Universidad del Valle
Programa Académico de Estadística
Estadística Aplicada III
Cali, marzo de 2019

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.

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

[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

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1 : \text{Al menos una } \tau_i \neq 0$$

Para el factor B:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$$
$$H_1 : \text{Al menos una } \beta_j \neq 0$$

También existe interés en determinar si los tratamientos de los factores A y B interactúan. Por lo tanto, también querría probarse

$$H_0 : (\tau\beta)_{ij} = 0 \text{ para todas las } i, j$$
$$H_1 : \text{Al menos una } (\tau\beta)_{ij} \neq 0$$

Todas las hipótesis anteriores se pueden probar realizando el análisis de varianza de dos factores como sigue:

Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F
A	a-1	SC_A	$CM_A = \frac{SC_A}{a-1}$	$F = \frac{CM_A}{CM_{error}}$
B	b-1	SC_B	$CM_B = \frac{SC_B}{b-1}$	$F = \frac{CM_B}{CM_{error}}$
AB	(a-1)(b-1)	SC_{AB}	$CM_{AB} = \frac{SC_{AB}}{(a-1)(b-1)}$	$F = \frac{CM_{AB}}{CM_{error}}$
Error	ab(n-1)	SC_{error}	$CM_{error} = \frac{SC_{error}}{ab(n-1)}$	
Total	abn-1	SC_{total}		

Donde la suma de cuadrados total es:

$$SC_{total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

La suma de cuadrados de los efectos principales son:

$$SC_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn} \quad \text{y} \quad SC_B = \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{abn}$$

La suma de cuadrados de la interacción es:

$$SC_{AB} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn} - SC_A - SC_B$$

Y, la suma de cuadrados del error se encuentra como:

$$SC_{error} = SC_{total} - \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn}$$

2.1. Ejemplo

Un ingeniero está diseñando una batería que se usará en un dispositivo que se someterá a variaciones de temperatura extremas. El único parámetro del diseño que puede seleccionar en este punto es el material de la placa o ánodo de la

batería, y tiene tres elecciones posibles. Cuando el dispositivo esté fabricado y se envíe a campo, el ingeniero no tendrá control sobre las temperaturas extremas en las que operará el dispositivo, pero sabe por experiencia que la temperatura probablemente afectará la vida efectiva de la batería. Sin embargo, la temperatura puede controlarse en el laboratorio donde se desarrolla el producto para fines de prueba.
El ingeniero decide probar los tres materiales de la placa con tres niveles de temperatura(15,70 y 125°F), ya que estos niveles de temperatura son consistentes con el medio ambiente donde se usará finalmente el producto. Se prueban cuatro baterías con cada combinación del material de la placa y la temperatura, y las 36 pruebas se corren de manera aleatoria. Los datos se presentan en la siguiente tabla.

Tipo de material	Temperatura (°F)					
	15		70		125	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

Se analizaron todos los supuestos y no se presenta un incumplimiento de alguno, por lo que se procede a plantear las hipótesis para los efectos de ambos tratamientos y para el efecto de la interacción, y a interpretar el análisis de varianza que se muestra en la siguiente tabla

Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F	Pr(>F)
Material	2	10639	5319.5	7.9308	0.001952 **
Temperatura	2	39197	19598.7	29.2195	1.762e-07 ***
Material:Temperatura	4	9638	2409.6	3.5925	0.017916 *
Residuals	27	18110	670.7		

La conclusión más importante que se obtiene del análisis de varianza es que el efecto de la interacción entre el material y la temperatura sobre la vida de las baterías(en horas) es significativo, esto quiere decir que en conjunto, la temperatura y el material afectan directamente la vida de las baterías. Ahora, analizando los factores individualmente, tanto el material como la temperatura son significativos, es decir, que cada uno por separado también tiene un efecto considerable sobre la vida de las baterías.

Posteriormente, se abordará el mismo ejemplo por medio de diseños factoriales fraccionados con el objetivo de realizar una comparación y poder observar cuánta información se pierde y cuánto varían los resultados con respecto al diseño factorial completo.

3. Diseños factoriales fraccionales

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.

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:

No. De factores	No. De casillas	Efectos principales	Interacciones entre factores de						
			1	3	4	5	6	7	8
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

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

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.

3.1. Fracción un medio de un diseño 2^k