UD 2 – Modelos paramétricos para la toma de decisiones

UD 2.1 – Análisis de la Varianza

(*ANalysis Of VAriance*)

ANOVA con varios factores controlados

Contenidos

- ANOVA con varios factores controlados
  - Efecto simple. Interacción
  - Estimación de efectos
  - Predicciones
  - Ejemplos
- Cuestiones adicionales en el ANOVA
- Ejercicios

ANOVA con varios factores controlados

- En el ejemplo de la primera unidad, se ha considerado **un único factor** (el proveedor) cuyo efecto sobre la variable estudiada (el equilibrado dinámico de los cigueñales suministrados) se quería investigar. Este factor tenía tres alternativas o variantes, pues se estudiaron tres proveedores diferentes

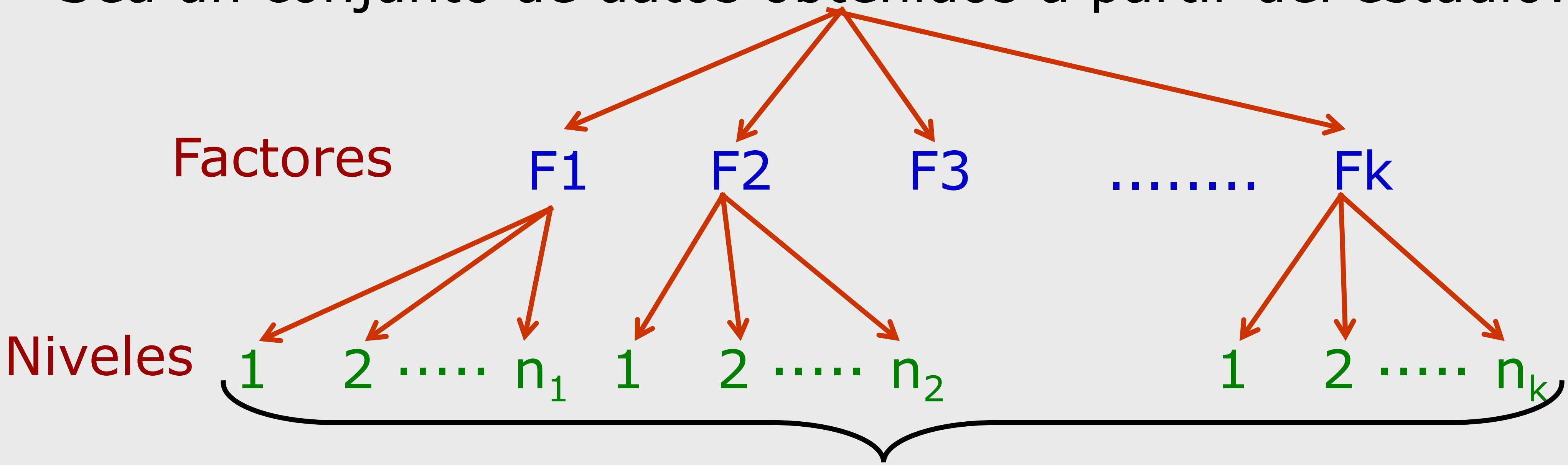
- Como ya se ha comentado, en la mayor parte de los problemas reales son muchos los **factores** que pueden afectar a los resultados de interés

⇩

ANOVA

ANOVA con varios factores controlados

- Sea un conjunto de datos obtenidos a partir del estudio:



Nº de posibles condiciones diferentes: $\mathbf{n_1 \times n_2 \times \ldots \times n_k}$

• Si cada **tratamiento** posible se ensaya un mismo número $\mathbf{n_r}$ de veces → Plan Factorial Equilibrado

• Si $\mathbf{n_r} = 1$ → Plan Factorial Equilibrado no replicado

Efectos simples

- **Ejemplo**: se dispone de datos de emisión de CO_2 (gr CO_2/Km) en motores de motocicletas de cierto cubicaje. Los datos proceden de un plan factorial equilibrado, con dos tipos de motores diferentes y con distinto número de cilindros:
  - Factor 1: Tipo de motor, con dos variantes
    - 1: Serie
    - 2: Preparado
  - Factor 2: Número de cilindros, con tres niveles
    - 1: bajo
    - 2: medio
    - 3: alto

Efectos simples

Se define sobre el promedio de las condiciones estudiadas de los restantes factores

- En el ejemplo, el efecto simple del factor **"Tipo de motor"** se medirá por la diferencia entre el promedio de CO_2 emitido con uno u otro tipo, **para el promedio de los tres niveles de cilindros** estudiados

- De forma análoga, el efecto simple del factor **"Nº de cilindros"** se medirá por las diferencias entre los promedios de CO_2 emitidos con los tres niveles del factor, **para el promedio de los dos tipos de motores** analizados

Interacciones dobles

Existe **interacción** entre dos factores cuando el efecto de uno de ellos es diferente según la variante considerada del otro factor

- En el ejemplo, existirá **interacción** entre los dos factores si, por ejemplo, la diferencia entre el promedio de CO_2 emitido por ambos tipos de motores es muy marcada si el número de cilindros del motor es elevado, pero es pequeña o inexistente si el número de cilindros es bajo

- De forma simétrica, existirá **interacción** entre los dos factores si , por ejemplo, la diferencia del promedio de CO_2 emitido entre motores con un número de cilindros alto y bajo es mucho más marcada trabajando con el tipo de motor preparado que con el serie

Interacciones dobles

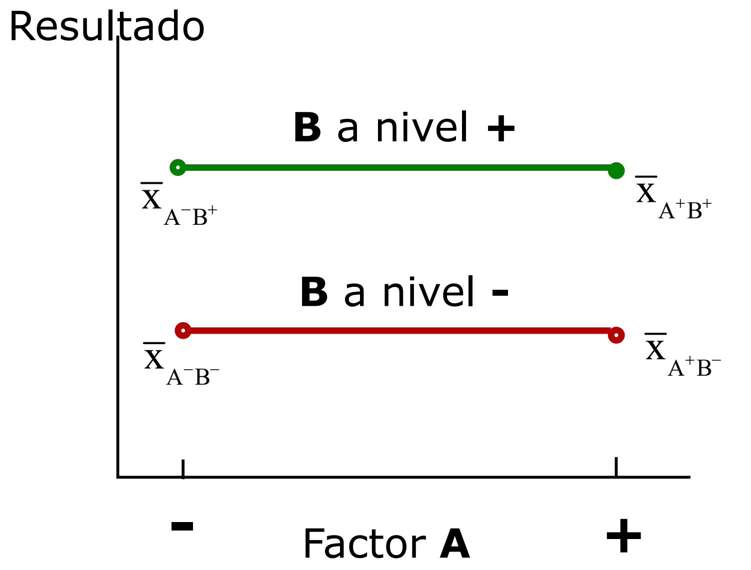
- Hay que tener en cuenta que si el efecto de un primer factor depende de la variante considerada del segundo, también necesariamente el efecto del segundo dependerá de la variante considerada del primero

- **Nota:** en estudios con más de dos factores puede plantearse también la existencia de interacciones de orden superior: triples, cuádruples, etc…

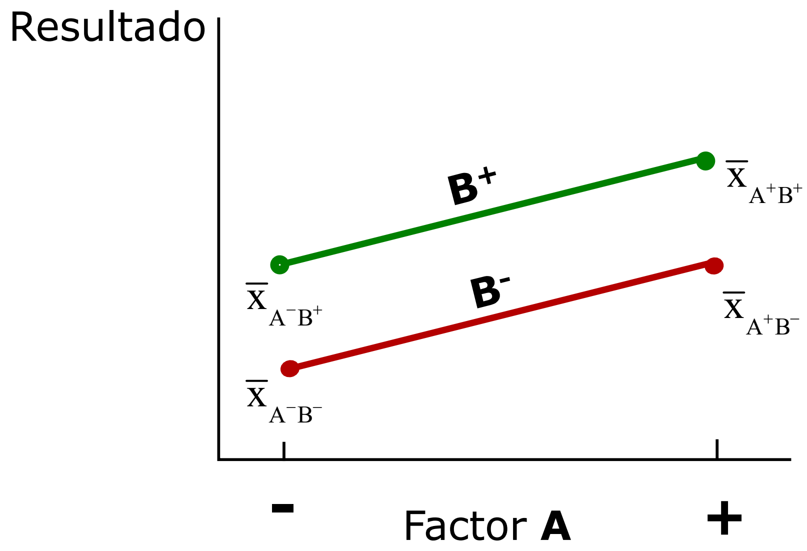
  Así, existirá una **interacción triple** entre tres factores A, B y C si, por ejemplo, hubiera una interacción doble entre A y B cuando C está a nivel bajo, pero no existiera dicha interacción A*B cuando C está a nivel alto

  (Las interacciones de orden superior a dos se presentan poco en la práctica, siendo además difíciles de interpretar)

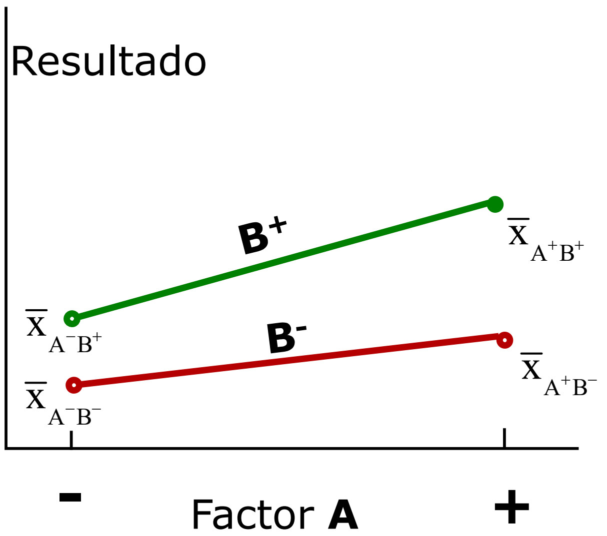
Interacciones dobles



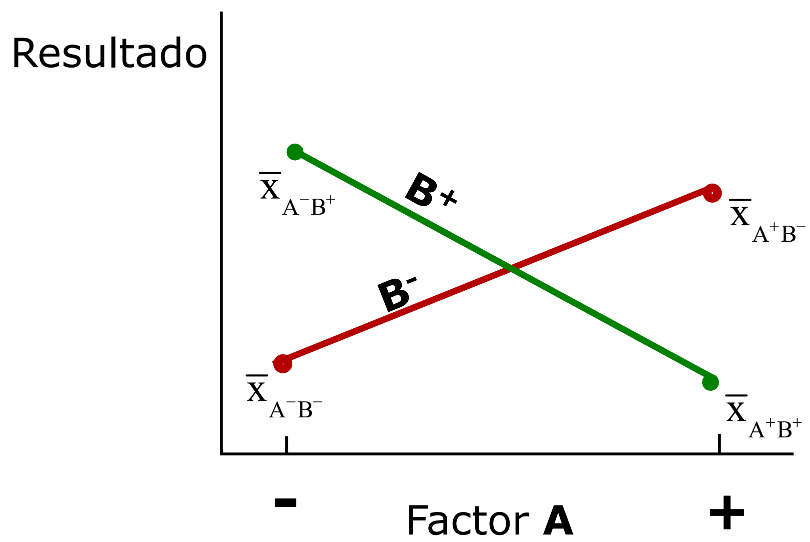
A no tiene efecto **B** tiene efecto
No hay interacción



A tiene efecto **B** tiene efecto
No hay interacción



A tiene efecto **B** tiene efecto
Ligera interacción



Fuerte interacción
(carece de sentido hablar de efectos simples)

Efectos simples. Interacciones

- Sean **2 factores**:

FI → **I** niveles o variantes

X

FJ → **J** niveles o variantes

Ix**J** tratamientos → Población (**IxJ** poblaciones)

$\mathbf{X_{ij}}$: variable asociada a la utilización de la variante/nivel i de FI con la variante/nivel j de FJ

$$\mathbf{X_{ij} \approx Normal(m_{ij}\ \sigma_{ij})}$$

Estimación de los efectos

- El **efecto simple** de un factor se define y estima sobre el **promedio** de las condiciones estudiadas para los otros factores

¿Cómo podemos **cuantificar** los efectos de los factores?

Estimación de los efectos

- Efecto de la variante **i** de un factor **FI**:

Diferencia entre el promedio m_i correspondiente a dicha variante i de FI y la media general m

$$Ef_i = \alpha_i = m_i - m$$

media poblacional de todos posibles resultados obtenidos en todas las pruebas en las que **FI** estaba al nivel **i**

media general de todas las pruebas

Estimación de los efectos

Estimación del efecto simple α_i

$$\boxed{a_i = \bar{x}_i - \bar{x}}$$

media de los resultados obtenidos en todas las pruebas en las que **FI** estaba al nivel **i**

media general de todas las pruebas

Estimación de los efectos

- Efecto de la variante **j** de un factor **FJ**:

Diferencia entre el promedio m_j correspondiente a dicha variante j de FJ y la media general m

$$Ef_j = \beta_j = m_j - m$$

media poblacional de todos posibles resultados obtenidos en todas las pruebas en las que **FJ** estaba al nivel **j**

media general de todas las pruebas

Estimación de los efectos

Estimación del efecto simple β_j

$$\boxed{b_j = \bar{x}_j - \bar{x}}$$

media de los resultados obtenidos en todas las pruebas en las que **FJ** estaba al nivel **j**

media general de todas las pruebas

Estimación de los efectos

¿Cómo cuantificar el efecto de una interacción?

- Si no existe interacción entre los dos factores, la media **m_{ij}** correspondiente a la combinación de la variante i del primero con la variante j del segundo resulta igual a la suma de la media general más los dos efectos simples:

$$\boxed{m_{ij} = m_{general} + \alpha_i + \beta_j}$$

Estimación de los efectos

¿Cómo cuantificar el efecto de una interacción?

- Si existe interacción entre los dos factores la igualdad anterior NO se verifica, definiéndose el efecto de la interacción entre ambas variantes **$(\alpha\beta)_{ij}$** como la diferencia entre los dos miembros de la ecuación anterior:

$$\boxed{(\alpha\beta)_{ij} = m_{ij} - \left(m_{general} + \alpha_i + \beta_j\right)}$$

Estimación de los efectos

$$(ab)_{ij} = \bar{X}_{ij} - a_i - b_j - \bar{X}$$

media de los resultados obtenidos en todas las pruebas en las que **FI** estaba al nivel **i** y **FJ** a nivel **j**

media general de todas las pruebas

Efecto estimado del factor **FI** al nivel **i**

Efecto estimado del factor **FJ** al nivel **j**

Estimación de los efectos

$$(ab)_{ij} = \bar{X}_{ij} - \bar{X}_{i} - \bar{X}_{j} + \bar{X}$$

media de los resultados obtenidos en todas las pruebas en las que **FI** estaba al nivel **i** y **FJ** a nivel **j**

media general de todas las pruebas

media de los resultados obtenidos en todas las pruebas en las que **FI** estaba al nivel **i**

media de los resultados obtenidos en todas las pruebas en las que **FJ** estaba al nivel **j**

Predicciones

- Si el objetivo de un estudio es **optimizar** cierto proceso a partir del análisis de datos obtenidos del mismo

➔ Selección de la **combinación óptima** de variantes de los distintos factores a partir de los resultados obtenidos en el análisis (**ANOVA**)

⇩

Condiciones Operativas

⇩

Predecir

⇩

¿ Cómo serán los resultados que se obtendrán en tales condiciones ?

Predicciones

¿Cómo realizar predicciones?

La predicción se lleva a cabo adicionando a la media general obtenida los valores estimados de los efectos correspondientes que hayan resultado significativos.

Así, si en el caso de tres factores hubieran resultado significativos los tres efectos simples y la interacción entre los dos primeros, la predicción sería:

$$m^*_{ijk} = \bar{X} + a_i + b_j + c_k + (ab)_{ij}$$

Contenidos

- ANOVA con varios factores controlados
  - Efecto simple. Interacción
  - Estimación de efectos
  - Predicciones
  - Ejemplos
- Cuestiones adicionales en el ANOVA
- Ejercicios

Ejemplo 1

- Se dispone de datos (fichero PINZADO_R.sf3) recogidos con el fin de estudiar el efecto sobre el hinchamiento post-extrusión, en la fabricación de botellas de polietileno de alta densidad de los factores:
  - **catalizador** utilizado en la obtención de dicho polietileno (Factor1 con **3 variantes**)
  - **tipo de molde** empleado (Factor2 con **2 variantes**)
- Los datos proceden de un Plan Factorial Equilibrado (3 x 2) con 10 replicaciones:
  - 3x2 = 6 tratamientos
  - 6 x 10 = 60 observaciones o datos

Ejemplo 1

- **Variable respuesta o dependiente**:

pinzado (hinchamiento)

La variable respuesta se determinó midiendo la longitud del pinzado en la base de cada botella (valores altos indican mayor hinchamiento, lo que es una característica desfavorable) → variable respuesta del tipo "cuanto más pequeño mejor"

OBJETIVOS: responder a….

1 ¿Qué factores tienen un efecto significativo sobre la variable de interés?

2 ¿Cuál es la naturaleza de estos efectos?

3 ¿Cuáles son las Condiciones Operativas Óptimas?

4 ¿Qué respuesta media cabe predecir en las condiciones óptimas encontradas?

5 ¿Hay efectos significativos sobre varianzas? ¿cuál es su naturaleza?

6 ¿Cuál es la varianza previsible de la respuesta en función de las condiciones operativas utilizadas?

Ejemplo 1

La tabla siguiente recoge los totales registrados en cada uno de los seis tratamientos ensayados:

TOTALES

		CATALIZADOR			
		A	**B**	**C**	
MOLDE	**1**	**915,93**	**916**	**953,71**	**2785,6**
	2	**871,93**	**885,15**	**904,98**	**2662,1**
		1787,9	**1801,2**	**1858,7**	**5448**

Ejemplo 1

Cálculo de las Sumas de Cuadrados y grados de libertad

$$SG = \frac{5448^2}{60} = 494678$$

$$SC_{Total} = \left(93^2 + 92^2 + ... + 91^2\right) - SG = 452,088 \qquad gl_{Totales} = 60 - 1 = 59$$

$$SC_{Molde} = \frac{2785,6^2 + 2662,1^2}{30} - SG = 254,48 \qquad gl_{Molde} = 2 - 1 = 1$$

$$SC_{Catalizador} = \frac{1787,9^2 + 1801,2^2 + 1858,7^2}{20} - SG = 141,75 \qquad gl_{Catalizador} = 3 - 1 = 2$$

$$SC_{Cat*Molde} = \frac{915,93^2 + ... + 904,98^2}{10} - SG - SC_{Catalizador} - SC_{Molde} = 5 \qquad gl_{Cat*Molde} = 1x2 = 2$$

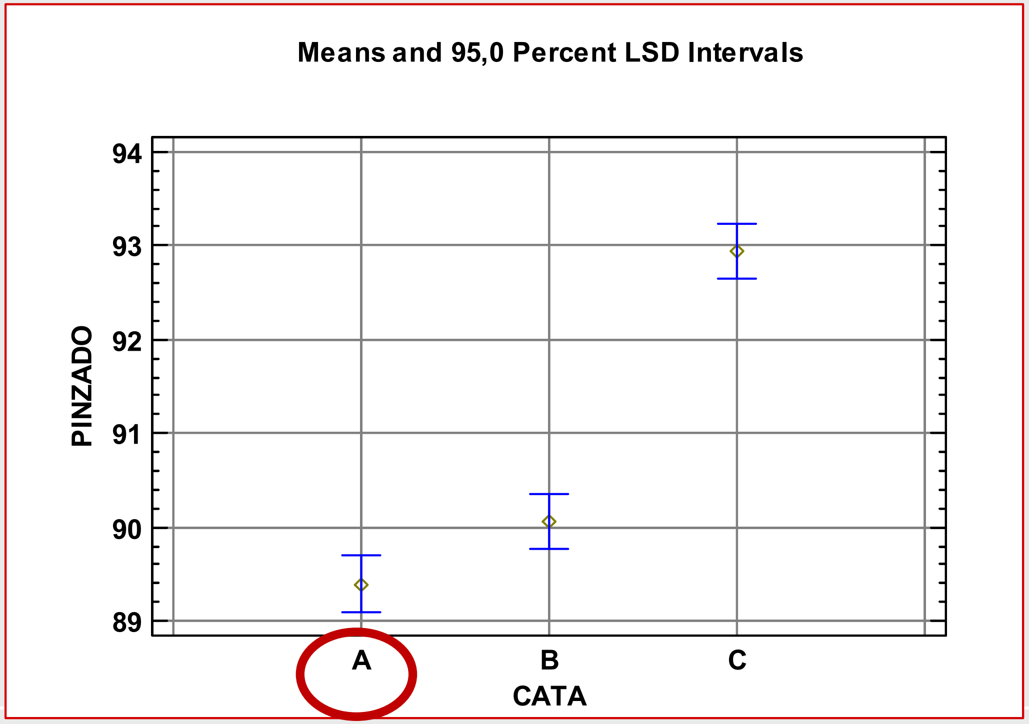
$$SC_{Residual} = SC_{Total} - SC_{Catalizador} - SC_{Molde} - SC_{Cat*Molde} = 50 \qquad gl_{Residual} = 59 - 1 - 2 - 2 = 54$$

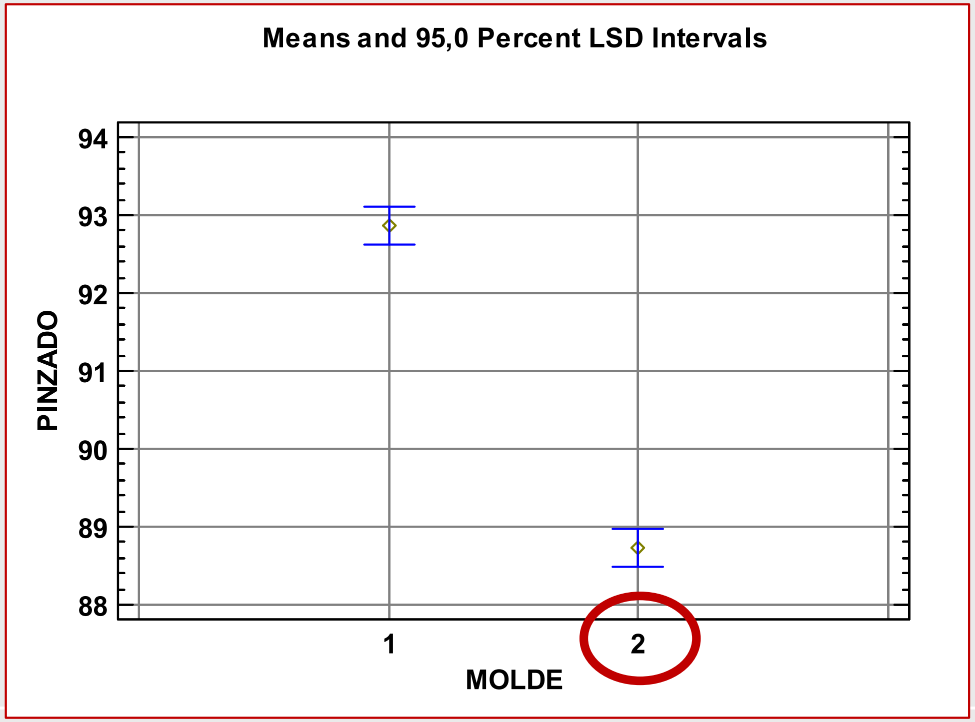
Ejemplo 1

Analysis of Variance for PINZADO - Type III Sums of Squares

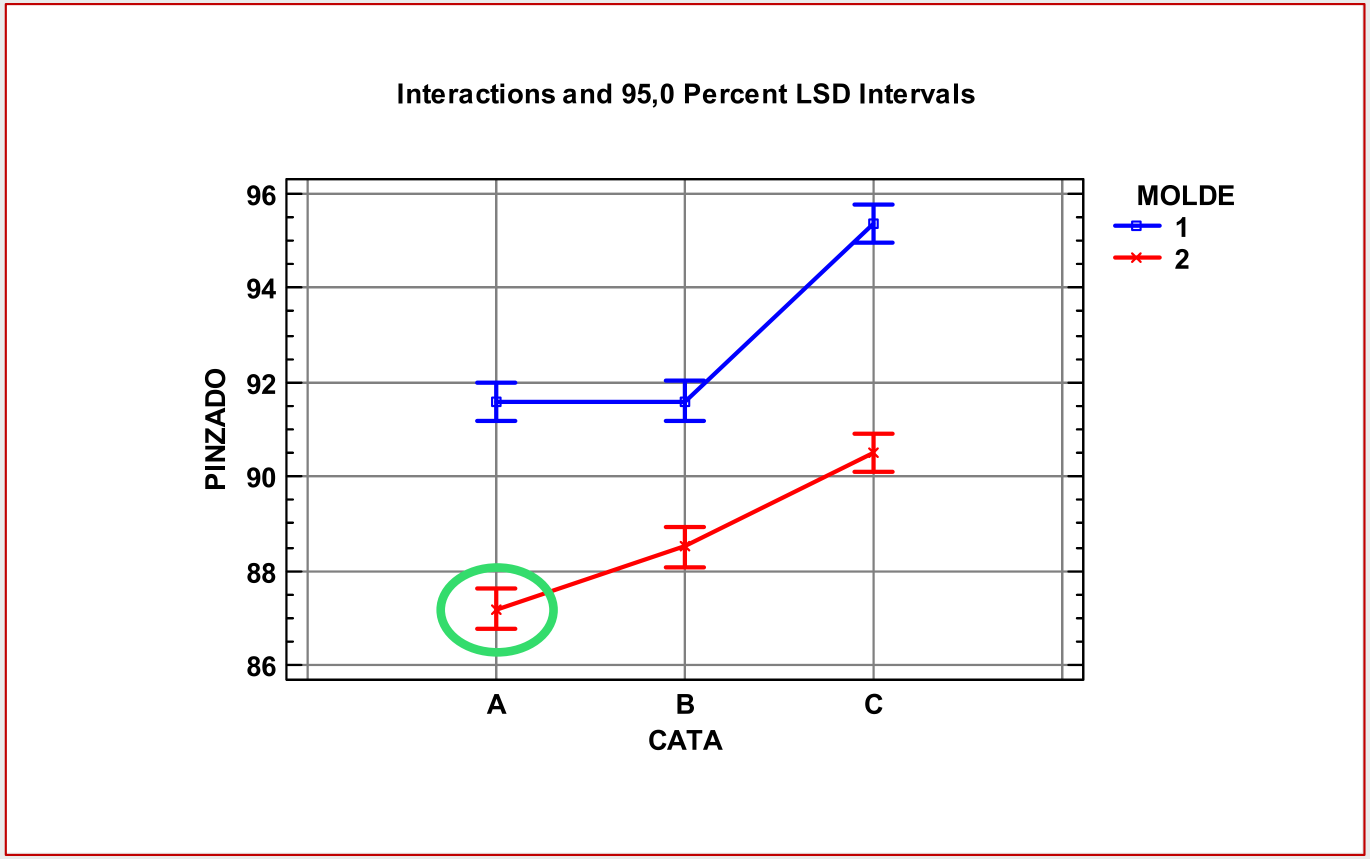
Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
MAIN EFFECTS					
A:CATA	141,75	2	70,875	80,95	0,0000
B:MOLDE	254,476	1	254,476	290,64	0,0000
INTERACTIONS					
AB	8,5826	2	4,2913	4,90	0,0111
RESIDUAL	47,28	54	0,875555		
TOTAL (CORRECTED)	452,088	59			

All F-ratios are based on the residual mean square error.





Ejemplo 1



El gráfico de la interacción constatada entre los dos factores indica que mientras la diferencia entre el catalizador C y el resto es patente para ambos moldes, la diferencia entre los catalizadores A y B sólo es significativa con el molde 2

Ejemplo

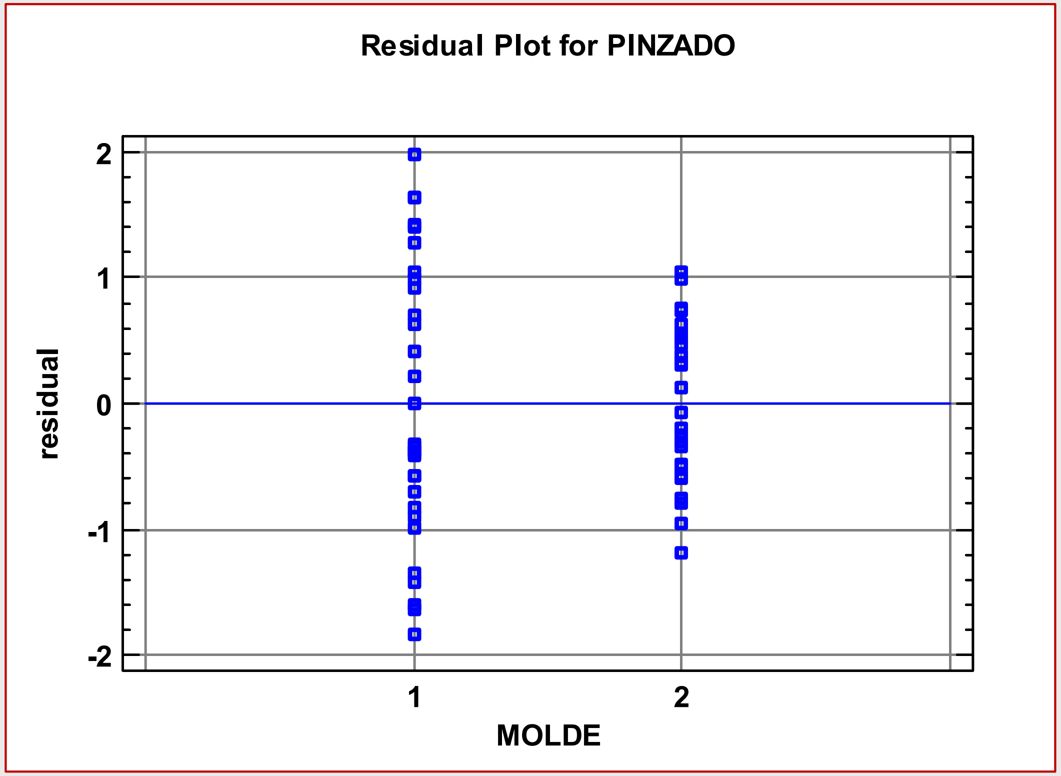
La tabla de los valores medios proporcionada por *Statgraphics* es la siguiente:

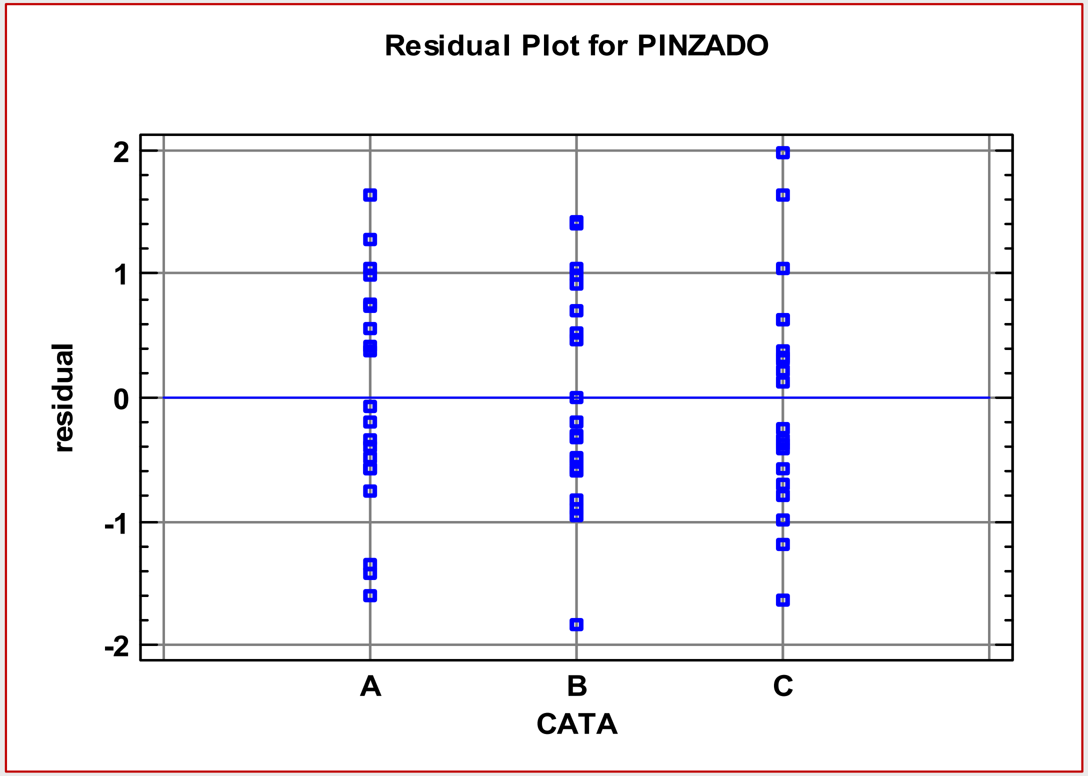
Table of Least Squares Means for PINZADO with 95,0% Confidence Intervals

Level	*Count*	*Mean*	*Stnd. Error*	*Lower Limit*	*Upper Limit*
GRAND MEAN	60	90,7948			
CATA					
A	20	89,3927	0,209231	88,9732	89,8122
B	20	90,0573	0,209231	89,6378	90,4768
C	20	92,9344	0,209231	92,5149	93,3538
MOLDE					
1	30	92,8542	0,170837	92,5117	93,1967
2	30	88,7353	0,170837	88,3928	89,0779
CATA by MOLDE					
A,1	10	91,5926	0,295898	90,9994	92,1858
A,2	10	87,1928	0,295898	86,5996	87,786
B,1	10	91,5995	0,295898	91,0063	92,1928
B,2	10	88,515	0,295898	87,9218	89,1083
C,1	10	95,3705	0,295898	94,7773	95,9637
C,2	10	90,4982	0,295898	89,905	91,0914

Ejemplo

Análisis de residuos





Ejemplo

Estudio de efectos sobre dispersión

Analysis of Variance for RESIDUALS ^2 - Type III Sums of Squares

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
MAIN EFFECTS					
A:CATA	0,0409174	2	0,0204587	0,03	0,9726
B:MOLDE	9,66637	1	9,66637	13,13	0,0006
INTERACTIONS					
AB	0,553052	2	0,276526	0,38	0,6887
RESIDUAL	39,759	54	0,736279		
TOTAL (CORRECTED)	50,0194	59			

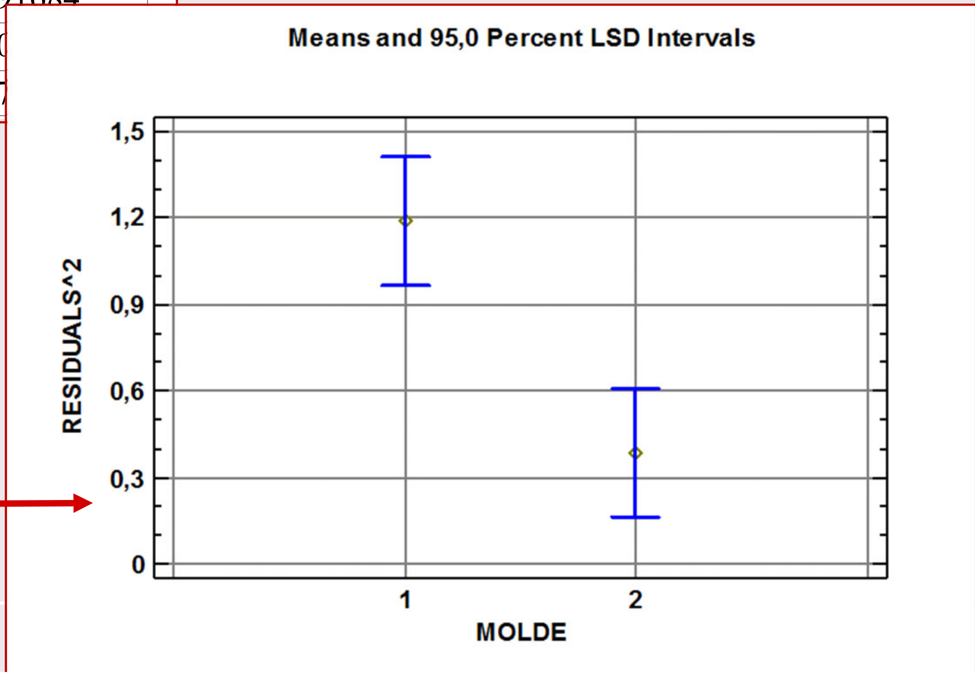
Este ANOVA pone de manifiesto el efecto significativo del factor MOLDE sobre la varianza de la respuesta

Ejemplo

Table of Least Squares Means for RESIDUALS^2 with 95,0% Confidence Intervals

			Stnd.	*Lower*	*Upper*
Level	*Count*	*Mean*	*Error*	*Limit*	*Limit*
GRAND MEAN	60	0,788			
CATA					
A	20	0,815999	0,19187	0,431322	1,20067
B	20	0,753144	0,19187	0,368468	1,13782
C	20	0,794856	0,19187	0,41018	1,17953
MOLDE					
1	30	$\boxed{1{,}18938}$	0,156661	0,875293	1,50347
2	30	$\boxed{0{,}386619}$	0,156661	0,0725325	0,700706
CATA by MOLDE					
A,1	10	1,34857	0,271345	0,804551	1,89258
A,2	10	0,283432	0,271345	-0,260582	0,827446
B,1	10	1,05862	0,271345	0,514604	1,60263
B,2	10	0,44767	0,271345	-0,0963444	0,991684
C,1	10	1,16096	0,271345	0,616942	1,70
C,2	10	0,428756	0,271345	-0,115258	0,97

Los resultados obtenidos en ambos análisis apuntan que con el Molde 1 se obtiene una variabilidad **más del doble** de la correspondiente al Molde 2



Ejemplo

Condiciones Operativas Óptimas (C.O.O.):

➔ Catalizador 1 (A) , Molde 2 (Pequeño)

Pinzado medio previsto en COO

Media general + Efecto de niveles (significat.)

$$90.8 + \underbrace{(89.4-90.8)}_{\text{Ef. Cataliz. A}} + \underbrace{(88.7-90.8)}_{\text{Ef. Molde 2}} + \underbrace{(87.2-89.4-88.7+90.8)}_{\text{Ef. Cataliz.A , Molde 2}} =$$

$= \underline{\mathbf{87.2}}$

Ejemplo

Condiciones Operativas Óptimas (C.O.O.):

➔ Catalizador 1 (A) , Molde 2 (Pequeño)

Varianza del pinzado prevista en COO

- Media prevista de residuos^2

  Media general + Efectos de niveles (sign.)

  $0.79 + \underbrace{(0.387-0.79)}_{\text{Ef. Molde 2}} = \mathbf{0.387}$

- Varianza estimada (s^2 corregida)

  0.387 x (60/54) = **0.43**

Ejemplo

- *En COO, ¿qué probabilidad tiene una botella tomada al azar de presentar un pinzado <u>inferior a 89 unidades</u>?*

- *¿Qué porcentaje de botellas presentarán un pinzado en COO <u>entre 86 y 89 unidades</u>?*

Ejemplo 2: ANOVA con 2 factores

- Objetivo: analizar el efecto que sobre el tiempo medio de respuesta de un sistema informático tienen dos factores :

  - **Factor cualitativo**: **FICHEROS** con 3 variantes codificadas como 1, 2 y 3 → **I=3**

    Distribución de los ficheros en los discos.

  - **Factor cuantitativo**: **BUFFERS** con 3 niveles

    10, 20, 30 → **J=3**

    Número de buffers en el sistema.

Tratamientos y pruebas

- Cada uno de los 9 tratamientos (9 combinaciones posibles) se ha probado 2 veces

↓

Plan Factorial Equilibrado y Replicado (k=2)

- Cada **prueba** consistió en un día completo:
  - obteniéndose los tiempos medios de respuesta
  - evaluados para un proceso estándar consistente en la compilación de un determinado programa en lenguaje C

- Los resultados se recogen en la siguiente tabla:

Resultados

		BUFFERS		
		10	20	30
FICHEROS	1	2'7 2'4	2'0 2'2	1'8 1'6
	2	3'1 3'2	2'7 2'5	2'2 1'9
	3	3'7 3'9	2'9 3'2	3'5 3'8

Resultados

Totales / Medias		BUFFERS			
		10	20	30	
FICHEROS	1	5’1 2’55	4’2 2’1	3’4 1’7	12’7 2’11
	2	6’3 3’15	5’2 2’6	4’1 2’05	15’6 2’6
	3	7’6 3’8	6’1 3’05	7’3 3’65	21 3’5
		19 3’167	15’5 2’583	14’8 2’467	49’3 2’74

Descomposición de la Interacción: un factor cuantitativo

Objetivos:

- Determinar **qué efectos resultan significativos** → Realizar un ANOVA incluyendo los dos factores (BUFFERS y FICHEROS) y su posible interacción
- Analizar el **efecto** del factor FICHEROS mediante los intervalos LSD
- Estudiar aproximadamente la **naturaleza** (lineal y/o cuadrática) del **efecto** de BUFFERS así como de su interacción con el factor FICHEROS a partir de los gráficos
- Constatar analíticamente los resultados anteriores obteniendo las SC asociadas a los efectos lineal y cuadrático de BUFFERS y a sus interacciones con FICHEROS y **estudiando su significación estadística** (esto último se estudiará mas adelante mediante un modelo de regresión)

ANOVA : Cálculo de las Sumas de Cuadrados

- Suma de Cuadrados Total (como siempre):

$$SCT = \sum_{\forall i,j} x_{i,j}^2 - SG =$$

$$= (2'7^2 + 2'4^2 \ldots + 3'8^2) - \frac{(2'7 + 2'4 \ldots + 3'8)^2}{18} = 8'74$$

- Sustraendo General (como siempre):

$$SG = \frac{(TG)^2}{IxJxK} = \frac{(49'3)^2}{18} = 135'03$$

(nº de datos totales – 1) $\leftrightarrow$

(IxJxK – 1)=(3x3x2 - 1)=17 g.l.

ANOVA: Sumas de Cuadrados y grados de libertad

$$SC_{Ficheros} = \frac{\sum (T_i)^2}{N_i} - SG$$

(nº de variantes – 1) $\leftrightarrow$

(I – 1)=(3-1)=2 g.l.

$$\mathbf{SC_{Ficheros}} = \frac{\sum (T_i)^2}{N_i} - SG = \frac{12'7^2 + 15'6^2 + 21^2}{6} - SG = \mathbf{5'91}$$

$$SC_{Buffers} = \frac{\sum (T_j)^2}{N_j} - SG$$

(nº de variantes – 1) $\leftrightarrow$

(J – 1)=(3-1)=2 g.l.

$$\mathbf{SC_{Buffers}} = \frac{\sum (T_j)^2}{N_j} - SG = \frac{19^2 + 15'5^2 + 14'8^2}{6} - SG = \mathbf{1'69}$$

✓

ANOVA: Sumas de Cuadrados y grados de libertad

$$SC_{FicherosxBuffers} = \frac{\sum (T_{ij})^2}{N_{ij}} - SG - SC_{Ficheros} - SC_{Buffers}$$

$(I - 1)x(J - 1) \leftrightarrow (3\text{-}1)x(3\text{-}1) = 2x2 = 4$ g.l.

$$SC_{FicherosxBuffers} = \frac{5'1^2 + 4'2^2 + 3'4^2 + \cdots + 7'3^2}{2} - $$

$$-SG - \mathbf{5'91} - \mathbf{1'69} = \mathbf{0'88}$$

Sumas de Cuadrados y g.l.

- **Suma de Cuadrados Total** $SCT = 8'74$

(nº de datos– 1) ↔ (**IxJxN – 1**)=(3x3x2 - 1)=**17 g.l.**

- **Suma de Cuadrados Factores**

$SC_{Ficheros} = 5'914$ ➡ (nº de variantes – 1) ↔ (**I – 1**)=(3-1)=**2 g.l.**

$SC_{Buffers} = 1'688$ ➡ (nº de niveles– 1) ↔ (**J – 1**)=(3-1)=**2 g.l.**

Sumas de Cuadrados y g.l.

- **Suma de Cuadrados Interacción**

SC Factor i x Factor j = SC_{FixFj}

$$SC_{FicherosxBuffers} = \mathbf{0'875}$$

$$(\mathbf{I} - 1)x(\mathbf{J} - 1) \leftrightarrow (3\text{-}1)x(3\text{-}1) = 2x2 = \mathbf{4\ g.l.}$$

… Más adelante veremos con más detalle qué representan las interacciones entre factores.

Sumas de Cuadrados y g.l.

- **Suma de Cuadrados Residual**

$$\boxed{SC_{Residual} = SC_{Total} - \sum SC_{Ef_p}}$$

$$SC_{Residual} = 8'74 - (5'914 + 1'568 + 0'875) = 0'265$$

- Los **g.l** de los que disponemos serían:
  - **SCT** → 18-1=17 g.l.
  - $SC_{FICHEROS}$ → 3-1=**2** g.l.
  - $SC_{BUFFERS}$ → 3-1=**2** g.l.
  - $SC_{FICHxBUFF}$ → 2x2=**4** g.l.

  8 g.l

  - **SCR** → 17-8=**9** g.l. → **Suficientes**

Tabla Resumen del ANOVA

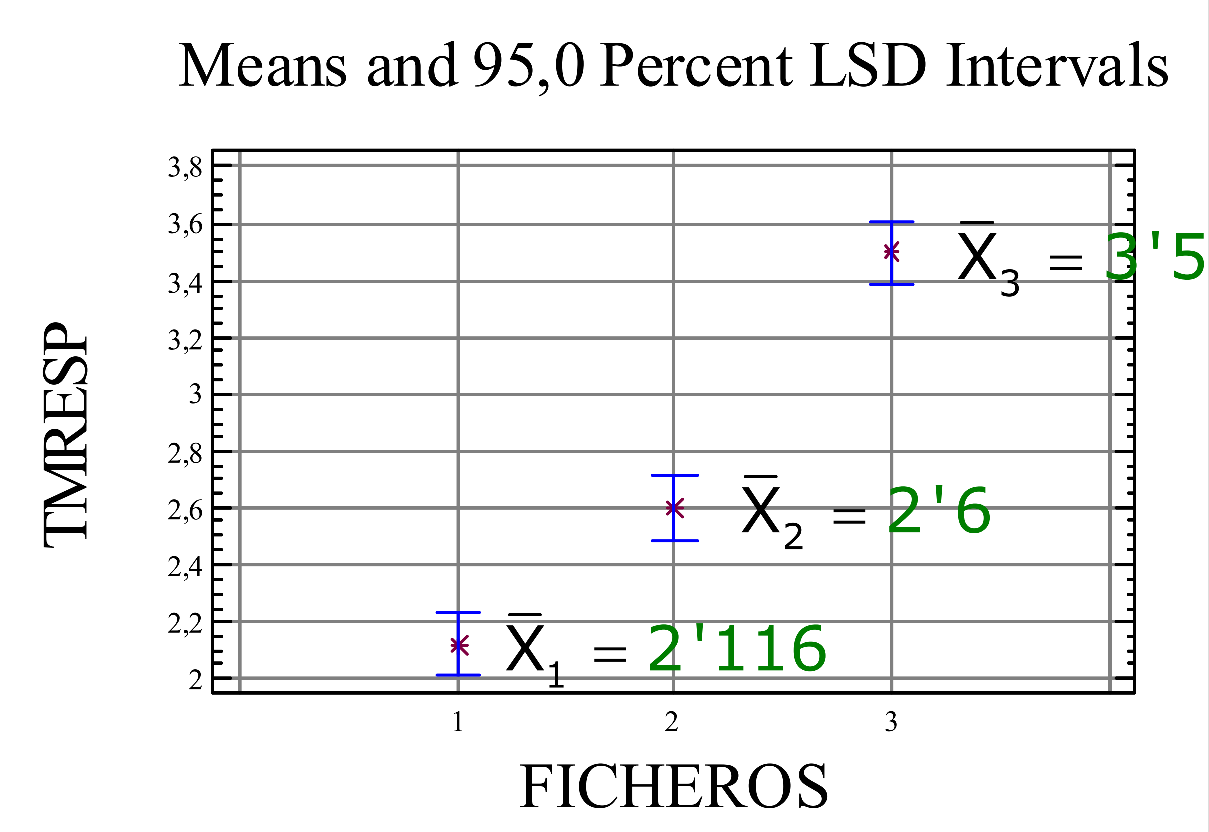
Origen Variación	SC	g.l.	CM	F ratio	F_tabla α=0'05
Total	8'743				
FICHEROS	5'914				
BUFFERS	1'688				
FICHxBUFF	0'876				
Residual					

Tabla Resumen del ANOVA

Origen Variación	SC	g.l.	CM	F ratio		F_tabla α=0'05
Total	**8'743**	**17**	**-**	**-**		
FICHEROS	**5'914**	**2**	**2'957**	**100'433**	>	**4'26** = $F_{2,9}$
BUFFERS	**1'688**	**2**	**0'844**	**28'66**	>	**4'26** = $F_{2,9}$
FICHxBUFF	**0'876**	**4**	**0'219**	**7'437**	>	**3'63** = $F_{4,9}$
Residual	**0'265**	**9**	**0'029**			

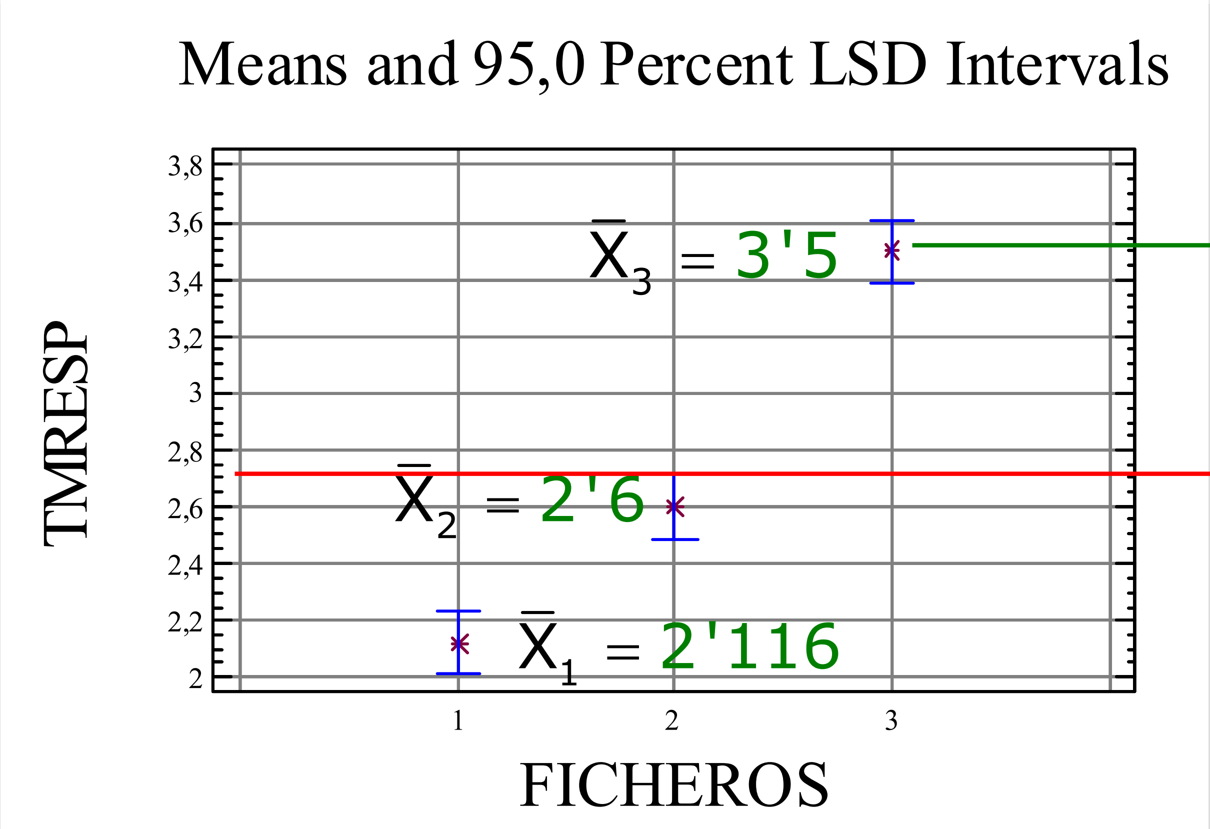
- Los efectos simples de FICHEROS y BUFFERS resultan significativos, pues su F_{ratio} es **mayor** que el valor en tablas de una $F_{2,9}^{(\alpha=0,05)}$.
- Por el mismo motivo, **también** es significativo el efecto de la interacción FICHEROSxBUFFERS ($F_{4,9}^{\alpha}$)

Análisis efecto FICHEROS: Intervalos LSD



Para el promedio de BUFFERS ensayados, con la distribución de FICHEROS 1 se obtienen Tiempos medios de Respuesta significativamente menores que para la distribución 2, y a su vez para ésta se tienen Tiempos de Respuesta significativamente más pequeños que para la configuración 3.

Efecto simple de un factor



$$Ef_{F3} = \bar{X}_3 - \bar{X} =$$
$$= 3,5 - 2,74 = 0,76$$

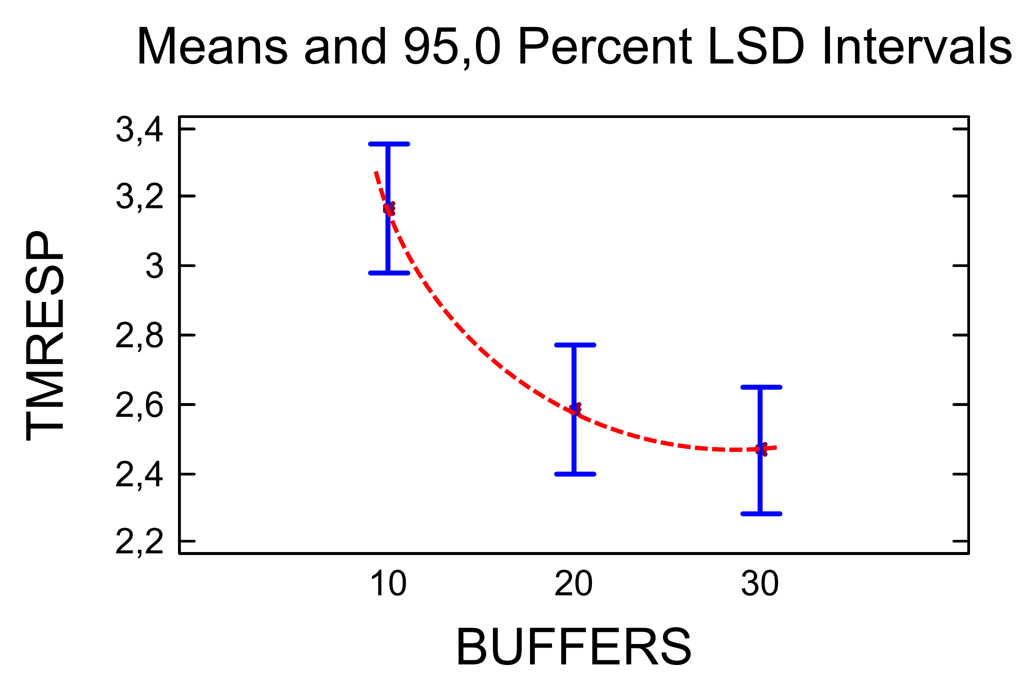
$$\bar{X} = 2'74$$

Media general

El efecto simple de un factor se define sobre el promedio de las condiciones estudiadas de los restantes factores

El efecto simple de la distribución de FICHEROS 3 hace aumentar el Tiempo MEDIO DE RESPUESTA en 0,76

EJEMPLO: Buffers y Ficheros



No tiene sentido decir que para 10 buffers el TM de Respuesta es mayor y que entre 20 y 30 buffers no hay diferencia, obteniéndose los menores TM de respuesta (aunque se ha de tener en cuenta).

Lo importante es si se aprecia o no una posible relación lineal o cuadrática para, en el caso de que ésta fuera significativa, obtener la expresión matemática correspondiente y obtener el **número óptimo de ficheros**, que puede ser no se haya contemplado en el experimento.

Análisis del efecto del factor BUFFERS

El factor número de BUFFERS ha resultado significativo en el ANOVA anterior, y se trata de un factor cuantitativo →

→ estudiaremos la naturaleza del mismo descomponiendo la $SC_{BUFFERS}$ en dos términos, cada uno con un grado de libertad, asociados a las componentes Lineal y Cuadrática del efecto del factor . (Técnica de Contrastes ortogonales)

BUFFERS ↔ $(J - 1)=(3-1)=2$ g.l.

Componente Lineal → 1 g.l.

Componente Cuad. → 1 g.l.

Contrastes Ortogonales: BUFFERS

- Posible efecto lineal → **$SC_{Z\ lineal}$** → 1 g.l
- Posible efecto cuadrático → **$SC_{Z\ Cuadrático}$** → 1 g.l

$$SC_Z = \frac{N_j \times Z^2}{\sum_{j=1}^{J} \lambda_j^2}$$

$$Z = \sum_{j=1}^{J} \lambda_j \bar{X}_j$$

Media de cada nivel j

Número de observ. con que se calcula cada $\bar{X}_j$

	3 niveles	4 niveles	5 niveles
λ_{Lineal}	-1 0 1	-3 -1 1 3	-2 -1 0 1 2
$\lambda_{Cuadrática}$	1 -2 1	1 -1 -1 1	2 -1 -2 -1 2
$\lambda_{Cúbica}$	----------	-1 3 -3 1	-1 2 0 -2 1

Contrastes Ortogonales: BUFFERS

	10	20	30	Z	SC_Z
Medias	3’17	2’58	2’47		
Lineal	-1	0	1	-0’7	1’47
Cuadrático	1	-2	1	0’46	0’23

$$Z_{lineal} = -1x3'17 + 0x2'58 + 1x2'47 = -0'7$$

$$Z_{Cuadrática} = 1x3'17 + (-2)x2'58 + 1x2'47 = 0'46$$

$$SC_{Z_{Lineal}} = \frac{6.(0'7)^2}{(-1)^2 + 0^2 + 1^2} = 1'47$$

$$SC_{Z_{Cuadratica}} = \frac{6.(0'46)^2}{1^2 + (-2)^2 + 1^2} = 0'23$$

Significación de las componentes lineal y cuadrática de la $SC_{BUFFERS}$

Origen Variación	SC	g.l.	CM	F ratio	F_tabla α=0'05
BUFFERS	**1'69**	**2**	**0'845**	**29'14**	**4'26 = $F_{2,9}$**
Ef. Lineal	**1'47**	**1**	**1'47**	**50'69** >	**5'12 = $F_{1,9}$**
Ef. Cuad.	**0'23**	**1**	**0'23**	**7'93** >	**5'12 = $F_{1,9}$**
Residual	**0'26**	**9**	**0'029**	**-**	

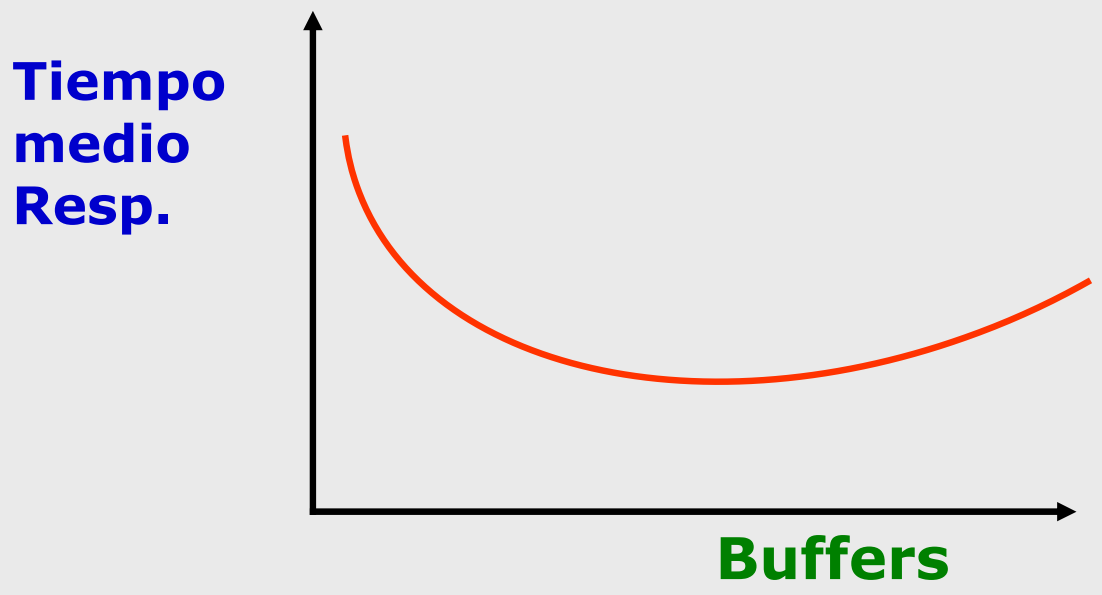
- Los **efectos** de las componentes **lineal** y **cuadrática** de **BUFFERS** son **significativos**, α=0'05

- RECORDAD:

$$SC_{BUFFERS} = SC_{Comp.\ Lineal} + SC_{Comp.\ Cuadrática}$$

Interpretación del efecto de las componentes lineal y cuadrática del factor BUFFERS

Z_{lineal} < 0 : A medida que aumenta el número de BUFFERS el tiempo medio de respuesta decrece linealmente, como **$Z_{Cuadrática}$ > 0** → A medida que aumenta el número de BUFFERS el tiempo medio de respuesta decrece cada vez más despacio



Significación de las componentes lineal y cuadrática de la $SC_{BUFFERS}$

Regresión Polinomial - TRES versus BUFFERS

Variable dependiente: TRES
Variable independiente: BUFFERS
Orden del polinomio = 2
Número de observaciones: 18

Parámetro	*Estimado*	*Error Estándar*	*Estadístico T*	*Valor-P*
CONSTANTE	4,21667	1,22041	3,45514	0,0035
BUFFERS	-0,128333	0,138583	-0,926039	0,3691
BUFFERS^2	0,00233333	0,00342904	0,680462	0,5066

Análisis de Varianza

Fuente	*Suma de Cuadrados*	*Gl*	*Cuadrado Medio*	*Razón-F*	*Valor-P*
Modelo	1,68778	2	0,843889	1,79	0,2002
Residual	7,055	15	0,470333		
Total (Corr.)	8,74278	17			

R-cuadrada = 19,3048 porciento
R-cuadrada (ajustada por g.l.) = 8,54547 porciento
Error estándar del est. = 0,685809
Error absoluto medio = 0,524074
Estadístico Durbin-Watson = 1,13119 (P=0,0052)

ANOVA para las Variables según Orden de Ajuste

Fuente	*Suma de Cuadrados*	*Gl*	*Cuadrado Medio*	*Razón-F*	*Valor-P*
BUFFERS	1,47	1	1,47	3,13	0,0974
BUFFERS^2	0,217778	1	0,217778	0,46	0,5066
Modelo	1,68778	2			

¡OJO!

Efecto de la interacción doble

Recuerda que…….

Cuando se estudia más de un factor, aparece el concepto de **INTERACCIÓN**

- Puede ser **doble**, **triple**, etc, según sea entre dos, tres, … factores. El estudio de interacciones de orden superior a 3 son difíciles de interpretar y rara vez tienen sentido

- Aparece cuando **el efecto de un determinado nivel (o variante) de un factor es diferente para cada nivel (o variante) del otro factor**

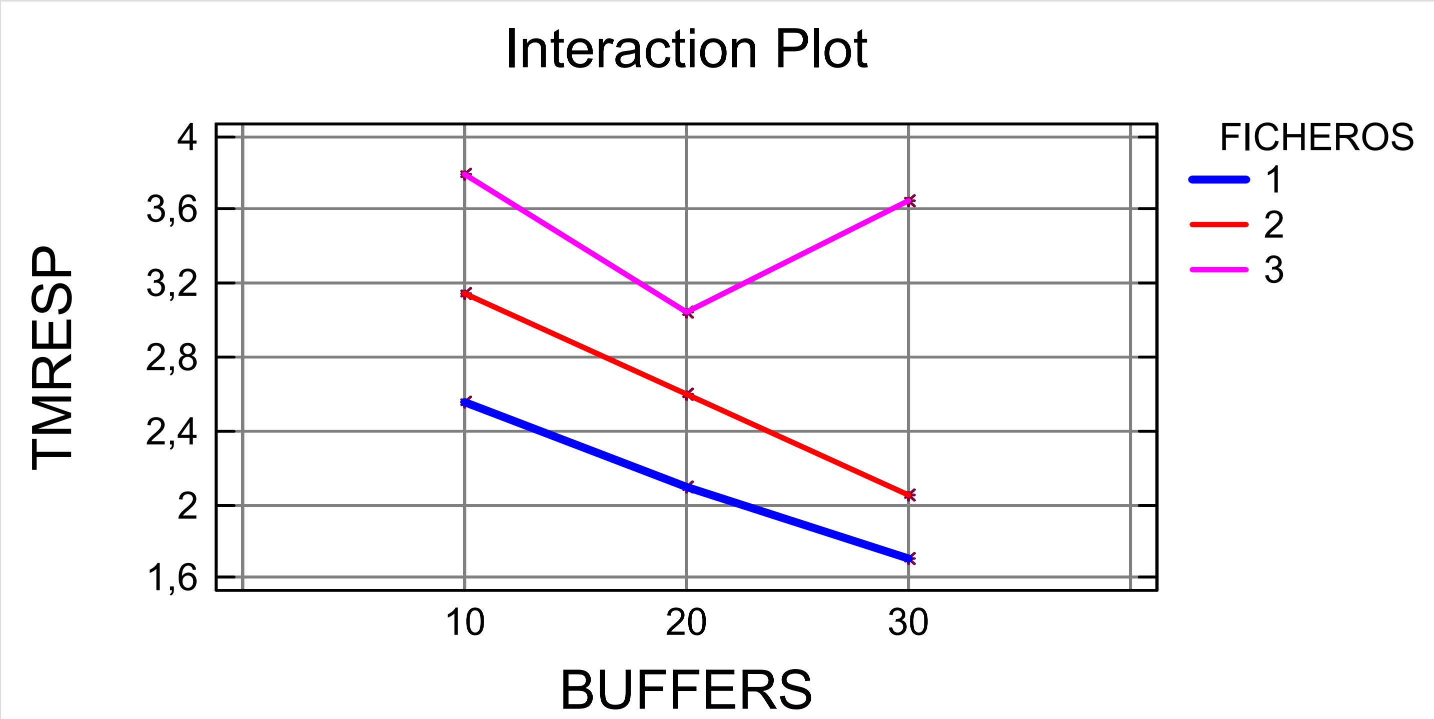
NOTA: Descomposición de las interacciones

➔ En general, si en un ANOVA en el que intervienen varios factores:

- resulta significativa una interacción doble en la que al menos uno de los factores implicados es de tipo cuantitativo y tiene más de dos niveles conviene interpretar con detalle la naturaleza de dicha interacción ➔ descomponiendo la SC de la misma en una serie de términos asociados a distintas componentes de la misma

- La descomposición se plantea de diferente forma según sea de naturaleza cuantitativa uno o ambos factores

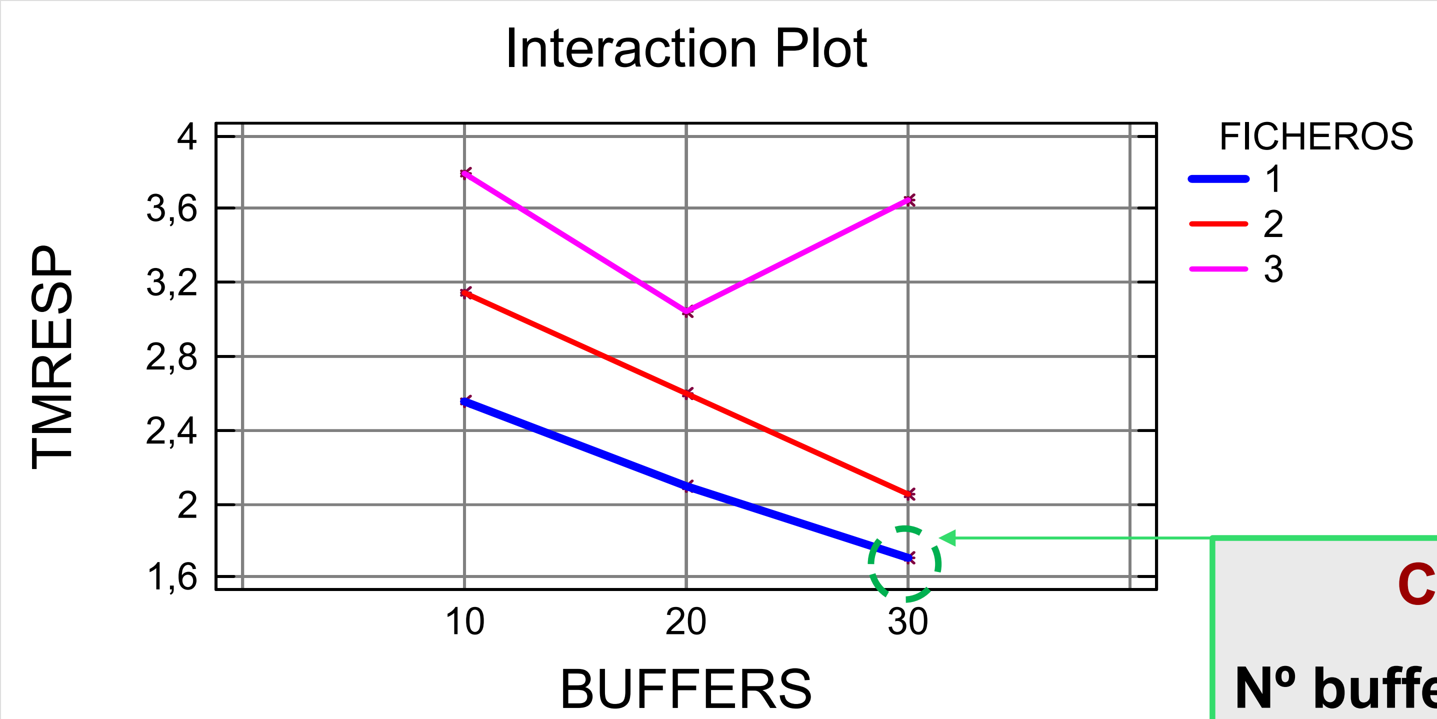
Efecto de la interacción doble



El efecto de la interacción existe porque el efecto del factor Buffers es distinto para las diferentes variantes de ficheros

Condiciones Operativas Óptimas

¿Qué número de buffers y qué protocolo se deberían usar para que el TM de Respuesta fuera el menor?



COO

Nº buffers = 30

Distribución de
Ficheros = 1

CONCLUSIONES:

- Para el promedio de **BUFFERS** ensayados, los ficheros de tipo 1 tienen un tiempo de respuesta en promedio inferior al de los ficheros de tipo 2, y éstos, a su vez, inferior a los del tipo 3.
- Para el promedio de **FICHEROS** ensayados, el tiempo de respuesta medio tiende a disminuir conforme aumenta el número de buffers, siendo esta disminución cada vez más pequeña conforme aumenta dicho número de buffers. Finalmente se alcanza un mínimo y el tiempo respuesta medio aumenta con el número de buffers.
- Respecto a la interacción **BUFFERS*FICHEROS**, el tiempo medio de respuesta tiende a disminuir linealmente conforme aumenta el número de buffers para los ficheros de tipo 1 y 2. No existe en este caso tendencia cuadrática. Sin embargo, para los ficheros de tipo 3 existe un efecto cuadrático positivo, no existiendo un efecto lineal significativo.

Ejemplo 3

- Se realizó un estudio en una acería para analizar la influencia sobre el **alargamiento máximo hasta la rotura** (variable E_{max}) en barras corrugadas de acero, de la **calidad del acero** (dos calidades: B400SD, tipo 1, y B500SD, tipo 2) y del **diámetro de la barra** (3 diámetros: 8mm, 16 mm y 24 mm)

  Para cada calidad y diámetro se seleccionaron aleatoriamente 5 barras, cada una de una colada diferente, determinándose en cada una el valor de E_{max} mediante un ensayo de tracción-deformación. Los resultados obtenidos se recogen en la tabla adjunta. (Se han incluido también en la misma, en negrita, los totales de los 5 datos de cada casilla, de las filas, de las columnas y el total general)

Ejemplo 3: Valores del alargamiento a la rotura en barras corrugadas de acero

		DIÁMETRO			
		8 mm	**16 mm**	**24 mm**	
CALIDAD	**1**	15.29	16.91	17.23	
		15.89	16.99	17.81	
		16.02	17.27	17.74	
		16.56	16.85	18.02	
		15.46	16.35	18.37	
		79.22	**84.37**	**89.17**	**252.76**
	2	12.07	12.92	13.30	
		12.42	13.01	12.82	
		12.73	12.21	12.49	
		13.02	13.49	13.55	
		12.05	14.01	14.53	
		62.29	**65.64**	**66.69**	**252.76**
		141.51	**150.01**	**155.86**	**447.38**

Ejemplo 3

Análisis de efectos sobre medias

Los cálculos para obtener la tabla resumen del ANOVA son los siguientes:

$$SG = \frac{447.38^2}{30} = 6671.63$$

$$SC_{Total} = 15.29^2 + 15.89^2 + ... + 14.53^2 - SG = 131.807 \qquad gl_{Totales} = 30 - 1 = 29$$

$$SC_{Calidad} = \frac{252.76^2 + 194.62^2}{15} - SG = 112.675 \qquad gl_{Calidad} = 2 - 1 = 1$$

$$SC_{Diametro} = \frac{141.51^2 + 150.01^2 + 155.86^2}{10} - SG = 10.413 \qquad gl_{Diametro} = 3 - 1 = 2$$

$$SC_{Cal*Diam} = \frac{79.22^2 + 84.37^2 + ... + 66.69^2}{10} - SG - SC_{Calidad} - SC_{Diametro} = 1.604 \qquad gl_{Cal*Diam} = 1x2 = 2$$

$$SC_{Residual} = SC_{Total} - SC_{Calidad} - SC_{Diametro} - SC_{Cal*Diam} = 7.115 \qquad gl_{Calidad} = 29 - 1 - 2 - 2 = 24$$

Ejemplo 3

La tabla resumen del ANOVA es, por tanto

Analysis of Variance for EMAX - Type III Sums of Squares

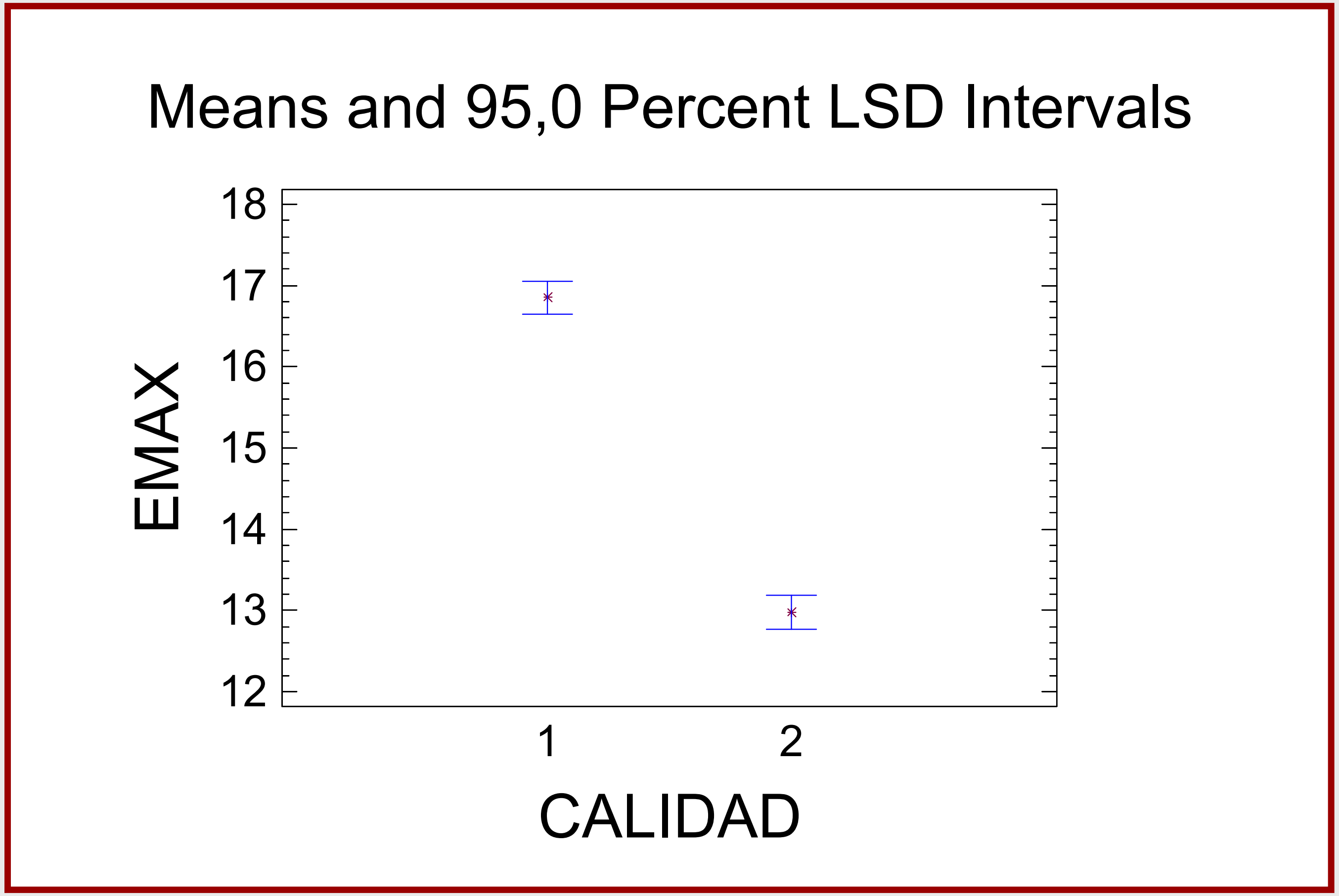
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:CALIDAD	112,675	1	112,675	380,08	0,0000
B:DIAMETRO	10,4132	2	5,20658	17,56	0,0000
INTERACTIONS					
AB	1,6035	2	0,80175	2,70	0,0873
RESIDUAL	7,1148	24	0,29645		
TOTAL (CORRECTED)	131,807	29			

All F-ratios are based on the residual mean square error.

Como se puede apreciar en la tabla, los efectos simples de “Calidad” y de “Diámetro” son claramente significativos (*p-values* muy bajos), mientras que la interacción es dudosa (*p-value* comprendido entre el 5% y el 10%)

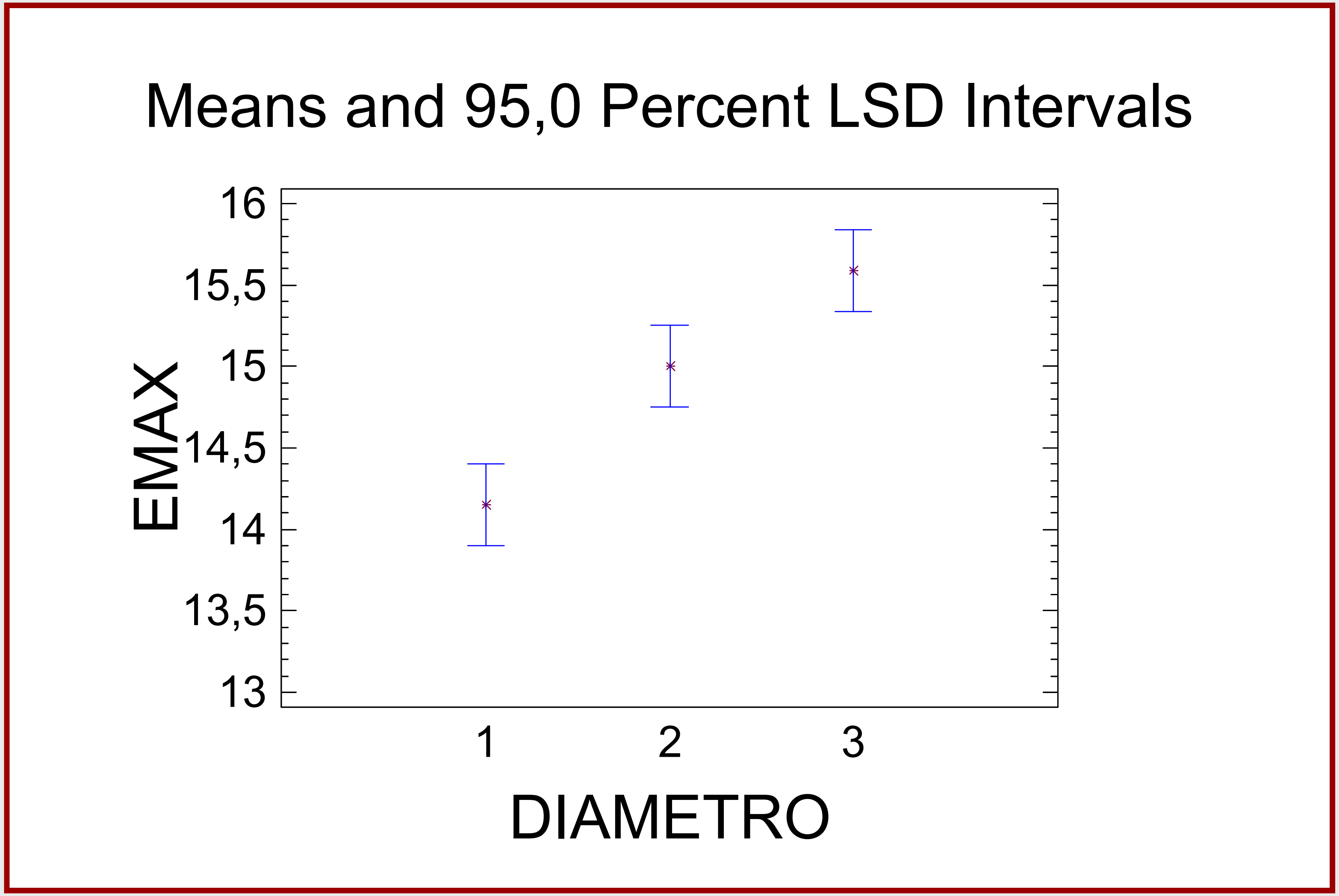
Ejemplo 3

Para precisar los efectos simples, se muestran los gráficos con los correspondientes intervalos LSD:



En el gráfico se aprecia que el E_{max} medio es significativamente más alto en la calidad 1 (B400SD) que en la 2 (B500SD)

Ejemplo 3



En el gráfico se aprecia que el E_{max} medio aumenta progresivamente al hacerlo el diámetro de la barra

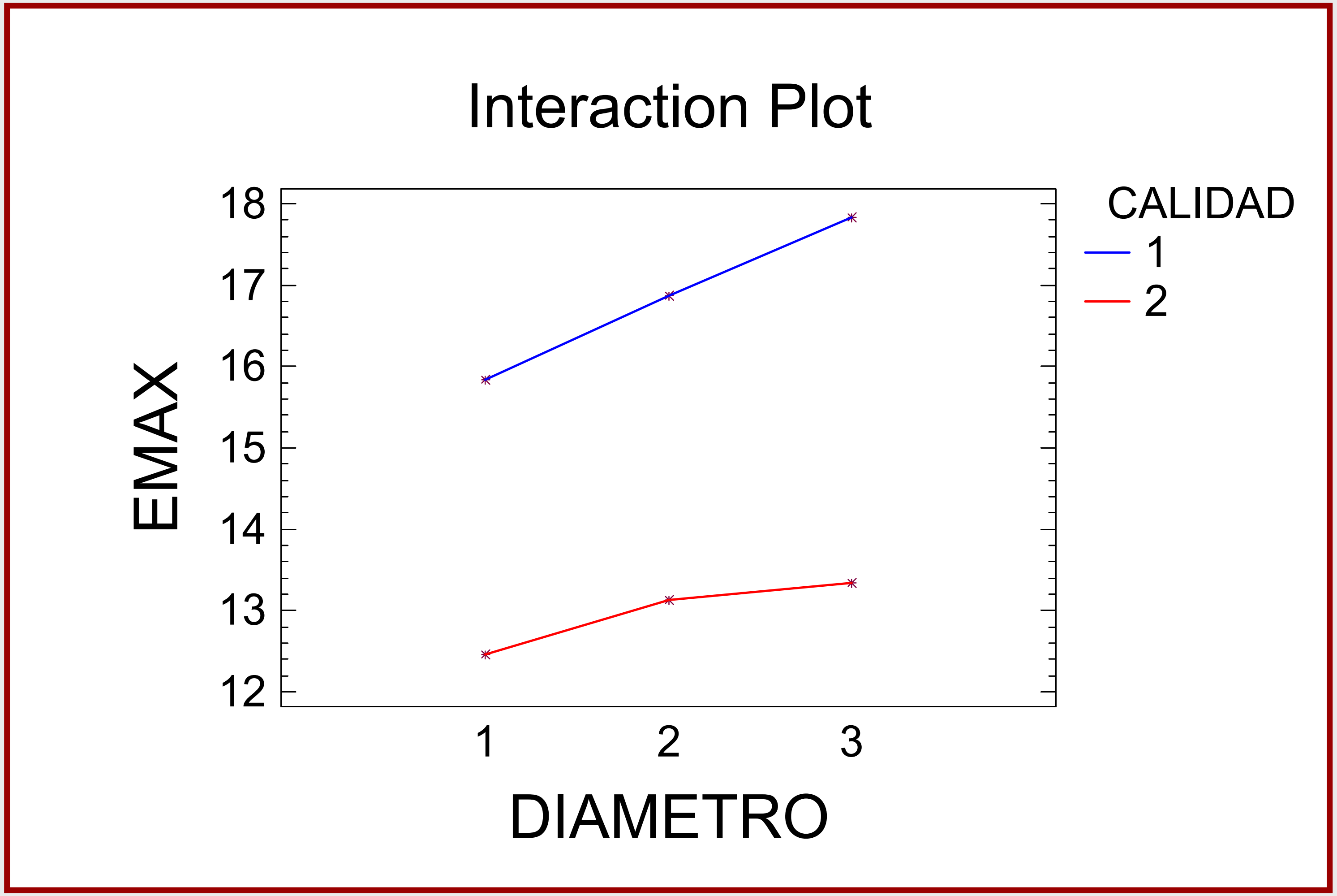
Ejemplo 3

¿Cómo podrías estudiar con detalle el efecto del factor Diámetro de las barras sobre el alargamiento hasta la rotura de las barras de acero?

Ejemplo 3

- Para analizar la interacción, que en este caso ha resultado dudosa, es útil el siguiente gráfico, obtenido mediante *Statgraphics*, que visualiza los valores medios obtenidos para las diferentes combinaciones de los factores "Calidad" con "Diámetro"

Ejemplo 3



En el gráfico parece apreciarse que el incremento en el promedio de E_{max} al aumentar los diámetros es más marcado con el acero de calidad 1 (B400SD) que en el de tipo 2 (B500SD)

Ejemplo 3

Análisis de efectos sobre varianzas

- También es interesante analizar si la dispersión, en torno a su media, de los valores de E_{max} es mayor en una calidad o en otra o en unos diámetros que en otros.

- Se trata de investigar el efecto de los factores estudiados sobre la varianza del E_{max} como complemento al estudio realizado de efectos sobre la media

- Este análisis puede llevarse a cabo de manera sencilla realizando un nuevo ANOVA en el que la variable respuesta sea el cuadrado de los residuos del análisis anterior

Ejemplo 3

Análisis de efectos sobre varianzas

Analysis of Variance for RESIDUALS^2 - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:CALIDAD	0,266752	1	0,266752	2,56	0,1226
B:DIAMETRO	0,106044	2	0,0530221	0,51	0,6075
INTERACTIONS					
AB	0,236566	2	0,118283	1,14	0,3380
RESIDUAL	2,50026	24	0,104178		
TOTAL (CORRECTED)	3,10963	29			

All F-ratios are based on the residual mean square error.

En la tabla resumen del ANOVA ninguno de los posibles efectos sobre la varianza resulta estadísticamente significativo, al ser todos los *p-values* superiores al 10%

Contenidos

- ANOVA con varios factores controlados
  - Efecto simple. Interacción
  - Estimación de efectos
  - Predicciones
  - Ejemplos
- Cuestiones adicionales en el ANOVA
- Ejercicios

Hipótesis básicas del ANOVA

- Si se estudian **2 factores**:

FI → **I** niveles (cuantitativo) o variantes (cualitativos)

FJ → **J** niveles (cuantitativo) o variantes (cualitativos)

X (variable respuesta)

Ix**J** tratamientos → **I**x**J** poblaciones

El valor constatado en el ensayo **n** de cada uno de los **I**x**J** tratamientos es el valor observado de una v.a → $\mathbf{X_{ijn}}$

Hipótesis básicas del ANOVA

- Los resultados obtenidos mediante el ANOVA se deducen de las siguientes hipótesis:

  - Independencia
  - Homocedasticidad (igualdad de varianzas)
  - Normalidad

- **¿**El incumplimiento de alguna de éstas hipótesis → Los resultados obtenidos mediante el ANOVA **no** son **válidos?**

Tengamos en cuenta…….

Ningunos datos reales seguirán exactamente un modelo matemático!!!!

Hasta qué punto las **conclusiones** que pueden obtenerse de nuestro análisis son <u>sensibles</u> al hecho de que las pautas de variabilidad constatadas en los datos <u>difieran marcadamente</u> de las postuladas por el modelo y <u>qué medidas</u> pueden tomarse en estos casos

Independencia

✓ Las **n** observaciones para cada combinación o tratamiento corresponden a individuos extraídos independientemente de la población considerada (es decir, constituyen una muestra de la población en estudio obtenida al azar, m.a.s.)

✓ Las observaciones obtenidas en la aplicación de los diferentes tratamientos son también independientes entre sí

- Incumplimiento: el ANOVA no es robusto frente a éste incumplimiento → Hipótesis más importante
- Responsabilidad del investigador
- Solución: **ALEATORIZACIÓN**

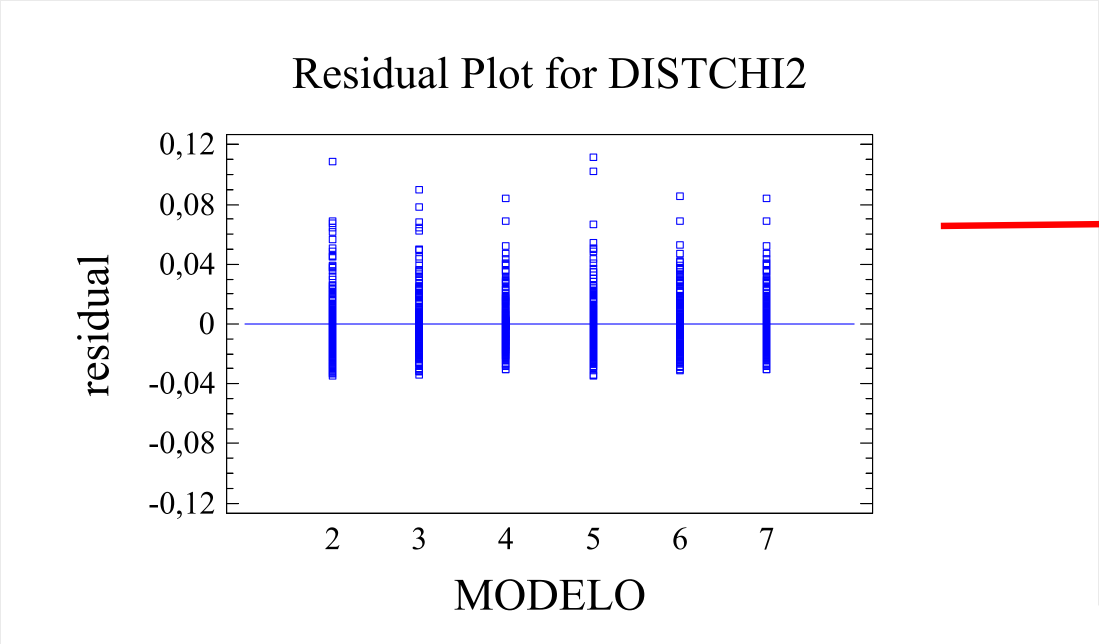
Homocedasticidad

✓ Las **IxJ poblaciones** estudiadas tienen la misma σ^2

✓ Una estimación de la σ^2 común en las poblaciones es el Cuadrado Medio Residual (CMR)

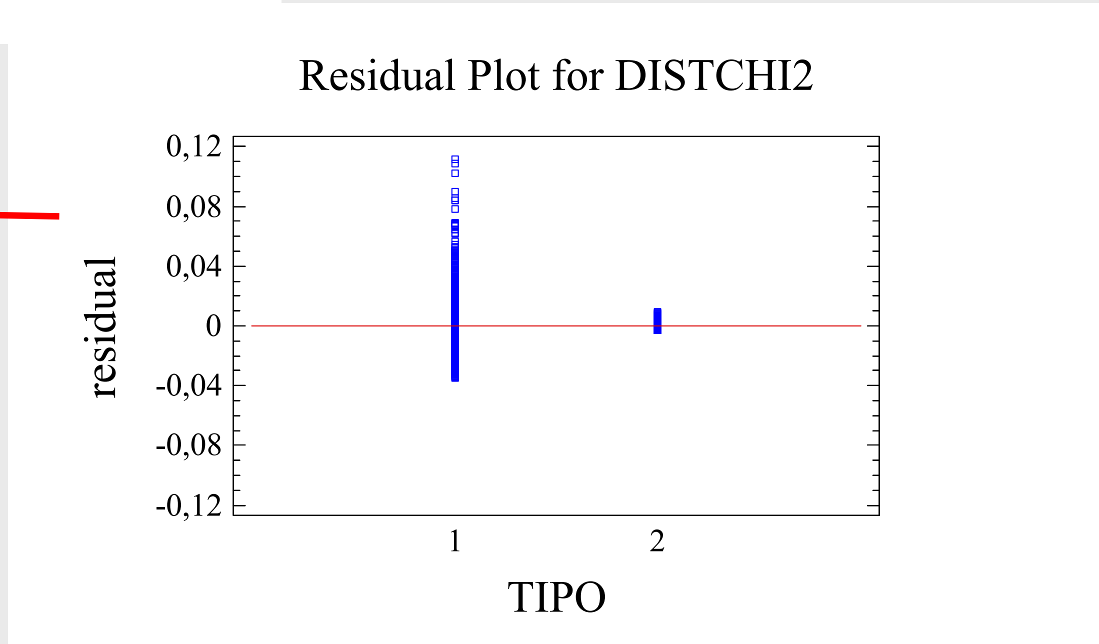
- Incumplimiento:
  - Existencia de efectos significativos sobre la dispersión → sobre la variabilidad de **X**
  - El ANOVA es moderadamente robusto frente a éste, especialmente si el diseño es equilibrado.
- Detección:
  - Gráficos de residuos frente a los niveles de los factores
  - ANOVA sobre el cuadrado de los residuos
- Solución: en algunos casos (heterocedasticidad por fenómenos de escala) → transformación logarítmica de la variable respuesta.

Homocedasticidad



Homocedasticidad

Heterocedasticidad



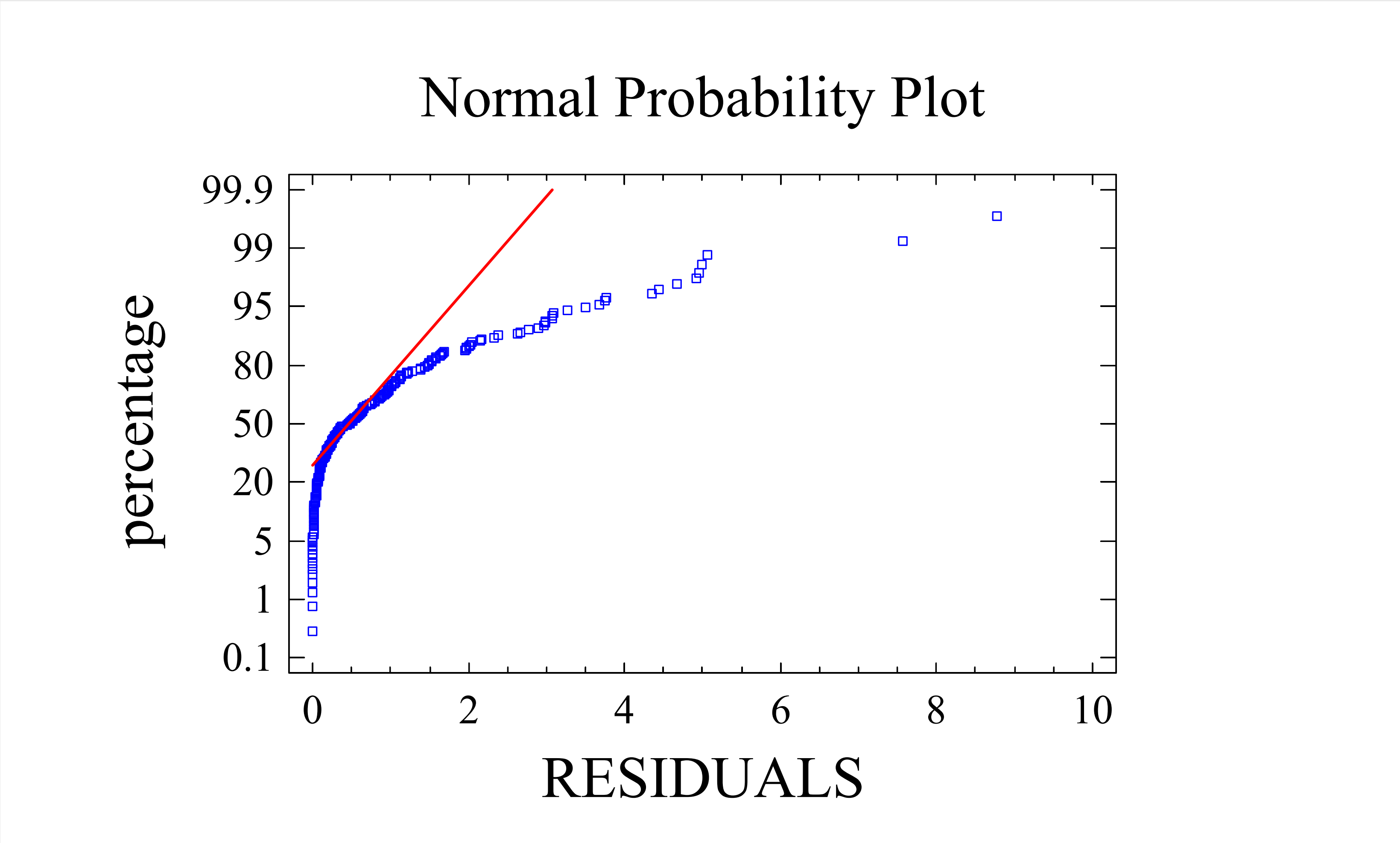
Normalidad

✓ Las **IxJ poblaciones** estudiadas son Normales

✓ Una estimación de la σ^2 común en las poblaciones es el Cuadrado Medio Residual (CMR)

- Incumplimiento:
  - El ANOVA es bastante robusto frente a éste
- Detección:
  - Gráficos de residuos en Papel Probabilístico Normal
- Solución: en casos de distribuciones con asimetría positiva → transformación logarítmica de la variable respuesta

Normalidad



Los residuos NO se alinean siguiendo una recta

Uso de transformaciones

- Transformación logarítmica (log X):
  - Asimetría positiva de las poblaciones
  - Asimetría positiva de las poblaciones y heterocedasticidad asociada a fenómenos de escala
  - Interpretación sencilla de los efectos estudiados de naturaleza multiplicativa
- Transformación: $\sqrt{X}$
  - Variable respuesta es el resultado de un conteo "*número de veces que....*"
- Transformación: $\text{arcsen}\sqrt{X}$
  - Variable respuesta es una proporción "*porcentaje de veces que....*"

Contenidos

- ANOVA con varios factores controlados
  - Efecto simple. Interacción
  - Estimación de efectos
  - Predicciones
  - Ejemplos
- Cuestiones adicionales en el ANOVA
- Ejercicios

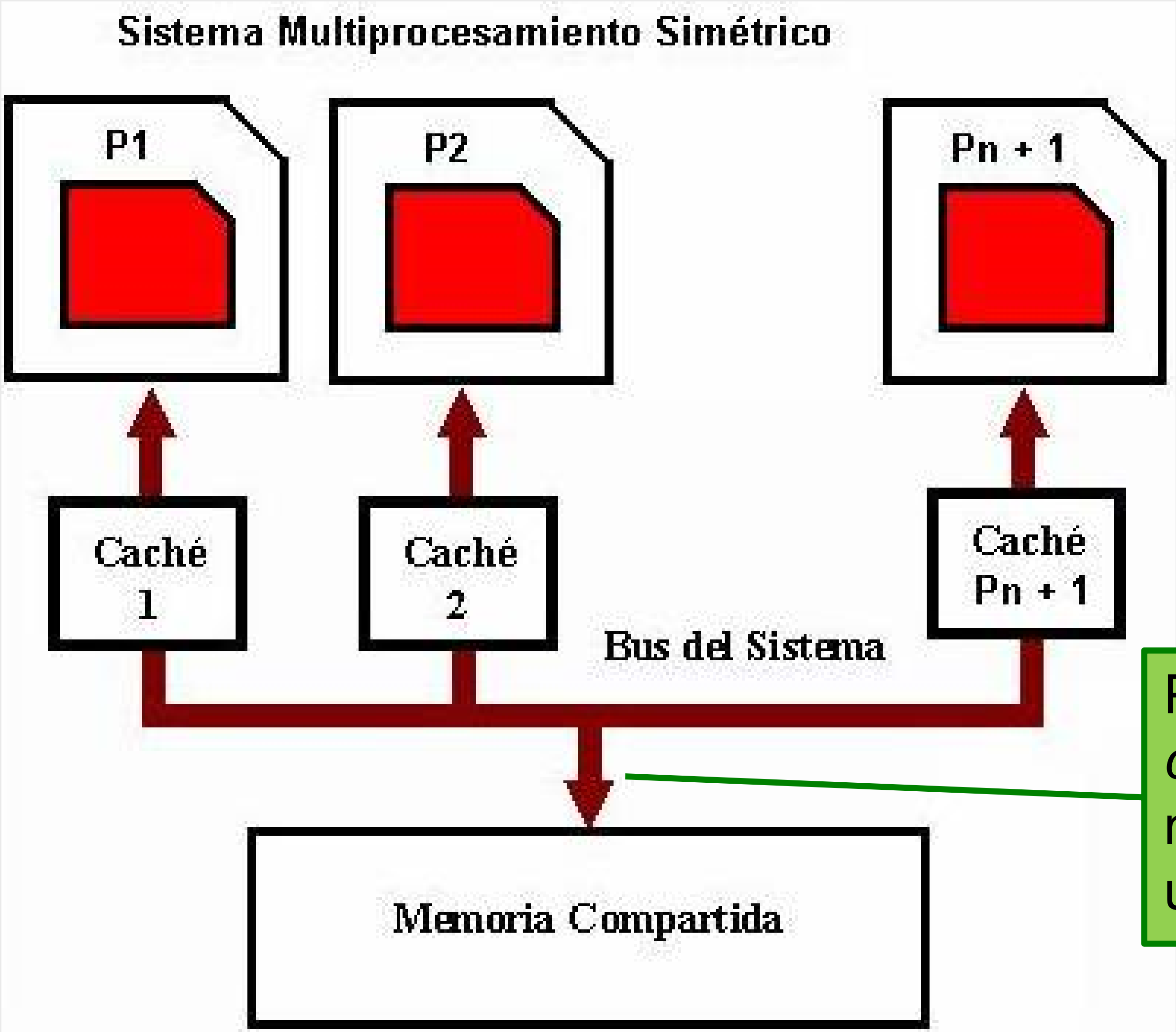
Ejercicio 1

Con el objeto de analizar el comportamiento de los sistemas de memoria caché en un tipo de multiprocesador se plantea llevar a cabo un estudio de la influencia de dos de las características más importantes de estos sistemas (Nº de procesadores y protocolo) sobre las prestaciones de los mismos (3 niveles y 3 variantes)

Cada uno de los 9 tratamientos se ensayó 3 veces, midiéndose en cada prueba la tasa de fallos (%) de los sistemas de memoria producida por la ejecución de un determinado programa tipo.

Los datos obtenidos se recogen en la siguiente tabla:

Ejercicio 1



CPU's

Para mantener la *coherencia* entre las memorias necesitamos un *protocolo*

Ejercicio 1

		PROTOCOLO (PROT)		
		MSI	MESI	DRAGON
Nº DE PROCESADORES (NPRO)	2	25,80	30,00	35,25
		26,25	32,25	29,10
		24,60	33,75	41,55
	4	32,40	40,05	48,45
		28,80	37,35	47,70
		30,90	37,05	45,75
	6	19,05	20,10	14,70
		15,60	19,95	16,65
		14,70	21,90	20,10

DATOS ADICIONALES
$SCT = 2.677,36$
$SC_{PROT} = 379,8$
$SC_{PROTxNPRO} = 212,92$
$CM_{NPRO} = 977,41$

Ejercicio 1

a) ¿Cuál es la variable respuesta?, ¿cuáles son los factores?, indica de qué tipo son.

b) Construye la tabla resumen del ANOVA e indica qué efectos han resultado significativos y por qué (α=0,05). Explica los cálculos realizados.

c) Analiza el efecto del tipo de protocolo utilizado mediante los intervalos LSD que se acompañan, e indica qué porcentaje de fallos se ha obtenido, en promedio, para cada variante.

Ejercicio 1

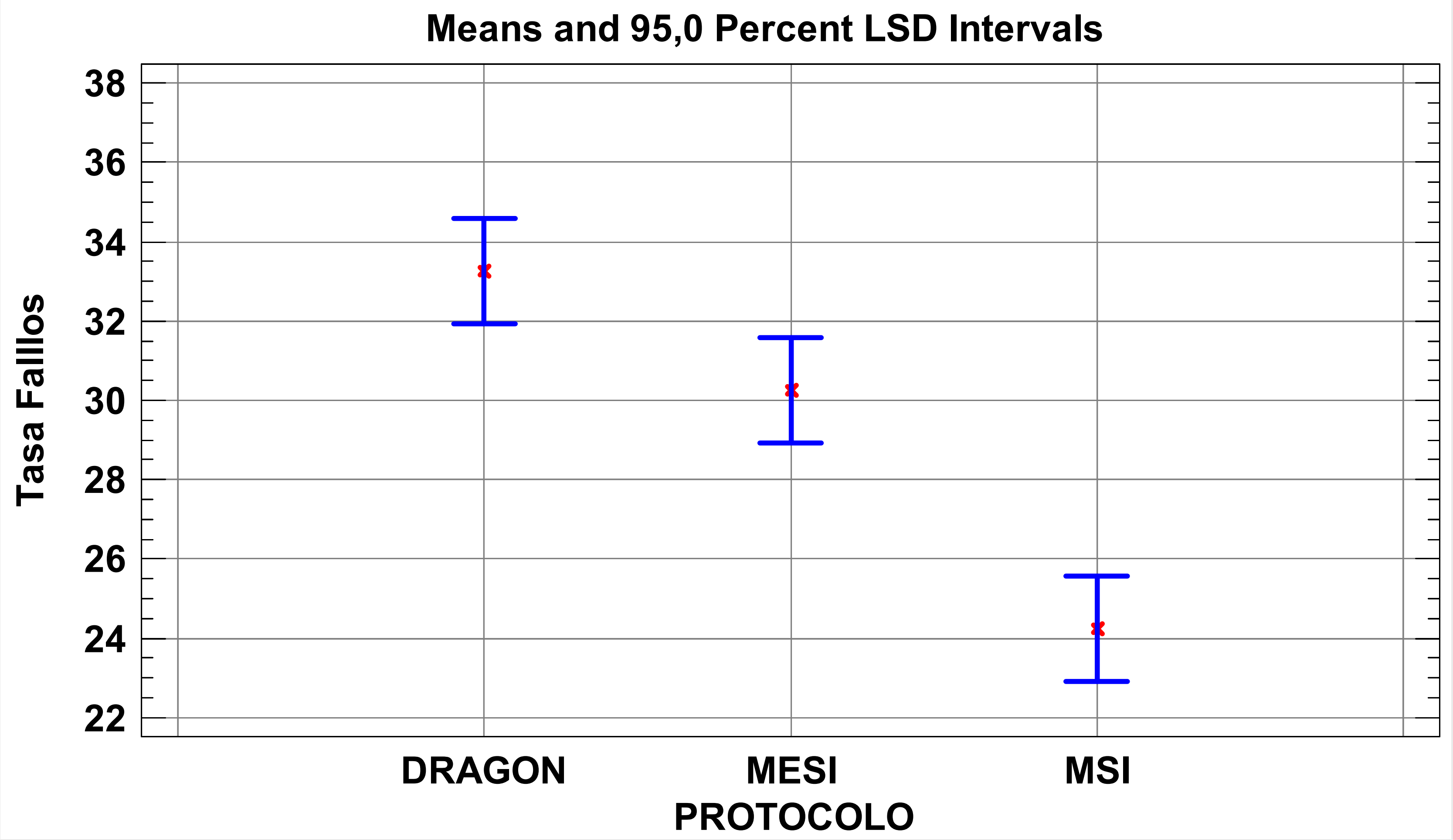
d) Estudia la naturaleza del efecto del número de procesadores a nivel descriptivo mediante los correspondientes gráficos de medias. ¿Existen indicios de una posible relación lineal o cuadrática (positiva o negativa) entre el porcentaje de fallos de los sistemas de memoria y el número de procesadores? Justifica la respuesta.

e) Interpreta gráficamente los gráficos de la interacción entre el tipo de protocolo y el número de procesadores.

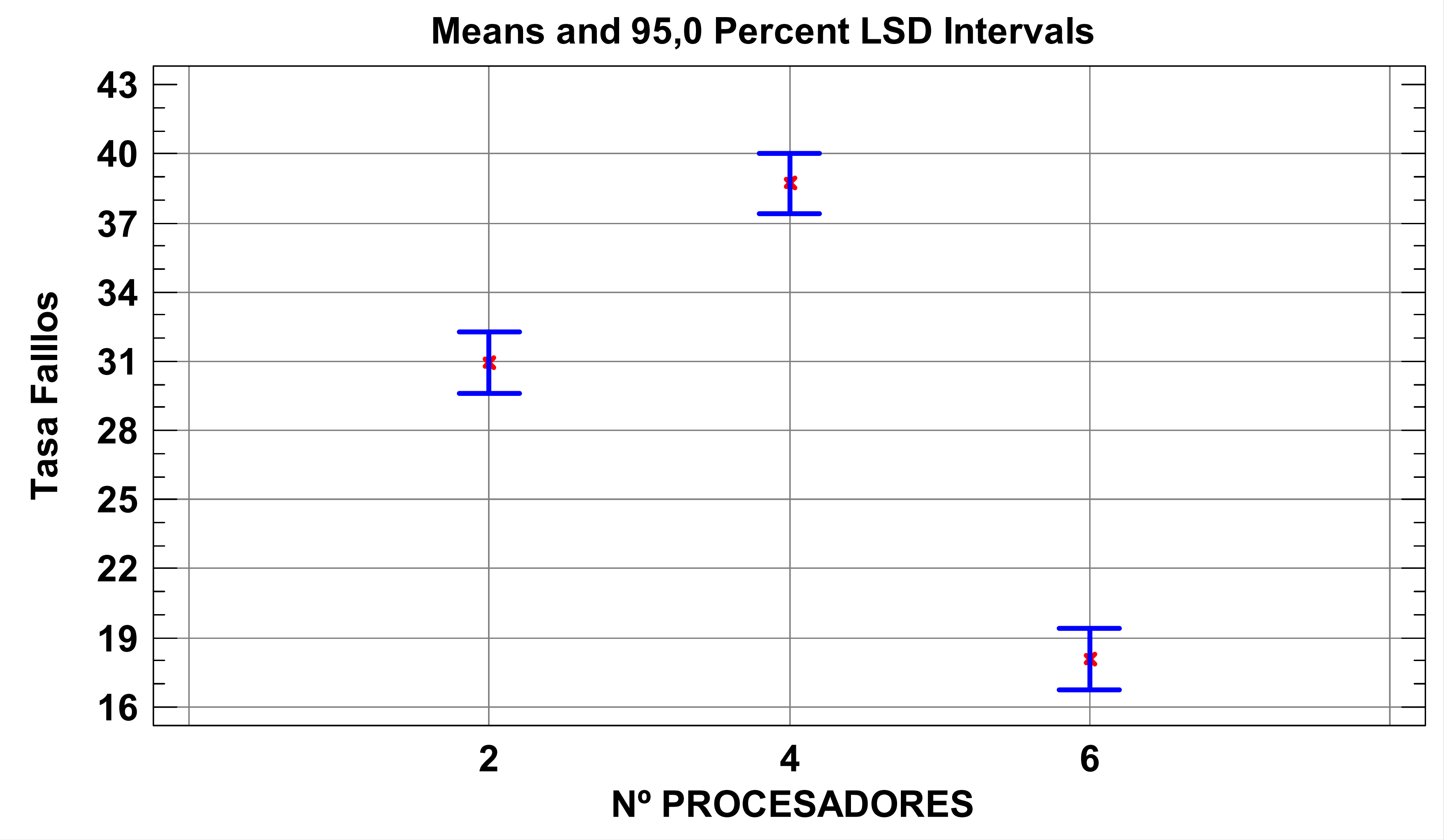
Ejercicio 1: Tabla Resumen del ANOVA

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:NPRO	1954,82	2	977,41	135,52	0,0000
B:PROT	379,805	2	189,902	26,33	0,0000
INTERACTIONS					
AB	212,915	4	53,2287	7,38	0,0011
RESIDUAL	129,825	18	7,2125		
TOTAL (CORRECTED)	2677,37	26			

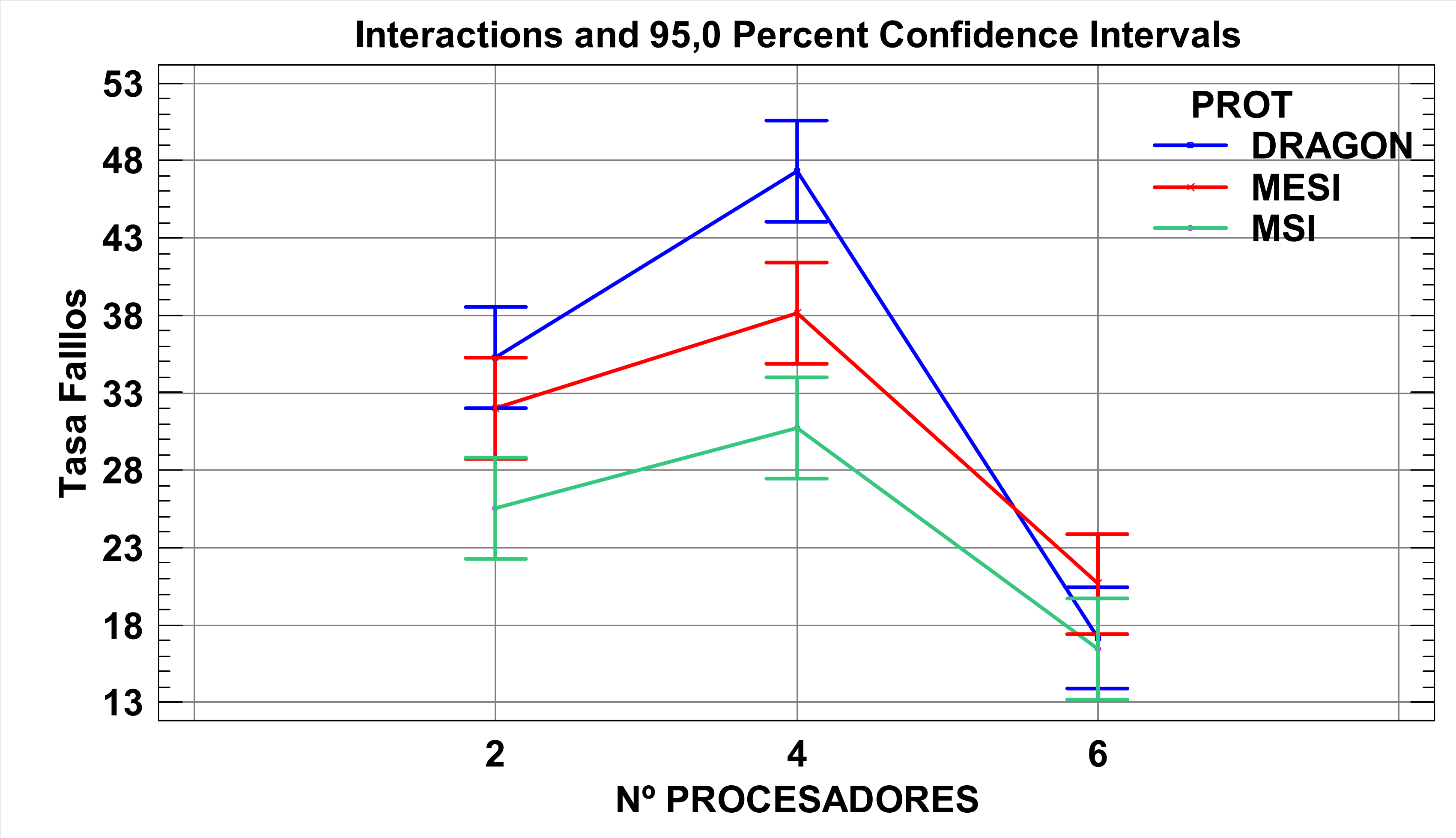
Ejercicio 1: Intervalos LSD



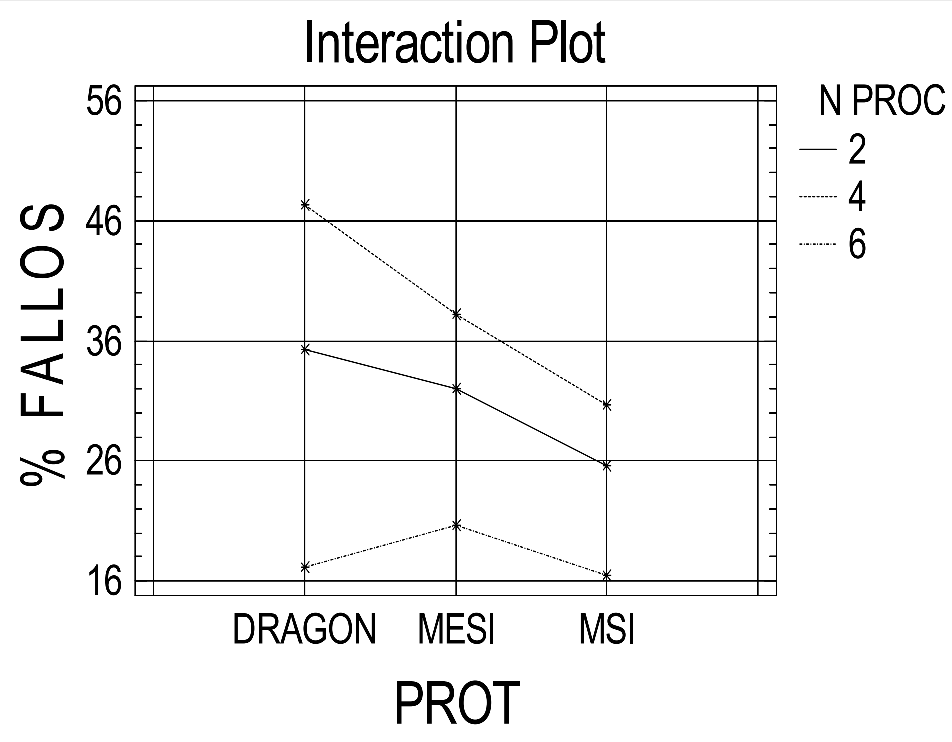
Ejercicio 1: Intervalos LSD

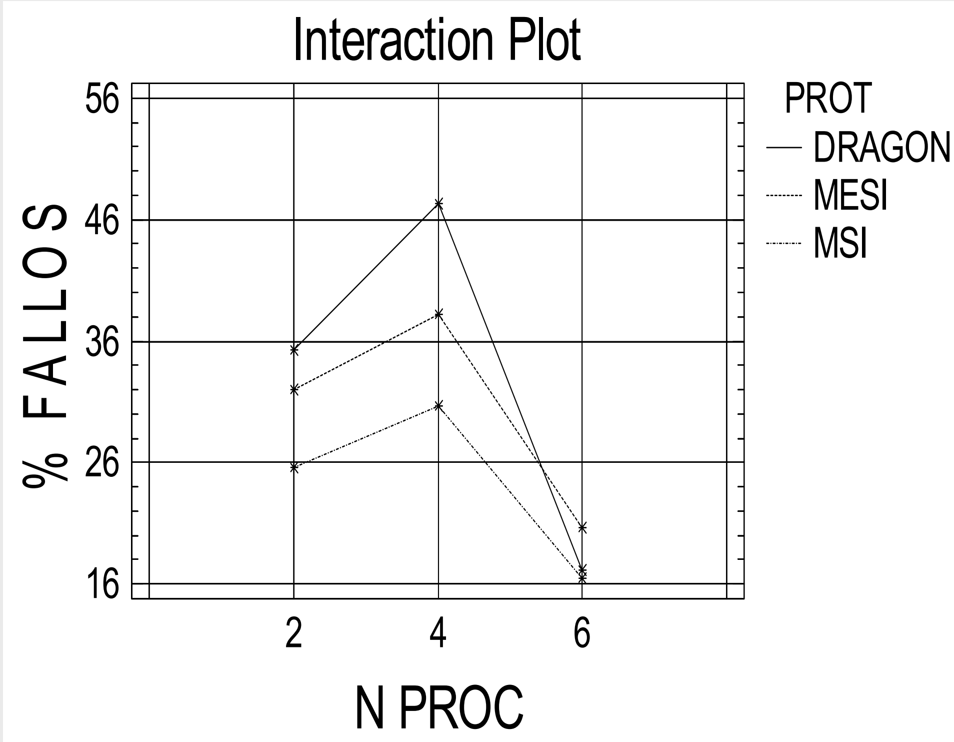


Ejercicio 1: Interacción doble



Ejercicio 1: Interacción doble





Ejercicio 2

Se ha llevado a cabo un diseño de experimentos con el objeto de conocer la posible influencia de dos tipos de puerto de conexión (codificados como A y B) y de tres niveles de memoria RAM (64, 128 y 192), sobre los tiempos medios de transmisión de un servidor. Los resultados del experimento, expresados en segundos por Mb de información, son los que se indican a continuación:

		RAM		
		64	**128**	**192**
Conexión	**A**	5,462	3,308	2,692
		5,769	3,923	2,923
		5,615	4,231	3,077
	B	6,154	5,231	4,307
		6,538	5,307	4,615
		6,077	5,538	4,692

Ejercicio 2

Se ha llevado a cabo un diseño de experimentos con el objeto de conocer la posible influencia de dos tipos de puerto de conexión (codificados como A y B) y de tres niveles de memoria RAM (64, 128 y 192), sobre los tiempos medios de transmisión de un servidor. Los resultados del experimento, expresados en segundos por Mb de información, son los que se indican a continuación:

		RAM		
		64	**128**	**192**
Conexión	**A**	5,462	3,308	2,692
		5,769	3,923	2,923
		5,615	4,231	3,077
	B	6,154	5,231	4,307
		6,538	5,307	4,615
		6,077	5,538	4,692

Ejercicio 2

a) Realizar un ANOVA para estudiar qué efectos resultan significativos sobre el tiempo medio de transmisión del servidor, teniendo en cuenta los dos factores en estudio y su interacción (α=0,1).

NOTA: SCT= 24,0078; SC_{CONX}=7,295; CM_{RAM}=7,493; CMR=0,0683.

b) Estudia la naturaleza Indicar cuál de los dos tipos de puerto resulta más interesante. Justifica el procedimiento utilizado para llegar a la decisión adoptada.

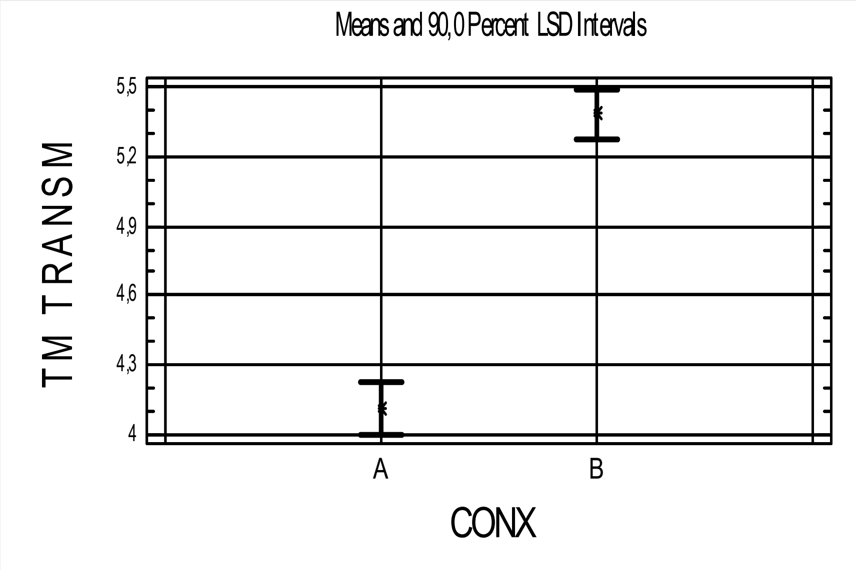
Ejercicio 2

c) del factor memoria a nivel descriptivo mediante los correspondientes gráficos de medias. ¿Existen indicios de una posible relación lineal o cuadrática (positiva o negativa) entre el tiempo medio de transmisión y el tamaño de memoria? Justifica la respuesta.

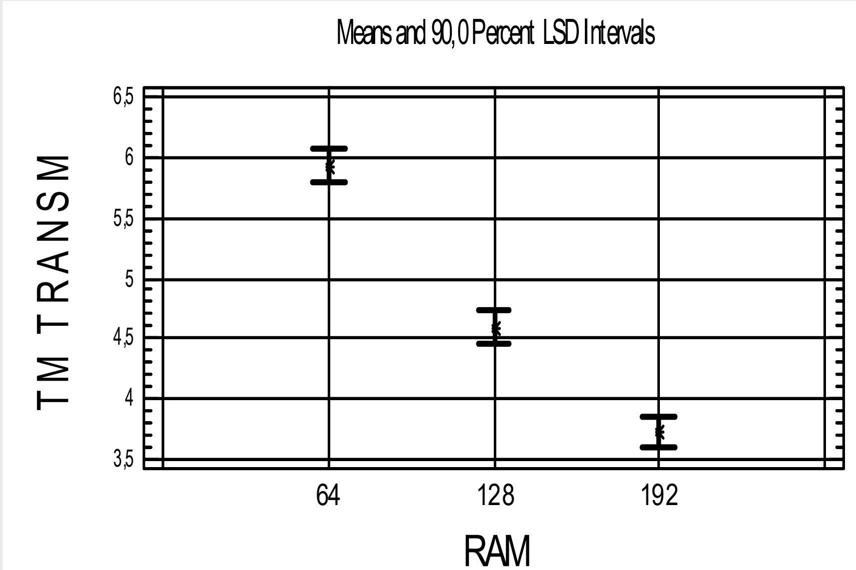
d) Interpreta gráficamente los gráficos de la interacción entre el tipo de conexión y el tamaño de memoria.

e) ¿Con qué tipo de conexión y cantidad de RAM se debería trabajar con el fin de obtener el menor tiempo medio de transmisión?. Justifica la respuesta.

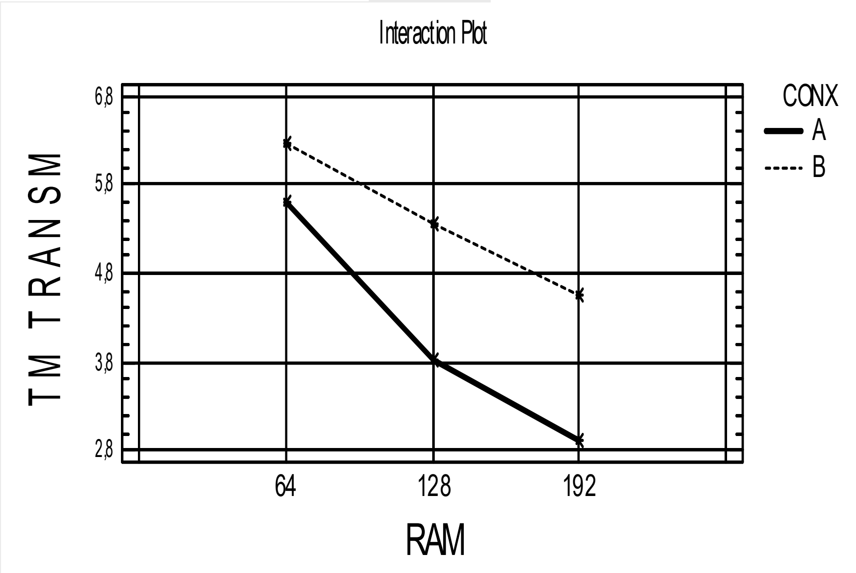
Ejercicio 2



Means and 90,0 Percent LSD Intervals
TM TRANSM
5,5
5,2
4,9
4,6
4,3
4
A
B
CONX



Means and 90,0 Percent LSD Intervals
TM TRANSM
6,5
6
5,5
5
4,5
4
3,5
64
128
192
RAM



Interaction Plot
TM TRANSM
6,8
5,8
4,8
3,8
2,8
64
128
192
RAM
CONX
A
B

Ejercicio 3

- Cierta fábrica desea mejorar la **resistencia a la torsión** de las adhesiones de componentes electrónicos sobre placas. Se estudian **dos tipos de pegamentos** (A y B) y **tres temperaturas de curado** (60ºC, 80ºC y 100ºC), analizándose dos componentes para cada uno de los 6 tratamientos posibles, obteniéndose los resultados que se presentan en la tabla siguiente. (Los datos también están disponibles en el fichero de *Statgraphics* *EJEM_PLANES_FAC.sgd*)

	60ºC		80ºC		100ºC	
Pegamento A	2,5	2,8	3,8	3,4	4,0	4,2
Pegamento B	1,6	1,2	3,2	2,8	4,3	4,7

Ejercicio 3

- Realizar el análisis estadístico de los datos respondiendo a las siguientes cuestiones:

    1) ¿Resultan significativos los factores analizados?

    2) ¿Cuáles serían las condiciones operativas óptimas que permiten aumentar la resistencia a la torsión de los componentes electrónicos?

    3) ¿Cuál será la resistencia media prevista en tales condiciones?

Ejercicio 4

- Se ha realizado una prueba de esfuerzo en diferentes individuos con el fin de estudiar la posible influencia de tres factores: nivel de grasa corporal, género y tabaco. La variable registrada fue el tiempo que los individuos permanecían corriendo en una cinta controlada. Cada uno de los tratamientos se ensayó en tres individuos diferentes. Los tiempos obtenidos en las pruebas se muestran en la tabla siguiente.

- Los factores y las diferentes alternativas analizadas:

  Nivel de grasa corporal: bajo, alto

  Género: masculino, femenino

  Historia del individuo como fumador: nada, poco, muy fumador

Ejercicio 4

		Género					
		Masculino			Femenino		
		Hª Fumador			Hª Fumador		
		Nada	Poco	Mucho	Nada	Poco	Mucho
Grasa corporal	Bajo	34, 32, 31	27, 24, 23	20, 21, 24	25, 35, 26	22, 22, 20	15, 10, 13
	Alto	21, 20,13	17, 17, 14	16, 15, 19	11, 16, 14	14, 7, 16	14, 9, 12

Fuente: *Kutner et al. (1996)*

1.- Analizar los resultados con la opción de ANOVA multifactorial (datos disponibles en el fichero *Test_Esfuerzo.sgd* de *Statgraphics Centurion*)

2.- De las características analizadas, ¿cuáles proporcionan un mayor tiempo en la prueba?

Bibliografía

R. Romero y L. Zúnica (2005). *Métodos Estadísticos en Ingeniería.* SPUPV, Valencia.

R. O. Kuehl (2000). *Diseño de Experimentos*. Thomson Learning, México D.F.

D. Montgomery (1991). *Diseño y Análisis de Experimentos.* Grupo Ed. Iberoamericana, México D.F.

G. Box, W. Hunter, J. Hunter (1988). *Estadística para investigadores*. Reverté, Barcelona.