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. Antecedentes

Para estudiar un poco como se han trabajado los diseños factoriales fraccionados se estudiaron tres artículos, el primero **Using fractional factorial designs with mixture constraints to improve nutritional value and sensory properties of processed food** escrito por **Rytz et al. (2017)** donde se propone un enfoque simple, genérico y eficiente para combinar los factores de proceso y mezcla en el mismo diseño al manejar fácilmente cualquier restricción de mezcla en el marco de los diseños factoriales fraccionarios, teniendo en cuenta que los productos alimenticios industriales se pueden considerar como mezclas procesadas de ingredientes para los cuales los parámetros del proceso y los parámetros de la mezcla se deben investigar simultáneamente al tratar de mejorar sus propiedades nutricionales y sensoriales; el segundo artículo **The use of nonregular fractional factorial designs in combination toxicity studies** escrito por **Phoa et al. (2009)** el cual muestra cuando hay interés en estudiar n productos químicos que utilizan x niveles de dosis cada uno, los diseños factoriales que requieren x^n grupos de tratamiento se han presentado como uno de los valiosos enfoques estadísticos para la evaluación de peligros de mezclas químicas, y permitir aplicaciones ejemplares y comparaciones de costo-eficiencia de diseños factoriales completos y diseños factoriales fraccionados regulares en estudios de toxicidad y por último, **Villegas & Baldemar (2013)** en su artículo **Aplicación de diseño de experimentos para el análisis de secado de un producto** exponen un estudio realizado sobre el pintado de un producto, para determinar en qué niveles de los factores se minimiza el tiempo de espera para el secado total. Se realizó un experimento factorial completo de 3 factores con 2 niveles cada uno, los factores que se utilizaron fueron tipo de pintura-esmalte a utilizar (estándar o de secado rápido), el método de secado (con o sin agua fría) y la cantidad de capas que se aplican (1 o 2 capas). Se realizaron 3 réplicas para quitar ambigüedades en las respuestas.

3. 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.

[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

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:

TABLA 1: Diseño 2^2 para ilustrar estimación de efectos

	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.

4. 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: Tabla estándar con interacciones de diseño 2^3

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: Análisis de varianza diseño 2^k

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	$n2^k - 1$

En este tipo de diseño es sencillo expresar los resultados del experimento en términos de un modelo de regresión. El modelo de regresión es

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

donde x_1 y x_2 son las variables codificadas que representa el nivel alto $+1$ o el nivel bajo -1 de cada factor y las β son los coeficientes de refresión, donde la estimaciones de estos coeficientes son la mitad de las estimaciones de los efectos de los factores correspondientes.

5. Diseños factoriales 2^k con una sola replica

Debido a que por lo general los recursos son limitados, el número de réplicas que el experimentador puede emplear quizás esté restringido. Con frecuencia, los recursos disponibles permiten hacer únicamente una sola réplica del diseño.
Un riesgo cuando se realiza un experimento que tiene una sola corrida para combinación de prueba es que el modelo puede ajustarse al ruido. Es decir, si la respuesta y es sumamente variable, pueden resultar conclusiones engañosas del experimento. Ahora bien, si hay menos variabilidad en la respuesta, la posibilidad de una conclusión errónea será más reducida. Otra forma de asegurarse de que se obtienen estimaciones confiables de los efectos es incrementando la distancia entre los niveles bajo (-) y alto (+) del factor.
Una sola réplica del diseño 2^k se denomina diseño factorial no replicado. Con una sola réplica, no se cuenta con ninguna estimación interna del error. Una forma d abordar este análisis consiste en suponer que algunas interacciones de orden superior son insignificantes y combinar sus cuadrados medios para estimar el error para lograr obtener la ANOVA.

Daniel sugiere examinar una gráfica de probabilidad normal de las estimaciones de los efectos. Los efectos que son insignificantes siguen una distribución normal con media cero y varianza σ^2, y tenderán a localizarse sobre una línea recta en esta gráfica, mientras que los efectos significativos tendrán medias diferentes de cero y no se localizarán sobre la línea recta. Por lo tanto, el modelo preliminar se especificará de tal modo que contenga aquellos efectos que aparentemente son diferentes de cero, con base en la gráfica de probabilidad normal. Los efectos aparentemente insignificantes se combinan como una estimación del error.

Posteriormente en la sección de los ejemplos se mostrará como es el análisis para este tipo de diseños.

6. 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: Relación efectos y casillas con respecto al número de factores

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.

6.1. 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: Agrupación signos positivos y negativos diseño 2^3

Combinación de tratamientos	Efecto factorial							
	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 5, 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. Esto se indica con la notación $L_A \rightarrow\ A + BC$, $L_B \rightarrow\ B + AC$ y $L_C \rightarrow C + AB$

7. Ejemplos

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

7.1. Diseño factorial completo

Según los datos de la tabla, se tiene que:

$$(1) = 9 \quad c = 16 \quad b = 34 \quad bc = 16 \quad a = 15 \quad ac = 9 \quad ab = 10 \quad abc = 30$$

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: Análisis de varianza factorial completo

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			

Podemos ver que la ANOVA para el diseño completo dio como resultado que la interacción entre los 3 factores es significativa, esto significa que en conjunto las 3 maquinas tienen un efecto significativo sobre la cantidad de juguetes fabricados.

7.2. Diseño fraccionado un medio

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. Como se utilizó la relación de definición de diseño $I = ABC$, se deducen los alias $A = BC$, $B = AC$ y $C = AB$ De esta forma si alguno de los Alias A,B,C es significativo podemos inferir que la interacción ABC también lo es.

La ANOVA de la fracción un medio del diseño 2^k es

TABLA 8: Análisis de varianza factorial fraccionado

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

Nuestra ANOVA para el ejemplo de la fracción de diseño dio como resultada que Alias $B = AC$ es significativo, esto quiere decir que la interacción entre los diferentes velocidades de cada estación tienen un efecto significativo sobre la cantidad de juguetes ensamblados, lo cual no es diferente el resultado que arrojó la ANOVA del diseño no fraccionado.

7.3. Ejemplo diseño factorial completo con una sola réplica

Un producto químico se fabrica en un envase presurizado. Se lleva a cabo un experimento factorial en la planta piloto para estudiar los factores que se piensa influyen en el índice de filtración de este producto. Los cuatro factores son la temperatura(A), la presión(B), la concentración del formaldehído(C) y la velocidad de agitación(D). Cada factor está presente con dos niveles. La matriz de diseño es la siguiente

TABLA 9: Matriz de diseño ejemplo con una sola réplica

Número de corrida	Factor A	B	C	D	Etiqueta de la corrida	Índice de filtración (gal/h)
1	-	-	-	-	(1)	45
2	+	-	-	-	a	71
3	-	+	-	-	b	48
4	+	+	-	-	ab	65
5	-	-	+	-	c	68
6	+	-	+	-	ac	60
7	-	+	+	-	bc	80
8	+	+	+	-	abc	65
9	-	-	-	+	d	43
10	+	-	-	+	ad	100
11	-	+	-	+	bd	45
12	+	+	-	+	abd	104
13	-	-	+	+	cd	75
14	+	-	+	+	acd	86
15	-	+	+	+	bcd	70
16	+	+	+	+	abcd	96

El análisis de estos datos se inicia construyendo la gráfica de probabilidad normal de la estimaciones de los efectos. Podemos observar en la gráfica 1 que los efectos significativos son los efectos principales A, C y D y las interacciones AC y AD. Puesto que B(presión) no es significativo y todas las interacciones en las que interviene B son insignificante, B puede descartarse del experimento, de tal modo que ahora si se puede tener una estimación del error ya que quedan grados de libertad para dicha estimación.

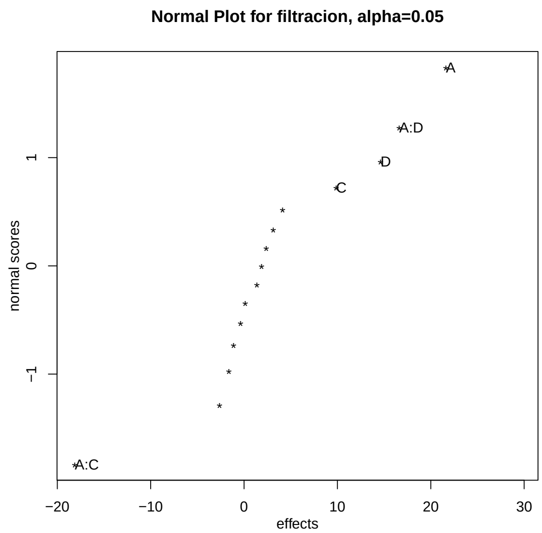

FIGURA 1: Gráfica QQplot de los efectos

La ANOVA da como resultado

TABLA 10: Análisis de varianza una sola réplica

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	F	Pr>F
A	1	1870.56	1870.56	83.3677	1.667e-05 ***
C	1	390.06	390.06	17.3844	0.0031244 **
D	1	855.56	855.56	38.1309	0.0002666 ***
A:C	1	1314.06	1314.06	58.5655	6.001e-05 ***
A:D	1	1105.56	1105.56	49.2730	0.0001105 ***
C:D	1	5.06	5.06	0.2256	0.6474830
A:C:D	1	10.56	10.56	0.4708	0.5120321
Residual	8	179.50	22.44		

La ANOVA dió como resultado que las interacciones AC y AD son significativas, sin embargo la interacción ACD no lo es, lo que nos dice que tanto al temperatura combinada con la concentración como la temperatura combinada con la velocidad de agitación tiene un efecto sobre el indice de filtración.
Dado que los únicos efectos significativos son $A = 21.625$,$C = 9.875$,$D = 14.625$,$AC = -18.125$ y $AD = 16.625$, el modelo estimado será

$$\hat{y} = 70.06 + \left(\frac{21.625}{2}\right) x_1 + \left(\frac{9.875}{2}\right) x_3 + \left(\frac{14.625}{2}\right) x_4 - \left(\frac{18.125}{2}\right) x_1x_3 + \left(\frac{16.625}{2}\right) x_1x_4$$

7.4. Ejemplo diseño fraccionado un medio con una sola réplica

Se retoma el ejemplo anterior y se propone una fracción un medio del diseño 2^4. Se usará el diseño 2^{4-1} con $I = ABCD$. Para construir el diseño, primero se apunta el diseño básico, el cual es un diseño 2^3. Para encontrar los niveles del cuarto factor, se resuelve $I = ABCD$ para D, o $D = ABC$. Por lo tanto, el nivel de D de cada corrida es el producto de los signos positivos y negativos de las columnas A,B y C. El procso se ilustra en la siguiente tabla

TABLA 11: Matriz de diseño 2^4 fraccionado un medio

Número de	Diseño básico				Combinación de	Índice de
corrida	A	B	C	D=ABC	tratamientos	filtración (gal/h)
1	-	-	-	-	(1)	45
2	+	-	-	+	ad	100
3	-	+	-	+	bd	45
4	+	+	-	-	ab	65
5	-	-	+	+	cd	75
6	+	-	+	-	ac	60
7	-	+	+	-	bc	80
8	+	+	+	+	abcd	96

Utilizando la relación de definición del diseño $I = ABCD$, podemos deducir que cada uno de los efectos principales es alias de una interacción de tres factores como sigue, $A = BCD$,$B = ACD$,$C = ABD$,$D = ABC$, y además, cada interacción de dos factores es alias de otra interacción de dos factores como sigue, $AB = CD$,$AC = BD$ y $BC = AD$. Los cuatro efectos principales más los 3 pares de alias de interacciones de dos factores representan los siete grados de libertad del diseño.
En la tabla **tal** se muestran las estimaciones de los efectos obtenidas de este diseño. Para mostrar los cálculos, la combinación lineal de las observaciones asociadas con el efecto de A es

$$L_A = \frac{1}{4}(-45 + 100 - 45 + 65 - 75 + 60 - 80 + 96) = 19 \rightarrow \quad A + BCD$$

Mientras que para el efecto AB se obtendría

$$L_{AB} = \frac{1}{4}(45 - 100 - 45 + 65 + 75 - 60 - 80 + 96) = -1 \rightarrow \quad AB + CD$$

Tabla 12: Estimaciones de los efectos y los alias

Estimación	**Estructura de los alias**
$L_A = 19$	$L_A \rightarrow \quad A + BCD$
$L_B = 1.5$	$L_B \rightarrow \quad B + ACD$
$L_C = 14$	$L_C \rightarrow \quad C + ABD$
$L_D = 16.5$	$L_D \rightarrow \quad D + ABC$
$L_{AB} = -1$	$L_{AB} \rightarrow \quad AB + CD$
$L_{AC} = -18.5$	$L_{AC} \rightarrow \quad AC + BD$
$L_{AD} = 19$	$L_{AD} \rightarrow \quad AD + BC$

De la tabla anterior, podemos concluir que los efectos principales A, C y D son grandes. Además, si A, C y D son los efectos principales importantes, entonces es lógico concluir que las dos cadenas de alias de interacciones $AC + BD$ y $AD + BC$ tienen efectos grandes, ya que las interacciones AC y AD también son significativas. Puesto que el factor B no es significativo, puede sacarse de consideración.
Con base en el análisis anterior, puede obtenerse ahora un modelo para predecir el índice de filtración en la región experimental. Este modelo es

$$\hat{y} = \hat{\beta_0} + \hat{\beta_1}x_1 + \hat{\beta_3}x_3 + \hat{\beta_4}x_4 + \hat{\beta_{13}}x_1x_3 + \hat{\beta_{14}}x_1x_4$$

Donde x_1,x_3 y x_4 son variables codificadas $(-1 \leq x_i \leq +1)$ y las β son los coeficientes de regresión que se obtienen dividiendo la estimación del efecto por 2. Por lo tanto, la ecuación es:

$$\hat{y} = 70.75 + \left(\frac{19}{2}\right)x_1 + \left(\frac{14}{2}\right)x_3 + \left(\frac{16.5}{2}\right)x_4 + \left(\frac{-18.5}{2}\right)x_1x_3 + \left(\frac{19}{2}\right)x_1x_4$$

8. Conclusión

La principal conclusión que obtenemos es que los diseños factoriales fraccionados son una alternativa muy útil cuando se tienen muchos factores y los recursos para llevar a cabo el experimento son limitados. Se pudo observar que en los dos ejemplos realizados, centrándonos en el primero, vimos que se llegó a la misma conclusión en la ANOVA con el factorial completo y el factorial fraccionado, de que la interacción de los 3 factores A,B y C es significativa. Con respecto al segundo ejemplo, se pudo observar que los resultados del modelo ajustado con los β_i estimados también son muy parecidos en el factorial completo y el fraccionado; esto nos lleva a concluir que los diseños factoriales fraccionados son muy importantes, ya que con la mitad o menos de las corridas se llegan a resultados muy semejantes a si se hiciera completo, esto es muy útil ya que en la practica la mayoría de las investigaciones o diseños que se plantean tienen restricciones de recursos y este método es una opción bastante confiable.

Apéndice A. Código R

```
#Kevin Steven García - 1533173
#Alejandro Vargas - 1525953
#Alejandro Soto - 1532457
#Diseños factoriales fraccionados

##Datos diseño para ejemplo 1
FA=c(rep("0",8),rep("1",8))
FC=c(rep(c("0","0","1","1"),4))
FB=c(rep(c(rep("0",4),rep("1",4)),2))
Respuesta=c(5,4,7,9,20,14,10,6,4,11,2,7,4,6,14,16)
datos4=data.frame(FA,FB,FC,Respuesta)
str(datos4)

#grafico de efectos principales
x11()
Efectos <- data.frame(FA,FB,FC,Respuesta)
plot.design(Efectos, fun="mean", main=" Gráfica de efectos principales", ylab= "Respuesta", xlab="Factor")

#MODELO FACTORIAL
mod4<-lm(Respuesta~FA+FB+FC+FA:FB+FA:FC+FB:FC+FA:FB:FC, data=datos4)
anova(mod4)

#fraccionado
#a mano seria tomar las que muestran signo positivo en la interaccion de todas las velocidades
#antes de hacer el experimento
library("AlgDesign")
levels.design = c(2,2,2)
f.design <- gen.factorial(levels.design)

str(f.design)

#vector con los signos de la interaccion ABC
tr=c(f.design$X1*f.design$X2*f.design$X3)

#cuando se tienen replicas
fr=c()
for (i in 1:8) {
  fr=c(fr,rep(tr[i],2))
}

#extraccion de los datos matriz fraccionada
nueva=datos4
j=c()
for (i in 1:length(fr)) {
  if(fr[i]!=1){
    j=c(j,i)
  }
  nueva=datos4[-j,]
}
colnames(nueva)=c("H","J","K","O")



#MODELO FACTORIAL
mod6<-lm(O~H+J+K, data=nueva)
anova(mod6)
```

```
#Ejemplo2
#Diseño factorial completo con una sola replica.
#Ingresamos los datos
A<-c(rep(c("0","1"),8))
B<-c(rep(c("0","0","1","1"),4))
C<-c(rep(c(rep("0",4),rep("1",4)),2))
D<-c(rep("0",8),rep("1",8))
filtracion<-c(45,71,48,65,68,60,80,65,43,100,45,104,75,86,70,96)
datos2=data.frame(A,B,C,D,filtracion)
str(datos2)
head(datos2)
mod3<-lm(filtracion~A+C+D+A:C+A:D+C:D+A:C:D, data=datos2)
anova(mod3)


library(FrF2)
Tabla <- FrF2(nruns = 16,
              nfactors = 4,
              factor.names = list(A=c("0","1"),
                                  B=c("0","1"),
                                  C=c("0","1"),
                                  D=c("0","1")),
              replications = 1, randomize = F)
Tabla <- add.response(design = Tabla, response = filtracion)
Tabla
#Análisis método de Daniel
#Gráfico manual
Efectos0<-c(21.625,3.125,9.875,14.625,0.125,-18.125,16.625,2.375,-0.375,
            -1.125,1.875,4.125,-1.625,-2.625,1.375)
x11()
qqnorm(Efectos0)
qqline(Efectos0)

#Gráfico R
x11()
DanielPlot(Tabla)

#Gráfico efectos principales
MEPlot(Tabla, lwd = 2)
abline(h=0, col="red")

## Gráficas de Interacciones
IAPlot(Tabla, lwd = 2)

# Gráfica de interacción triple
filtracion1<-filtracion[1:8]
A1<-A[1:8]
B1<-B[1:8]
C1<-C[1:8]
filtracion2<-filtracion[9:16]
A2<-A[9:16]
B2<-B[9:16]
C2<-C[9:16]
x11()
par(mfrow=c(1,2))
cubePlot(obj = filtracion1,
         eff1 = A1,
         eff2 = B1,
         eff3 = C1,
```

```
        main = " Gráfica de interacción triple con factor D bajo")
cubePlot(obj = filtracion2,
        eff1 = A2,
        eff2 = B2,
        eff3 = C2,
        main = " Gráfica de interacción triple con factor D alto",round = -1)

#Fraccionado con una sola replica
#a mano seria tomar las que muestran signo positivo en la interaccion de todas las velocidades
#antes de hacer el experimento
library("AlgDesign")
levels.design1 = c(2,2,2,2)
f.design1 <- gen.factorial(levels.design1)
str(f.design1)

#vector con los signos de la interaccion ABC
tr1=c(f.design1$X1*f.design1$X2*f.design1$X3*f.design1$X4)

#extraccion de los datos matriz fraccionada
j=c()
for (i in 1:length(tr1)) {
  if(tr1[i]!=1){
    j=c(j,i)
  }
}
nueva1=datos2[-j,]
colnames(nueva1)=c("A","B","C","D","O")

Efectos<-c(19,1.5,14,16.5,-1,-18.5,19)
x11()
qqnorm(Efectos)
qqline(Efectos)
text(c(0.7,-0.3,-0.07,0.42,-0.68,-1.28,1.27),Efectos,c("A","B","C","D","AB","AC","AD"))
```

Referencias

Castro, E. C. (n.d.), 'Una sola réplica en el diseño 2^k'.
*https://slideplayer.es/slide/6143452/

Gromping, U. (2014), 'R package FrF2 for creating and analyzing fractional factorial 2-level designs', *Journal of Statistical Software* **56**(1), 1–56.
*http://www.jstatsoft.org/v56/i01/

Kuehl, R. O. (2001), *Diseño de experimentos*.

Montgomery, D. C. (2004), *Diseño y análisis de experimentos*.

Phoa, F. K., Xu, H. & Wong, W. K. (2009), 'The use of nonregular fractional factorial designs in combination toxicity studies', *Food and Chemical Toxicology* **47**(9), 2183 – 2188.
*http://www.sciencedirect.com/science/article/pii/S0278691509002701

Pulido, H. G. & de la Vara Salazar, R. (2008), *Análisis y diseño de experimentos*.

R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*https://www.R-project.org/

Rytz, A., Moser, M., Lepage, M., Mokdad, C., Perrot, M., Antille, N. & Pineau, N. (2017), 'Using fractional factorial designs with mixture constraints to improve nutritional value and sensory properties of processed food', *Food Quality and Preference* **58**, 71 – 75.
*http://www.sciencedirect.com/science/article/pii/S0950329317300150

Villagarcía, T. (n.d.), 'Diseños factoriales a dos niveles'.

Villegas, G. & Baldemar, J. (2013), 'Aplicación de diseño de experimentos para el análisis de secado de un producto (experiment design application for analysis of the drying a product)', *Innovaciones de negocios* **10**(19), 145–158.

Wheeler, B. (2014), *AlgDesign: Algorithmic Experimental Design*. R package version 1.1-7.3. *https://CRAN.R-project.org/package=AlgDesign