})$$

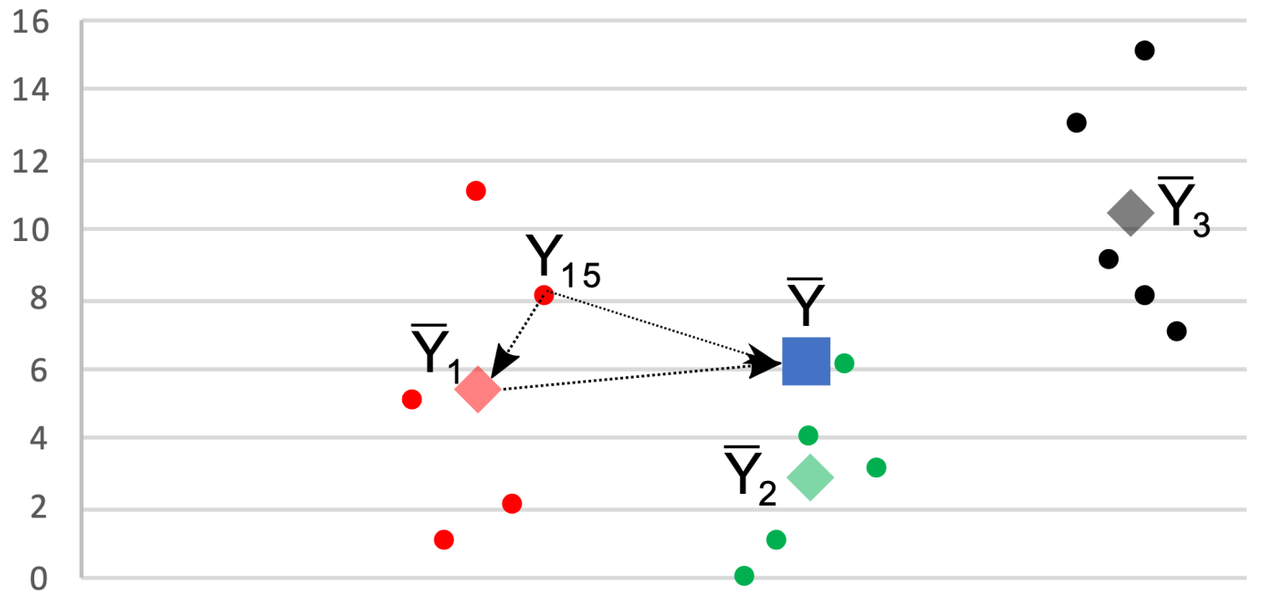
II. Fundamentos del Análisis de la Varianza

$$Y_{gi} = \mu + \alpha_g + e_{gi}$$

Desviación residual

$$(Y_{gi} - \overline{Y}) = (Y_{gi} - \overline{Y}_g) + (\overline{Y}_g - \overline{Y})$$

Desviación explicada por el factor



Si elevamos al cuadrado ambos miembros obtenemos:

$$(Y_{gi} - \overline{Y})^2 = (Y_{gi} - \overline{Y}_g)^2 + (\overline{Y}_g - \overline{Y})^2 + 2\,(Y_{gi} - \overline{Y}_g)\,(\overline{Y}_g - \overline{Y})$$

II. Fundamentos del Análisis de la Varianza

Grados de Libertad (df) y Cuadrados Medios (MS)

Los grados de libertad están asociados con cada fuente de variabilidad

Grados de Libertad Entre Grupos

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

Grados de Libertad Intra Grupos

Grados de Libertad Total

II. Fundamentos del Análisis de la Varianza

Grados de Libertad (df) y Cuadrados Medios (MS)

Los grados de libertad están asociados con cada fuente de variabilidad

k = número de grupos; grados de libertad k – 1; 3 – 1 = 2

Grados de Libertad Entre Grupos

			6.2

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

II. Fundamentos del Análisis de la Varianza

Grados de Libertad (df) y Cuadrados Medios (MS)

Los grados de libertad están asociados con cada fuente de variabilidad

k = número de grupos; grados de libertad k – 1; 3 – 1 = 2

Grados de Libertad Entre Grupos

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

X			**6.2**

II. Fundamentos del Análisis de la Varianza

Grados de Libertad (df) y Cuadrados Medios (MS)

Los grados de libertad están asociados con cada fuente de variabilidad

k = número de grupos; grados de libertad k – 1; 3 – 1 = 2

Grados de Libertad Entre Grupos

X	X		6.2

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

II. Fundamentos del Análisis de la Varianza

Grados de Libertad (df) y Cuadrados Medios (MS)

Los grados de libertad están asociados con cada fuente de variabilidad

k = número de grupos; grados de libertad k – 1; 3 – 1 = 2

Grados de Libertad Entre Grupos

X	X	?	6.2

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

II. Fundamentos del Análisis de la Varianza

Grados de Libertad (df) y Cuadrados Medios (MS)

Los grados de libertad están asociados con cada fuente de variabilidad

grados de libertad; (k(n-1)); 3(5 – 1) = 12

Grados de Libertad Entre Grupos

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

Grados de Libertad Intra Grupos

X	**X**	**X**
X	X	X
X	X	X
X	X	X
?	?	?
5.4	2.8	10.4

II. Fundamentos del Análisis de la Varianza

Grados de Libertad (df) y Cuadrados Medios (MS)

Los grados de libertad están asociados con cada fuente de variabilidad

grados de libertad; n-1; n - 1= 14

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

Grados de Libertad Total

X	**X**	**X**
X	X	X
X	X	X
X	X	X
?	X	X
6.2		

II. Fundamentos del Análisis de la Varianza

Grados de Libertad (df) y Cuadrados Medios (MS)

Los cuadrados medios (MS) son la varianza de cada una de las fuentes de variación

No son aditivos

SS_F / df $MS_F = 149.2 / 2 = 74.6$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

SS_E / df $MS_E = 139.2 / 12 = 11.6$



III. F ratio y Tamaño del Efecto

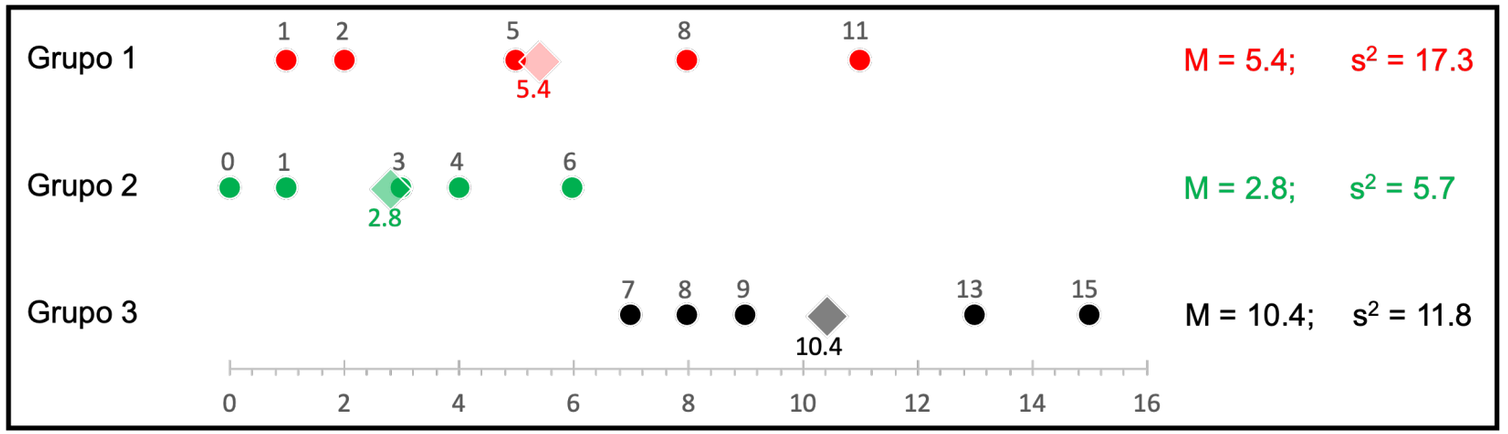
- **Estadístico F**

Ratio entre los Cuadrados Medios del Factor y los Cuadrados Medios Residuales

III. F ratio y Tamaño del Efecto

Estadístico F

Ratio entre los Cuadrados Medios del Factor y los Cuadrados Medios Residuales



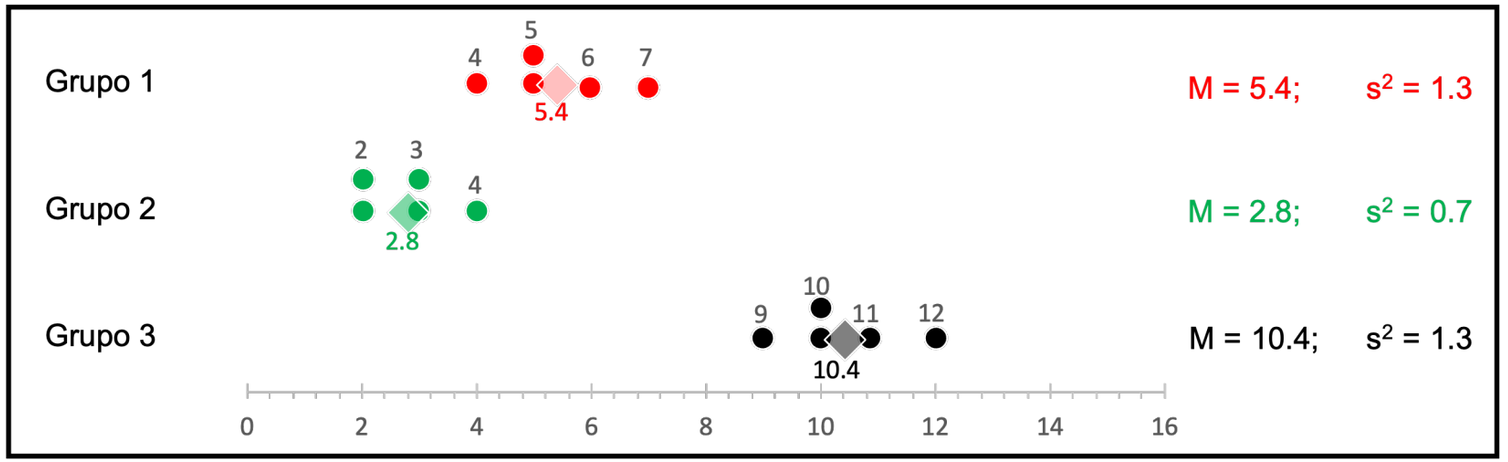
Varianza entre grupos $s^2 = 74.6$

Varianza intra grupos

$s^2 = 11.6$

$$\frac{17.3 + 5.7 + 11.8}{3}$$

$$F = \frac{74.6}{11.6} = 6.43$$



Varianza entre grupos $s^2 = 74.6$

Varianza intra grupos

$s^2 = 1.10$

$$\frac{1.3 + 0.7 + 1.3}{3}$$

$$F = \frac{74.6}{1.10} = 67.82$$

III. F ratio y Tamaño del Efecto

Estadístico F

Ratio entre los Cuadrados Medios del Factor y los Cuadrados Medios Residuales

$$F = \frac{Cuadrados\ Medios\ Tratamiento}{Cuadrados\ Medios\ Error}$$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

$$F = \frac{74.6}{11.6}$$

III. F ratio y Tamaño del Efecto

Estadístico F

Ratio entre los Cuadrados Medios del Factor y los Cuadrados Medios Residuales

$$F = \frac{Cuadrados\ Medios\ Tratamiento}{Cuadrados\ Medios\ Error}$$

El valor obtenido en este caso es 6.43

¿Cómo evaluar si es estadísticamente significativo o no lo es?

III. F ratio y Tamaño del Efecto

Estadístico F

Ratio entre los Cuadrados Medios del Factor y los Cuadrados Medios Residuales

$$F = \frac{Cuadrados\ Medios\ Tratamiento}{Cuadrados\ Medios\ Error}$$

El valor obtenido en este caso es 6.43

Cuál sería el valor medio de F esperado para nuestro ejemplo 12 / 12 -2 = 1.2

III. F ratio y Tamaño del Efecto

Tamaño del Efecto

El tamaño del efecto es la medida de la fuerza de la relación entre las variables

Las dos medidas más empleadas para evaluar el tamaño del efecto son:

- eta cuadrado (η^2)
- omega Cuadrado (ω2)

Eta Cuadrado (η^2)

Es también conocido como el coeficiente de determinación (R^2)

Viene dado por:

$$\eta^2 = \frac{Suma\ de\ Cuadrados\ del\ Factor}{Suma\ de\ Cuadrados\ Total}$$

III. F ratio y Tamaño del Efecto

Eta Cuadrado (η^2)

Es también conocido como el coeficiente de determinación (R^2)

Viene dado por:

$$\eta^2 = \frac{Suma\ de\ Cuadrados\ del\ Factor}{Suma\ de\ Cuadrados\ Total}$$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

$$\eta^2 = \frac{149.2}{288.4}$$

$$\eta^2 = 0.52$$

Es la proporción de la varianza total que es explicada por el modelo

III. F ratio y Tamaño del Efecto

Omega Cuadrado (ω^2)

Mientras que eta cuadrado se basa en los datos de la muestra omega trata de estimar el tamaño del efecto para la población

Viene dado por:

$$\omega^2 = \frac{Suma\ de\ Cuadrados\ del\ Factor - (glf)\ MCR}{Suma\ de\ Cuadrados\ Total + MCR}$$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

$$\omega^2 = \frac{149.2 - (2 * 11.6)}{288.4 + 11.6}$$

$$\omega^2 = 0.42$$

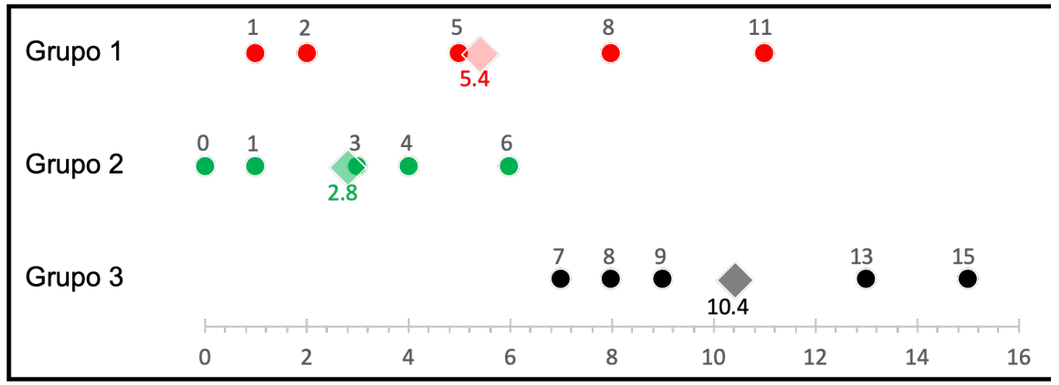
.01 efecto pequeño
.06 efecto medio
.14 efecto grande

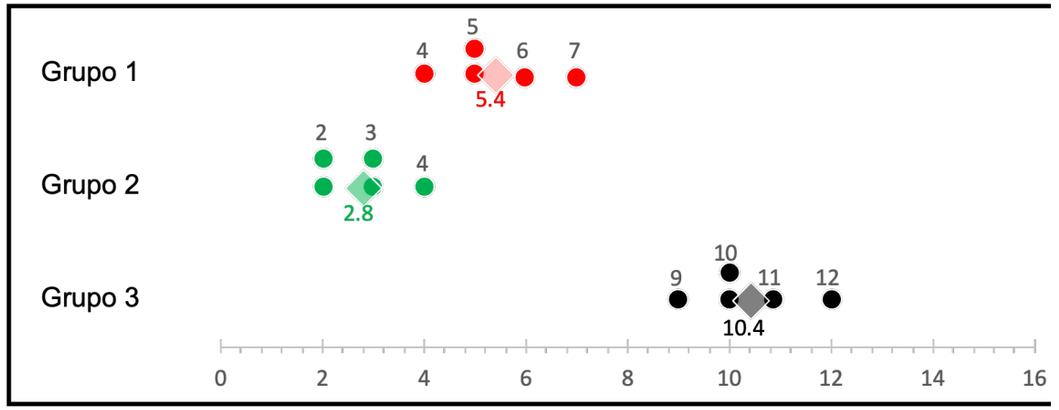
III. F ratio y Tamaño del Efecto

Ejercicio 3

Calcule el valor de F y el tamaño del efecto de los siguientes datos

Yi	Y_G1	Y_G2	Y_G3
Y_1	4	2	9
Y_2	5	2	10
Y_3	5	3	10
Y_4	6	3	11
Y_5	7	4	12





Compare los resultados con el ejemplo anterior

Razone el resultado



IV. Supuestos ANOVA

IV. Supuestos ANOVA

Supuestos de Modelo

- Normalidad de la variable de respuesta
  - La evaluación de este supuesto puede realizarse con:
    - si n < 50 → prueba de Shapiro-Wilk
    - si n > 50 → prueba de Kolmogorov-Smirnov
  - ANOVA es bastante robusto ante violaciones de este supuesto
- Homogeneidad de las varianzas o varianza constante para cada nivel del factor
  - Si los tamaños muestrales son iguales, la violación de este supuesto no es grave
    - Prueba de Levene (más robusta: Brown–Forsythe — mediana)
- Independencia de las observaciones
    - Prueba de Durbin-Watson (Valores DW próximos a 2 se consideran óptimos)

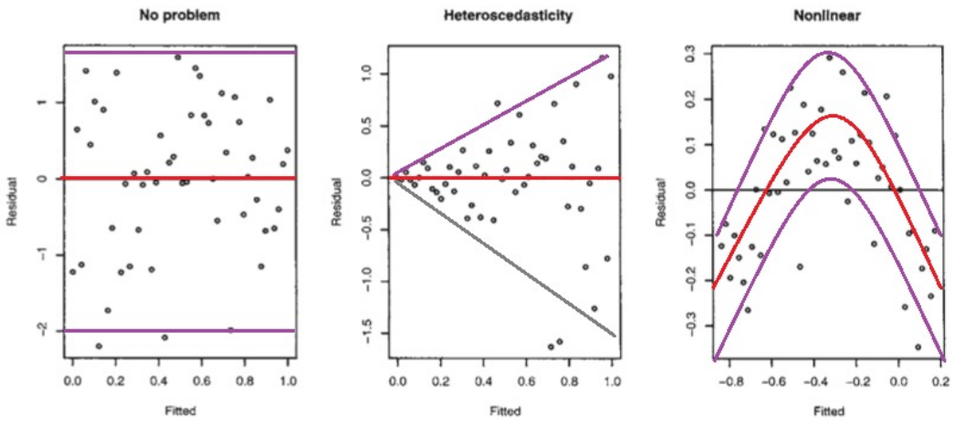
IV. Supuestos ANOVA

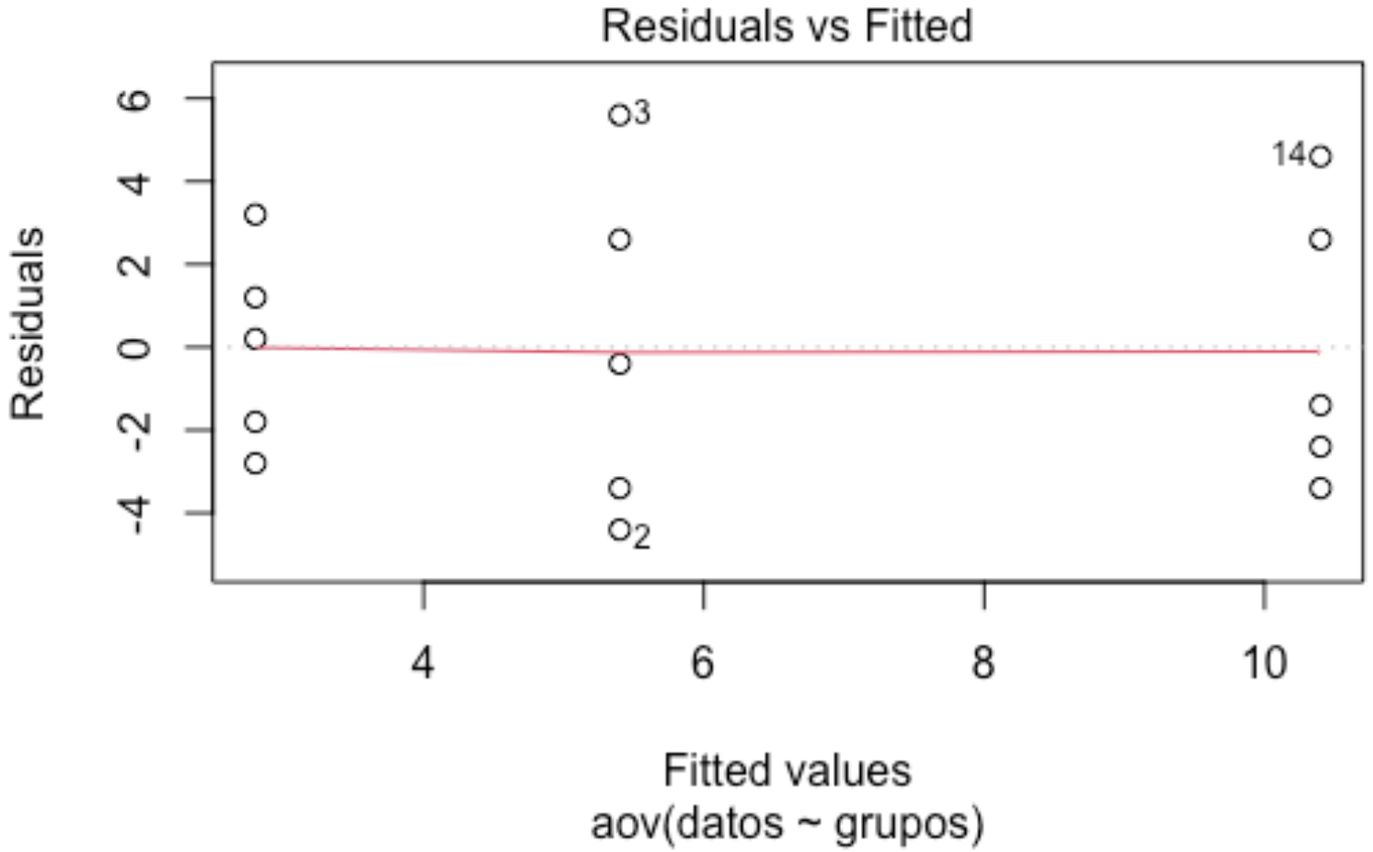
Independencia de los errores

Los errores asociados a las puntuaciones son independientes

Se asume que no hay un patrón

Si los errores fueran mucho mayores en alguno de los grupos no habría independencia





IV. Supuestos ANOVA

Distribución aproximadamente normal de los errores

Los errores se distribuyen de forma aproximadamente normal

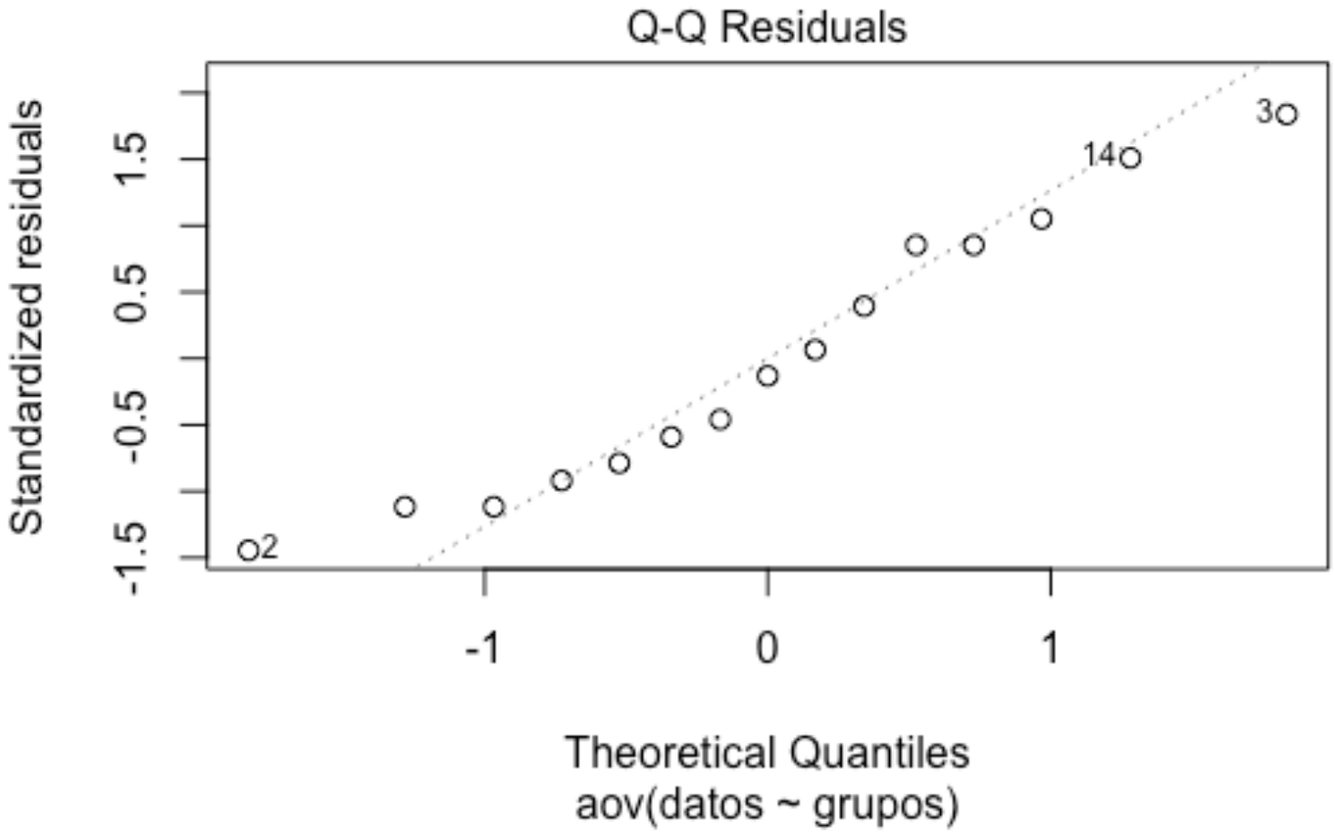
El gráfico de cuantiles (Normal Q-Q plot)

Permite evaluar la desviación de la normalidad

En función del alejamiento de los puntos respecto a la diagonal

En este caso se ajustan bastante bien

Los puntos 3 y 2 son una excepción



IV. Supuestos ANOVA

Homocedasticidad

Las varianzas en cada nivel del factor deben ser aproximadamente iguales

Se puede comprobar con la prueba de Levene

Prueba la hipótesis nula de que las varianzas de los grupos son iguales

Básicamente es una prueba ANOVA sobre los residuos absolutos

Si la prueba da un resultado significativo las varianzas son significativamente distintas

Ejercicio 4.

Compruebe, mediante la prueba de Levene, si las varianzas del ejemplo anterior son homogéneas entre grupos

IV. Supuestos ANOVA

Violaciones de los supuestos

ANOVA es bastante robusto ante la violación del supuesto de normalidad cuando el tamaño de los grupos es igual

Lo mismo respecto a la homogeneidad de varianzas

La violación de la independencia entre observaciones incrementa el error Tipo I

Medidas si no se cumplen los supuestos

Si no hay homogeneidad de varianzas → Welch's *F*

Si no hay normalidad → Alternativas no paramétricas (Kruskal-Wallis)

Si no hay normalidad → Métodos Robustos (Wilcox, 2005)



V. Pruebas Post Hoc

V. Pruebas Post Hoc

La Prueba F nos indica si hay o no diferencias entre grupos

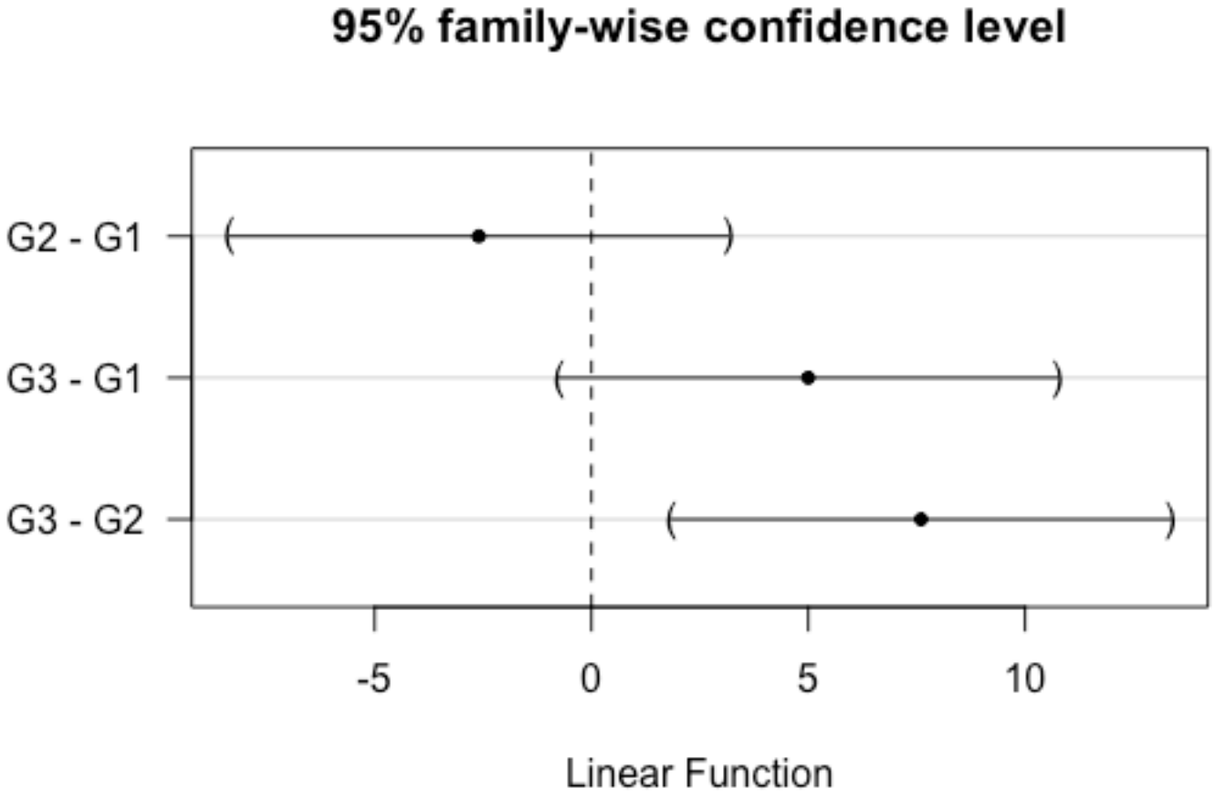
Pero no nos dice entre qué grupos hay diferencias

Solo sabemos si las diferencias entre las medias son estadísticamente significativas

Es necesario continuar el análisis para saber entre qué grupos hay diferencias

Para ello hacemos comparaciones *Post hoc*

Se trata de comparar las medias entre grupos dos a dos sin incrementar el error Tipo I



V. Pruebas Post Hoc

Estas pruebas controlan la inflación del error Tipo I

De tal forma que este error se mantiene en .05

La forma más sencilla de hacerlo es dividir alfa por el número de comparaciones:

$p_{crit} = \frac{\alpha}{k}$ Este método se conoce como la corrección de Bonferroni

El problema de este método es que se pierde potencia estadística

Se incrementa la probabilidad de rechazar un efecto significativo que sí existe (error Tipo II)

Por lo que han surgido otros métodos, ie. Holm:

$p_{crit} = \frac{\alpha}{j}$

Se ordenan las comparaciones y se les asigna un número de orden (j)

A la mayor probabilidad se le asigna 1 de tal forma que 0.05 / 1 = 0.05

Y la menor probabilidad es igual a la fórmula de Bonferroni

V. Pruebas Post Hoc

Resumen

El error Tipo I y la potencia estadística están relacionados

Si la probabilidad de error Tipo I es muy pequeña la probabilidad de error Tipo II aumenta

Cuando los tamaños de los grupos son iguales y las varianzas similares conviene usar:

- Tukey si el número de comparaciones es grande
- Bonferroni si el número de comparaciones es pequeño

Si los tamaños de los grupos son distintos:

- Gabriel o Hochberg

Número de comparaciones posible = $\frac{a\ (a-1)}{2}$

V. Pruebas Post Hoc

Resumen

Si no hay homocedasticidad (no es posible asumir igualdad de varianzas):

- Games-Howell
- Dunnett's C
- Dunnett's T3
- Tamhane T2

- Generalmente Games-Howel es más potente que Dunnett's C
- Con mayor número de grados de libertad Dunnett's C es más potente que Dunnett's T3
- Con menor número de grados de libertad Dunnett's T3 es más potente que Dunnett's C
- Tamhane T2 es el más conservador de los 4

V. Pruebas Post Hoc

Resumen

Si hay homocedasticidad (es posible asumir igualdad de varianzas) y grupos iguales:

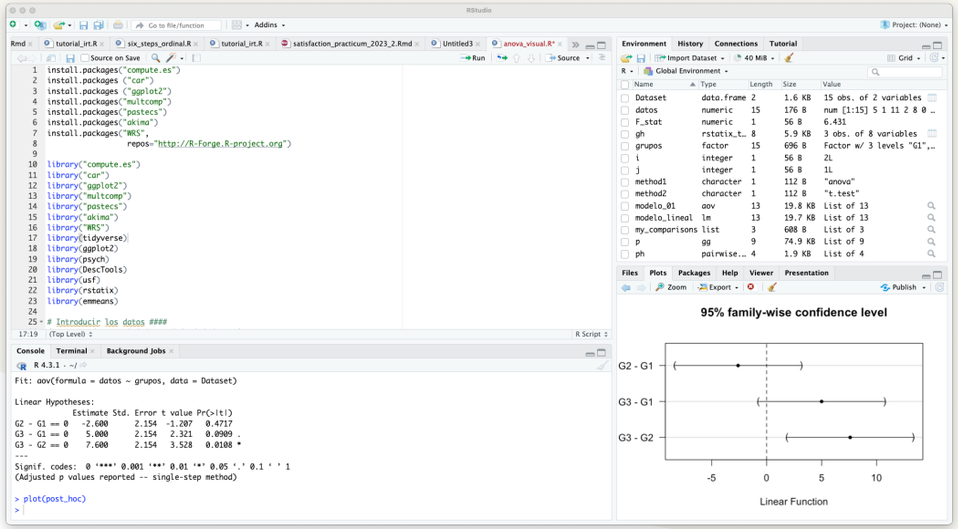


- LSD de Fisher es el más potente pero no controla el error Tipo I
- Algunos autores recomiendan no usar Tukey-b (Keppel and Wickens, 2004)
- Scheffe's es el más conservador

VI. Cálculo de ANOVA con R y SPSS

Conjunto de datos

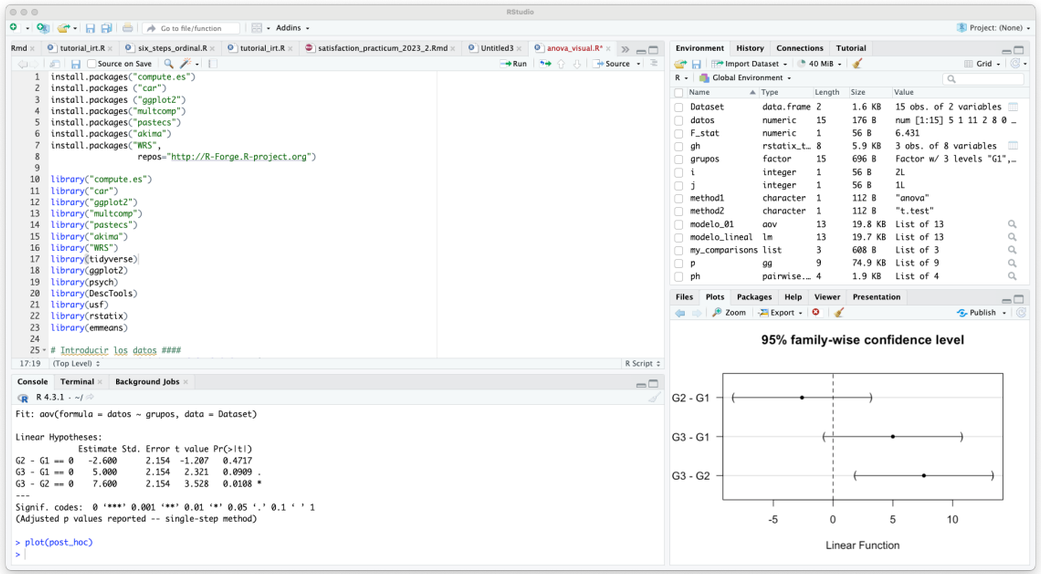
Yi	Y_G1	Y_G2	Y_G3
Y_1	4	2	9
Y_2	5	2	10
Y_3	5	3	10
Y_4	6	3	11
Y_5	7	4	12
Medias	5,4	2,8	10,4



VI. Cálculo de ANOVA con R y SPSS

Conjunto de datos

Yi	Y_G1	Y_G2	Y_G3
Y_1	4	2	9
Y_2	5	2	10
Y_3	5	3	10
Y_4	6	3	11
Y_5	7	4	12
Medias	5,4	2,8	10,4



VI. Cálculo de ANOVA con R y SPSS

R



Utilizaremos R-Studio

Paquetes para ANOVA de una vía:

```
library(compute.es)
library(car)
library(ggplot2)
library(multcomp)
library(pastecs)
library(WRS)
```

Instalación de los paquetes:

```
install.packages("compute.es")
install.packages("car")
install.packages("ggplot2")
install. packages("multcomp")
install.packages("pastecs")
install. packages("WRS",
repos="http://R-Forge.R-project.org")
```

VI. Cálculo de ANOVA con R y SPSS

R

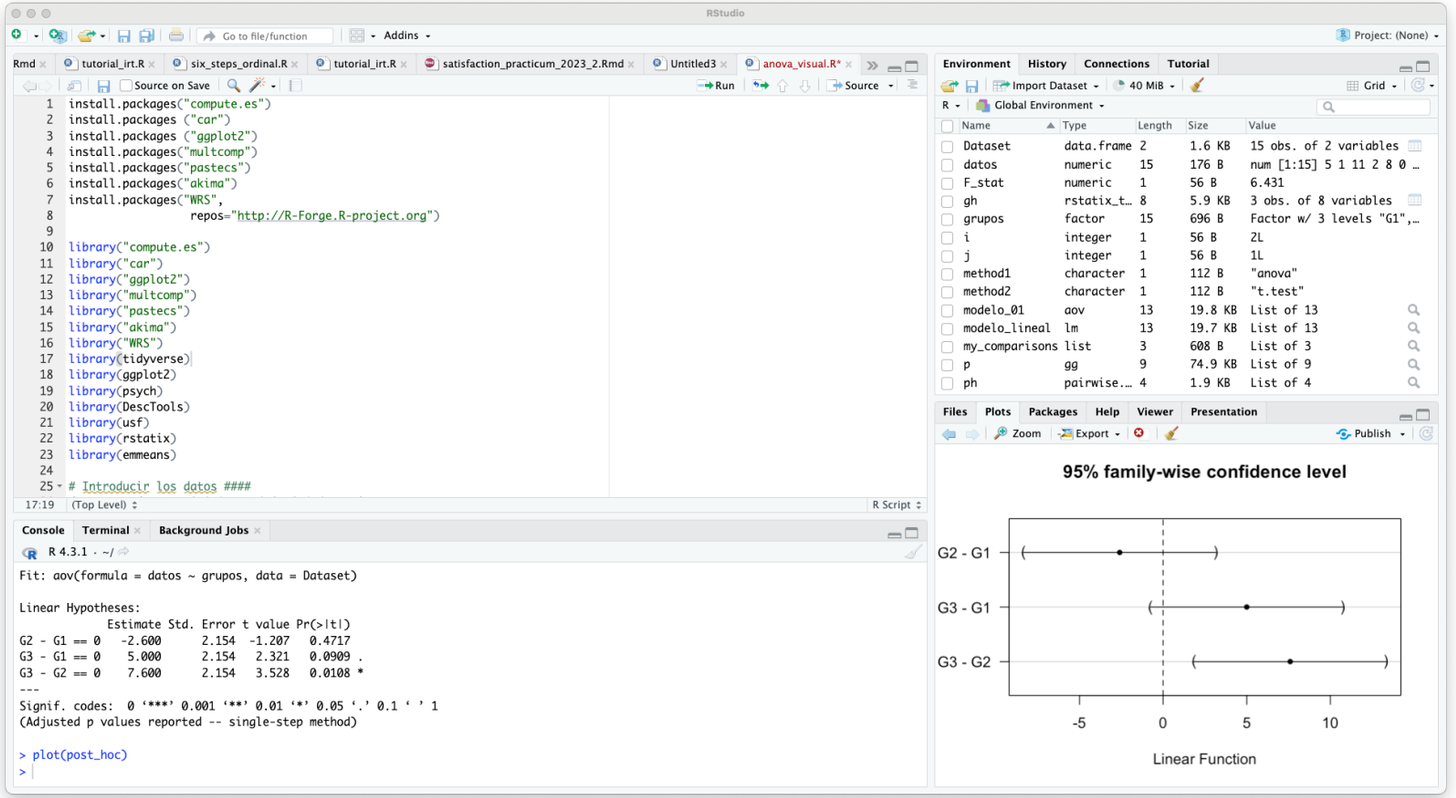
Procedimiento general

1º Introducir los datos

2º Explorar el conjunto de datos

3º Calcular el Análisis de la Varianza

4º Calcular los contrastes post hoc



VI. Cálculo de ANOVA con R y SPSS

R

Procedimiento general

1º Instalar los paquetes necesarios

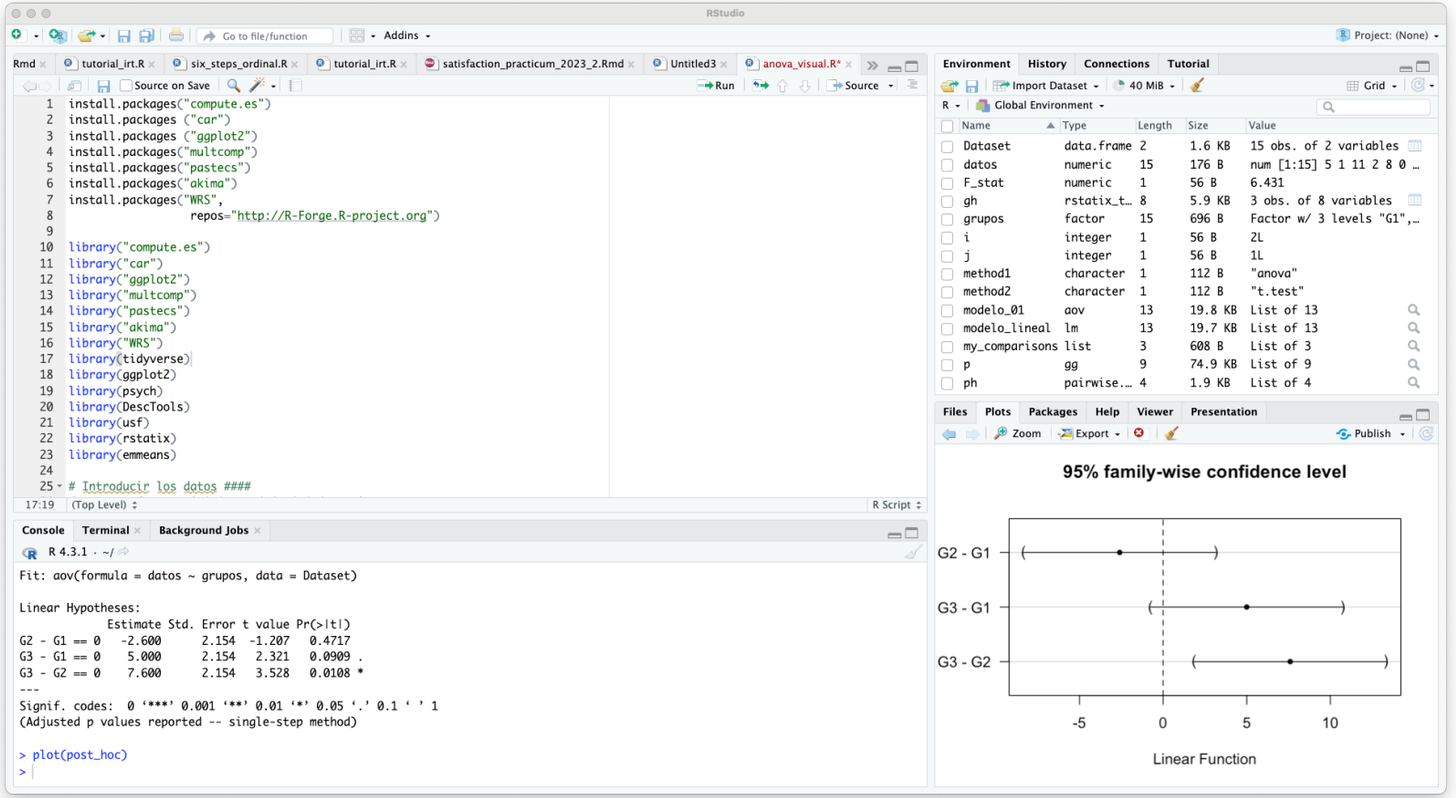
2º Cargar los paquetes

3º Introducir los datos

4º Explorar el conjunto de datos

5º Análisis de la Varianza

6º Calcular los contrastes post hoc



VI. Cálculo de ANOVA con R y SPSS

R

Procedimiento general

1º Instalar los paquetes necesarios

2º Cargar los paquetes

```
install.packages("compute.es")
install.packages ("car")
install.packages ("ggplot2")
install.packages("multcomp")
install.packages("pastecs")
install.packages("akima")
install.packages("WRS", repos="http://R-Forge.R-project.org")

library("compute.es")
library("car")
library("ggplot2")
library("multcomp")
library("pastecs")
library("akima")
library("WRS")
```

```
install.packages("compute.es")
install.packages ("car")
install.packages ("ggplot2")
install.packages("multcomp")
install.packages("pastecs")
install.packages("akima")
install.packages("WRS",
    repos="http://R-Forge.R-project.org")

library("compute.es")
library("car")
library("ggplot2")
library("multcomp")
library("pastecs")
library("akima")
library("WRS")
```

VI. Cálculo de ANOVA con R y SPSS

R

3º Introducir los datos (1)



```
# Introducir los datos ####
datos <- c(5,1,11,2,8,0,1,4,6,3,13,9,8,15,7)

# Introducir los grupos ####
grupos <- gl(3,5, labels = c("G1", "G2", "G3"))

# Crear el conjunto de datos ####
Dataset <- data.frame(grupos, datos)

# Podemos ver el conjunto de datos
View(Dataset)
```


```
# Introducir los datos ####
datos <- c(5,1,11,2,8,0,1,4,6,3,13,9,8,15,7)

# Introducir los grupos ####
grupos <- gl(3,5, labels = c("G1", "G2", "G3"))

# Crear el conjunto de datos ####
Dataset <- data.frame(grupos, datos)

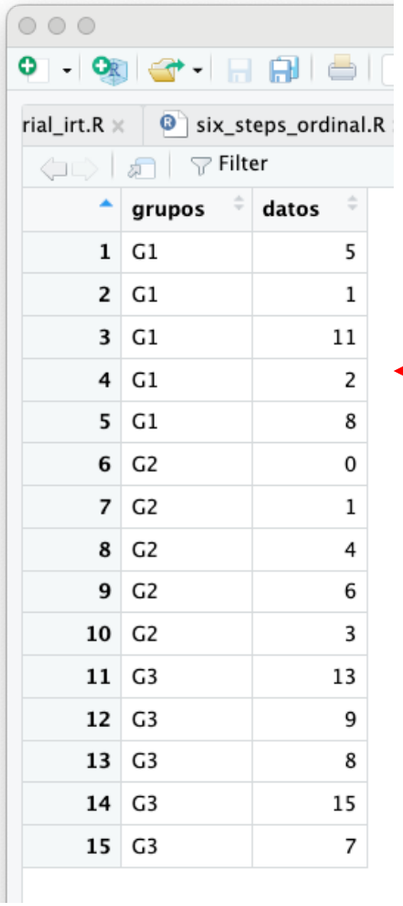
# Podemos ver el conjunto de datos
View(Dataset)
```

VI. Cálculo de ANOVA con R y SPSS

R

3º Introducir los datos (2)

Hemos pasado del formato “wide” que teníamos en Excel al formato “long”



	grupos	datos
1	G1	5
2	G1	1
3	G1	11
4	G1	2
5	G1	8
6	G2	0
7	G2	1
8	G2	4
9	G2	6
10	G2	3
11	G3	13
12	G3	9
13	G3	8
14	G3	15
15	G3	7

CÁLCULO DE LA VARIANZA

Yi	Y_G1	Y_G2	Y_G3
Y_1	5	0	13
Y_2	1	1	9
Y_3	11	4	8
Y_4	2	6	15
Y_5	8	3	7

VI. Cálculo de ANOVA con R y SPSS

R

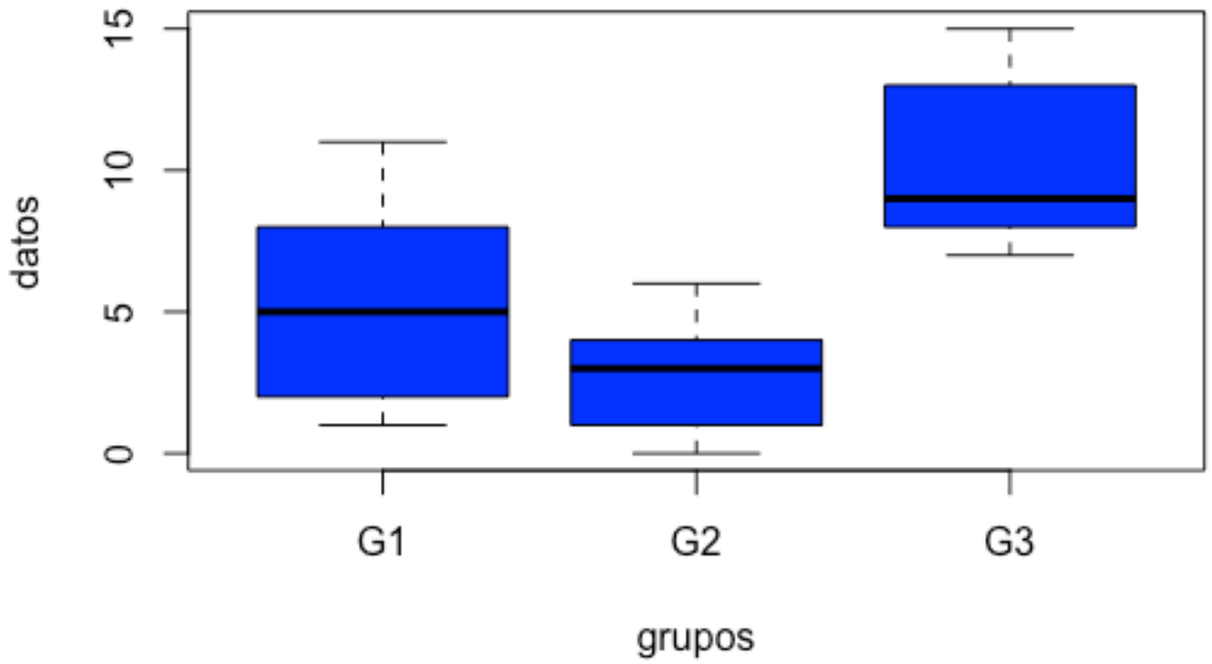
4º Explorar el conjunto de datos (1)

```
# Explorar los datos visualmente
plot(Dataset, pch = 16, col = "blue")
```

```
# Hacer un análisis descriptivo ####
by(Dataset$datos, Dataset$grupos, stat.desc)

describeBy(Dataset, Dataset$grupos)
```

```
group: G1
        vars n mean   sd median trimmed  mad min max range skew kurtosis   se
grupos*    1 5  1.0 0.00      1     1.0 0.00   1   1     0  NaN      NaN 0.00
datos      2 5  5.4 4.16      5     5.4 4.45   1  11    10 0.19    -1.97 1.86
------------------------------------------------------------------------------
group: G2
        vars n mean   sd median trimmed  mad min max range skew kurtosis   se
grupos*    1 5  2.0 0.00      2     2.0 0.00   2   2     0  NaN      NaN 0.00
datos      2 5  2.8 2.39      3     2.8 2.97   0   6     6  0.1     -1.9 1.07
------------------------------------------------------------------------------
group: G3
        vars n mean   sd median trimmed  mad min max range skew kurtosis   se
grupos*    1 5  3.0 0.00      3     3.0 0.00   3   3     0  NaN      NaN 0.00
datos      2 5 10.4 3.44      9    10.4 2.97   7  15     8 0.29    -2.05 1.54
```



VI. Cálculo de ANOVA con R y SPSS

R

4º Explorar el conjunto de datos (2)

```
# Prueba de homogeneidad de varianzas de Levene
leveneTest (Dataset$datos, Dataset$grupos,
            center = median)
```

```
> leveneTest(Dataset$datos, Dataset$grupos, center = mean)
Levene's Test for Homogeneity of Variance (center = mean)
      Df F value Pr(>F)
group  2  1.2275 0.3273
      12
```

Es el mismo resultado que obtuvimos con excel

La prueba de Levene arrojó un resultado **no significativo**

Podemos concluir que las varianzas en los tres grupos son similares

VI. Cálculo de ANOVA con R y SPSS

R

5º Calcular el Análisis de la Varianza (1)

```
modelo_01 <- aov(datos ~ grupos, data = Dataset)
summary(modelo_01)
```

```
             Df     Sum Sq    Mean Sq   F value   Pr(>F)
grupos        2      149.2     74.6        6.431     0.0126 *
Residuals    12      139.2     11.6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149,2	2	74,6	6,43103448	0,0126434	3,88529383
Within Groups	139,2	12	11,6			
Total	288,4	14				

VI. Cálculo de ANOVA con R y SPSS

R

5º Calcular el Análisis de la Varianza (Welch's F) (2)



Calcular el modelo ANOVA con varianzas desiguales (Welsch's F):

```
oneway.test(datos ~ grupos, data = Dataset, var.equal = FALSE))
```

```
> oneway.test(datos ~ grupos, data = Dataset, var.equal = FALSE)

        One-way analysis of means (not assuming equal variances)

data:  datos and grupos
F = 7.5855, num df = 2.0000, denom df = 7.5733, p-value = 0.01554
```

Corrige los grados de libertad del denominador

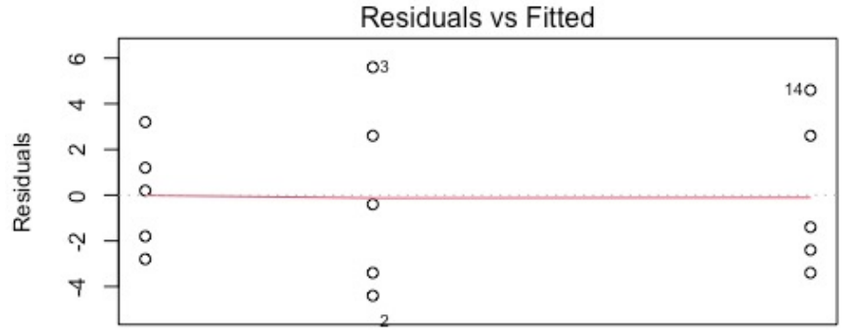
El resultado sigue siendo significativo

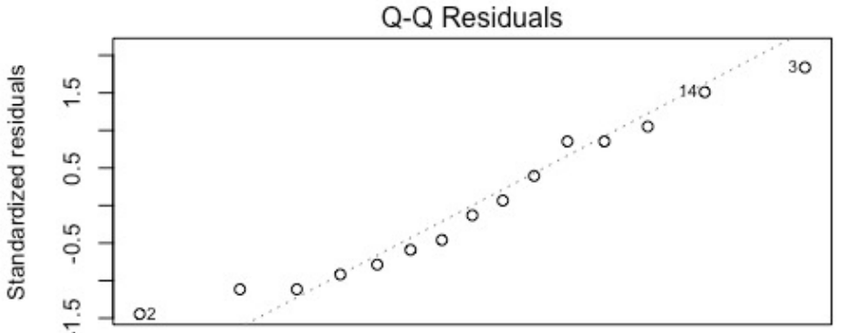
VI. Cálculo de ANOVA con R y SPSS

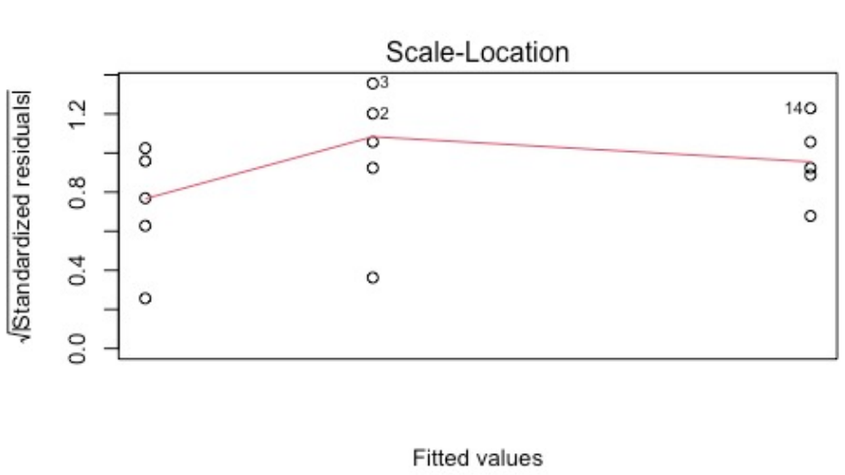
R

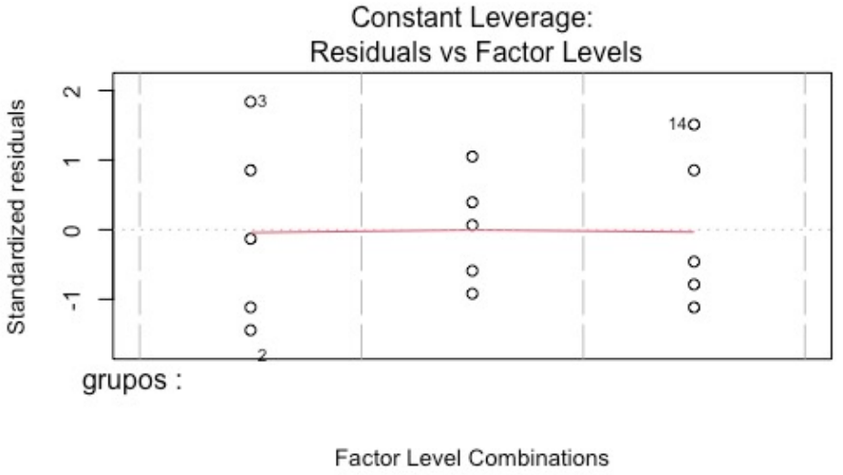
5º Calcular el Análisis de la Varianza — Diagnosis y validación del modelo (3)

```
plot(modelo_01)
```









VI. Cálculo de ANOVA con R y SPSS

R

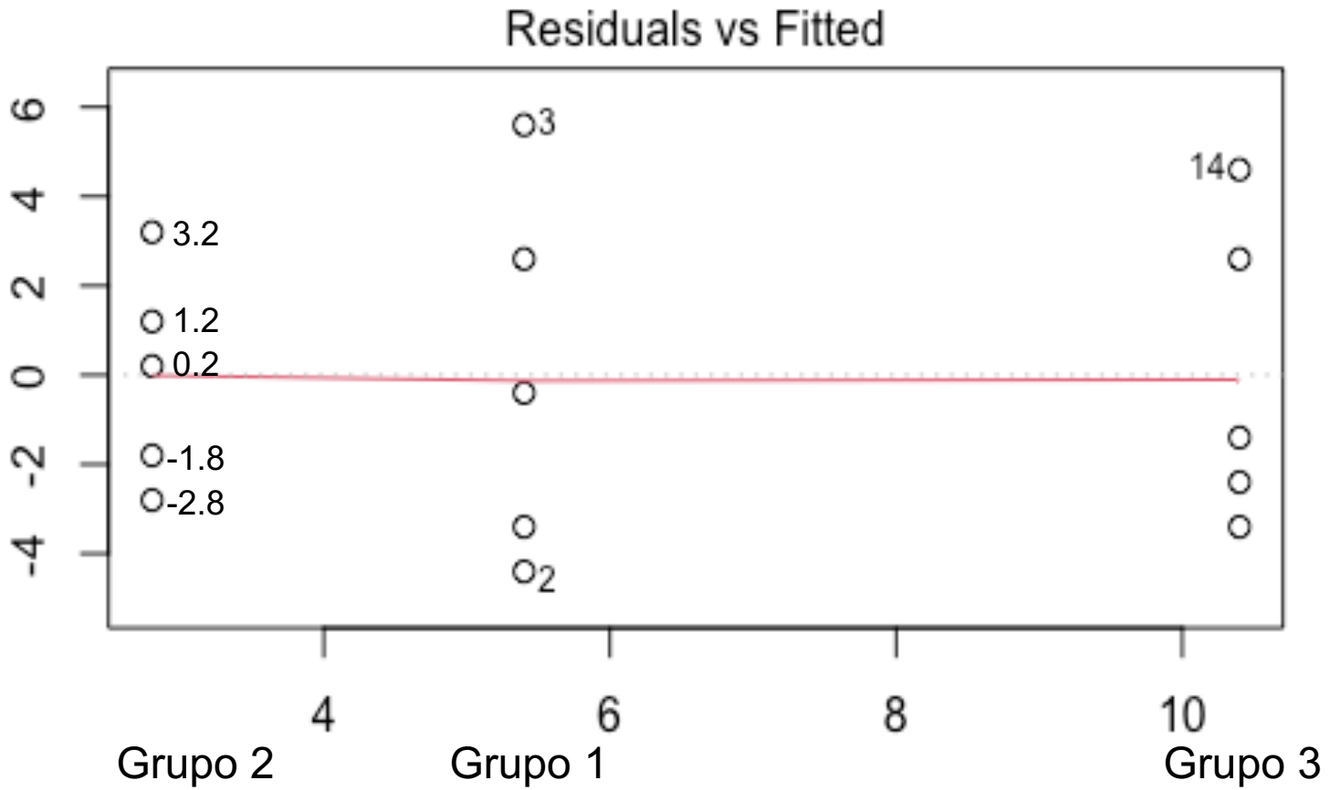
5º Calcular el Análisis de la Varianza — Diagnosis y validación del modelo (4)

```
plot(modelo_01, which = 1)
```

Este gráfico muestra los residuos en el eje de ordenadas

Residuos Grupo 2

Y_G2	Residuos
0	-2,8
1	-1,8
4	1,2
6	3,2
3	0,2
2,8	



VI. Cálculo de ANOVA con R y SPSS

R

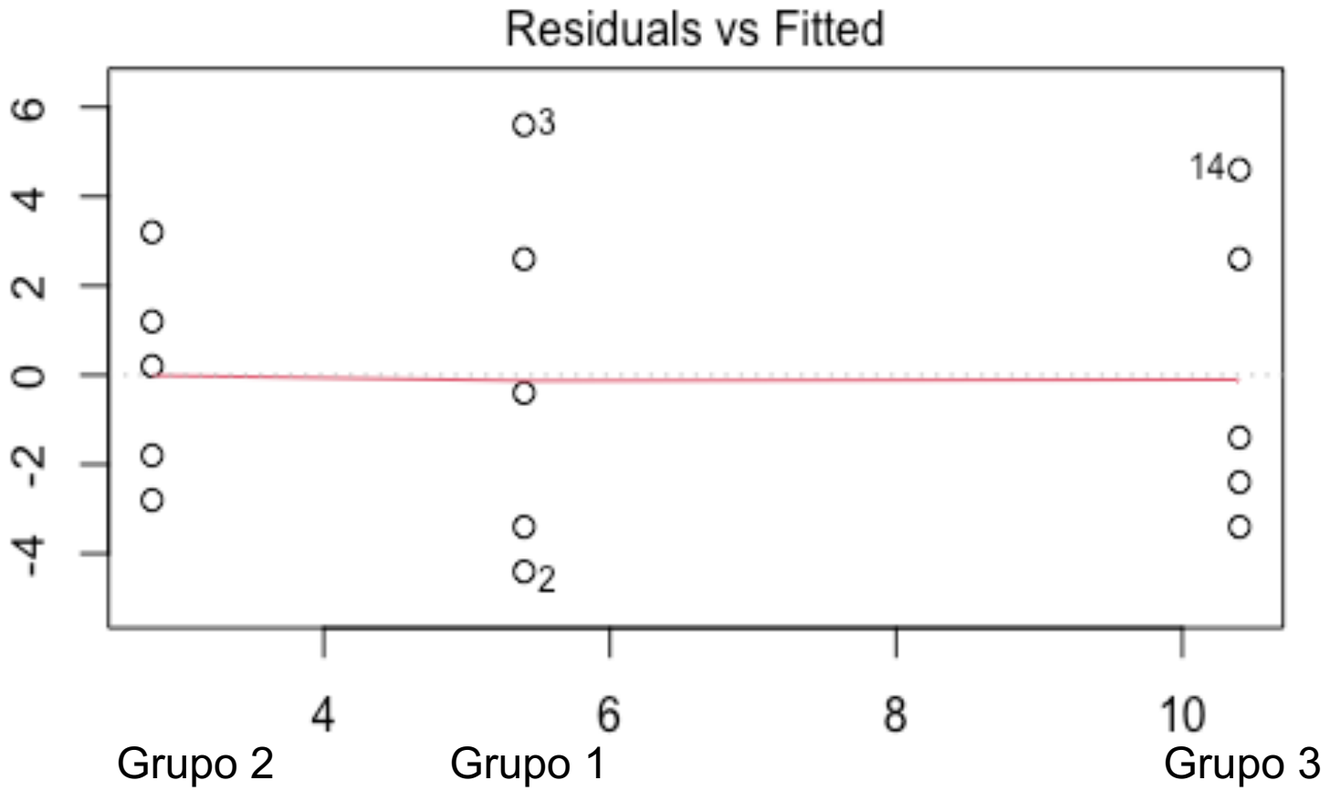
5º Calcular el Análisis de la Varianza — Diagnosis y validación del modelo (5)

```
plot(modelo_01, which = 1)
```

Este gráfico muestra los residuos en el eje de ordenadas

Residuos Grupo 1

Y_G1	Residuos
5	-0,4
1	-4,4
11	5,6
2	-3,4
8	2,6
5,4	

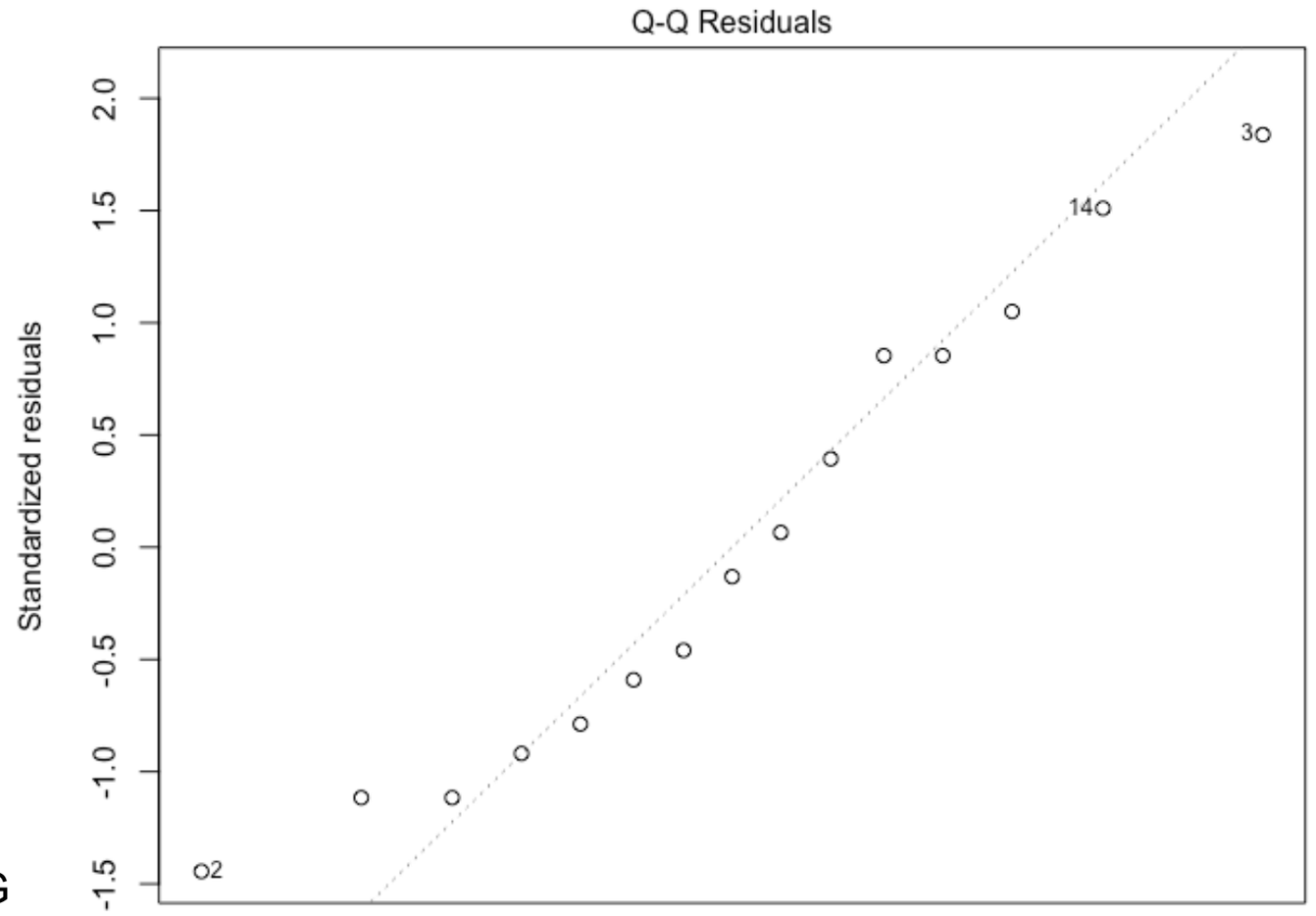


VI. Cálculo de ANOVA con R y SPSS

R

5º Calcular el Análisis de la Varianza — Diagnosis y validación del modelo (6)

```
plot(modelo_01, which = 2)
```



G

VI. Cálculo de ANOVA con R y SPSS

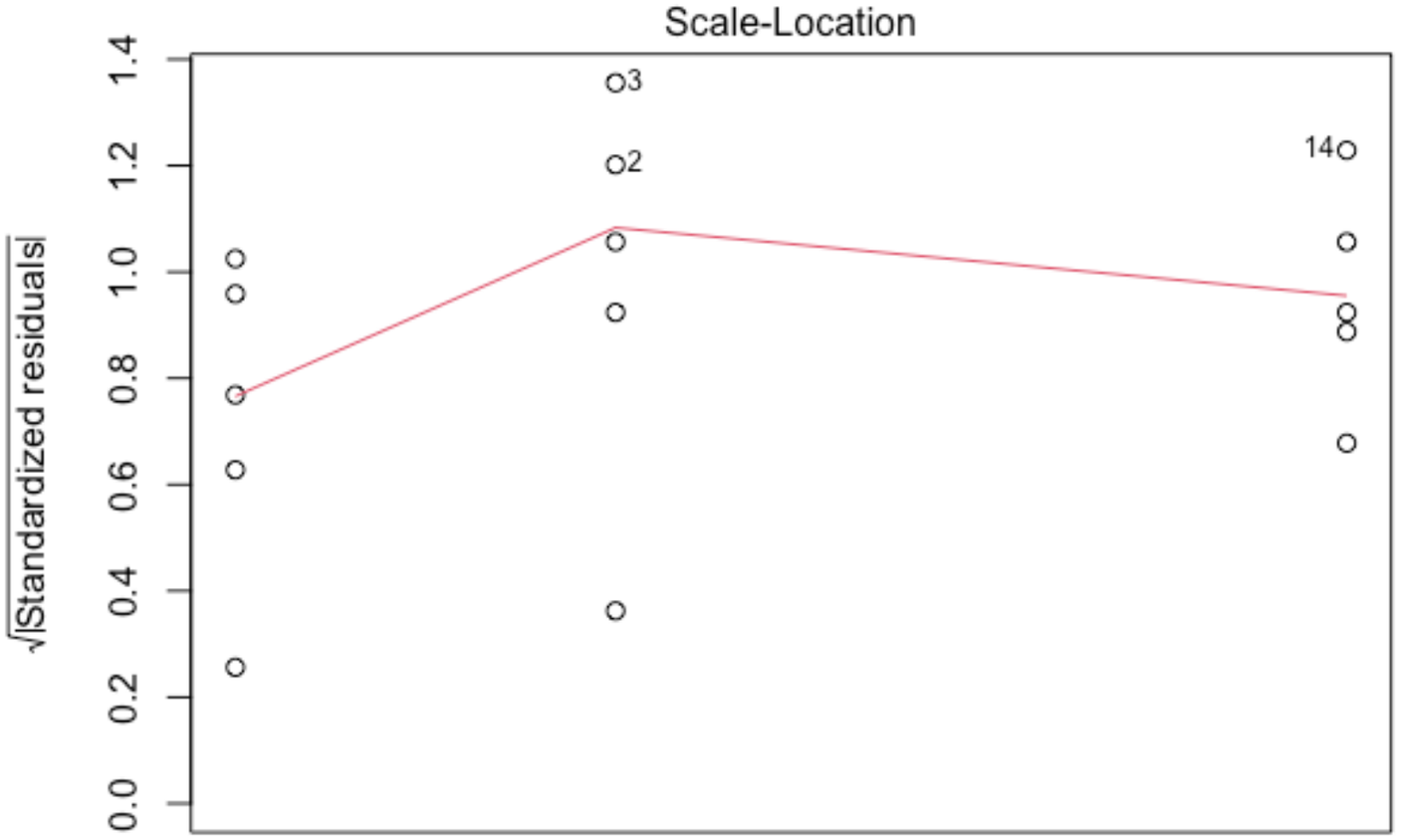
R

5º Calcular el Análisis de la Varianza — Diagnosis y validación del modelo (7)

```
plot(modelo_01, which = 3)
```

Valores superiores a 2 sugerirían heterogeneidad de varianzas

No se observa una tendencia clara de los residuos en función del grupo



VI. Cálculo de ANOVA con R y SPSS

R

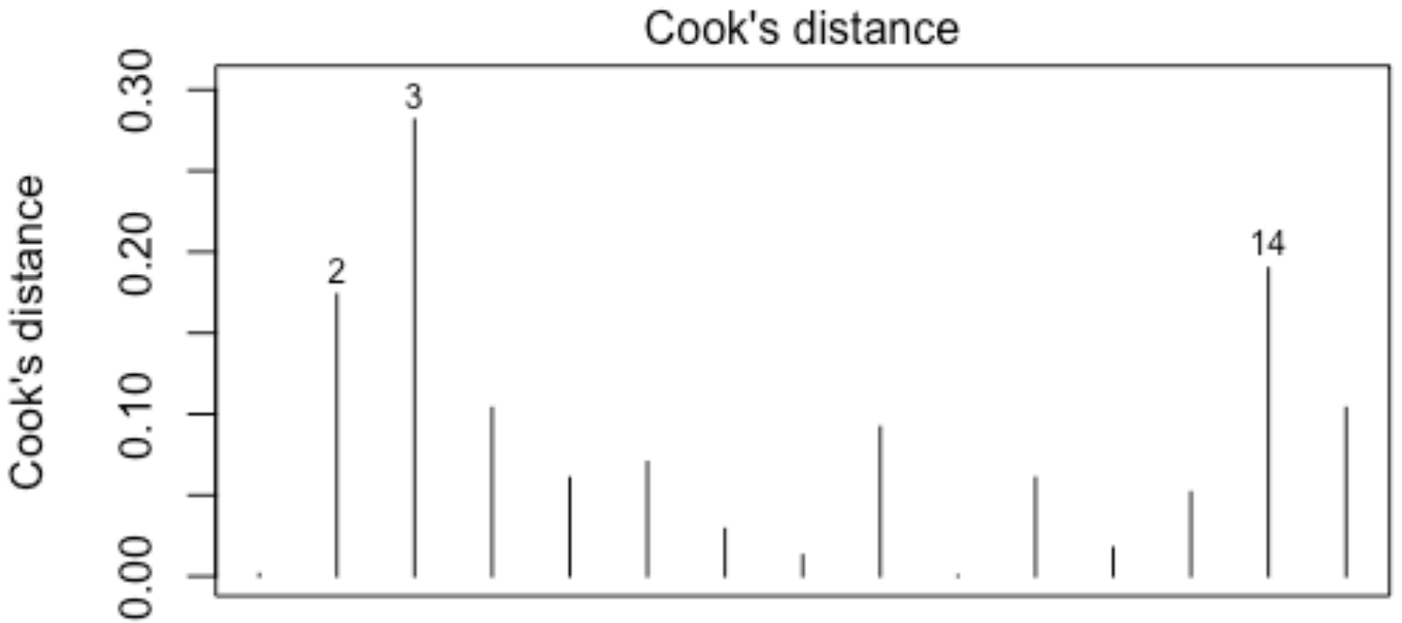
5º Calcular el Análisis de la Varianza — Diagnosis y validación del modelo (8)

Distancias de Cook

```
plot(modelo_01, which = 4)
```

Miden el impacto de cada observación en la capacidad predictiva del modelo

Distancias de Cook inferiores a 1 son adecuadas



VI. Cálculo de ANOVA con R y SPSS

R

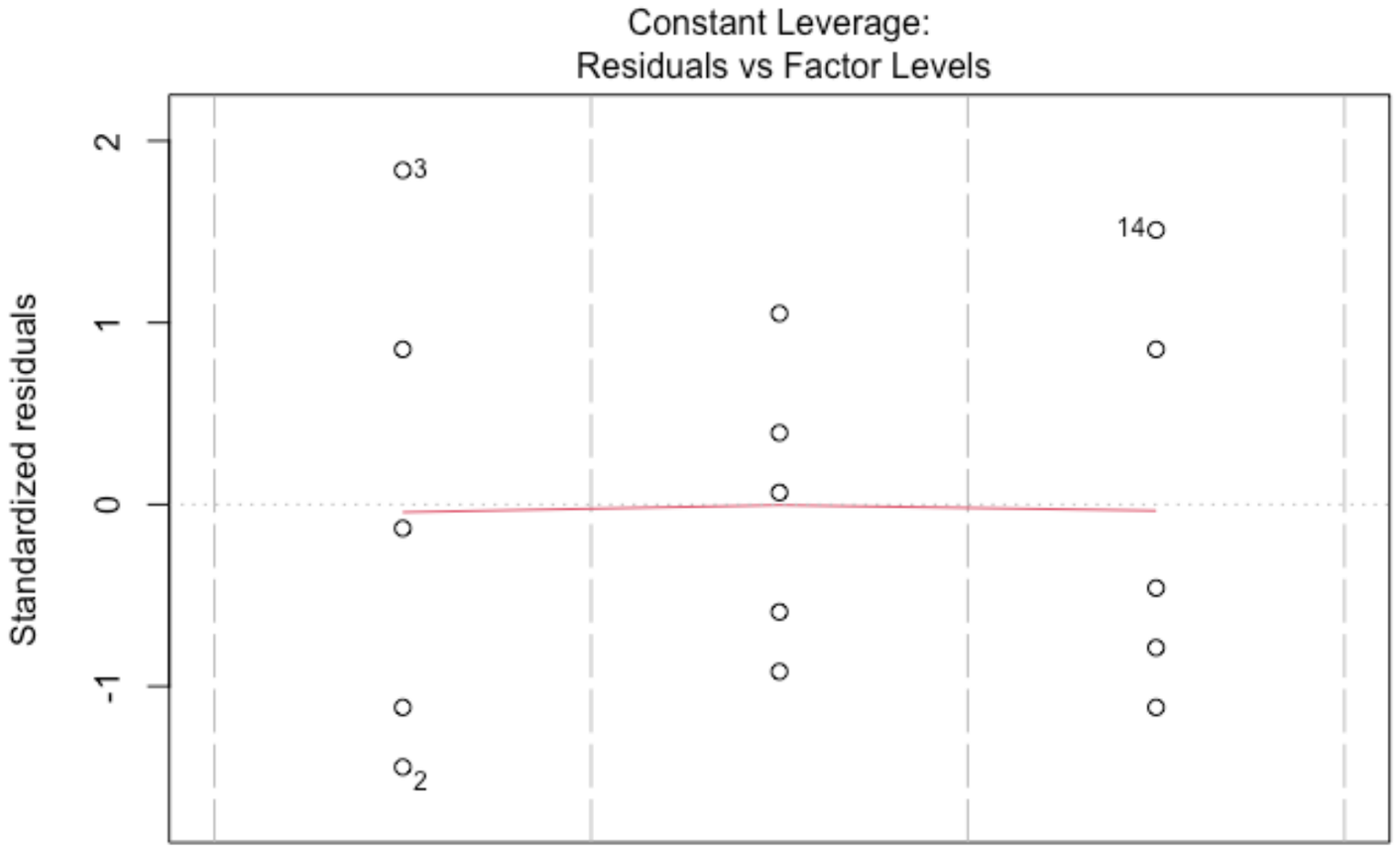
5º Calcular el Análisis de la Varianza — Diagnosis y validación del modelo (9)

```
plot(modelo_01, which = 5)
```

Menos del 5 % de los residuos estandarizados deben ser superiores a 2

Menos del 1 % de los residuos estandarizados deben ser superiores a 2.5

Los residuos estandarizados superiores a 3 serían outliers



La aproximación de la recta central al valor 0 sugiere que no hay puntos influyentes

VI. Cálculo de ANOVA con R y SPSS

R

6º Calcular los contrastes post hoc (1)

```
post_hoc <- glht(modelo_01, linfct = mcp(grupos = "Tukey"))

summary(post_hoc)

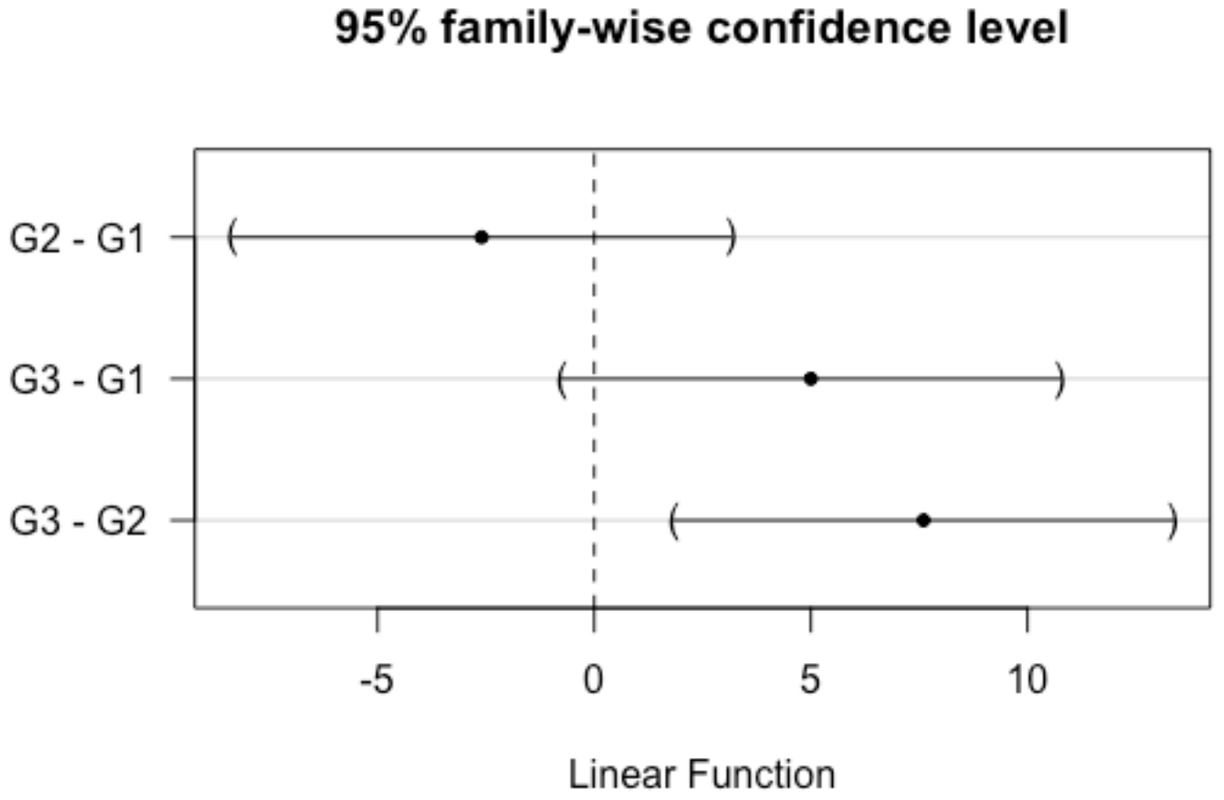
plot(post_hoc)
```

```
        Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = datos ~ grupos, data = Dataset)

Linear Hypotheses:
             Estimate Std. Error t value Pr(>|t|)
G2 - G1 == 0   -2.600      2.154  -1.207   0.4717
G3 - G1 == 0    5.000      2.154   2.321   0.0911 .
G3 - G2 == 0    7.600      2.154   3.528   0.0108 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```



VI. Cálculo de ANOVA con R y SPSS

R

6º Calcular los contrastes post hoc (2)

Bonferroni: #### Si el número de comparaciones es pequeño

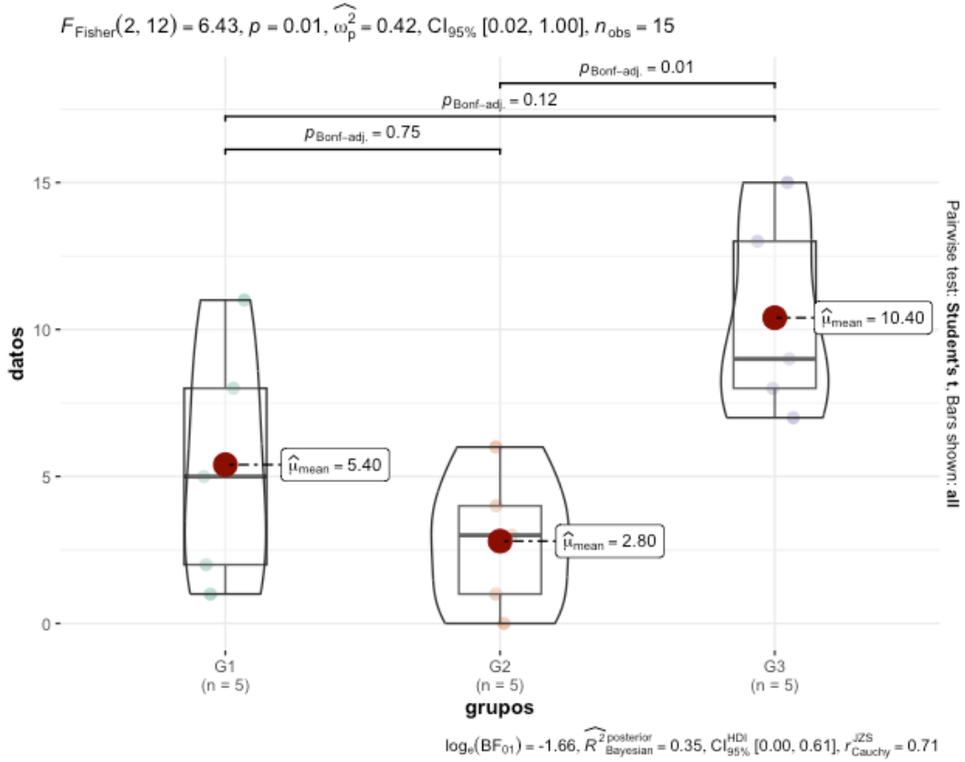
```
pairwise.t.test(Dataset$datos, Dataset$grupos, p.adj='bonferroni')
```

```
        Pairwise comparisons using t tests with pooled SD

data:  Dataset$datos and Dataset$grupos

   G1    G2
G2 0.752 -
G3 0.116 0.012

P value adjustment method: bonferroni
```



VI. Cálculo de ANOVA con R y SPSS

R



Ejercicio 5

Realice el análisis de la varianza de los siguientes datos en R

Yi	Y_G1	Y_G2	Y_G3
Y_1	4	2	9
Y_2	5	2	10
Y_3	5	3	10
Y_4	6	3	11
Y_5	7	4	12

VI. Cálculo de ANOVA con R y SPSS

SPSS

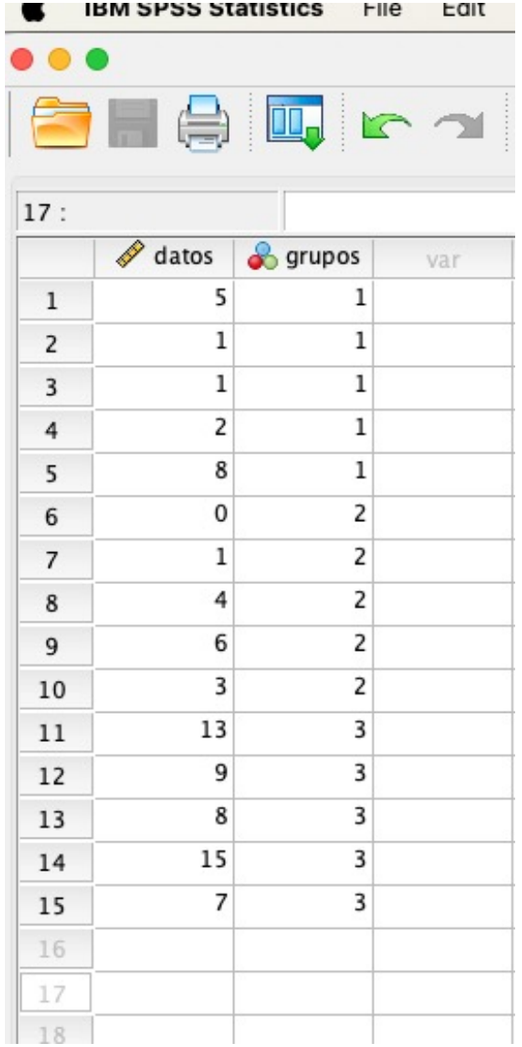
1º Archivo de datos

En la primera columna tenemos las puntuaciones

En la segunda columna los grupos

De nuevo utilizamos el formato “long”

anova.sav

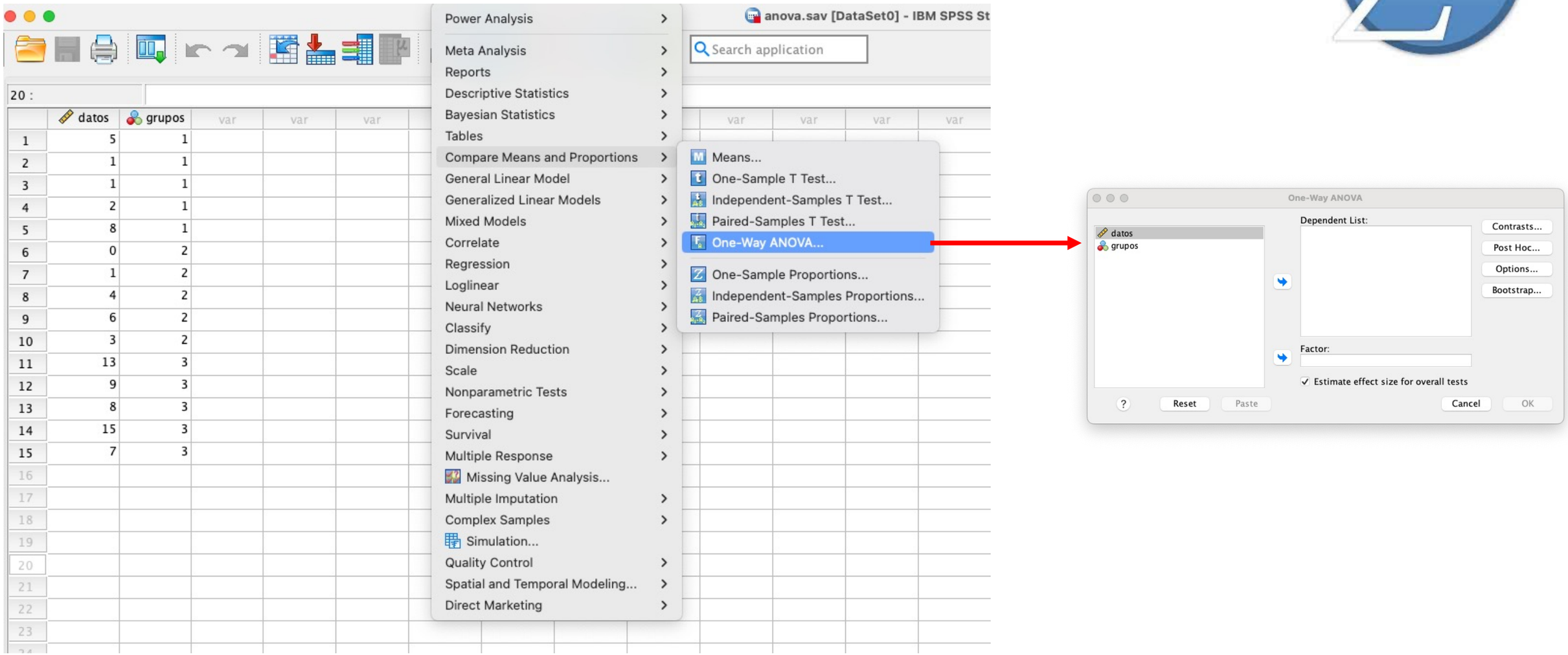


	datos	grupos	var
1	5	1	
2	1	1	
3	1	1	
4	2	1	
5	8	1	
6	0	2	
7	1	2	
8	4	2	
9	6	2	
10	3	2	
11	13	3	
12	9	3	
13	8	3	
14	15	3	
15	7	3	
16			
17			
18			

VI. Cálculo de ANOVA con R y SPSS

SPSS

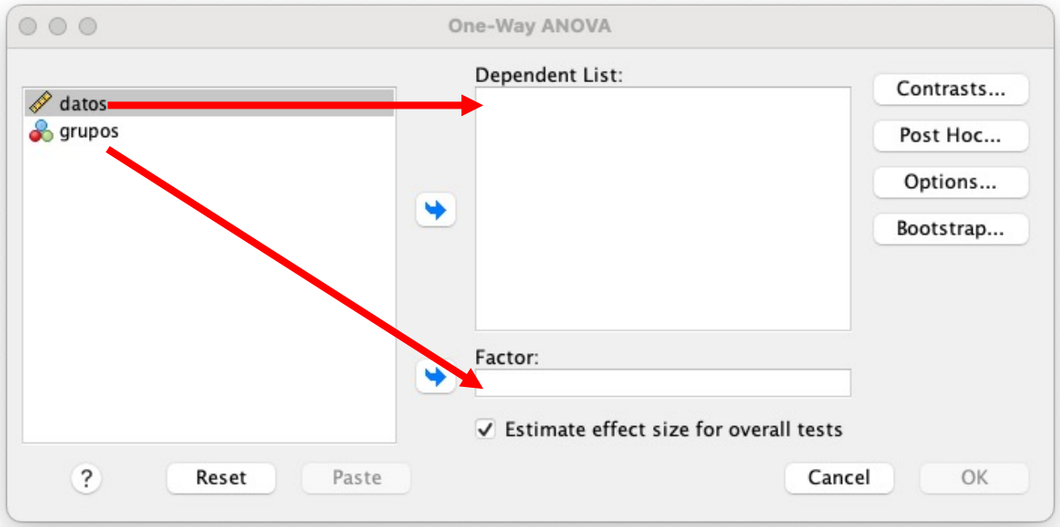
2º Analizar > Comparar Medias y Proporciones > ANOVA de una vía (1)

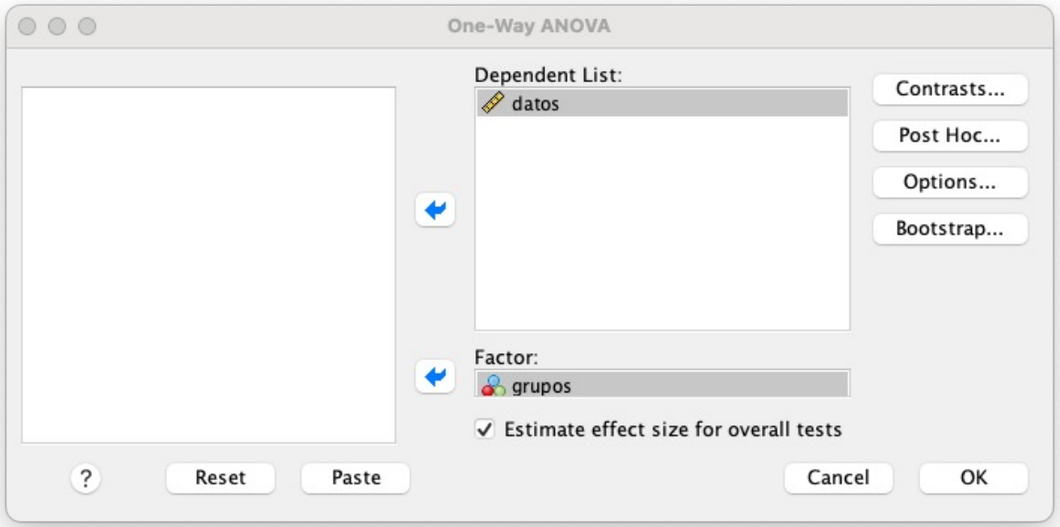


VI. Cálculo de ANOVA con R y SPSS

SPSS

2º Analizar > Comparar Medias y Proporciones > ANOVA de una vía (2)

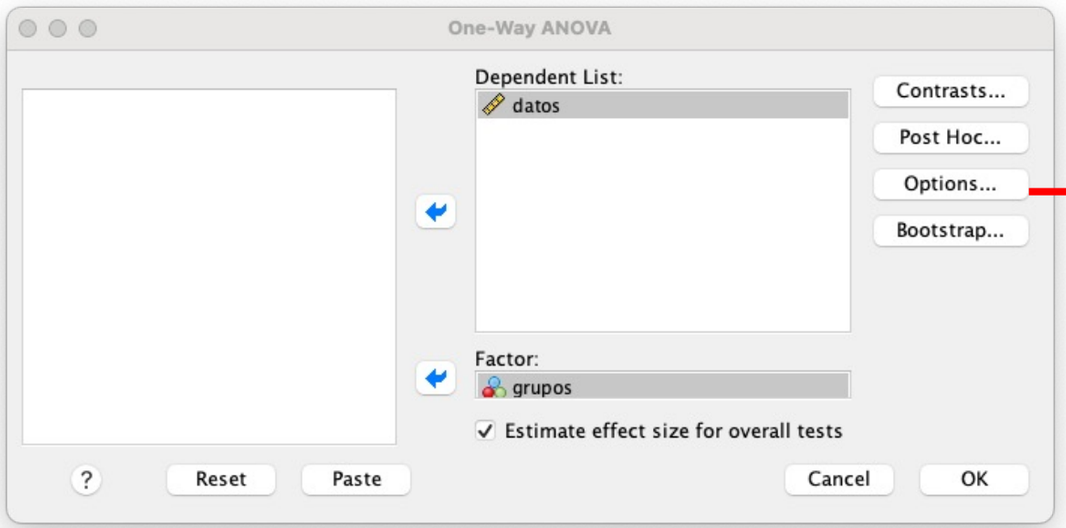


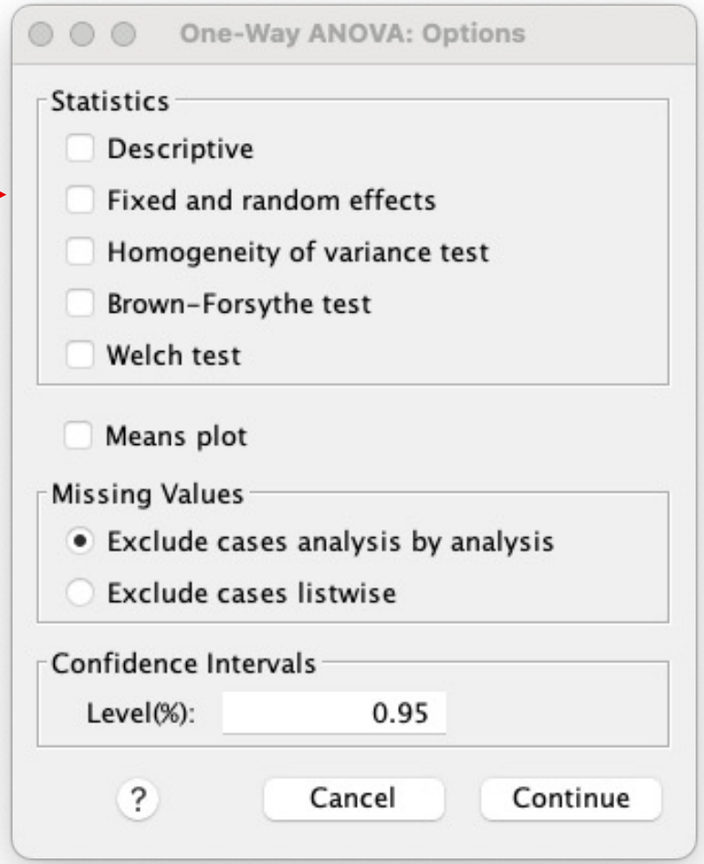


VI. Cálculo de ANOVA con R y SPSS

SPSS

2º Analizar > Comparar Medias y Proporciones > ANOVA de una vía (3)

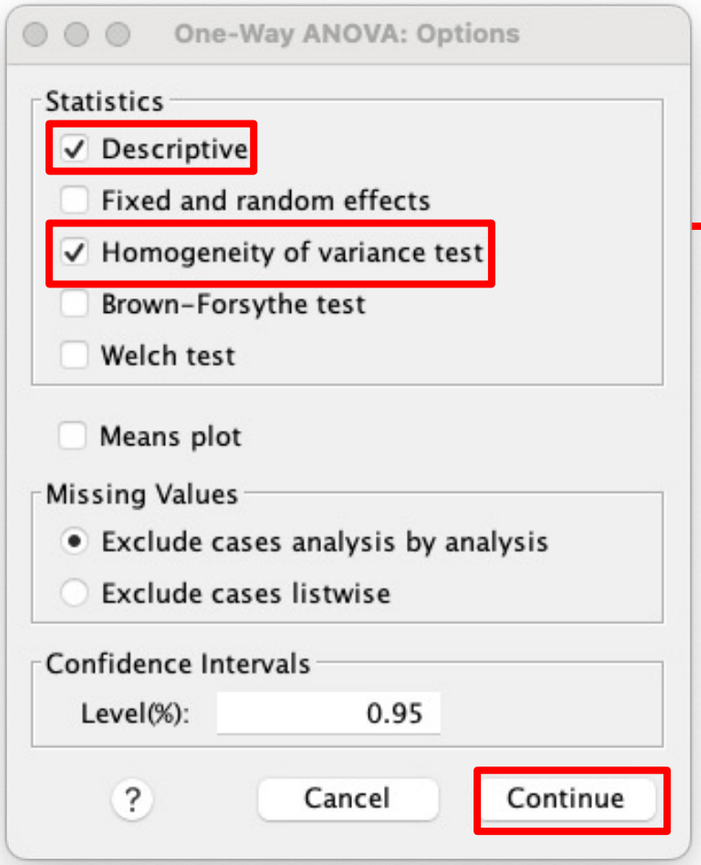


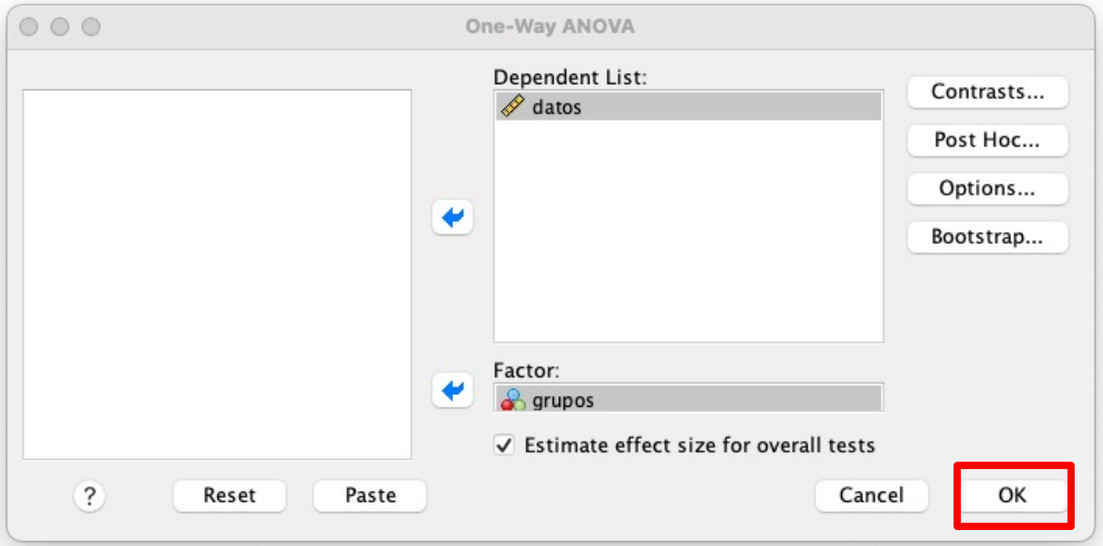


VI. Cálculo de ANOVA con R y SPSS

SPSS

2º Analizar > Comparar Medias y Proporciones > ANOVA de una vía (4)





Inicialmente asumimos que las varianzas de los grupos son iguales
Si el resultado de la prueba de Levene indica lo contrario rehacemos el análisis

VI. Cálculo de ANOVA con R y SPSS

SPSS

2º Analizar > Comparar Medias y Proporciones > ANOVA de una vía (5)

➔ **Oneway**

Descriptives

datos

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
G1	5	5.40	4.159	1.860	.24	10.56	1	11
G2	5	2.80	2.387	1.068	-.16	5.76	0	6
G3	5	10.40	3.435	1.536	6.13	14.67	7	15
Total	15	6.20	4.539	1.172	3.69	8.71	0	15

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
datos	Based on Mean	1.227	2	12	.327
	Based on Median	.607	2	12	.561
	Based on Median and with adjusted df	.607	2	10.156	.564
	Based on trimmed mean	1.161	2	12	.346

No se viola el supuesto de igualdad de varianzas

Brown–Forsythe (1974) (Desviaciones absolutas de la mediana)
Esta prueba es algo más robusta

VI. Cálculo de ANOVA con R y SPSS

SPSS

2º Analizar > Comparar Medias y Proporciones > ANOVA de una vía (6)

ANOVA

datos

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	149.200	2	74.600	6.431	.013
Within Groups	139.200	12	11.600		
Total	288.400	14			

→ Las medias no son iguales

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval Lower	Upper
datos	Eta-squared	.517	.036	.696
	Epsilon-squared	.437	-.124	.645
	Omega-squared Fixed-effect	.420	-.115	.629
	Omega-squared Random-effect	.266	-.054	.459

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

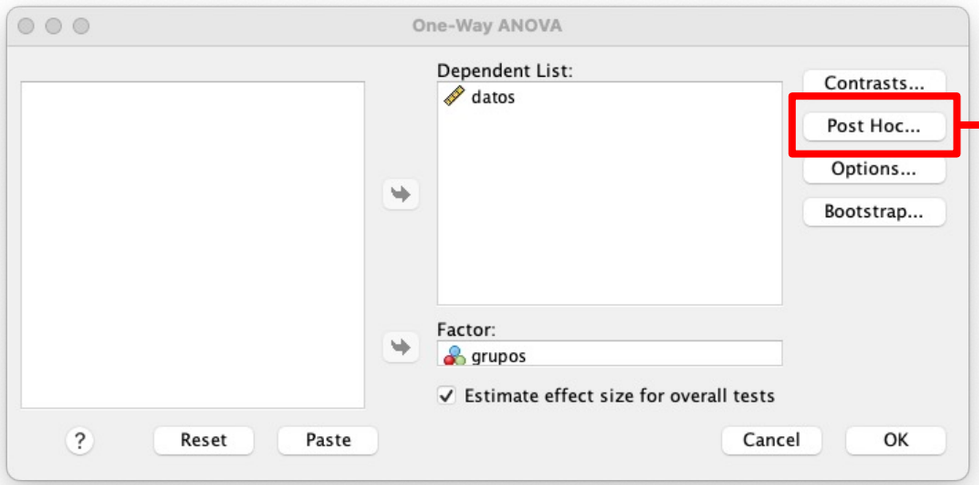
b. Negative but less biased estimates are retained, not rounded to zero.

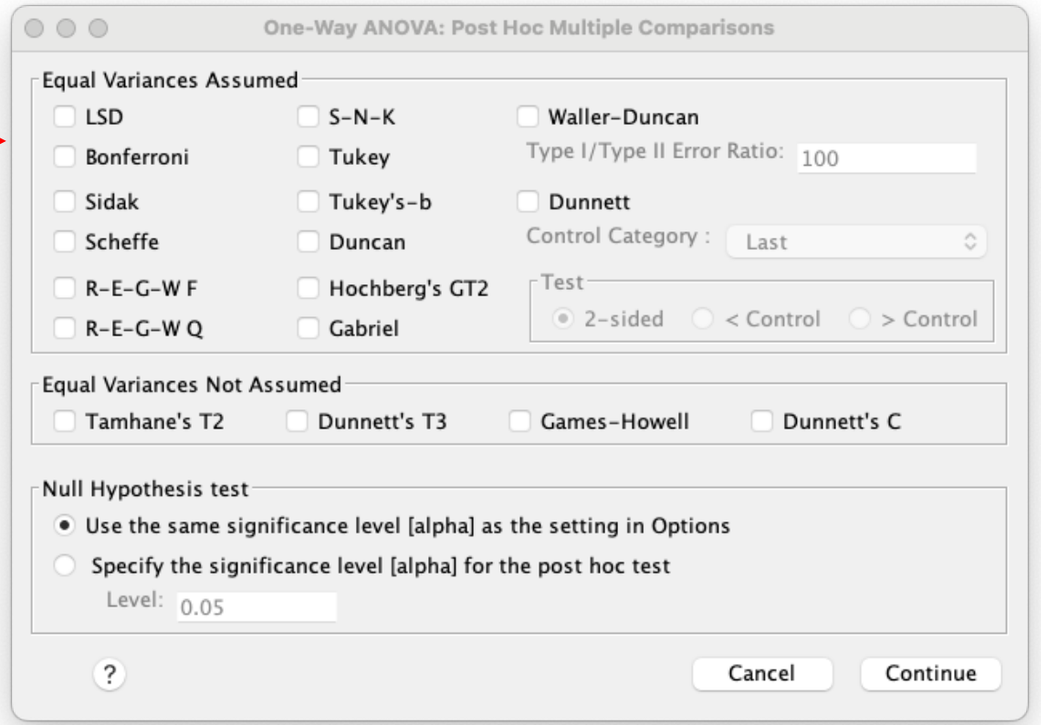
→ Estimaciones del tamaño del efecto

VI. Cálculo de ANOVA con R y SPSS

SPSS

3º Comparaciones Múltiples (1)





VI. Cálculo de ANOVA con R y SPSS

SPSS

3º Comparaciones Múltiples (2)

Post Hoc Tests

Multiple Comparisons

Dependent Variable: datos

	(I) grupos	(J) grupos	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	G1	G2	2.600	2.154	.472	-3.15	8.35
		G3	-5.000	2.154	.091	-10.75	.75
	G2	G1	-2.600	2.154	.472	-8.35	3.15
		G3	-7.600*	2.154	.011	-13.35	-1.85
	G3	G1	5.000	2.154	.091	-.75	10.75
		G2	7.600*	2.154	.011	1.85	13.35
Bonferroni	G1	G2	2.600	2.154	.752	-3.39	8.59
		G3	-5.000	2.154	.116	-10.99	.99
	G2	G1	-2.600	2.154	.752	-8.59	3.39
		G3	-7.600*	2.154	.012	-13.59	-1.61
	G3	G1	5.000	2.154	.116	-.99	10.99
		G2	7.600*	2.154	.012	1.61	13.59

*. The mean difference is significant at the 0.05 level.

VI. Cálculo de ANOVA con R y SPSS

SPSS

3º Comparaciones Múltiples (2)

Post Hoc Tests

Multiple Comparisons

Dependent Variable: datos

	(I) grupos	(J) grupos	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	G1	G2	2.600	2.154	.472	-3.15	8.35
		G3	-5.000	2.154	.091	-10.75	.75
	G2	G1	-2.600	2.154	.472	-8.35	3.15
		G3	-7.600*	2.154	.011	-13.35	-1.85
	G3	G1	5.000	2.154	.091	-.75	10.75
		G2	7.600*	2.154	.011	1.85	13.35
Bonferroni	G1	G2	2.600	2.154	.752	-3.39	8.59
		G3	-5.000	2.154	.116	-10.99	.99
	G2	G1	-2.600	2.154	.752	-8.59	3.39
		G3	-7.600*	2.154	.012	-13.59	-1.61
	G3	G1	5.000	2.154	.116	-.99	10.99
		G2	7.600*	2.154	.012	1.61	13.59

*. The mean difference is significant at the 0.05 level.

Homogeneous Subsets

datos

	grupos	N	Subset for alpha = 0.05	
			1	2
Tukey HSD[a]	G2	5	2.80	
	G1	5	5.40	5.40
	G3	5		10.40
	Sig.		.472	.091

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

VI. Cálculo de ANOVA con R y SPSS

SPSS

Realice el análisis de la varianza con los siguientes datos en SPSS

Yi	Y_G1	Y_G2	Y_G3
Y_1	4	2	9
Y_2	5	2	10
Y_3	5	3	10
Y_4	6	3	11
Y_5	7	4	12



VII. Informe de Resultados

VII. Informe de Resultados

En el informe de resultados debe aparecer:

- Los niveles de la variable independiente
- Las medias y desviaciones estándar de los grupos
- La dirección de la diferencia.
- El valor de F y los grados de libertad en los que se evaluó
- Si la probabilidad alcanzó o no su nivel alfa: $p < .05$.
- El tamaño del efecto y el estadístico utilizado para realizar la evaluación

VII. Informe de Resultados

Un ejemplo con nuestros datos:

Participaron 15 deportistas que se asignaron al azar a uno de los tres grupos de entrenamiento: (a) grupo 1, entrenamiento de fuerza tradicional; (b) grupo 2, entrenamiento pliométrico y; (c) grupo 3, entrenamiento combinado. Se evaluó su nivel de … tras el entrenamiento mediante la prueba … Se observaron diferencias significativas entre las puntuaciones medias de los tres grupos con un tamaño del efecto grande ($F(2,12) = 6.43$, $p < .05$, $\omega^2 = .42$). Los participantes en el grupo 3, de entrenamiento combinado obtuvieron una puntuación media ($M = 10.4$, $SD = 3.43$) significativamente superior que los participantes en el grupo 1 de entrenamiento de fuerza tradicional ($M = 2.8$, $SD = 2.39$). Sin embargo, no se observaron diferencias significativas en las puntuaciones obtenidas entre los participantes del grupo 3 y los del grupo 2 de entrenamiento pliométrico ($M = 5.4$, $SD = 4.16$).

VII. Informe de Resultados

Un ejemplo con nuestros datos:

Participaron 15 deportistas que se asignaron al azar a uno de los tres grupos de entrenamiento: (a) grupo 1, entrenamiento de fuerza tradicional; (b) grupo 2, entrenamiento pliométrico y; (c) grupo 3, entrenamiento combinado. Se evaluó su nivel de … tras el entrenamiento mediante la prueba … Se observaron diferencias significativas entre las puntuaciones medias de los tres grupos con un tamaño del efecto grande

$$(F(2,12) = 6.43, p < .05, \omega^2 = .42).$$

Los participantes en el grupo 3 de entrenamiento combinado obtuvieron una puntuación media (M = 10.4, SD = 3.43) significativamente superior que los participantes en el grupo 1 de entrenamiento de fuerza tradicional (M = 2.8, SD = 2.39). Sin embargo, no se observaron diferencias significativas en las puntuaciones obtenidas entre los participantes del grupo 3 y los del grupo 2 de entrenamiento pliométrico (M = 5.4, SD = 4.16).

gl entre grupos, gl intra grupos

valor de F

tamaño del efecto

Referencias

Agresti, A., & Franklin, C. (2014). *Statistics The Art & Science of learning from Data*. Pearson.

Doncaster, C. P., & Davey, A. J. H. (2007). *Analysis of variance and covariance: How to choose and construct models for the life sciences* (Vol. 140, Issue 4). Cambridge.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. John wiley & Sons.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). Elsevier.