Capítulo 10

Confusión en experimentos factoriales

10.1. Introducción

Los diseños factoriales 2^k y 3^k son de gran importancia práctica, como se encuentra en el capítulo 9, estos arreglos se emplean ampliamente en los estudios de investigación como un medio para tratar los efectos de varios factores de tratamiento en forma simultánea en el mismo experimento. Los factoriales 2^k y 3^k tienen k factores a dos y tres niveles, respectivamente; conforme el número de factores aumenta, el número de combinaciones de tratamientos aumenta con rapidez, cuando esto sucede se debe hacer uso de los diseños en bloques incompletos para controlar el error experimental.

El uso de bloques incompletos para garantizar homogeneidad dentro de las unidades experimentales cuando hay arreglos con muchos tratamientos obliga a hacer uso de las siguientes técnicas de reducción del tamaño:

i. Confusión

ii. Replicación fraccionada con el fin de disminuir costos en el experimento.

El principio de confusión tiene la base en que ciertas interacciones de poca importancia se pueden sacrificar, de manera que la imprecisión resultante del uso de bloques grandes y heterogéneos se concentra en esas interacciones (las confundidas) en vez de afectar el resto de los efectos e interacciones que son de mayor interés para la investigación.

El principio de confusión consiste en formar bloques incompletos de tal modo que los efectos de interés sean ortogonales con bloques y que algunos efectos o interacciones de poco interés práctico queden confundidos con bloques. La idea original de éste análisis fue propuesta por Yates (1934).

De esta manera, la *confusión* es una técnica de diseño mediante la cual un experimento factorial completo se distribuye en bloques, donde el tamaño del bloque es menor que el número de combinaciones de los tratamientos de una réplica.

La confusión hace uso de la teoría de bloque generador, grupos y campos de Galois (ver anexo 9.6.2), puesto que los tratamientos que forman un bloque se determinan con base en éstos desarrollos matemáticos.

La siguientes definiciones son importantes para el entendimiento de los resultados presentados a lo largo de este capítulo.

Definición 10.1. Se llama *réplica* del experimento a un conjunto de bloques en el cual están todos los tratamientos una sola vez.

Definición 10.2. Si un efecto esta confundido en todas las réplicas, se dice que hay *confusión total*. Acá necesariamente hay sacrificio de efectos principales o interacciones (que fueron confundidos con bloques).

Definición 10.3. Cuando un efecto o interacción se confunde con bloques en algunas replicaciones, se dice que hay *confusión parcial*. En éste caso se tiene información de todo el conjunto de tratamientos en el arreglo factorial.

En algunas áreas de la experimentación, como en la agronomía, biología e industria, se pueden tener varios tamaños de UE. Si un bloque es una UE más grande con error experimental mayor que las UE que lo constituyen, se podría estudiar efectos totalmente confundidos, aunque con menor grado de precisión a los otros.

Definición 10.4. Las variaciones que se consideran entre bloques se denominan *error interbloque*. Al error que proviene de la variación entre UE dentro de bloques se le llama *error intrabloque*.

Las interacciones entre dos factores se llaman de primer orden; las de tres factores se llaman de segundo orden y así sucesivamente.

Está claro que para la conformación de los bloques, conviene confundir los efectos que no interesen estudiar o no tengan interés práctico en la investigación. Como las interacciones de orden superior por lo general son negligibles, estas deben ser confundidas con los bloques. Además, en el caso de experimentos con muchos factores, los efectos principales, las interacciones de dos factores y otras interacciones de menor orden son los efectos de mayor interés. Al confundir las interacciones de orden más alto los demás efectos se estiman sin penalización.

Una justificación al hecho anterior está amparada por la serie de expansión de Taylor. En este caso, si y es la característica en estudio, la cual está en función de los niveles de los factores, es decir $f(a, b, c, \dots)$, al expandir, en series de Taylor, esta función puede ser escrita como

$$\begin{aligned} y = f(a, b, c, \dots) = \mu &+ (\alpha_1 a + \beta_1 b + \gamma_1 c + \cdots) + (\alpha_2 a^2 + \beta_2 b^2 + \gamma_2 c^2 + \cdots) \\ &+ (\alpha\beta)_{11} ab + (\alpha\gamma)_{11} ac + \cdots + residuo \end{aligned}$$

donde; μ es el valor medio de los factores, $\alpha_1 a + \beta_1 b + \gamma_1 c + \cdots$ es la magnitud de los efectos lineales y en $(\alpha_2 a^2 + \beta_2 b^2 + \gamma_2 c^2 + \cdots) + (\alpha\beta)_{11} ab + (\alpha\gamma)_{11} ac + \cdots$ se tienen los efectos cuadráticos y todas las interacciones de primer, segundo, ... orden.

Finalmente, se tiene un residuo cuyo valor depende de la aproximación que se requiera al verdadero valor y, como los residuos convergen a cero es claro que en valor absoluto cada término de orden superior que se agrega a la expansión en la serie es cada vez menor.

Puede observarse que si el desarrollo en serie del valor de y, incluye solamente los efectos lineales, tal consideración conduce a una primera aproximación del valor con la característica en estudio, un mayor refinamiento se produce cuando se incluyen los efectos cuadráticos y las interacciones de segundo orden, etc.

Como en los experimentos factoriales 2^k no es posible estimar efectos cuadráticos, los efectos lineales son de hecho los factores principales, el término dentro de los paréntesis contendría las interacciones de dos factores, de tres factores, etc. De aquí es evidente considerar más importante los efectos principales de los factores que las interacciones de primer orden, éstas últimas más importantes que las de segundo orden, y así sucesivamente.

10.2. Confusión en series 2^k

En esta sección, se considera la construcción y el análisis del sistema factorial 2^k en 2^p bloques incompletos, donde $p < k$. Estos diseños pueden correrse en dos bloques, en cuatro bloques, en ocho bloques y más, según los criterios de confusión.

10.2.1. Confusión del diseño factorial 2^k en dos bloques

Para una mayor compresión de este procedimiento se ilustra con un diseño factorial 2^3. Si en este arreglo, bloques de ocho UE homogéneos no son factibles, se puede trabajar con bloques de cuatro UE que sí sean homogéneos. De esta forma, los tratamientos se asignarán a los bloques según el siguiente esquema:

Bloque I		Bloque II	
a	100	(1)	000
b	010	ab	110
c	001	ac	101
abc	111	bc	011
$(ABC)_{i+j+k=1}$		$(ABC)_{i+j+k=0}$	

El esquema anterior se llama un *esquema básico*, este se replica R-veces, generando $2R$ bloques y aleatorizando los tratamientos en los bloques. El modelo usual es el de bloques aleatorizados, dado por

$$y_{sl} = \mu + \tau_s + \delta_l + e_{sl} \tag{10.1}$$

con $s = 1, \ldots, t$, $l = 1, \ldots, 2R$ y, μ, τ_s y δ_l los efectos de la media, tratamientos que se descomponen en efectos principales e interacciones y bloques, respectivamente.

El anterior modelo se puede expresar como

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \delta_l + e_{ijkl} \tag{10.2}$$

en términos de (10.1),

$$\tau_s = \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}, \quad i, j, k = 0, 1; \quad l = 1, 2, \ldots, 2R.$$

Otra forma adicional de presentar el modelo (10.1) es

$$y_{ijkhd} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \eta_h + \omega_{d(h)} + e_{ijkhd} \tag{10.3}$$

donde, $\delta_l = \eta_h + \omega_{d(h)}$ con $h = 1, 2, \ldots, R$ y $d = 1, 2$.

La variación entre $A_{i=0}$ y $A_{i=1}$ es la misma que se tendría sin efecto de bloques. Por lo tanto, el factor A se estima como

$$\begin{aligned} A &= \frac{1}{4R}[110 + 101 + 100 + 111 - 000 - 010 - 001 - 011] \mod 2 \\ &= \frac{1}{4R}[A_{i=1} - A_{i=0}] \mod 2. \end{aligned}$$

Es decir, la estimación del factor A no se ve afectada por el hecho de tener bloques incompletos. El estimador de A es ortogonal a bloques, en el sentido de que dos tratamientos en el bloque I tienen signo negativo y dos tratamientos tienen signo positivo, lo mismo sucede en el bloque II, por lo tanto el estimador de A no contiene efectos aditivos de bloques.

De igual manera B y C son ortogonales con ABC y en consecuencia con los bloques.

Para ilustrar la ortogonalidad con las interacciones de dos factores, se considera la interacción AB,

$$SC(AB) = \frac{[(AB)_{i+j=0}]^2 + [(AB)_{i+j=1}]^2]}{4R} - \frac{y^2_{....}}{8R}$$

con $(AB)_{i+j=0} = (000+110+001+111)$ y $(AB)_{i+j=1} = (100+010+101+011)$. En estos totales se tienen $2\sum_{l=1}^{2R}\delta_l$ entonces no se afectará la variabilidad entre los totales $(AB)_0$ y $(AB)_1$ ya que ambos se ven incrementados por $2\sum_{l=1}^{2R}\delta_l$.

De igual forma, resultarán ortogonales con bloques las interacciones AC y BC. La confusión se ve claramente al considerar los totales que miden el efecto ABC, es decir,

$$\begin{aligned}
ABC &= \frac{1}{4R}[100+010+001+111-000-110-101-011] \\
&= \frac{1}{4R}[a+b+c+abc-(1)-ab-ac-bc] \\
&= \frac{1}{4R}[(ABC)_{i+j+k=1} - (ABC)_{i+j+k=0}]
\end{aligned}$$

y el estimador de la diferencia entre bloques es

$$Bloques = \frac{1}{4R}\left[\sum bloques_1 - \sum bloques_2\right].$$

Las estimaciones en las dos anteriores expresiones coinciden, lo cual muestra que el efecto de la interacción triple ABC, queda confundido con el bloque en el modelo (10.2). La varianza de estimación de cualquiera de los efectos factoriales es $\frac{(\sigma^2)'}{2R}$.

Nota. Si el experimento se planea en bloques completamente aleatorizados todos los efectos factoriales serían estimables, y la varianza de estimación de cualquiera de ellos en $\frac{\sigma^2}{2R}$, donde σ^2 es la varianza por UE cuando el experimento es DBCA.

Si $(\sigma^2)' \geq \sigma^2$, la eficiencia relativa (ER) de estimación del DBI contra el DBCA, se define por el cociente

$$ER = \frac{(\sigma^2)'}{\sigma^2}.$$

En general se tiene mayor precisión de estimación en DBI a causa de la disminución de la varianza intrabloque (dentro de bloques); al tener en cada bloque menos UE's, entonces es más homogéneo. Sin embargo, debe tenerse claro que así sean más eficientes los bloques incompletos, conducen a cierta pérdida de información con relación a los efectos factoriales de interés.

Las sumas de cuadrados se calculan de la manera usual para el análisis de varianza excepto que se excluye la partición de la suma de cuadrados para el efecto de interacción confundido con los bloques. La suma de cuadrados del bloque incluirá el efecto factorial confundido.

Las fuentes de variación y los grados de libertad para el análisis de varianza en el factorial 2^3 con 2 bloques incompletos en cada una de las R réplicas se presentan en la tabla 10.1. Como ABC se confunde con los bloques, la suma de cuadrados para los bloques incluye el efecto ABC.

Usualmente la variación entre bloques no se descompone como se indica en la tabla 10.1; sin embargo, en algunos casos, se requiere tener información sobre la interacción ABC, aunque con mucho menos precisión que sin usar la confusión. Así se descomponen los grados de libertad y las sumas de cuadrados de bloques en la forma indicada en dicha tabla. La SCE interbloque se obtiene por diferencia, esto es

$$SCE(\text{interbloque}) = SC(Bloques) - SC(Replicas) - SC(ABC).$$

El error interbloque es el que se genera por la variación entre bloques dentro de réplicas. Esta descomposición casi no se usa, excepto cuando se confunden efectos principales generando el llamado arreglo en parcelas divididas, diseño que se discute con mayor detalle en el capítulo 11.

C. de V.	gl
Bloques	$2R-1$
Réplicas	$R-1$
Bloques(Réplicas)	$R(b-1)=R$
ABC	1
Error Interbloque	
Réplica × ABC	$R-1$
Tratamientos	6
A	1
B	1
AB	1
C	1
AC	1
BC	1
Error intrabloque	$6(R-1)$
Total	$8R-1$

Tabla 10.1. Análisis de varianza para un factorial 2^3 con $b = 2$ bloques incompletos en cada una de los R grupos de réplicas.

***Ejemplo* 10.1.** *Una investigación busca mejorar la pureza de un producto químico, para ello se involucra la influencia de tres factores: Tasa de agitación (A), concentración del compuesto base (B) y concentración del reactivo (C). El químico estableció un experimento con un diseño factorial con factores a dos niveles para un arreglo factorial* 2^3.

Se realizaron tres réplicas del experimento, pero sólo podía realizar cuatro corridas del proceso químico en un día, esto llevó a que cada réplica debía correrse en dos días diferentes (bloques).

En la conducción del experimento, se construyo un diseño de bloques incompletos confundiendo la interacción de orden tres ABC *con los bloques. En este caso el contraste de definición es* $L = x_1 + x_2 + x_3$ *y el diseño se ilustra en la figura 10.1.*

En la figura 10.1, $\circ$ *representa las corridas con* $(ABC)_0$ *y* $\bullet$ *las corridas con* $(ABC)_1$.

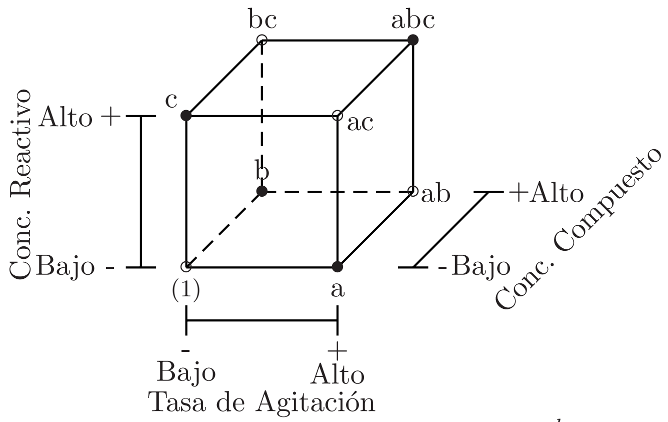

Figura 10.1. Vista geométrica del diseño factorial 2^k en dos bloques.

Las combinaciones de tratamientos dentro de cada bloque se muestran en la tabla 10.2.

Réplica I		Réplica II		Réplica III	
Bloque 1 $(ABC)_1$	Bloque 2 $(ABC)_0$	Bloque 3 $(ABC)_0$	Bloque 4 $(ABC)_1$	Bloque 5 $(ABC)_0$	Bloque 6 $(ABC)_1$
001 : 44,5	000 : 46,8	101 : 49,8	001 : 55,5	011 : 53,2	100 : 69,5
010 : 44,2	011 : 44,5	110 : 52,0	100 : 59,8	101 : 57,2	010 : 62,8
100 : 60,1	101 : 57,0	011 : 48,8	010 : 56,0	000 : 56,0	001 : 55,0
111 : 48,8	110 : 58,5	000 : 51,5	111 : 58,5	110 : 59,0	111 : 53,8

Tabla 10.2. Arreglo de tratamientos para la pureza observada de un producto químico en un factorial 2^3 confundido totalmente.

En este experimento, se tienen dos pasos en el proceso de la aleatorización:

a) Decidir qué bloques llevan $(ABC)_0$ y qué bloques llevan $(ABC)_1$, y

b) Decidir la aleatorización de los tratamientos dentro de cada bloque.

Con los datos presentados en la tabla 10.2 se realiza el análisis de varianza usual. El efecto de tratamientos se descompone en seis contrastes ortogonales que representan los efectos principales y las interacciones, como se muestra

	Totales de tratamiento									
	000	001	010	011	100	101	110	111		
Efecto	154,3	155.0	163.0	146.5	189.4	164.0	169.5	161.1	$\sum +$	$\sum -$
Total	+	+	+	+	+	+	+	+	1302,8	
A	-	-	-	-	+	+	+	+	684,0	618,8
B	-	-	+	+	-	-	+	+	640,1	662,7
AB	+	+	-	-	-	-	+	+	639,9	662,9
C	-	+	-	+	-	+	-	+	626,6	676,2
AC	+	-	+	-	-	+	-	+	642,4	660,4
BC	+	-	-	+	+	-	-	+	651,3	651,5
ABC	-	+	+	-	+	-	-	+	668,5	634,3

Tabla 10.3. Arreglo de signos para el diseño del ejemplo 10.1.

en la tabla 10.3.

Al utilizar los totales de las combinaciones de los tratamientos que se muestran en la tabla 10.3, los efectos de los factores pueden estimarse como en un diseño 2^k, usando los procedimientos presentados en el capítulo 9.

El efecto del bloque confundido con la interacción ABC se calcula por la diferencia en la respuesta promedio entre los dos bloques, a través de todas las réplicas.

$$\begin{aligned} \textit{Efecto del bloque} &= \bar{y}_{\textit{Bloque 1}} - \bar{y}_{\textit{Bloque 2}} = \frac{1}{12}(ABC_1 - ABC_0) = ABC \\ &= \frac{1}{12}(668{,}5 - 634{,}3) = 2{,}85. \end{aligned}$$

En la tabla 10.4 se resume el análisis de varianza de este experimento. Las sumas de cuadrados de cada uno de los efectos no confundidos se obtienen con los resultados del capítulo 9, estos fueron los siguientes:

$$
\begin{aligned}
SC(Bloques) =& \sum_{l=1}^{B} \frac{B_l^2}{2^{k-1}} - \frac{y_{\cdots\cdot}^2}{2^k R}\\
=& \frac{197{,}6^2 + 206{,}8^2 + 202{,}1^2 + 229{,}8^2 + 225{,}4^2 + 241{,}1^2}{4}\\
& - \frac{1302,8^2}{24} = 379{,}378
\end{aligned}
$$

$$
\begin{aligned}
SC(Répli) =& \sum_{h=1}^{R} \frac{R_h^2}{2^k} - \frac{y_{\cdots\cdot\cdot}^2}{2^k R}\\
=& \frac{404{,}4^2 + 431{,}9^2 + 466{,}5^2}{8} - \frac{1302,8^2}{24} = 242{,}076
\end{aligned}
$$

$$
\begin{aligned}
SC(Bloque(Répli)) =& \sum_{h=1}^{R}\left(\sum_{d=1}^{b} \frac{y_{\cdots hd}^2}{2^{k-1}} - \frac{y_{\cdots h\cdot}^2}{2^k}\right)\\
=& \frac{197{,}6^2 + 206{,}8^2}{4} - \frac{404{,}4^2}{8} + \frac{202{,}1^2 + 229{,}8^2}{4} - \frac{431{,}9^2}{8}\\
& + \frac{225{,}4^2 + 241{,}1^2}{4} - \frac{466{,}5^2}{8} = 137{,}303\\
SC(ABC) =& 2^2(3)(2{,}85)^2 = 48{,}735\\
SC(Répli \times ABC) =& SC(Bloque) - SC(Répli) - SC(ABC) = 88,567\\
SC(Total) =& 44{,}5^2 + 44{,}2^2 + \cdots + 53{,}8^2 - \frac{1302,8^2}{24} = 893{,}633
\end{aligned}
$$

A partir de los resultados obtenidos en la tabla 10.4, se concluye que hay diferencias entre los bloques al igual que entre tratamientos. Esta última es causada únicamente por el efecto que los factores tasa de agitación (A) y concentración del reactivo (C) tienen sobre la pureza del producto químico, los demás efectos principales e interacciones no son significativos. Adicionalmente, las diferencias que se ocasionan en la pureza se ven influenciadas en mayor parte, por la diferencia que hay entre los niveles de la tasa de agitación, más que por los niveles de concentración del reactivo.

C. de V.	gl		SC	CM	F
Bloques	5		379,379	75,875	5,12
Réplicas		2	242,076	121,038	
Bloques(Réplicas)		3	137,303	45,767	
ABC		1	48,735	48,735	
Error Interbloque					
Réplica $\times$ ABC		2	88,567	44,283	
Tratamientos	6		336,46	56,076	3,78
A		1	177,127	177,127	11,95
B		1	21,282	21,282	1,44
AB		1	22,042	22,042	1,49
C		1	102,507	102,507	6,92
AC		1	13,500	13,500	0,91
BC		1	0,002	0,002	0,00
Error intrabloque	12		177,797	14,816	
Total	23		893,633		

Tabla 10.4. Análisis de varianza para los datos de la pureza de un producto químico en un factorial 2^3 confundido totalmente.

10.2.2. Confusión del diseño factorial 2^k en cuatro bloques

Es posible construir diseños factoriales 2^k confundido en cuatro bloques con 2^{k-2} observaciones cada uno. Estos diseños son particularmente útiles en situaciones en las que el número de factores es moderadamente grande, por ejemplo $k \geq 4$, y el tamaño de los bloques es relativamente pequeño.

Como un ejemplo, para ilustrar el procedimiento, considere el diseño factorial 2^4. Si cada bloque incluye únicamente cuatro tratamientos, entonces debe utilizarse cuatro bloques. La construcción de este diseño es relativamente simple. Se confunde con el efecto de bloque las interacciones ABC y BCD, estos efectos tienen dos contrastes de definición asociadas con ellos

$$\begin{aligned} L_1 &= x_1 + x_2 + x_3 \\ L_2 &= x_2 + x_3 + x_4 \end{aligned}$$

El total de tratamientos se presenta en el siguiente arreglo, a partir del cual se genera la primera combinación de tratamientos que queda asignada al primer bloque (bloque generador):