DISEÑOS FACTORIALES COMPLETOS

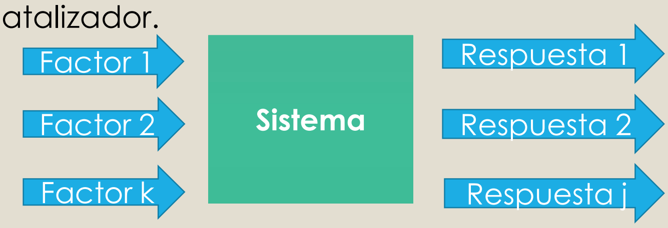
Profa. Dra. Kelly Johana Dussán Medina
kelly.medina@unesp.br

unesp
Instituto de Química
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE LA REPÚBLICA URUGUAY

1

Como variar todo al mismo tiempo

- Un de los problemas más común para quien realiza experimentos es determinar la influencia de una o más variables sobre otra variable de interés.
- Por ejemplo, podemos querer saber cómo el rendimiento afecta si se varia la temperatura o se usa un catalizador diferente. Nos interesa averiguar cómo depende la respuesta (el rendimiento de la reacción) de los factores temperatura y catalizador.



2

2

El sistema actúa como una función - en principio desconocida - que opera sobre las variables de entrada (los factores) y produce como salida las respuestas observadas.

El objetivo de la persona que realiza los experimentos es descubrir una función o, al menos, obtener una aproximación satisfactoria a ella.

Con este conocimiento, podemos comprender mejor la naturaleza de la reacción estudiada y elegir así las mejores condiciones de funcionamiento del sistema.

Al planificar cualquier experimento, lo primero que tenemos que hacer es **decidir** qué **factores** y **respuestas** nos interesan.

Los factores, en general, son las variables que el experimentador puede **controlar**. Pueden ser cualitativas, como el tipo de catalizador, o cuantitativas, como la temperatura.

3

3

A veces, en un experimento, sabemos que hay factores que pueden afectar las respuestas, pero no podemos o no nos interesa controlarlos.

Las respuestas son las variables de salida del sistema que nos interesan y que se verán o no afectadas por los cambios introducidos en los factores. También pueden ser cualitativas o cuantitativas. Dependiendo del problema, podemos tener varias respuestas de interés, que pueden necesitar ser consideradas simultáneamente.

Una vez identificados todos los factores y respuestas, nuestro siguiente paso consiste en definir claramente el objetivo que queremos alcanzar con los experimentos, para poder elegir a continuación el mejor tipo de diseño.

4

4

- Por ejemplo, nuestro químico puede preguntarse:
  - **Cambiar el catalizador** por otro más barato no **disminuirá** el **rendimiento** de la reacción**.**
  - Averiguar qué **temperatura** debe utilizarse para obtener el **máximo rendimiento.**
  - Cuando se pueden **variar los factores** sin alterar el rendimiento o la calidad del producto final.
- La planificación de los experimentos, es decir, la especificación detallada de todas las operaciones experimentales que deben llevarse a cabo dependerá del objetivo que se quiera alcanzar. Objetivos diferentes requerirán diseños diferentes.
- Los diseños factoriales de dos niveles son muy útiles en investigaciones preliminares, cuando queremos saber si determinados factores influyen o no en la respuesta.

5

5

Número relativamente elevado de factores

Algunos de ellos no tienen una influencia significativa en la respuesta.

En este caso, un diseño de experimentos completo no sería la mejor opción.

En primer lugar, decida qué factores merecen un estudio más detallado.

Podemos utilizar el Diseño Factorial Fraccionario.

6

6

Diseño factorial Completo 2^2

- Para llevar a cabo un diseño factorial completo, empezamos por especificar los niveles de cada factor, es decir, los valores de los factores (o clases, en los casos cualitativos) que vamos a utilizar para realizar los experimentos.
- Por ejemplo, vamos a estudiar el efecto del factor de la temperatura en 4 niveles, 50 °C, 60 °C, 70 °C e 80 °C, y el efecto del catalizado em 3 niveles, los catalizadores A, B y C.
- Para hacer un **diseño factorial completo**, debemos realizar experimentos con todas las combinaciones posibles de niveles de los factores. Cada uno de estos experimentos, es un ensayo experimental.

7

7

4 niveles factor A × 3 niveles factor B = 12 ensayos ≠

Fatorial 4 x 3

- n_1 nível del fator 1, n_2 nível del fator 2, … e n_k del factor k, el diseño de experimentos será un factorial $n_1 \times n_2 \times \ldots \times n_k$.
- Número mínimo de ensayos necesarios para un diseño factorial completo. Si se estima el error experimental a partir de ensayos repetidos, se necesitan más experimentos.
- Para estudiar el efecto de cualquier factor sobre una respuesta dada, necesitamos variar su nivel (manipularlo), y observar el resultado que esta variación produce sobre la respuesta.

8

8

Podemos concluir que el diseño de experimentos más sencillo es aquel en que todos los factores se estudian en sólo dos niveles.

Para k factores, **k variables** controladas por el experimentador, un diseño completo de dos niveles requiere la realización de 2 x 2 x ... x 2 = **2^k ensayos** diferentes, siendo llamado de **diseño factorial completo 2^k**

9

de niveles → 2^k ← # de factores

- Analizaremos el efecto de la temperatura y el tipo de catalizador en el rendimiento de una reacción.
- Comenzamos escogiendo los niveles:

Temperatura → Nivel Alto (+) 60°C
Temperatura → Nivel Bajo (-) 40°C

Catalizador → A (-)
Catalizador → B (+)

- Para realizar el diseño 2^2, debemos realizar ensayos y registrar las respuestas observadas (rendimientos) en todas las cuatro posibles combinaciones de los niveles escogidos.
- La lista de estas combinaciones es llamada de **matriz de diseño**.

10

$2^2 = 4$ (ensayos)

Ejemplo 1 - Matriz 2^2

Ensayo	Temperatura	Catalizador	Respuesta
1	-1	-1	
2	+1	-1	
3	-1	+1	
4	+1	+1	

Ensayo	Temperatura (°C)	Catalizador	Rendimiento (%)		Media
1	40	A	57	61	59
2	60	A	92	88	90
3	40	B	55	53	54
4	60	B	66	70	68

Los ensayos se realizaron dos veces, produciendo 8 respuestas. Así, podemos estimar el error experimental de una respuesta individual. La magnitud de este error es importante para decidir si existen o no efectos significativos que podamos atribuir a los factores.

Cálculo dos efectos

- Cuando el efecto de una variable depende del nivel de la otra, decimos que las dos variables interactúan, y podemos calcular el valor del efecto de interacción entre ellas.
- El efecto principal de la temperatura es, por definición, la media de sus efectos sobre los dos niveles del catalizador. Utilizando la letra T para representar este efecto, y siendo $\bar{y}_i$ la respuesta media observada en el i-ésimo ensayo.

$$T = \frac{(\bar{y}_2 + \bar{y}_4)}{2} - \frac{(\bar{y}_1 + \bar{y}_3)}{2} = \frac{(90 + 68)}{2} - \frac{(59 + 54)}{2} = 22{,}5\%$$

- Este valor indica que el rendimiento de la reacción aumenta 22,5% cuando la temperatura sube de su nivel más bajo a su nivel más alto.
- Debemos interpretar los efectos de los dos factores conjuntamente, para que no quede ninguna duda sobre la interacción entre ellos.
- El efecto principal T es la diferencia entre la respuesta de la media en el nivel superior y la respuesta en el nivel inferior de este factor

$$T = \bar{y}_{+} - \bar{y}_{-}$$

- Para el catalizador

$$C = \overbrace{\frac{(\bar{y}_3 + \bar{y}_4)}{2}}^{+} - \overbrace{\frac{(\bar{y}_1 + \bar{y}_2)}{2}}^{-} = \frac{(54 + 68)}{2} - \frac{(59 + 90)}{2} = -13{,}5\%$$

13

13

- Obsérvese que el efecto es negativo. Cuando sustituimos el catalizador A por el catalizador B, el rendimiento disminuye 13,5%.
- Ahora tenemos que calcular la interacción entre los efectos. La mitad de la diferencia entre los dos factores es el efecto de interacción.
- Utilizando T x C, o simplemente TC, para representar este efecto

$$TC = \overbrace{\frac{(\bar{y}_1 + \bar{y}_4)}{2}}^{+} - \overbrace{\frac{(\bar{y}_2 + \bar{y}_3)}{2}}^{-} = \frac{(59 + 68)}{2} - \frac{(90 + 54)}{2} = -8{,}5\%$$

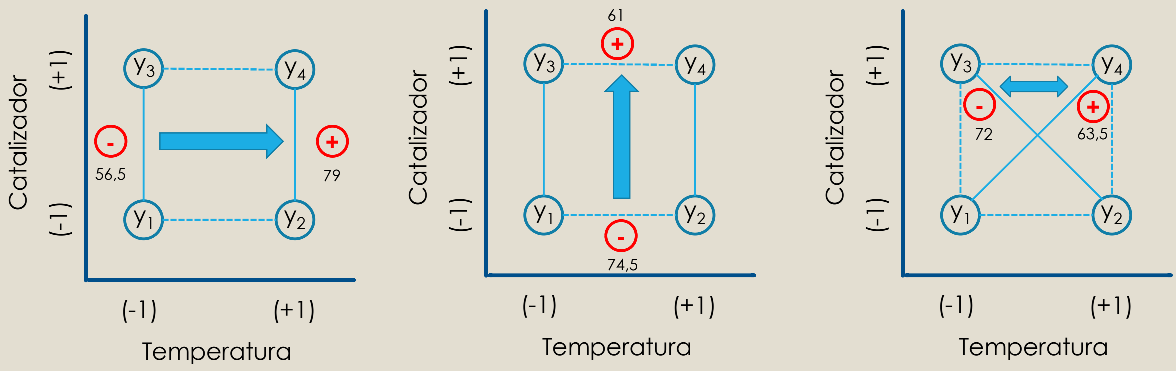
- Para calcular cualquier efecto utilizamos todas las respuestas observadas. Cada efecto es la diferencia entre dos medias. La mitad de las observaciones contribuyen a uno de los promedios, y la mitad restante aparece en el otro promedio. Las respuestas obtenidas nunca son ociosas.

14

14

Interpretación geométrica de los efectos

Podemos representar los efectos en un sistema cartesiano, con un eje para cada factor



Error experimental

Para obtener la desviación estándar de una observación, "s", promediamos las varianzas de las dos muestras ponderadas por sus respectivos grados de libertad:

$$s^2 = \frac{(N_A - 1)\, s_A^2 + (N_B - 1)\, s_B^2}{(N_A - 1) + (N_B - 1)}$$

Cada uno de los ensayos se realizó dos veces y, por tanto, proporciona el cálculo de la varianza con un solo grado de libertad.

Para obtener un cálculo conjunto con 4 grados de libertad, calculamos la media de todos los experimentos, considerando los respectivos grados de libertad.

$$s^2 = \frac{(N_1-1)\ s_1^2 + (N_2-1)\ s_2^2 + (N_3-1)\ s_3^2 + (N_4-1)\ s_4^2}{(N_1-1)+(N_2-1)+(N_3-1)+(N_4-1)}$$

$$s_1^2 = \frac{1}{(N_1-1)}\sum(y_i-\bar{y})^2 = \frac{(57-59)^2+(61-59)^2}{2-1} = 8$$

$$s^2 = \frac{(1\times 8)+(1\times 8)+(1\times 2)+(1\times 8)}{1+1+1+1} = 6{,}5$$

La raíz cuadrada de este valor, obtenemos con cuatro grados de libertad la desviación estándar asociada a una observación, es decir, el error experimental característico, denominado error estándar.

$$s = \sqrt{6{,}5} = 2{,}55\%$$

17

17

- Cuando el número de repeticiones es el mismo en todos los ensayos, la estimación de la varianza experimental es simplemente la media aritmética de las varianzas observadas en los ensayos individuales.
- En el caso general, si cada ensayo se repite "ni" veces y hay "m" ensayos diferentes, la estimación conjunta de la varianza experimental vendrá dada por

$$s^2 = \frac{\nu_1\ s_1^2 + \nu_2\ s_2^2 + \cdots + \nu_m\ s_m^2}{\nu_1+\nu_2+\cdots+\nu_m}$$

Donde $\nu_i = n_i - 1$ son los grados de libertad de s_i^2, la estimación de la variancia del i-ésimo ensayo.

18

18

- Cada uno de los efectos calculados anteriormente son combinaciones lineales de cuatro valores, con coeficientes ai iguales a +1/2 y -1/2. Debido a la autenticidad de las repeticiones y al orden aleatorio en que se realizaron los ensayos, estos valores deben ser estadísticamente independientes.

$$C = \frac{(\bar{y}_3 + \bar{y}_4)}{2} - \frac{(\bar{y}_1 + \bar{y}_2)}{2}$$

- Asumiendo también que tienen la misma varianza poblacional

$$\sigma_{\hat{y}}^2 = \sum_i a_i^2\, \sigma_i^2$$

$$\hat{V}(efeito) = \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}\right)\sigma_{\bar{y}}^2 = \sigma_{\bar{y}}^2$$

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{N}$$

σ^2 es la varianza de una observación individual.

19

19

Utilizando la estimación s^2 = 6,5 en lugar de σ^2, obtenemos finalmente una estimación, con cuatro grados de libertad, del error estándar de un efecto:

$$s(efeito) = \sqrt{\frac{\sigma^2}{2}} = 1{,}80\%$$

Otra forma de obtener el error estándar de un efecto es utilizar la ecuación:

$$T = \bar{y}_+ - \bar{y}_-$$

Cada media contén N/2 observaciones.

$$\sigma^2(efeito) = \sigma^2(\bar{y}_+ - \bar{y}_-) = \sigma^2(\bar{y}_+) + \sigma^2(\bar{y}_-) = \frac{\sigma^2}{(N/2)} + \frac{\sigma^2}{(N/2)} = \frac{4\,\sigma^2}{N}$$

Que es 4 veces la variancia de la media, $\frac{\sigma^2}{N}$. Calculando la raíz cuadrada, tenemos:

$$\sigma(efeito) = \frac{2\,\sigma}{\sqrt{N}} = 2\,\sigma_{\bar{y}}$$

20

20

- Como el error estándar, podemos construir intervalos de confianza para los valores de los efectos, utilizando la distribución de Student

$$\hat{\eta} - t_{\nu} \times s(efeito) < \eta < \hat{\eta} + t_{\nu} \times s(efeito)$$

η es el valor verdadero de un efecto, es decir, el valor poblacional, y el acento circunflejo para indicar la estimación de este valor obtenida a partir de las pruebas realizadas en el experimento.

- En la práctica, la ecuación implica que sólo debemos considerar estadísticamente significativos aquellos efectos cuyas estimaciones sean mayores en valor absoluto que el producto del error estándar del efecto por el punto de la distribución de Student, porque sólo así el intervalo de confianza no incluirá el valor cero.

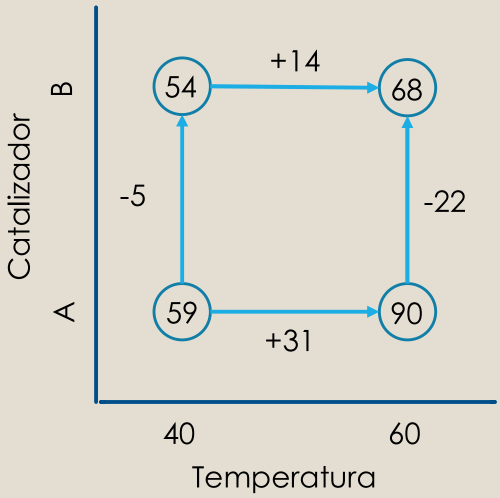
Interpretación de los resultados

Media Global	67,75 ± 0,9
Efectos Principales	
T	22,5 ± 1,8
C	-13,5 ± 1,8
Efectos de Interacción	
TC	-8,5 ± 1,8

Inicialmente, debemos decidir cuáles de los efectos calculados son significativamente diferentes de cero y, por tanto, dignos de interpretación.

Según la ecuación $\hat{\eta} - t_{\nu} \times s(efeito) < \eta < \hat{\eta} + t_{\nu} \times s(efeito)$, sólo consideramos estadísticamente significativo, con un 95% de confianza, un efecto cuyo valor absoluto sea mayor que t_4 x s(efeito) = 2,776 x 1,8% = 5,0%. Por tanto, todos son significativos.

- Como el efecto de interacción es significativo, los efectos principales deben interpretarse conjuntamente. La mejor manera de hacerlo es dibujar un diagrama que contenga las respuestas medias en todas las combinaciones de niveles de variables.



1. Aumentar la temperatura incrementa el rendimiento de la reacción, pero este efecto es mucho más pronunciado con el catalizador A que con el catalizador B.
2. Sustituyendo el catalizador A por el catalizador B reducimos el rendimiento de la reacción, y este efecto es mucho más significativo a 60°C que a 40°C.
3. Los mayores rendimientos se obtienen con el catalizador A y a una temperatura de 60°C.

23

23

Modelo estadístico

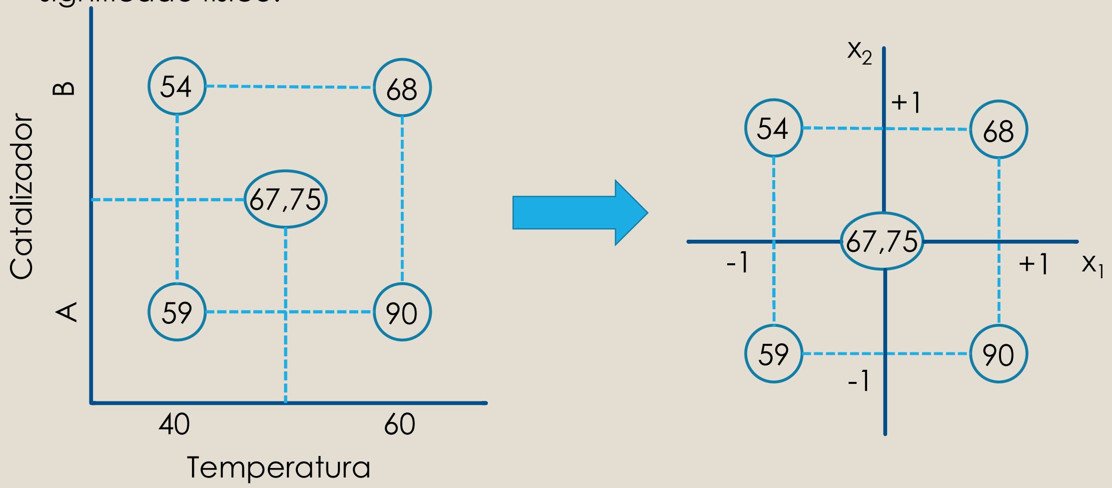
- En el algoritmo utilizado para calcular los efectos, los valores reales de los niveles de los factores se sustituyeron por +1 y -1.
- Esto corresponde a **una codificación** de las variables originales.
- Para transformar los valores 40°C y 60°C en -1 y +1, basta con restar el valor medio, 50°C, de cada uno de ellos y dividir el resultado por la mitad de la amplitud de la variación, que es la diferencia entre el valor superior y el valor inferior.

$$Valor\ Codificado = \frac{Valor\ Real - Punto\ Medio}{Intervalo/2} = \frac{40 - \left(\frac{40+60}{2}\right)}{\left(\frac{60-40}{2}\right)} = -1$$

24

24

- Esto significa situar el origen del eje de la temperatura en el valor intermedio, 50°C, y definir una nueva escala, en la que cada unidad corresponde a 10°C.
- Del mismo modo, la codificación hará que el origen del eje de los catalizadores se centre entre los catalizadores A y B, en una especie de "nivel cero" sin ningún significado físico.



- Con la codificación, cada efecto corresponde ahora siempre a una variación de dos unidades del factor correspondiente, ya que el nivel del factor varía de -1 a + 1. Por unidad de x_1 y x_2 los efectos son la mitad de los valores que calculamos anteriormente.
- Decir que el efecto de la temperatura es del 22,50% cuando T pasa de 40°C a 60°C es lo mismo que decir que este efecto es del 11,50% por unidad de x_1.
- Dividiendo por dos los tres efectos calculados anteriormente, obtenemos los nuevos valores de 11,25% (temperatura), -6,75% (catalizador) y -4,25% (interacción).
- El modelo estadístico utilizado para describir las respuestas de un diseño factorial se formula en términos de los efectos por unidad de x_1 y x_2.

- Para un diseño de experimentos 2^2, la respuesta observada en el nivel (x_1 y x_2) se considera una variable aleatoria $y(x_1, x_2)$. Esta variable se distribuye en torno a una cierta media poblacional η (x_1, x_2), con una cierta varianza poblacional.
- En un diseño de experimentos 2^2, nuestro modelo postula que la media poblacional está adecuadamente representada por la expresión:

$$\hat{y}(x_1, x_2) = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2$$

El acento circunflejo nos recuerda que no se trata de un valor poblacional, sino de una estimación. Los coeficientes b_0, b_1, b_2 y b_{12} se denominan estimadores de los parámetros poblacionales.

En nuestro caso,

$$\hat{y}(x_1, x_2) = 67{,}75 + 11{,}25\, x_1 - 6{,}75\, x_2 - 4{,}25\, x_1 x_2$$

- Nuestra estimación del rendimiento para el ensayo 1 es del 59%, mientras que los valores realmente observados en los experimentos individuales fueron del 57% y 61%. Por lo tanto, la predicción realizada por el modelo para el ensayo 1 deja dos residuos y_1 - $\hat{y}_1$ = 57-59 = -2% e y_1 - $\hat{y}_1$ = 61-59 = 2%.
- Esta diferencia entre los valores observados y los valores predichos aparece siempre que utilizamos un modelo con un número de parámetros inferior al número total de observaciones.
- La estimación de una observación es función de dos variables independientes, con cuatro coeficientes por determinar, que estiman los cuatro parámetros del modelo.
- Los residuos aparecen porque este modelo se ajusta a 8 observaciones. Si sólo hubiera habido 4 observaciones, el ajuste habría sido perfecto y los residuos serían todos cero.

- Es importante señalar que los residuos no son independientes, porque las ecuaciones utilizadas para calcular las estimaciones de los parámetros eliminan cuatro grados de libertad de las observaciones originales. Esto deja sólo 4 grados de libertad para el conjunto de 8 residuos.
- El análisis de los residuos es esencial si queremos evaluar hasta qué punto un modelo se ajusta a las observaciones. Un residuo individual anormalmente alto, por ejemplo, podría significar la presencia de una observación anómala, quizá causada por un error grosero, y podría llevarnos a la conclusión de que deberíamos repetir este ensayo.
- En un modelo bien ajustado, el comportamiento de los residuos no debería ser incompatible con lo que cabría esperar de errores aleatorios.
- Sin embargo, analizar los residuos sólo tiene sentido cuando el número de grados de libertad del conjunto de residuos es relativamente alto.

29

29

Diseño factorial Completo 2^3

$2^3 = 8$ (ensayos)

Ensayo	1	2	3	Rendimiento (%)		Media
1	-	-	-	56 (7)	52 (12)	54,0
2	+	-	-	85 (9)	88 (10)	86,5
3	-	+	-	49 (11)	47 (15)	48,0
4	+	+	-	64 (2)	62 (1)	63,0
5	-	-	+	65 (13)	61 (5)	63,0
6	+	-	+	92 (6)	95 (16)	93,5
7	-	+	+	57 (14)	60 (3)	58,5
8	+	+	+	70 (8)	74 (4)	72,0

30

30

Coeficientes de contraste para un factorial 2^3

1	2	3	12	13	23	123	Media
-	-	-	+	+	+	-	54,0
+	-	-	-	-	+	+	86,5
-	+	-	-	+	-	+	48,0
+	+	-	+	-	-	-	63,0
-	-	+	+	-	-	+	63,0
+	-	+	-	+	-	-	93,5
-	+	+	-	-	+	-	58,5
+	+	+	+	+	+	+	72,0

31

31

Cálculo de los efectos

- El divisor de la media es 8 y 4 para cada uno de los efectos.

$$E_1 = \bar{y}_{1+} - \bar{y}_{2-} = \frac{86{,}5 + 63{,}0 + 93{,}5 + 72{,}0}{4} - \frac{54{,}0 + 48{,}0 + 63{,}0 + 58{,}5}{4}$$

$$E_1 = 22{,}88$$

Media	67,31
1	22,88
2	-13,88
3	8,88
12	-8,63
13	-0,88
23	0,88
123	0,13

32

32

Estimación del error

Como las observaciones individuales se realizaron todas dos veces, podemos utilizar la siguiente ecuación para calcular la estimación de la varianza de una observación individual

$$\hat{V}(y) = s^2 = \sum d_i^2/2N$$

Donde d_i es la diferencia entre dos observaciones correspondientes al ensayo i-ésima.

Sustituyendo los valores numéricos y haciendo N = 8, encontramos $s^2 \approx 5{,}2$.

En un diseño factorial 2^3, cada efecto es una combinación lineal de 8 valores, con coeficientes ±1/4. Suponiendo que estos valores son independientes, podemos utilizar la ecuación que estima la varianza de un efecto.

33

33

Ahora hacemos $a_i^2 = 1/16$, para i = 1,2, ..., 8. Cada uno de los ocho valores de la combinación es, a su vez, la media de otros dos, porque los ensayos se hicieron dos veces. Si la varianza de una observación individual se estima en 5,2, la varianza de la media de las observaciones será 5,2/2.

$$\sigma_y^2 = \sum_i a_i^2\,\sigma_i^2$$

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{2}$$

$$\hat{V}(efeito) = \left(\frac{1}{16} + \cdots + \frac{1}{16}\right) \times \left(\frac{5{,}2}{2}\right) = \left(\frac{8}{16}\right) \times \left(\frac{5{,}2}{2}\right) = 1{,}30$$

El error estándar de un efecto es la raíz cuadrada de este valor, que es aproximadamente 1,14%. El error estándar del rendimiento medio global será la mitad de este valor, 0,57%, porque los coeficientes de la combinación lineal en este caso son todos iguales a ±1/8, en lugar de ±1/4.

34

34

Media	67,3 ± 0,57
Efectos Principales	
1 (Temperatura)	22,9 ± 1,1
2 (Catalizador)	-13,9 ± 1,1
3 (Concentración)	8,9 ± 1,1
Interacciones de dos factores	
12	-8,6 ± 1,1
13	-0,9 ± 1,1
23	0,9 ± 1,1
Interacción de tres factores	
123	0,1 ± 1,1

Diagrama para interpretar los efectos de la temperatura y el catalizador en el diseño de experimentos 2^3

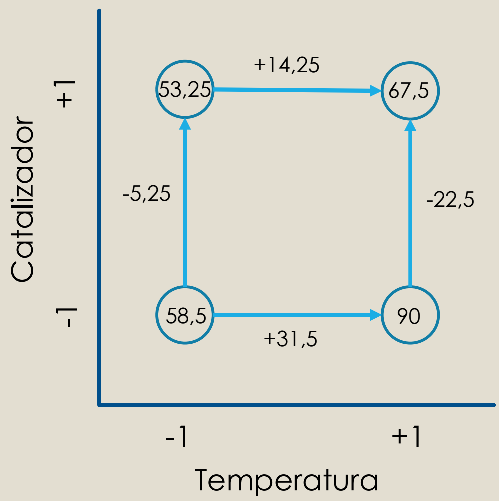
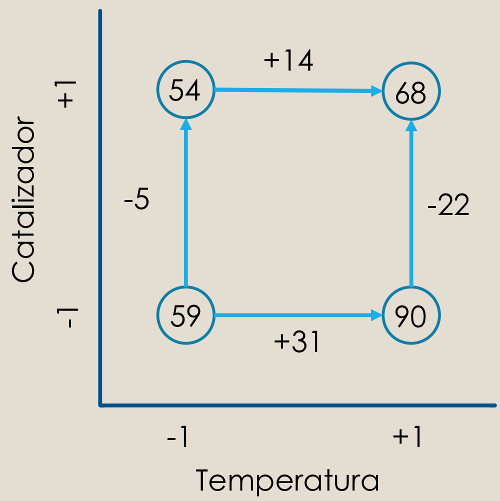


Diagrama para interpretar los efectos de la temperatura y el catalizador en el diseño de experimentos 2^2

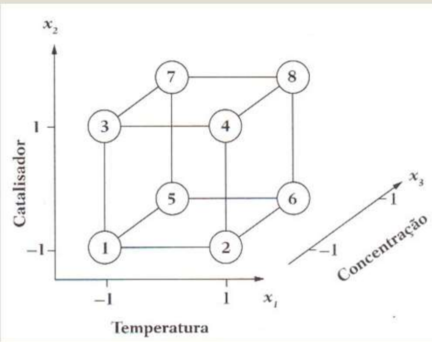


Interpretación de los resultados

Inicialmente, debemos decidir cuáles de los efectos calculados son significativamente diferentes de cero y, por tanto, dignos de interpretación.

- Según la ecuación $\hat{\eta} - t_\nu \times s(efeito) < \eta < \hat{\eta} + t_\nu \times s(efeito)$, sólo consideramos estadísticamente significativo, con un 95% de confianza, un efecto cuyo valor absoluto sea mayor que t_8 x s(efecto) = 2,306 x 1,1% = 2,63%. Así, las interacciones **13, 23 y 123** no son significativas.
- Cuando se aumenta la concentración de 1,0 M a 1,5 M, se produce un aumento medio de alrededor del 9% en el rendimiento, y no hay pruebas de que este aumento dependa de los niveles de las demás variables en el intervalo experimental investigado.

- Los efectos calculados en un factorial 2^3 también pueden interpretarse como contrastes geométricos. .
- Con tres factores en lugar de dos, la figura básica será un cubo y ya no un cuadrado. Los ocho ensayos de la matriz de diseño de experimentos corresponden a los vértices del cubo. Los efectos principales y las interacciones de dos factores son contrastes entre dos planos, que podemos identificar examinando los coeficientes de contraste.



Ensaio	1	2	3	12	13	23	123
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

El modelo estadístico

El modelo puede construirse por analogía con lo propuesto anteriormente, salvo que ahora las variables codificadas son tres: x_1, x_2 e x_3.

$$\hat{y}(x_1, x_2, x_3) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{12}\ x_1x_2x_3$$

b_0	67,3
b_1	11,4
b_2	-6,9
b_3	4,4
b_{12}	-4,3
b_{13}	-0,4
b_{23}	0,4
b_{123}	0,07

39

39

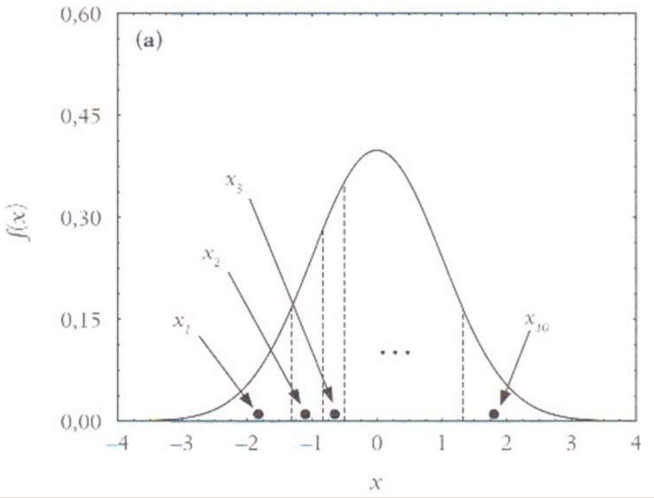
Análisis mediante gráficos normales

- Técnica alternativa para intentar distinguir, en los resultados de un ejercicio de diseño de experimentos, los valores que corresponden realmente a los efectos de los que sólo se deben al ruido.
- Su funcionamiento se basa en la noción de probabilidad acumulada.
- Se construirá organizando los elementos en orden ascendente.
- Como el muestreo ha sido aleatorio, podemos considerar que cada elemento representa una porción equivalente al 100%/#elementos.
- Por ejemplo, para 10 elementos, cada pedazo equivale al 10% del área total de la distribución. El 1er elemento, que es el más pequeño, representa el primer 10%; el 2º representa entre el 10 y el 20%, y así sucesivamente.

40

40

Amostragem aleatória de dez elementos numa distribuição normal padronizada.
Cada elemento representa uma região cuja área é igual a 1/10 da área total sob a curva.



41

41

Exemplo 4

- Ahora tenemos que asociar a cada punto la probabilidad acumulada del centro del intervalo que representa. Así, para el 1er punto, que está en el intervalo de 0 a 10%, la probabilidad acumulada sería del 5%, para el 2º punto sería del 15%...
- Para un factorial 2^4, podemos considerarlos como una muestra aleatoria extraída de una distribución aproximadamente normal.
- Podemos dibujar un gráfico normal de los 15 valores y utilizarlos para probar la hipótesis de que no hay efectos.

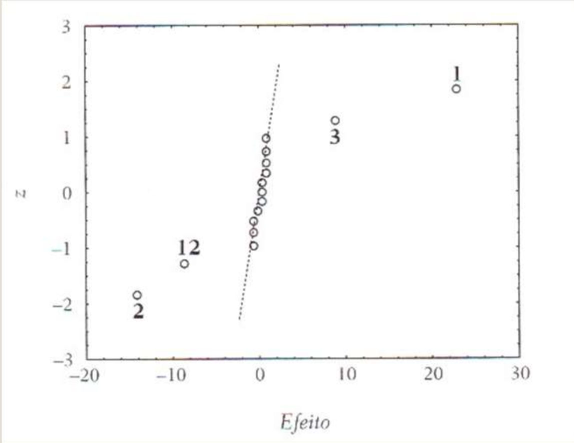
Correspondência entre os efeitos calculados para o planejamento 2^4 e os valores de probabilidade cumulativa

Ordem	Efeito	Região de probabilidade cumulativa (%)	Ponto central	z
1	−14,125	0-6,67	3,33	−1,838
2	−8,625	6,67-13,33	10,00	−1,282
3	−0,625	13,33-20,00	16,67	−0,9673
4	−0,625	20,00-26,67	23,33	−0,7280
5	−0,625	26,67-33,33	30,00	−0,5244
6	−0,125	33,33-40,00	36,67	−0,3406
7	0,375	40,00-46,67	43,33	−0,1680
8	0,375	46,67-53,33	50,00	0,00
9	0,375	53,33-60,00	56,67	0,1680
10	0,875	60,00-66,67	63,33	0,3406
11	0,875	66,67-73,33	70,00	0,5244
12	0,875	73,33-80,00	76,67	0,7280
13	0,875	80,00-86,67	83,33	0,9673
14	8,875	86,67-93,33	90,00	1,282
15	22,875	93,33-100,00	96,67	1,838

42

42

- Los puntos centrales se ajustan muy bien a una línea recta que cruza la probabilidad acumulada del 50% prácticamente en el punto cero del eje de abscisas. Consideramos que estos puntos proceden de una población normal con media cero. En otras palabras, representan "efectos" sin importancia física.
- Los demás efectos están muy alejados de la región central y, por tanto, son efectos significativos.



43

43

RESUMIENDO

44

44

Diseños Experimentales Completos 2^k

\# de niveles → 2^k ← # de factores

MATRIZ DEL DISEÑO DE EXPERIMENTOS

$2^2 = 4$ (ensayos)

Ensaio	1	2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

$2^3 = 8$ (ensayos)

Ensaio	1	2	3
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

45

45

MATRIZ DEL DISEÑO DE EXPERIMENTOS

$2^4 = 16$ (ensayos)

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

46

46

Cálculo dos Efectos

$$Efecto = \bar{y}_{+} - \bar{y}_{-}$$

Diseño Experimental 2^2

$$Factor(1) = \frac{(\bar{y}_2 + \bar{y}_4)}{2} - \frac{(\bar{y}_1 + \bar{y}_3)}{2}$$

Diseño Experimental 2^3

$$Factor(1) = \frac{(\bar{y}_2 + \bar{y}_4 + \bar{y}_6 + \bar{y}_8)}{4} - \frac{(\bar{y}_1 + \bar{y}_3 + \bar{y}_5 + \bar{y}_7)}{4}$$

Diseño Experimental 2^4

$$Factor(1) = \frac{(\bar{y}_2 + \bar{y}_4 + \bar{y}_6 + \bar{y}_8 + \bar{y}_{10} + \bar{y}_{12} + \bar{y}_{14} + \bar{y}_{16})}{8} - \frac{(\bar{y}_1 + \bar{y}_3 + \bar{y}_5 + \bar{y}_7 + \bar{y}_9 + \bar{y}_{11} + \bar{y}_{13} + \bar{y}_{15})}{8}$$

47

47

Interacciones

2^2

Ensaio	1	2	12
1	-1	-1	+1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	+1

$$Interacción(12) = \frac{(\bar{y}_1 + \bar{y}_4)}{2} - \frac{(\bar{y}_2 + \bar{y}_3)}{2}$$

48

48

2^3

Ensaio	1	2	3	12	13	23	123
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

$$Interacción(12) = \frac{(\bar{y}_1 + \bar{y}_4 + \bar{y}_5 + \bar{y}_8)}{4} - \frac{(\bar{y}_2 + \bar{y}_3 + \bar{y}_6 + \bar{y}_7)}{4}$$

49

49

2^4

Ensaio	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234
1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1	+1
2	+1	-1	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	-1	-1
3	-1	+1	-1	-1	-1	+1	+1	-1	-1	+1	+1	+1	-1	+1	-1
4	+1	+1	-1	-1	+1	-1	-1	-1	-1	+1	-1	-1	+1	+1	+1
5	-1	-1	+1	-1	+1	-1	+1	-1	+1	-1	+1	-1	+1	+1	-1
6	+1	-1	+1	-1	-1	+1	-1	-1	+1	-1	-1	+1	-1	+1	+1
7	-1	+1	+1	-1	-1	-1	+1	+1	-1	-1	-1	+1	+1	-1	+1
8	+1	+1	+1	-1	+1	+1	-1	+1	-1	-1	+1	-1	-1	-1	-1
9	-1	-1	-1	+1	+1	+1	-1	+1	-1	-1	-1	+1	+1	+1	-1
10	+1	-1	-1	+1	-1	-1	+1	+1	-1	-1	+1	-1	-1	+1	+1
11	-1	+1	-1	+1	-1	+1	-1	-1	+1	-1	+1	-1	+1	-1	+1
12	+1	+1	-1	+1	+1	-1	+1	-1	+1	-1	-1	+1	-1	-1	-1
13	-1	-1	+1	+1	+1	-1	-1	-1	-1	+1	+1	+1	-1	-1	+1
14	+1	-1	+1	+1	-1	+1	+1	-1	-1	+1	-1	-1	+1	-1	-1
15	-1	+1	+1	+1	-1	-1	-1	+1	+1	+1	-1	-1	-1	+1	-1
16	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

$$Interacción(12) = \frac{(\bar{y}_1 + \bar{y}_4 + \bar{y}_5 + \bar{y}_8 + \bar{y}_9 + \bar{y}_{12} + \bar{y}_{13} + \bar{y}_{16})}{8} - \frac{(\bar{y}_2 + \bar{y}_3 + \bar{y}_6 + \bar{y}_7 + \bar{y}_{10} + \bar{y}_{11} + \bar{y}_{14} + \bar{y}_{15})}{8}$$

50

50

Estimación Error experimental

Varianza experimental

$$s^2 = \frac{\nu_1\ s_1^2 + \nu_2\ s_2^2 + \cdots + \nu_m\ s_m^2}{\nu_1 + \nu_2 + \cdots + \nu_m}$$

2^2

$$s^2 = \frac{(N_1 - 1)\ s_1^2 + (N_2 - 1)\ s_2^2 + (N_3 - 1)\ s_3^2 + (N_4 - 1)\ s_4^2}{(N_1 - 1) + (N_2 - 1) + (N_3 - 1) + (N_4 - 1)}$$

$$s_1^2 = \frac{1}{(N_1 - 1)} \sum (y_i - \bar{y})^2$$

N = número de repeticiones por ensayo

ν_m = Grados de Liberdade

51

51

2^3

$$s^2 = \frac{(N_1 - 1)\ s_1^2 + (N_2 - 1)\ s_2^2 + (N_3 - 1)\ s_3^2 + (N_4 - 1)\ s_4^2 + (N_5 - 1)\ s_5^2 + (N_6 - 1)\ s_6^2 + (N_7 - 1)\ s_7^2 + (N_8 - 1)\ s_8^2}{(N_1 - 1) + (N_2 - 1) + (N_3 - 1) + (N_4 - 1) + (N_5 - 1) + (N_6 - 1) + (N_7 - 1) + (N_8 - 1)}$$

Se todos los ensayos son realizados dos vezes, N - 1 = 1 = N_1 = N_2 = N_3 = ... = N_m

$$s^2 = \frac{(1)\ s_1^2 + (1)\ s_2^2 + (1)\ s_3^2 + (1)\ s_4^2 + (1)\ s_5^2 + (1)\ s_6^2 + (1)\ s_7^2 + (1)\ s_8^2}{(1) + (1) + (1) + (1) + (1) + (1) + (1) + (1)}$$

Grados de Liberdade = 8

La raíz cuadrada de la varianza experimental es el error estándar (%)

$$s = \sqrt{s^2}$$

52

52

Coeficientes de la combinación lineal

2^2 = +1/2 e -1/2

2^3 = +1/4 e -1/4

2^4 = +1/8 e -1/8

Varianza de la población

$$\sigma_y^2 = \sum_i a_i^2\, \sigma_i^2$$

53

Varianza del efecto

2^2

$$\hat{V}(efeito) = \left(\frac{1}{2}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{2}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{2}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{2}\right)^2 \sigma_{\bar{y}}^2$$

$$\hat{V}(efeito) = \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}\right)\sigma_{\bar{y}}^2 = \sigma_{\bar{y}}^2$$

Recordando que la varianza de la media es: $\sigma_{\bar{y}}^2 = \frac{\sigma^2}{N}$; N = número de repeticiones por ensayos.

2^3

$$\hat{V}(efeito) = \left(\frac{1}{4}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{4}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{4}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{4}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{4}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{4}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{4}\right)^2 \sigma_{\bar{y}}^2 + \left(\frac{1}{4}\right)^2 \sigma_{\bar{y}}^2$$

$$\hat{V}(efeito) = \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16}\right)\sigma_{\bar{y}}^2 = \frac{1}{2}\sigma_{\bar{y}}^2$$

54

Error estándar de un efecto

$$s(efeito) = \sqrt{\frac{\sigma^2}{N}} = \frac{\sigma}{\sqrt{N}}$$

N = número de repeticiones por ensayo

Erro estándar de la media

Cada media tiene N/2 observaciones.

$$\sigma^2(efeito) = \sigma^2(\bar{y}_+ - \bar{y}_-) = \sigma^2(\bar{y}_+) + \sigma^2(\bar{y}_-) = \frac{\sigma^2}{(N/2)} + \frac{\sigma^2}{(N/2)} = \frac{4\,\sigma^2}{N}$$

Que es 4 veces la variancia de la media, $\frac{\sigma^2}{N}$. Calculando raíz cuadrada, tenemos:

$$\sigma(efeito) = \frac{2\,\sigma}{\sqrt{N}} = 2\,\sigma_{\bar{y}}$$

55

55

Teniendo en cuenta que las observaciones se hicieron dos vezes

Otra forma de calcular la estimación de la varianza de una observación individual es:

$$\hat{V}(y) = s^2 = \sum d_i^2/2N$$

d_i es la diferencia entre dos observaciones correspondientes al i-ésimo ensayo.

N = número de ensayos del diseño de experimentos.

56

56

Intervalos de confianza

$$\hat{\eta} - t_{\nu} \times s(efecto) < \eta < \hat{\eta} + t_{\nu} \times s(efecto)$$

η es el valor verdadero de un efecto (valor poblacional), y el acento circunflejo para indicar la estimación de este valor obtenida a partir de los ensayos realizados en el experimento.

En la práctica, la ecuación implica que sólo debemos considerar estadísticamente significativos aquellos efectos cuyas estimaciones sean mayores en valor absoluto que el producto del error estándar por el punto de la distribución de Student, porque sólo entonces el intervalo de confianza no incluirá el valor cero..

En otras palabras, un intervalo de confianza que contenga el valor cero indica que no es significativo.

57

57

2^2

$$\hat{y}(x_1, x_2) = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2$$

2^3

$$\hat{y}(x_1, x_2, x_3) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3$$

2^4

$$\hat{y}(x_1, x_2, x_3, x_4) \\ = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4\,x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 \\ + b_{24}x_2x_4 + b_{34}\,x_3x_4 + b_{123}x_1x_2x_3 + b_{124}x_1x_2x_4 + b_{134}x_1x_3x_4 + b_{23}\,\,x_2x_3x_4 + b_{12}\quad x_1x_2x_3x_4$$

Los valores de los coeficientes del modelo propuesto, b_1, b_2, b_3 ... son iguales a la mitad de los valores de los efectos que calculamos anteriormente.

58

58