

Diseño Factorial 2^k

1 INTRODUCCIÓN

Los diseños factoriales son ampliamente utilizados en experimentos en los que intervienen varios factores para estudiar el efecto conjunto de éstos sobre una respuesta. En el Capítulos anteriores se presentaron métodos generales para, el análisis de diseños factoriales. Sin embargo, existen varios casos especiales del diseño factorial general que resultan importantes porque se usan ampliamente en el trabajo de investigación, y porque constituyen la base para otros diseños de gran valor práctico.

El más importante de estos casos especiales ocurre cuando se tienen k factores, cada uno con dos niveles. Estos niveles pueden ser cuantitativos como sería el caso de dos valores de temperatura, presión o tiempo. También pueden ser cualitativos como sería el caso de dos máquinas, dos operadores, los niveles "superior" e "inferior" de un factor o, quizás, la ausencia o presencia de un factor. Una réplica completa de tal diseño requiere que se recopilen $2 \times 2 \times \dots \times 2 = 2^k$ observaciones y se conoce como diseño factorial 2^k . El segundo caso especial es el de k factores con tres niveles cada uno, conocido como diseño factorial 3^k .

En este capítulo se presentan métodos especiales para el análisis de estas dos útiles series de diseños. Se supone, 1) que los factores son fijos, 2) que los diseños son completamente aleatorizados, y 3) que se satisface la suposición usual de normalidad.

El diseño 2^k es particularmente útil en las primeras fases del trabajo experimental, cuando es probable que haya muchos factores por investigar. Conlleva el menor número de corridas con las cuales pueden estudiarse k factores en un diseño factorial completo. Debido a que sólo hay dos niveles para cada factor, debe suponerse que la respuesta es aproximadamente lineal en el intervalo de los niveles elegidos de los factores

2 EL DISEÑO 2^2

El primer diseño de la serie 2^k es aquel que tiene sólo dos factores, A y B , cada uno con dos niveles. Este diseño se conoce como diseño factorial 2^2 . Arbitrariamente, los niveles del factor pueden llamarse "inferior" y "superior". A modo de ejemplo considere una investigación llevada a cabo para estudiar el efecto que tienen la concentración de un

reactivo y la presencia de un catalizador sobre el tiempo de reacción de un proceso químico. Sea la concentración del reactivo el factor A con dos niveles de interés, 15 y 20%. El catalizador constituye el factor B ; el nivel alto (o superior) denota el uso de dos sacos de catalizador y el nivel bajo (o inferior) denota el uso de sólo un saco. El experimento se realiza ("réplica" o "repite") tres veces, y los datos son como sigue:

Combinación de tratamientos	Réplica			Total
	I	II	III	
A bata B baja	28	25	27	80
A alta, B baja	36	32	32	100
A baja, B alta	18	19	23	60
A alta, B alta	31	30	29	90

En la Figura 9-1 se presentan gráficamente las combinaciones de tratamientos para este diseño. Por convención, el efecto de un factor se denota por la letra latina mayúscula. De este modo, " A " se refiere al efecto del factor A , " B " se refiere al efecto del factor B , y " AB " se refiere a la interacción AB . En el diseño 2^2 , los niveles bajo y alto de A y B se denotan por "-" y "+", respectivamente, en los ejes A y B . Así, - en el eje A representa el nivel bajo de concentración (15%), mientras que + representa el nivel alto (25%), y - en el eje B representa el nivel bajo de catalizador mientras que + denota el nivel alto.

Las cuatro combinaciones de tratamientos en el diseño suelen representarse por letras minúsculas, como se muestra en la Fig. 9-1. En esta figura se aprecia que el nivel superior de cualquier factor de una combinación de tratamientos está representado por la presencia de la letra minúscula correspondiente, mientras que la ausencia de esta última representa el nivel inferior del factor. Así a representa la combinación de tratamientos, en la que A se encuentra en el nivel superior y B en el inferior; b representa aquella en la que A se halla en el nivel inferior y B en el superior, y ab representa a ambos factores en el nivel superior. Por convención (1) se usa para representar a ambos factores en el nivel inferior. Esta notación se usará a lo largo de toda la serie 2^k .

El efecto promedio de un factor se define como el cambio en la respuesta producida por un cambio en el nivel de ese factor, promediado sobre los niveles del otro factor. Como se ilustra en la Figura 9-1, las letras minúsculas (1), a , b y ab también se usan para representar

los totales de las n réplicas de las combinaciones de tratamientos correspondientes. Ahora bien, el efecto de A en el nivel inferior de B es $[a - (1)]/n$, mientras que en el nivel

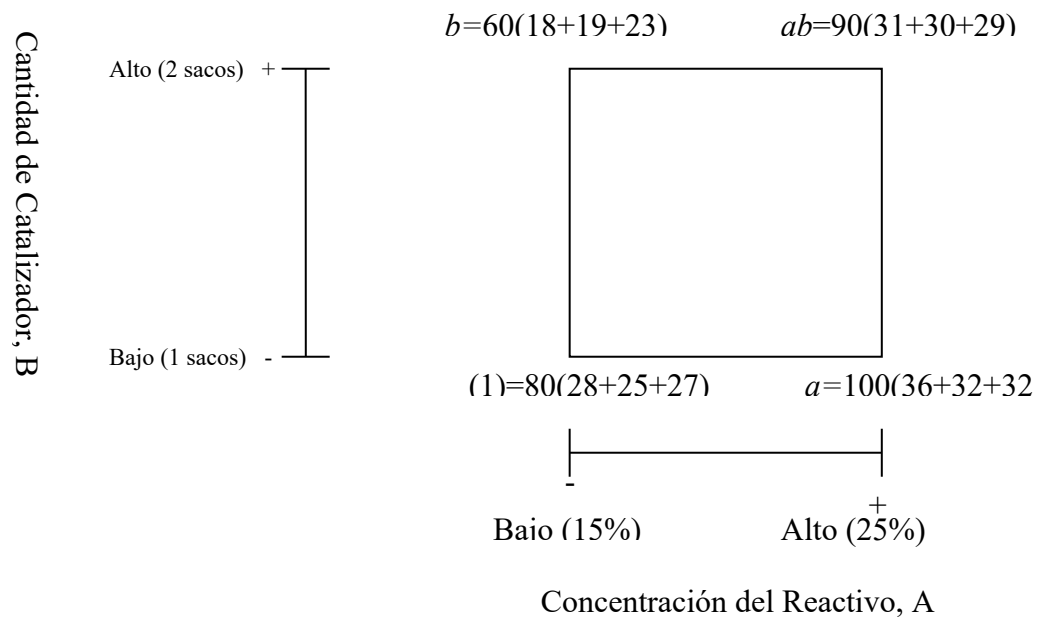


Figura 9-1 Combinaciones de tratamientos en el diseño 2²

superior de B es $[ab - b]/n$. Tomando el promedio de estas dos cantidades se obtiene

$$\begin{aligned}
 A &= \frac{1}{2n} \{ [ab - b] + [a - (1)] \} \\
 &= \frac{1}{2n} [ab + a - b - (1)]
 \end{aligned}
 \tag{9-1}$$

El efecto promedio de B se determina a partir de su efecto en el nivel inferior de A (esto es $[b - (1)]/n$, y de su efecto en el nivel superior de A (que es igual a $[ab - a]/n$) obteniéndose

$$\begin{aligned}
 B &= \frac{1}{2n} \{ [ab - a] + [b - (1)] \} \\
 &= \frac{1}{2n} [ab + b - a + (1)]
 \end{aligned}
 \tag{9-2}$$

El efecto de interacción AB se define como la diferencia promedio entre el efecto de A en el nivel superior de B y su efecto en el nivel inferior de B . Así,

$$\begin{aligned}
 AB &= \frac{1}{2n} \{ [ab - b] - [a - (1)] \} \\
 &= \frac{1}{2n} [ab + (1) - a - b]
 \end{aligned}
 \tag{9-3}$$

Por otro lado se puede definir AB como la diferencia promedio entre el efecto de B en el nivel superior de A y el efecto de B en el nivel inferior de A . Esto conduce a la misma Ecuación 9-3.

Las fórmulas para los efectos de A , B y AB pueden deducirse por otro método. El efecto de A puede hallarse como la diferencia en la respuesta promedio de las dos combinaciones de tratamientos en la mitad derecha de la Fig. 9-1 (que llamaremos \bar{Y}_{A^+} , puesto que es la respuesta promedio para las combinaciones de tratamientos a las que A se encuentra en el nivel alto) y las dos combinaciones de tratamientos en la mitad izquierda (o \bar{Y}_{A^-}) Esto es,

$$\begin{aligned}
 A &= \bar{Y}_{A^+} - \bar{Y}_{A^-} \\
 &= \frac{ab + a}{2n} - \frac{b + 1}{2n} \\
 &= \frac{1}{2n} [ab + a - b - (1)]
 \end{aligned}$$

Éste es exactamente el mismo resultado que en la Ecuación 9-1. El efecto de B (Ecuación 9-2) se encuentra como la diferencia entre el promedio de las dos combinaciones de tratamientos en la parte superior del cuadrado \bar{Y}_{B^+} y el promedio de las dos combinaciones de tratamientos en la parte inferior (\bar{Y}_{B^-}), o

$$\begin{aligned}
 B &= \bar{Y}_{B^+} - \bar{Y}_{B^-} \\
 &= \frac{ab+b}{2n} - \frac{a+1}{2n} \\
 &= \frac{1}{2n} [ab+b-a-(1)]
 \end{aligned}$$

Finalmente, el efecto de interacción AB es el promedio de las combinaciones de tratamientos en la diagonal de derecha a izquierda del cuadrado $[ab \text{ y } (1)]$ menos el promedio de las combinaciones de tratamientos en la diagonal de izquierda a derecha (a y b), o

$$\begin{aligned}
 AB &= \frac{ab+(1)}{2n} - \frac{a+b}{2n} \\
 &= \frac{1}{2n} [ab+(1)-a-b]
 \end{aligned}$$

que es una expresión idéntica a la Ecuación 9-3.

Con los datos que aparecen en la Fig. 9-1, las estimaciones de los efectos promedio son

$$A = \frac{1}{2(3)}(90 + 100 - 60 - 80) = 8.33$$

$$B = \frac{1}{2(3)}(90 + 60 - 100 - 80) = -5.00$$

$$AB = \frac{1}{2(3)}(90 + 80 - 100 - 60) = 1.67$$

El efecto de A (concentración de reactivo) es positivo; esto sugiere que elevar A del nivel bajo (15%) al nivel alto (25%) incrementará el rendimiento. El efecto de B (catalizador) es negativo; esto sugiere que elevar la cantidad de catalizador agregada al proceso reducirá el rendimiento. Al parecer, el efecto de interacción es pequeño comparado con los dos efectos principales.

En muchos experimentos que implican diseños 2^k se examina la *magnitud* y la *dirección* de los efectos de los factores para determinar cuáles variables es probable que

sean importantes. Por lo general puede emplearse el análisis de variancia para confirmar esta interpretación. En el diseño 2^k existen algunos métodos rápidos especiales para realizar los cálculos del análisis de variancia.

Consideremos la suma de cuadrados para A , B y AB . Obsérvese a partir de la Ecuación 9-1 que se utiliza un contraste para estimar A ; esto es,

$$\text{Contraste}_A = ab + a - b - (1) \quad (9-4)$$

Este contraste suele llamarse efecto *total* de A . A partir de las ecuaciones 9-2 y 9-3, puede apreciarse que también se usan contrastes para estimar B y AB . Además, estos tres contrastes son ortogonales. La suma de cuadrados de cualquiera de ellos puede calcularse usando la Ecuación 3-19. Esta ecuación establece que la suma de cuadrados de un contraste es igual al contraste elevado al cuadrado entre el producto del número de las observaciones de cada total del contraste por la suma de los cuadrados de los coeficientes del mismo. En consecuencia, se obtiene que las sumas de cuadrados de A , B y AB son

$$SS_A = \frac{[ab + a - b - (1)]^2}{n(4)} \quad (9-5)$$

$$SS_B = \frac{[ab + b - a - (1)]^2}{n(4)} \quad (9-6)$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{n(4)} \quad (9-6)$$

Con los datos de la Fig. 9-1, las sumas de cuadrados se pueden calcular aplicando las Ecuaciones 9-5, 9-6 Y 9-7, obteniéndose

$$SS_A = \frac{(50)^2}{4(3)} = 208.33$$

$$SS_B = \frac{(-30)^2}{4(3)} = 75.00 \quad (9-8)$$

$$SS_{AB} = \frac{(10)^2}{4(3)} = 8.33$$

La suma total de cuadrados se determina de la manera usual mediante

$$SS_T = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^n y_{ijk}^2 - \frac{Y_{\square\square\square}^2}{4n} \quad (9-9)$$

En general SS_T tiene $4n - 1$ grados de libertad. La suma de cuadrados del error, con $4(n - 1)$ grados de libertad, se puede calcular en la forma usual, por diferencia, mediante

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} \quad (9-10)$$

Con los datos de la Fig. 9-1 Y los resultados de la Ecuación 9-8 se obtiene

$$SS_T = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^n y_{ijk}^2 - \frac{Y_{\square\square\square}^2}{4n} = 9398.00 - 9075.00 = 323.00$$

Y

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} = 323.00 - 208.33 - 75.00 - 8.33 = 31.34$$

El análisis de variancia completo se presenta en la Tabla 9-1. Ambos efectos principales son significativos al 1 %.

A menudo es conveniente escribir las combinaciones de tratamientos en el orden (1), a , b y ab . Este orden se conoce como *orden estándar*. Cuando se utiliza es posible apreciar que los coeficientes de los contrastes usados para estimar los efectos son

Efectos	(1)	a	b	ab
<i>A</i> :	-1	+1	-1	+1
<i>B</i> :	-1	-1	+1	+1
<i>AB</i> :	+1	-1	-1	+1

Tabla 9-1 Análisis de variancia para los datos de la Figura 9-1

Fuente de Variación	Sumas de Cuadrados	Grados de libertad	Media de Cuadrados	F_0
<i>A</i>	208.33	1	208.33	53.15 ^a
<i>B</i>	75.00	1	75.00	19.13 ^a
<i>AB</i>	8.33	1	8.33	2.13
Residual	31.34	8	3.92	
Total	323.00	11		

^a Significativo al 1%

Tabla 9-2 Signos Algebraicos para calcular los efectos en un diseño 2^2

Combinación de Tratamientos	Efecto Factorial			
	<i>I</i>	<i>A</i>	<i>B</i>	<i>AB</i>
(1)	+	-	-	+
<i>a</i>	+	+	-	-
<i>b</i>	+	-	+	-
<i>ab</i>	+	+	+	+

Obsérvese que los coeficientes de los contrastes usados para estimar la interacción son iguales al producto de los coeficientes correspondientes a los dos efectos principales. Los coeficientes de los contrastes siempre son + 1 o -1 y se puede usar una tabla de signos positivos y negativos como la mostrado en la Tabla 9-2 para determinar el signo apropiado de cada combinación de tratamientos. En el encabezado de las columnas de la Tabla 9-2 se encuentran los efectos principales (*A* y *B*), la interacción *AB*, e *I*, que representa el total o el

promedio de todo el experimento. Se observa que la columna encabezada por I se compone sólo de signos positivos. Los renglones corresponden a las combinaciones de tratamientos. Para encontrar un contraste con el fin de estimar cualquier efecto, simplemente se multiplican los signos de la columna apropiada de la tabla por la correspondiente combinación de tratamientos, y se suman. Por ejemplo, el contraste para estimar A es $-(1) + a - b + ab$, lo cual concuerda con la ecuación 9-1.

Análisis de Residuos. Es fácil calcular los residuos de un diseño 2^k por medio de un modelo de regresión. Para el experimento del proceso químico, el modelo de regresión es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

donde: x_1 es una variable codificada que representa la concentración de reactivo y x_2 es otra variable codificada que representa la cantidad de catalizador; las β son coeficientes de regresión. La relación entre las variables *naturales* (la concentración de reactivo y la cantidad de catalizador) y las variables codificadas está dada por

$$x_1 = \frac{Conc - (Conc_{baja} + Conc_{alta})/2}{(Conc_{alta} - Conc_{baja})/2}$$

$$x_2 = \frac{Catal - (Catal_{baja} + Catal_{alta})/2}{(Catal_{alta} - Catal_{baja})/2}$$

Cuando las variables naturales sólo tienen dos niveles, esta codificación produce la familiar notación ± 1 para los niveles de las variables codificadas. A fin de ilustrar esto en el caso de nuestro ejemplo, obsérvese que.

$$x_1 = \frac{Conc - (15 + 25)/2}{(25 - 15)/2}$$

$$= \frac{Conc - 20}{5}$$

Por tanto, si la concentración se encuentra en el nivel alto ($Conc = 25\%$), entonces $x_1 = +1$, mientras que si está en su nivel, bajo ($Conc = 15\%$), entonces $x_1 = -1$, Además,

$$x_2 = \frac{Catal - (1+2)/2}{(2-1)/2} = \frac{Catal - 1.5}{0.5}$$

De este modo, si el catalizador se halla en el nivel alto ($Catal = 2$ sacos), entonces $x_2 = +1$, mientras que si está en el nivel bajo ($Catal = 1$ saco), entonces $x_2 = -1$,

El modelo de regresión ajustado es

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)x_1 + \left(\frac{-5.00}{2}\right)x_2$$

donde la ordenada al origen es el gran promedio de las 12 observaciones, y los coeficientes de regresión $\hat{\beta}_1$ y $\hat{\beta}_2$ son un medio de las estimaciones de los efectos de los factores correspondientes. La causa de que el coeficiente de regresión sea la mitad de la estimación del efecto es que un coeficiente de regresión mide el efecto de un cambio unitario en x sobre la media de y , y la estimación del efecto se basa en un cambio de dos unidades (de -1 a +1).

Este modelo puede utilizarse para generar los valores predichos de y y en los cuatro puntos del diseño. Por ejemplo, cuando la concentración de reactivo se encuentra en el nivel bajo ($x_1 = -1$) y el catalizador está también en el nivel bajo ($x_2 = -1$), el rendimiento predicho es

$$\begin{aligned}\hat{y} &= 27.5 + \left(\frac{8.33}{2}\right)(-1) + \left(\frac{-5.00}{2}\right)(-1) \\ &= 25.832\end{aligned}$$

Existen tres observaciones para esta combinación de tratamientos, y los residuos son:

$$e_1 = 28 - 25.835 = 2.165$$

$$e_2 = 25 - 25.835 = -0.835$$

$$e_3 = 27 - 25.835 = 1.165$$

Los demás valores y residuos predichos se calcula de modo similar. Para el nivel alto de concentración de reactivo y el nivel bajo de catalizador

$$\begin{aligned}\hat{y} &= 27.5 + \left(\frac{8.33}{2}\right)(+1) + \left(\frac{-5.00}{2}\right)(-1) \\ &= 34.165\end{aligned}$$

$$e_4 = 36 - 34.165 = 1.835$$

$$e_5 = 32 - 34.165 = -2.165$$

$$e_6 = 32 - 34.165 = -2.165$$

Para el nivel bajo de concentración de reactivo y el nivel alto de catalizador:

$$\begin{aligned}\hat{y} &= 27.5 + \left(\frac{8.33}{2}\right)(-1) + \left(\frac{-5.00}{2}\right)(+1) \\ &= 20.835\end{aligned}$$

$$e_7 = 18 - 20.835 = -2.835$$

$$e_8 = 19 - 20.835 = -1.835$$

$$e_9 = 23 - 20.835 = 2.165$$

Finalmente, para el nivel alto de ambos factores:

$$\begin{aligned}\hat{y} &= 27.5 + \left(\frac{8.33}{2}\right)(+1) + \left(\frac{-5.00}{2}\right)(+1) \\ &= 29.165\end{aligned}$$

$$e_{10} = 31 - 29.165 = 1.835$$

$$e_{11} = 30 - 29.165 = 0.835$$

$$e_{12} = 29 - 29.165 = -0.165$$

En la Fig. 9-2 se presenta una gráfica de probabilidad normal de estos residuos contra los dos factores, concentración de reactivo y cantidad de catalizador. Al parecer estas gráficas indican resultados satisfactorios, por lo que no tenemos motivos para sospechar que existan problemas en cuanto a la validez de nuestras conclusiones.

