UD5: INFERENCIA

Parte 1: Distribuciones en el muestreo

Parte 2: Inferencia sobre una población

Comparación de 2 poblaciones

Parte 3: ANOVA (Análisis de la Varianza)

Parte 4: Regresión

UD 5 parte 3

ANOVA

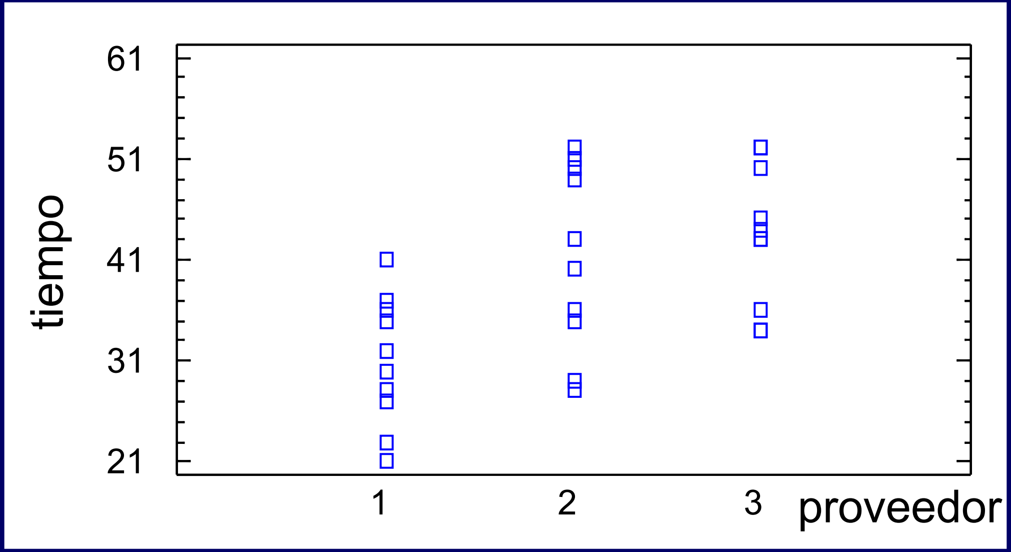
ANÁLISIS DE LA VARIANZA

Una empresa de ordenadores tiene que comprar las baterías para un nuevo modelo de portátil. Hay 3 posibles proveedores: S1, S2 y S3.

Para estudiar cuál es el más conveniente, la empresa compra 10 baterías de cada uno y determina el siguiente parámetro:

1 factor (proveedor) con 3 variantes:

RESULTADOS OBTENIDOS Tiempo de funcionamiento operativo de la batería, en meses (hasta que la batería dura 2 horas)	S1	S2	S3
	23	35	50
	28	36	43
	21	29	36
	27	40	34
	35	43	45
	41	49	52
	37	51	52
	30	28	43
	32	50	44
	36	52	34



¿Qué proveedor se debería elegir?

H_0: $m_1 = m_2 = m_3$

Media: 31 41.3 43.3

Precio: 70 80 85

¿Hay suficiente evidencia para afirmar que las baterías del proveedor 3 duran más, y por tanto serían más convenientes a pesar de su mayor precio?

¿Sería posible analizar estos datos comparando S1 frente a S2 ; S1 frente a S3 ; S2 frente a S3 ?

Con 5 proveedores, ¿cuántas parejas habría que comparar?

COMPARAR TODAS LAS POSIBLES PAREJAS NO ES RECOMENDABLE

ANalysis Of VAriance (ANOVA) Desarrollado en ~1930 por R. A. Fisher

HERRAMIENTA POTENTE PARA ESTUDIAR EL EFECTO DE UNO O MÁS FACTORES EN LA MEDIA DE UNA VARIABLE

IDEA INTUITIVA DEL ANOVA

Residuo = $X_{observado} - X_{predicho}$ → **$X_{observado} = X_{predicho}$ + residuo**

$$\mathbf{x}_{obs} - \bar{\mathbf{x}} = \mathbf{x}_{pred} - \bar{\mathbf{x}} + \text{resid}$$

Media de todos los datos

$$\sum^{n}(\mathbf{x}_{obs} - \bar{\mathbf{x}})^2 = \sum^{n}(\mathbf{x}_{pred} - \bar{\mathbf{x}})^2 + \sum(\text{resid})^2$$

Suma de cuadrados (SC): $SC_{TOTAL} = SC_{FACTOR} + SC_{RESIDUAL}$

$$\frac{\textit{suma de cuadrados}}{\textit{grados de libertad}} = \textit{Cuadrado Medio}$$

SC / g.l. = CM

$g.l._{TOTAL} = g.l._{FACTOR} + g.l._{RESIDUAL}$

CM = SC / g.l.

$$CM_{TOTAL} = \frac{SC_{TOTAL}}{N-1} = \frac{\sum (x_i - \bar{x})^2}{N-1} = s^2$$

Grados de libertad totales = nº total de datos - 1

$$CM_{RESIDUAL} = \frac{SC_{RESID}}{g.l._{residual}}$$

$$CM_{FACTOR} = \frac{SC_{FACTOR}}{g.l._{FACTOR}}$$

Grados libertad$_{FACTOR}$ = Nº de variantes - 1

Comparación de la varianza en 2 poblaciones Normales:

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \sim F_{N_1-1, N_2-1}$$

Asumiendo $\sigma^2_1 = \sigma^2_2$:

$$s_1^2 / s_2^2 \approx F_{n_1-1, n_2-1}$$

El C. Medio es más o menos una varianza:

$$\frac{CM_1}{CM_2} \cong \frac{s_1^2}{s_2^2} \approx F$$

(una medida de dispersión)

TEST F

¿Hay diferencias entre la duración media de las baterías de los 3 proveedores?

HIPÓTESIS NULA: $\mathbf{H_0 : m_1 = m_2 = m_3}$ (no hay diferencias)

SI LA HIPÓTESIS NULA ES CIERTA:

1) LAS MEDIAS MUESTRALES $\bar{x}_1$, $\bar{x}_2$ y $\bar{x}_3$ SERÁN "SIMILARES" (Y TAMBIÉN "SIMILARES" A LA MEDIA GENERAL $\bar{\bar{x}}$)

2) LA SUMA DE CUADRADOS DEL FACTOR SERÁ "PEQUEÑA", PERO… ¿QUÉ DEBE CONSIDERARSE COMO "PEQUEÑA" ?

3) $CM_{residual}$ ES UNA ESTIMACIÓN DE σ^2 ($\sigma^2_1 = \sigma^2_2 = \sigma^2_3$)

4) CM_{factor} ES TAMBIÉN UN ESTIMADOR DE σ^2 , INDEPENDIENTE DE CM_{RESID}, Y POR TANTO:

$$\text{F-ratio} = (CM_{factor} / CM_{residual}) = S_f^2 / S_r^2 \sim F_{2,27}$$

T: tiempo funcion. batería

$T_1 \sim N(m_1, \sigma^2)$	$T_2 \sim N(m_2, \sigma^2)$	$T_3 \sim N(m_3, \sigma^2)$

$$S_r^2 = \frac{SC_{RES}}{g.l._{res}} = CM_{RES} \quad \text{(estimación de } \sigma^2\text{)}$$

$$\frac{CM_{FACTOR}}{CM_{RESID}} = \text{F-ratio}$$

- SI **NO** HAY DIFFERENCIAS ENTRE PROVEEDORES ($m_1 = m_2 = m_3$)

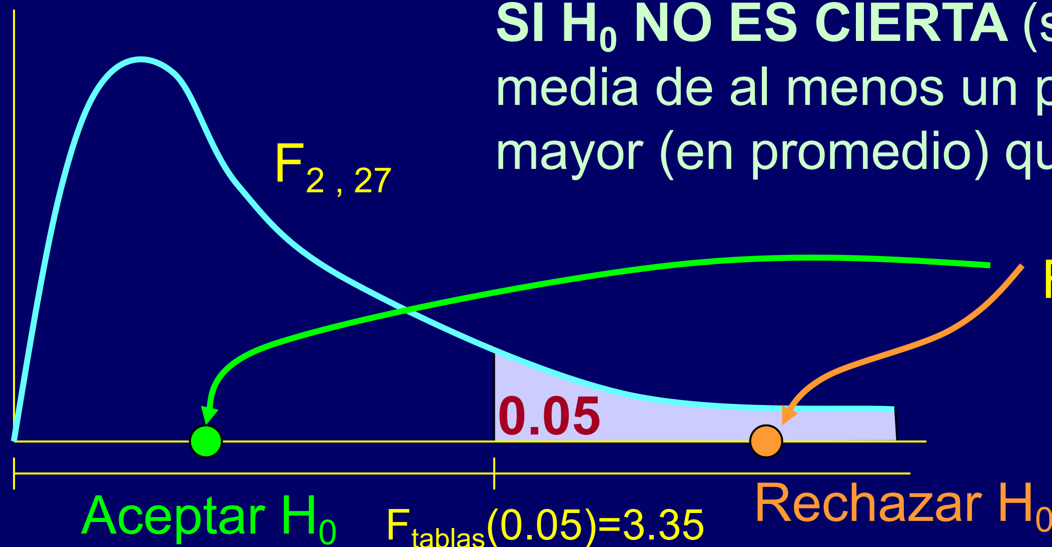
$$\frac{SC_{factor}}{g.l._{factor}} = CM_{FACTOR} = S_f^2 \quad \text{(estimación de } \sigma^2\text{)} \longrightarrow \frac{CM_{FACTOR}}{CM_{RESID}} \approx 1$$

- SI HAY DIFFERENCIAS ENTRE PROVEEDORES:

$$\frac{SC_{factor}}{g.l._{factor}} = CM_{FACT} = S_f^2 \text{ tenderá a ser mayor que } \sigma^2 \longrightarrow$$

F-ratio tiende a ser > 1

SI H_0 NO ES CIERTA (si hay diferencias en la media de al menos un proveedor): F-ratio será mayor (en promedio) que $F_{2,27}$



$$\text{F-ratio} = \frac{CM_{factor}}{CM_{residual}}$$

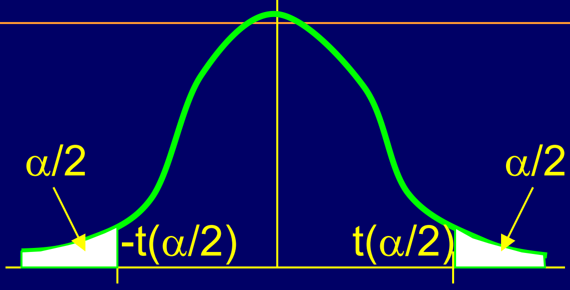
INTERVALOS LSD PARA COMPARACIÓN DE MEDIAS

SI EL TEST F RESULTA ESTADÍSTICAMENTE SIGNIFICATIVO:

Rechazar H_0 ($m_1 = m_2 = m_3$) → Aceptar H_1 (al menos una m_i es distinta del resto)

¿CUÁL ES DISTINTA ? → OBTENER INTERVALOS **LSD** (“LEAST SIGNIFICANCE DIFFERENCE”) PARA LAS MEDIAS MUESTRALES DE CADA PROVEEDOR

$$\bar{x}_i \pm \frac{\sqrt{2}}{2} \cdot t^{\alpha/2}_{g.l.resid} \sqrt{\frac{CM_{resid}}{K_i}}$$



K: Nº de repeticiones del proveedor i

$t^{\alpha/2}_{gl.res}$: valor crítico de la tabla-t

LA DIFERENCIA ENTRE 2 MEDIAS SERÁ SIGNIFICATIVA SI SUS INTERVALOS LSD **NO SE SOLAPAN**

... DE LO CONTRARIO, HAY SUFICIENTE EVIDENCIA PARA DECIR QUE LA DIFERENCIA ES ESTADÍSTICAMENTE SIGNIFICATIVA

LOS INTERVALOS LSD **NO** SON INTERVALOS DE CONFIANZA PARA LA MEDIA

ANOVA con 1 Factor (3 variantes)

(ejemplo de las baterías) T: tiempo de funcionamiento (meses)

3 posibles proveedores: S_A, S_B y S_C.

Tiempo – baterías de S_A:	14	11	15	12	13	$\bar{x}_A = 13$
Tiempo – baterías de S_B:	16	13	17	15	14	$\bar{x}_B = 15$
Tiempo – baterías de S_C:	18	16	19	20	17	$\bar{x}_C = 18$

¿Son las diferencias estadísticamente significativas (α=0.05)?

H_0: $m_A = m_B = m_C$

Media general: $\bar{\bar{x}} = 15.33$

$$SC_{TOTAL} = \sum_{i=1}^{i=15} (x_i - 15.33)^2 = (14 - 15.33)^2 + \ldots + (17 - 15.33)^2 = 93.33$$

$SC_{RESIDUAL} = SC_{TOTAL} - SC_{FACTOR}$

ANOVA con 1 Factor (3 variantes)

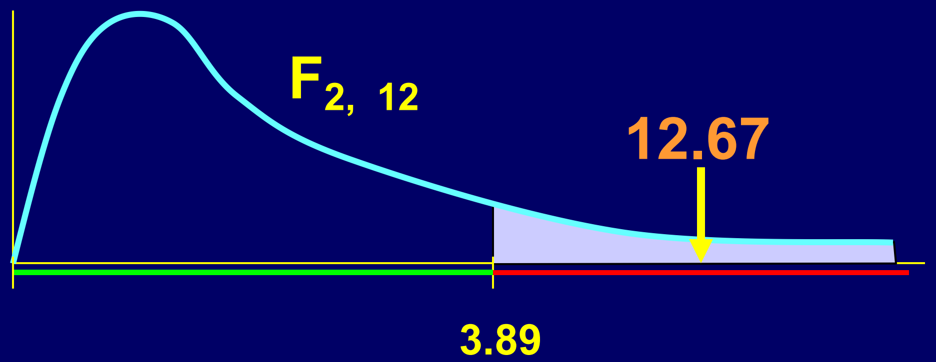
factor	X_{obs}	X_{pred}	$(x_{pred} - \bar{\bar{x}})^2$
A	14 11 15 12 13	13	$5 \cdot (13 - 15.33)^2$
B	16 13 17 15 14	15	$5 \cdot (15 - 15.33)^2$
C	18 16 19 20 17	18	$5 \cdot (18 - 15.33)^2$

$\bar{\bar{x}} = 15.33$

suma:

$63.33 = SC_{FACTOR}$

Fuente variac.	S.C.	g.l.	C.M.	F_{ratio}
FACTOR	63.33	2	31.67	12.67
RESIDUAL	30	12	2.5	
TOTAL	93.33	14		



* Valor crítico de tablas:

$F_{2,12}(0.05) = 3.89 < 12.67$

Rechazar H_0

ANOVA con 1 Factor (3 variantes)

Rechazar H_0

H_0: $m_A = m_B = m_C$

H_1: $m_A \neq m_B \neq m_C$

H_1: $(m_A = m_B) \neq m_C$

H_1: $(m_A = m_C) \neq m_B$

H_1: $(m_B = m_C) \neq m_A$

Al menos una media es significativamente distinta del resto

Intervalos LSD (Least Significant Differences):

$$\bar{x}_i \pm \frac{\sqrt{2}}{2} \cdot t_{g.l.resid}^{\alpha/2} \sqrt{\frac{CM_{resid}}{K}}$$

K: nº de datos usados en calcular $\overline{x_i}$

g.l.res: Grados de libertad residuales

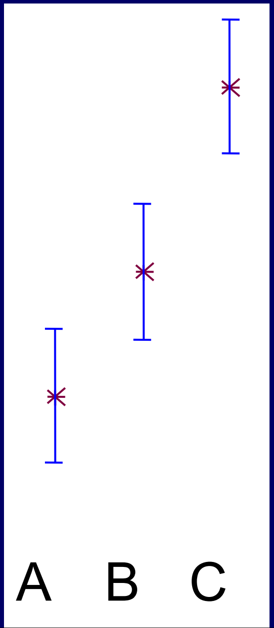
$$t_{12}^{0.025} = 2.179$$

LSD_A: $13 \pm (\sqrt{2}/2) \cdot 2.179 \cdot \sqrt{2.5/5}$

13 ± 1.09 → **[11.91, 14.09]**

LSD_B: 15 ± 1.09 **[13.91, 16.09]**

LSD_C: 18 ± 1.09 **[16.91, 19.09]**



Solución:

$(m_A = m_B) < m_C$

ANÁLISIS DE RESIDUOS

Es una parte importante de ANOVA y regresión.

Cualquier análisis estadístico de datos reales siempre debe completarse con el análisis de residuos.

RESIDUO de la observación ij = $X_{ij} - \overline{X}_i$

Los residuos reflejan el efecto en la variable respuesta de todos los factores no controlados en el experimento.

En ANOVA, estudiar los residuos permite comprobar si:

- Los datos se distribuyen normalmente (no hay datos anómalos)

- La varianza poblacional de todos los proveedores es la misma

ESTUDIO DEL EFECTO SOBRE DISPERSIÓN

¿La varianza de los datos de todos los proveedores es la misma?

$T_1 = N(m_1 , \sigma_1^2)$	$T_2 = N(m_2 , \sigma_2^2)$	$T_3 = N(m_3 , \sigma_3^2)$

$$H_0: \quad \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

PREGUNTA:

$$s^2_{\text{proveedor}_i} \propto \sum_j \frac{e_{ij}^2}{N_i} = \frac{SCresid_i}{N_i}$$

¿Hay relación entre la <u>media de los residuos al cuadrado</u> de un proveedor y la varianza de ese proveedor?

ANOVA es una técnica potente para estudiar diferencias en la media de una variable entre distintas poblaciones

MÉTODO PARA ESTUDIAR EFECTOS SOBRE DISPERSIÓN:

1.- **ANOVA para estudiar efectos sobre la media.**

2.- **Calcular los residuos de todas las observaciones.**

3.- **Guardar los residuos al cuadrado**

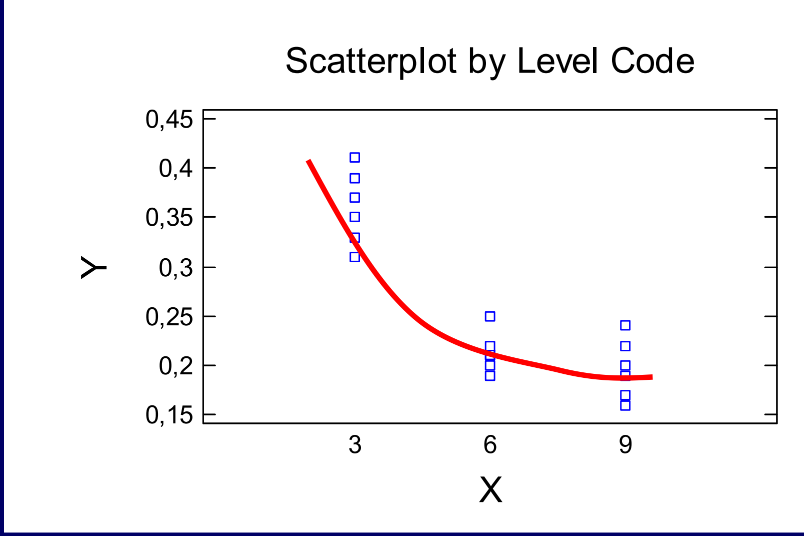
4.- **Realizar un ANOVA considerando como variable respuesta los residuos al cuadrado**

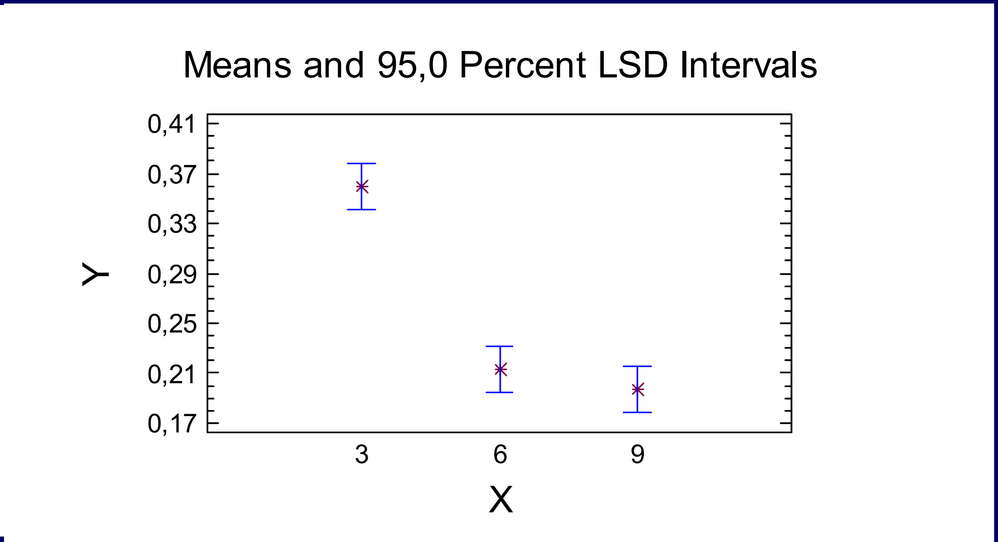
ANOVA CON UN FACTOR CUANTITATIVO

Se ha realizado un experimento para obtener la dosis óptima de plaguicida (ensayada a 3 niveles) para controlar una plaga. La variable respuesta es una medida de la población del insecto. RESULTADOS:

Dosis = 3	Dosis = 6	Dosis = 9
0.31	0.20	0.20
0.35	0.22	0.19
0.39	0.21	0.22
0.37	0.25	0.16
0.33	0.21	0.17
0.41	0.19	0.24

¿Qué dosis recomendarías para minimizar Y?





¿Qué dosis recomendarías para minimizar Y?

Según los intervalos LSD, dosis = 6 ó dosis = 9 (las diferencias entre ambas dosis no son estadísticamente significativas)

PERO si el factor es cuantitativo, interesa estudiar si hay "relación" entre X e Y (no interesa comparar entre los niveles ensayados)

ANOVA => H_0: $m_{dosis3} = m_{dosis6} = m_{dosis9}$

REGRESIÓN => $Y = a + b \cdot X + c \cdot X^2$ H_0: $b = 0$ H_0: $c = 0$

¡ Tiene más sentido analizar los datos con regresión !

ANOVA

con >1 factor

ANOVA CON 3 FACTORES

Se ha realizado un experimento para estudiar las condiciones que maximizan un parámetro de calidad (Y) de un producto químico. Se han considerado tres factores experimentales:

- Catalizador (2 tipos ensayados: A y B)
- Proveedor de la materia prima (s1 y s2)
- Temperatura de la reacción (2 niveles ensayados: 70ºC y 80ºC)

catalizador	proveedor	temperatura	Y
A	s1	70	15
B	s1	70	20
A	s2	70	16
B	s2	70	14
A	s1	80	17
B	s1	80	23
A	s2	80	10
B	s2	80	13

¿Qué condiciones operativas deberían usarse para maximizar Y?

Un ingeniero opina que, como el mayor valor es 23, las condiciones óptimas serían: Catalizador B, prov. s1, Tª = 80

¿Es ésta la "mejor" condición operativa?

NO: puede que el efecto de todos los factores no sea estadísticamente significativo

cataliz.	X_{obs}	$(x_{obs} - \bar{\bar{x}})^2$	X_{pred}	$(x_{pred} - \bar{\bar{x}})^2$
A	15	$(15-16)^2$	14.5	$(14.5-16)^2$
A	16	$(16-16)^2$	14.5	$(14.5-16)^2$
A	17	$(17-16)^2$	14.5	$(14.5-16)^2$
A	10	$(10-16)^2$	14.5	$(14.5-16)^2$
B	20	$(20-16)^2$	17.5	$(17.5-16)^2$
B	14	$(14-16)^2$	17.5	$(17.5-16)^2$
B	23	$(23-16)^2$	17.5	$(17.5-16)^2$
B	13	$(13-16)^2$	17.5	$(17.5-16)^2$
suma:		$116 = SC_{TOTAL}$		$18 = SC_{CATALIZADOR}$

$\bar{\bar{x}} = 16$

$\bar{x}_A = 14.5$

$\bar{x}_B = 17.5$

X_{obs}	prov.	X_{pred}	$(x_{pred} - \bar{\bar{x}})^2$	Temp.	X_{pred}	$(x_{pred} - \bar{\bar{x}})^2$
15	s1	18.75	$(18.75 - 16)^2$	70	16.25	$(16.25 - 16)^2$
20	s1	18.75	$(18.75 - 16)^2$	70	16.25	$(16.25 - 16)^2$
17	s1	18.75	$(18.75 - 16)^2$	80	15.75	$(15.75 - 16)^2$
23	s1	18.75	$(18.75 - 16)^2$	80	15.75	$(15.75 - 16)^2$
16	s2	13.25	$(13.25 - 16)^2$	70	16.25	$(16.25 - 16)^2$
14	s2	13.25	$(13.25 - 16)^2$	70	16.25	$(16.25 - 16)^2$
10	s2	13.25	$(13.25 - 16)^2$	80	15.75	$(15.75 - 16)^2$
13	s2	13.25	$(13.25 - 16)^2$	80	15.75	$(15.75 - 16)^2$
	$SC_{PROV} = 60.5$			$SC_{TEMP} = 0.5$		

FACTOR	S.C.	g.l.	C.M.	F_{ratio}	
catalizador	18	1	18	1.95	<<7.71
proveedor	60.5	1	60.5	6.54	<7.71
temperatura	0.5	1	0.5	0.05	<<7.71
RESIDUAL	37	4	9.25		
TOTAL	116	7			

$F_{1;4}^{0.05} = 7.71$

Tabla del ANOVA después de eliminar factores no significativos:

FACTOR	S.C.	g.l.	C.M.	F_{ratio}	
proveedor	60.5	1	60.5	6.54	>5.99
RESIDUAL	55.5	**6**	9.25		
TOTAL	116	7			

$F_{1;6}^{0.05} = 5.99$

El factor proveedor es estadísticamente significativo (α=0.05)

Condiciones óptimas para maximizar Y: usar s1 como proveedor de la materia prima; usar Tª 70 ó 80; usar catalizador A ó B.

Valor medio de Y usando proveedor s1: $\overline{x}_{S1} = 18.75$

A continuación, hay que estudiar si:

- Los residuos siguen una distrib. Normal (ver si hay outliers)

- La interacción entre proveedor y Tª o catalizador es estadísticamente significativa

¿ QUÉ ES UNA INTERACCIÓN ?

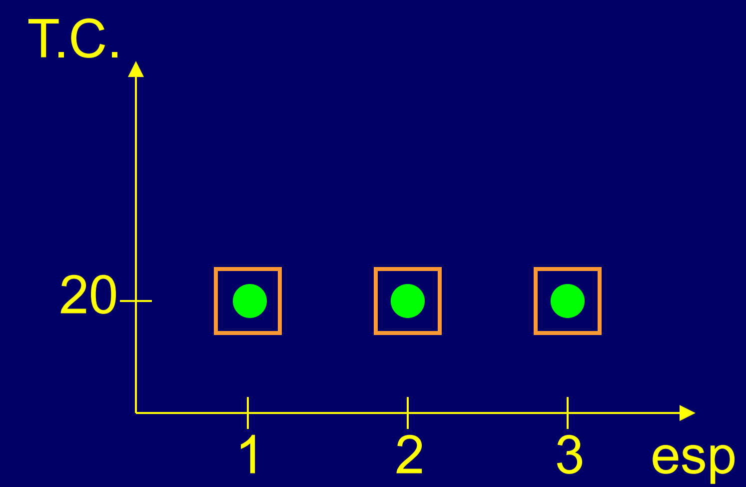
EJEMPLO INTUITIVO:

EFECTO DEL ESPESOR DE LA CAPA DE PINTURA Y EL TIPO DE METAL EN LA TASA DE CORROSIÓN (T.C.) DE 12 PROTOTIPOS

CASO A:

ESPESOR CAPA PINTURA

		1		2		3	
TIPO METAL	1	20	20	20	20	20	20
	2	20	20	20	20	20	20



SUMA DE CUADRADOS TOTAL:

$$\sum_{ijk}\left(x_{ijk}-\overline{x}\right)^2=0$$

i: espesor

j: tipo metal

k: repetición

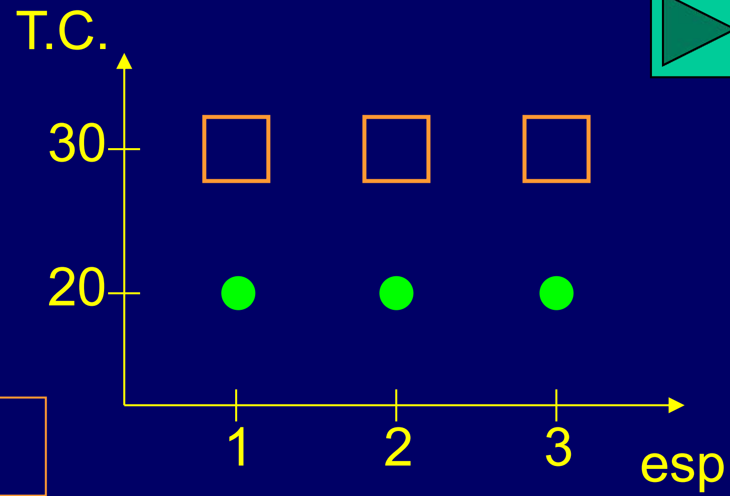
¡ SI NO HAY VARIANZA, NADA TIENE INFLUENCIA !

CASO B:

ESPESOR CAPA PINTURA

		1		2		3	
TIPO METAL	1	20	20	20	20	20	20
	2	30	30	30	30	30	30

$$SC_{TOTAL} = SC_{METAL}$$

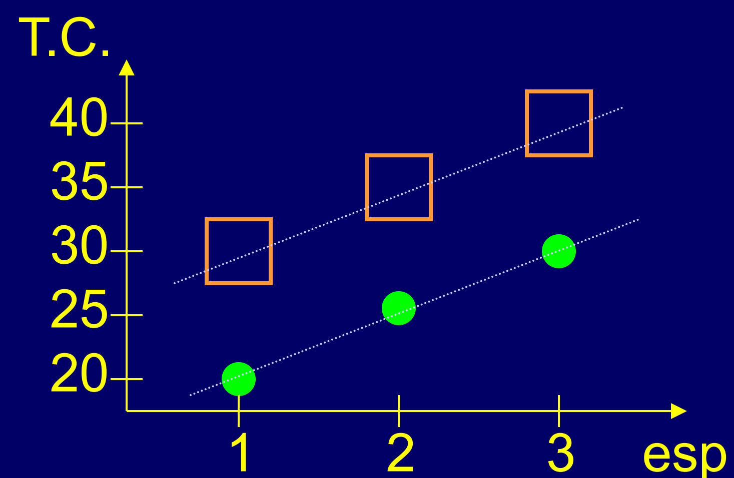


El factor "tipo de metal" tiene efecto en la media.

CASO C:

ESPESOR CAPA PINTURA

		1		2		3	
TIPO METAL	1	20	20	25	25	30	30
	2	30	30	35	35	40	40

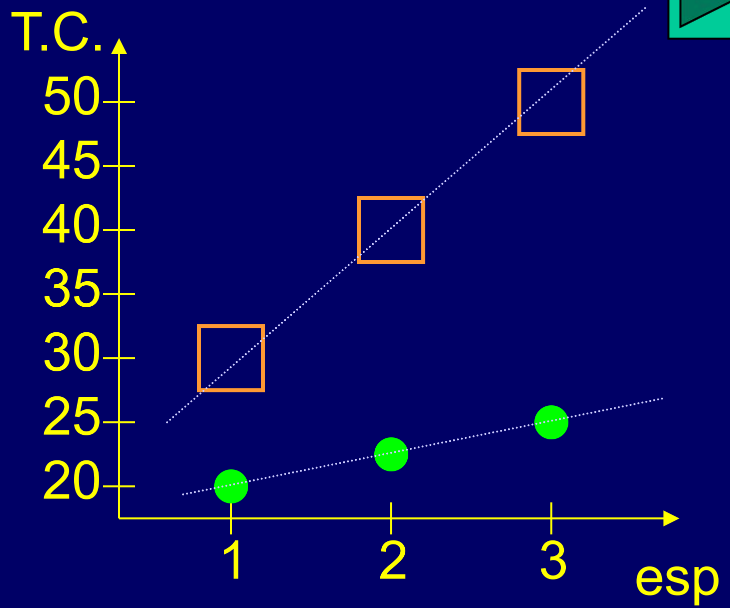


HAY VARIABILIDAD ($SC_{TOTAL} \neq 0$) DEBIDA AL EFECTO DEL TIPO DE METAL Y DEL ESPESOR

$$SC_{TOTAL} = SC_{METAL} + SC_{ESPESOR}$$

CASO D:

TIPO METAL	ESPESOR CAPA PINTURA 1		2		3	
1	20	20	25	25	30	30
2	30	30	40	40	50	50



HAY VARIABILIDAD DEBIDA A AMBOS FACTORES

Y A SU **INTERACCIÓN**

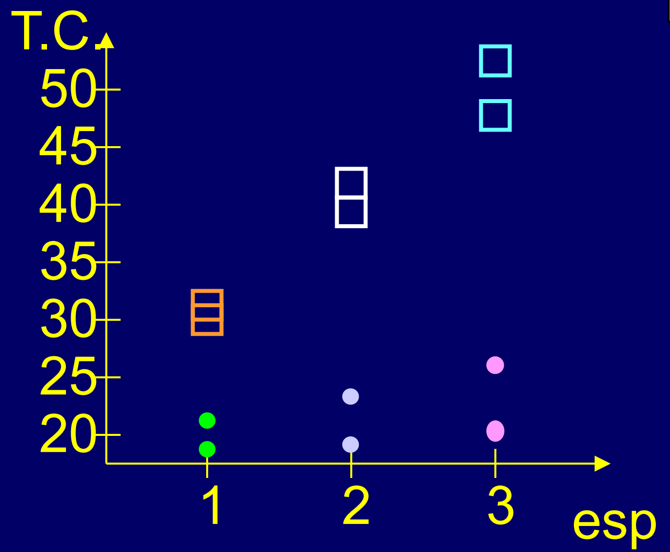
$$SC_{TOTAL} = SC_{METAL} + SC_{ESPESOR} + SC_{INTERACCIÓN}$$

CASO E:

ESPESOR CAPA PINTURA

		1		2		3	
TIPO METAL	1	19	21	27	24	28	32
	2	30	31	42	39	47	51

(EL ÚNICO CASO REALISTA)

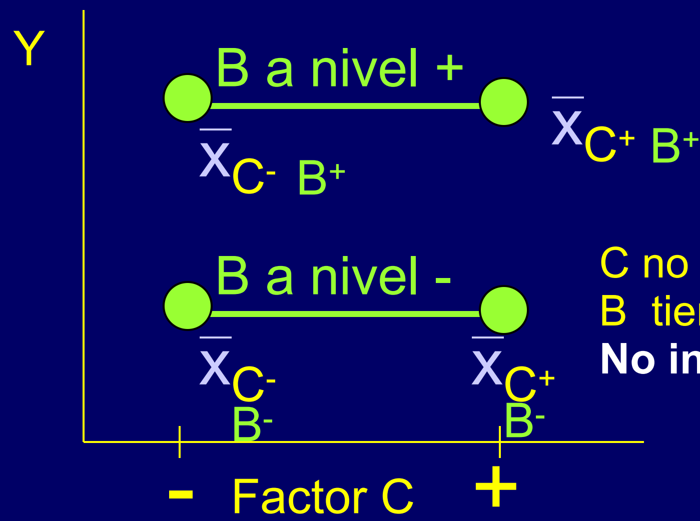


EL EFECTO DE LOS FACTORES ESTÁ PARCIALMENTE ENMASCARADO POR LA VARIABILIDAD RESIDUAL DEBIDA A LOS FACTORES NO CONTROLADOS,

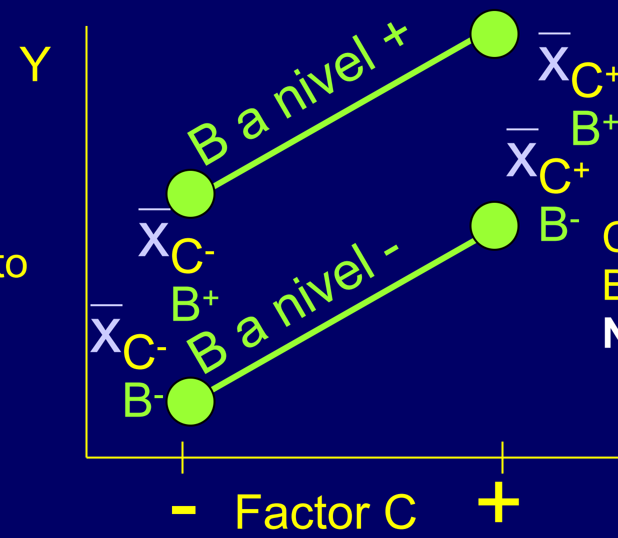
…YA QUE LA CORROSIÓN DE LOS PROTOTIPOS ENSAYADOS CON EL MISMO METAL Y ESPESOR NO ES EXACTAMENTE LA MISMA

$$SC_{TOTAL} = SC_{METAL} + SC_{ESPESOR} + SC_{INTERACCIÓN} + SC_{RESIDUAL}$$

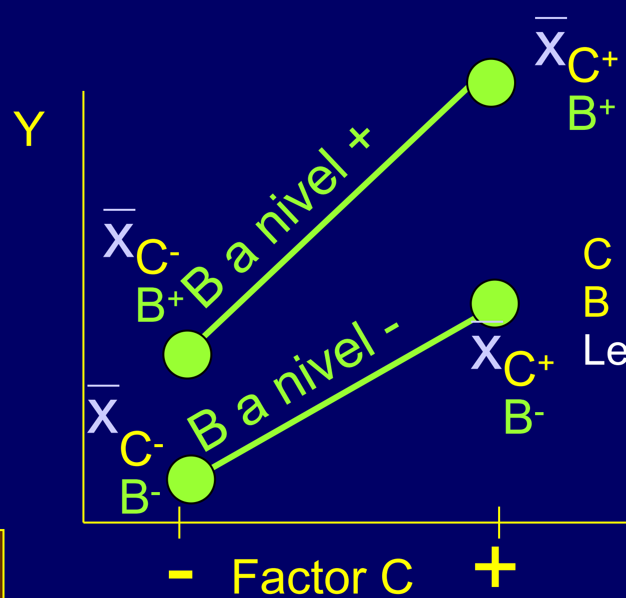
DEFINICIÓN: Hay interacción entre 2 factores si el efecto de uno de ellos es distinto en función del nivel del otro factor.



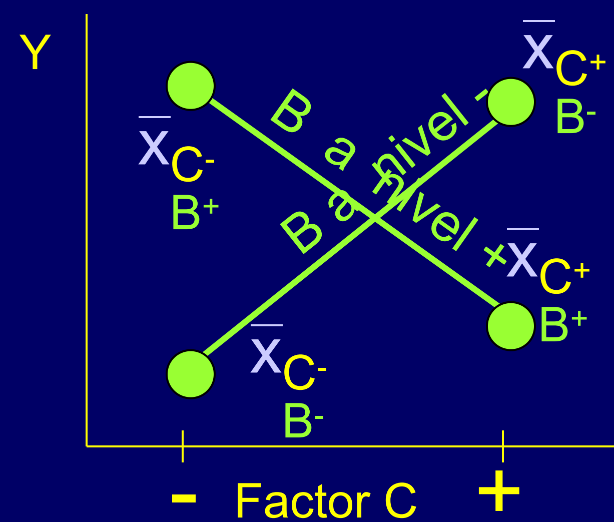
C no tiene efecto
B tiene efecto
No interacción



C tiene efecto
B tiene efecto
No interacción



C tiene efecto
B tiene efecto
Leve interacción



Fuerte interacción
(discutir los efectos simples no tiene sentido)

EJERCICIO: ANOVA de 2 factores con interacción

Una empresa farmacéutica pretende mejorar el rendimiento de un proceso de fermentación (variable Y). Para ello, se lleva a cabo un experimento con 2 factores a 2 niveles:

- **Concentración de nitrógeno (2 niveles ensayados: 40 y 50)**
- **Tipo de materia prima (2 tipos ensayados: A o B)**

RESULTADOS

concentr.	mat. prima	Y	
40	**A**	46	**42**
40	**A**	38	
40	**B**	42	**38**
40	**B**	34	
50	**A**	32	**30**
50	**A**	28	
50	**B**	51	**50**
50	**B**	49	

¿Qué concentración y tipo de mat. prima se debería emplear para maximizar el rendimiento?

Mayor valor =>
Solución: conc = 50
mat. prima = B

Como el mayor valor medio es 50, un ingeniero opina que las condiciones óptimas serían: conc = 50 y mat. prima = B

¿Es ésta la "mejor" condición operativa?

¡ SÓLO si el efecto de ambos factores o de su interacción es estadísticamente significativo !

concentr	X_{obs}	$(x_{obs} - \bar{\bar{x}})^2$	X_{pred}	$(x_{pred} - \bar{\bar{x}})^2$
40	46	$(46 - 40)^2$	40	$(40 - 40)^2$
40	38	$(38 - 40)^2$	40	$(40 - 40)^2$
40	42	$(42 - 40)^2$	40	$(40 - 40)^2$
40	34	$(34 - 40)^2$	40	$(40 - 40)^2$
50	32	$(32 - 40)^2$	40	$(40 - 40)^2$
50	28	$(28 - 40)^2$	40	$(40 - 40)^2$
50	51	$(51 - 40)^2$	40	$(40 - 40)^2$
50	49	$(49 - 40)^2$	40	$(40 - 40)^2$

$\bar{\bar{x}} = 40$

$\bar{x}_{40} = 40$

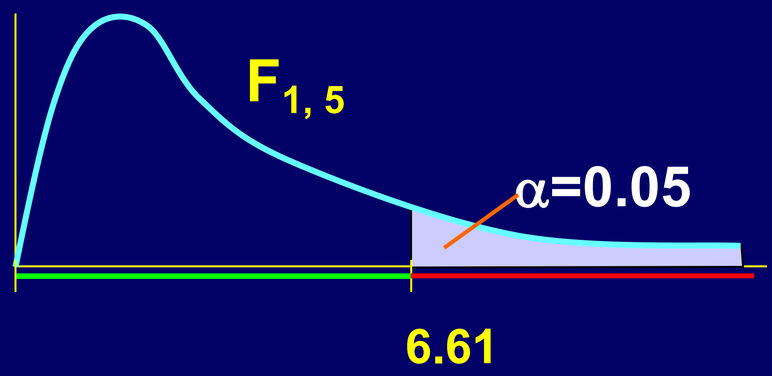
$\bar{x}_{50} = 40$

suma: $490 = SC_{TOTAL}$ $0 = SC_{CONCENTR}$

m_pr	X_{obs}	X_{pred}	$(x_{pred} - \bar{\bar{x}})^2$
A	46	36	$(36 - 40)^2$
A	38	36	$(36 - 40)^2$
A	32	36	$(36 - 40)^2$
A	28	36	$(36 - 40)^2$
B	42	44	$(44 - 40)^2$
B	34	44	$(44 - 40)^2$
B	51	44	$(44 - 40)^2$
B	49	44	$(44 - 40)^2$

$SC_{M\_PR} = 128$

$\bar{x}_A = 36$ $\bar{x}_B = 44$



FACTOR	S.C.	g.l.	C.M.	F_{ratio}	
concentración	0	1	0	0	
mat. prima	128	1	128	1.77	<<6.61
RESIDUAL	362	**5**	72.4		
TOTAL	490	7			

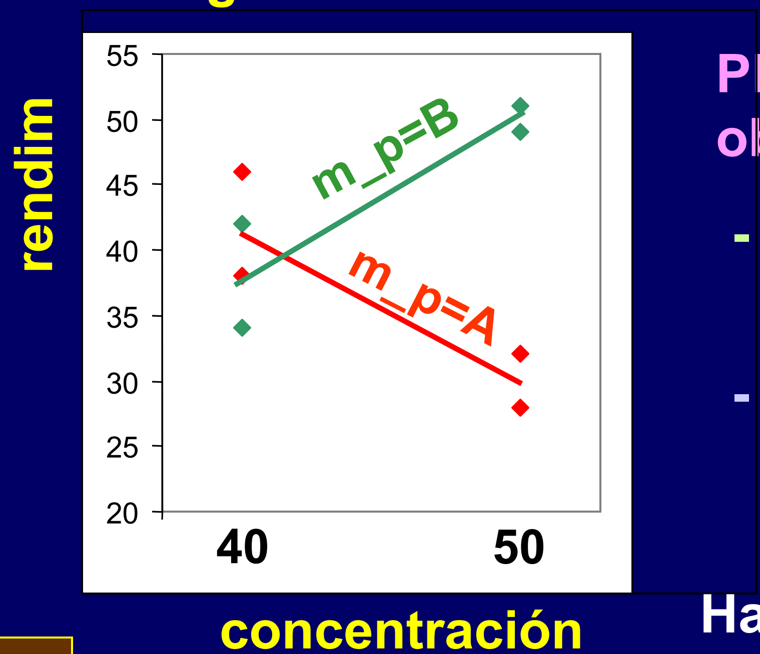
Siguiente paso: descartar “concentración” y repetir el ANOVA, pero el factor mat. prima no resulta estadísticamente significativo.

PREGUNTA: ¿cuáles son las condiciones óptimas en este caso?

El análisis no ha terminado hasta que se compruebe:

- Si los residuos siguen una distribución Normal (sin outliers)
- Si la interacción entre ambos factores es estadístic. significativa

Estudio gráfico de la interacción:



PREGUNTA: ¿piensas que la variabilidad observada se ha obtenido por azar?

- SÍ (interacción no significativa):
  condiciones óptimas = cualquiera
- NO (interacción significativa):
  condiciones óptimas = 50 ; B

Hay que calcular la suma de cuadrados de la interacción e incluirla en la tabla del ANOVA

Conc	m_pr	X_{obs}	X_{pred}	$(x_{pred} - \bar{\bar{x}})^2$
40	A	46	42	$(42 - 40)^2$
40	A	38	42	$(42 - 40)^2$
40	B	42	38	$(38 - 40)^2$
40	B	34	38	$(38 - 40)^2$
50	A	32	30	$(30 - 40)^2$
50	A	28	30	$(30 - 40)^2$
50	B	51	50	$(50 - 40)^2$
50	B	49	50	$(50 - 40)^2$

$\Sigma = 416 =$

$= SC_C + SC_{MP} + SC_{C \cdot MP}$

$SC_{CONC \cdot M\_PR} = 416 - 128 - 0 = 288$

FACTOR	S.C.	g.l.	C.M.	F_{ratio}	
concentr.	0	1	0	0	<<7.71
mat. prima	128	1	128	6.92	<7.71
conc x m_pr	288	1·1	288	15.6	>7.71
RESIDUAL	74	4	18.5		
TOTAL	490	7			

$F_{1;4}^{0.05} = 7.71$

La interacción entre concentración y materia prima es estadísticamente significativa (α=0.05) aunque el efecto simple de ambos factores no lo es.

Para maximizar el rendimiento, la empresa debería usar:

Concentración de nitrógeno = 50; materia prima de tipo B

Rendimiento medio esperado en esas condiciones = (51+49)/2 = 50

REGLA: si una interacción doble es significativa, ambos factores deben siempre incluirse en la tabla del ANOVA.

REGLA: grados de libertad de CONCxMAT_PR = $(g.l._{CONC})\cdot(g.l._{M\_PR})$

Con más de 2 factores, las interacciones triples también deberían estudiarse, aunque generalmente no son significativas.

DISEÑO DE EXPERIMENTOS

EXPERIMENTO: CONJUNTO DE **PRUEBAS** REALIZADAS PARA OBTENER INFORMACIÓN DE UN PRODUCTO O PROCESO.

Enfoque común usado para optimizar procesos industriales:

A	B	C	Respuesta
-	-	-	..
-	-	-	..
-	-	-	..
-	-	-	..
+	-	-	
+	-	-	
+	-	-	
+	-	-	

A^+ mejor que A^-

A	B	C	Respuesta
+	-	-	..
+	-	-	..
+	-	-	..
+	-	-	..
+	+	-	
+	+	-	
+	+	-	
+	+	-	

B^+ mejor que B^-

A	B	C	Respuesta
+	+	-	..
+	+	-	..
+	+	-	..
+	+	-	..
+	+	+	
+	+	+	
+	+	+	
+	+	+	

C^+ mejor que C^-

Nº total de pruebas experimentales = 24

Condición Operativa Óptima (COO) : $A^+ B^+ C^+$ pero…

¿se han ensayado todas las posibles combinaciones?

ESTE ENFOQUE (modificar un factor cada vez) NO ES EFICIENTE:

- **Muy caro (requiere muchas pruebas).**
- **No garantiza obtener las condiciones operativas óptimas.**
- **No es posible estudiar el efecto de las interacciones.**

Enfoque ALTERNATIVO, mucho más eficiente:

¡ USAR UN **DISEÑO DE EXPERIMENTOS FACTORIAL** PARA ESTUDIAR EL EFECTO SIMULTÁNEO DE TODOS LOS FACTORES!

A	B	C
-	-	-
+	-	-
-	+	-
+	+	-
-	-	+
+	-	+
-	+	+
+	+	+

- TODAS LAS CONDICIONES OPERATIVAS SE HAN ENSAYADO
- SÓLO 8 PRUEBAS SON NECESARIAS (frente a las 24 del enfoque tradicional)
- TODOS LOS EFECTOS SON ORTOGONALES Y ES POSIBLE ESTUDIAR EL EFECTO DE LAS INTERACCIONES

FACTORES CONTROLADOS: Parámetros del proceso en estudio que se ensayan a 2 o más niveles o variantes, para investigar el efecto sobre la variable respuesta.

DISEÑO DE EXPERIMENTOS

Factores cuantitativos => variantes ; Factores cualitativos => niveles

Diseño experimental con 3 factores: F1 con 2 variantes (A, B), F2 con 2 niveles (2, 3), F3 con 2 variantes (X, Y)

Variable respuesta (medida experimentalmete):

F1	F2	F3	VR
A	2	X	3.5
B	2	X	4.9
A	3	X	3.2
B	3	X	6.2
A	2	Y	5.3
B	2	Y	4.7
A	3	Y	2.9
B	3	Y	4.8

- **Tratamientos**: combinaciones distintas de variantes/niveles: $2 \cdot 2 \cdot 2 = 8$

- **Repeticiones / réplicas**: número de veces que se ensaya cada tratamiento

- **Ensayos**: pruebas experimentales realizadas

Nº ensayos = Nº tratamientos x repeticiones

Diseño factoral no replicado (8 tratamientos, 8 pruebas)

DISEÑO FACTORIAL EQUILIBRADO: todos los tratamientos se ensayan N_R veces

Si N_R=1: diseño factorial no replicado

PROPIEDAD: TODOS los factores e interacciones son ortogonales entre sí

PASOS PARA DISEÑAR UN EXPERIMENTO

0.- DEFINIR EL PROBLEMA Y ESTABLECER EL OBJETIVO

1.- IDENTIFICAR LA VARIABLE RESPUESTA

2.- SELECCIONAR LOS FACTORES Y DECIDIR NIVELES / VARIANTES

4.- DEFINIR EN QUÉ CONSISTIRÁ CADA PRUEBA INDIVIDUAL

5.- DECIDIR EL NÚMERO DE ENSAYOS A REALIZAR

5'.- DECIDIR EL TRATAMIENTO A USAR EN CADA PRUEBA

6.- ORGANIZAR TODO EL TRABAJO; REALIZAR EL EXPERIMENTO

7.- ANALIZAR ESTADÍSTICAMENTE LOS RESULTADOS

IMPORTANTE: LAS PRUEBAS SIEMPRE DEBEN REALIZARSE EN ORDEN ALEATORIO

(para evitar confusión de efectos con factores no controlados)

ANÁLISIS DE RESULTADOS:

1) Identificar con ANOVA qué efectos son significativos

2) Interpretar la naturaleza de los efectos significativos:

- en factores CUALITATIVOS : usar INTERVALOS LSD

- en factores CUANTITATIVOS ensayados a >2 niveles: el mejor método es regresión lineal múltiple

3) Análisis de los residuos:

- DETECCIÓN DE OUTLIERS (datos anómalos)

- EFECTOS SOBRE LA VARIANZA DE LA VARIABLE RESPUESTA

FACTORES A 3 NIVELES:

Los experimentos requieren mayor nº de pruebas PERO proporcionan más información sobre efectos lineales / cuadráticos que permiten estimar mejor las condiciones operativas óptimas.

El análisis de datos con ANOVA asume 3 hipótesis:

1) Las poblaciones correspondientes a todos los tratamientos ensayados tienen la MISMA varianza.

(ANOVA es bastante robusto si las varianzas no son todas iguales)

2) NORMALIDAD:

La variable respuesta es Normal en todos los tratamientos

- **Si la distribución es muy asimétrica => usar transformaciones (ej. trabajar con log(X) o $X^{0.5}$ en lugar de X)**

3) INDEPENDENCIA:

¡ La más importante !

Las observaciones de cada tratamiento corresponden a individuos extraídos aleatoriamente de la población considerada.

Los datos resultantes de tratamientos distintos son independientes entre sí.

Para que se cumpla esta hipótesis es necesario aleatorizar el orden de todas las pruebas experimentales

ORTOGONALIDAD

F1	F2	F3	VR
A	2	X	3.5
B	3	X	4.9
A	2	X	3.2
B	3	X	6.2
A	2	Y	5.3
B	3	Y	4.7
A	2	Y	2.9
B	3	Y	4.8

3 factores, 8 tratamientos, 8 ensayos, pero...

¿es un buen diseño experimental?

El efecto de F1 está totalmente confundido con el efecto del factor F2 => MAL DISEÑO

Si este efecto es estadístic. significativo, NO es posible saber si se debe a F1, F2 o ambos.

F1	F2	F3	VR
A	2	X	3.5
B	3	X	4.9
A	2	X	3.2
B	2	X	6.2
A	3	Y	5.3
B	3	Y	4.7
A	2	Y	2.9
B	3	Y	4.8

¿Es un buen diseño experimental?

- La variante A de F1 se ha ensayado 3 veces a F2=2 y una vez a F2=3
- La variante B de F1 se ha ensayado una vez a F2=2 y 3 veces a F2=3

F1=A ⟷ F2=2 F1=B ⟷ F2=3

El efecto del factor F1 está parcialmente confundido con el efecto del factor F2

PREGUNTA: ¿es éste un buen diseño experimental?

F1	F2	VR
2	A	3.5
2	B	4.9
2	A	3.2
2	B	6.2
2	A	5.3
2	B	4.7
3	A	2.9
3	B	4.8
3	A	5.4
3	B	2.7

El diseño es correcto, pero F1 se ha ensayado 6 veces a nivel 2 y 4 veces a nivel 3.

DISEÑO NO EQUILIBRADO

Si cada tratamiento se ensaya el mismo número de veces => diseño factorial equilibrado

Un diseño equilibrado es mejor, a no ser que $\sigma^2_{F1=2} > \sigma^2_{F1=3}$

F1=2
- 3 veces a F2=A (50%=A)
- 3 veces a F2=B (50%=B)

F1=3
- 2 veces a F2=A (50%=A)
- 2 veces a F2=B (50%=B)

Misma proporción:

"F1 y F2 son ortogonales"

(F1 y F2 no están confundidos)

Dados 2 factores (F_I con I variantes y F_J con J variantes):
el efecto simple de ambos factores será ortogonal si en cada variante de I, las proporciones de las J variantes son iguales.

Consecuentemente: es posible separar el efecto de cada factor sobre la variable en estudio

PREGUNTA: ¿es éste un buen diseño?

Diseño con 3 factores: A (2 niveles), B (3 niveles), C (2 niveles):

A	B	C
1	1	1
1	2	2
1	3	2
2	1	1
2	2	2
2	3	2

¿Son A y B ortogonales?

¿Son A y C ortogonales?

¿Son B y C ortogonales?

PREGUNTA:

¿El efecto simple de F3 es ortogonal a la interacción F1 x F2?

F1	F2	F3	VR
A	2	X	3.5
B	2	X	4.9
A	3	X	3.2
B	3	X	6.2
A	2	Y	5.3
B	2	Y	4.7
A	3	Y	2.9
B	3	Y	4.8

F3=X
- 1 vez a F1=A; F2=2 (25%=A·2)
- 1 vez a F1=B; F2=2 (25%=B·2)
- 1 vez a F2=A; F2=3 (25%=A·3)
- 1 vez a F2=B; F2=3 (25%=B·3)

F3=Y
- 1 vez a F1=A; F2=2 (25%=A·2)
- 1 vez a F1=B; F2=2 (25%=B·2)
- 1 vez a F2=A; F2=3 (25%=A·3)
- 1 vez a F2=B; F2=3 (25%=B·3)

Misma proporción: "F3 es ortogonal a la interacción F1 x F2"

(F3 y la interacción F1 x F2 no están confundidos)

Ejercicio: en un plan factorial 2^4, comprobar que AxB es ortogonal a CxD; y que AxBxC es ortogonal a D.

PREGUNTA:

DISEÑO-1		
F1	F2	F3
1	1	1
2	1	1
1	2	1
2	2	1
1	1	2
2	1	2
1	2	2
2	2	2

En DISEÑO-1, comprobar que:

F1 es orhogonal a F2

F1 es ortogonal a F3

F2 es ortogonal a F3

F1 es ortogonal a F2xF3

F2 es ortogonal a F1xF3

F3 es ortogonal a F1xF2

DISEÑO-2			
F1	F2	F3	F4
1	1	1	1
1	2	2	2
1	3	3	3
2	1	2	3
2	2	3	1
2	3	2	2
3	1	3	2
3	2	1	3
3	3	2	1

En DISEÑO-2, comprobar que son ortogonales: F1-F2, F1-F3, F1-F4, F2-F3, F2-F4 y F3-F4.

¿Son los efectos simples ortogonales a las interac. dobles?

¿Cuántas pruebas tendría el diseño “completo”?

¿Qué ocurre si 2 efectos están confundidos?

(por ejemplo el efecto simple de A con la interacción B·C):

- No es posible separar el efecto de A frente a B·C

- Se podría asumir que el efecto de A es más importante que el de B·C, pero esto puede ser falso en la práctica.

Pregunta: ¿es éste un buen diseño?

F_1	F_2	F_3	VR
A	2	X	3.5
B	3	X	4.9
B	2	Y	3.2
A	3	Y	6.2

Comprueba la ortogonalidad entre F1 y F2; F1 y F3 ; F2 y F3

El efecto de F3 está confundido con el efecto de F1xF2, pero este diseño es el mejor posible con sólo 4 pruebas.

Diseños Factoriales 2^k

Todos los factores se ensayan a 2 niveles/variantes

- Experimentos más sencillos, al tener menor nº de pruebas.
- Los efectos de factores e interac. son fáciles de interpretar.
- No se pueden estudiar los posibles efectos cuadráticos.

Diseño 2^K: es posible calcular 2^K-1 efectos a partir de los datos

Relevantes en la práctica

K EFECTOS SIMPLES (cada uno con 1 grado de lib.)

$\binom{K}{2}$ INTERACCIONES DOBLES (cada una con 1x1 = 1 g.l.)

$\binom{K}{3}$ INTERACCIONES TRIPLES (cada una con 1x1x1 = 1 g.l.)

....

$\binom{K}{K}=1$ INTERACCIÓN DE ORDEN K (con 1 g.l.)

EJEMPLO: **En un plan 2^4 es posible calcular 4 efectos simples de factores, 6 interacciones dobles, 4 int. triples y 1 int. cuádruple = 15 efectos**

Se recomienda que g.l. residual ≥ 10 (al menos g.l. res. ≥ 4)

FRACCIONES FACTORIALES 2^{k-1}

K factores se ensayan a 2 niveles con la MITAD de pruebas (es decir, 2^{K-1} pruebas) que serían necesarias en el diseño completo.

Para construir un diseño 2^{K-1} :

1) Comenzar con un diseño completo 2^K

2) Multiplicar el signo de todos los factores (interacción de mayor orden)

3) Seleccionar aquellas pruebas con el signo " + " (o -)

En un diseño 2^{k-1} :

- No es posible estudiar la interacción de mayor orden, pero este efecto es seguramente el menos importante.

- El efecto simple de un factor está confundido con la interacción de orden (k-1) de los restantes factores.

- La interacción doble de 2 factores está confundida con la interacción de orden (k-2) de los restantes factores.

- Dos efectos estarán completamente confundidos si sus signos son los mismos en todas las pruebas (para interacciones, multiplicar los signos de los factores).

Diseño completo con 4 factores a 2 niveles. ¿Qué tratamientos deberían seleccionarse si queremos realizar sólo 8 pruebas?

	A	B	C	D	ABCD
1	-	-	-	-	+
2					
3					
4	+	+	-	-	+
5					
6	+	-	+	-	+
7	-	+	+	-	+
8					
9					
10	+	-	-	+	+
11	-	+	-	+	+
12					
13	-	-	+	+	+
14					
15					
16	+	+	+	+	+

Éste es el mejor diseño posible con 8 pruebas porque:

a) Cada factor se ha ensayado 4 veces a nivel (-) y 4 veces a nivel (+)

b) El efecto de cada factor es ortogonal a los restantes factores y a las interacciones dobles.

EJERCICIO: comprobar que el efecto simple de A es ortogonal a la interac. B·C

Comprobar que el efecto simple de A no es ortogonal a la interacción B·C·D

Pruebas seleccionadas: ABCD = (+)

Ejemplo: en un diseño 2^6 (6 factores, 64 pruebas), es posible estimar 63 efectos: 6 efectos simples, 15 interacciones dobles, 20 int. triples, 15 int. cuádruples, 6 int de 5º orden, 1 int. 6º orden

¡PERO la mayoría (interacciones de orden >2) serán irrelevantes!

Interesa reducir el número de ensayos "sacrificando" cierta precisión y la posibilidad de estudiar interacciones de orden alto.

UNA BUENA FRACCIÓN FACTORIAL:

A) NUNCA debería confundir efectos simples entre sí

B) Debería intentar evitar (si es posible) la confusión entre efectos simples e interacciones dobles.

(ya que éstas son relevantes a veces, y su confusión produciría una interpretación errónea de los resultados)

C) Si es posible, NO debería confundir interacciones dobles entre sí (para que puedan estudiarse).

ORTHOGONAL ARRAYS L_8 , L_{16} , L_{18} , L_{26}

Son tablas propuestas para simplificar el diseño de fracciones factoriales

Desarrollados en Japón por el profesor G. Taguchi

Orthogonal array L8:

permite estudiar hasta 7 factores con 8 pruebas:

	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

$L_8 (2^{7-4})$

L_{16} (2^{15-11})

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1

ORTOGONAL ARRAY L_{18}

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

L_{18} $(2^1 * 3^7)$

18 pruebas en lugar de 4,374

Los efectos simples ocupan

1+7x2 = 15 g.l., y los 2 g.l. restantes están asociados a la interacción F1xF2

EJERCICIO: comprobar que la interacción F1xF2 es ortogonal a todos los efectos simples.

EJERCICIO: comprobar que F1xF3 NO es ortogonal al efecto simple de F4

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	2	2	2	2	2
3	1	1	1	1	3	3	3	3	3	3	3	3	3
4	1	2	2	2	1	1	1	2	2	2	3	3	3
5	1	2	2	2	2	2	2	3	3	3	1	1	1
6	1	2	2	2	3	3	3	1	1	1	2	2	2
7	1	3	3	3	1	1	1	3	3	3	2	2	2
8	1	3	3	3	2	2	2	1	1	1	3	3	3
9	1	3	3	3	3	3	3	2	2	2	1	1	1
10	2	1	2	3	1	2	3	1	2	3	1	2	3
11	2	1	2	3	2	3	1	2	3	1	2	3	1
12	2	1	2	3	3	1	2	3	1	2	3	1	2
13	2	2	3	1	1	2	3	2	3	1	3	1	2
14	2	2	3	1	2	3	1	3	1	2	1	2	3
15	2	2	3	1	3	1	2	1	2	3	2	3	1
16	2	3	1	2	1	2	3	3	1	2	2	3	1
17	2	3	1	2	2	3	1	1	2	3	3	1	2
18	2	3	1	2	3	1	2	2	3	1	1	2	3
19	3	1	3	2	1	3	2	1	3	2	1	3	2
20	3	1	3	2	2	1	3	2	1	3	2	1	3
21	3	1	3	2	3	2	1	3	2	1	3	2	1
22	3	2	1	3	1	3	2	2	1	3	3	2	1
23	3	2	1	3	2	1	3	3	2	1	1	3	2
24	3	2	1	3	3	2	1	1	3	2	2	1	3
25	3	3	2	1	1	3	2	3	2	1	2	1	3
26	3	3	2	1	2	1	3	1	3	2	3	2	1
27	3	3	2	1	3	2	1	2	1	3	1	3	2

ORTOGONAL ARRAY L_{27}

27 pruebas en lugar de 1.584.323

Diseño completamente saturado: los efectos simples ocupan 13·2=26 grados de libertad, que son los g.l. totales disponibles (27-1)

EJERCICIOS

Datos en PoliformaT:

recursos \ Transparencias \ English group \ UD 5 DOE-data.xls

EJERCICIO-1

Se han ensayado tres variedades de naranjo (A, B, C) y distintas alturas de poda para estudiar su rendimiento. Resultados:

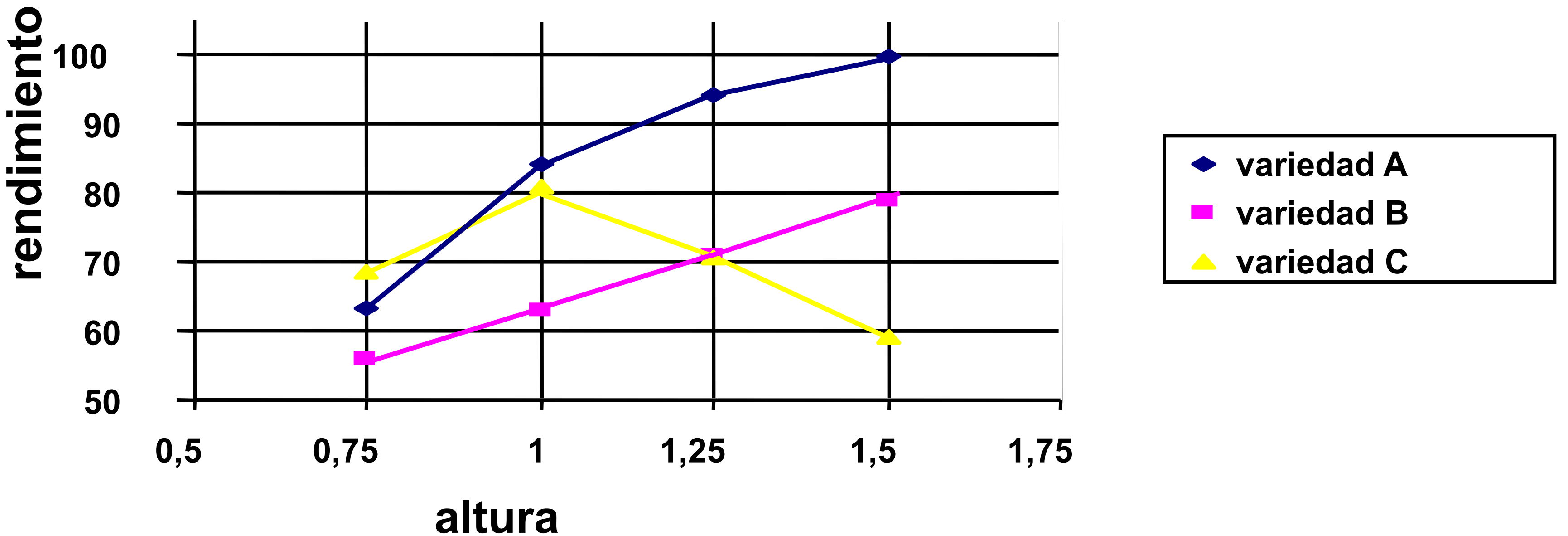
	A	B	C	$\bar{X}_j$
Altura 0.75	68 60 62 $\bar{X}_{11} = 63.33$	52 55 61 $\bar{X}_{21} = 56$	66 72 68 $\bar{X}_{31} = 68.67$	$\bar{X}_1 = 62.67$
Altura 1	91 75 86 $\bar{X}_{12} = 84$	62 67 60 $\bar{X}_{22} = 63$	83 82 78 $\bar{X}_{32} = 81$	$\bar{X}_2 = 76$
Altura 1.25	90 98 94 $\bar{X}_{13} = 94$	64 75 74 $\bar{X}_{23} = 71$	72 66 74 $\bar{X}_{33} = 70.67$	$\bar{X}_3 = 78.56$
Altura 1.50	105 95 99 $\bar{X}_{14} = 99.67$	68 85 83 $\bar{X}_{24} = 78.67$	61 58 58 $\bar{X}_{34} = 59$	$\bar{X}_4 = 79.11$

Tabla resumen del ANOVA:

Multifactor ANOVA - yield

Analysis of Variance for yield - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:variedad	2287,17	2	1143,58	42,93	0,0000
B:altura_poda	1613,64	3	537,88	20,19	0,0000
INTERACTIONS					
AB	2284,61	6	380,769	14,29	0,0000
RESIDUAL	639,333	24	26,6389		
TOTAL (CORRECTED)	6824,75	35			



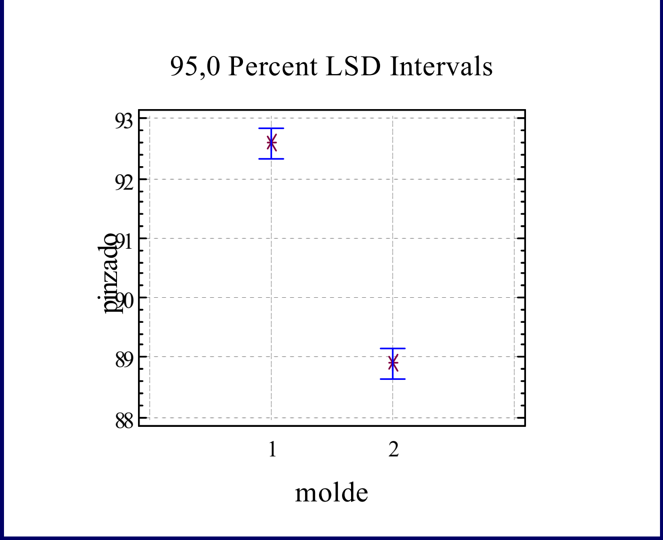
EJERCICIO-2:

DISEÑO REPLICADO

Plan factorial equilibrado 3x2 con 10 repeticiones

Experimento para investigar el efecto del catalizador y tipo de molde en un parámetro de calidad de botellas de plástico.

Objetivo: minimizar la variable respuesta.

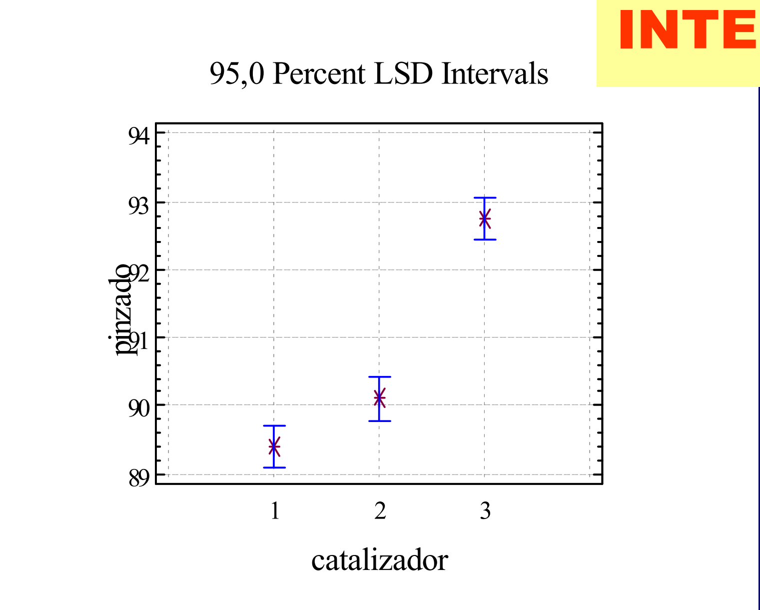


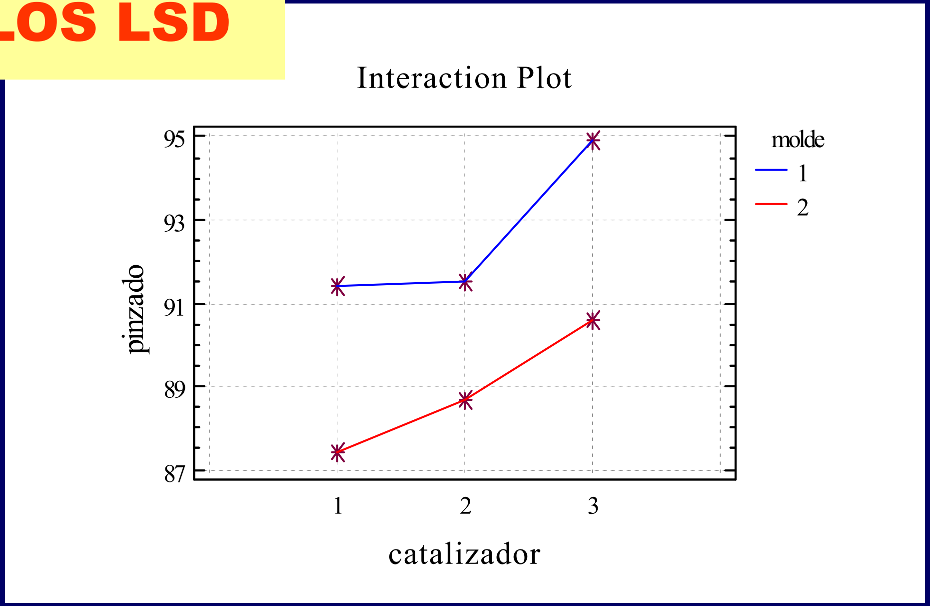
Molde: \ Catalizador:	A	B	C
1	93	92	95
	93	94	94
	90	90	94
	91	91	94
	92	90	94
	91	91	97
	90	92	95
	91	92	96
	93	92	94
	90	91	96
2	88	90	91
	88	88	90
	87	88	92
	87	88	90
	88	89	91
	87	90	89
	87	89	90
	87	88	91
	87	88	91
	88	89	91

Analysis of Variance for pinzado - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A: catalizador	124,9	2	62,45	63,99	0,0000
B: molde	205,35	1	205,35	210,42	0,0000
INTERACTIONS					
AB	6,3	2	3,15	3,23	0,0474
RESIDUAL	52,7	54	0,975926		
TOTAL (CORRECTED)	389,25	59			

INTERVALOS LSD



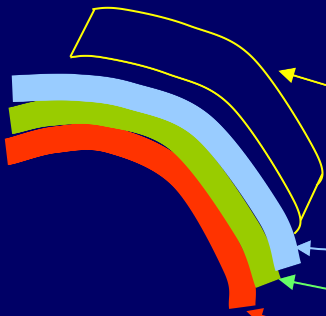


Condic. Operativas Óptimas (para minimizar): Cataliz (1), molde (2)

EJERCICIO-3: DISEÑO 2^3 NO REPLICADO

VARIABLE RESPUESTA: una medida del grado de corrosión de la chapa metálica de un vehículo (la cual debe ser lo más baja posible).

FACTORES:
- Electrocoat
- Primer
- Esmalte



NIVELES para cada factor: dos espesores distintos (2 y 5 micras)

EC	PR	ES	corrosión
-	-	-	14
+	-	-	10
-	+	-	8
+	+	-	6
-	-	+	12
+	-	+	4
-	+	+	6
+	+	+	2

NÚMERO DE PRUEBAS = 8 :

2^3 no replicado

$$SC_{TOTAL} = \sum SC_{EFECTOS} + SC_{RESIDUAL}$$

ANÁLISIS DE LA VARIANZA

¿Qué efectos son estadísticamente significativos?

EFECTO	S.C.	g.l.	C.M.	F-ratio	F-tablas	P-valor
TOTAL	115.5	7	-	-		
EC	40.5	1	40.5	16.2	7.71	0.0158
PR	40.5	1	40.5	16.2	7.71	0.0158
ES	24.5	1	24.5	9.8	7.71	0.0352
RESIDUAL	10	4	2.5			

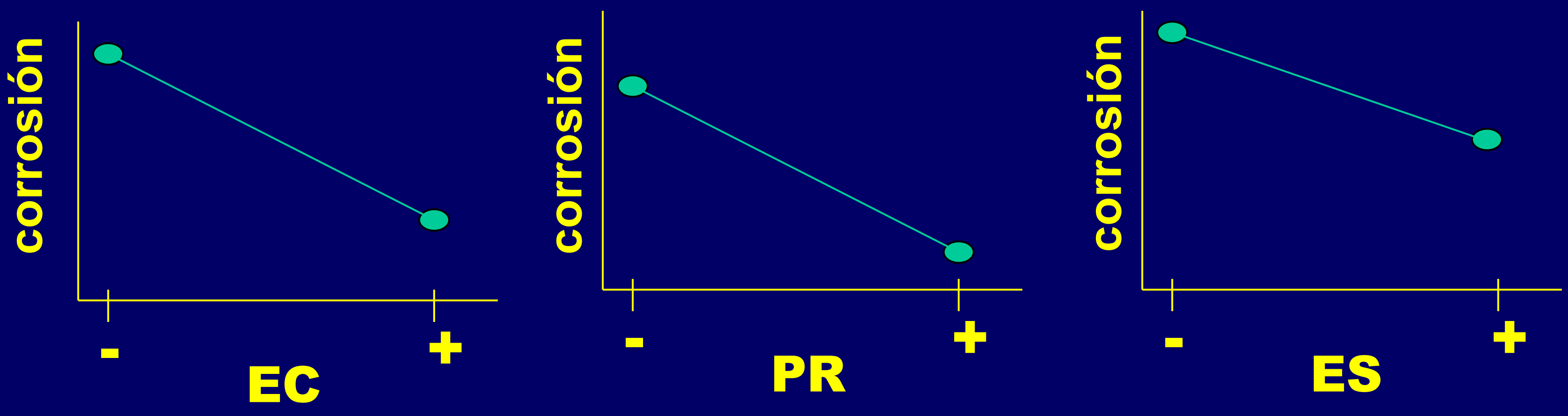
SE RECOMIENDA QUE g.l.res ≥ 4

¿Cómo conseguir g.l. para la SCresidual? → Agrupando en la $SC_{residual}$ los efectos menores

CONDICIONES OPERATIVAS ÓPTIMAS

Combinación de los factores significativos que optimizan (maximizan o minimizan) la variable respueta.

C.O.O. que minimizan la corrosión: EC^+ PR^+ ES^+ (espesores mayores)



ANÁLISIS DE RESIDUOS

Residuo para la observación *i* :

e_i = (valor observado)$_i$ – (media del tratamiento)$_i$

IMPORTANTE PARA: - Detectar datos anómalos

- Comprobar la hipótesis de normalidad
- Estudiar efectos sobre la varianza de la corrosión.

EJERCICIO-4: DISEÑO REPLICADO 2^2

En una reacción química, el objetivo es mejorar la calidad de una mezcla usada para incorporar cierto aditivo a un polímero.

Variable respuesta: % aditivo **Objetivo: maximizar**

2 FACTORES:
- **Velocidad de agitación en la mezcla**
- **Tiempo de agitación**

		TIEMPO: 3 '	TIEMPO: 6 '
VELOCIDAD	600 rpm	17.2 17.0 17.1 $\overline{X} = 17.1$	16.4 16.8 15.6 $\overline{X} = 16.267$
	1000 rpm	18.7 19.0 18.6 $\overline{X} = 18.767$	19.4 17.7 17.4 $\overline{X} = 18.167$

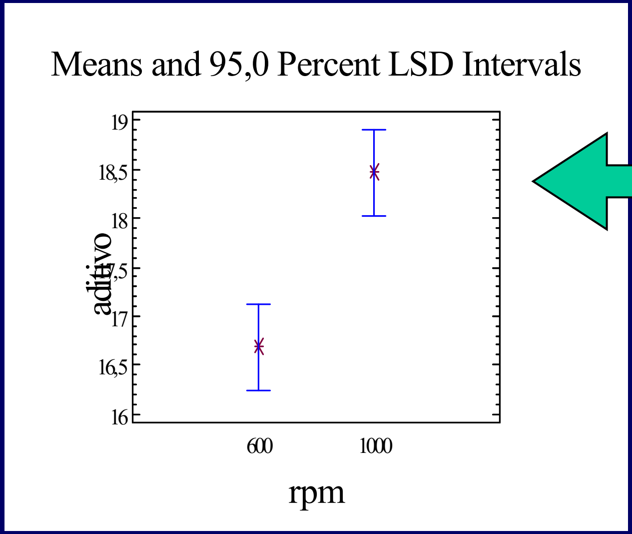
Analysis of Variance for aditivo

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:tiempo	1,54083	1	1,54083	3,88	0,0845
B:velocidad	9,54083	1	9,54083	24,00	0,0012
AB:tiemp*veloc	0,04083	1	0,04083	0,10	0,7568
Total error	3,18	8	0,3975		
Total (corr.)	14,3025	11			

Efecto "casi" significativo

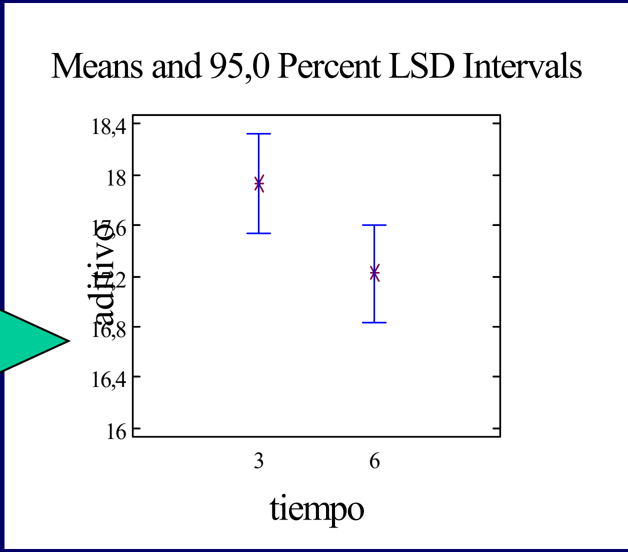
Claramente significativo

No hay evidencia de interacción entre los factores



EFECTO DE:
VELOCIDAD

TIEMPO



Como el efecto del tiempo es casi significativo, es conveniente tenerlo en cuenta.

¿Cuáles serían las C.O.O.?

EJERCICIO-5: Fracción Factorial 2^{4-1}

Ensayo para estudiar el efecto de 4 factores en la estabilidad de un producto químico.

Objetivo: maximizar la estabilidad

	Variable	-	+
1	Concentración de ácido (%)	20	30
2	Concentración de catalizador (%)	1	2
3	Temperatura (° C)	100	150
4	Concentración de monómeros	25	50

ensayo	V1	V2	V3	V4	ESTABILIDAD
1	-	-	-	-	20
2	+	-	-	+	14
3	-	+	-	+	17
4	+	+	-	-	10
5	-	-	+	+	19
6	+	-	+	-	13
7	-	+	+	-	14
8	+	+	+	+	10

Analysis of Variance for STABILITY - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:ACIDO	66,125	1	66,125	144,27	0,0012
B:TEMPERATURA	3,125	1	3,125	6,82	0,0796
C:MONOMERO	1,125	1	1,125	2,45	0,2152
D:CATALIZADOR	28,125	1	28,125	61,36	0,0043
RESIDUAL	1,375	3	0,458333		
TOTAL (CORRECTED)	99,875	7			

Analysis of Variance for STABILITY - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:ACIDO	66,125	1	66,125	58,78	0,0006
B:CATALIZADOR	28,125	1	28,125	25,00	0,0041
RESIDUAL	5,625	5	1,125		
TOTAL (CORRECTED)	99,875	7			

Ejercicio: determinar las condiciones operativas óptimas

EJERCICIO-6

Se estudian 5 factores para mejorar el rendimiento de un proceso químico (objetivo: maximizar la variable respuesta).

Trat.	resp.	A	B	C	D	E
1	15.6	-	-	-	-	-
2	13.5	+	-	-	-	-
3	16.3	-	+	-	-	-
4	17.1	+	+	-	-	-
5	26.8	-	-	+	-	-
6	25	+	-	+	-	-
7	30	-	+	+	-	-
8	28.9	+	+	+	-	-
9	15.4	-	-	-	+	-
10	12.7	+	-	-	+	-
11	15.3	-	+	-	+	-
12	15.9	+	+	-	+	-
13	20.3	-	-	+	+	-
14	21.3	+	-	+	+	-
15	27	-	+	+	+	-
16	24.1	+	+	+	+	-
17	28.9	-	-	-	-	+
18	29	+	-	-	-	+
19	33.7	-	+	-	-	+
20	33.6	+	+	-	-	+

Trat.	resp.	A	B	C	D	E
21	47.4	-	-	+	-	+
22	44.2	+	-	+	-	+
23	52.6	-	+	+	-	+
24	46.2	+	+	+	-	+
25	27.8	-	-	-	+	+
26	29.5	+	-	-	+	+
27	30.1	-	+	-	+	+
28	29.6	+	+	-	+	+
29	35.9	-	-	+	+	+
30	36.4	+	-	+	+	+
31	41	-	+	+	+	+
32	38.6	+	+	+	+	+

1) Resultados del ANOVA con los datos de los 32 ensayos:

Source of variation	Sum of Squares	Df	Mean Square	F-Ratio	Sig. Level
TOTAL	3514.4847	31			
B	79.0653	1	79.0653	21.656	.0001
C	1031.7153	1	1031.7153	282.35	.0000
D	144.0753	1	144.0753	39.462	.0000
E	2101.1403	1	2101.1403	575.499	.0000
C*D	63.562812	1	63.562812	17.410	.0003
RESIDUAL	94.925625	26	3.6509856		

2) ANOVA si SÓLO las 16 pruebas en rojo se hubiesen realizado:

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
B	47,2656	1			
C	530,151	1			
D	60,4506	1			
E	1028,81	1			
INTERACTIONS					
BC	50,0556	1			
RESIDUAL					
TOTAL (CORRECTED)	1774,76				

(completar la tabla y obtener conclusiones)

- Comprobar que los efectos relevantes son casi los mismos con 16 pruebas

- Obtener las condiciones operativas óptimas (solución: $B^+ C^+ D^- E^+$)

EJERCICIO-7: FRACCIÓN FACTORIAL 2^{6-3}

Experimento para mejorar la resistencia frente a la corrosión de chapas metálicas.

6 factores estudiados (todos a 2 niveles)

	A	B	C	D	E	F	RESP.
1	-	-	-	+	+	+	8
2	+	-	-	-	-	+	6
3	-	+	-	-	+	-	7
4	+	+	-	+	-	-	7
5	-	-	+	+	-	-	10
6	+	-	+	-	+	-	8
7	-	+	+	-	-	+	8
8	+	+	+	+	+	+	9

FACTOR	SC	g.l.	CM	$F_{CALCULADA}$
A	1.125	1	1.125	9
B	0.125	1	0.125	1
C	6.125	1	6.125	49
D	3.125	1	3.125	25
E	0.125	1	0.125	1
F	0.125	1	0.125	1
RESIDUAL	0.125	1	0.125	

Conviene agrupar en la $SC_{RESIDUAL}$ aquellos factores claramente no significativos para tener al menos 4 grados de lib. residuales

FACTOR	SC	g.l.	CM	$F_{CALCULADA}$
A	1.125	1	1.125	9
C	6.125	1	6.125	49
D	3.125	1	3.125	25
RESIDUAL	0.5	4	0.125	

$F_{1,4}(\alpha=0.05) = 7.71$

¿Qué efectos son estadísticamente significativos?

Determinar las condiciones óptimas para maximizar la resistencia

EJERCICIO-8: FRACCIÓN FACTORIAL 2^{6-2}

Experimento para estudiar el efecto de 6 parámetros de un proceso químico sobre la resistencia en un adhesivo (objetivo: maximizar)

VARIABLE (unidades)	NIVEL	
	-1	1
SACAROSA (GR)	43	71
PARAFORMOL (GRS)	30	42
NaOH (GR)	6	10
AGUA (GR)	16	20
MAX. TEMPER. (ºC)	80	90
TIEMPO	25	35

X_1	X_2	X_3	X_4	X_5	X_6	y
-1	-1	-1	-1	-1	-1	162
1	-1	-1	-1	1	-1	146
-1	1	-1	-1	1	1	182
1	1	-1	-1	-1	1	133
-1	-1	1	-1	1	1	228
1	-1	1	-1	-1	1	143
-1	1	1	-1	-1	-1	223
1	1	1	-1	1	-1	172
-1	-1	-1	1	-1	1	168
1	-1	-1	1	1	1	128
-1	1	-1	1	1	-1	175
1	1	-1	1	-1	-1	186
-1	-1	1	1	1	-1	197
1	-1	1	1	-1	-1	175
-1	1	1	1	-1	1	196
1	1	1	1	1	1	173

Los efectos simples NO están confundidos con las interacciones dobles

Es posible calcular en total 15 efectos (N-1):

- **6 EFECTOS SIMPLES**
  (confundidas con interacciones de orden > 2)
- **7 INTERACCIONES DOBLES**
  (confundidas con interacciones de orden > 2)
- **2 INTERACCIONES DE ORDEN >2**

Analysis of Variance for Y - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:F1	4726,56	1	4726,56	14,23	0,0023
B:F3	3220,56	1	3220,56	9,69	0,0082
RESIDUAL	4319,31	13	332,255		
TOTAL (CORRECTED)	12266,4	15			

Se debería comprobar que ninguna interacción doble es significativa

Determinar las condic. operativas óptimas para maximizar la resistencia

(solución: F1 a nivel '-' y F3 a nivel '+')

EJERCICIO-9

En un cierto proceso, el tiempo es un parámetro de calidad que se pretende minimizar.

FACTORES ESTUDIADOS:

· 3 a 2 NIVELES (A C E)

· 5 a 3 NIVELES (B D F G H)

Diseño empleado: L_{18} adaptando las columnas 3ª y 5ª para acomodar factores a 3 niveles

Ensayo	A	B	C	D	E	F	G	H	TIEMPO
1	1	1	1	1	1	1	1	1	852
2	1	1	2	2	2	2	2	2	540
3	1	1	2	3	2	3	3	3	417
4	1	2	1	1	2	2	3	3	1282
5	1	2	2	2	2	3	1	1	505
6	1	2	2	3	1	1	2	2	445
7	1	3	1	2	1	3	2	3	852
8	1	3	2	3	2	1	3	1	482
9	1	3	2	1	2	2	1	2	707
10	2	1	1	3	2	2	2	1	492
11	2	1	2	1	1	3	3	2	975
12	2	1	2	2	2	1	1	3	450
13	2	2	1	2	2	1	3	2	722
14	2	2	2	3	1	2	1	3	402
15	2	2	2	1	2	3	2	1	732
16	2	3	1	3	2	3	1	2	482
17	2	3	2	1	2	1	2	3	855
18	2	3	2	2	1	2	3	1	515

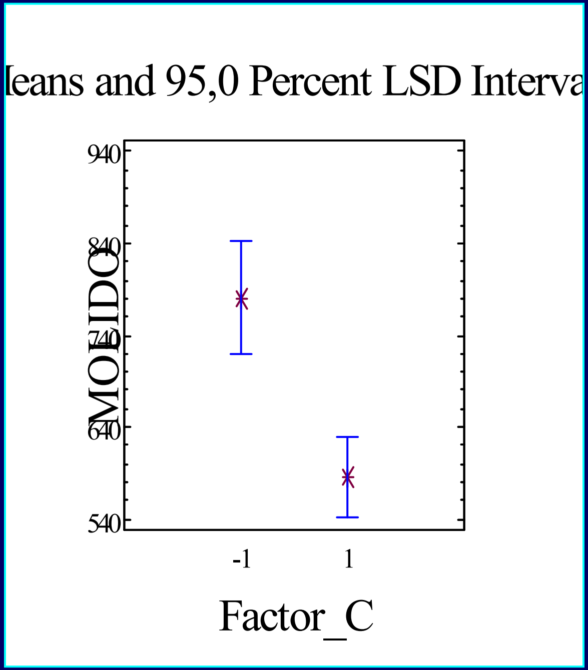
Analysis of Variance for TIME - Type III Sums of Squares

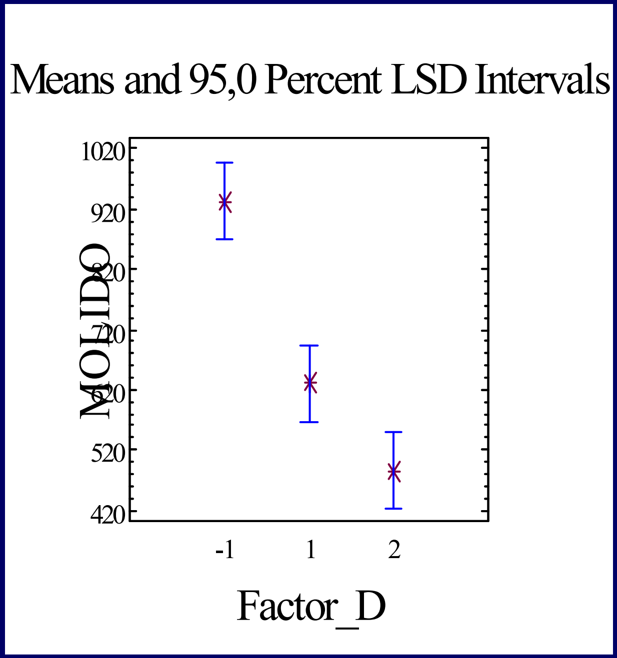
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:Factor_A	11602,7	1	11602,7	1,00	0,3729
B:Factor_B	10942,1	2	5471,06	0,47	0,6536
C:Factor_C	151970,0	1	151970,0	13,16	**0,0222**
D:Factor_D	625208,0	2	312604,0	27,07	**0,0047**
E:Factor_E	4807,11	1	4807,11	0,42	0,5539
F:Factor_F	2372,11	2	1186,06	0,10	0,9047
G:Factor_G	82548,8	2	41274,4	3,57	**0,1287**
H:Factor_H	38778,8	2	19389,4	1,68	0,2955
RESIDUAL	46186,5	4	11546,6		
TOTAL (CORRECTED)	974416,0	17			

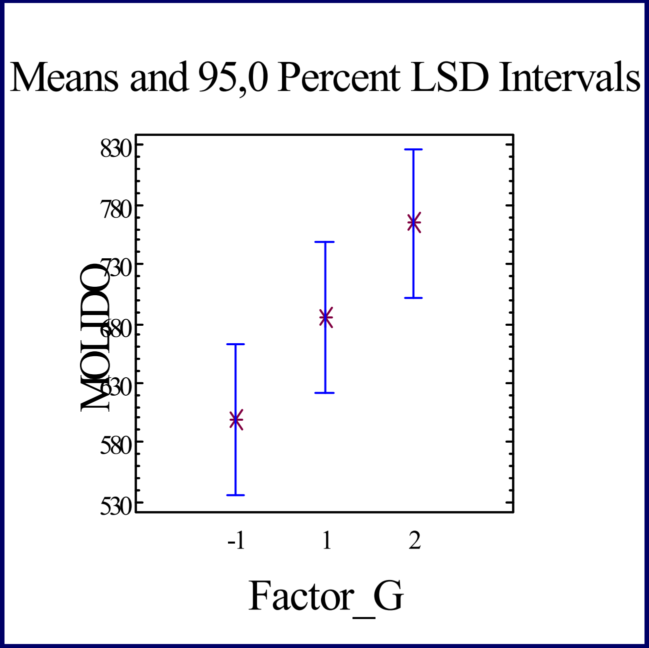
Analysis of Variance for TIME - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:Factor_D	625208,0	2	312604,0	32,71	0,0000
B:Factor_G	82548,8	2	41274,4	4,32	0,0387
C:Factor_C	151970,0	1	151970,0	15,90	0,0018
RESIDUAL	114689,0	12	9557,45		
TOTAL (CORRECTED)	974416,0	17			

Los factores C, D, G son estadísticamente significativos







Condiciones operativas óptimas para minimizar: C (2) D (3) G (1)

Se observa cierto efecto cuadrático en el factor D

(se podría estudiar con regresión lineal múltiple)

El efecto de G es claramente lineal (no cuadrático) en el intervalo ensayado. Por tanto, si fuera posible, sería conveniente disminuir más el nivel del factor G.