

DISEÑO EXPERIMENTAL Y OPTIMIZACIÓN DE SISTEMAS CON MÚLTIPLES RESPUESTAS

Parte 2: Selección de factores

Héctor Goicoechea

E-mail: hgoico@fcb.unl.edu.ar

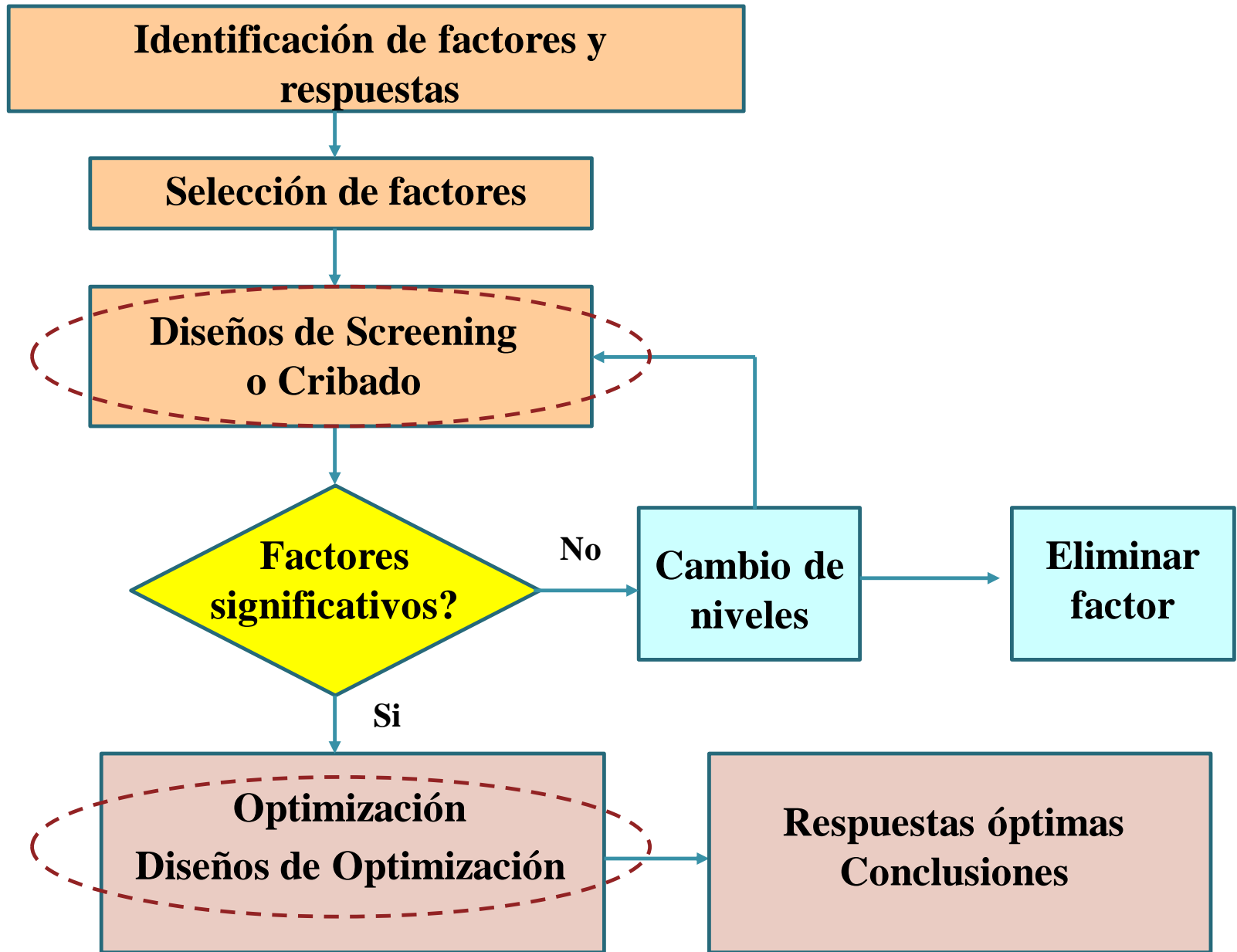
<http://www.fcb.unl.edu.ar/laboratorios/ladaq/>



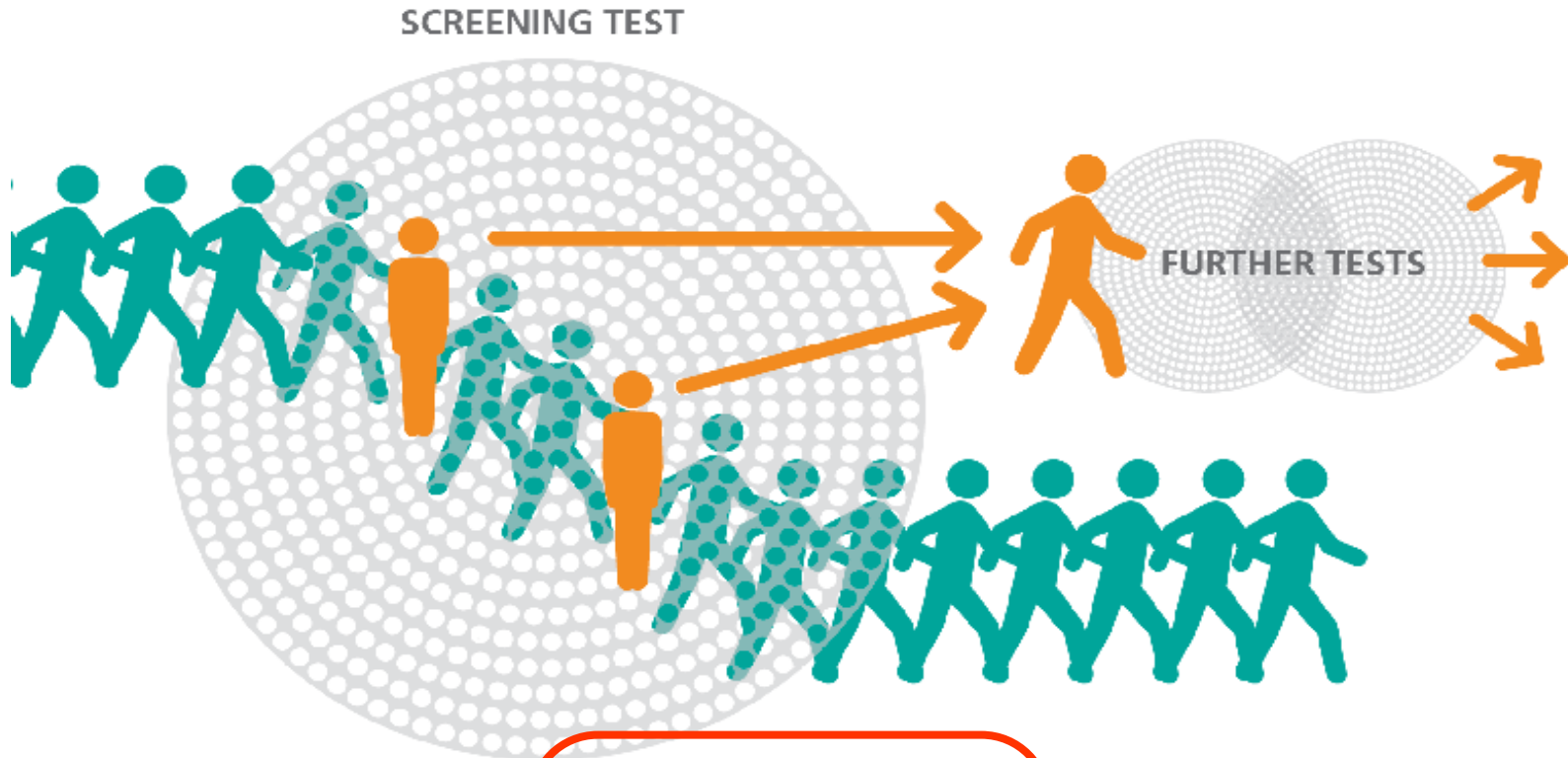
Facultad de Bioquímica y
Ciencias Biológicas



Universidad Nacional del
Litoral



Etapa de *screening*: selección de factores



Investigar todos los factores

- Muchas pruebas experimentales
- Mucho tiempo
- Mucho dinero

Diseños de *screening* para la selección de factores

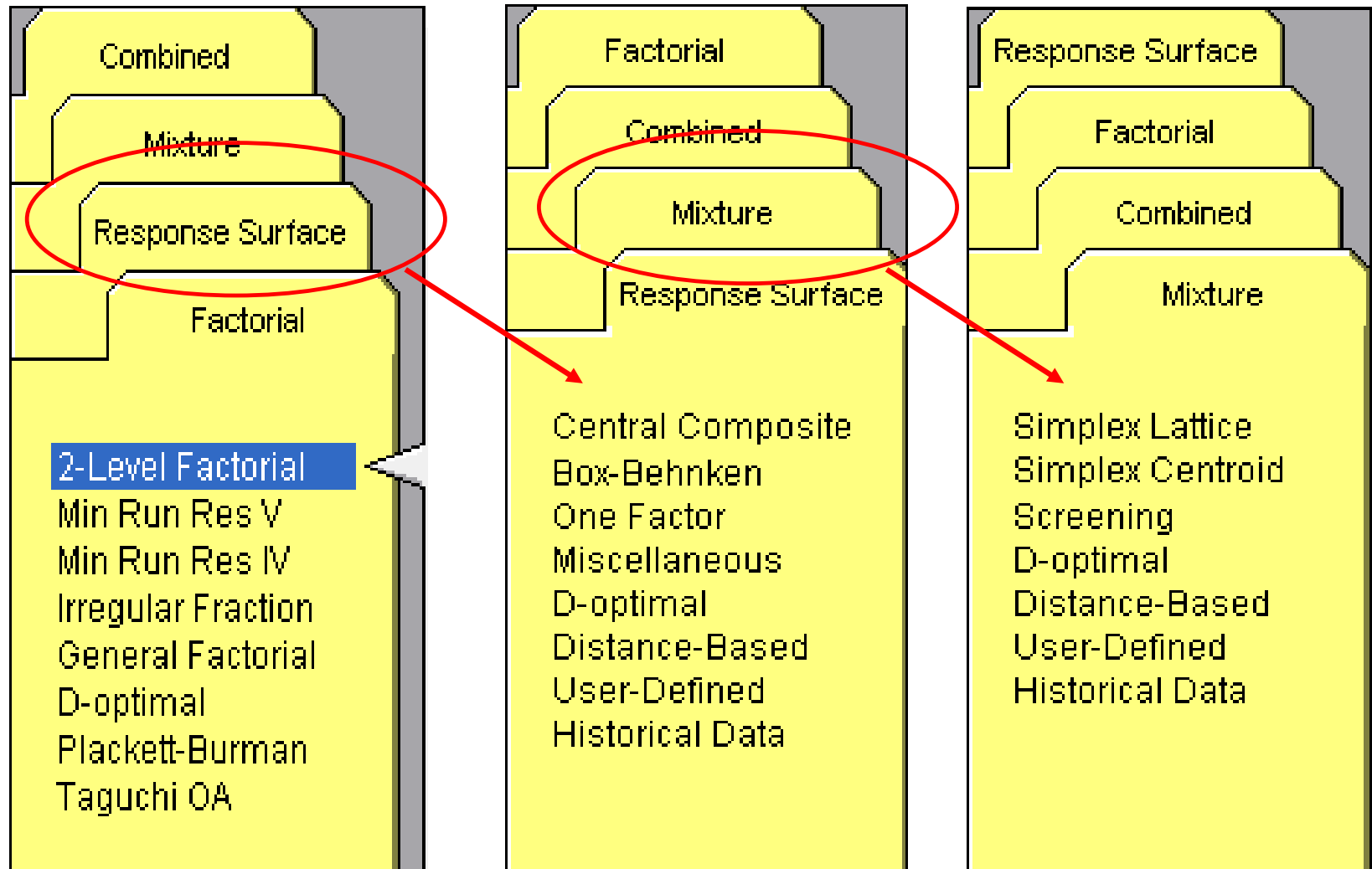
Primera etapa del análisis

- Delimitar el **problema**, definir la **hipótesis**, establecer el **objetivo** del experimento-resultado deseado.
 - Seleccionar la **variable respuesta** que representa al problema que hay que resolver y determinar **cómo se va a medir** de manera confiable.

Primera etapa del análisis

- Determinar **qué factores se van a estudiar**, para analizar su influencia sobre la respuesta, sobre la base de información y experiencia previa.
- Seleccionar el **rango experimental a estudiar** y el **diseño** de experimentos adecuado.
- **Realizar el experimento** y seleccionar los factores que tengan una influencia significativa sobre la respuesta.

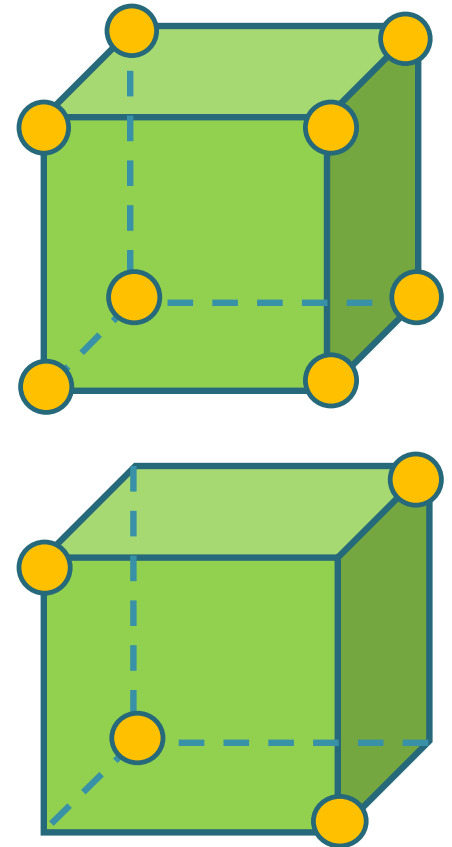
Diseños experimentales: gran variedad



Etapa de screening: Selección de factores

Diseños más usados para realizar experimentos de previsualización o *screening* para factores numéricos y categóricos

- Factorial completo o total a dos niveles (2^k)
- Factorial fraccionado a dos niveles (2^{k-p})
- Plackett-Burman (muchos factores)



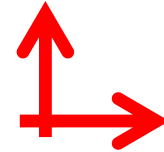
Selección del diseño

Aspectos que influyen en la selección de un diseño

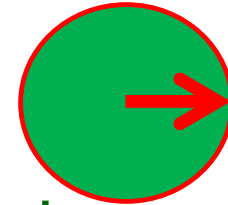
- Naturaleza del problema, conocimiento previo y tipo de información que se desea obtener.
- Número y tipo de factores e interacciones que se deben estudiar.
 - Restricciones operativas, de costo y tiempo.
 - Necesidad de bloqueo.
- Características o propiedades de cada diseño.
- Facilidad de comprensión e implementación.

Características de los diseños

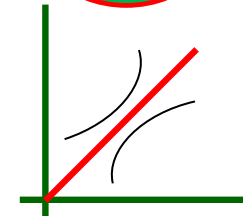
Ortogonalidad



Rotabilidad



Error estándar-leverage



Eficiencia

Resolución

ORTOGONALIDAD

- Los coeficientes estimados en el modelo ajustado no están correlacionados entre si.
- Las columnas de la matriz del diseño deben estar formadas por vectores independientes entre si.



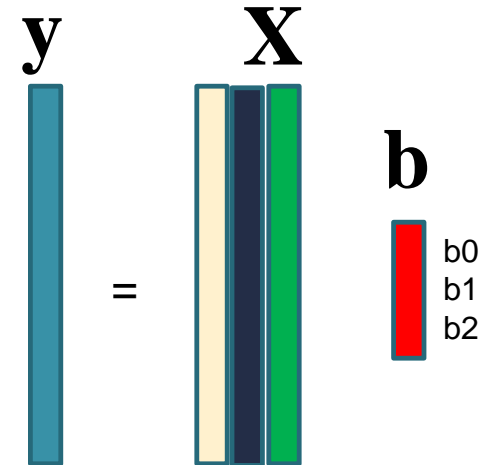
**Minimizar la Varianza de los
Coeficientes de Regresión**

Modelo para un sistema lineal de 2 componentes:

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$y = \mathbf{X} \mathbf{b} + \mathbf{e} = \mathbf{y}_{\text{pred}} + \mathbf{e}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \mathbf{b}$$



$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

$$\mathbf{b} = [b_0 ; b_1 ; b_2]$$

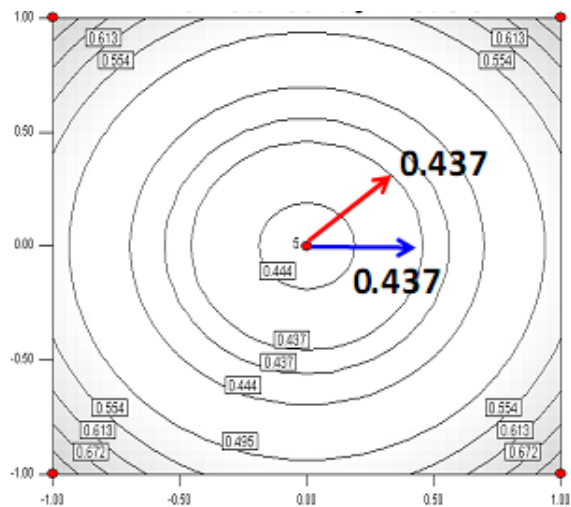
$$\mathbf{y}_{\text{pred}} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \rightarrow \mathbf{y}_{\text{pred}} = \mathbf{H} \mathbf{y} \text{ (} \mathbf{H} \text{ es conocida como matriz "hat" por sombrero)}$$

$$\mathbf{X} \mathbf{b} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

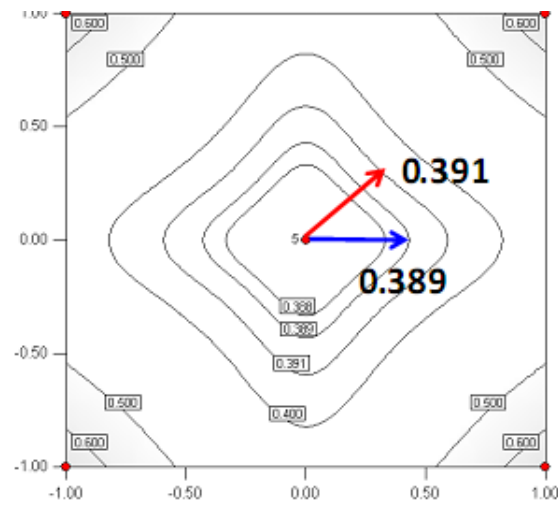
Pero $\mathbf{X}^T \mathbf{X}$ es singular en un diseño cuyas columnas están correlacionadas

ROTABILIDAD

La **varianza de la respuesta predicha** en un punto cualquiera del espacio experimental es solamente función de la distancia al punto del centro del diseño.



Rotable

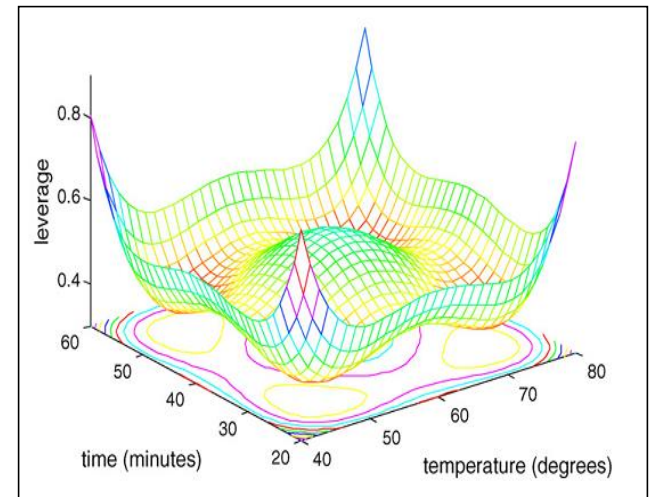


No Rotable

Proporcionar un **error de predicción** estable en el entorno experimental

LEVERAGE (PALANCA)

- Depende de la **matriz experimental** y del **modelo seleccionado** (construcción de \mathbf{X}).
- No depende del error experimental de la respuesta y debe evaluarse antes de experimentar.
- La varianza de la respuesta estimada en un punto se obtiene multiplicando el leverage por la varianza experimental.



EFICIENCIA

- **Cociente entre coeficientes estimados y puntos experimentales.**
- Se necesitan como mínimo la **misma cantidad de puntos experimentales** diferentes en el diseño **que coeficientes a estimar.**
- Para evaluar la falta de ajuste se hacen repeticiones de un punto del diseño.

RESOLUCIÓN

Diseños factoriales fraccionados

Indica el nivel de confusiones que se presentan en la estimación de efectos, dando una idea de que tan bien pueden estimarse los efectos potencialmente importantes mediante el diseño.

Confusiones

Diseños factoriales fraccionados

Cuando dos o más factores, con nombres distintos, comparten la misma columna de signos (-1 y +1), no se sabe que factor es el responsable del efecto estimado.

Hay confusión

Experimento	FACTOR A	FACTOR B	Interac. A×B
1	-1	-1	+1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	+1

DISEÑOS FACTORIALES

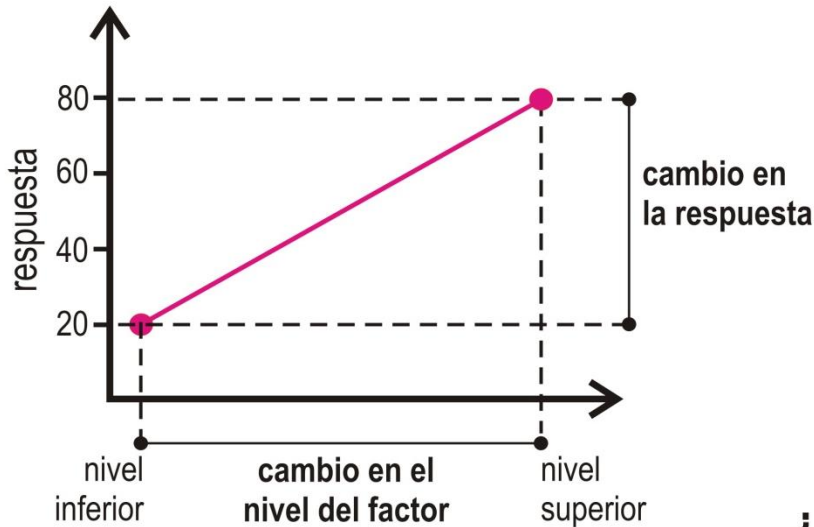
Diseños factoriales completos a dos niveles: 2^k

- Se investigan todas las posibles combinaciones de los niveles de los factores.
 - La cantidad de puntos experimentales esta dada por 2^k
 - Útiles para $2 \leq k \leq 5$ (4 a 32 tratamientos)
- Permiten estimar los efectos de todos los factores principales y sus interacciones.
 - Son diseños ortogonales y rotables.

El número de experimentos crece rápidamente con el número de factores

Diseños factoriales completos a dos niveles

Efecto de un factor – Representación gráfica



$$\text{EFECTO del factor} = (\text{respuesta nivel superior}) - (\text{respuesta nivel inferior})$$

80

20

$$\text{EFECTO del factor} = 60$$

¿Que forma tiene la respuesta entre los dos niveles de medición?

aproximación
LINEAL

Efectos estimables en los diseños factoriales completos a dos niveles

Factores	Diseño	Experimentos	Efectos estimables
A B	2^2	4	A B AB
A B C	2^3	8	A B C AB AC BC ABC
A B C D	2^4	16	A B C D AB AC AD BC BD CD ACD ABC BCD ABD ABCD
A B C D E	2^5	32	A B C D E AB AC AD AE BC BD BE CD CE DE ABC ABD ABE ACD ACE ADE BCD BCE CDE BDE ABCD ABCE BCDE ACDE ABDE ABCDE

Diseños factoriales completos a dos niveles

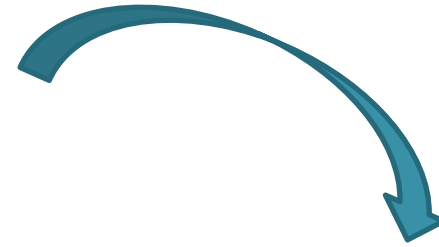
La construcción de la matriz del diseño (k columnas y 2^k renglones considerando una réplica) se hace alternando los signos $-$ y $+$ en la primer columna, dos menos y dos más en la segunda, cuatro menos y cuatro más en la tercera y así sucesivamente.

	Experimento	Factor A	Factor B	Factor C	Factor D	Factor E
2^2	1	-1	-1	-1	-1	-1
	2	+1	-1	-1	-1	-1
	3	-1	+1	-1	-1	-1
	4	+1	+1	-1	-1	-1
2^3	5	-1	-1	+1	-1	-1
	6	+1	-1	+1	-1	-1
	7	-1	+1	+1	-1	-1
	8	+1	+1	+1	-1	-1
2^4	9	-1	-1	-1	+1	-1
	10	+1	-1	-1	+1	-1
	11	-1	+1	-1	+1	-1
	12	+1	+1	-1	+1	-1

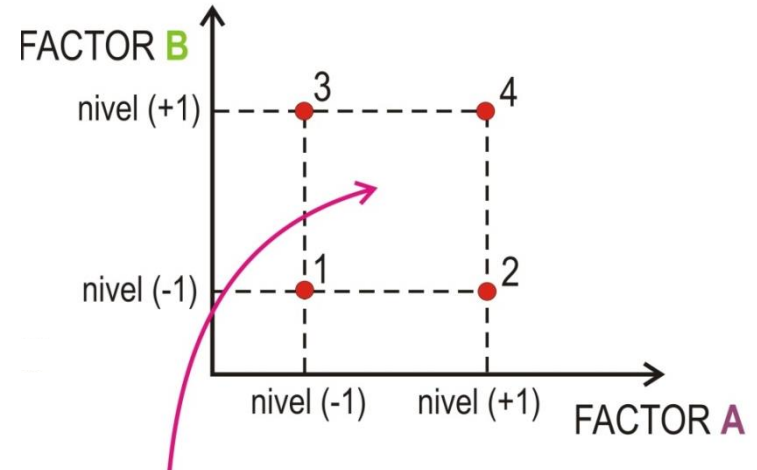
Diseños factoriales completos a dos niveles

Matriz del diseño 2^2

Experimento	FACTOR A	FACTOR B
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1



Representación geométrica



región experimental
donde son válidas las
conclusiones del experimento

Diseños factoriales completos a dos niveles

Efecto de un factor

$$\text{EFECTO} = \frac{\sum(\text{Rtas nivel superior})}{n^{\circ} \text{ de respuestas}} - \frac{\sum(\text{Rtas nivel inferior})}{n^{\circ} \text{ de respuestas}}$$

diferencia entre las medias aritméticas de los niveles del factor

Planilla experimental

Experimento	FACTOR A	FACTOR B	Respuesta
1	-1	-1	40
2	+1	-1	80
3	-1	+1	36
4	+1	+1	70

$$\begin{aligned}\text{EFECTO Factor A} &= \frac{(80 + 70)}{2} - \frac{(40 + 36)}{2} \\ \text{EFECTO A} &= (75) - (38) = \mathbf{37}\end{aligned}$$

$$\begin{aligned}\text{EFECTO Factor B} &= \frac{(36 + 70)}{2} - \frac{(40 + 80)}{2} \\ \text{EFECTO B} &= (53) - (60) = \mathbf{-7}\end{aligned}$$

Diseños factoriales completos a dos niveles

Efecto de la interacción

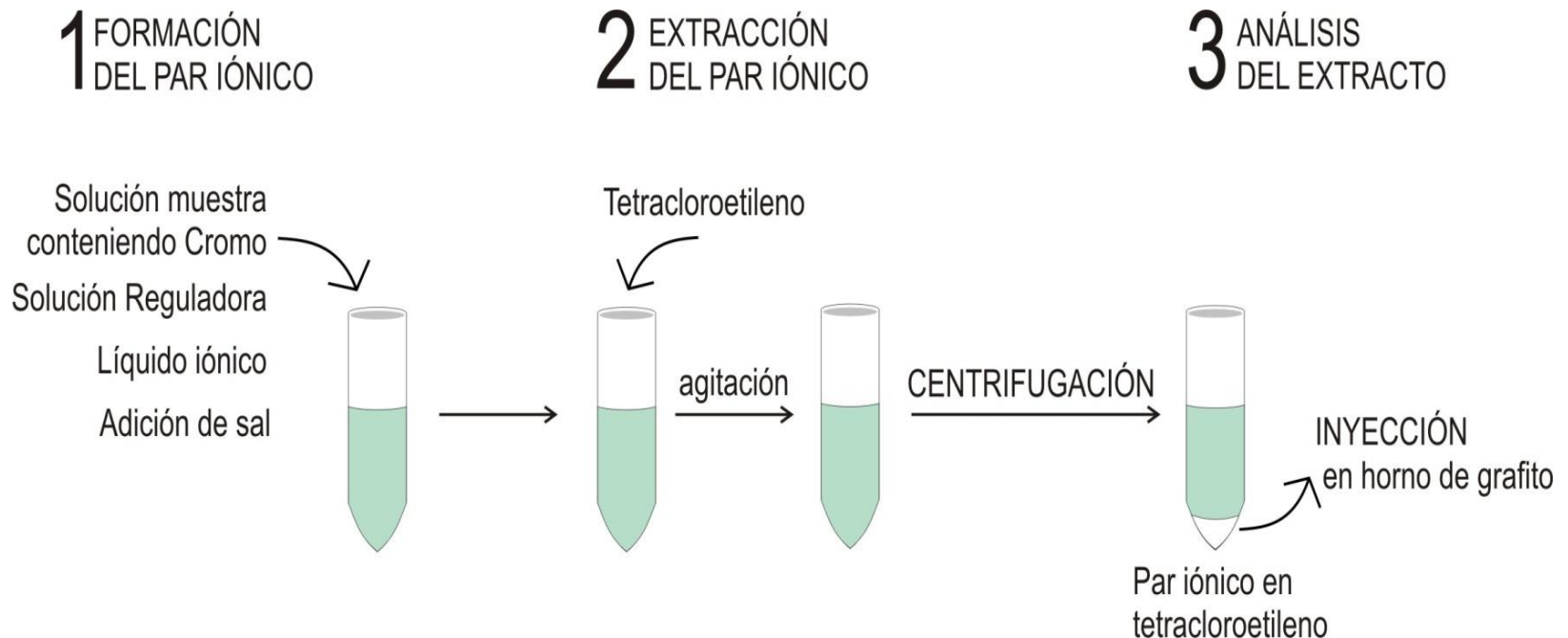
Experimento	FACTOR A	FACTOR B	Interac. A×B	Respuesta
1	-1	-1	+1	40
2	+1	-1	-1	80
3	-1	+1	-1	36
4	+1	+1	+1	70

$$\text{EFECTO } A \times B = \frac{(40 + 70)}{2} - \frac{(80 + 36)}{2}$$

$$\text{EFECTO } A \times B = (55) - (58) = -3 \rightarrow \text{Efecto de la interacción entre A y B}$$

Ejemplo:

Se necesita incrementar la eficiencia de un procedimiento de extracción de cromo (VI) en muestras de agua de río utilizando la metodología de microextracción en fase líquida



A microextraction procedure based on an ionic liquid as an ion-pairing agent optimized using a design of experiments for chromium species separation and determination in water samples. *Anal. Methods*, 2013, 5, 5065. P Berton, L Vera Candioti, H Goicoechea, R Wuilloud

EXTRACCIÓN de cromo de una muestra acuosa

FACTORES posibles

Factor 1: volumen de la muestra

Factor 2: volumen de la solución reguladora (SR)

Factor 3: concentración de la SR

Factor 4: pH de la SR

Factor 5: volumen de líquido iónico (LI)

Factor 6: concentración del LI

Factor 7: adición de sal

Factor 8: tiempo de formación del par iónico

Factor 9: tipo de solvente de extracción

Factor 10: volumen de solvente de extracción

Factor 11: tiempo de extracción

Factor 12: tipo de agitación

Factor 13: velocidad de centrifugación

Factor 14: tiempo de centrifugación

Diseño factorial completo:

$$2^k = 2^{14} = 16384 \text{ experimentos}$$

Si consideramos sólo 7 factores: $2^7 = 128$ experimentos



**¿Cuántos experimentos
debo realizar?**



SON MUCHOS!!!!

Considerando un diseño 2^7 con 128 experimentos:

Los 7 FACTORES son: A B C D E F G

Qué efectos se pueden estimar con 128 experimentos?

7 efectos simples: efectos principales A B C D E F G

21 interacciones dobles AB CE GA FD

35 interacciones de tercer orden DBE CFG CAB

35 interacciones de cuarto orden ABEG DCFE

21 interacciones de quinto orden GADBE

7 interacciones de sexto orden AGFEDC

1 interacción de séptimo orden ABCDEFG

Se pueden
estimar 127
efectos

¿tienen sentido práctico las
interacciones de orden elevado?

**Hay que encontrar una estrategia que permita reducir de
manera importante el número de tratamientos
experimentales**

Diseños factoriales fraccionados

- Se investigan **ALGUNAS** de todas las posibles combinaciones de los niveles de los factores

La cantidad de experimentos a realizar esta dado por:

$$2^{k-p}$$

p indica cuantas veces se fracciona o reduce el diseño completo a la mitad

- 2^{k-1} el diseño se reduce a la mitad
- 2^{k-2} el diseño se reduce a la cuarta parte
- 2^{k-3} el diseño se reduce a la octava parte

Diseños factoriales fraccionados

Se basa en
dos principios

PRINCIPIO de PARETO

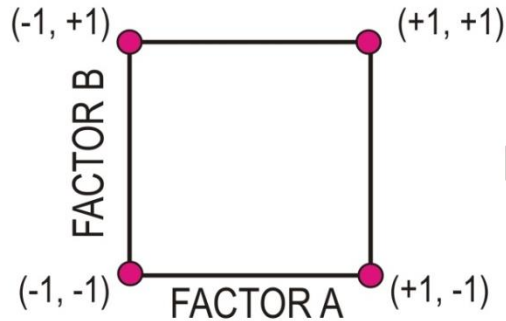
Al incluir una elevada cantidad de factores en la fase inicial de un experimento, unos pocos son los responsables de la mayor variabilidad de la respuesta.

JERARQUIZACIÓN de EFECTOS

Son más importantes los efectos principales, seguidos por las interacciones dobles, triples, etc.

Diseños factoriales fraccionados

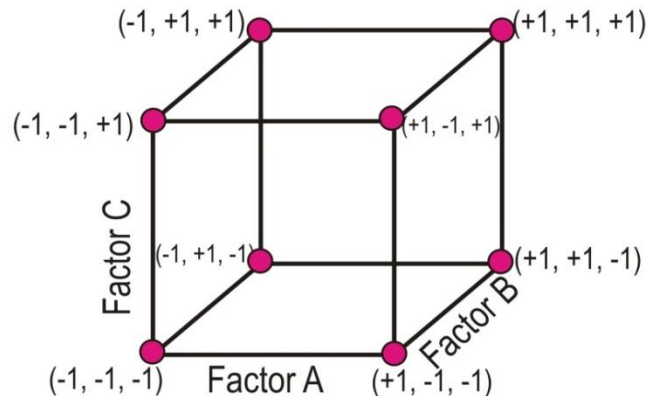
2^2



No se puede fraccionar

COMPLETO

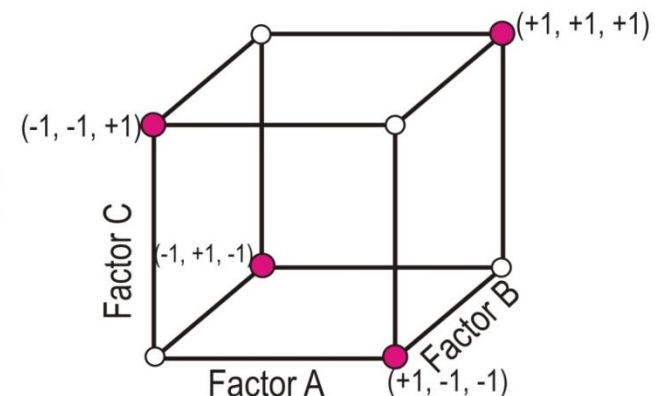
2^3



8 experimentos

FRACCIONADO

2^{3-1}



4 experimentos

Mientras más grande es el valor de k , mayor es el grado de fraccionamiento que admite el diseño

Diseños factoriales fraccionados

¿Qué información obtengo?

Ejemplo para 7 factores

efectos principales +
interacciones dobles

interacciones triples,
cuádruples, etc.

Diseño Factorial	Experimentos	Efectos estimables	Efectos no ignorables	Efectos ignorables
Completo 2^7	128	127	28	99
Fraccionado 2^{7-1}	64	63	28	35
Fraccionado 2^{7-2}	32	31	28	3
Fraccionado 2^{7-3}	16	15	28	—
Fraccionado 2^{7-4}	8	7	28	—

con pocos experimentos **se pierde información** que puede ser relevante

exceso de información

Quando realizo muchos experimentos y estimo sólo los efectos potencialmente importantes, tengo más grados de libertad para estimar el error aleatorio

Diseños factoriales fraccionados

“Aliasing”

Efectos alias son efectos que se confunden y no es posible separarlos.

Para interpretar los efectos alias es necesario suponer que solo uno de ellos es responsable del efecto observado y que los demás efectos son nulos.

Se utiliza el principio de jerarquía.

Diseños factoriales fraccionados

“Aliasing”

No es una buena estrategia utilizar diseños donde se alían dos efectos que son potencialmente importantes, tales como efectos principales y las interacciones dobles.

Diseños factoriales fraccionados

Diseño 2^{4-1}

1. Se escribe el diseño 2^{k-p} como si fuese el factorial completo para $k - p$ factores

Experimento	Factor A	Factor B	Factor C
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

Planilla experimental para un diseño factorial completo 2^3

¿Como construyo la columna del cuarto factor?

2. Se eligen p **generadores** iniciales

es una interacción del más alto orden posible que va a dar origen a la columna de signos de un factor

La interacción del más alto orden en un diseño 2^3 es AxBxC que sería el generador del nuevo factor

Diseños factoriales fraccionados

Diseño 2^{4-1}

3. Para los últimos p factores las columnas de signos se obtienen multiplicando las columnas que indican los generadores

Experimento	Factor A	Factor B	Factor C	Factor D
1	-1	-1	-1	-1
2	+1	-1	-1	+1
3	-1	+1	-1	+1
4	+1	+1	-1	-1
5	-1	-1	+1	+1
6	+1	-1	+1	-1
7	-1	+1	+1	-1
8	+1	+1	+1	+1

interacción AxBxC



Diseños factoriales fraccionados

Diseño 2^{5-2}

1. Construyo la matriz experimental con un diseño factorial completo 2^3 (factores A, B y C)
2. Como debo generar dos columnas más (factores D y E) necesito 2 **generadores** que serán interacciones dobles

Interacción AxB

Interacción AxC

Experim.	Factor A	Factor B	Factor C	Factor D	Factor E
1	+1	+1	-1	+1	-1
2	-1	-1	-1	+1	+1
3	-1	+1	-1	-1	+1
4	+1	-1	+1	-1	+1
5	+1	+1	+1	+1	+1
6	+1	-1	-1	-1	-1
7	-1	+1	+1	-1	-1
8	-1	-1	+1	+1	-1

Diseños factoriales fraccionados

Generadores

$$D = AB$$

$$E = AC$$

Estructuras de confusión en un diseño 2^{5-2}

efecto de A = efecto de A + efecto de BD+ efecto de CE

efecto de B = efecto de B + efecto de AD+ efecto de CDE

efecto de C = efecto de C + efecto de AE+ efecto de BDE

efecto de D = efecto de D + efecto de AB+ efecto de BCE

efecto de E = efecto de E + efecto de AC+ efecto de BCD

Cuando un diseño se fracciona mucho no hay confianza en la estimación de los efectos principales

Resolución de los diseños factoriales fraccionados

 2^6

Factorial completo

Permite estimar todos los efectos principales y sus interacciones sin confusiones.

 2^{6-1}_{VI}

Resolución V o mayor

Los efectos principales y las interacciones dobles se confunden con interacciones triples o de orden mayor.

 2^{6-2}_{IV}

Resolución IV

Los efectos principales no se confunden entre ellos ni con interacciones dobles, pero si estas entre si.

 2^{6-3}_{III}

Resolución III

Los efectos principales no se confunden entre ellos pero hay efectos principales que se confunden con interacciones dobles.

- Combined
- Mixture
- Response Surface
- Factorial

2-Level Factorial Design

Design for 2 to 21 factors where each factor is varied over 2 levels. Useful for estimating main effects and interactions. Fractional factorials can be resolution: Green = Res V or higher, Yellow = Res IV, and Red = Res III.

- 2-Level Factorial
- Min Run Res V
- Min Run Res IV
- Irregular Fraction
- General Factorial
- D-optimal
- Plackett-Burman
- Taguchi OA

		Number of Factors												
		2	3	4	5	6	7	8	9	10	11	12	13	14
Runs	4	2^2	2^{3-1}_{III}											
	8		2^3	2^{4-1}_{IV}	2^{5-2}_{III}	2^{6-3}_{III}	2^{7-4}_{III}							
	16			2^4	2^{5-1}_V	2^{6-2}_{IV}	2^{7-3}_{IV}	2^{8-4}_{IV}	2^{9-5}_{III}	2^{10-6}_{III}	2^{11-7}_{III}	2^{12-8}_{III}	2^{13-9}_{III}	2^{14-10}_{III}
	32				2^5	2^{6-1}_VI	2^{7-2}_{IV}	2^{8-3}_{IV}	2^{9-4}_{IV}	2^{10-5}_{IV}	2^{11-6}_{IV}	2^{12-7}_{IV}	2^{13-8}_{IV}	2^{14-9}_{IV}
	64					2^6	2^{7-1}_VII	2^{8-2}_V	2^{9-3}_{IV}	2^{10-4}_{IV}	2^{11-5}_{IV}	2^{12-6}_{IV}	2^{13-7}_{IV}	2^{14-8}_{IV}
	128						2^7	2^{8-1}_VIII	2^{9-2}_VI	2^{10-3}_V	2^{11-4}_V	2^{12-5}_{IV}	2^{13-6}_{IV}	2^{14-7}_{IV}
	256							2^8	2^{9-1}_IX	2^{10-2}_VI	2^{11-3}_VI	2^{12-4}_VI	2^{13-5}_V	2^{14-6}_V
	512								2^9	2^{10-1}_X	2^{11-2}_VII	2^{12-3}_VI	2^{13-4}_VI	2^{14-5}_VI

Diseños factoriales fraccionados

Se busca la máxima resolución posible con un n° razonable de corridas experimentales y de gasto de recursos

$$2^{6-3}_{III}$$

Resolución III: no recomendados. En algunos casos es arriesgado suponer de antemano que ninguna interacción doble está activa. Se usan cuando hay muchos factores o cuando cada corrida es demasiado cara.

$$2^{6-2}_{IV}$$

Resolución IV: para $5 \leq k \leq 15$ existen diseños que no requieren más de 32 ensayos y proporcionan información de todos los efectos principales y de algunas de las interacciones dobles.

$$2^{6-1}_{VI}$$

Resolución V y VI: cuando se pueden hacer más ensayos.

Un caso especial: **PLACKETT-BURMAN**

- Es un tipo especial de diseño factorial fraccionario (altamente fraccionado con resolución III).
 - Es útil cuando existen múltiples factores.
- El diseño mínimo permite estudiar hasta 11 factores con 12 experimentos.
- Las interacciones dobles tienen un complejo alias con los términos principales.

DISEÑO DE PLACKETT-BURMAN

- Permite estudiar efectos principales, suponiendo que no existen interacciones entre los factores.
 - Si hay interacciones, el diseño puede fallar, es decir, puede considerar significativos efectos que no lo son y viceversa.
- En general, cuando hay interacciones importantes, el ajuste de PB es malo.
- ¿Hay alguna manera de superar este problema?



Contents lists available at SciVerse ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech



Significant factors selection in the chemical and enzymatic hydrolysis of lignocellulosic residues by a genetic algorithm analysis and comparison with the standard Plackett–Burman methodology

Pablo C. Giordano^{a,b}, Alejandro J. Beccaria^b, Héctor C. Goicoechea^{a,c,*}

^aLaboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, CC 242 (S3000ZAA) Santa Fe, Argentina

^bLaboratorio de Fermentaciones, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, CC 242 (S3000ZAA) Santa Fe, Argentina

^cConsejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Avda. Rivadavia 1917, CP C1033AAJ Buenos Aires, Argentina

ARTICLE INFO

Article history:

Received 9 June 2011

Received in revised form 8 August 2011

Accepted 1 September 2011

Available online 18 September 2011

Keywords:

Genetic algorithm

Plackett–Burman design

Hydrolysis

Carbohydrates

Significant factors

ABSTRACT

A comparison between the classic Plackett–Burman design (PB) ANOVA analysis and a genetic algorithm (GA) approach to identify significant factors have been carried out. This comparison was made by applying both analyses to data obtained from the experimental results when optimizing both chemical and enzymatic hydrolysis of three lignocellulosic feedstocks (corn and wheat bran, and pine sawdust) by a PB experimental design.

Depending on the kind of biomass and the hydrolysis being considered, different results were obtained. Interestingly, some interactions were found to be significant by the GA approach and allowed to identify significant factors, that otherwise, based only in the classic PB analysis, would have not been taken into account in a further optimization step. Improvements in the fitting of c.a. 80% were obtained when comparing the coefficient of determination (R^2) computed for both methods.

© 2011 Elsevier Ltd. All rights reserved.

DISEÑO DE PLACKETT-BURMAN

Factores	Experimentos
≤ 11	12
12-19	20
20-23	24
24-27	28
28-31	32

El n° de puntos del diseño N es múltiplo de cuatro.

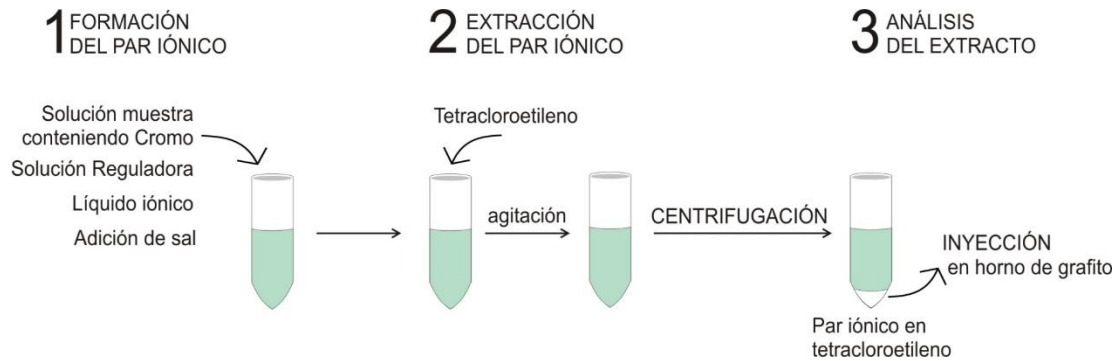
Se pueden estudiar hasta k factores en N experimentos donde $k = N - 1$.

DISEÑO DE PLACKETT-BURMAN

Std	Run	Block	Factor 1 A:Tiempo 1	Factor 2 B:Temp 1	Factor 3 C:Temp 2	Factor 4 D:Tiempo 2	Factor 5 E:Temp 3	Factor 6 F:Agitacion	Factor 7 G:Centrif.	Factor 8 H:Lavado	Factor 9 J:Dummy 1	Factor 10 K:Dummy 2	Factor 11 L:Dummy 3	Response 1 Respuesta
10	1	Block 1	-1.00	1.00	1.00	1.00	-1.00	F1	-1.00	H2	-1.00	1.00	1.00	
	9	Block 1	1.00	1.00	1.00	-1.00	-1.00	F1	1.00	H1	1.00	1.00	-1.00	
	4	Block 1	-1.00	1.00	-1.00	1.00	1.00	F1	1.00	H2	1.00	-1.00	-1.00	
	12	Block 1	-1.00	-1.00	-1.00	-1.00	-1.00	F1	-1.00	H1	-1.00	-1.00	-1.00	
	8	Block 1	1.00	1.00	-1.00	-1.00	-1.00	F2	-1.00	H2	1.00	-1.00	1.00	
	11	Block 1	1.00	-1.00	1.00	1.00	1.00	F1	-1.00	H1	1.00	-1.00	1.00	
	1	Block 1	1.00	1.00	-1.00	1.00	1.00	F2	-1.00	H1	-1.00	1.00	-1.00	
	3	Block 1	1.00	-1.00	1.00	1.00	-1.00	F2	1.00	H2	-1.00	-1.00	-1.00	
	2	Block 1	-1.00	1.00	1.00	-1.00	1.00	F2	1.00	H1	-1.00	-1.00	1.00	
	5	Block 1	-1.00	-1.00	1.00	-1.00	1.00	F2	-1.00	H2	1.00	1.00	-1.00	
	7	Block 1	1.00	-1.00	-1.00	-1.00	1.00	F1	1.00	H2	-1.00	1.00	1.00	
	6	Block 1	-1.00	-1.00	-1.00	1.00	-1.00	F2	1.00	H1	1.00	1.00	1.00	

Cuando hay menos factores que el máximo a estudiar (por ejemplo 8 factores y doce experimentos (11 permitidos para el estudio):

Variables dummy (se usan para estimar el error)



Se evaluará el efecto de **9 factores** sobre la cantidad de cromo (VI) extraído

Factor 1: concentración de la SR

Factor 2: volumen de líquido iónico (LI)

Factor 3: adición de sal

Factor 4: tiempo de formación del par iónico

Factor 5: volumen de solvente de extracción

Factor 6: tiempo de extracción

Factor 7: tipo de agitación

Factor 8: velocidad de centrifugación

Factor 9: tiempo de centrifugación

- o **SI**: se desea evaluar si 9 factores son significativos para una respuesta.

¿Qué diseño emplearía?	Experimentos	Bloques posibles
Factorial completo	512	1, 2, 4 y 8
Factorial fraccionado 2^{9-1} RIX	256	1, 2, 4 y 8
Factorial fraccionado 2^{9-2} RVI	128	1, 2, 4 y 8
Factorial fraccionado 2^{9-3} RIV	64	1, 2, 4 y 8
Factorial fraccionado 2^{9-4} RIV	32	1, 2, 4 y 8
Factorial fraccionado 2^{9-5} RIII	16	1, 2 y 4
Plackett-Burman	12	1

buenas alternativas

El diseño factorial fraccionado con RIII tiene estructuras de confusión más simples que un PB con RIII

Confusiones en un diseño Factorial Fraccionado 2^{9-5} con RIII

efecto de A = efecto de **A** + efecto de **FJ** + efecto de interacciones triples

efecto de B = efecto de **B** + efecto de **GJ** + efecto de interacciones triples

efecto de C = efecto de **C** + efecto de **HJ** + efecto de interacciones triples

efecto de D = efecto de **D** + efecto de **EJ** + efecto de interacciones triples

efecto de E = efecto de **E** + efecto de **DJ** + efecto de interacciones triples

efecto de F = efecto de **F** + efecto de **AJ** + efecto de interacciones triples

efecto de G = efecto de **G** + efecto de **BJ** + efecto de interacciones triples

efecto de H = efecto de **F** + efecto de **CJ** + efecto de interacciones triples

efecto de J = efecto de **J** + efecto de interacciones **AF + BG + CH + DE**

En base a las estructuras de confusión puedo ubicar los factores en las columnas adecuadas
Si el **factor A** tiene como estructura de confusión a la **interacción FJ**,
es conveniente colocar en las columnas F y J factores que no interaccionen con A

Diseño de Plackett-Burman para el problema

Experimento	Conc. SR (mM)	Volumen (μL)	Tiempo (min)	Respuesta
1	-1	-1	-1	72
2	+1	-1	-1	51
3	+1	-1	+1	50
4	-1	+1	+1	63
5	+1	+1	+1	65
6	-1	+1	+1	90
7	+1	+1	-1	60
8	+1	-1	+1	72
9	-1	+1	-1	82
10	+1	+1	-1	70
11	-1	-1	+1	57
12	-1	-1	-1	59

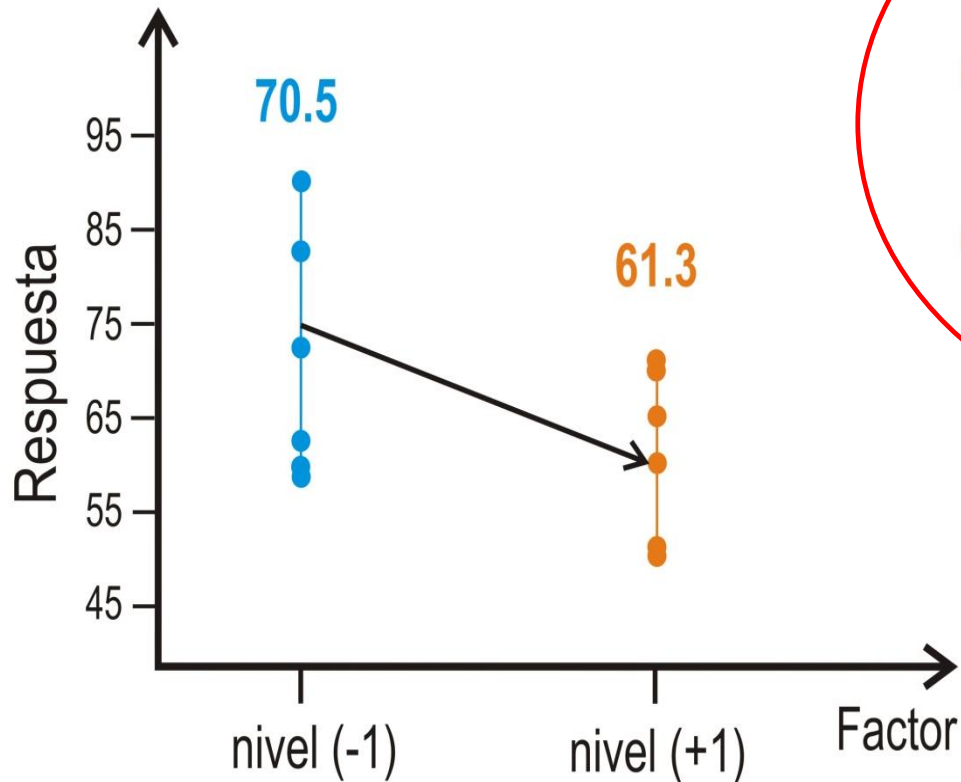
...

Estimación de efectos

Experimento	Conc. SR (mM)	Volumen (μL)	Tiempo (min)	Respuesta
1	-1	-1	-1	72
2	+1	-1	-1	51
3	+1	-1	+1	50
4	-1	+1	+1	63
5	+1	+1	+1	65
6	-1	+1	+1	90
7	+1	+1	-1	60
8	+1	-1	+1	72
9	-1	+1	-1	82
10	+1	+1	-1	70
11	-1	-1	+1	57
12	-1	-1	-1	59
Promedio nivel (+1)	61.3	71.7	66.2	Media general 65.9
Promedio nivel (-1)	70.5	60.2	65.7	
Efecto	-9.2	11.5	0.5	

Estimación de efectos

FACTOR: Concentración de la SR

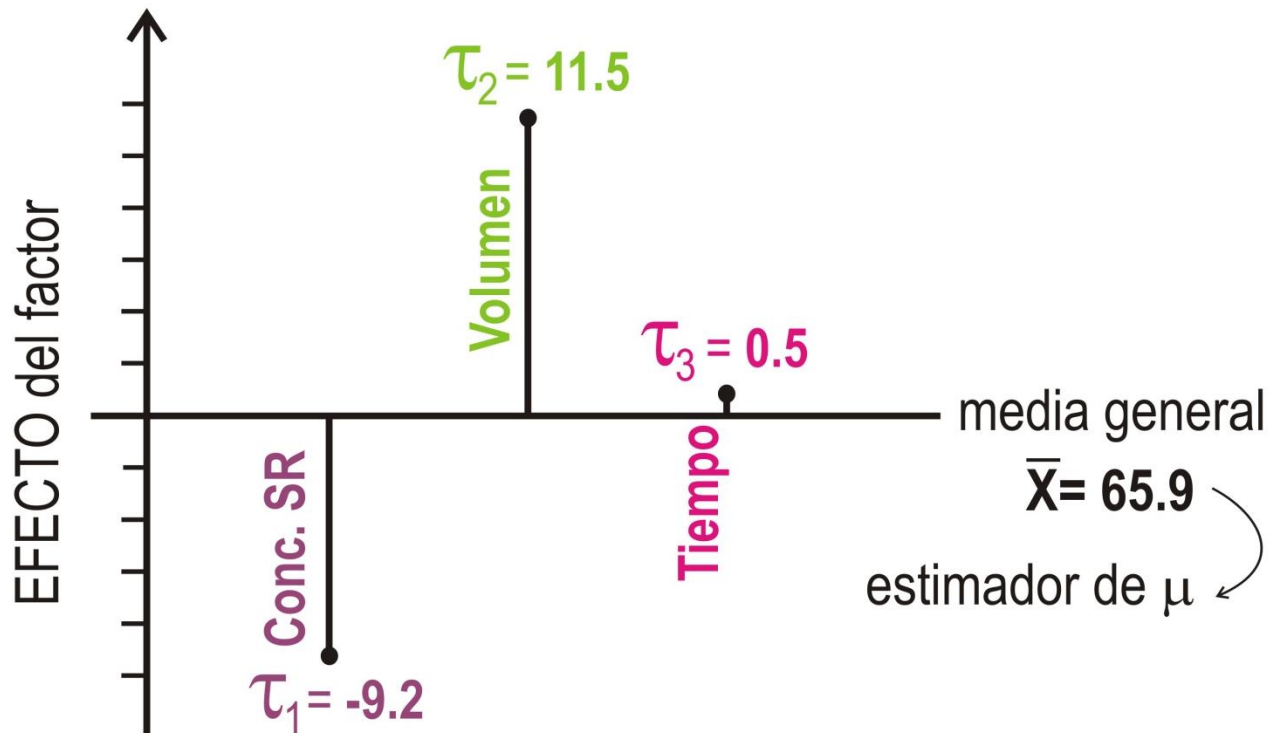


Se puede observar:

- la variabilidad en la respuesta para cada nivel del factor
- la tendencia en la respuesta: positiva o negativa

Estimación de efectos

Construcción de un modelo matemático que explique la variación



Resultados

¿Qué factor es significativo?

Factor	Efecto
Conc. SR (mM)	-9.2
Volumen (μ L)	11.5
Tiempo (min)	0.5

Pregunta:

Los efectos ¿son estadísticamente significativos?

La respuesta se puede obtener de dos maneras:

- Gráfica
- ANOVA

Evaluación de los efectos

Opciones gráficas

Se decide cuáles de los efectos principales, interacciones dobles y triples se pueden enviar al error. La SC_{error} en un ANOVA posterior contendrá los efectos “excluidos”.

Gráfica de Pareto

Gráfica de Probabilidad Normal (Daniel)

Gráfica de Probabilidad Media Normal (Daniel)

Opciones gráficas, más ANOVA

Opciones gráficas



Design-Expert® Software
Viscosidad

- ▲ Error from replicates
- A: Ingrediente A
- B: Ingrediente B
- C: Ingrediente C
- Positive Effects
- Negative Effects

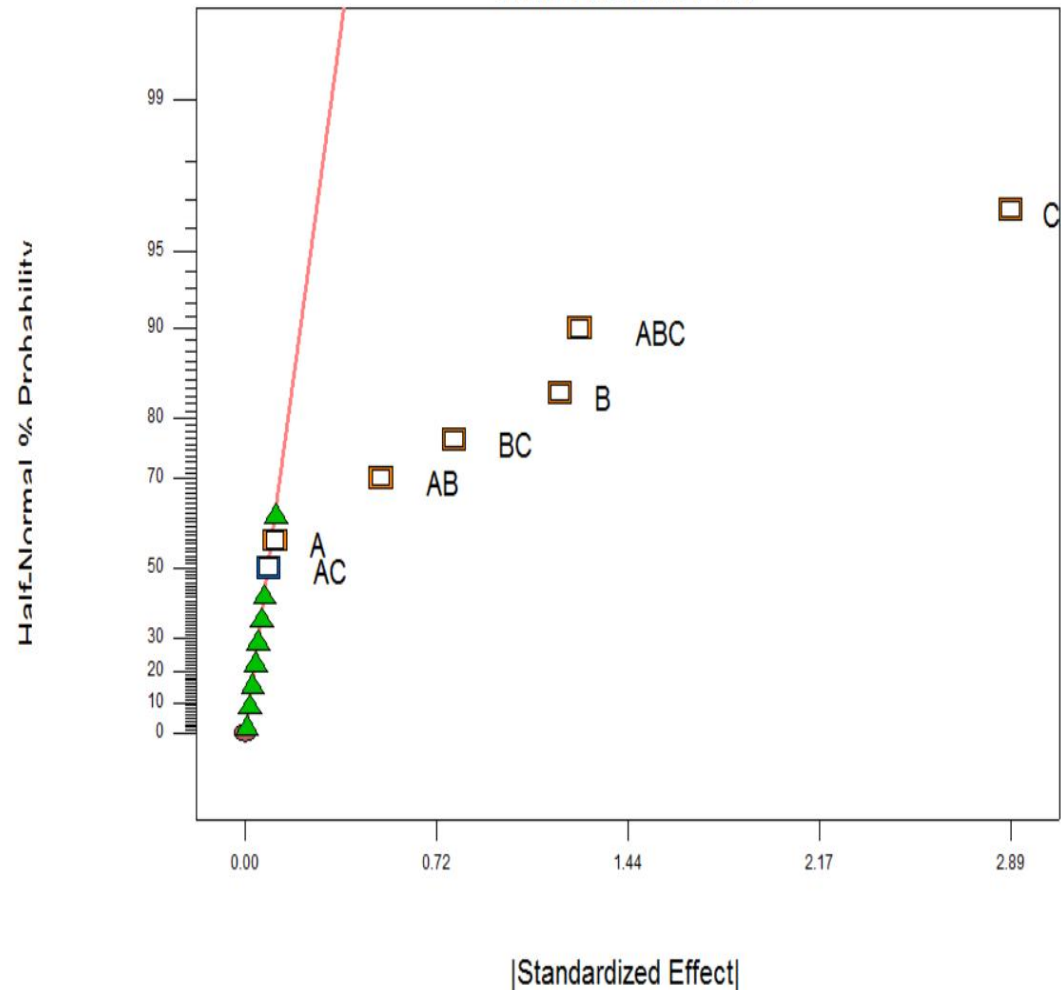
Effects Tool

- Half-Normal
- Normal Plot
- Pareto Chart
- Effects List

Clear Selection

Recalculate

Half-Normal Plot



Opciones gráficas

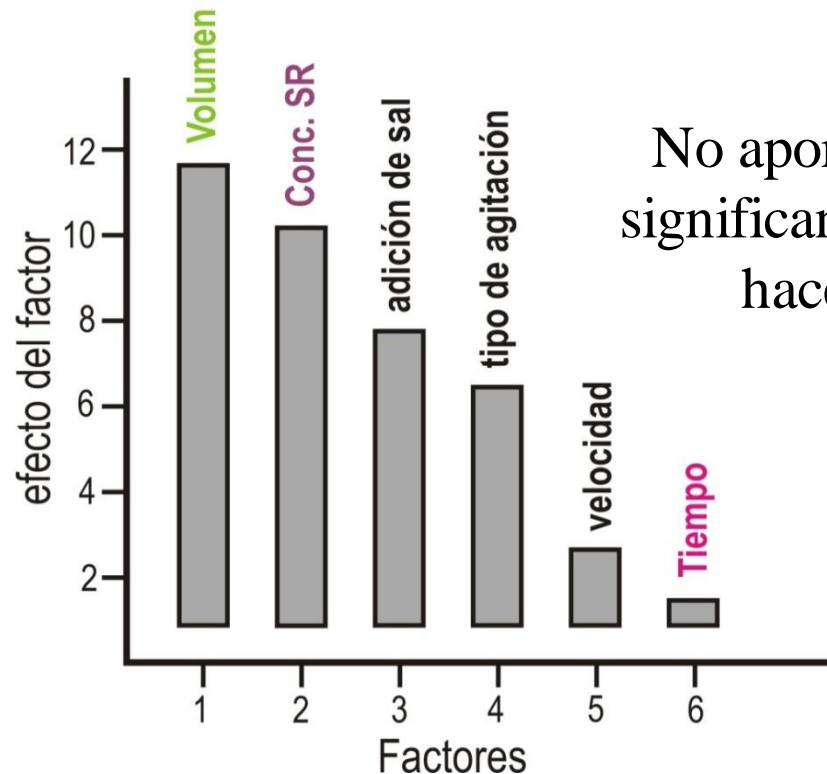
Diagrama de Pareto

Gráfico que permite discriminar entre las causas más importantes de un problema y las que afectan menos.

Se basa en la idea de que unos pocos factores son los responsables de los cambios significativos en la respuesta.

Opciones gráficas - Pareto

Pareto permite tener una idea visual rápida sobre que factor es más influyente.

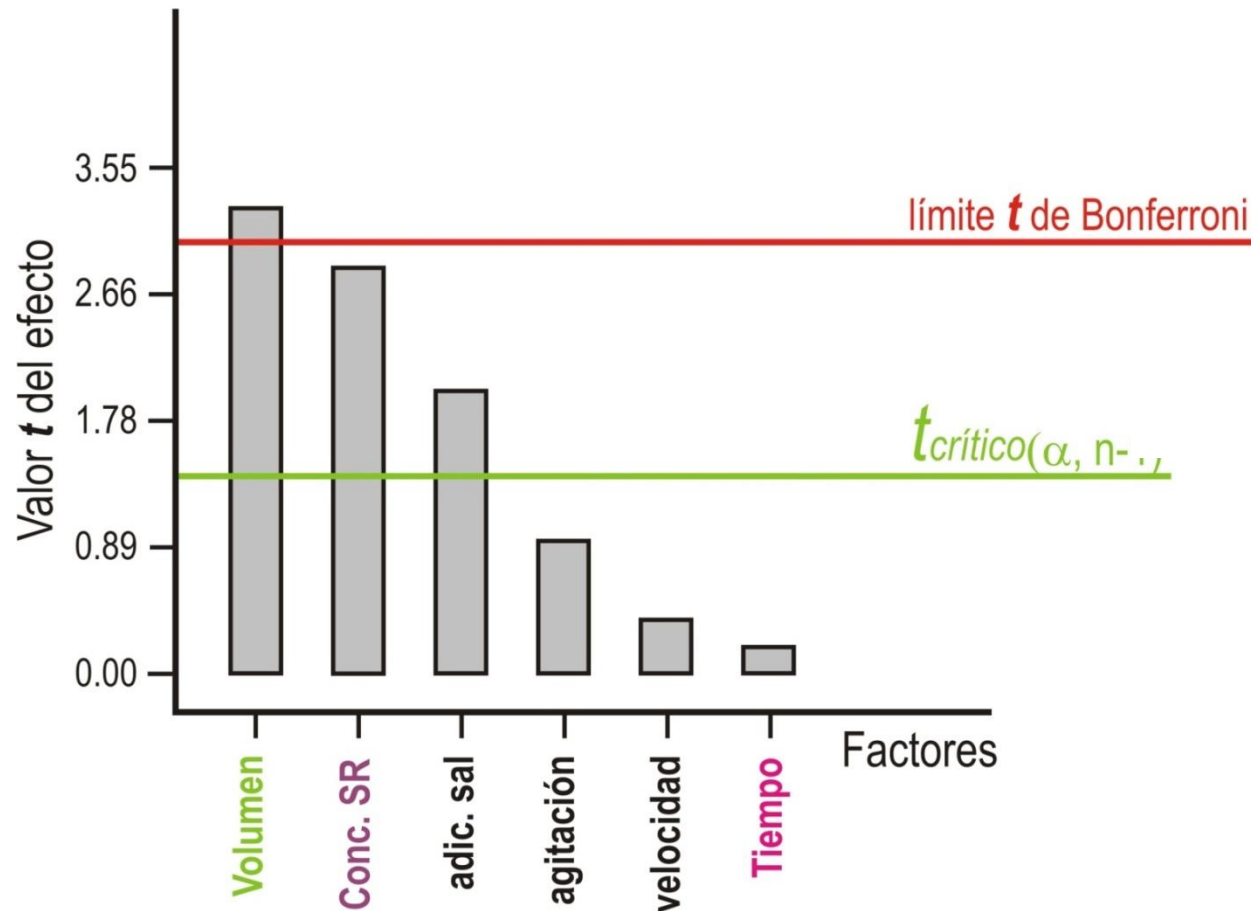


No aporta información sobre su significancia, es decir que hay que hacer un test estadístico.

Opciones gráficas

Diagrama de Pareto – Método de Bonferroni

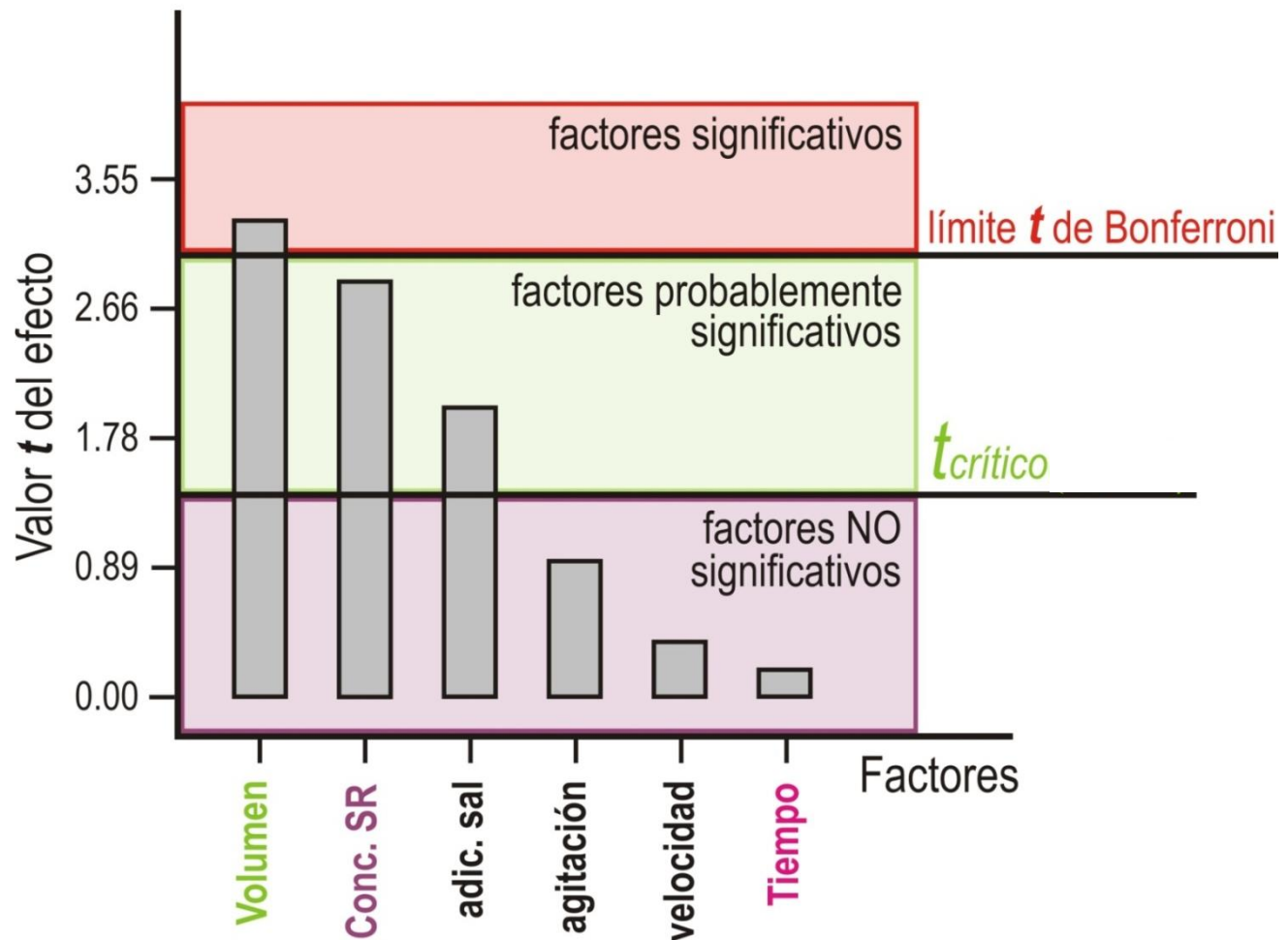
Representación gráfica de los valores t de los efectos



Opciones gráficas

Diagrama de Pareto – Método de Bonferroni

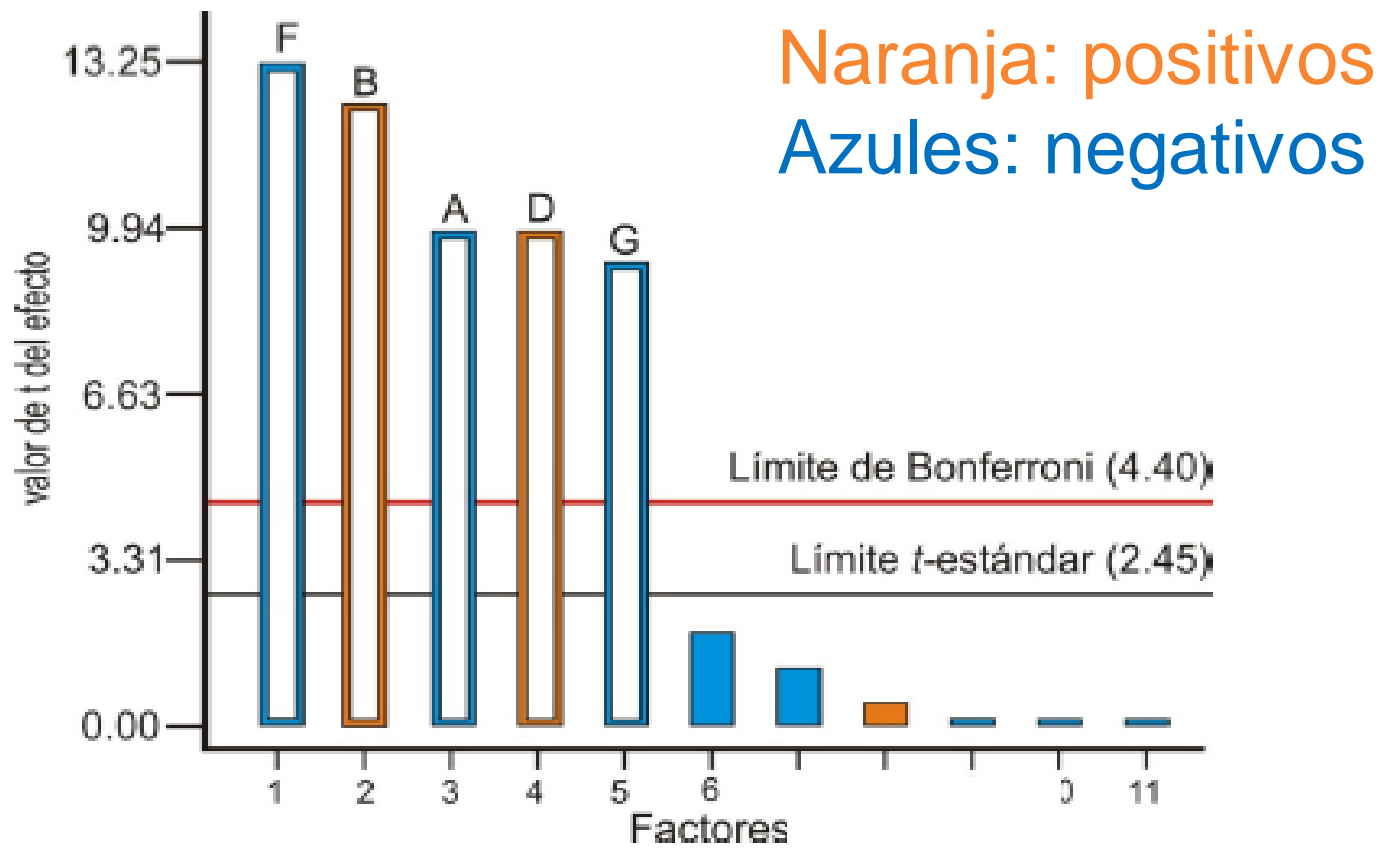
Representación gráfica de los valores t de los efectos



Modelo

Respuesta = media general + efectos principales de los factores + efectos de interacción + error

$$x_{ij} = \mu + \tau_i + \gamma_i + \varepsilon_{ij}$$



Opciones gráficas

DIAGRAMA de PROBABILIDAD NORMAL

GRÁFICA
DE
DANIEL

Gráfico que permite evaluar si un conjunto de datos proviene o no de una **distribución normal**

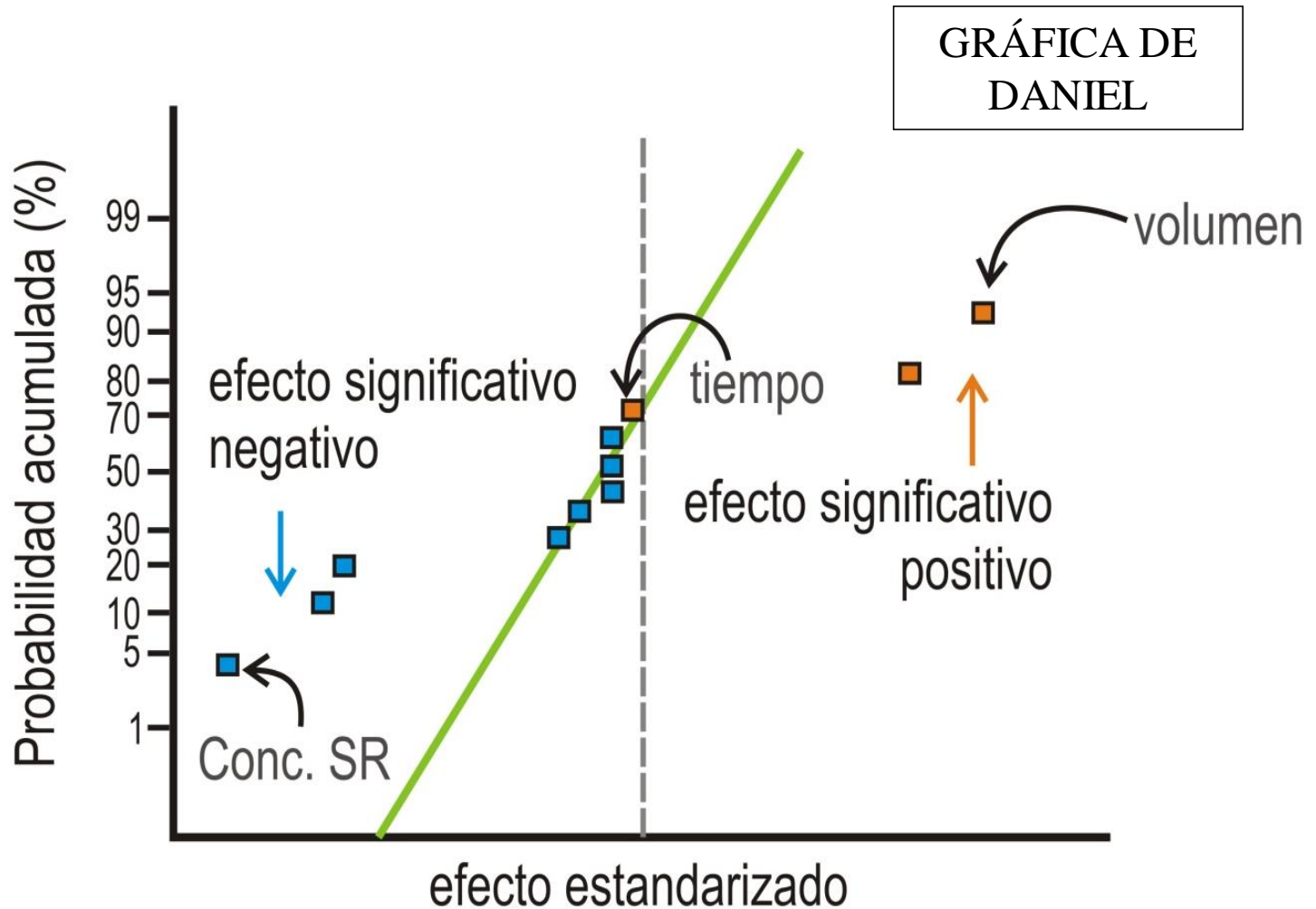
Se grafican los efectos experimentales vs. los efectos teóricos que se obtendrían de una distribución gaussiana (normal)

Los efectos de los factores que se alejan de la línea recta, no forman parte de una distribución normal, por lo tanto **son significativos.**

Importante: es una interpretación contraria a la que se hará para residuos en una parte posterior del estudio.

Opciones gráficas

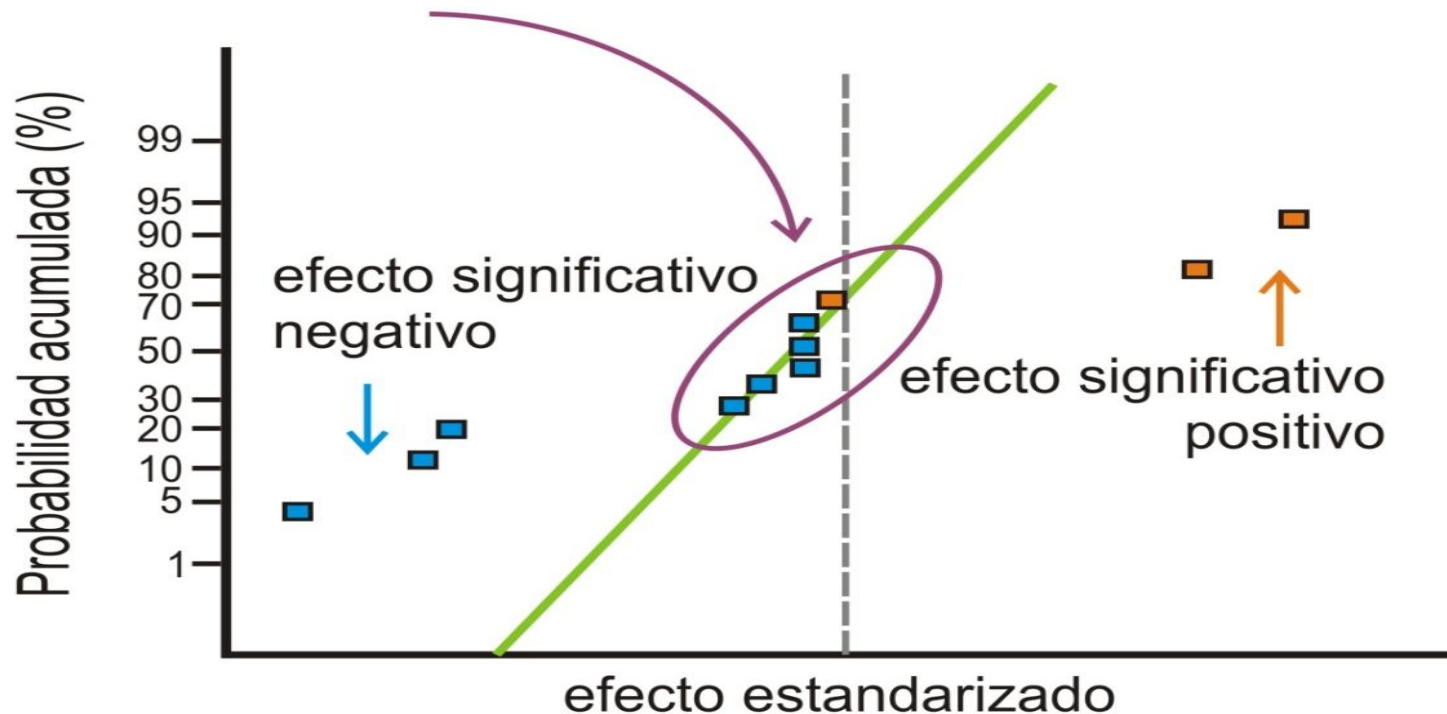
GRÁFICO DE EFECTOS EN PAPEL PROBABILÍSTICO NORMAL



Opciones gráficas

Test de Shapiro Wilk

Evalúa si los efectos de los factores sobre la línea recta, provienen de una distribución normal

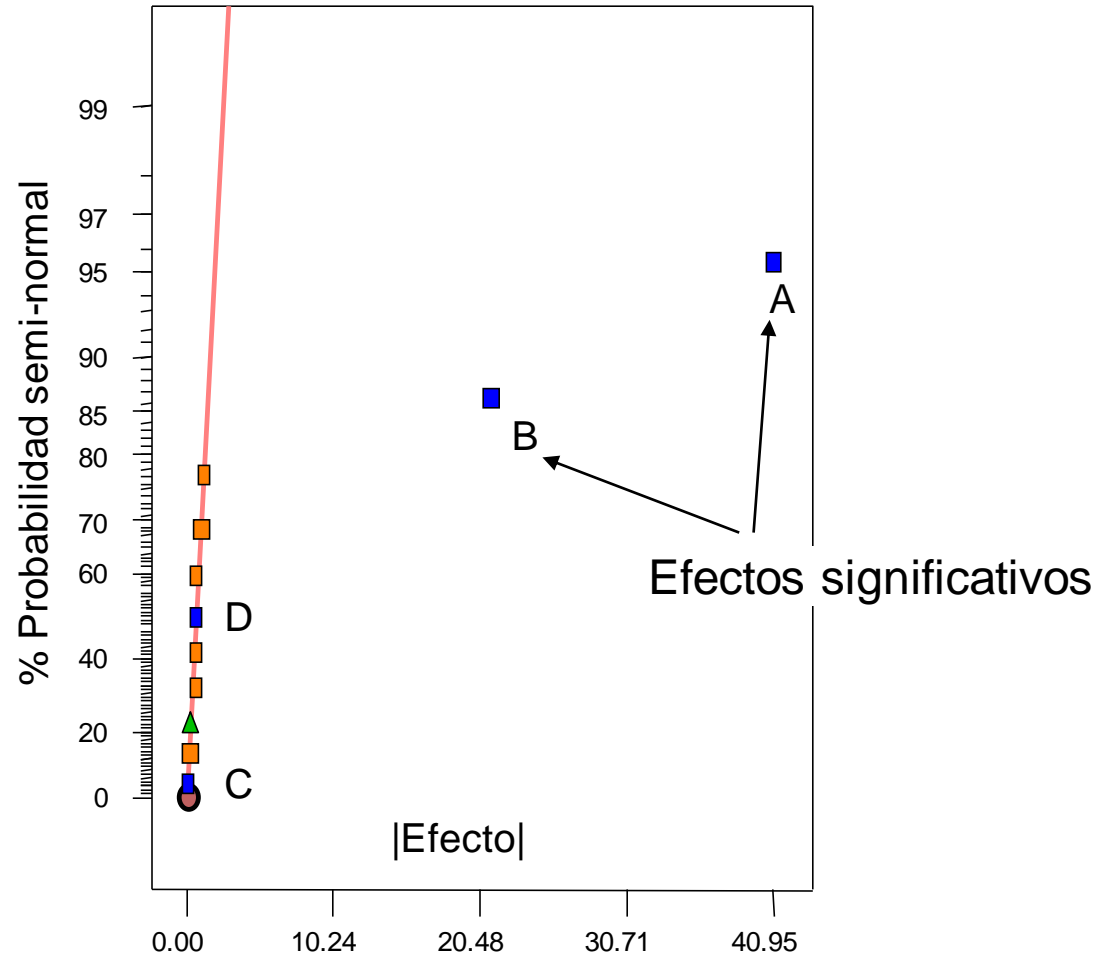


El test arroja un **valor-p** que indica la probabilidad de que la H_0 se cumpla (los factores provienen de una distribución normal y sus efectos no son significativos)

Opciones gráficas

Gráfico Semi-Normal

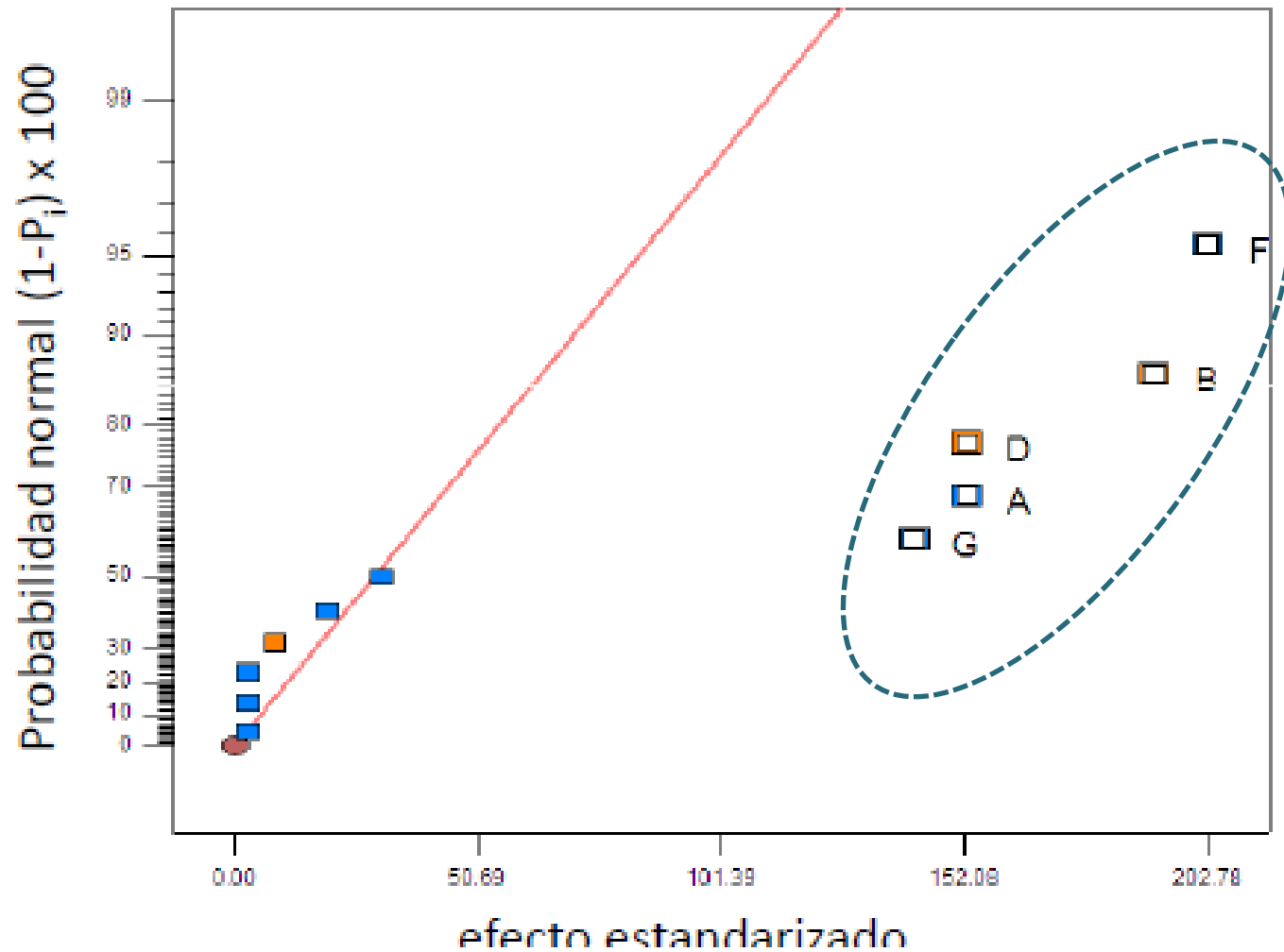
Papel
probabilístico
medio normal



Solo utiliza la parte positiva de la distribución normal estándar

Opciones gráficas

Gráfica de probabilidad media normal



ANOVA: Prueba de hipótesis

Hipótesis nula: los efectos de los factores provienen de una población con distribución normal (media $\mu=0$ y varianza σ^2) y, por lo tanto, no hay variación de la respuesta por el cambio del nivel del factor: **los factores no son significativos**

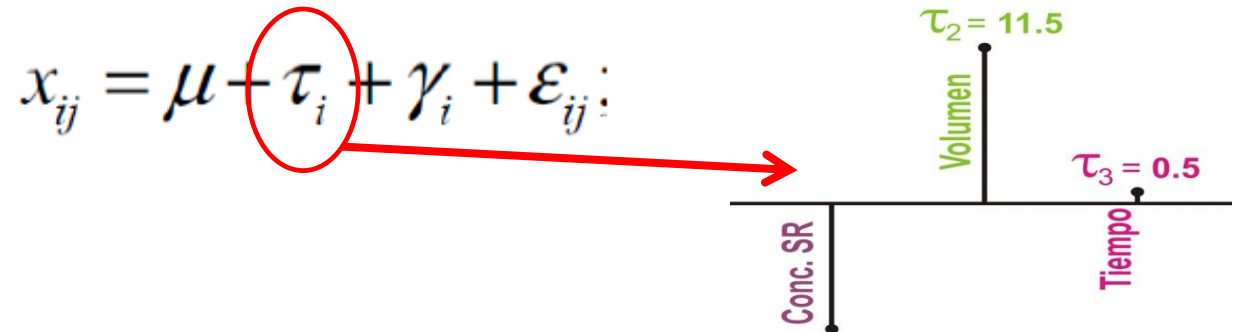
MÉTODOS ESTADÍSTICOS PARAMÉTRICOS (ANOVA)

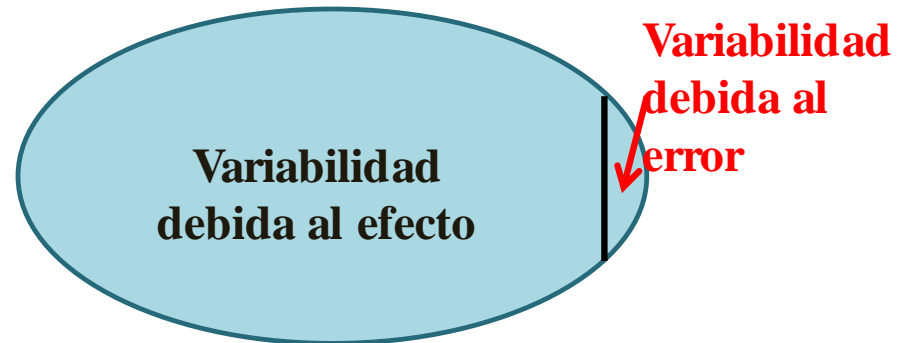
Permite responder a la **hipótesis** inicial con un **nivel de confianza**

→ **Cuál es la probabilidad de que H_0 es verdadera?**

Modelo de efectos: ANOVA

Respuesta = media general + efectos principales de los factores + efectos de interacción + error

$$x_{ij} = \mu + \tau_i + \gamma_i + \varepsilon_{ij}$$




El objetivo es:

- Seleccionar aquellos efectos de factores o interacciones que sean significativos y que formarán parte del modelo, explicando el comportamiento de la respuesta y ...
- Aquellos que sean no significativos que pueden utilizarse para conformar el error.

Ejemplo: uso de un diseño completo y réplica

En una fábrica de dispositivos electrónicos hay roturas y se piensa que hay **tres factores** (temperaturas de la etapa de procesamiento) responsables del problema.

Para determinar si estos factores son significativos y posteriormente encontrar una combinación de niveles donde la cantidad de **piezas rotas sea mínima**, se decide correr un diseño factorial 2^3 con dos réplicas, trabajando con una respuesta que es la proporción de piezas rotas.

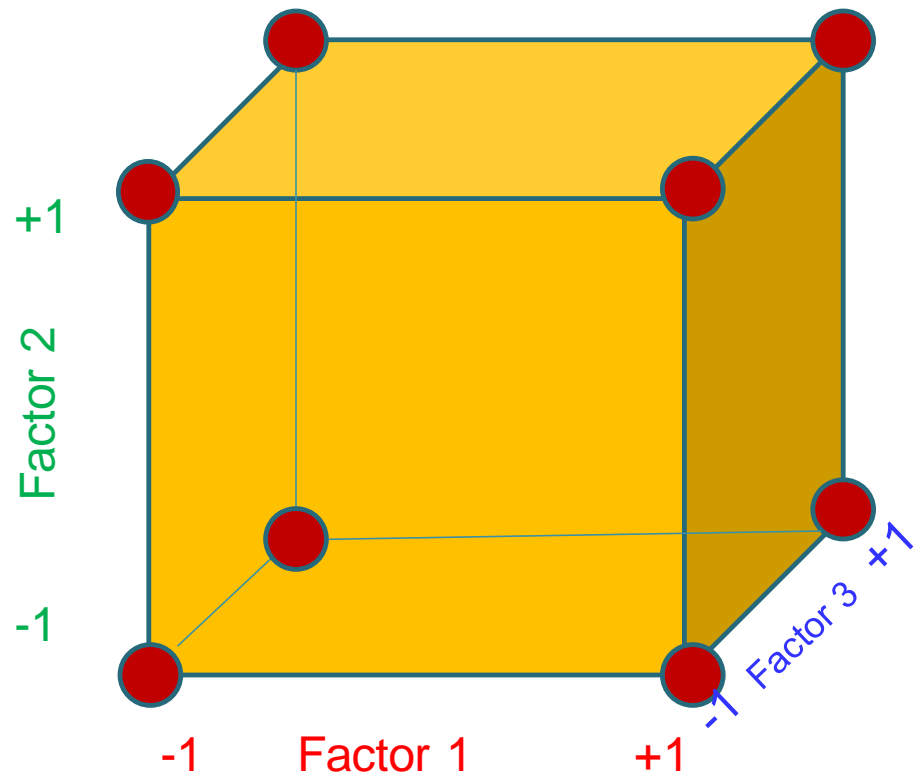
Diseño factorial 2^3

T1: Temperatura de grabado (-3 a -1 °C)

T2: Temperatura de piraña (60 a 98 °C)

T3: Temperatura de agua (20 a 78 °C)

Factor 1 A:T1 °C	Factor 2 B:T2 °C	Factor 3 C:T3 °C	Response 1 Piezas rotas
-1.00	-1.00	-1.00	40
-1.00	-1.00	-1.00	32
1.00	-1.00	-1.00	12
1.00	-1.00	-1.00	8
-1.00	1.00	-1.00	36
-1.00	1.00	-1.00	28
1.00	1.00	-1.00	0
1.00	1.00	-1.00	0
-1.00	-1.00	1.00	20
-1.00	-1.00	1.00	20
1.00	-1.00	1.00	0
1.00	-1.00	1.00	16
-1.00	1.00	1.00	16
-1.00	1.00	1.00	8
1.00	1.00	1.00	4
1.00	1.00	1.00	4



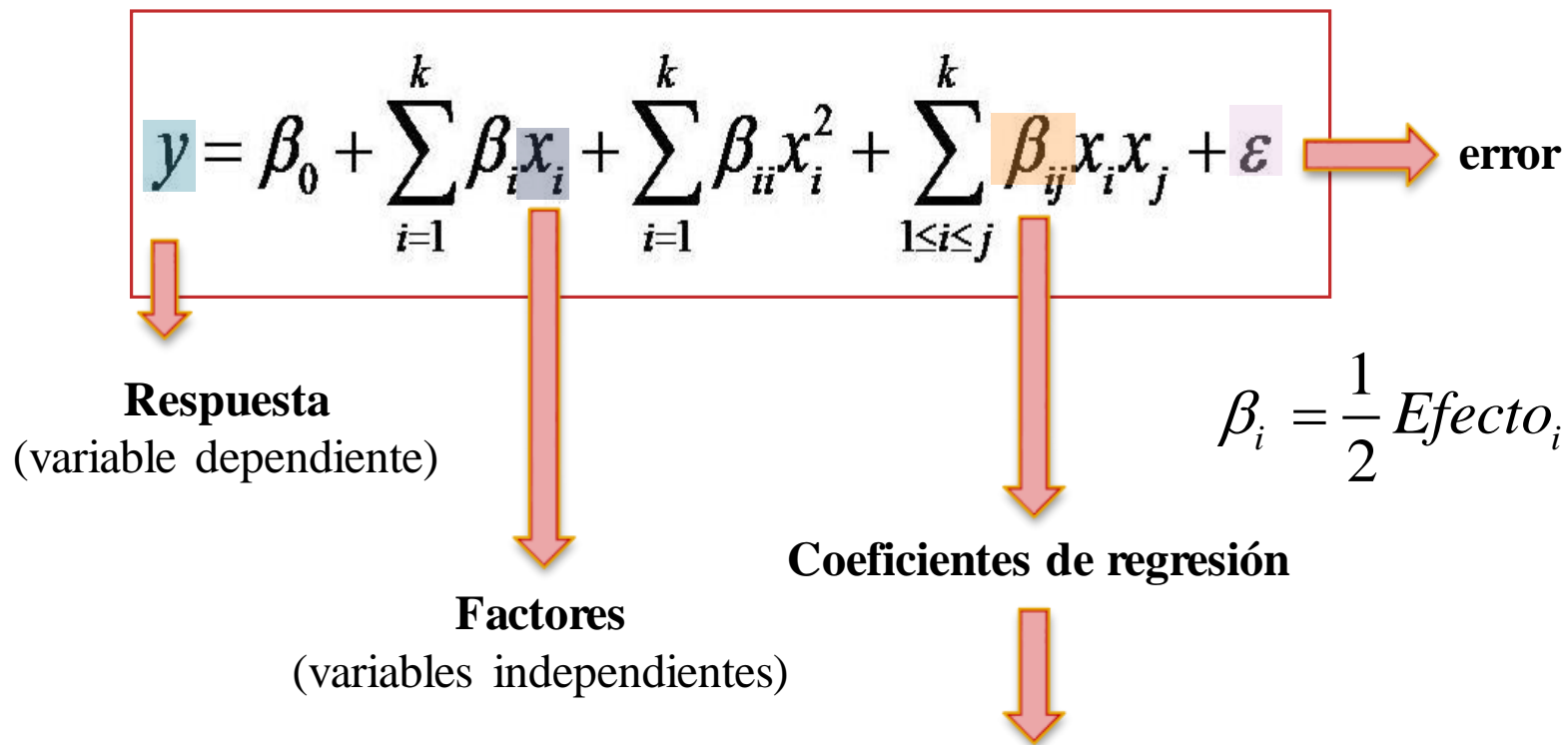
Forma de operar:

- 1- Hacer un ANOVA con todos los efectos.
- 2- Hacer un ANOVA que sólo incluya los términos significativos, mandando al error los efectos que claramente no son significativos.
- 3- Revalorar los términos que estaban en situación dudosa

Construcción del Modelo

Regresión lineal múltiple

Ecuación de regresión



Método de los cuadrados mínimos

Construcción del Modelo

Coeficientes

- Importancia del efecto
- Cantidad de factores que intervienen en el comportamiento de la respuesta

Complejidad

- Tipo de relación que existe entre la respuesta y esos factores
- Uso que se le va a dar al modelo

Polinomios más
simples

Polinomios más
complejos

Construcción del Modelo

¿Para que construimos modelos matemáticos?

- Seleccionar factores significativos
- Evaluar interacciones entre los factores
- Explorar el comportamiento de la respuesta en un entorno experimental acotado
- Describir lo más exactamente posible el comportamiento de la respuesta
- Encontrar un óptimo para la respuesta

Polinomios simples

Modelos Lineales

Modelos con Interacciones

Modelos Cuadráticos

Modelos Cúbicos

Polinomios complejos

Construcción del Modelo

¿Cómo obtenemos el mejor modelo a partir de los datos experimentales recolectados con un diseño de experimentos?

1- Ajustar los distintos tipos de modelos posibles a los datos

- Lineal
- Lineal con Interacción
- Cuadrático

2- Para cada uno de los modelos calcular la significancia de la regresión (idoneidad del modelo), la falta de ajuste y el coeficiente de determinación (R^2_{aj})

Construcción del Modelo

Significancia de la Regresión $F_0 = \frac{CM_R}{CM_E} > F_{0.05, k, n-k-1}$

Falta de Ajuste $F_0 = \frac{CM_{LOF}}{CM_{EPuro}} < F_{0.05, m-2, n-m}$

3- Seleccionar el modelo con mayor idoneidad, menos falta de ajuste y mayor (R^2_{aj})

¿Cómo se calcula una varianza?

Analizar el cálculo de la desviación estándar

x_1

x_2

x_3

.

.

.

x_n

$$s = \frac{\sum_{i=1}^n (x_p - x_i)^2}{n - 1}$$

$$\text{Varianza} = s^2$$

x_{promedio}

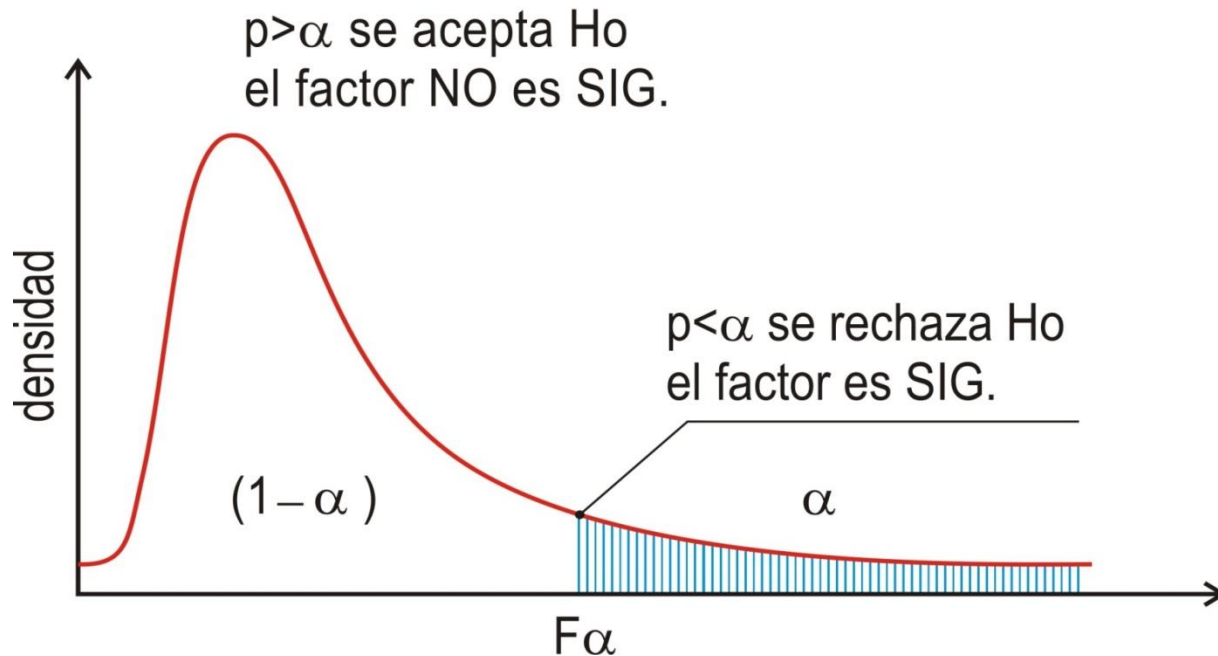
n

Test de hipótesis para comparar varianzas: prueba F

Hipótesis: Nula (H_0): $s_1^2 = s_2^2$

Alternativa (H_1): $s_1^2 > s_2^2$

Test estadístico F :



Comparación de dos varianzas muestrales

Se calcula el estadístico F

$$F_c = \frac{s_1^2}{s_2^2}$$

Siendo $s_1 > s_2$

$$F_{(n_1-1), (n_2-1), \alpha}$$

Si F calculado es **menor** al F tabulado, las series son homogéneas u homocedásticas

Si F calculado es **mayor** al F tabulado, las series son heterogéneas o heterocedásticas

Si la probabilidad correspondiente a ese F calculado es menor a 0.05, las series son heterogéneas o heterocedásticas

Distribución F Valores críticos de F para una prueba de una cola ($P = 0.05$)

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.660
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.803
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619	4.558
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.874
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.445
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218	3.150
9	5.117	4.256	3.863	3.633	3.482	3.374	3.291	3.228	3.179	3.137	3.073	3.006	2.936
10	4.965	4.103	3.708	3.478	3.326	3.217	3.133	3.072	3.020	2.978	2.913	2.845	2.774
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.646
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.544
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.459
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.388
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403	2.328
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.276
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.308	2.230
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.191
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.234	2.155
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.124

ν_1 = número de grados de libertad del numerador y ν_2 = número de grados de libertad del denominador.

Tabla 6.8 ANOVA completo para el ejemplo de obleas.

Efectos	SC	GL	CM	F_0	Valor-p
A: T_Grab	0.001521	1	0.001521	52.45	0.0001
B: T_Pira	0.000169	1	0.000169	5.83	0.0422
C: T_Agua	0.000289	1	0.000289	9.97	0.0135
AB	0.000001	1	0.000001	0.03	0.8573
AC	0.000361	1	0.000361	12.45	0.0078
BC	0.000001	1	0.000001	0.03	0.8573
ABC	0.000025	1	0.000025	0.86	0.3803
Error	0.000232	8	0.000029		
Total	0.002599	15			
$R^2 = 91.1$		$R^2_{aj} = 83.3$			

$$R^2 = \frac{(SC_T - SC_E)}{SC_T} \times 100 = \frac{SC_M}{SC_T} \times 100$$

$$R^2_{aj.} = \frac{(CM_T - CM_E)}{CM_T} \times 100$$

Cuando hay muchos términos en el modelo, se prefiere R^2_{aj} sobre R^2 (para no incrementar en forma artificial con cada término que se agrega). R^2_{aj} baja con cada término artificial que se agrega.

Se cumple: $0 \leq R^2_{aj} \leq R^2 \leq 1$

Se espera que sea al menos igual a 0.7

Tabla 6.9 El mejor ANOVA para el ejemplo de obleas.

Efectos	SC	GL	CM	F_0	Valor-p
A: T_Grab	0.001521	1	0.001521	64.60	0.0000
B: T_Pira	0.000169	1	0.000169	7.18	0.0214
C: T_Agua	0.000289	1	0.000289	12.27	0.0049
AC	0.000361	1	0.000361	15.33	0.0024
Error	0.000259	11	0.0000235		
Total	0.002599	15			
$R^2 = 90.0$		$R^2_{aj} = 86.4$			

$$R^2 = 91.1 \quad R^2_{aj} = 83.3$$

Calidad del ajuste

Verificación de los supuestos del ANOVA

Conclusiones

ANOVA

Cálculo de Contraste y Suma de cuadrados

Contraste

$$C = \sum_{i=1}^{2^k} c_i Y_i$$

coeficientes

Conc. SR (mM)	Volumen (μL)	Tiempo (min)	Respuesta
-1	-1	-1	72
+1	-1	-1	51
+1	-1	+1	50
-1	+1	+1	63
+1	+1	+1	65
-1	+1	+1	90
+1	+1	-1	60
+1	-1	+1	72
-1	+1	-1	82
+1	+1	-1	70
-1	-1	+1	57
-1	-1	-1	59
368	430	397	sumatoria (+1)
423	361	394	sumatoria (-1)
-55	69	3	contraste
252.08	396.75	0.75	suma de cuadrados

Suma de cuadrados

$$SC_F = \frac{\sum_{i=1}^{2^k} (c_i Y_i)^2}{n \sum_{i=1}^{2^k} (c_i)^2}$$

$$SC_F = \frac{(\text{contraste})^2}{\text{n}^\circ \text{ de experimentos}}$$

ANOVA

ANOVA - Cálculo de la suma de cuadrados totales

Conc. SR (mM)	Volumen (μL)	Tiempo (min)	Respuesta	$(X_{ij} - \bar{X})^2$
-1	-1	-1	72	31.01
+1	-1	-1	51	222.51
+1	-1	+1	50	253.34
-1	+1	+1	63	8.51
+1	+1	+1	65	0.84
-1	+1	+1	90	580.01
+1	+1	-1	60	35.01
+1	-1	+1	72	37.01
-1	+1	-1	82	258.67
+1	+1	-1	70	16.67
-1	-1	+1	57	79.51
-1	-1	-1	59	47.84

$$\bar{X} = 65.92$$

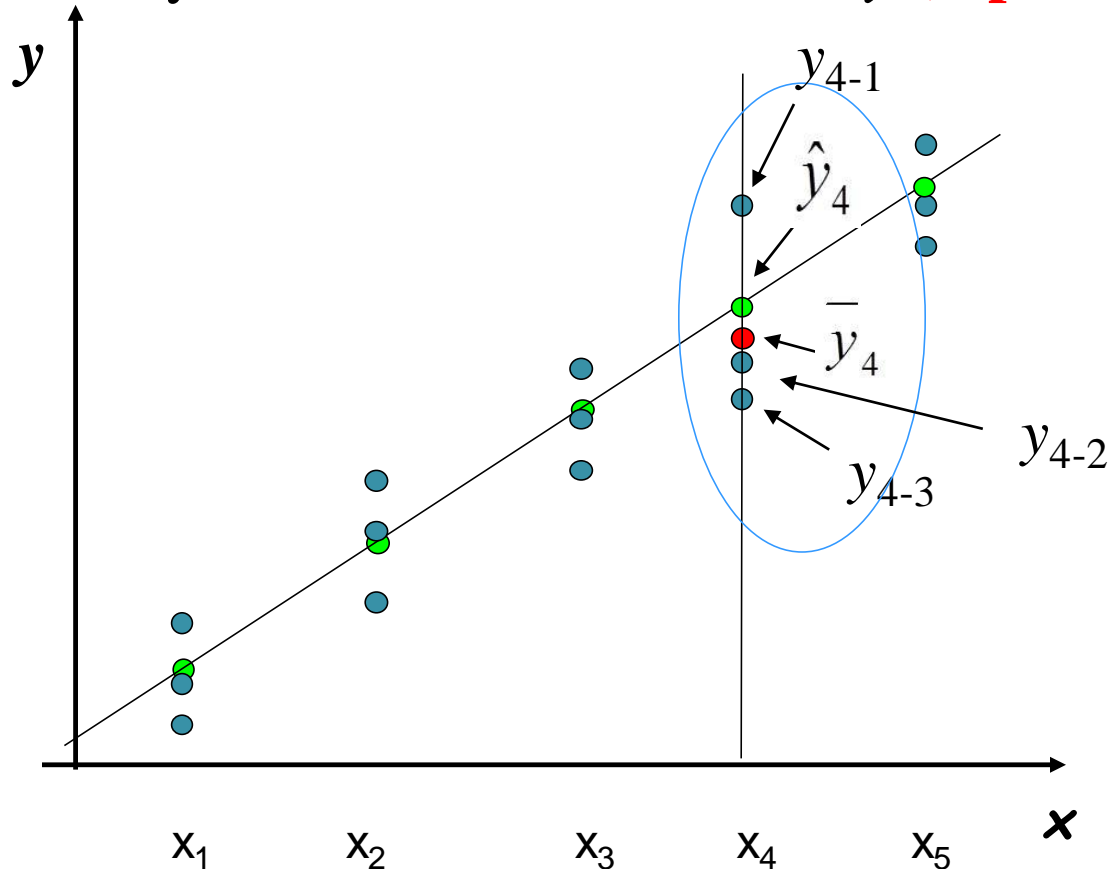
$$SC_T = \sum (x_i - \bar{x})^2 = 1576.9$$

Resumen del análisis

Parámetro	suma de cuadrados (SC)	grados de libertad (gl)	cuadrado medio (CM)	Prueba F	valor p
Modelo	1561.75	5	312.35	123.57	
Conc. SR (mM)	252.08	1	252.08	99.73	< 0.0001
Volumen (μL)	396.75	1	396.75	156.96	< 0.0001
Tiempo formación (min)	0.75	1	0.75	0.30	0.6376
Adición de sal (g/L)	252.08	1	252.08	99.73	< 0.0001
Volumen solvente (μL)	0.08	1	0.08	0.03	0.8745
Tiempo extracción (min)	444.08	1	444.08	175.68	< 0.0001
Tipo de agitación	216.75	1	216.75	85.95	< 0.0001
Veloc. centrifugación (rpm)	10.08	1	10.08	4.01	0.1778
Tpo centrifugación (rpm)	0.08	1	0.08	0.03	0.8745
Residuo	15.08	6	2.51		
Total	1576.92				

Calidad del ajuste del modelo de regresión lineal simple

Prueba de falta de ajuste: es necesario que para cada valor de x haya varias observaciones de y (**repeticiones**)



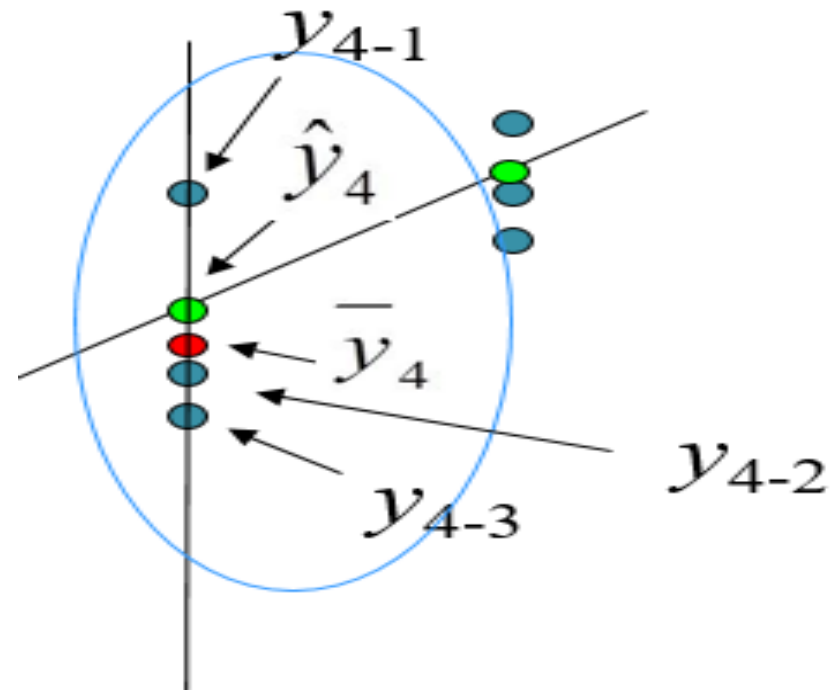
H_0 : el modelo se ajusta de manera adecuada a los datos

H_1 : el modelo no se ajusta en forma satisfactoria

$$SC_E = SC_{EP} + SC_{FA} \rightarrow SC_{FA} = SC_E - SC_{EP}$$

Para m niveles distintos de x y n_i repeticiones dentro de cada nivel:

$$SC_{EP} = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$



$$F_{cal} = \frac{SC_{FA}/(m-2)}{SC_{EP}/(N-m)} = \frac{CM_{FA}}{CM_{EP}}$$

Si $F_{cal} > F_{(\alpha, m-2, N-m)}$ o si $p < 0.05$

Se rechaza H_0

Verificación de los supuestos del ANOVA

Evaluación de los Modelos

- ¿Se cumplen los supuestos en los que se basa la regresión por cuadrados mínimos?

- **¿Cuáles son los supuestos?**

Normalidad Homocedasticidad Independencia

- ¿Provee una aproximación adecuada a la función real?

- Si el modelo no es adecuado y no está correctamente ajustado:
 - La selección de factores no será confiable.
 - La exploración y optimización del sistema no será adecuada.

Análisis de Residuos

Mucha utilidad en la comprobación de los supuestos

RESIDUO

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

**Diferencia entre la
respuesta
observada y la
predicha por el
modelo**

Experimento	Respuesta	Rta predicha	Residuo
1	72	70.52	1.48
2	51	52.82	-1.82
3	50	49.82	0.18
4	63	61.32	1.68
5	65	64.32	0.68
6	90	91.22	-1.22
7	60	60.62	-0.62
8	72	70.52	1.48
9	82	82.72	-0.72
10	70	69.82	0.18
11	57	58.32	-1.32
12	59	59.02	-0.02

Design-Expert® Software
Viscosidad

Diagnostics T... ✕

Diagnostics Influence

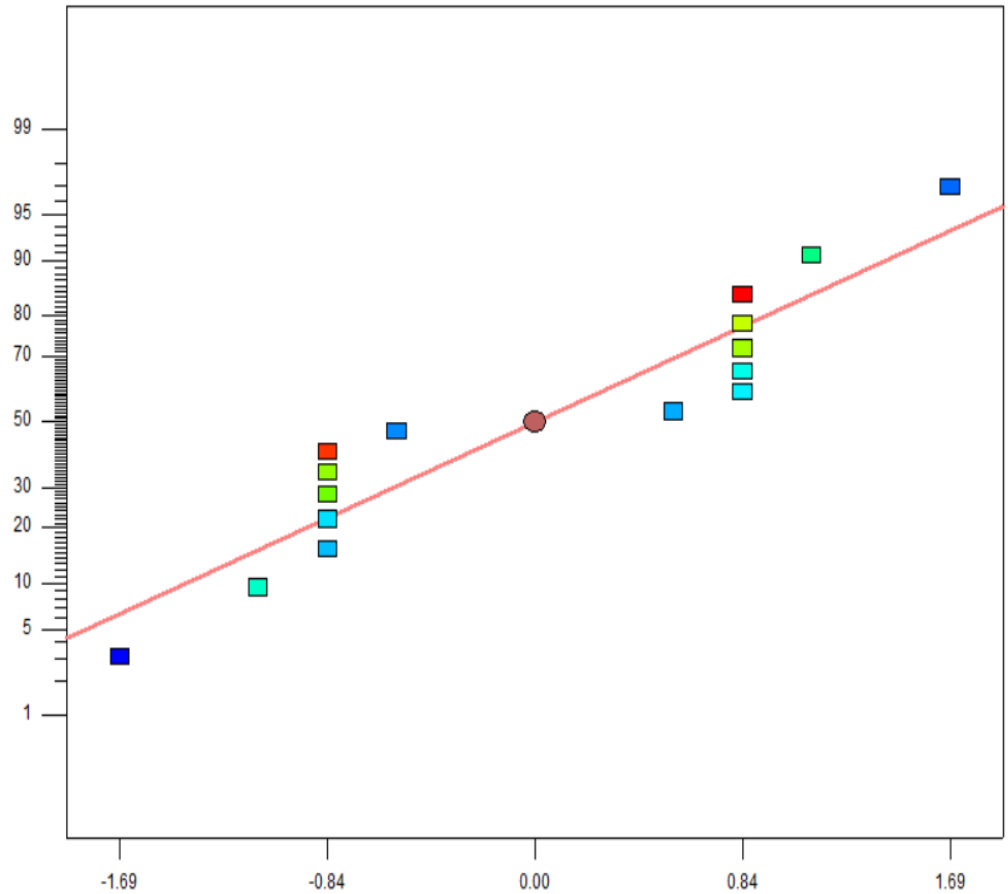
Studentized

Normal Plot
 e_j vs. Pred.
 e_j vs. Run
 Pred. vs. Actual
 Box Cox
 e_j vs. Factor

Color by Viscosidad ▾

Clear Points

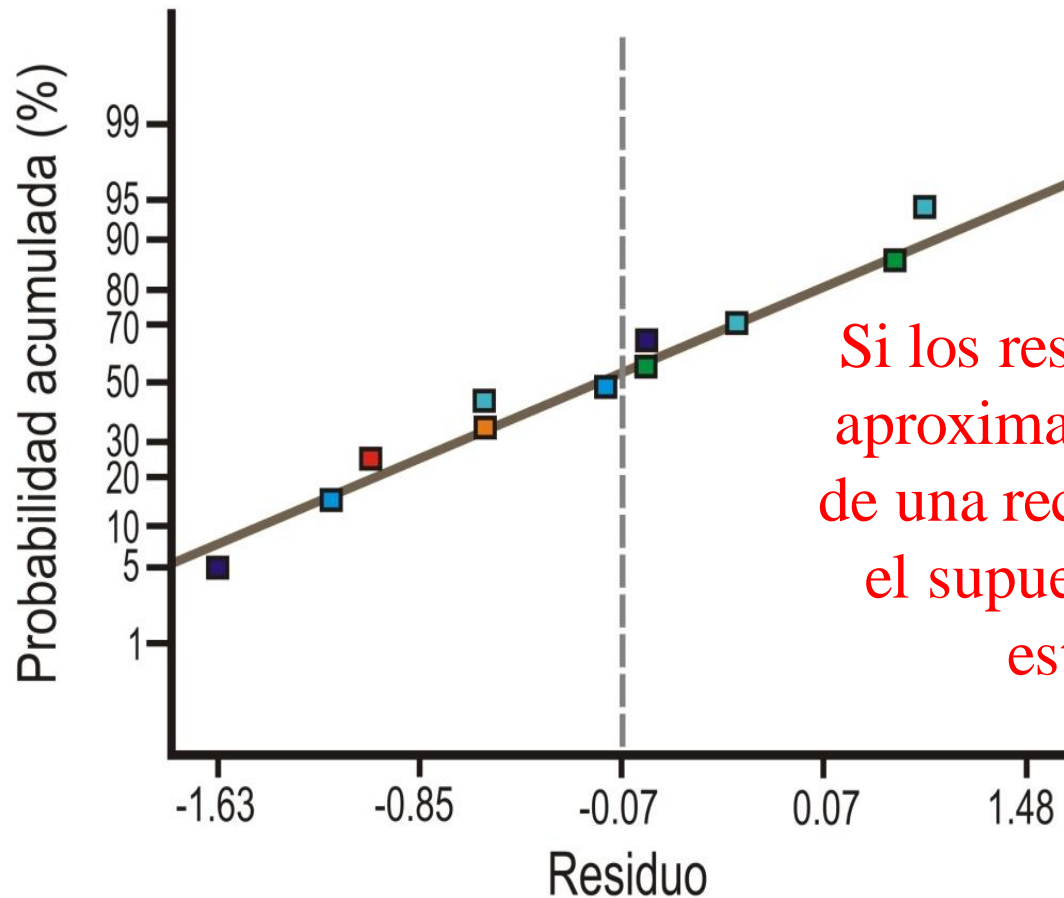
Normal Plot of Residuals



Internally Studentized Residuals

Normalidad de los residuos

Gráfico de probabilidad normal

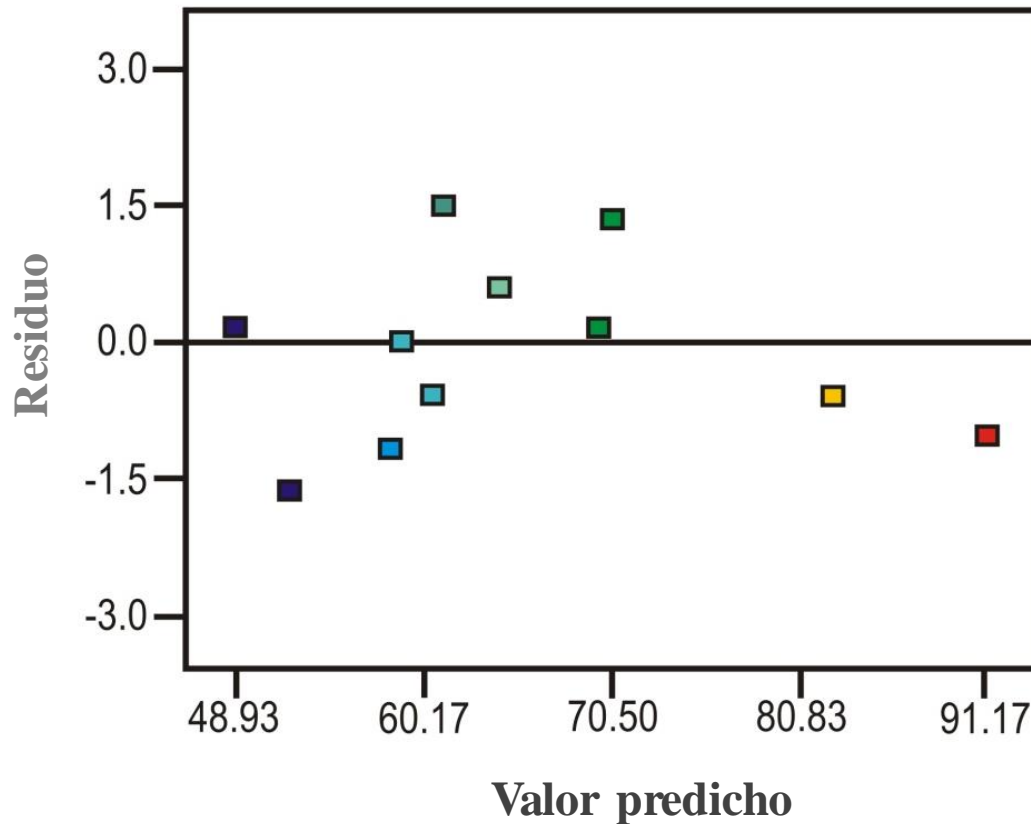


Si los residuos se acomodan aproximadamente a lo largo de una recta, se considera que el supuesto de normalidad está satisfecho.

Recordar que al analizar efectos le dábamos una interpretación diferente (transparencia N° 54)

Homocedasticidad de los residuos

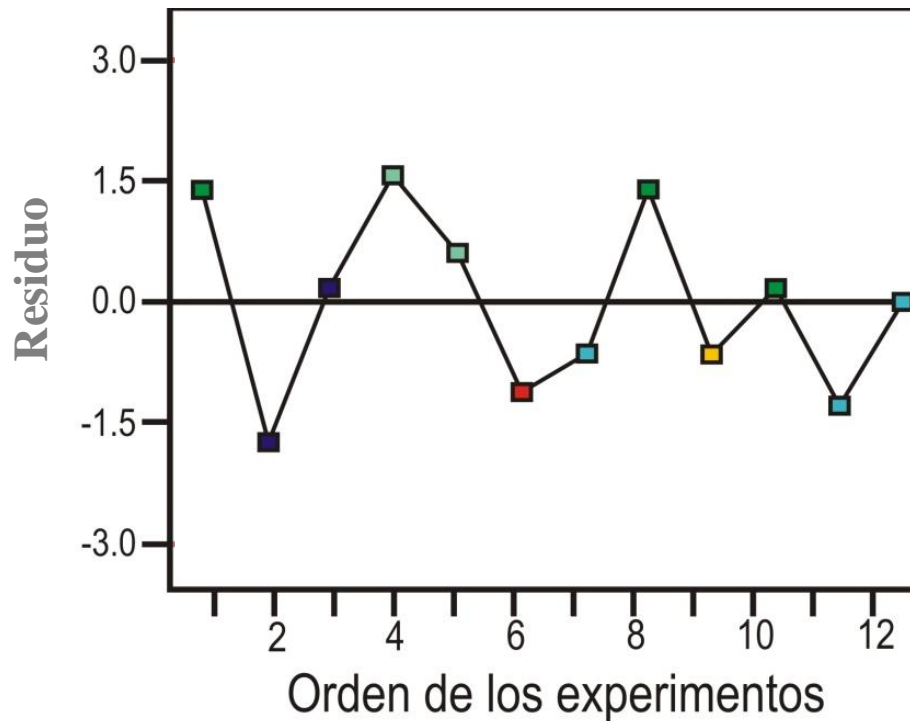
Gráfico de residuos vs. valor predicho



Si los residuos se distribuyen aleatoriamente, se considera que el supuesto de homocedasticidad está satisfecho.

Independencia de los residuos

Gráfico de residuos vs. orden del experimento



Si los residuos se distribuyen aleatoriamente, se considera que el supuesto de independencia está satisfecho.

Otros enfoques y usos: Escalado de residuos para la detección de *Outliers*

Residuo estandarizado

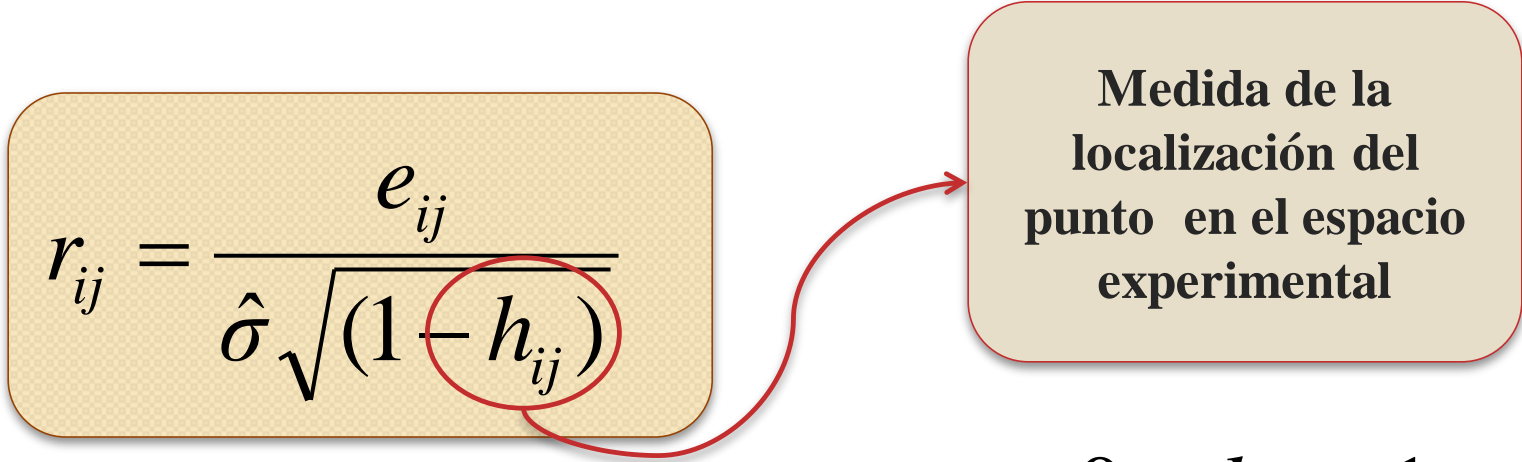
$$d_{ij} = \frac{e_{ij}}{\hat{\sigma}}$$

$$\hat{\sigma} = \sqrt{MC_E}$$

Útil para detectar atípicos (*outlier*):
residuo estandarizado mucho más
grande que los demás

Otros enfoques y usos: Escalado de residuos para la detección de *Outliers*

Residuo **estudentizado**

$$r_{ij} = \frac{e_{ij}}{\hat{\sigma} \sqrt{(1 - h_{ij})}}$$


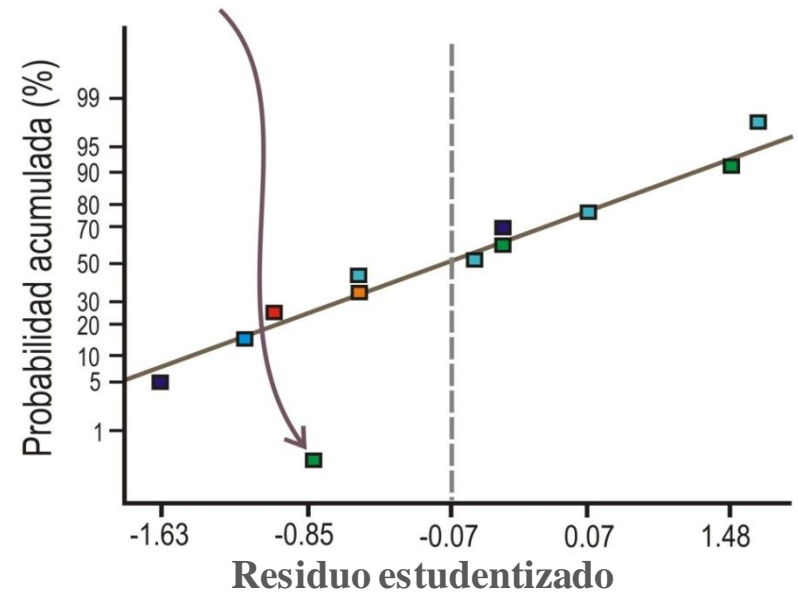
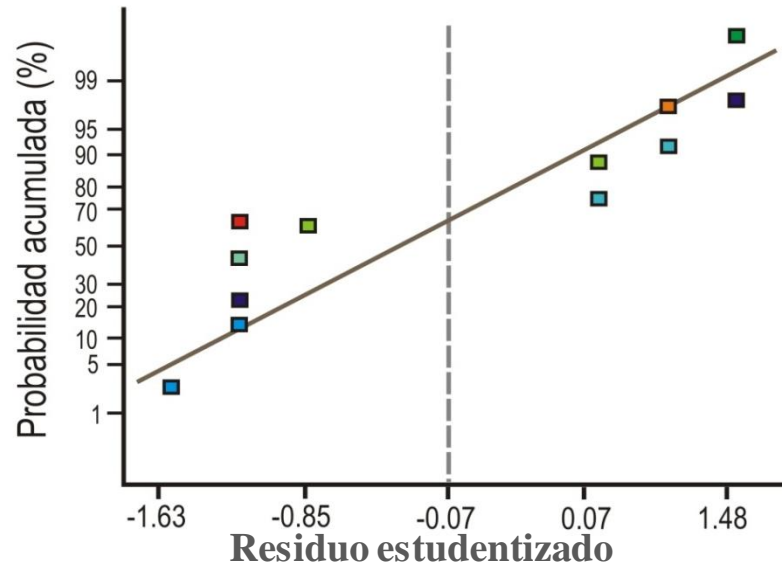
Medida de la
localización del
punto en el espacio
experimental

$$0 \leq h_{ij} \leq 1$$

**Más confiable para
detectar atípicos**

Escalado de residuos

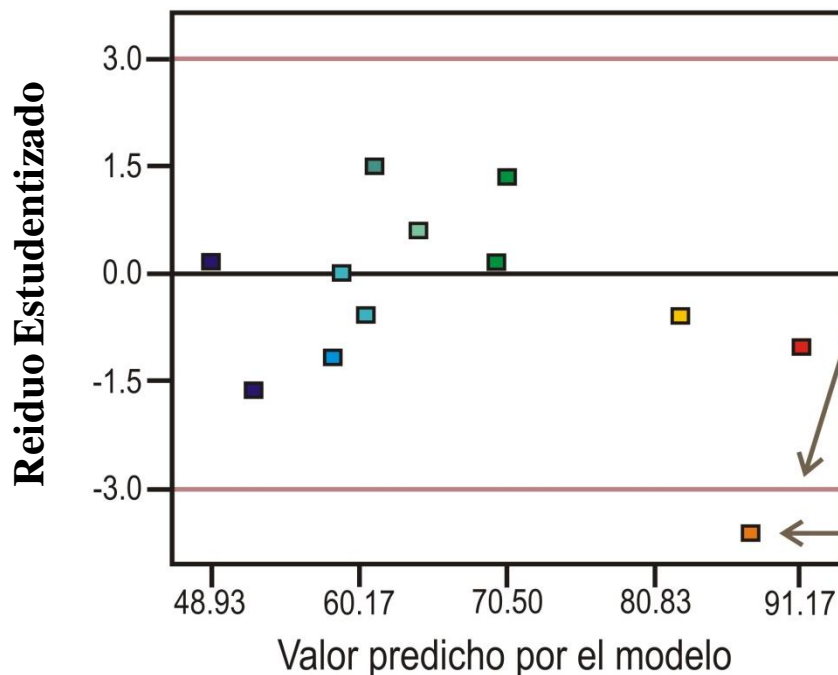
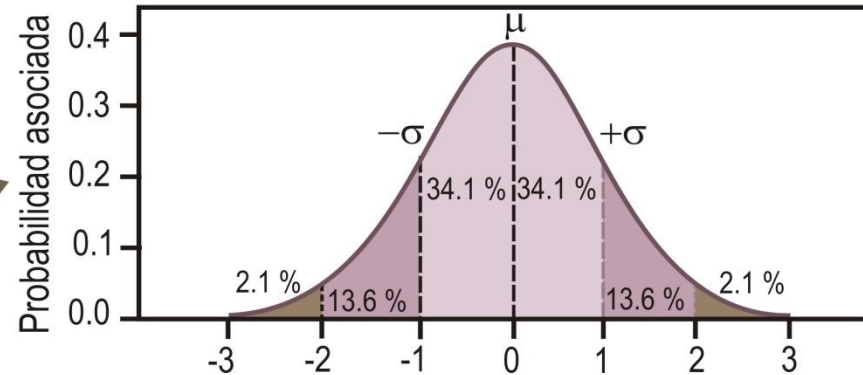
Gráficas de residuo estudentizados



Escalado de residuos

Gráficas de residuos escalados

Los residuos escalados (d_{ij}) deben ser $N(0, \sigma^2)$



límites permitidos de fluctuación de los residuos estandarizados

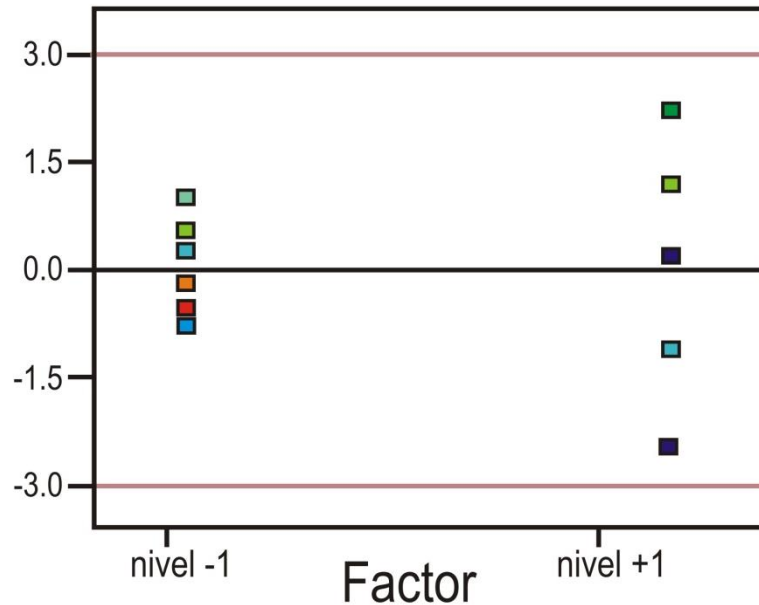
- 68.3 % de los d_{ij} debe estar entre ± 1
- 95.4 % de los d_{ij} debe estar entre ± 2
- 99.7 % de los d_{ij} debe estar entre ± 3

punto atípico

3 o 4 desviaciones estándar del origen

Escalado de residuos

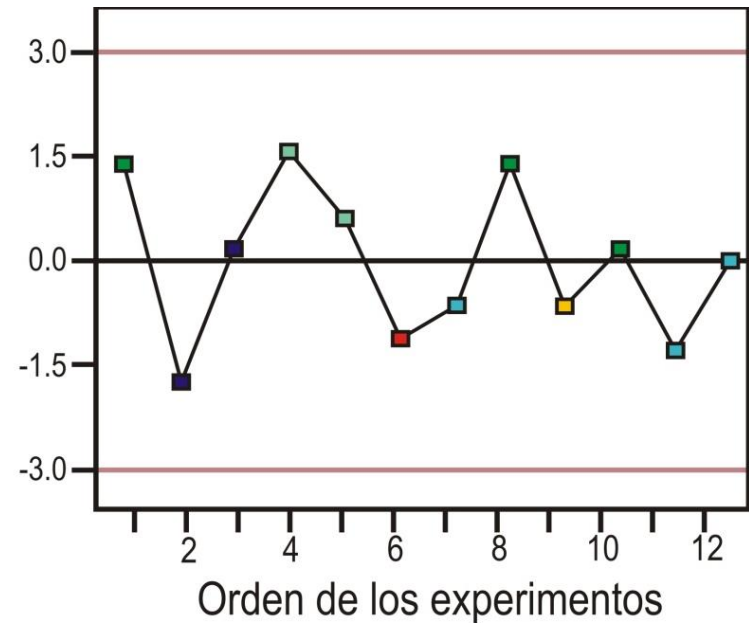
Gráficas de residuos estudentizados – Aplicación de pruebas estadísticas



Test para homogeneidad de varianzas
Prueba de Bartlett y Levene

Prueba de Durbin-Watson

Diagnostica la presencia de correlación
entre los residuos consecutivos



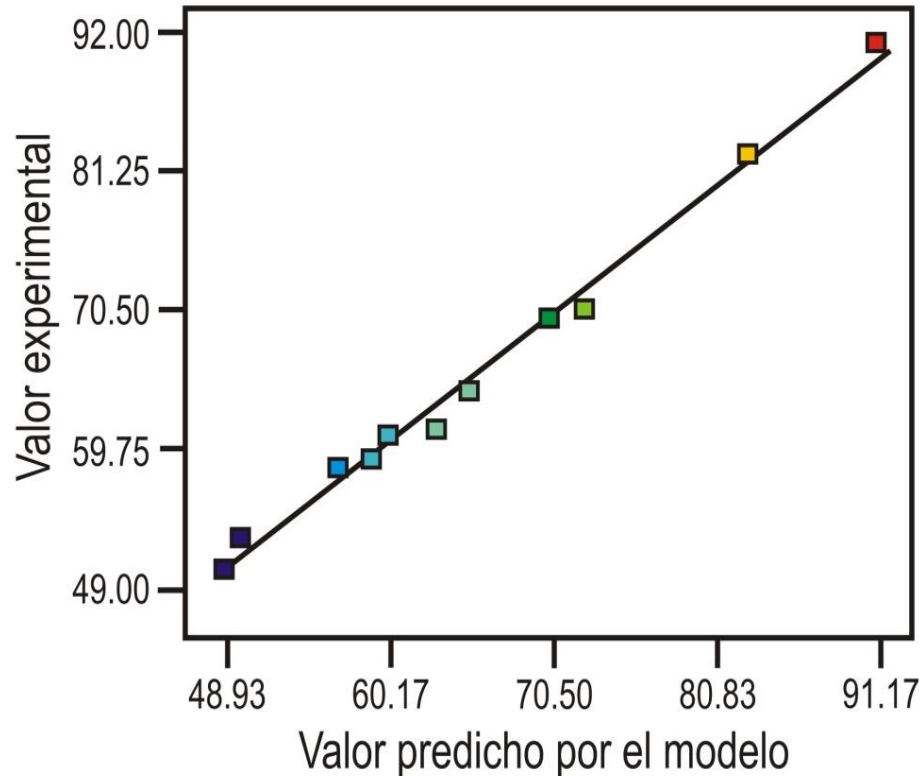
Posibles causas de puntos atípicos

- Errores experimentales.
- Errores de transcripción.
- Ubicación en una región del espacio experimental en la que el modelo, que aparentemente estaba bien ajustado, no está haciendo una buena predicción de la respuesta.

Otra gráfica útil: Capacidad predictiva del modelo

Aproximación a la función real

Gráfica que muestra la bondad de ajuste: cuanto más se aproximen los datos a la diagonal, mejor es el modelo



La recta debe tener pendiente 1 y ordenada al origen cero

Se pueden detectar datos que no son bien predichos por el modelo

Incumplimiento de los supuestos

Cuando los residuos provienen de una muestra pequeña, pueden esperarse pequeños desvíos de la normalidad y la homocedasticidad.

¿Qué hacemos si no podemos ajustar adecuadamente un modelo a los datos experimentales?

¿Qué pasa si los residuos son marcadamente?

No normales

Heterocedásticos

Correlacionados



Transformar la respuesta
Métodos no paramétricos (ANN)

Transformación de la respuesta

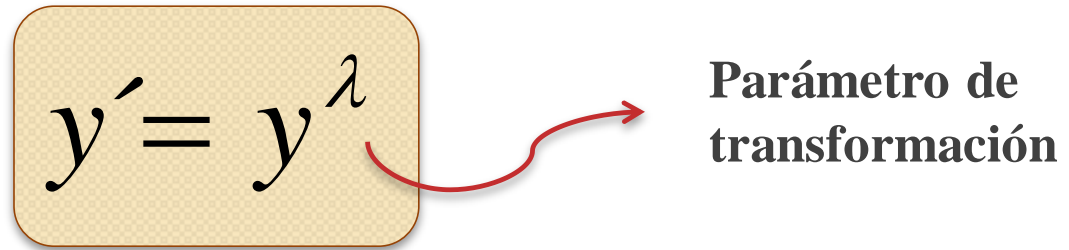
- Estabilizar la varianza de la respuesta.
- Lograr que la distribución de la respuesta sea cercana a la normal.
- Mejorar el ajuste del modelo a los datos.

Transformación de la respuesta

Se aplica una función matemática a la respuesta para obtener una respuesta transformada


$$y' = y^\lambda$$

Parámetro de transformación



- La transformación elegida se basa en el método de la máxima probabilidad.
- Se realizan análisis de varianza para los modelos obtenidos al utilizar diversos valores para *lambda* y se selecciona el que produce menor suma de cuadrados del error.

Transformación de la respuesta

$$y + k > 0$$


Si la respuesta tiene valores negativos debe sumarse una constante para lograr que todos los datos sean positivos antes de aplicar la transformación.



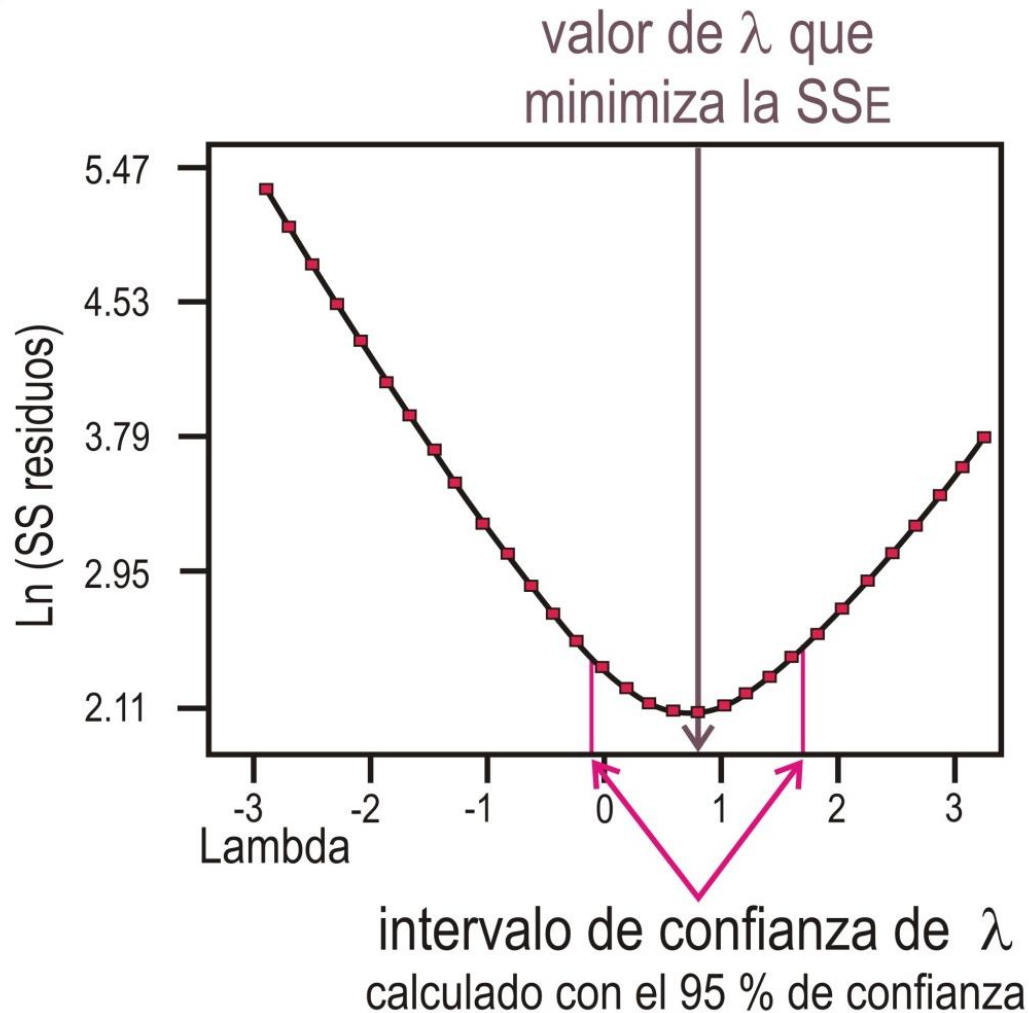
λ

Valores de -3 a +3

En general se prueban entre 10 y 20 valores de λ y se construye una gráfica.

Transformación de la respuesta

GRÁFICA de BOX-COX



Si el intervalo
calculado incluye al
1
no es necesario
transformar los
datos.

Transformaciones más frecuentes

Square Root (lambda = 0.5)

$$y' = \sqrt{y + k}$$

Reciprocal Square Root
(lambda = -0.5)

$$y' = \frac{1}{\sqrt{y + k}}$$

Inverse (lambda = -1.0)

$$y' = \frac{1}{y + k}$$

Natural Log (lambda = 0.0)

$$y' = \ln(y + k)$$

Log10 (lambda = 0.0)

$$y' = \log_{10}(y + k)$$

Transformación de la respuesta

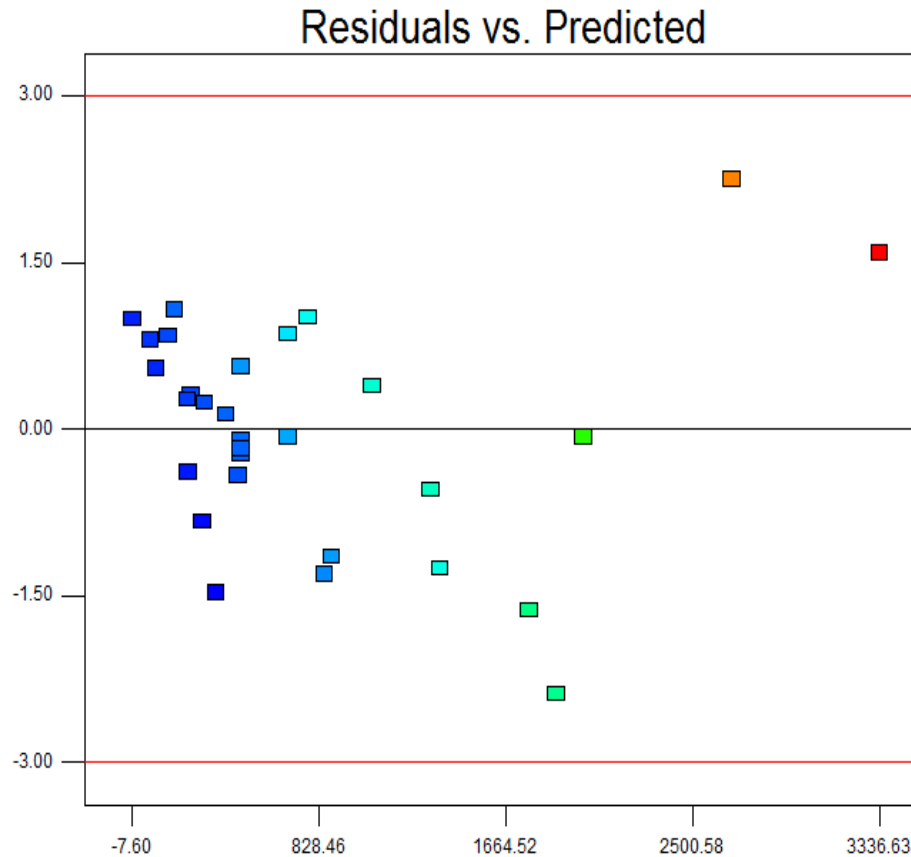
Primer ajuste de un modelo para estudiar el comportamiento de una respuesta en función de tres factores

ANOVA for Response Surface Reduced Quadratic Model

Analysis of variance table [Partial sum of squares - Type III]

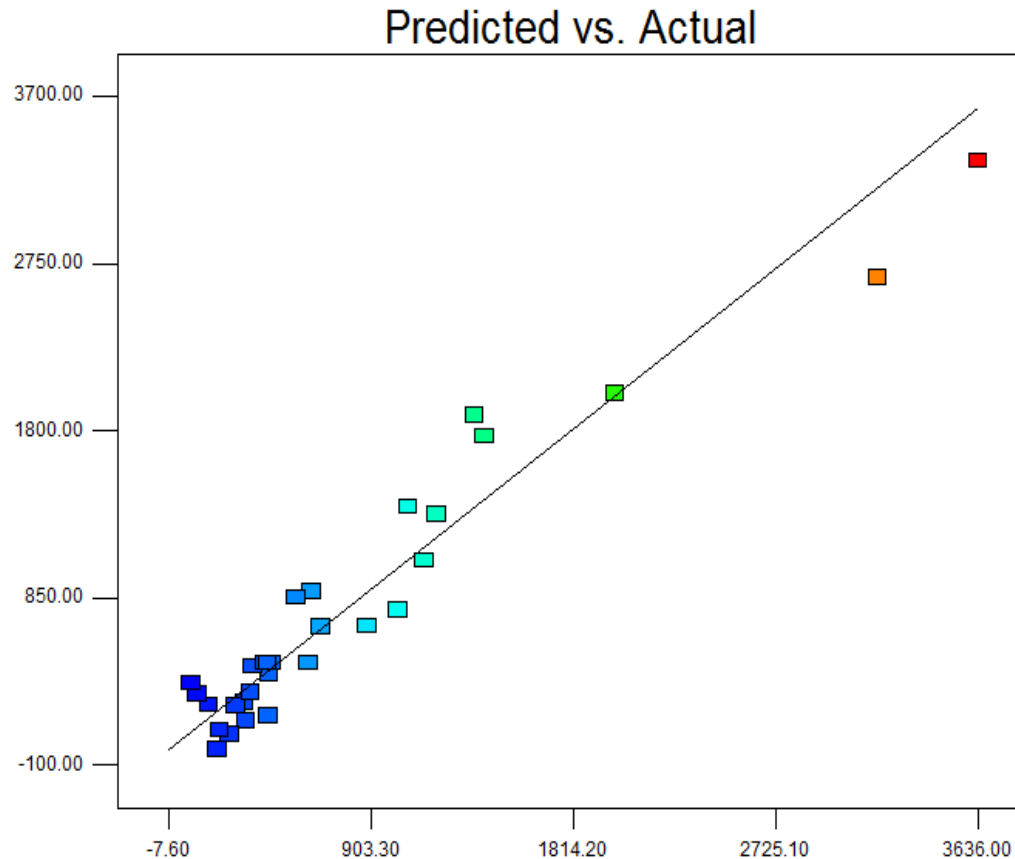
Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	
Model	1.899E+007	8	2.373E+006	34.03	< 0.0001	significant
A-A	7.579E+006	1	7.579E+006	108.68	< 0.0001	
B-B	5.169E+006	1	5.169E+006	74.12	< 0.0001	
C-C	1.616E+006	1	1.616E+006	23.18	< 0.0001	
AB	2.501E+006	1	2.501E+006	35.86	< 0.0001	
AC	5.755E+005	1	5.755E+005	8.25	0.0091	
BC	2.454E+005	1	2.454E+005	3.52	0.0747	
A ²	4.079E+005	1	4.079E+005	5.85	0.0247	
B ²	6.803E+005	1	6.803E+005	9.75	0.0051	
Residual	1.465E+006	21	69738.39			
Lack of Fit	1.439E+006	18	79968.96	9.57	0.0436	significant
Pure Error	25064.75	3	8354.92			
Cor Total	2.045E+007	29				

Transformación de la respuesta



El análisis de gráficas de residuos demuestra heterocedasticidad

Transformación de la respuesta

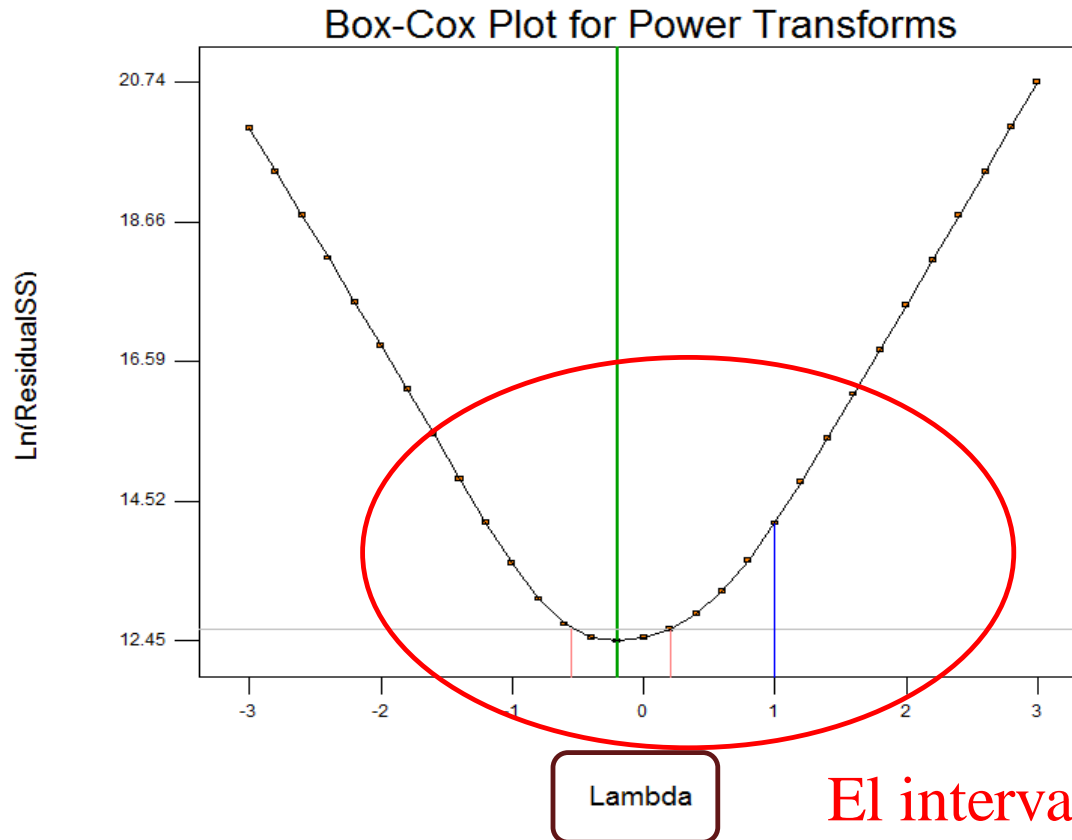


Mala predicción.

El modelo no es capaz de predecir bien los datos

No hay un buen ajuste.

Transformación de la respuesta



El intervalo de confianza
no incluye al uno

Se sugiere transformación logarítmica
(esto lo determina el programa usado)

Transformación de la respuesta

Ajuste con los datos transformados

$$y' = \ln y$$

ANOVA for Response Surface Linear Model

Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value	
Model	22.08	3	7.36	185.99	< 0.0001	significant
A-A	12.25	1	12.25	309.46	< 0.0001	
B-B	7.17	1	7.17	181.11	< 0.0001	
C-C	2.67	1	2.67	67.40	< 0.0001	
Residual	1.03	26	0.040			
Lack of Fit	0.94	23	0.041	1.33	0.4688	not significant
Pure Error	0.092	3	0.031			
Cor Total	23.11	29				

¡Se logra ajustar adecuadamente un modelo mas sencillo!

Otras gráfica útiles para la detección de puntos atípicos e influyentes

Residuo estudentizado internamente

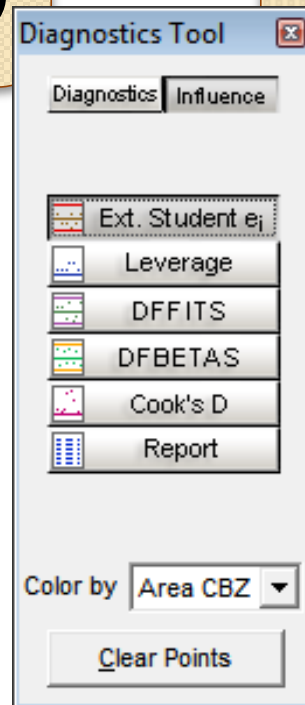
$$r_{ij} = \frac{e_{ij}}{\hat{\sigma} \sqrt{(1 - h_{ij})}}$$

Estimada internamente como CM_E obtenido al ajustar el modelo con **TODOS** los datos

Residuo estudentizado externamente

$$t_{ij} = \frac{e_{ij}}{\hat{\sigma}_{(ij)} \sqrt{(1 - h_{ij})}}$$

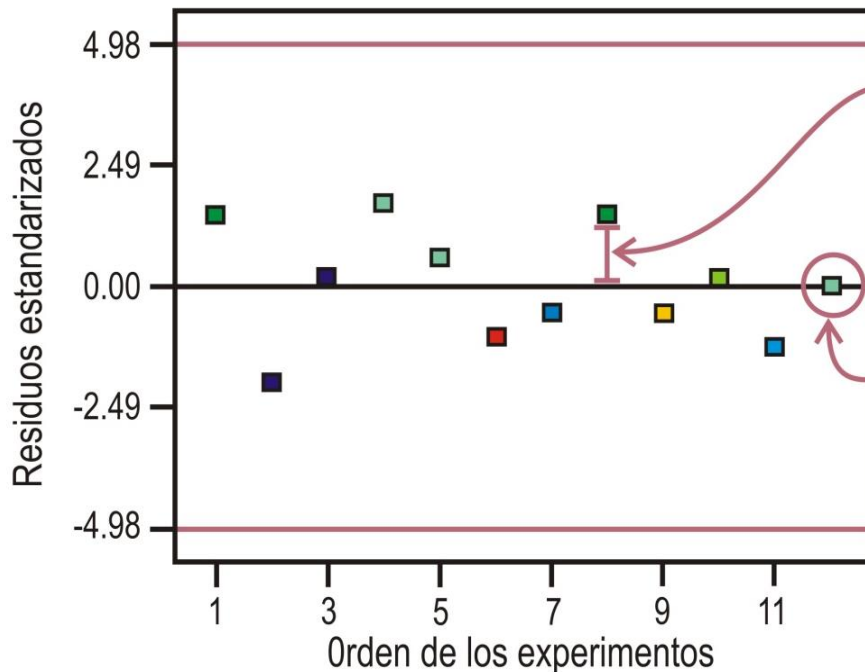
Estimada como CM_E al ajustar el modelo cuando se **quita el dato** y_{ij}



Puntos atípicos e influyentes

Residuo estudentizado externamente \rightarrow **OUTLIERT**

Medida en términos de ' t ' de cuanto se desvía el valor experimental del valor predicho, después de eliminar un experimento dado.



error que habría en la estimación del punto 8 si elimino este punto del análisis

si elimino el exp.12 prácticamente no habría error en su estimación usando los coeficientes calculados con los demás experimentos.

Si el residuo estudentizado externamente supera el límite dado por las líneas rojas se considera un punto atípico.

Puntos influyentes

Puntos que tienen una influencia desproporcionada en el modelo.

Suele suceder en estos casos que los parámetros del modelo dependen más de la influencia de este punto que del conjunto formado por el resto de los puntos.

LEVERAGE

$$h_{ii}$$

Elemento de la diagonal
de la matriz H

$$\hat{y} = Hy$$

Matriz que determina la varianza y covarianza de la respuesta predicha y de los residuos

El promedio de los elementos de la diagonal es:

$$\frac{\sum_{i=1}^n h_{ii}}{n} = \frac{p}{n}$$

factores considerados en el modelo

numero de puntos experimentales



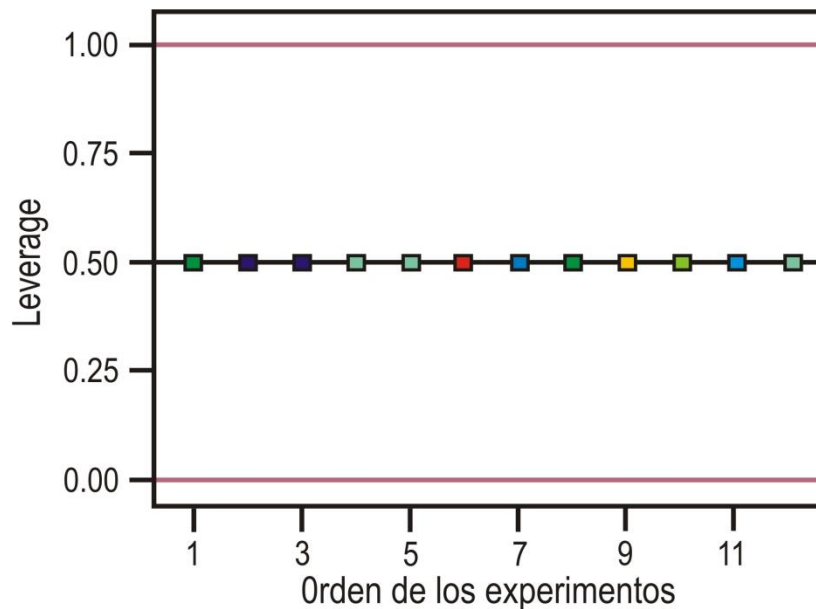
Leverage promedio

Puntos influyentes

LEVERAGE

Es un número entre 0.0 y 1.0 que indica como una medición puntual influye en la predicción por el modelo. Si el leverage es igual 1.0 indica que el valor predicho es igual al experimental y el residuo es cero

Un punto experimental es un posible **punto influyente** cuando tiene un leverage grande



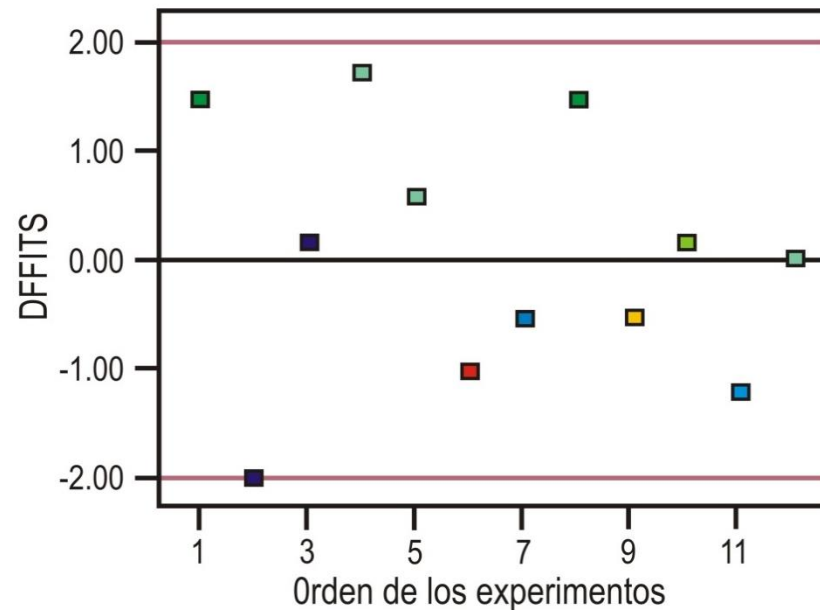
Depende de la matriz experimental y del modelo que se quiere ajustar.

Es decir que se puede conocer antes de realizar los experimentos.

Puntos influyentes

DFFITS (Difference between fitted values test)

Medida de cuantas desviaciones estándar se desvía el valor experimental del valor predicho, después de la supresión de un experimento determinado



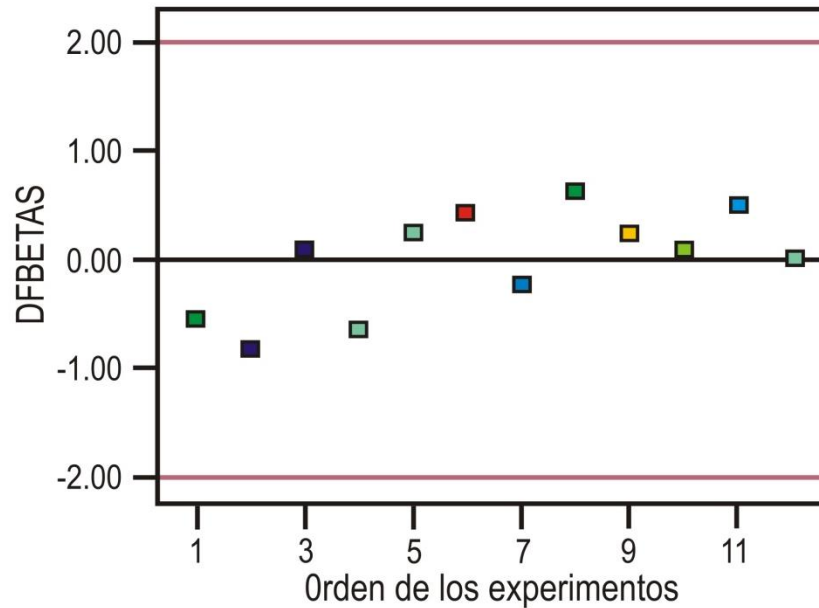
Un punto es influyente cuando está fuera de los límites ± 2 desv. est.

Es la misma gráfica que la de residuos externos estandarizados magnificada por el **leverage**

Puntos Influyentes

DFFBETAS

Mide cuanto cambia el coeficiente de regresión estimado, en unidades de desviaciones estándar, si se elimina un experimento.



Un punto es influyente cuando está fuera de los límites ± 2 desv. est.

Este parámetro se calcula para cada coeficiente del modelo.

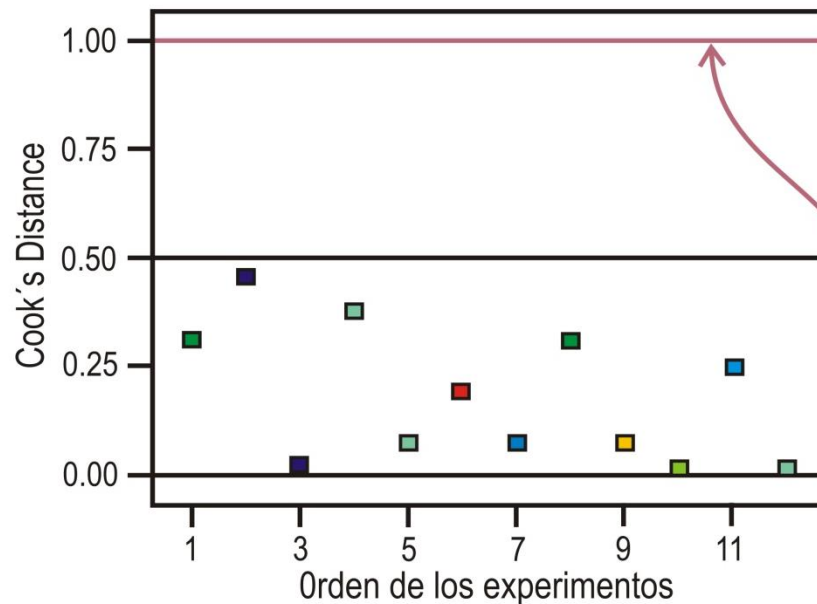
Puntos Influyentes

Distancia de Cook

Es una medida de cuanto puede cambiar la regresión si un experimento se omite del modelo

$\hat{Y} = X b \rightarrow$ vector de valores predichos por el modelo

$\hat{Y}_{(i)} = X b_{(i)} \rightarrow$ vector de valores predichos por el modelo cuando se elimina el *i*-ésimo punto



La distancia de Cook es el cuadrado de la distancia entre los vectores \hat{Y} y $\hat{Y}_{(i)}$ dividido por ps^2

Si $D > 1.0$ entonces el punto se considera influyente

Valores altos están asociados con alto leverage y alto residuo estudentizado.

Gráfica	Evaluación	Respuesta esperada	Respuesta No esperada	Acción
Probabilidad Normal	Distribución normal de res.	Línea recta	Curva forma de «S»	Transformación de respuesta
Residuos vs. Predicho	Constancia de varianza	Dispersión al azar	Varianza expandida	Transformación de respuesta
Residuos vs. Orden	Outliers	Dispersión al azar	Tendencia	Ramdomización y bloqueo
Residuos vs. Factor	Constancia de varianza dependiendo nivel del factor	Dispersión al azar	Curvatura pronunciada	Nuevo modelo de regresión
Residuo Estudentizado Externamente	Outliers	≤ 3.5 estándar desviación		Rechazar
Leverage	Influencia del punto experimental en el error de predicción	≤ 1		Agregar puntos
DFFITS	Influencia del punto experimental en predicción	$< 2/\sqrt{P/N}$ P: parámetros N: núm. Exp.		Agregar puntos
DFBETAS	Influencia de punto experimental en coeficiente	$< 2/\sqrt{N}$		Agregar puntos
Distancia de Cook	Outliers	< 2 promedio Cook		Rechazar

Modelo Definitivo

Modelo Reducido y Jerárquico

Significancia de la Regresión

Falta de Ajuste



Ajustar el Modelo

**Verificación de los
Supuestos del ANOVA**



Eliminar puntos con
residuos atípicos

Aplicar transformaciones

Puntos Influyentes



Eliminar puntos influyentes



Modelo Definitivo

¡Importante!

El análisis de los residuos (para ver supuestos del modelo) se realiza en forma iterativa con el análisis de los efectos (tanto para esta etapa como para la siguiente (optimización))



Problema

(D. Montgomery, 'Diseño y análisis de experimentos', 1991)

Se utiliza una máquina para alisar la superficie de trabajo de una hélice. Se quiere saber que parámetros influyen en la desviación del perfil

Parámetros (-1 y +1):

A: Desplazamiento en eje x (0.001 plg)	0	15
B: Desplazamiento en eje y (0.001 plg)	0	15
C: Desplazamiento en eje z (0.001 plg)	0	15
D: Fabricante de la herramienta	1	2
E: Desplazamiento en eje a (0.001 °)	0	30
F: Velocidad del uso (%)	90	110
G: Altura del montaje (0.001 plg)	0	15
H: Rapidez de alimentación	90	110

- Se usa como variable respuesta la desviación estándar de la diferencia entre perfil real y especificado.
 - Como la máquina tiene 4 husos, se eligen 4 bloques.
 - Se quieren estudiar factores principales e interacciones.

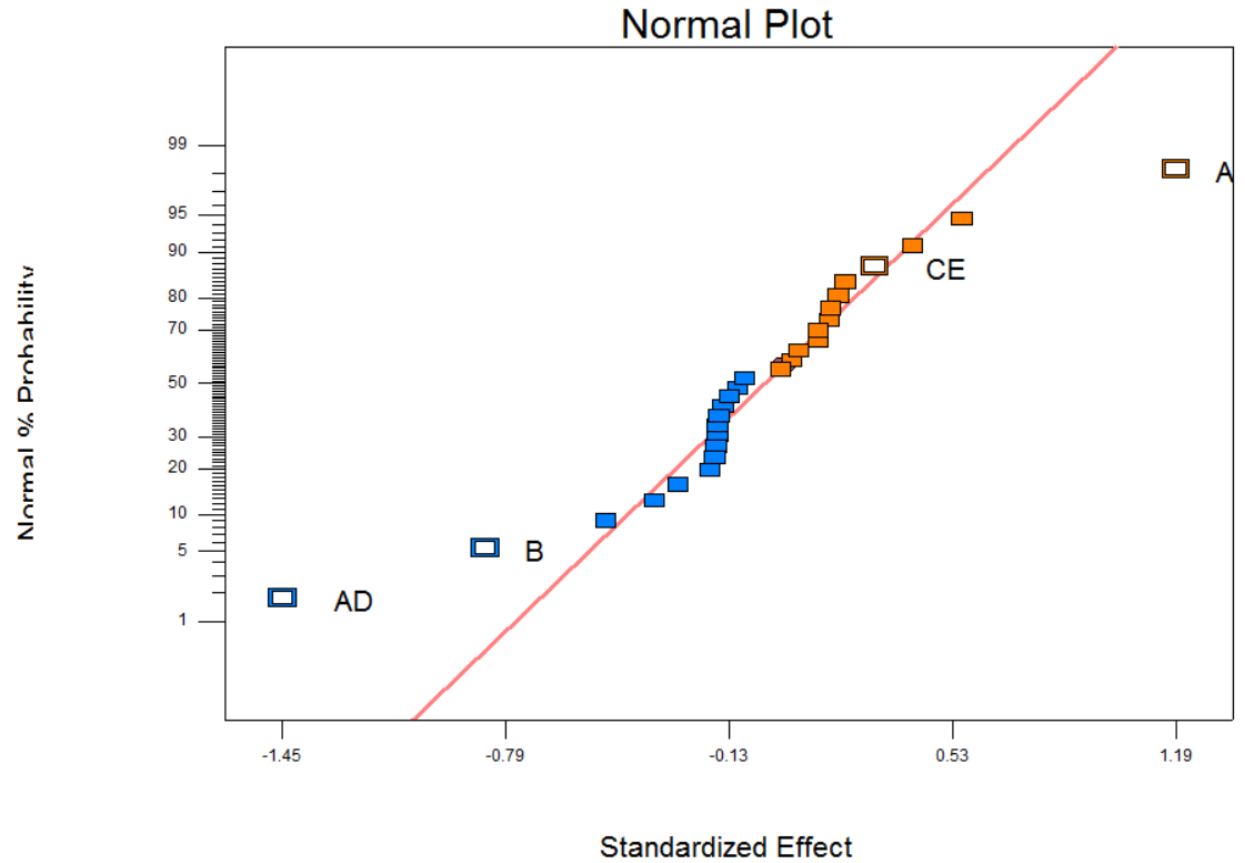
Tabla 8-16 El diseño 2^{8-3} en cuatro bloques del ejemplo 8-6

Corrida	Diseño básico					$F = ABC$	$G = ABD$	$H = BCDE$	Bloque	Orden real de las corridas	Desviación estándar ($\times 10^3$ pulg)
	A	B	C	D	E						
1	-	-	-	-	-	-	-	+	3	18	2.76
2	+	-	-	-	-	+	+	+	2	16	6.18
3	-	+	-	-	-	+	+	-	4	29	2.43
4	+	+	-	-	-	-	-	-	1	4	4.01
5	-	-	+	-	-	+	-	-	1	6	2.48
6	+	-	+	-	-	-	+	-	4	26	5.91
7	-	+	+	-	-	-	+	+	2	14	2.39
8	+	+	+	-	-	+	-	+	3	22	3.35
9	-	-	-	+	-	-	+	-	1	8	4.40
10	+	-	-	+	-	+	-	-	4	32	4.10
11	-	+	-	+	-	+	-	+	2	15	3.22
12	+	+	-	+	-	-	+	+	3	19	3.78
13	-	-	+	+	-	+	+	+	3	24	5.32
14	+	-	+	+	-	-	-	+	2	11	3.87
15	-	+	+	+	-	-	-	-	4	27	3.03
16	+	+	+	+	-	+	+	-	1	3	2.95
17	-	-	-	-	+	-	-	-	2	10	2.64
18	+	-	-	-	+	+	+	-	3	21	5.50
19	-	+	-	-	+	+	+	+	1	7	2.24
20	+	+	-	-	+	-	-	+	4	28	4.28
21	-	-	+	-	+	+	-	+	4	30	2.57
22	+	-	+	-	+	-	+	+	1	2	5.37
23	-	+	+	-	+	-	+	-	3	17	2.11
24	+	+	+	-	+	+	-	-	2	13	4.18
25	-	-	-	+	+	-	+	+	4	25	3.96
26	+	-	-	+	+	+	-	+	1	1	3.27
27	-	+	-	+	+	+	-	-	3	23	3.41
28	+	+	-	+	+	-	+	-	2	12	4.30
29	-	-	+	+	+	+	+	-	2	9	4.44
30	+	-	+	+	+	-	-	-	3	20	3.65
31	-	+	+	+	+	-	-	+	1	5	4.41
32	+	+	+	+	+	+	+	+	4	31	3.40

Design-Expert® Software
Std Dev

Shapiro-Wilk test
W-value = 0.960
p-value = 0.430

- A: A
- B: B
- C: C
- D: D
- E: E
- F: F
- G: G
- H: H
- Positive Effects
- Negative Effects



ANOVA for selected factorial model

Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	
Block	1.52	3	0.51			
Model	28.43	4	7.11	24.47	< 0.0001	significant
A-A	8.34	1	8.34	28.72	< 0.0001	
B-B	5.02	1	5.02	17.27	0.0004	
D-D	0.44	1	0.44	1.52	0.2295	
AD	14.76	1	14.76	50.81	< 0.0001	
Residual	6.97	24	0.29			
Cor Total	36.93	31				

The Model F-value of 24.47 implies the model is significant. There is only a 0.01% chance that a "Model F-Value" this large could occur due to noise.

Values of "Prob > F" less than 0.0500 indicate model terms are significant.

In this case A, B, AD are significant model terms.

Values greater than 0.1000 indicate the model terms are not significant.

If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model.

Std. Dev.	0.54	R-Squared	0.8031
Mean	3.75	Adj R-Squared	0.7703
C.V. %	14.38	Pred R-Squared	0.6513

Confusiones

Notes for Helices_signos

- Design (Actual)
- Summary
- Graph Columns

Transform Effects ANOVA Diagnostics Model Graph

Term	Aliases
Intercept	H

AB	
AC	AB BF CF DG EGH
AD	H AB BG CF DG ABH CFH DGH EFH
AE	H AB CF DG ABH CFH CGH DFH DGH

AB	
AC	AB BF CF DG EGH
AD	H AB BG CF DG ABH CFH DGH EFH
AE	H AB CF DG ABH CFH CGH DFH DGH
AF	H BC DEH
AG	H AB BD CF DG ABH CEH CFH DGH
AH	H AB CF DG ABH CEG CFH DEF DGH
BC	
BD	
BE	H AB CF DG ABH CDH CFH DGH FGH
BF	
BG	
BH	H AB CF DG ABH CDE CFH DGH EFG
CD	H AB CF DG FG ABH BEH CFH DGH
CE	H AB CF DG ABH AGH BDH CFH DGH
CF	
CG	H AB CF DF DG ABH AEH CFH DGH
CH	H AB CF DG ABH AEG BDE CFH DGH
DE	H AFH BCH
DF	
DG	
DH	H AB CF DG ABH AEF BCE CFH DGH
EF	H AB CF DG ABH ADH BGH CFH DGH
EG	AB CF DG ACH BFH
EH	H AB CF DG ABH ACG ADF BCD BFG
FG	
FH	H AB CF DG ABH ADE BEG CFH DGH

Effects Tool

- Half-Normal
- Normal Plot
- Pareto Chart
- Effects List
- Alias List

Clear Selection

Recalculate

Se debe analizar si la interacción AD es factible o si es alguno de los alias

Problema

En una fábrica de semiconductores se quiere mejorar el rendimiento usando DOE. Los factores que podrían tener mayor influencia sobre la variable respuesta (rendimiento) son cinco y se decide correr un diseño 2^5 con una sola réplica (32 tratamientos)

A = Nivel de la abertura (pequeña – grande)

B = Tiempo de exposición (20% abajo – 20% arriba)

C = Tiempo de revelado (30 seg – 45 seg)

D = Dimensión de la máscara (pequeña – grande)

E = Tiempo de grabado (14.5 min – 15.5 min)

¿Es posible incluir los 31 efectos en el análisis?

NO, el ANOVA quedaría incompleto ya que no habría grados de libertad para el error.

Se hace un ANOVA preliminar, donde se mandan al error las interacciones de tres factores en adelante:

Tabla 6.14 Datos acomodados en el orden estándar (Yates).

$(1) = 7$	$d = 8$	$e = 18$	$de = 6$
$a = 9$	$ad = 10$	$ae = 12$	$ade = 10$
$b = 34$	$bd = 32$	$be = 35$	$bde = 30$
$ab = 55$	$abd = 50$	$abe = 52$	$abde = 53$
$c = 16$	$cd = 18$	$ce = 15$	$cde = 15$
$ac = 20$	$acd = 21$	$ace = 22$	$acde = 20$
$bc = 40$	$bcd = 44$	$bce = 45$	$bcde = 41$
$abc = 60$	$abcd = 61$	$abce = 65$	$abcde = 63$

	Std	Run	Block	Factor 1 A:A	Factor 2 B:B	Factor 3 C:C	Factor 4 D:D	Factor 5 E:E	Response 1 Rendimiento
	1	29	Block 1	-1.00	-1.00	-1.00	-1.00	-1.00	7
	2	14	Block 1	1.00	-1.00	-1.00	-1.00	-1.00	9
	3	2	Block 1	-1.00	1.00	-1.00	-1.00	-1.00	34
	4	30	Block 1	1.00	1.00	-1.00	-1.00	-1.00	55
	5	10	Block 1	-1.00	-1.00	1.00	-1.00	-1.00	16
	6	28	Block 1	1.00	-1.00	1.00	-1.00	-1.00	20
	7	3	Block 1	-1.00	1.00	1.00	-1.00	-1.00	40
	8	9	Block 1	1.00	1.00	1.00	-1.00	-1.00	60
	9	24	Block 1	-1.00	-1.00	-1.00	1.00	-1.00	8
	10	7	Block 1	1.00	-1.00	-1.00	1.00	-1.00	10
	11	15	Block 1	-1.00	1.00	-1.00	1.00	-1.00	32
	12	21	Block 1	1.00	1.00	-1.00	1.00	-1.00	50
	13	8	Block 1	-1.00	-1.00	1.00	1.00	-1.00	18
	14	27	Block 1	1.00	-1.00	1.00	1.00	-1.00	21
	15	6	Block 1	-1.00	1.00	1.00	1.00	-1.00	44
	16	5	Block 1	1.00	1.00	1.00	1.00	-1.00	61
	17	20	Block 1	-1.00	-1.00	-1.00	-1.00	1.00	18
	18	22	Block 1	1.00	-1.00	-1.00	-1.00	1.00	12
	19	23	Block 1	-1.00	1.00	-1.00	-1.00	1.00	35
	20	1	Block 1	1.00	1.00	-1.00	-1.00	1.00	52
	21	25	Block 1	-1.00	-1.00	1.00	-1.00	1.00	15
	22	12	Block 1	1.00	-1.00	1.00	-1.00	1.00	22
	23	19	Block 1	-1.00	1.00	1.00	-1.00	1.00	45
	24	11	Block 1	1.00	1.00	1.00	-1.00	1.00	65
	25	4	Block 1	-1.00	-1.00	-1.00	1.00	1.00	6
	26	16	Block 1	1.00	-1.00	-1.00	1.00	1.00	10
	27	31	Block 1	-1.00	1.00	-1.00	1.00	1.00	30
	28	17	Block 1	1.00	1.00	-1.00	1.00	1.00	53
	29	32	Block 1	-1.00	-1.00	1.00	1.00	1.00	15
	30	26	Block 1	1.00	-1.00	1.00	1.00	1.00	20
	31	13	Block 1	-1.00	1.00	1.00	1.00	1.00	41
	32	18	Block 1	1.00	1.00	1.00	1.00	1.00	63

Tabla 6.16 ANOVA preliminar para los semiconductores.

<i>FV</i>	<i>SC</i>	<i>GL</i>	<i>CM</i>	F_0	Valor- <i>p</i>
<i>A</i> : Abertu	1 116.28	1	1 116.28	449.32	0.0000
<i>B</i> : T-expo	9 214.03	1	9 214.03	3 708.79	0.0000
<i>C</i> : T-revel	750.78	1	750.78	302.20	0.0000
<i>D</i> : máscara	5.28	1	5.28	2.13	0.1642
<i>E</i> : T-grab	1.53	1	1.53	0.62	0.4439
<i>AB</i>	504.03	1	504.3	202.88	0.0000
<i>AC</i>	1.53	1	1.53	0.62	0.4439
<i>AD</i>	0.03	1	0.03	0.01	0.9121
<i>AE</i>	7.03	1	7.03	2.83	0.1119
<i>BC</i>	0.03	1	0.03	0.01	0.9121
<i>BD</i>	3.78	1	3.78	1.52	0.2351
<i>BE</i>	2.53	1	2.53	1.02	0.3278
<i>CD</i>	5.28	1	5.28	2.13	0.1642
<i>CE</i>	0.78	1	0.78	0.31	0.5827
<i>DE</i>	11.28	1	11.28	5.54	0.0490
Total error	39.75	16	2.48		
Total	11 664.0	31			

Eliminando aquellos factores e interacciones con $p > 0.05$:

Use your mouse to right click on individual cells for definitions.

Response 1 Rendimiento

ANOVA for selected factorial model

Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	significant
Model	11148.19	6	1858.03	286.68	< 0.0001	significant
A-A	1001.28	1	1001.28	154.49	< 0.0001	
B-B	8877.78	1	8877.78	1369.76	< 0.0001	
C-C	657.03	1	657.03	101.37	< 0.0001	
D-D	16.53	1	16.53	2.55	0.1228	
E-E	9.03	1	9.03	1.39	0.2489	
AB	586.53	1	586.53	90.50	< 0.0001	
Residual	162.03	25	6.48			
Cor Total	11310.22	31				

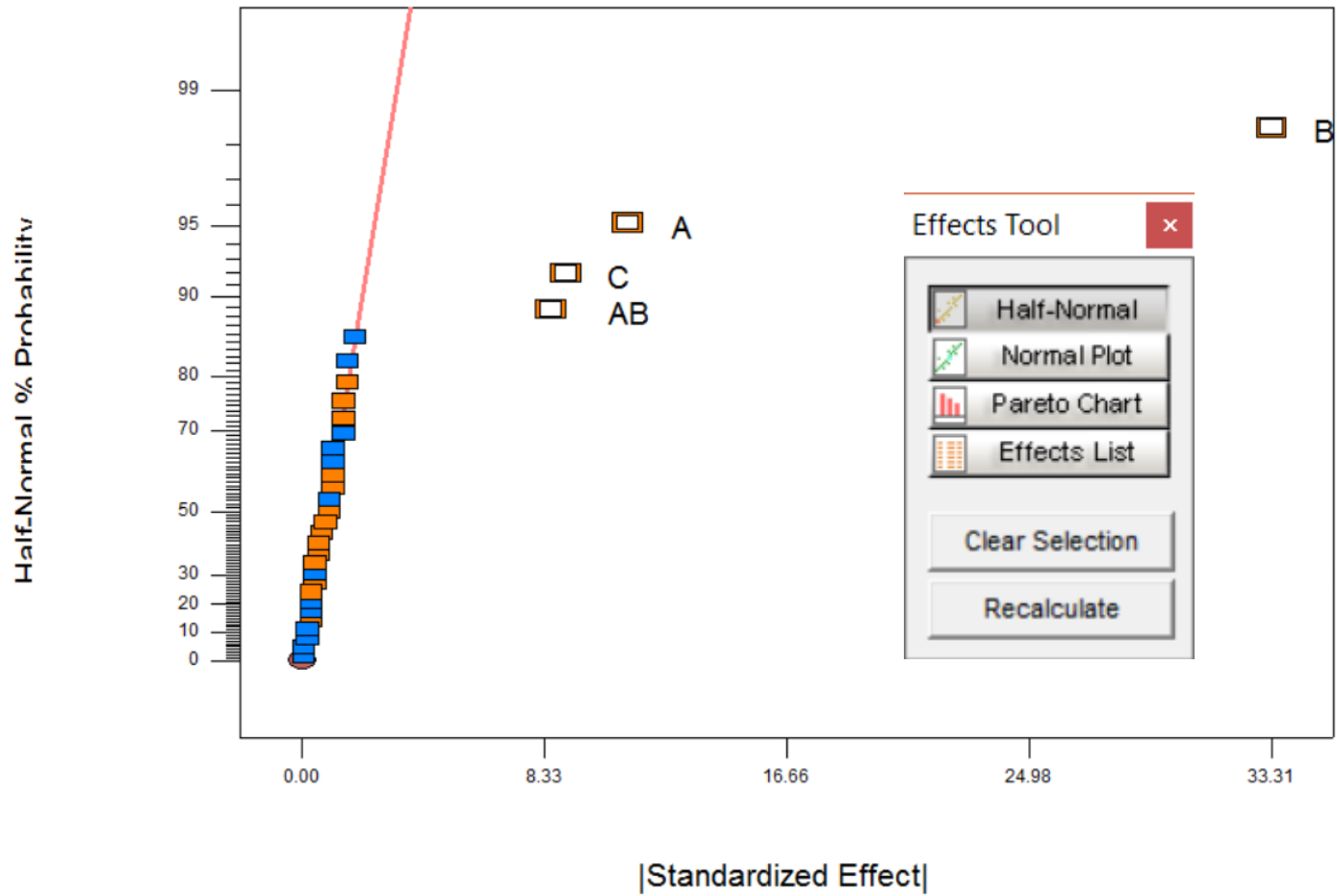
The Model F-value of 286.68 implies the model is significant. There is only a 0.01% chance that a "Model F-Value" this large could occur due to noise.

Values of "Prob > F" less than 0.0500 indicate model terms are significant. In this case A, B, C, AB are significant model terms.

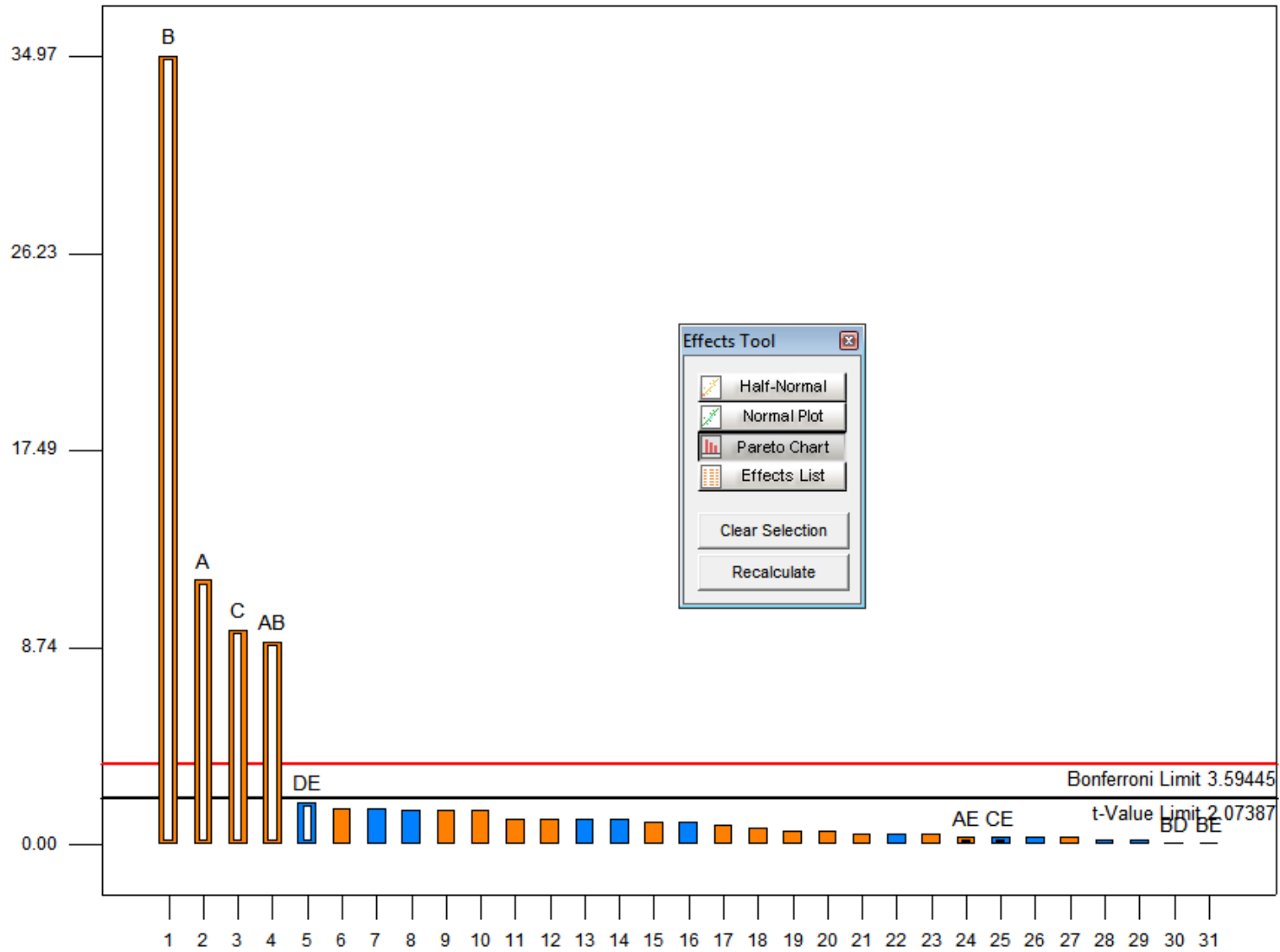
Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model.

Std. Dev.	2.55	R-Squared	0.9857
Mean	30.84	Adj R-Squared	0.9822
...

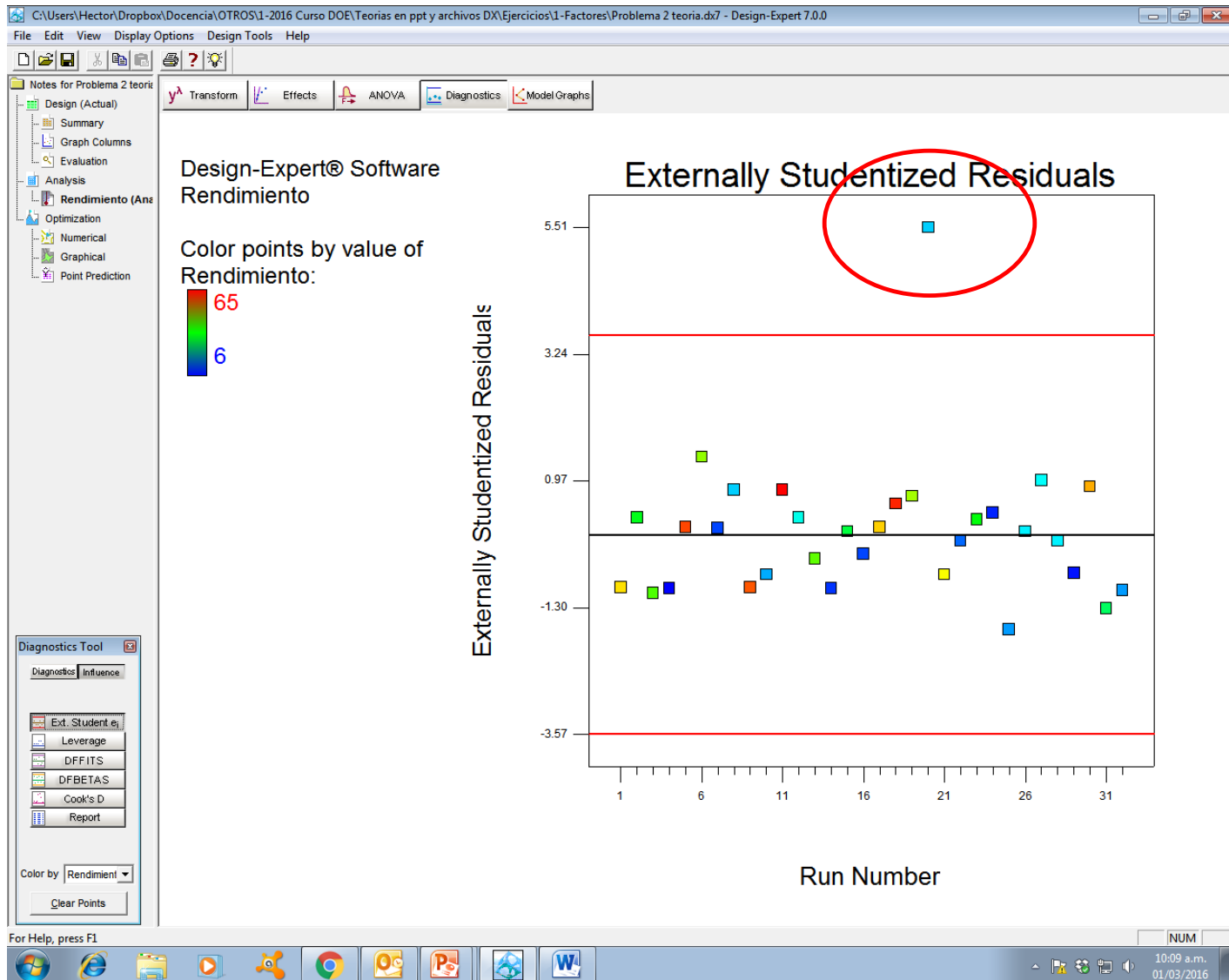
Half-Normal Plot



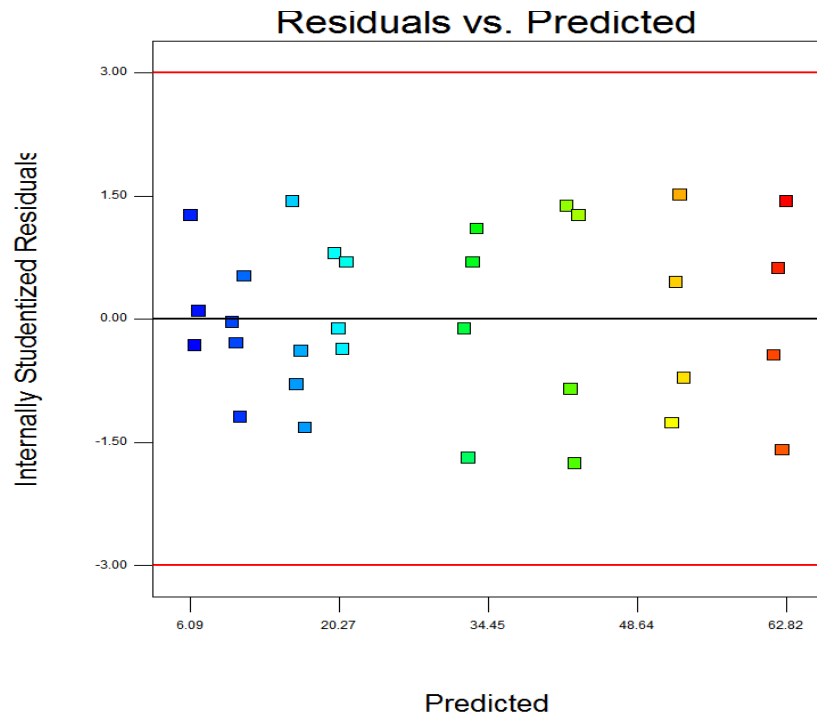
Pareto Chart



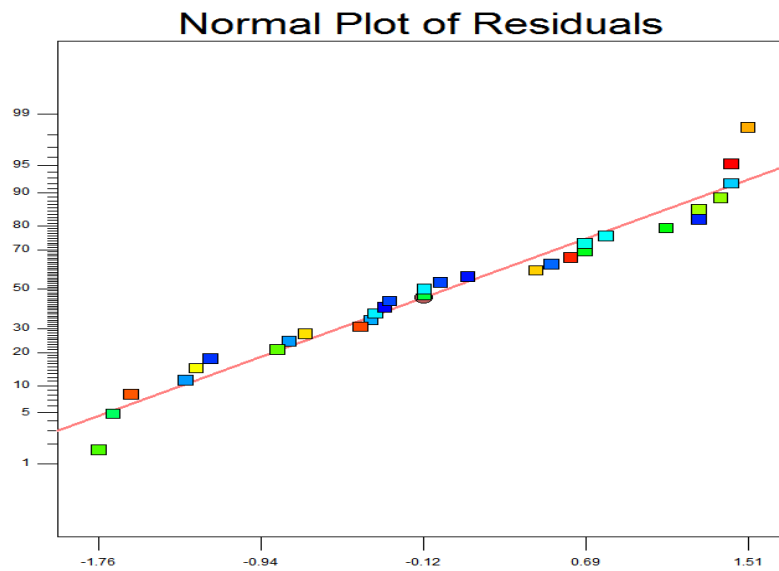
Detección de un punto influyente...

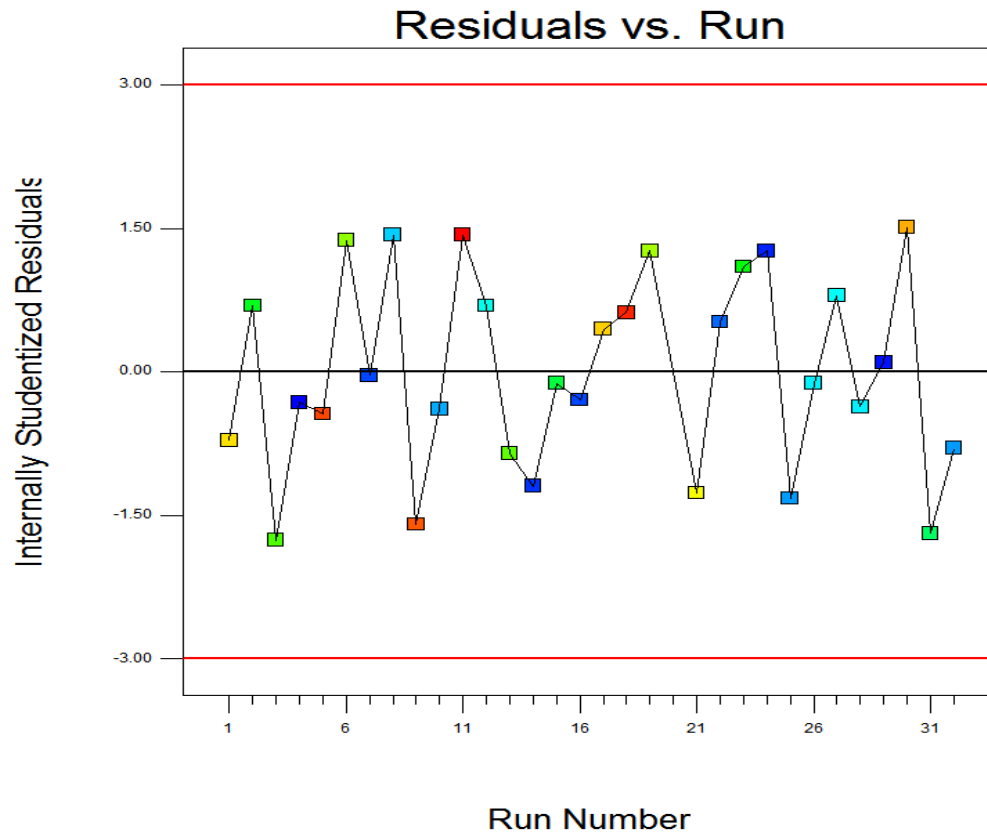


Verificación de los supuestos del ANOVA



El supuesto de normalidad se cumple





El supuesto de independencia se cumple

La situación con los efectos es tan contundente que aún una violación clara de los supuestos difícilmente cambiaría las conclusiones

ETAPA-SCREENING Resumen

OPCIONES GRÁFICAS

ANOVA PRELIMINAR



ANOVA

Es el mejor ANOVA-modelo final donde sólo se incluyen términos significativos, o ¿hay que seguir excluyendo efectos?



- ✓ Los grados de libertad del error deben ser al menos 8 para tener un ANOVA confiable.
- ✓ Indicadores del ajuste (R cuadrado ajustado).



Verificación de los supuestos del modelo que corresponde al mejor ANOVA



Conclusiones

Cálculo de los coeficientes del modelo

Factores		Efecto	Coeficientes
x1	Conc. SR (mM)	-9.2	-4.60
x2	Volumen (μL)	11.5	5.75
x3	Adición de sal (g/L)	9.2	4.60
x4	Tiempo extracción (min)	-12.2	-6.10
x5	Tipo de agitación	-8.5	-4.25
Media general			65.92

$$Y = 65.92 - 4.6x_1 + 5.75x_2 + 4.6x_3 - 6.1x_4 - 4.25x_5$$

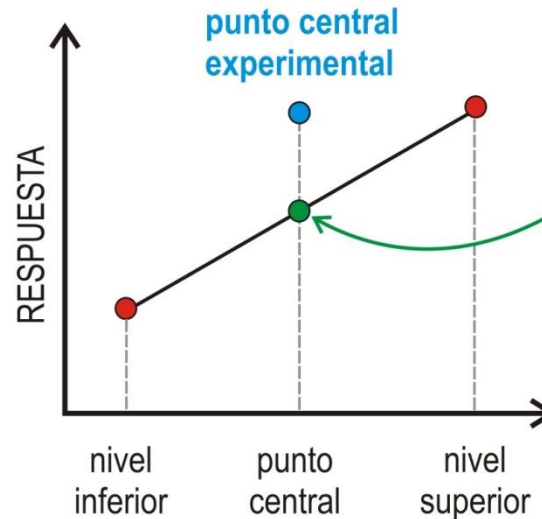
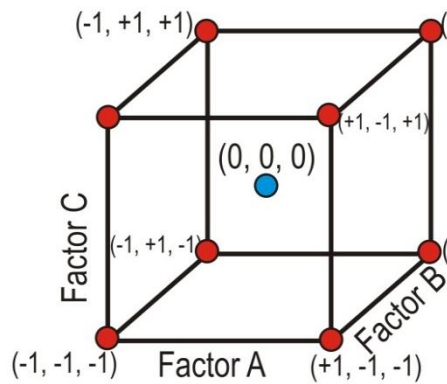
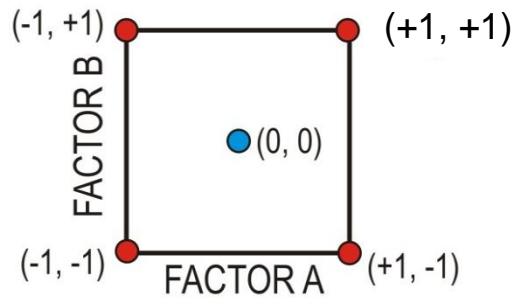
valores codificados
entre -1 y +1

Para que es necesario conocer el modelo de regresión?

para poder conocer la respuesta predicha por el modelo (\hat{Y})
y de allí conocer el **residuo** ($Y_{\text{obs.}} - \hat{Y}_{\text{predicha}}$)

Puntos al centro

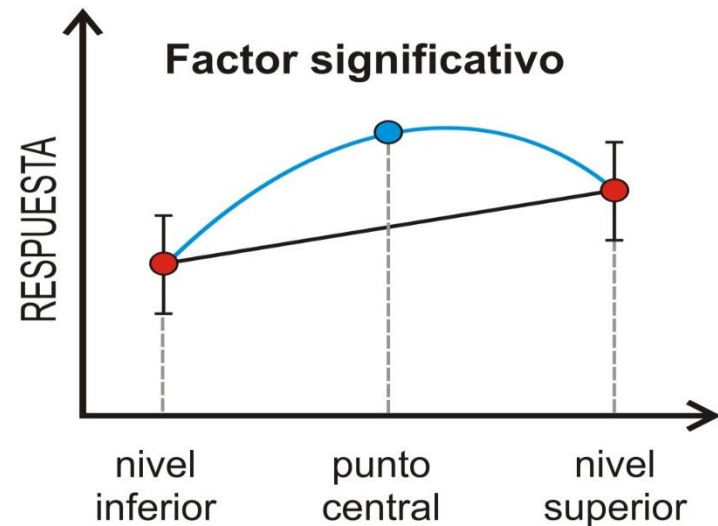
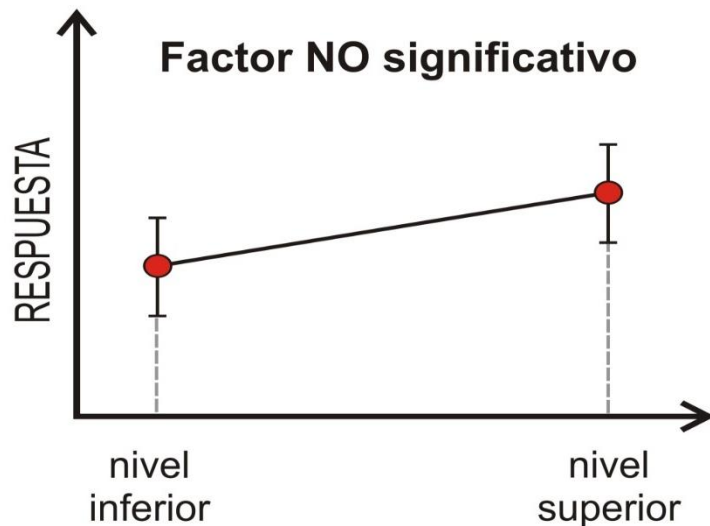
Tratamiento formado por la combinación del nivel intermedio o medio de todos los factores del diseño



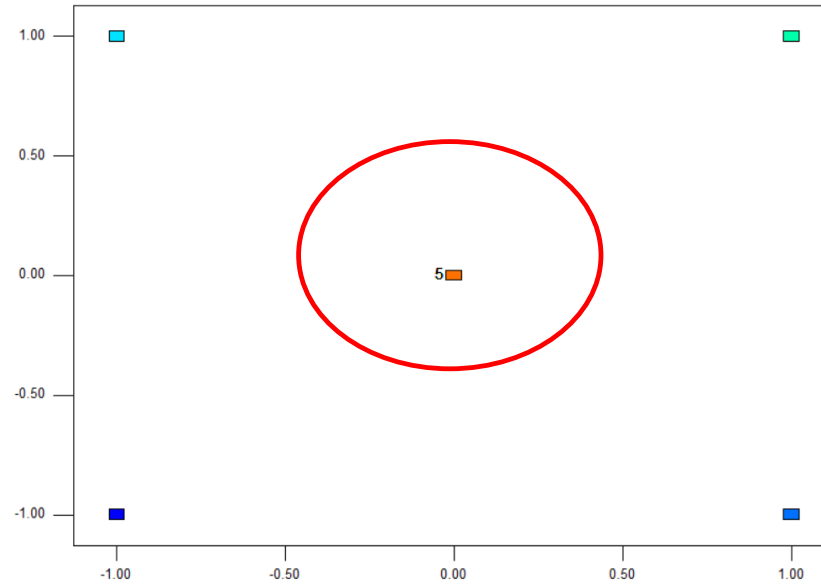
2. estimación del valor de la respuesta en el punto central considerando un comportamiento lineal entre el nivel inferior y superior

Puntos al centro

- Grados de libertad adicionales para el error en la tabla de ANOVA
- ## ANOVA
- Las repeticiones al centro permiten detectar la posible presencia de curvatura



¿Como evaluar la curvatura?



Use your mouse to right click on individual cells for definitions.

Response 1 R1

ANOVA for selected factorial model

Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value	Prob > F	significant
Model	2.83	3	0.94	21.92	0.0060		significant
A-A	0.42	1	0.42	9.83	0.0350		
B-B	2.40	1	2.40	55.87	0.0017		
AB	2.300E-003	1	2.300E-003	0.058	0.8273		
Curvature	2.722E-003	1	2.722E-003	0.063	0.8137		not significant
Pure Error	0.17	4	0.043				
Cor Total	3.00	8					

- Primera etapa concluida, ya se conocen los factores que influyen significativamente.
- Ahora estamos en condiciones de avanzar hacia la optimización

