Diseños factoriales fraccionados

Kevin Steven García[a], Alejandro Vargas[b], Alejandro Soto[c]

1. Introducción

Para entender y abordar el tema correspondiente a los diseños factoriales fraccionados, debemos tener claro el concepto y funcionamiento de un diseño factorial en general, por ello, se dará una breve introducción sobre este tipo de diseño, ya que este es la base de nuestro tema en cuestión.

Los diseños factoriales surgen ante la necesidad de estudiar conjuntamente varios factores obedeciendo a la posibilidad de que el efecto de un factor cambie según los niveles de otros factores, esto es, que los factores interactúen. Estos diseños también se usan cuando se quiere optimizar la respuesta o variable dependiente, es decir, se quiere encontrar la combinación de niveles de los factores que producen un valor óptimo de la variable dependiente.
Si se investiga un factor por separado, el resultado puede ser diferente al estudio conjunto y es mucho más difícil describir el comportamiento general del proceso o encontrar el óptimo.

2. Diseños factoriales

Un diseño factorial es un tipo de experimento cuyo diseño permite estudiar los efectos que varios factores pueden tener en una respuesta. Al realizar un experimento, variar los niveles de todos los factores al mismo tiempo en lugar de uno a la vez, permite estudiar las interacciones entre los factores.

Los diseños factoriales producen experimentos más eficientes, pues cada observación proporciona información sobre todos los factores, y es factible ver las respuestas de un factor en diferentes niveles de otro factor en el mismo experimento. La respuesta a cualquier factor observado en diferentes condiciones indica si los factores actúan en las unidades experimentales de manera independiente. La interacción entre factores ocurre cuando su actuación no es independiente.

Suponga que tenemos dos factores, A con a niveles y B con b niveles. Las observaciones de un experimento factorial pueden describirse con un modelo. Hay varias formas de escribir el modelo de un experimento factorial. La forma más utilizada es el modelo de los efectos:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., n \end{cases}$$

Donde μ es el efecto promedio global, τ_i es el efecto del nivel i-ésimo del factor A de los renglones, β_j es el efecto del nivel j-ésimo del factor B de las columnas, $(\tau\beta)_{ij}$ es el efecto de la interacción entre τ_i y β_j, y ε_{ijk} es un componente del error aleatorio. Puesto que hay n replicas del experimento, hay abn observaciones en total.

El efecto de un factor se define como el cambio en la respuesta producido por un cambio en el nivel del factor, a este efecto se le conoce como efecto principal porque se refiere a los factores de interés primario en el experimento. Entonces si tenemos dos factores A y B, como en la siguiente tabla:

[a]Código: 1533173. E-mail: kevin.chica@correounivalle.edu.co
[b]Código: 1525953. E-mail: jose.alejandro.vargas@correounivalle.edu.co
[c]Código: 1532457. E-mail: asotomurillo@gmail.com

TABLA 1

	Factor A	
Factor B	**bajo**	**alto**
bajo	a	b
alto	c	d

El efecto principal del factor A es la diferencia entre la respuesta promedio con el nivel alto de A y la respuesta promedio con el nivel bajo de A. Esto es:

$$A = \frac{b+d}{2} - \frac{a+c}{2}$$

Análogamente, el efecto principal del factor B es:

$$B = \frac{c+d}{2} - \frac{a+b}{2}$$

Y, el efecto de la interacción se define como la diferencia promedio de los dos efectos del factor A, es decir:

$$AB = \frac{(d-c)-(b-a)}{2}$$

En el diseño factorial de dos factores, ambos factores, A y B como la interacción AB son de igual interés, por lo que se prueban las hipótesis acerca de la igualdad de los efectos de los tratamientos de cada factor y de su interacción. Se tienen 3 hipótesis:

$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$ $H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$ $H_0 : (\tau\beta)_{ij} = 0$ para todas las i, j

H_1 :Al menos una $\tau_i \neq 0$ H_1 :Al menos una $\beta_j \neq 0$ H_1 :Al menos una $(\tau\beta)_{ij} \neq 0$

Todas las hipótesis anteriores se pueden probar realizando el análisis de varianza respectivo.

3. Diseños factoriales 2^k

Existen varios casos especiales del diseño factorial general, los cuales son muy importantes debido a su uso generalizado en el trabajo de investigación. El más importante de estos casos especiales es el diseño de k factores, donde cada factor cuenta con sólo dos niveles. Estos niveles pueden ser tanto cuantitativos como cualitativos. Una réplica completa de este diseño requiere $2x2x\cdots x2 = 2^k$ observaciones y se le llama diseño factorial 2^k.

El diseño 2^k es de particular utilidad en las etapas iniciales del trabajo experimental, cuando probablemente se estén investigando muchos factores. Este diseño proporciona el menor número de corridas con las que pueden estudiarse k factores en un diseño factorial completo.

El modelo estadístico para un diseño 2^k incluiría k efectos principales, $\binom{k}{2}$ interacciones de dos factores, $\binom{k}{3}$ interacciones de tres factores,..., y una interacción de k factores. Es decir, para un diseño 2^k el modelo completo contendría $2^k - 1$ efectos.

El enfoque general para el análisis estadístico del diseño 2^k se resume en los siguientes pasos:

1. Estimar los efectos de los factores
2. Formar el modelo inicial
3. Realizar las pruebas estadísticas
4. Refinar el modelo
5. Analizar los residuales
6. Interpretar los resultados

En este diseño, la tabla estándar se construye empezando con el primer factor con el signo(-) y alterna signos (-) y (+). El segundo factor cambia de signo cada dos observaciones (2^1), el tercer factor cada cuatro (2^2) y el factor k-ésimo cada 2^{k-1}

observaciones. las interacciones de los factores tienen los signos que se obtienen de multiplicar los signos de los factores implicados. Por ejemplo, en el diseño 2^3 la tabla estándar con las interacciones es:

Tabla 2

A	**B**	**C**	**AB**	**AC**	**BC**	**ABC**	**Respuesta y**
-	-	-	+	+	+	-	$y_{111} = (1)$
+	-	-	-	-	+	+	$y_{211} = a$
-	+	-	-	+	-	+	$y_{121} = b$
+	+	-	+	-	-	-	$y_{221} = ab$
-	-	+	+	-	-	+	$y_{112} = c$
+	-	+	-	+	-	-	$y_{212} = ac$
-	+	+	-	-	+	-	$y_{122} = bc$
+	+	+	+	+	+	+	$y_{222} = abc$

En la tabla se logra apreciar que el nivel alto de cualquiera de los factores en una combinación de tratamientos se denota por la letra minúscula correspondiente y que el nivel bajo de un factor en una combinación de tratamientos se denota por la ausencia de la letra respectiva. Por lo tanto, para un diseño 2^2 a representa la combinación de tratamientos con A en el nivel alto y B en el nivel bajo, b representa A en el nivel bajo y B en el nivel alto, y ab representa ambos factores en el nivel alto. Por convención, se usa (1) para denotar que ambos factores están en el nivel bajo. Esta notación se utiliza en todos los diseños 2^k.
Ahora, teniendo en cuenta la notación anterior y que n es el número de réplicas hechas con la combinación de los tratamientos.Para estimar un efecto o calcular la suma de cuadrados de un efecto, primero debe determinarse el contraste asociado con ese efecto. En general, el contraste del efecto $AB\cdots K$ se determina expandiendo el miembro derecho de

$$Contraste_{AB\cdots K} = (a \pm 1)(b \pm 1)\cdots(k \pm 1)$$

Una vez se han calculado los contrastes de los efectos, pueden estimarse los efectos y calcular las sumas de cuadrados de acuerdo con

$$AB\cdots K = \frac{2}{n2^k}(Contraste_{AB\cdots K})$$

y

$$SS_{AB\cdots K} = \frac{1}{n2^k}(Contraste_{AB\cdots K})^2$$

Entonces el análisis de varianza de un diseño 2^k es

Tabla 3

Fuente de variación	**Suma de cuadrados**	**Grados de libertad**
A	SS_A	1
B	SS_B	1
$\vdots$	$\vdots$	$\vdots$
K	SS_K	1
AB	SS_{AB}	1
AC	SS_{AC}	1
$\vdots$	$\vdots$	$\vdots$
JK	SS_{JK}	1
ABC	SS_{ABC}	1
ABD	SS_{ABD}	1
$\vdots$	$\vdots$	$\vdots$
IJK	SS_{IJK}	1
$\vdots$	$\vdots$	$\vdots$
$ABC\cdots K$	$SS_{ABC\cdots K}$	1
Error	SS_E	$2^k(n-1)$
Total	SS_T	$n2^k - 1$

4. Diseños factoriales fraccionados

En un diseño factorial conforme el número de factores del experimento crece, el número de casillas o condiciones experimentales (y por lo tanto el número de lecturas o pruebas necesarias), crece exponencialmente, también crece el número de efectos a evaluar (interacciones principalmente) y el número de efectos y casillas varía con el número de factores en una relación como se muestra en la tabla siguiente para un experimento factorial 2^k.

Tabla 4: Add caption

No. De factores	No. De casillas	Efectos principales	Interacciones entre factores						
			2	3	4	5	6	7	8
2	4	2	1						
3	8	3	3	1					
4	16	4	6	4	1				
5	32	5	10	10	5	1			
6	64	6	15	20	15	6	1		
7	128	7	21	35	35	27	7	1	
8	256	8	28	58	70	56	28	8	1

Observando esto cuando se tienen 8 factores, existen 256 condiciones experimentales esto implica que al hacer una replicación de todo el experimento se requiere un numero de observaciones igual a 256 y esto en caso de que se decida tomar una replica por celda, si se deciden tomar dos estaríamos hablando de 512 observaciones la cual es una cantidad excesiva.

Por otro lado, se necesitan 128 observaciones para un experimento con 7 factores por que se deben evaluar 127 posibles efectos (que son los grados de libertad totales en 128 observaciones) de estos efectos 7 son los factores principales, 21 interacciones de 2 factores, 35 de tres, 35 de cuatro, 27 de cinco en cinco, 7 de seis en seis y una interacción de 7 factores. En general el número de interacciones de k factores tomados r en r es:

$$kCr = \frac{k!}{r!(k-r)!}$$

El concepto de replicación fraccionada se basa en 3 hipótesis:

1. Las interacciones de tres o más factores son sumamente raras en la práctica, por lo que en general se pueden suponer como no existentes.
2. En un experimento de varios factores lo más probable es que solo algunos de ellos sean relevantes para la variable de respuesta.
3. La mayor parte del efecto se debe a los factores principales y algunas interacciones de dos factores.

Lo anterior implica que por ejemplo para siete factores son necesarios probablemente solo 28 grados de libertad (7 factores principales y 21 interacciones de dos factores), y esto equivale a solo 29 unidades de información y no 128 como en el experimento original. Esto quiere decir que no es necesario el correr una replicación completa de todo el experimento cuando el número de factores crece, sino solamente algunas casillas o condiciones experimentales.

Cuando solamente una parte de las posibles casillas se prueban, se dice que se tiene una replicación fraccionada del experimento.

Las preguntas que surgen son:

1. ¿Cuántas y cuales casillas probar?
2. ¿Cómo analizar los resultados?
3. ¿Qué información se pierde?

El objetivo del diseño factorial fraccionado (replicación fraccionada) es darle respuesta a estos interrogantes.

4.1. Fracción un medio de un diseño 2^k

En este informe nos enfocaremos en este tipo de diseño, donde se utiliza la mitad de las corridas del diseño completo 2^k, ya que este es el diseño más importante y más utilizado en la práctica.
Para explicar más fácil este método nos centraremos en un ejemplo en la que tres factores, cada uno con dos niveles, son de interés. Suponiendo que los experimentadores no están en posición de correr las $2^3 = 8$ combinaciones de tratamientos, la lógica sugeriría disminuir las corridas. Si se pueden llevar cuatro corridas, esto sugiere una fracción un medio del diseño 2^3. Puesto que el diseño contiene $2^{3-1} = 4$ combinaciones de tratamientos, es común llamar diseño 2^{3-1} a una fracción un medio del diseño 2^3.
Supongamos que se seleccionan las cuatro combinaciones de tratamientos a,b,c y abc como la fracción un medio con la que se trabajará. En la siguiente tabla se muestra la agrupación de signos positivos y negativos del diseño 2^3 la cuál se utilizará para explicar el proceso.

TABLA 5: Add caption

Combinación de	Efecto factorial							
tratamientos	I	A	B	C	AB	AC	BC	ABC
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
c	+	-	-	+	+	-	-	+
abc	+	+	+	+	+	+	+	+
ab	+	+	+	-	+	-	-	-
ac	+	+	-	+	-	+	-	-
bc	+	-	+	+	-	-	+	-
(1)	+	-	-	-	+	+	+	-

Observemos que el diseño 2^{3-1} se forma seleccionando sólo las combinaciones de tratamientos que tienen signo positivo en la columna ABC. Por lo tanto, a ABC se le llama generador de esta fracción particular. Además, se observa que la columna identidad I también es totalmente positiva, por lo que a

$$I = ABC$$

se le llama la relación de definición del diseño. En general, la relación de definición de un diseño factorial fraccionado será siempre el conjunto de todas las columnas que son iguales a la columna identidad I.

Con respecto a la tabla 7, se observa que las combinaciones lineales de las observaciones usadas para estimar los efectos principales de A, B y C son

$$\begin{aligned} l_A &= \tfrac{1}{2}(a - b - c + abc) \\ l_B &= \tfrac{1}{2}(-a + b - c + abc) \\ l_C &= \tfrac{1}{2}(-a - b + c + abc) \end{aligned}$$

también se observa que las combinaciones lineales de las observaciones usadas para estimar las interacciones de dos factores son:

$$\begin{aligned} l_{BC} &= \tfrac{1}{2}(a - b - c + abc) \\ l_{AC} &= \tfrac{1}{2}(-a + b - c + abc) \\ l_{AB} &= \tfrac{1}{2}(-a - b + c + abc) \end{aligned}$$

Por lo tanto $l_A = l_{BC}$,$l_B = l_{AC}$ y $l_C = l_{AB}$; por consiguiente, es imposible diferenciar entre A y BC, entre B y AC y entre C y AB. De hecho, cuando se estiman A, B y C, se están estimando en realidad $A+BC$, $B+AC$ y $C+AB$. A dos o más efectos que tienen esta propiedad se les llama alias. En este ejemplo, A y BC son alias, al igual que B y AC y C y AB.

5. Ejemplo

Una empresa que fabrica juguetes tiene 3 estaciones de ensamble (A,B,C) y cada estación funciona en dos velocidades distintas (Alta y Baja), se desea conocer el efecto que tienen en conjunto las diferentes velocidades sobre la cantidad de juguetes ensamblados en un día. Los datos se muestran en la siguiente tabla

Tabla 6: Datos ejemplo diseño 2^3

	ESTACIÓN B			
	BAJA		ALTA	
ESTACIÓN A	ESTACIÓN C		ESTACIÓN C	
	BAJA	ALTA	BAJA	ALTA
BAJA	4	7	20	10
	5	9	14	6
ALTA	4	2	4	14
	11	7	6	16

5.1. Factorial completo

Según los datos de la tabla, se tiene que:

$$\begin{array}{cccc} (1) = 9 & c = 16 & b = 34 & bc = 16 \\ a = 15 & ac = 9 & ab = 10 & abc = 30 \end{array}$$

Y los efectos serán

$$\begin{gathered} A = \tfrac{1}{8}[15 - 9 + 10 - 34 + 9 - 16 + 30 - 16] = -\tfrac{11}{8} = -1.375 \\ B = \tfrac{1}{8}[34 + 10 + 16 + 30 - (9 + 15 + 16 + 9)] = \tfrac{41}{8} = 5.125 \\ C = \tfrac{1}{8}[16 + 9 + 16 + 30 - (9 + 15 + 34 + 10)] = \tfrac{3}{8} = 0.375 \\ AB = \tfrac{1}{8}[9 + 10 + 16 + 30 - (15 + 34 + 9 + 16)] = -\tfrac{9}{8} = -1.125 \\ AC = \tfrac{1}{8}[9 + 34 + 9 + 30 - (15 + 10 + 16 + 16)] = \tfrac{25}{8} = 3.125 \\ BC = \tfrac{1}{8}[9 + 15 + 16 + 30 - (34 + 10 + 16 + 9)] = \tfrac{1}{8} = 0.125 \\ ABC = \tfrac{1}{8}[30 + 15 + 34 + 16 - (10 + 9 + 16 + 9)] = \tfrac{51}{8} = 6.375 \end{gathered}$$

Y las sumas de cuadrados serán

$$\begin{gathered} SC_A = \tfrac{1}{16}(-11)^2 = 7.56 \\ SC_B = \tfrac{1}{16}(41)^2 = 105.06 \\ SC_C = \tfrac{1}{16}(3)^2 = 0.56 \\ SC_{AB} = \tfrac{1}{16}(9)^2 = 5.06 \\ SC_{AC} = \tfrac{1}{16}(25)^2 = 39.06 \\ SC_{BC} = \tfrac{1}{16}(1)^2 = 0.06 \\ SC_{ABC} = \tfrac{1}{16}(51)^2 = 162.56 \\ SC_{total} = (4^2 + 5^2 + \cdots + 14^2 + 16^2) - \tfrac{1}{16}139^2 = 389.44 \\ SC_{error} = 69.52 \end{gathered}$$

Teniendo en cuenta lo anterior, la tabla de análisis de varianza es

Tabla 7

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	F	Pr>F
Factor A	1	7.56	7.56	0.87	0.3781
Factor B	1	105.06	105.06	12.09	0.0083*
Factor C	1	0.56	0.56	0.06	0.8055
Interacción AB	1	5.06	5.06	0.58	0.4671
Interacción AC	1	39.06	39.06	4.49	0.0667
Interacción BC	1	0.06	0.06	0.01	0.9345
Interacción ABC	1	162.56	162.56	18.71	0.0025*
Residual	8	69.52	8.62		
Total	15	389.44			

5.2. Ejemplo fraccionado

Como se mostró en la tabla 5, se seleccionan los factores e interacciones tales que el signo de la interacción ABC es positivo. En este caso, la fracción un medio del diseño 2^3 está compuesto por los factores A, B, C y la interacción ABC.
La ANOVA de la fracción un medio del diseño 2^k es

Tabla 8

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	F	Pr>F
H	1	3.125	3.125	0.2688	0.6315
J	1	136.125	136.125	11.7097	0.02673*
K	1	1.125	1.125	0.0968	0.77127
Residual	4	46.5	11.625		
Total	7	186.875			