UNIDAD DIDÁCTICA 5

INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

(3ª parte: Introducción al Análisis de la Varianza)

OBJETIVO

En este apartado de la UD 5 se introduce el **Análisis de la Varianza** (**ANOVA**):

1. para estudiar el efecto que uno o dos factores tienen sobre la media de una variable respuesta.

2. Se aplicar el método a partir de cálculos con Statgraphics.

Contenidos

Contenidos

Contenidos

4. ANOVA con dos factores

   4.1 Concepto de efecto simple y de interacción doble

   4.2 Un ejemplo

   4.3 Descomposición de la suma de cuadrados. Test F

   4.4 Análisis con Statgraphics

Introducción

- El ANOVA consiste en descomponer la variabilidad total observada en unos datos en:
  - términos asociados a los efectos de cada factor estudiado y a sus posibles interacciones,
  - más una parte residual, con la que después se compararán las primeras para estudiar su significación estadística.

Idea intuitiva del ANOVA
Ejemplo:

Se estudian los efectos que el tipo de procesador y la temperatura de la sala tienen sobre la velocidad de ejecución de un tipo de programa.

Se van a comparar tres tipos de procesador (A, B y C) y dos temperaturas (15 y 40).

Se decide ejecutar 3 programas por cada una de las 6 combinaciones posibles de procesador y temperatura.

- La velocidad de ejecución es la **variable respuesta**.

- Hay dos **factores**:
  - el tipo de procesador, que es un **factor cualitativo** para el que se analizan **3 variantes**,
  - la temperatura, que es un **factor cuantitativo** para el que se analizan **2 niveles**.

- En total hay 6 **tratamientos** posibles, obtenidos al combinar los 3 tipos de procesador con las 2 temperaturas.
- **A cada tratamiento le corresponde una población** en la que está definida la variable respuesta.
- Así, la primera población esta constituida por todos los programas que pueden ejecutarse con el procesador tipo A y la temperatura 15.
- Los 3 programas asignados a cada tratamiento constituyen una **muestra** de la población correspondiente. El conjunto de los resultados de las 6 muestras son los **datos** disponibles para el estudio.

Las cuestiones que se estudian son:

- En promedio para las 2 temperaturas estudiadas ¿hay diferencia entre las velocidades medias de los 3 procesadores? (¿Existe un **efecto del factor procesador** sobre la media de la respuesta?)

- En promedio para los 3 procesadores estudiados, ¿varía la velocidad media al variar la temperatura? (¿Existe un **efecto del factor temperatura** sobre la media de la respuesta?)

- El efecto de la temperatura sobre la velocidad ¿es diferente en el procesador B que en el procesador A o C? (¿Existe una **interacción entre los efectos de los dos factores temperatura y procesador**?)

- Se aplica el **Análisis de la Varianza** (**ANOVA**) para determinar qué efectos son significativos.

Caso 1º: los 18 valores X_{ijk} han resultado idénticos. ¡No hay variabilidad en los datos! La variabilidad total se mide en el ANOVA por la Suma de Cuadrados Total,

	Proc. A	Proc. B	Proc. C
T=15	5 5 5	5 5 5	5 5 5
T=40	5 5 5	5 5 5	5 5 5

En este caso 1º se tendrá que $SC_{total} = \sum_{i,j,k} (X_{ijk} - \overline{X})^2$ es cero.

Consecuencia: ni la temperatura ni el procesador influyen en la velocidad.

Caso 2º: ¡Sí que hay variabilidad en los datos!

	Proc. A	Proc. B	Proc. C
T=15	6 6 6	6 6 6	6 6 6
T=40	4 4 4	4 4 4	4 4 4

$$SC_{total} = \sum_{i,j,k} \left(x_{ijk} - \bar{x} \right)^2 = 18$$

- Se debe al efecto de la temperatura sobre la media

- En la tabla resumen del ANOVA:

$SC_{total}=18$ $SC_{temp}= 18$ $SC_{proc}=0$

Caso 3º: la SC_{total} es ahora:

$$SC_{total} = \sum_{i,j,k} (X_{ijk} - \overline{X})^2 = 66$$

	Proc. A			Proc. B			Proc. C		
T=15	5	5	5	7	7	7	9	9	9
T=40	3	3	3	5	5	5	7	7	7

- Calculando las Sumas de Cuadrados asociadas a cada efecto se obtiene:

$SC_{total}=66$ $SC_{temp}= 18$ $SC_{proc}=48$ $SC_{interacción}=0$

Caso 4º: la variabilidad total,

$$SC_{total} = \sum_{i,j,k} (X_{ijk} - \overline{X})^2 = 88{,}5$$

	Proc. A	Proc. B	Proc. C
T=15	4 4 4	9 9 9	8 8 8
T=40	3 3 3	4 4 4	5 5 5

- Calculando las Sumas de Cuadrados asociadas a cada efecto se obtiene:

SC_{total}=88,5 SC_{temp}= 40,5 SC_{proc}=36 $SC_{interacción}$=12

Caso 5º: Los efectos de los factores vienen parcialmente enmascarados por la variabilidad residual originada por factores no controlados.

	Proc. A	Proc. B	Proc. C
T=15	3,2 2,9 3,8	8,4 9,7 8,6	7,7 7,9 6,8
T=40	2,9 3,2 3,5	3,6 2,7 4,2	5,7 5,6 5,9

- Los programas ejecutados con idénticos temperatura y procesador no son exactamente iguales y, por tanto, no dan exactamente las mismas velocidades.

- Calculando las Sumas de Cuadrados asociadas a cada efecto se obtendría:

$SC_{total}=91,885$ $SC_{temp}= 26,1606$ $SC_{proc}=40,17$ $SC_{interac}=22,1011$ $SC_{residual}=3,453$

Idea intuitiva del ANOVA

- Los "grados de libertad" totales son siempre el número de datos menos 1 (18-1=17 en el ejemplo),
- Se realiza una descomposición de estos grados de libertad en términos asociados a cada efecto.
- Los grados de libertad asociados al efecto de un factor son el número de variantes del factor menos 1 (gl_{temp}=2-1=1 y gl_{proc}=3-1=2).
- Los grados de libertad de una interacción doble, son el producto de los grados de libertad de los factores correspondientes (gl_{interc}=1x2=2).
- Los grados de libertad residuales se calculan:

  gl_{resid}=17-1-2-2=12

Idea intuitiva del ANOVA

- La comparación de la "varianza" asociada a cada efecto con la varianza residual permite estudiar si dicho efecto es significativo.
- Dicha varianza se estima dividiendo su Suma de Cuadrados por sus correspondientes grados de libertad. El resultado se denomina **Cuadrado Medio**.

(El CM_{total} es la varianza de los datos).

Cuadrado Medio Residual $= \dfrac{\text{SCResidual}}{\text{g.lib.residuales}}$

Es una estimación de la varianza σ^2 de cada una de las poblaciones en estudio, asumiendo que dichas poblaciones tienen todas la misma σ^2 (o del promedio de dichas varianzas en el caso de que difieran de unas poblaciones a otras).

El Cuadrado Medio de un efecto:

- Es también una estimación de σ^2 (independiente de la anterior) si el efecto no existe en la población. En ese caso la estimación del CM_{efecto} no diferirá de $CM_{residual}$.
- Cuando es admisible que existe el efecto en la población, su cuadrado medio es mayor que σ^2 y por tanto, será significativamente mayor que el $CM_{residual}$.

- Para estudiar si el CM_{efecto} difiere significativamente del $CM_{residual}$, lo que implicaría la existencia de un efecto real a nivel poblacional, se comprueba si el cociente $CM_{efecto}/CM_{residual}$ (al que se denomina F-ratio) es <u>demasiado elevado</u> para ser una F de Fisher con los grados de libertad correspondientes.

- Se calcula el valor *p-value* asociado (que no es más que la probabilidad de que una F, con los grados de libertad correspondiente, sea mayor que el valor obtenido para la F_{ratio}).

- Los resultados se disponen en una tabla resumen del ANOVA, tal como se recoge a continuación para los datos del Caso 5º:

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Temperatura	26,1606	1	26,1606	90,91	0,0000
Procesador	40,17	2	20,085	69,79	0,0000
Temp*Proc	22,1011	2	11,0506	38,40	0,0000
RESIDUAL	3,45333	12	0,287778		
TOTAL	91,885	17			

- Del análisis de la tabla se desprende que los efectos simples de temperatura y de procesador son muy significativos estadísticamente (*p-values* menores que 0,01).
- El efecto de la interacción temperatura*procesador es también significativo estadísticamente (*p-value* inferior a 0,01).

ANOVA con un factor a dos niveles

Para comparar los tiempos medios de depuración de dos sistemas se han tomado dos muestras aleatorias de tamaño 10.

Los datos (en minutos) se dan en la tabla siguiente.

Resultados

	Sistema 1	Sistema 2
	63	62
	61	59
	57	61
	61	60
	58	58
	60	63
	64	61
	60	58
	59	63
	58	60
Media $\bar{X}$	60,1	60,5
Desv.Típica s	2,23	1,84

Fundamento del análisis estadístico

Pregunta fundamental que se pretende contestar en el estudio realizado:

¿hay diferencias entre ambos sistemas en la distribución del tiempo medio de depuración?

Dicho de otra forma:

¿es admisible la hipótesis $m_1=m_2$, o por el contrario, los datos evidencian de forma significativa que m_2 es diferente de m_1?

Aplicando el ANOVA es posible determinar si la H_0: $m_1=m_2$ es admisible

Descomposición de la suma de cuadrados. Test F

- De acuerdo con las ideas intuitivas dadas en el apartado 2, la variabilidad total de los 20 datos se descompondrá según el siguiente esquema:

Variabilidad total en los datos =	**Variabilidad debida a diferencias entre sistemas (efecto del factor sistema)**	**+**	**Variabilidad residual (variabilidad dentro de cada sistema)**

Descomposición de la suma de cuadrados. Test F

Suma de Cuadrados Total: mide la variabilidad total en los datos, y no es más que la suma de los cuadrados de las desviaciones de cada dato **respecto a la media general** del experimento.

SCTotal=76,2

Esta SCTotal tiene un número de grados de libertad que es igual al número total de datos menos uno: gltotal=20-1=**19**.

Descomposición de la suma de cuadrados. Test F

Suma de Cuadrados del Factor: mide la variabilidad en los datos asociada al efecto del factor Sistema sobre la media (o sea la diferencia de medias entre los 2 sistemas).

SCFactor=0,8

La SCFactor tiene un número de grados de libertad que es igual al número de sistemas menos uno: glfactor=2-1=**1**.

Descomposición de la suma de cuadrados. Test F

Suma de Cuadrados Residual: mide la variabilidad debida a factores no controlados.

Se obtiene hallando la suma de los cuadrados de las desviaciones de cada dato **respecto a la media de la configuración correspondiente**.

Descomposición de la suma de cuadrados. Test F

Suma de Cuadrados Residual:

Se verifica: SCTotal=SCFactor+SCResidual

SCResidual=SCTotal-SCFactor=76,2-0,8=75,4

Los grados de libertad asociados a la SCResidual se obtienen por diferencia entre los gltotal y los glfactor (glresid=19-1=**18**)

Descomposición de la suma de cuadrados. Test F

Test F

- La hipótesis nula es que no hay diferencias ($m_1 = m_2$).
- Si H_0 es cierta

F-ratio=CMFactor/CMResidual

se distribuye como una F de Fisher con 1 y 18 grados de libertad.

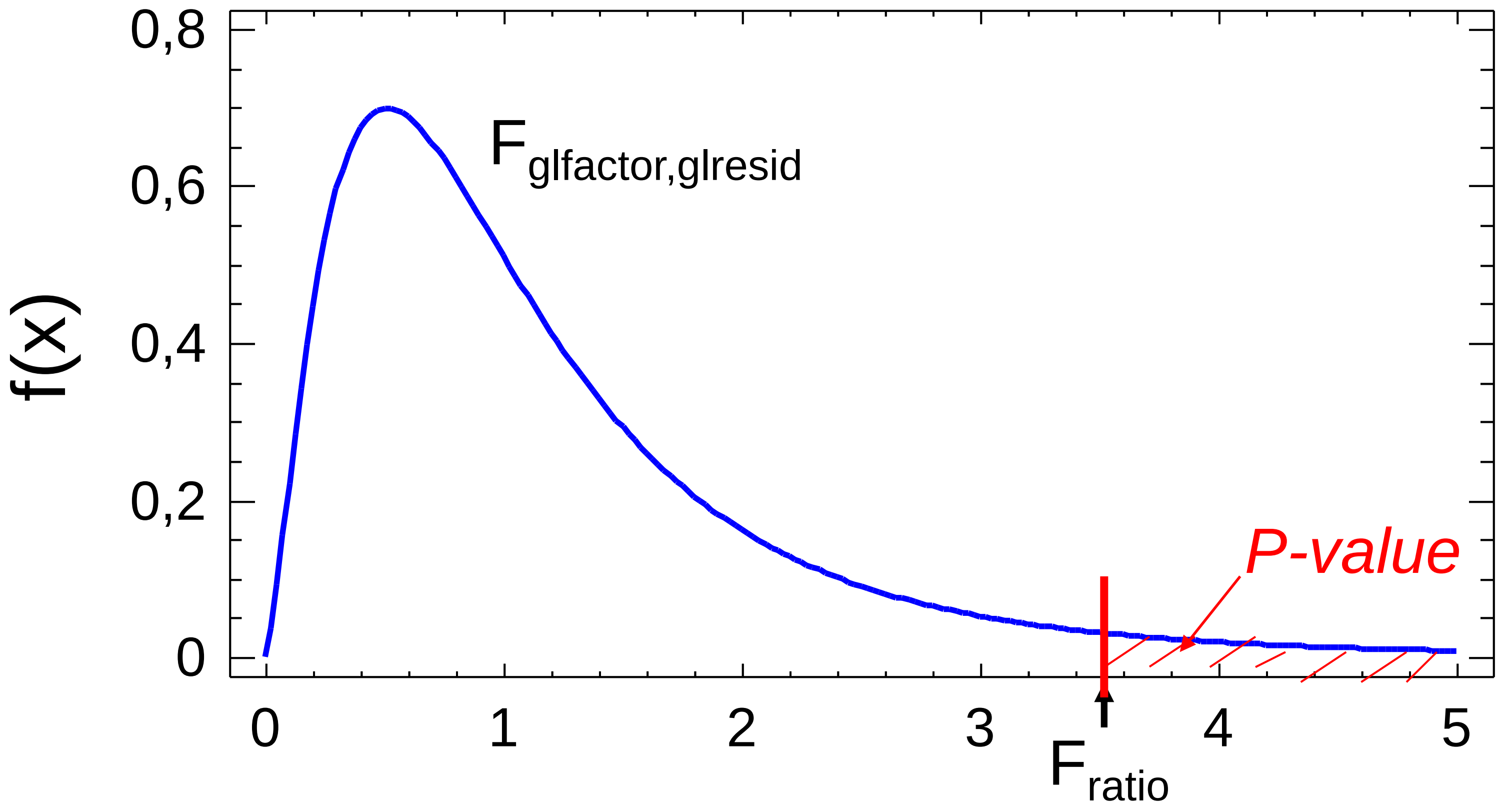
- Si no es cierta H_0 (los dos sistemas tienen medias diferentes) la F-ratio tiende a ser mayor que $F_{1,18}$

Descomposición de la Suma de Cuadrados. Test F

1. Si se dispone de una tabla F, se obtiene de la misma el valor F(α). Si F-ratio > F(α) se rechaza la hipótesis nula y se dice que el efecto es significativo.

2. Prob($F_{glfactor,\ glresid}$ > F-ratio)= *p-value*. Si *p-value* <α se rechaza la hipótesis nula y se dice que el efecto es significativo estadísticamente.

Descomposición de la Suma de Cuadrados. Test F



Descomposición de la Suma de Cuadrados. Test F

Los resultados se sintetizan en la **Tabla Resumen del ANOVA**

Origen variabilidad	Sumas de Cuadrados	Grados libertad	Cuadrado Medio	F-ratio
Total	76,2	19		
Sistema	0,8	1	0,8	0,19
Residual	75,4	18	4,1888	

Descomposición de la Suma de Cuadrados. Test F

- Efecto del sistema F-ratio=0,19.
- En la distribución F con 1 y 18 grados de libertad F(5%)=4,41 y F(1%)=8,29.
- Como F-ratio=**0,19** < 4,41, el efecto del sistema sobre el tiempo medio de depuración no es significativo.
- Por tanto no hay diferencias entre las dos medias.

ANOVA con un factor a más de dos niveles

Un ejemplo:

Se desea estudiar el efecto que la configuración (tres posibles A, B y C) tiene sobre el rendimiento medio.

Cada configuración se ha ensayado 8 veces.

ANOVA con un factor a 3 niveles

Los valores del rendimiento han sido:

Configuración A	Configuración B	Configuración C
48,5	27,4	48,8
38,4	29,9	51,8
51,2	26,8	40,8
43	34,4	39,3
50,8	36,1	34,9
38,7	18	45,7
61,6	31,1	32,4
33,3	32,2	47,2

Descomposición de la suma de cuadrados. Test F

- La variabilidad total de los 24 datos se descompone :

Variabilidad total en los datos =	**Variabilidad debida a diferencias entre configuraciones (efecto del factor configuración)**	**+**	**Variabilidad residual (variabilidad dentro de cada configuración)**

Descomposición de la Suma de Cuadrados. Test F

Los resultados se sintetizan en la **Tabla Resumen del ANOVA**

Origen variabilidad	Sumas de Cuadrados	Grados libertad	Cuadrado Medio	F-ratio
Total	2317,52	23		
Configuración	1184,43	2	592,215	10,97
Residual	1133,09	21	53,96	

Descomposición de la Suma de Cuadrados. Test F

- Efecto de la configuración F-ratio=10,97.
- En la distribución F con 2 y 21 grados de libertad F(5%)=3,47 y F(1%)=5,78.
- Como F-ratio=**10,97** > 5,78, el efecto de la configuración sobre el rendimiento medio es muy significativo.
- Por tanto hay diferencias de rendimiento medio entre las configuraciones.

Nota sobre el análisis de residuos

- El residuo de un dato puede definirse como: la diferencia entre el valor realmente obtenido (el dato) y el valor medio predecible para el tratamiento correspondiente a esa observación a partir del conjunto de los datos disponibles:

residuo = dato - (predicción media para el tratamiento)

Nota sobre el análisis de residuos

- El residuo refleja cómo han afectado a la observación el conjunto de factores no contemplados explícitamente en el estudio.
- Statgraphics proporciona como complemento del ANOVA, los valores de los residuos de cada observación y diferentes gráficos para analizarlos.
- Si se cumplen todas las hipótesis básicas del ANOVA, los residuos deben tener distribución normal, con media cero y varianza igual a la σ^2 residual.

Nota sobre el análisis de residuos

Ejemplo

- Se introduce en los datos del ejemplo del rendimiento en función de la configuración, una anomalía: al último dato de configuración B se le suman 30 unidades “62,2“.

Nota sobre el análisis de residuos

Configuración A	Configuración B	Configuración C
48,5	27,4	48,8
38,4	29,9	51,8
51,2	26,8	40,8
43	34,4	39,3
50,8	36,1	34,9
38,7	18	45,7
61,6	31,1	32,4
33,3	**62,2**	47,2

La Tabla resumen del ANOVA con este dato erróneo resulta:

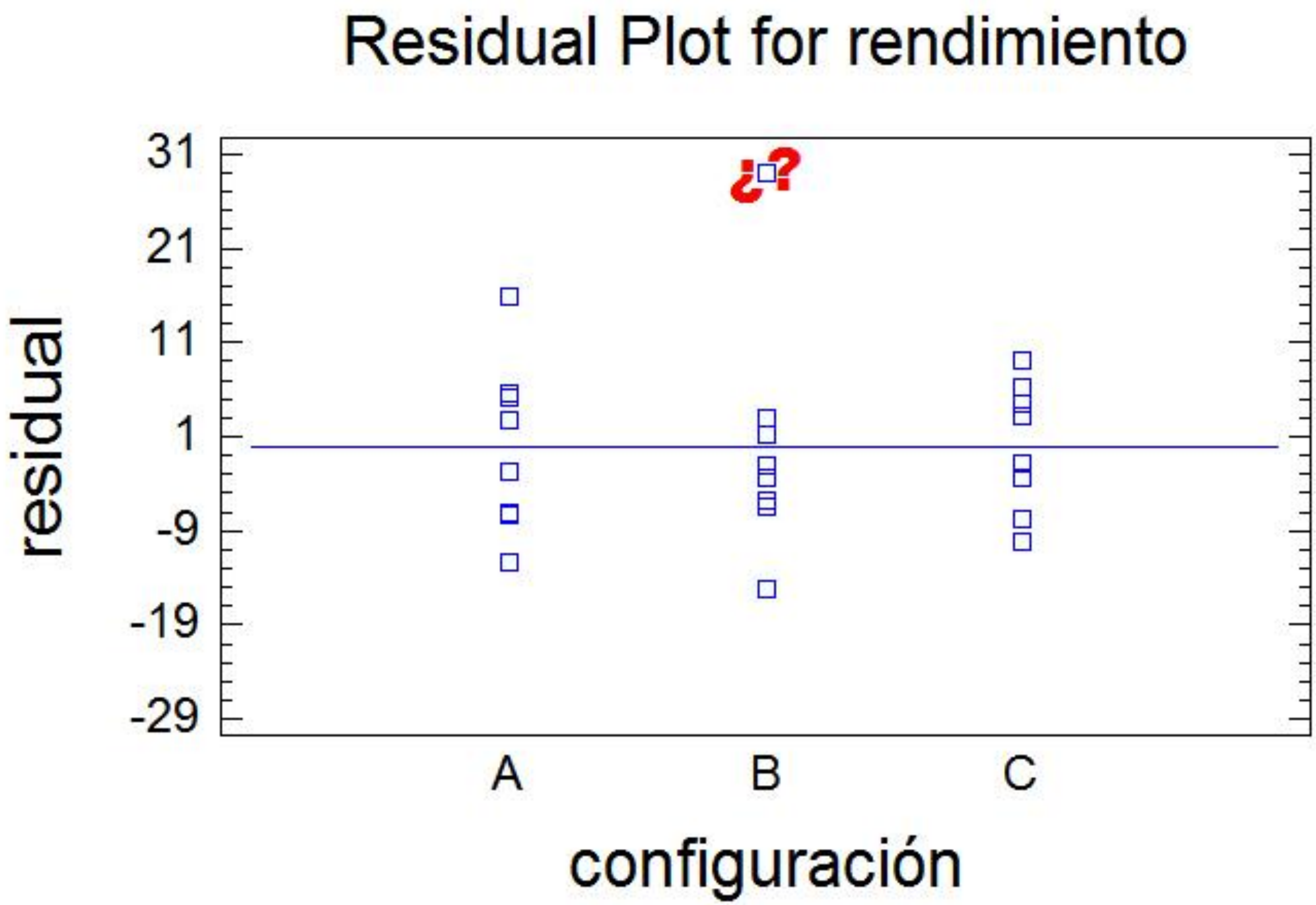
Nota sobre el análisis de residuos

Origen variabilidad	Sumas de Cuadrados	Grados libertad	Cuadrado Medio	F-ratio
Total	2756,27	23		
Configuración	672,93	2	336,465	3,39
Residual	2083,34	21	99,2065	

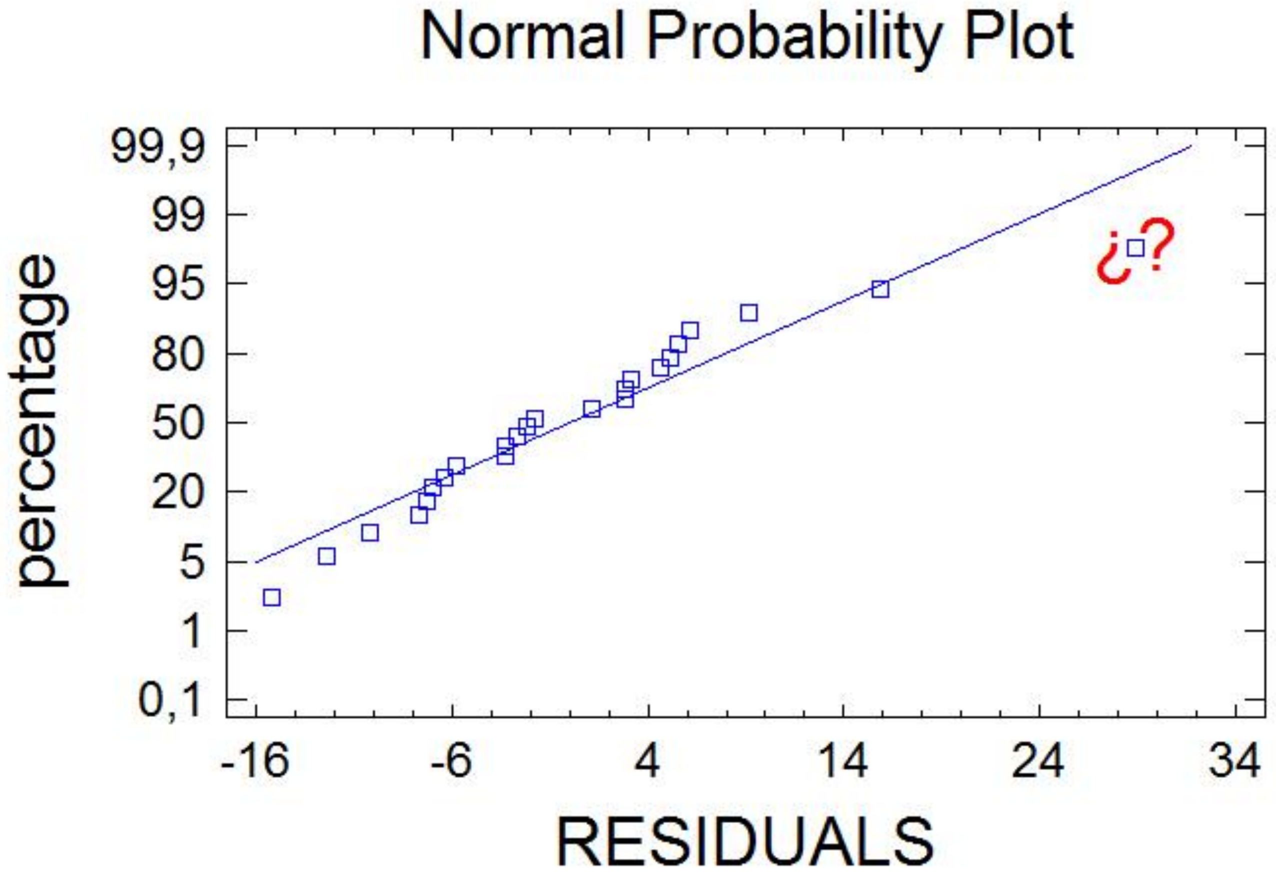
Nota sobre el análisis de residuos

- Comparando los resultados con los obtenidos con el dato correcto se aprecian grandes diferencias: la SCResidual ha pasado de valer 1133,09 a 2083,34, y la F-ratio ha pasado de 10,97 (muy significativa) a 3,39<F(5%)=3,47 (no significativa).
- Al hacer un gráfico de los residuos en función de la configuración se detecta con claridad la presencia de un residuo anómalo, que también se observa en una representación en papel probabilístico normal de los residuos.

Nota sobre el análisis de residuos



Nota sobre el análisis de residuos



Comparación de medias. Intervalos L.S.D.

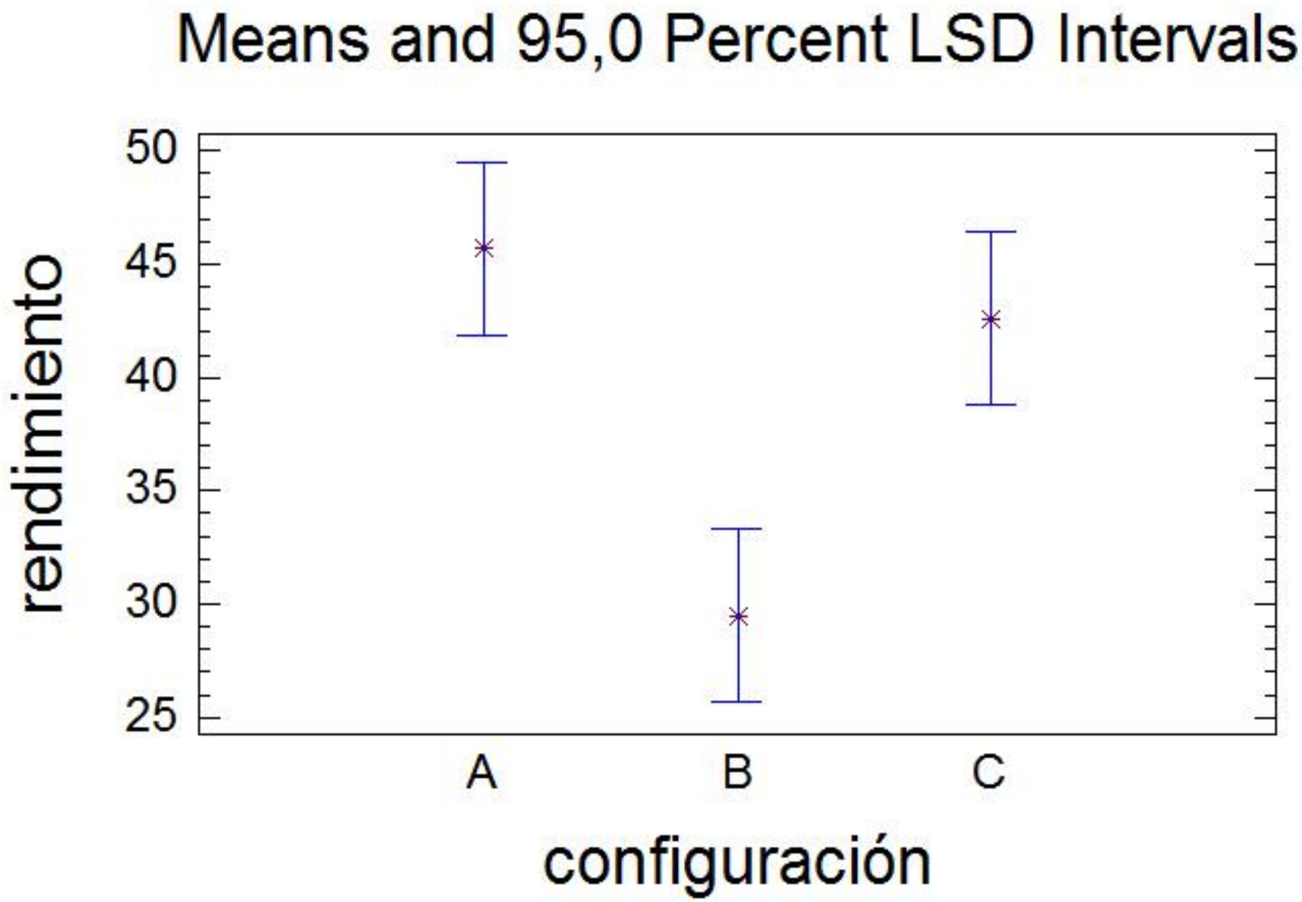
Ejemplo: el factor configuración tiene más de 2 variantes (A, B y C).

- Ha resultado muy significativo.

- Para estudiar entre qué configuraciones hay diferencias se va a aplicar intervalos LSD.

- Los intervalos se representan en la siguiente figura.

Comparación de medias. Intervalos L.S.D.



Comparación de medias. Intervalos L.S.D.

- Existe diferencia significativa entre la media de la configuración B y las de las otras dos configuraciones.
- El rendimiento medio es significativamente menor con la configuración B (su intervalo LSD no se solapa con los otros dos).
- No es significativa la diferencia al respecto entre las configuraciones A y C, dando mayor el rendimiento medio con ambas.

Análisis con Statgraphics

- Para obtener la tabla del ANOVA, el análisis de residuos y los intervalos LSD con Statgraphics se puede utilizar cualquier de las dos opciones siguientes:

  Compare...Analysis of Variance…One-Way ANOVA

  Compare…Analysis of Variance…Multifactor ANOVA

- La Tabla de ANOVA con la segunda opción resulta:

Análisis mediante el Statgraphics

Analysis of Variance for rendimiento

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:configuración	1184,43	2	592,215	10,98	0,0005
RESIDUAL	1133,09	21	53,9565		
TOTAL	2317,52	23			

El P-Value resulta <0,01, por lo que el efecto de la configuración sobre la media del rendimiento es muy significativo

Nota sobre factores cuantitativos

<u>Ejemplo</u>: El tiempo medio de respuesta de unas unidades de disco de un tipo determinado fluctúa en función de la carga. Con el fin de determinar esta fluctuación se ha realizado un experimento con tres niveles de carga (3, 5 y 7), realizando seis pruebas por cada uno de dichos niveles.

La tabla siguiente muestra los resultados.

Nota sobre factores cuantitativos

Carga=3	Carga=5	Carga=7
49,6	56	55,9
50,3	58,9	53,5
48,4	58,2	55,2
48,2	54,5	57,6
45,6	55,1	57
48,1	56,7	55,1

Nota sobre factores cuantitativos

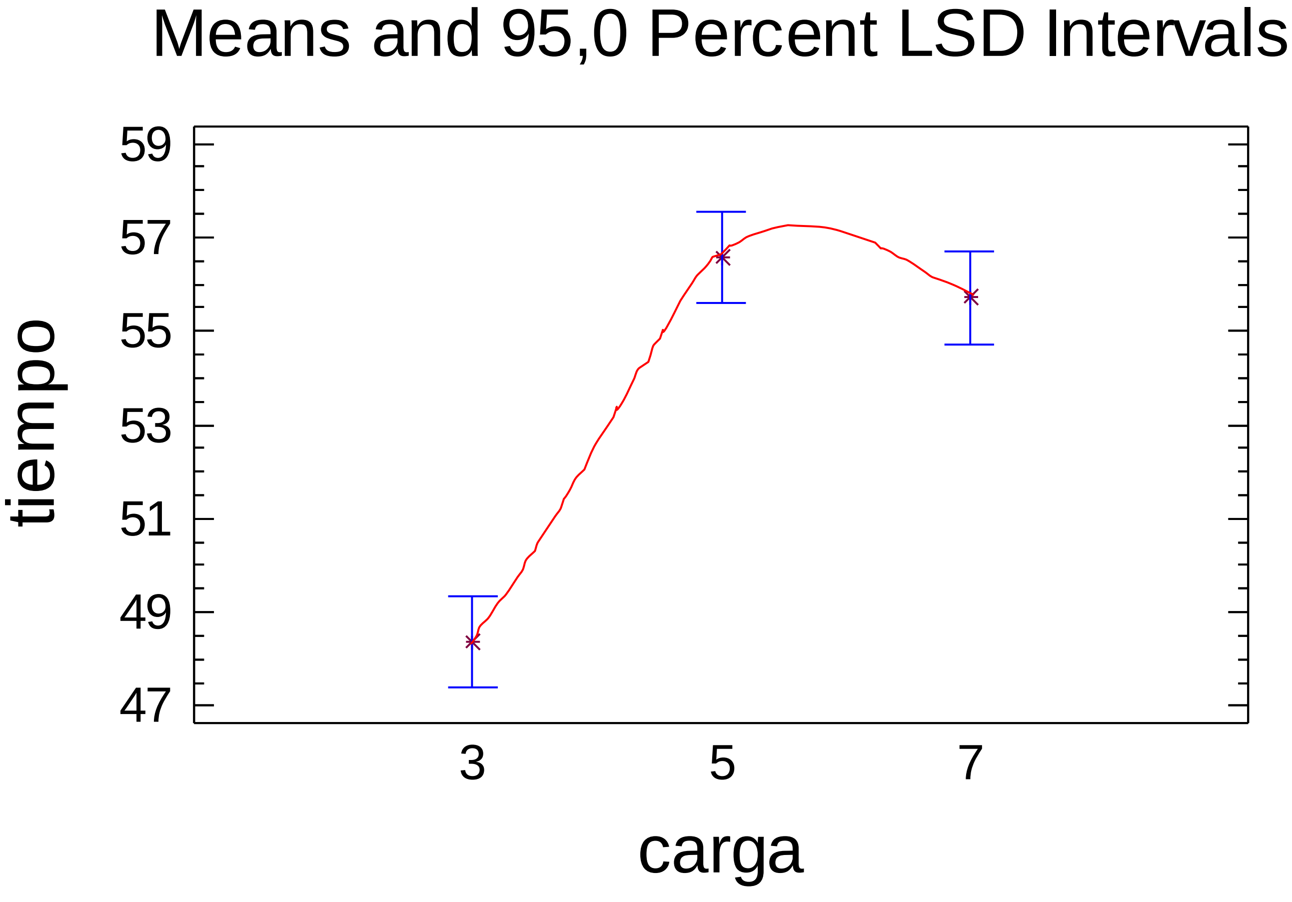
- El cuadro resumen del ANOVA, calculado con STATGRAPHICS, se recoge a continuación.
- Indica que el efecto de la carga sobre el tiempo medio de respuesta es muy significativo:

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
carga	243,97	2	121,985	47,31	0,0000<0,01
RESIDUAL	38,675	15	2,57833		
TOTAL	282,645	17			

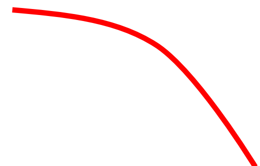
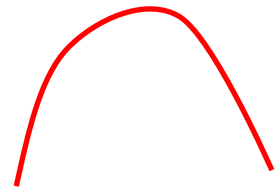
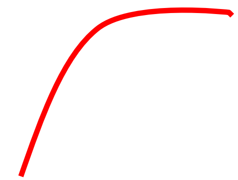
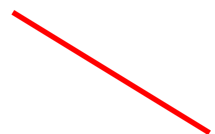
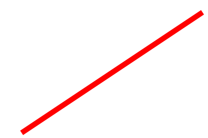
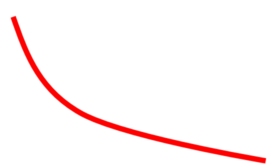
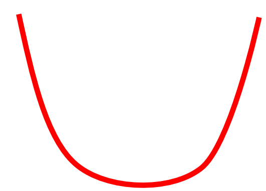
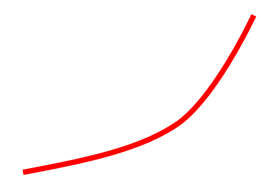
Nota sobre factores cuantitativos

- El gráfico de intervalos LSD para las medias de las 3 cargas se da a continuación.
- Se aprecia en la curva ajustada en la figura que, en el rango de cargas estudiado, el tiempo medio de respuesta tiene una tendencia general a crecer al hacerlo la carga (**efecto lineal positivo**), pero que este crecimiento es cada vez más lento llegando incluso a decrecer para cargas altas (**curvatura negativa**).
- Se observa en la figura que el tiempo de respuesta máximo se obtendría para una carga entre 5 y 7. (Esto puede precisarse más ajustando un modelo de regresión como los que se estudiarán en el tema correspondiente).

Nota sobre los factores cuantitativos



Nota sobre los factores cuantitativos

Posibles respuestas		Efecto lineal		
		<0	0	>0
Efecto cuadrático	<0			
	0			
	>0			

ANOVA con dos factores

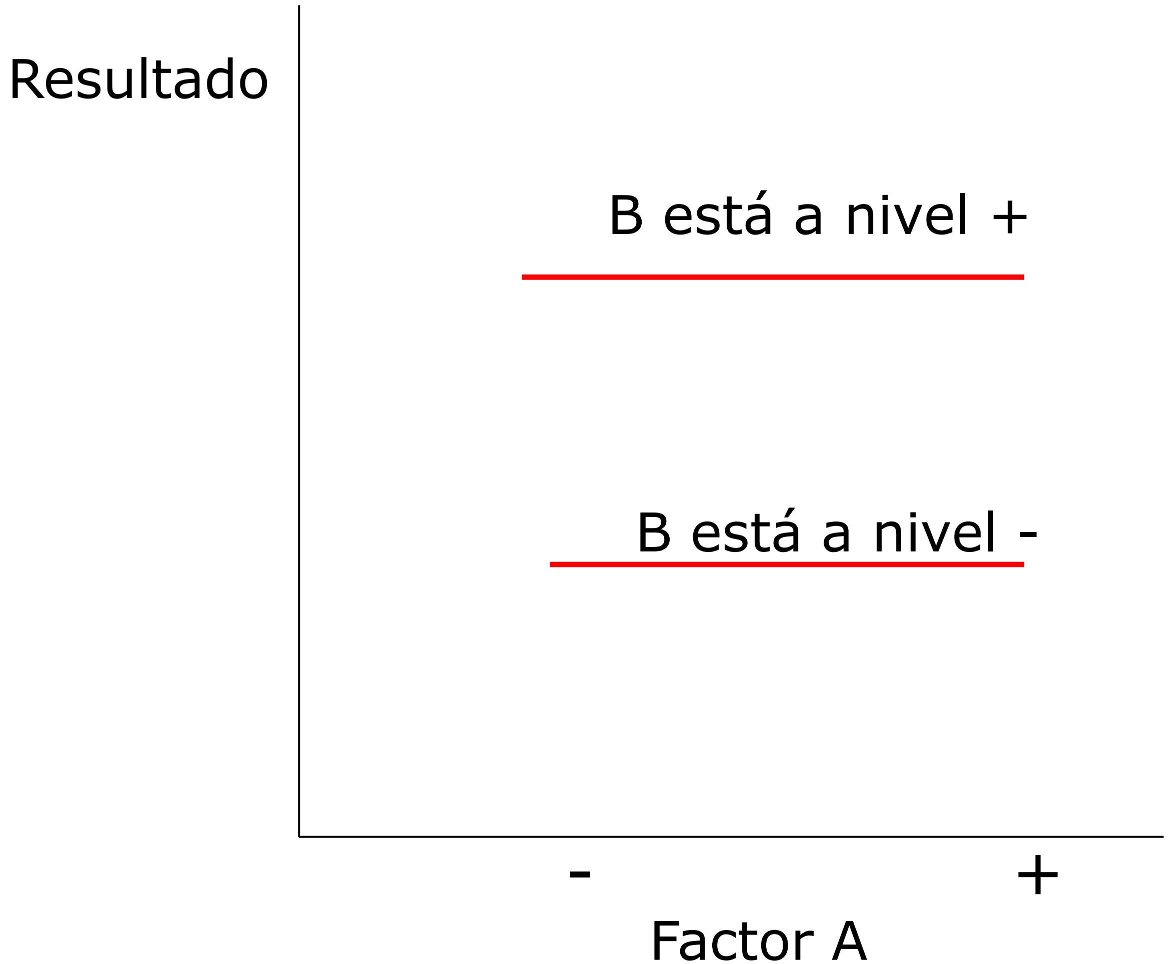
- Cuando se estudia sólo un único factor, el "efecto" del mismo sobre la media de la variable estudiada hace referencia a la existencia de diferencias entre las medias de las poblaciones asociadas a las diferentes variantes del factor.
- Cuando se estudian dos factores, aparecen los conceptos de "efectos simples" y de "interacción doble“.

Interacciones dobles

- Hay interacción doble, si el efecto de un factor cambia según la variante considerada del otro factor.

- Las figuras siguientes representan los distintos casos que se pueden dar entre un factor A y un factor B, cada uno a dos niveles, respecto a la interacción doble entre ellos.

Cuatro posibles situaciones sobre existencia de interacción doble

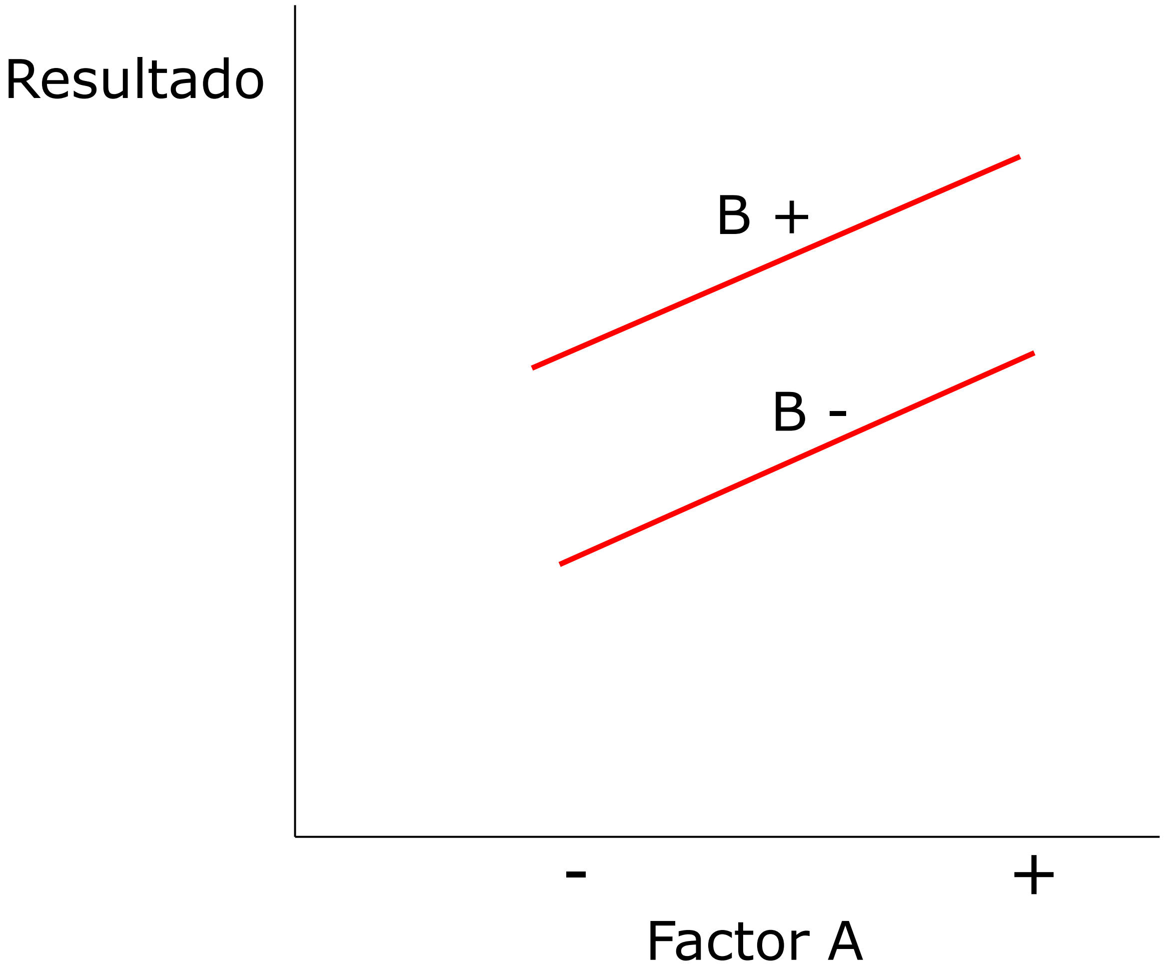


A no tiene efecto

B tiene efecto

No interacción

Cuatro posibles situaciones sobre existencia de interacción doble

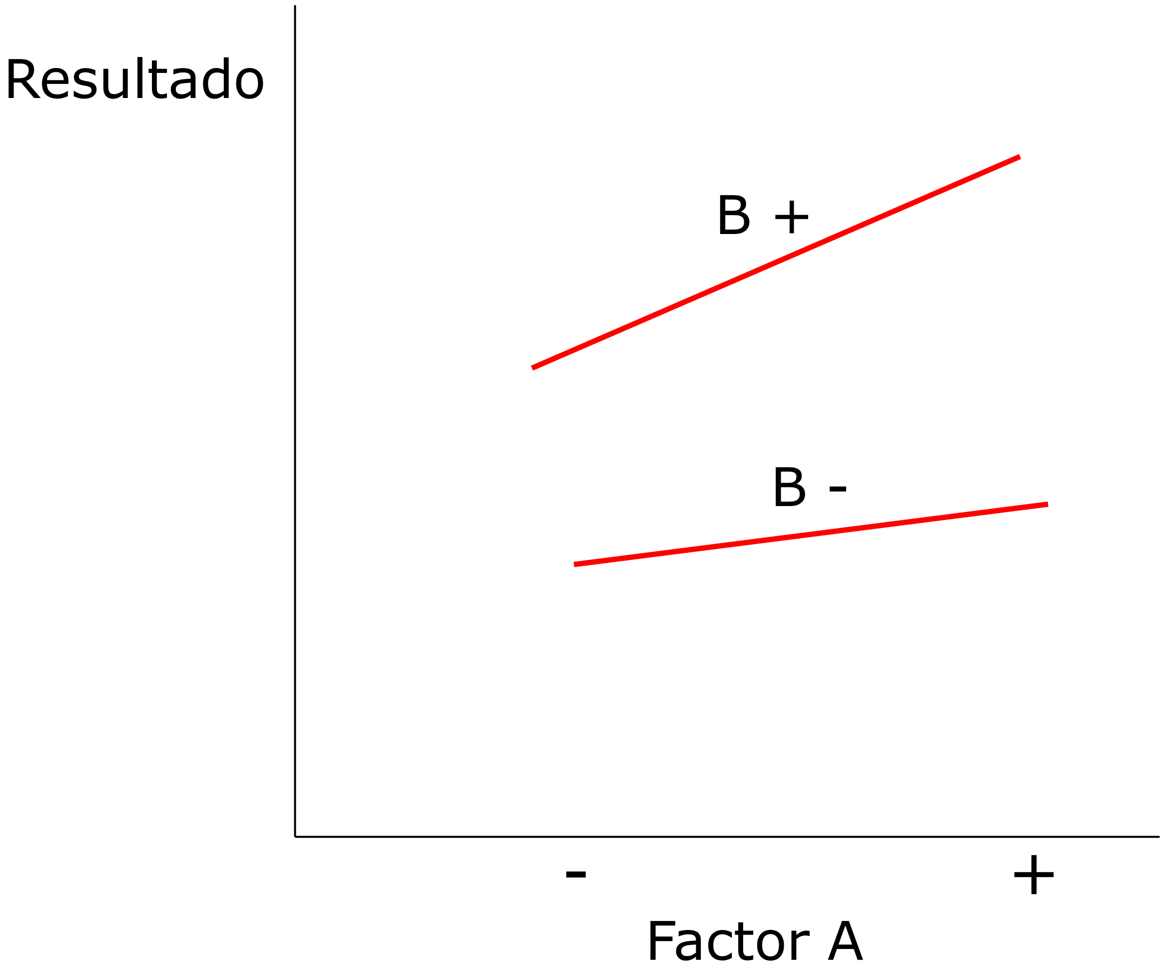


A tiene efecto

B tiene efecto

No interacción

Cuatro posibles situaciones sobre existencia de interacción doble

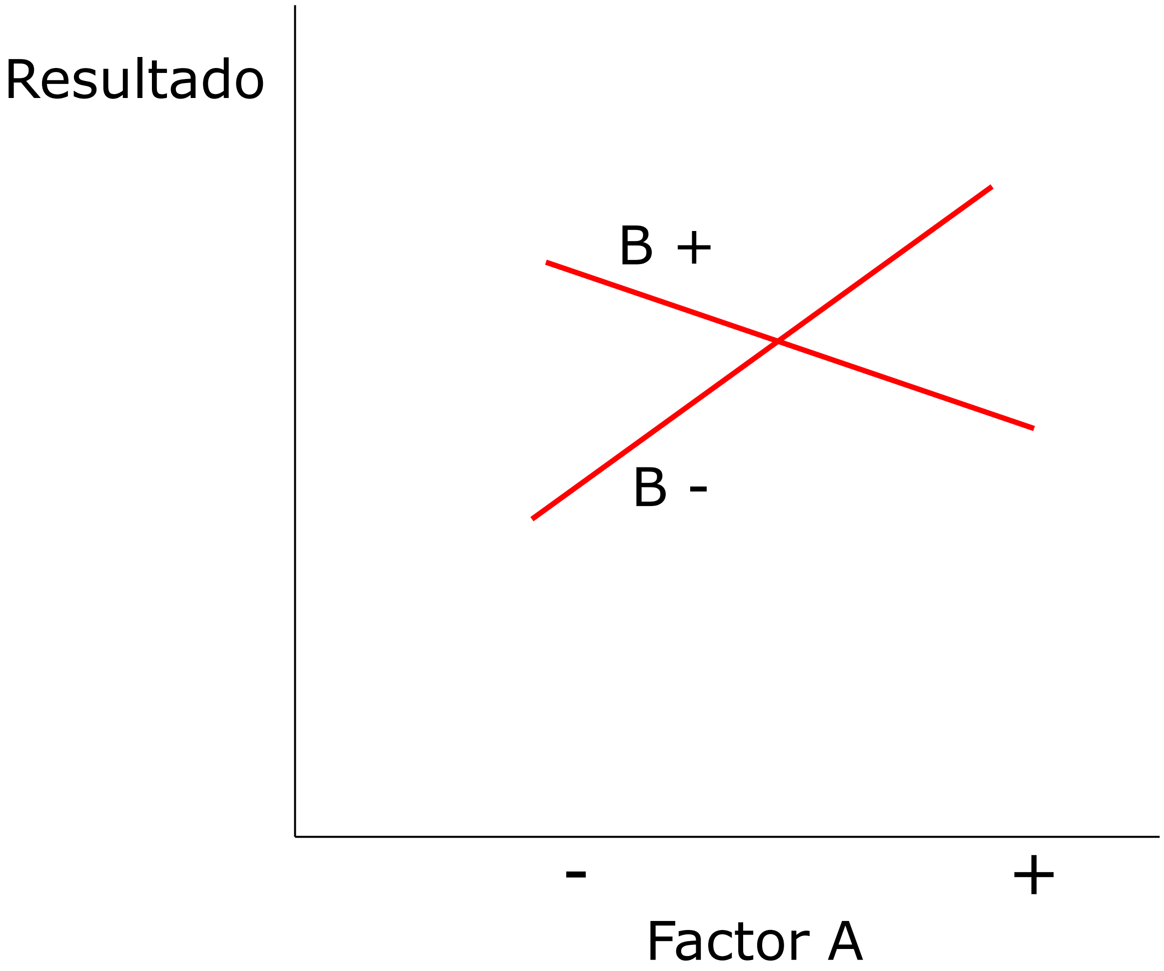


A tiene efecto

B tiene efecto

Ligera interacción

Cuatro posibles situaciones sobre existencia de interacción doble



Fuerte interacción
(Carece de sentido hablar de efectos simples)

Un ejemplo

En el ejemplo del principio:

Se comparan tres tipos de procesador (A, B y C) y dos temperaturas (15 y 40). Variable respuesta: velocidad de ejecución de un tipo de programa.

Hay tres datos por combinación de los dos factores.

Un ejemplo

- Los datos son (caso 5º):

	Proc. A	Proc. B	Proc. C
T=15	3,2 2,9 3,8	8,4 9,7 8,6	7,7 7,9 6,8
T=40	2,9 3,2 3,5	3,6 2,7 4,2	5,7 5,6 5,9

Cuadro resumen del ANOVA del ejemplo:

Origen variabilidad	Sumas de Cuadrados	Grados libertad	Cuadrado Medio	F-ratio
Total	91,8	17		
Temper.	26,16	1	26,16	93,43
Procesad.	40,16	2	20,08	71,71
TempxProc	22,1	2	11,05	39,46
Residual	3,4	12	0,28	

- Efecto de temperatura, F_{ratio}= 93,43

- En la distribución F con 1 y 12 grados de libertad el valor F(5%)= 4,75 y F(1%)=9,33

- Como 93,43 >>> 9,33 $\Rightarrow$ El efecto de temperatura es muy significativo estadísticamente.
- El factor temperatura tiene efecto sobre la velocidad media de ejecución.

- Para procesador F_{ratio}=71,71
- Se compara con la distribución $F_{2,12}$
- Como F(5%)=3,89, F(1%)=6,93 y 71,71 >>>6,93 $\Rightarrow$ es admisible que el efecto del tipo de procesador es muy significativo estadísticamente.
- El procesador influye sobre la media de la velocidad de ejecución.

- Para la interacción temperaturaxprocesador F_{ratio}=39,46.
- Se compara también con la distribución $F_{2,12}$.

  Como 39,46 >>> 6,93 $\Rightarrow$ El efecto de la interacción es muy significativo estadísticamente.
- La interacción también influye significativamente sobre el valor medio de la velocidad de ejecución.

Análisis con Statgraphics

- Para obtener la tabla del ANOVA y realizar el test F con más de un factor con el programa Statgraphics se puede utilizar la opción:

Compare….Analysis of Variance….Multifactor ANOVA

Análisis con Statgraphics

- Esta opción proporciona con el ejemplo la tabla ANOVA:

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value	
Temperatura	26,1606	1	26,1606	90,91	0,0000	<0,01
Procesador	40,17	2	20,085	69,79	0,0000	<0,01
Temp*Proc	22,1011	2	11,0506	38,40	0,0000	<0,01
RESIDUAL	3,45333	12	0,287778			
TOTAL	91,885	17				

La conclusión sería que el efecto de la temperatura, del procesador y de su interacción sobre el valor medio de la velocidad de ejecución, son muy significativos estadísticamente (P-Value < 0,01 para los tres efectos).

Análisis con Statgraphics

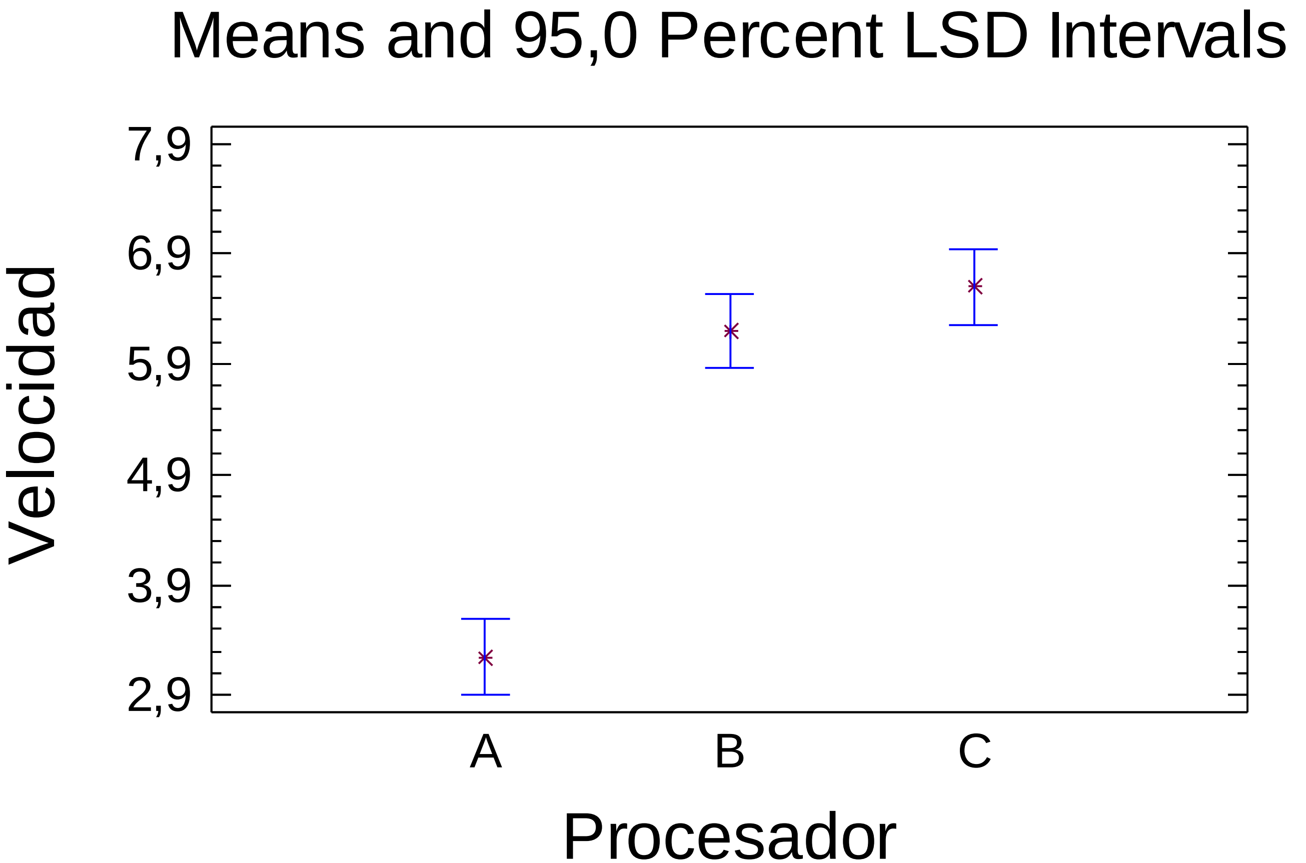
- Además de la tabla resumen del ANOVA, la opción utilizada permite obtener los intervalos LSD para factores cualitativos, gráficos de medias para factores cuantitativos, para la interacción y análisis de residuos.

Análisis con Statgraphics

Ejemplo: el factor procesador tiene más de 2 variantes (3 tipos A, B y C).

- Ha resultado muy significativo.
- Para estudiar entre qué tipos de procesador hay diferencias se va a aplicar intervalos LSD.
- Los intervalos se representan en la siguiente figura obtenida con Statgraphics.

Análisis con Statgraphics



Análisis con Statgraphics

- Puede constatarse que existe una diferencia significativa entre la media del tipo de procesador A y las de los otros dos tipos de procesadores.
- La velocidad media de ejecución es significativamente menor con procesador tipo A.
- No es significativa la diferencia al respecto entre estos los procesadores tipo B y C, dando mayor la velocidad media con ambos.

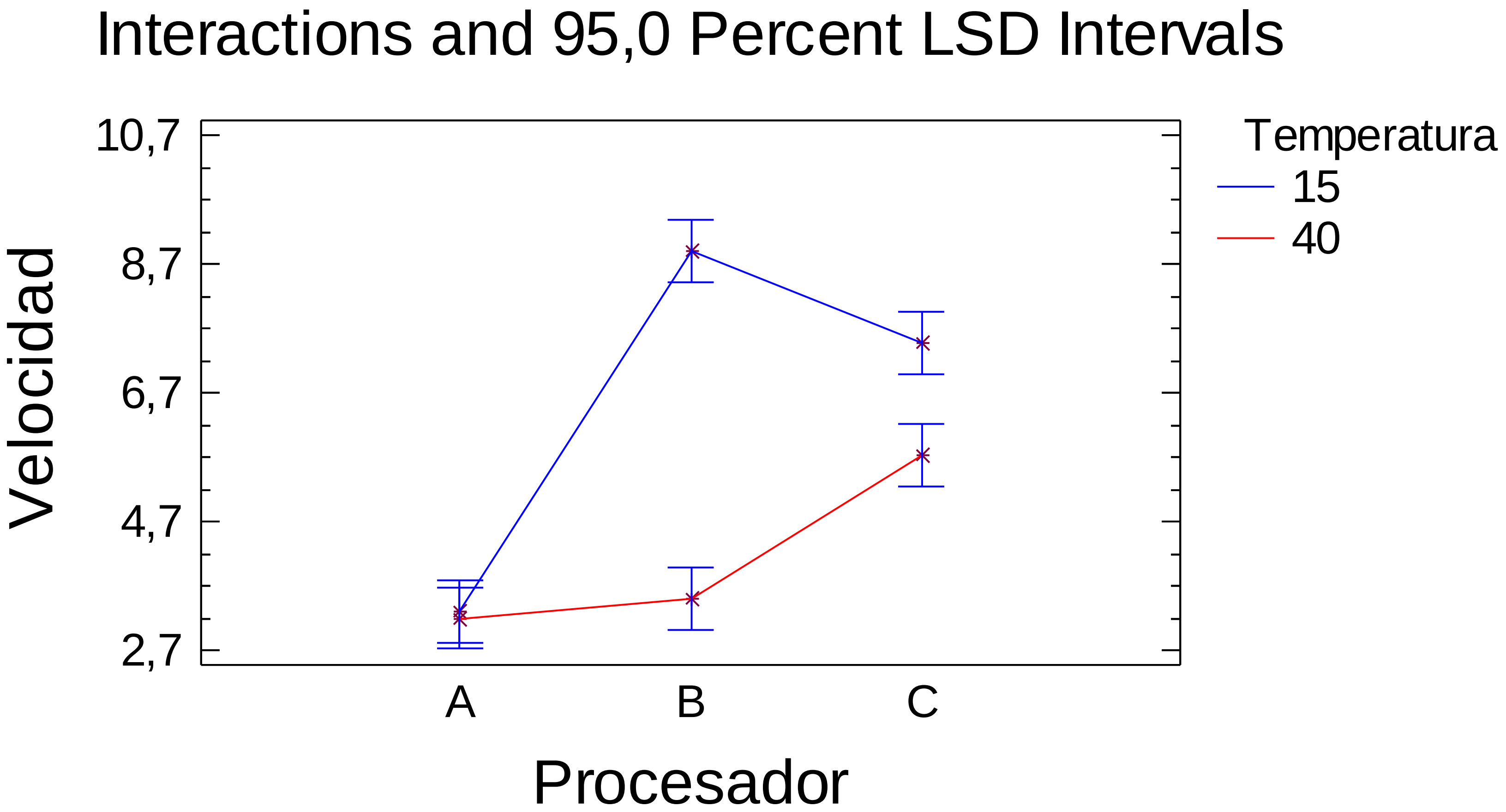
Análisis con Statgraphics

- Una forma sencilla de interpretar a nivel descriptivo las interacciones dobles, es utilizar un gráfico en el que se representen los valores medios obtenidos para las variantes (o niveles) de un factor, diferenciados según las variantes (o niveles) del otro factor.
- Estos gráficos los proporciona Statgraphics, teniendo la opción de elegir cuál de los dos factores se pone en el eje de abscisas.

Análisis con Statgraphics

- Como ejemplo se recoge a continuación el gráfico de la interacción temperatura x procesador del primer ejemplo (dicha interacción resultó muy significativa en la tabla del ANOVA).
- Se observa que con T=15 hay diferencias significativas entre los tres tipos de procesadores, siendo la velocidad media mayor con el B.
- Si embargo con T=40 los procesadores A y B no difieren entre sí, y la velocidad media es mayor con C.

Interpretación de las interacciones dobles



Ejercicio de autoevaluación

Se ha realizado un experimento para estudiar el efecto del sistema operativo utilizado (A, B y C) y el tamaño de programa (64, 128 y 192 Kb), sobre el tiempo medio, en segundos, que tardan en compilarse dichos programas en un determinado lenguaje.

En dicho experimento se realizaron 2 pruebas para cada una de las combinaciones posibles.

Ejercicio de autoevaluación

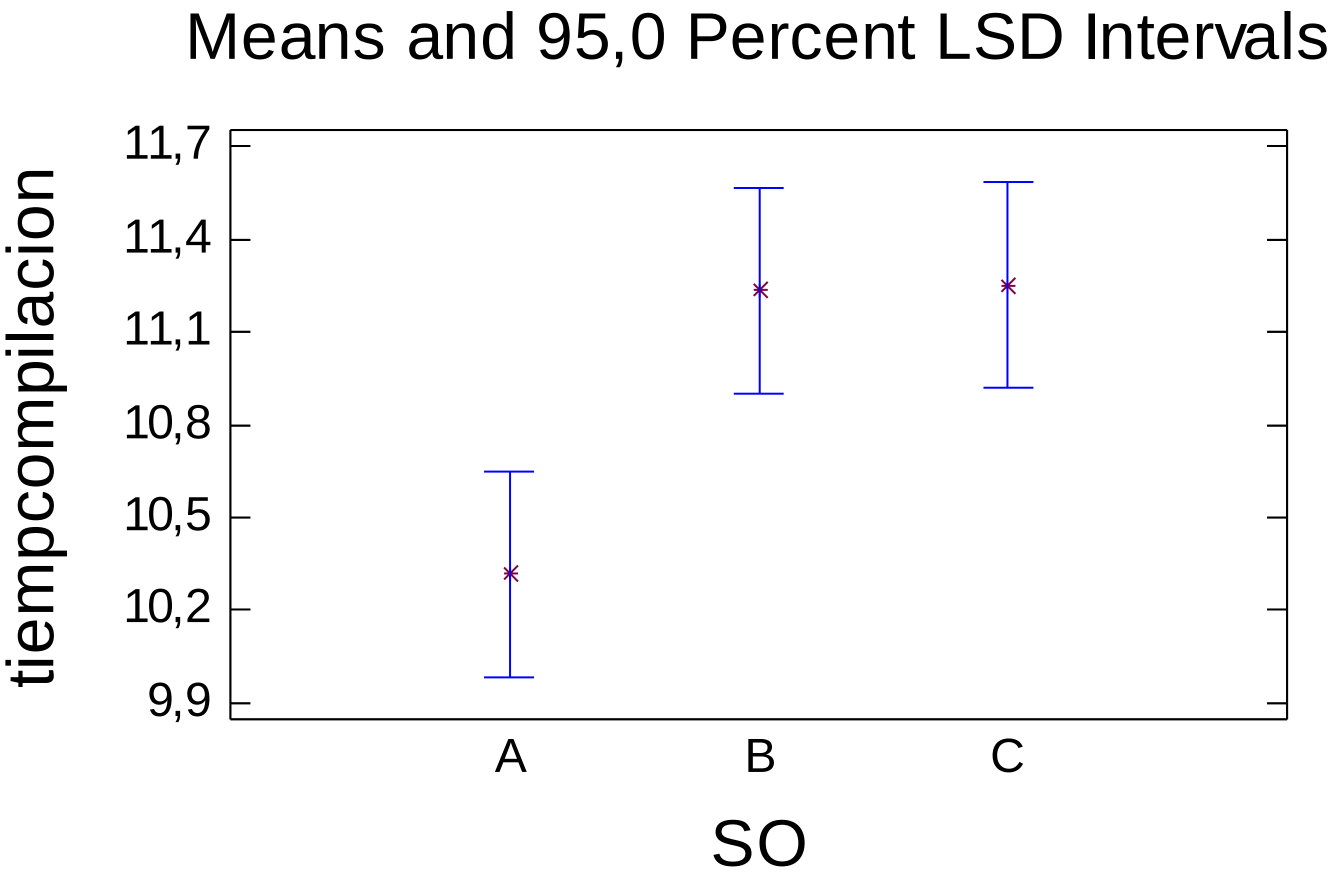
	Tamaño		
Sist.Operat.	64 Kb	128 Kb	192 Kb
A	10,2 10	9,5 10,7	11 10,5
B	12,3 11,9	9,9 10,6	10,7 12
C	11,3 10,9	11,7 12	11,1 10,5

Ejercicio de autoevaluación

a) Construye el cuadro resumen del ANOVA y estudia si son significativos los efectos de los factores estudiados y su interacción (α=5%). (Algunos resultados de la tabla del ANOVA: $SC_{total}=10{,}96$, $SC_{SO}=3{,}42$ $SC_{tamaño}=0{,}41$, $SC_{SOxtamaño}=4{,}78$)

b) Interpreta el efecto de sistema operativo a partir de los intervalos LSD del gráfico siguiente.

Ejercicio de autoevaluación



Ejercicio de evaluación

c) Interpreta a nivel descriptivo la interacción con el gráfico de medias siguiente: