Diseño de Experimentos - 3007340
DOE - Diseño factoriales 2^k - Parte I: Diseños replicados

Nelfi González Alvarez

Profesora Asociada Escuela de Estadística

e-mail: ngonzale@unal.edu.co

Facultad de Ciencias, Universidad Nacional de Colombia Sede Medellín

Escuela de Estadística

Semestre 02 de 2021

Contenido I

Contenido

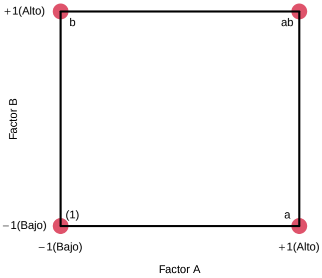
Introducción

- Se estudian los efectos de k factores cada uno en 2 niveles
- Los niveles son definidos como **bajo**, denotado también por -1 (será el nivel 1 del factor), y **alto**, denotado también por $+1$ (será el nivel 2 del factor).
- Los niveles pueden ser cualitativos o cuantitativos (lo suficientemente separados para notar cambio en la media, pero dentro de la región de operación admisible en el proceso)
- Todos los efectos posibles son de tipo fijos

Con $k = 2$ factores, los tratamientos son denotados como muestra la siguiente tabla

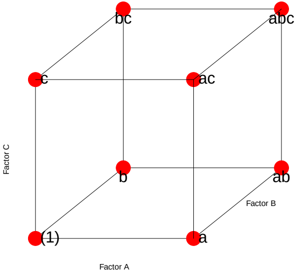
Tratamiento	A	B	Combinación
(1)	$-$	$-$	A bajo, B bajo
a	$+$	$-$	A alto, B bajo
b	$-$	$+$	A bajo, B alto
ab	$+$	$+$	A alto, B alto

En la siguiente figura se representan los 2^2 tratamientos para un diseño con dos factores A y B.



En un 2^3, los tratamientos son:

Tratamiento	A	B	C	Combinación
(1)	$-$	$-$	$-$	A bajo, B bajo, C bajo
a	$+$	$-$	$-$	A alto, B bajo, C bajo
b	$-$	$+$	$-$	A bajo, B alto, C bajo
ab	$+$	$+$	$-$	A alto, B alto, C bajo
c	$-$	$-$	$+$	A bajo, B bajo, C alto
ac	$+$	$-$	$+$	A alto, B bajo, C alto
bc	$-$	$+$	$+$	A bajo, B alto, C alto
abc	$+$	$+$	$+$	A alto, B alto, C alto



Nota 1.1

Más adelante se presentará la regla con la cual se codifican los tratamientos a través de cadenas de caracteres con las letras latinas minúsculas que se ha usado en la tabla anterior.

Contenido

Experimentos factoriales 2^2

Consideraciones:

- Cada uno de los cuatro tratamientos posibles son replicados n veces
- El modelo ANOVA sería el mismo para dos factores de efectos fijos con interacción, en un DCA

$$\begin{aligned}&Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \text{ con } \varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2),\\&\sum_{i=1}^{2} \alpha_i = 0,\ \sum_{j=1}^{2} \beta_j = 0,\\&\sum_{i=1}^{2} (\alpha\beta)_{ij} = 0 \text{ para } j = 1, 2,\ \sum_{j=1}^{2} (\alpha\beta)_{ij} = 0 \text{ para } i = 1, 2.\end{aligned} \tag{1}$$

- Sin embargo, trabajaremos con el modelo de regresión equivalente,

$$Y_l = \beta_0 + \beta_1 X_{1l} + \beta_2 X_{2l} + \beta_{12} X_{1l} X_{2l} + E_l,\ E_l \overset{iid}{\sim} N(0,\sigma^2),\ l = 1,\ldots,N \quad (2)$$

con

$$X_1 = \begin{cases} -1 & \text{si A está en su nivel bajo} \\ +1 & \text{si A está en su nivel alto} \end{cases} \quad X_2 = \begin{cases} -1 & \text{si B está en su nivel bajo} \\ +1 & \text{si B está en su nivel alto} \end{cases} \quad (3)$$

$N = 2^2 n$ es el total de observaciones en el experimento. Los parámetros de este modelo de regresión están relacionados con los del modelo ANOVA, así

Parámetro Anova	Parámetro MRLM
μ	β_0
α_2	β_1
β_2	β_2
$(\alpha\beta)_{22}$	β_{12}

Esta equivalencia es obtenida igualando las expresiones de la respuesta media en los dos modelos, en el tratamiento A en nivel alto ($i = 2$): $X_1 = +1$, y B en nivel alto ($j = 2$): $X_2 = +1$, es decir $\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} = \beta_0 + \beta_1 + \beta_2 + \beta_{12}$.

1. Para la equivalencia de los demás parámetros del modelo ANOVA con los del MRLM, puede hacerlo a través de la comparación de las medias de la respuesta en cada uno de los demás tratamientos, o bien, recordar que por las restricciones lineales en el ANOVA,
   - En el Anova $\alpha_1 = -\alpha_2$, entonces $\alpha_1 = -\beta_1$ del MRLM.
   - En el Anova $\beta_1 = -\beta_2$, entonces $\beta_1 = -\beta_2$ del MRLM.
   - En el Anova $(\alpha\beta)_{12} = -(\alpha\beta)_{22}$, entonces $(\alpha\beta)_{12} = -\beta_{12}$ del MRLM.
   - En el Anova $(\alpha\beta)_{21} = -(\alpha\beta)_{22}$, entonces $(\alpha\beta)_{21} = -\beta_{12}$ del MRLM.
   - En el Anova $(\alpha\beta)_{11} = -(\alpha\beta)_{12}$, entonces $(\alpha\beta)_{11} = \beta_{12}$ del MRLM.
2. Note además que los grados de libertad del SSE en el ANOVA y en el MRLM son iguales y corresponden a $2^2(n-1)$, pues en el ANOVA son calculados como $ab(n-1) = 2^2(n-1)$, con $a = b = 2$, y en el MRLM son calculados como $N - \#$ parámetros $= 2^2 n - 4 = 2^2(n-1)$.

Estimación

La estimación del modelo ANOVA se puede realizar a través de la estimación por MCO del modelo de RLM equivalente, sin embargo, vamos a recurrir al uso de ***contrastes de totales de tratamientos*** (o sea contraste de las sumas de las respuestas en los tratamientos). Para ello consideraremos la siguiente notación,

Notación de Yates	Total representado	Puntos factoriales
(1)	Total en el nivel bajo de A y nivel bajo de B	$(-1,-1)$
a	Total en el nivel alto de A y nivel bajo de B	$(+1,-1)$
b	Total en el nivel bajo de A y nivel alto de B	$(-1,+1)$
ab	Total en el nivel alto de A y nivel alto de B	$(+1,+1)$

Es decir, la notación dada representa tanto al tratamiento como al valor del total de la respuesta en el tratamiento.

Medias muestrales en función de la notación de totales de tratamientos

		B: -1	B: $+1$	Media
A	-1	(1)	b	$\bar{Y}_{1\bullet\bullet} = \frac{(1)+b}{2n}$
	$+1$	a	ab	$\bar{Y}_{2\bullet\bullet} = \frac{a+ab}{2n}$
Media		$\bar{Y}_{\bullet 1\bullet} = \frac{(1)+a}{2n}$	$\bar{Y}_{\bullet 2\bullet} = \frac{ab+b}{2n}$	$\bar{Y}_{\bullet\bullet\bullet} = \frac{(1)+a+b+ab}{2^2 n}$

Estimación de efectos a través de contrastes de totales

Total	Coeficientes			Contrastes de totales de tratamientos		
	A	B	$AB = A\times B$	A	B	AB
(1)	-1	-1	$+1$	$-(1)$	$-(1)$	$+(1)$
a	$+1$	-1	-1	$+a$	$-a$	$-a$
b	-1	$+1$	-1	$-b$	$+b$	$-b$
ab	$+1$	$+1$	$+1$	$+ab$	$+ab$	$+ab$
Valor contrastes de totales				$\widehat{w}_A = ab + a - b - (1)$	$\widehat{w}_B = ab - a + b - (1)$	$\widehat{w}_{AB} = ab - a - b + (1)$
Efectos estimados				$\widehat{\alpha}_A = \frac{ab+a-b-(1)}{2n}$	$\widehat{\alpha}_B = \frac{ab-a+b-(1)}{2n}$	$\widehat{\alpha}_{AB} = \frac{ab-a-b+(1)}{2n}$

Note la relación entre los efectos estimados y el contraste de totales, en cada caso:

$$\text{Efecto estimado} = \frac{\text{Contraste totales}}{2n} \tag{4}$$

- **Efecto estimado de A:**

$$\widehat{\alpha}_A = \frac{ab + a - b - (1)}{2n} = \bar{Y}_{2\bullet\bullet} - \bar{Y}_{1\bullet\bullet} = \widehat{\alpha}_2 - \widehat{\alpha}_1 = 2\widehat{\alpha}_2 \Longrightarrow \widehat{\alpha}_2 = \widehat{\alpha}_A/2. \tag{5}$$

- **Efecto estimado de B:**

$$\widehat{\alpha}_B = \bar{Y}_{\bullet 2 \bullet} - \bar{Y}_{\bullet 1 \bullet} = \widehat{\beta}_2 - \widehat{\beta}_1 = 2\widehat{\beta}_2 \Longrightarrow \widehat{\beta}_2 = \widehat{\alpha}_B/2. \tag{6}$$

- **Efecto estimado de interacción AB:**

$$\widehat{\alpha}_{AB} = 2(\bar{Y}_{22\bullet} - \bar{Y}_{2\bullet\bullet} - \bar{Y}_{\bullet 2 \bullet} + \bar{Y}_{\bullet\bullet\bullet}) = 2\widehat{(\alpha\beta)}_{22} \Longrightarrow \widehat{(\alpha\beta)}_{22} = \widehat{\alpha}_{AB}/2. \tag{7}$$

También podemos escribir

$$\widehat{\alpha}_{AB} = \underbrace{\frac{1}{2}\left(\bar{Y}_{22\bullet} - \bar{Y}_{12\bullet}\right)}_{\textbf{Efecto de A en nivel alto de B}} - \underbrace{\frac{1}{2}\left(\bar{Y}_{21\bullet} - \bar{Y}_{11\bullet}\right)}_{\textbf{Efecto de A en nivel bajo de B}} \tag{8}$$

Podemos decir que con relación al modelo Anova,

- $\widehat{\alpha}_A$ estima a $\alpha_A = \mu_{2\bullet} - \mu_{1\bullet} = 2\alpha_2$, la diferencia entre las medias del nivel alto y bajo del Factor A.
- $\widehat{\alpha}_B$ estima a $\alpha_B = \mu_{\bullet 2} - \mu_{\bullet 1} = 2\beta_1$, la diferencia entre las medias del nivel alto y bajo del Factor B.
- $\widehat{\alpha}_{AB}$ estima a $2(\mu_{22} - \mu_{2\bullet} - \mu_{\bullet 2} + \mu) = 2(\alpha\beta)_{22}$.

Dadas las restricciones lineales, los demás efectos principales y de interacciones también resultan estimados.

A través de las relaciones entre los parámetros del modelo Anova con los del MRLM equivalente, estos últimos también resultan estimados a través de $\widehat{\alpha}_A$, $\widehat{\alpha}_B$ y $\widehat{\alpha}_{AB}$. Resumiendo,

Parámetro estimado en el Anova	Parámetro estimado en el MRLM	Equivalencia con efectos estimados
$\widehat{\alpha}_2$	$\widehat{\beta}_1$	$\widehat{\alpha}_A/2$
$\widehat{\beta}_2$	$\widehat{\beta}_2$	$\widehat{\alpha}_B/2$
$\widehat{(\alpha\beta)}_{22}$	$\widehat{\beta}_{12}$	$\widehat{\alpha}_{AB}/2$

Ejemplo

Se lleva a cabo un experimento para aumentar la capacidad de adhesión de productos de caucho. Se fabrican 8 productos con el nuevo aditivo y otros 8 sin éste. Las capacidades de adhesión se registran a continuación ($n = 4$ por tratamiento):

A: Aditivo	B: Temperatura (°C)			
	50 ($X_2 = -1$)		60 ($X_2 = +1$)	
Sin aditivo ($X_1 = -1$)	2.3	2.9	3.4	3.7
	3.1	3.2	3.6	3.2
Con aditivo ($X_1 = +1$)	4.3	3.9	3.8	3.8
	3.9	4.2	3.9	3.5
$\bar{Y}_{\bullet\bullet\bullet} = 3.54375$				

Totales tratamientos		Coeficientes			Contrastes de totales de tratamientos		
Notación	valor	A	B	AB	totales A	totales B	totales AB
(1)	11.5	−1	-1	+1	-11.5	-11.5	11.5
a	16.3	+1	−1	−1	16.3	-16.3	-16.3
b	13.9	−1	1	−1	-13.9	13.9	-13.9
ab	15.0	+1	+1	+1	15.0	15.0	15.0
Valor contrastes de totales					$\widehat{w}_A = 5.9$	$\widehat{w}_B = 1.1$	$\widehat{w}_{AB} = -3.7$
Efectos estimados: $\widehat{\alpha} = \frac{\widehat{w}}{2n}$					$\widehat{\alpha}_A = 0.7375$	$\widehat{\alpha}_B = 0.1375$	$\widehat{\alpha}_{AB} = -0.4625$

Estimación del MRLM en ec. (2),

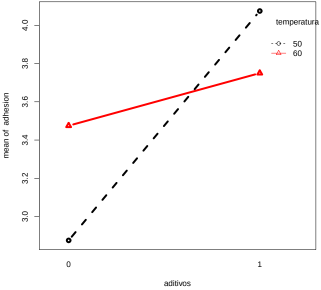
```
Call: lm(formula = adhesion ~ X1 * X2)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.54375    0.06663  53.182 1.29e-15 ***
X1           0.36875    0.06663   5.534 0.000129 ***
X2           0.06875    0.06663   1.032 0.322534
X1:X2       -0.23125    0.06663  -3.470 0.004627 **
```

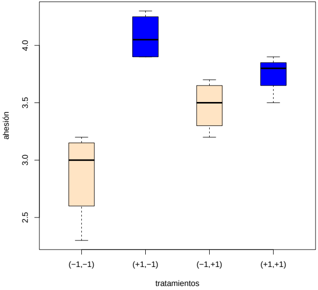
Le estimación por MCO del modelo en la ec. (2), corresponde a

$$\widehat{Y} = 3.54375 + 0.36875X_1 + 0.06875X_2 - 0.23125X_1 * X_2,$$

observe que efectivamente los valores estimados por MCO delos parámetros corresponden a la mitad de los efectos estimados, mientras que la estimación del intercepto es el promedio de los $N = 16$ datos.

El gráfico de perfiles de medias de tratamientos muestra que hay interacción entre los dos factores y dado que los perfiles se cruzan, la interacción puede tener potencialmente la posibilidad de enmascarar la significancia de los efectos principales de los factores (lo cual se verifica en el test de significancia de β_2 en el MRLM). En los boxplots, vemos que la diferencia entre medias del factor Aditivo es mucho mayor cuando el factor temperatura está en su nivel bajo, esto mismo se puede verificar sobre los perfiles de medias: mayor distancia vertical entre los extremos del perfil para la temperatura de 50°C que entre los extremos del perfil de medias para la temperatura de 60°C.





Si no se usa aditivo, entonces para obtener una mayor adhesión, es mejor la temperatura de 60°C, pero si se usa aditivo, es mejor la temperatura de 50°C para mayor adhesión. Finalmente, si el objetivo es escoger el tratamiento que dé mayor adhesión, entre los tratamientos estudiados, se escogería aplicar el aditivo a una temperatura de 50°C.

Código R

```
library(daewr);library(rsm)
library(gmodels);library(car)
rm(list=ls(all=TRUE))
#Función usuario para transformar predictores
#a unidades codificadas de -1 y 1
codificación=function(factor){
fcod=(factor-mean(factor))/((max(factor)-min(factor))/2)
fcod
}

#Lectura de los datos
#Se leyeron niveles del factor A: como 0 (sin aditivos) y
#1 (con aditivos)
datos=data.frame(scan(what=list(aditivos=0,temperatura=0,
                 adhesion=0)))
0 50 2.3
0 50 2.9
0 50 3.1
0 50 3.2
0 60 3.4
0 60 3.7
0 60 3.6
0 60 3.2
1 50 4.3
1 50 3.9
1 50 3.9
1 50 4.2
1 60 3.8
1 60 3.8
1 60 3.9
1 60 3.5

attach(datos)
X1=codificación(factor=aditivos)
X2=codificación(factor=temperatura)

data.frame(aditivos,X1,temperatura,X2)
mean(adhesion) #estimación media global

#Gráficos descriptivos
interaction.plot(aditivos,temperatura,adhesion,type="b",
                  pch=c(1,2),col=c("black","red"),lwd=4)

boxplot(adhesion~aditivos*temperatura,boxwex=0.4,
          col=c("bisque","blue"),
          xlab="tratamientos",ylab="ahesión",
          names=c("(-1,-1)","(+1,-1)","(-1,+1)","(+1,+1)"))

#Ajuste modelo Anova
MAnova=aov(adhesion~aditivos*temperatura)
anova(MAnova)

#Ajuste MRLM equivalente, con variables codificadas
MRLM=lm(adhesion~X1*X2)
summary(MRLM)

Anova(MRLM)
Anova(MRLM, type="III")

detach(datos)
```

ANOVA

Como el MRLM en (2) es equivalente al modelo factorial de los factores A, B con interacción, y hay una descomposición ortogonal de la suma de cuadrados totales entre las sumas de cuadrados SSA, SSB, SS(AB), tenemos que, las sumas de cuadrados tipo I, II y III del MRLM son iguales entre sí y equivalen a las sumas de cuadrados en el modelo Anova,

Anova en modelo factorial 2^2

Fuente	gl	SC	CM	F	VP
A	1	SSA	SSA	$F_A = \frac{SSA}{MSE}$	$P\left(f_{1,\text{dfe}} > F_A\right)$
B	1	SSB	SSB	$F_B = \frac{SSB}{MSE}$	$P\left(f_{1,\text{dfe}} > F_B\right)$
AB	1	$SS(AB)$	$F_{AB} = SS(AB)$	$F_{AB} = \frac{SS(AB)}{MSE}$	$P\left(f_{1,\text{dfe}} > F_{AB}\right)$
Error	dfe	SSE	MSE		

Anova con sumas de cuadrados tipo III en el MRLM equivalente

Fuente	gl	SC	CM	F	VP
X1	1	$SSR(X1\|X2, X1*X2)$	$SSR(X1\|X2, X1*X2)$	$\frac{SSR(X1\|X2,X1*X2)}{MSE}$	$P\left(f_{1,\text{dfe}} > F_{X_1}\right)$
X2	1	$SSR(X2\|X1, X1*X2)$	$SSR(X2\|X1, X1*X2)$	$\frac{SSR(X2\|X1,X1*X2)}{MSE}$	$P\left(f_{1,\text{dfe}} > F_{X_2}\right)$
X1*X2	1	$SSR(X1*X2\|X1, X2)$	$SSR(X1*X2\|X1, X2)$	$\frac{SSR(X1*X2\|X1,X2)}{MSE}$	$P\left(f_{1,\text{dfe}} > F_{X_1X_2}\right)$
Error	dfe	SSE	MSE		

$\text{dfe} = 2^2(n-1)$

$$SST = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} Y_{ijk}^2 - \frac{Y_{\bullet\bullet\bullet}^2}{2^2 n} \tag{9}$$

$$SSA = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} \widehat{\alpha}_i^2 = n\widehat{\alpha}_A^2 \tag{10}$$

$$SSB = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} \widehat{\beta}_j^2 = n\widehat{\alpha}_B^2 \tag{11}$$

$$SS(AB) = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} \widehat{\alpha\beta}_{ij}^2 = n\widehat{\alpha}_{AB}^2 \tag{12}$$

$$SSE = SST - SSA - SSB - SS(AB) \tag{13}$$

donde $Y_{\bullet\bullet\bullet} = \sum_{k=1}^{n} Y_{ijk}$. En general las SS en el Anova debidas a los efectos principales y de interacciones presentes en los experimentos factoriales 2^k, es una suma de cuadrados debidas a contrastes de totales de tratamientos observados, es decir, son de la forma

$$SS_w = \frac{\widehat{w}^2}{n\sum_{l=1}^{2^k} c_l^2} = \frac{\widehat{w}^2}{n2^k} \tag{14}$$

donde los $c_l \in \{-1, +1\}$ son los pesos que recibe cada total en el contraste respectivo, y por tanto, $\sum_{l=1}^{2^k} c_l^2 = \sum_{l=1}^{2^k} 1 = 2^k$.

Ahora observe la equivalencia entre tests de hipótesis

Fuente	ANOVA	MRLM	Test sobre efectos
A	$H_0: \alpha_1 = \alpha_2 = 0$ $H_1: \alpha_i \neq 0$ para algún $i = 1, 2$ $F_0 = \frac{MSA}{MSE}$	$H_0:\ \beta_1 = 0$ $H_1:\ \beta_1 \neq 0$ $F_0 = \frac{SSR(X_1 \mid X_2, X_1 * X_2)}{MSE}$	$H_0: \alpha_A = 0$ $H_1: \alpha_A \neq 0$ $F_A = \frac{SS_{w_A}}{MSE}$
B	$H_0: \beta_1 = \beta_2 = 0$ $H_1: \beta_j \neq 0$ para algún $j = 1, 2$ $F_0 = \frac{MSB}{MSE}$	$H_0:\ \beta_2 = 0$ $H_1:\ \beta_2 \neq 0$ $F_0 = \frac{SSR(X_2 \mid X_1, X_1 * X_2)}{MSE}$	$H_0: \alpha_B = 0$ $H_1: \alpha_B \neq 0$ $F_B = \frac{SS_{w_B}}{MSE}$
AB	$H_0: (\alpha\beta)_{ij} = 0\ \forall\ i, j = 1, 2$ $H_1: (\alpha\beta)_{ij} \neq 0$ para algún $i, j = 1, 2$ $F_0 = \frac{MS(AB)}{MSE}$	$H_0:\ \beta_{12} = 0$ $H_1:\ \beta_{12} \neq 0$ $F_0 = \frac{SSR(X_1 * X_2 \mid X_1, X_2)}{MSE}$	$H_0: \alpha_{AB} = 0$ $H_1: \alpha_{AB} \neq 0$ $F_{AB} = \frac{SS_{w_{AB}}}{MSE}$

En todos los casos la distribución del estadístico es $f_{1, 2^2(n-1)}$

En el experimento sobre la capacidad de ahesión de productos de caucho,

```
> MAnova=aov(adhesion~aditivos*temperatura)
> anova(MAnova)
Analysis of Variance Table
Response: adhesion
                     Df  Sum Sq Mean Sq F value   Pr(>F)
aditivos              1 2.17563 2.17563 30.6246 0.000129 ***
temperatura           1 0.07563 0.07563  1.0645 0.322534
aditivos:temperatura  1 0.85563 0.85563 12.0440 0.004627 **
Residuals            12 0.85250 0.07104
#-----------------------------------------------------------
> MRLM=lm(adhesion~X1*X2)
> Anova(MRLM)
Anova Table (Type II tests)
Response: adhesion
           Sum Sq Df F value   Pr(>F)
X1        2.17563  1 30.6246 0.000129 ***
X2        0.07563  1  1.0645 0.322534
X1:X2     0.85563  1 12.0440 0.004627 **
Residuals 0.85250 12
#-----------------------------------------------------------
> Anova(MRLM, type="III")
Anova Table (Type III tests)
Response: adhesion
             Sum Sq Df   F value    Pr(>F)
(Intercept) 200.931  1 2828.3490 1.285e-15 ***
X1            2.176  1   30.6246  0.000129 ***
X2            0.076  1    1.0645  0.322534
X1:X2         0.856  1   12.0440  0.004627 **
Residuals     0.853 12
```

Compare las sumas de cuadrados en la Anova del modelo factorial con las sumas de cuadrados Tipo II y tipo III en el MRLM, así como el valor del SSE, MSE, estadísticos F y valores P. En la siguiente Tabla se han calculado las SS_w según la ec. (14), es decir $SS_w = \dfrac{\widehat{w}^2}{2^k n}$, con $k = 2$ y $n = 4$,

	A	B	AB
contr. Totales $\widehat{w}$	5.9	1.1	-3.7
efectos $\widehat{\alpha}$	0.7375	0.1375	-0.4625
SS_w	2.175625	0.075625	0.855625
$SST = 3.959375$			

Entonces,

$$SSE = 3.959375 - 2.175625 - 0.075625 - 0.855625 = 0.8525$$

de donde,

Anova experimento 2^2

Fuente	gl	SS	CM	F0
A	1	2.175625	2.175625	30.624633
B	1	0.075625	0.075625	1.064516
AB	1	0.855625	0.855625	12.043988
Error	12	0.8525	0.071042	

Contenido

Experimento factorial 2^k general

Considere ahora el caso general, es decir, con $k \geq 2$ factores de efectos fijos, cada uno con dos niveles: bajo (-1) y alto $(+1)$.

Construcción de la matriz de signos

Esta matriz de 2^k filas se construye así:

1. Columna 1: Corresponde al factor A. Se llena alternando los signos $-$ y $+$ hasta completar los 2^k renglones.
2. Columna 2: Corresponde al factor B. Se alternan dos signos $-$ con dos signos $+$, hasta completar los 2^k renglones.
3. Columna 3: Corresponde al factor C. Se alternan cuatro signos $-$ con cuatro signos $+$, hasta completar los 2^k renglones.
4. Así sucesivamente, hasta la k-ésima columna compuesta de 2^{k-1} signos $-$ seguidos de 2^{k-1} signos $+$.
5. Los signos en las columnas de lasinteracciones interacciones resultan de multiplicar las columnas de los factores que aparecen en cada tipo de interacción.

Tabla 1: Familia de diseños factoriales 2^k, $k \leq 5$. Adaptado de Gutiérrez y de la Vara Salazar (2012)

Tratamiento	Notación de Yates	A	B	C	D	E	Tratamiento	Notación de Yates	A	B	C	D	E
1	(1)	−	−	−	−	−	17	e	−	−	−	−	+
2	a	+	−	−	−	−	18	ae	+	−	−	−	+
3	b	−	+	−	−	−	19	be	−	+	−	−	+
4	ab	+	+	−	−	−	20	abe	+	+	−	−	+
5	c	−	−	+	−	−	21	ce	−	−	+	−	+
6	ac	+	−	+	−	−	22	ace	+	−	+	−	+
7	bc	−	+	+	−	−	23	bce	−	+	+	−	+
8	abc	+	+	+	−	−	24	abce	+	+	+	−	+
9	d	−	−	−	+	−	25	de	−	−	−	+	+
10	ad	+	−	−	+	−	26	ade	+	−	−	+	+
11	bd	−	+	−	+	−	27	bde	−	+	−	+	+
12	abd	+	+	−	+	−	28	abde	+	+	−	+	+
13	cd	−	−	+	+	−	29	cde	−	−	+	+	+
14	acd	+	−	+	+	−	30	acde	+	−	+	+	+
15	bcd	−	+	+	+	−	31	bcde	−	+	+	+	+
16	abcd	+	+	+	+	−	32	abcde	+	+	+	+	+

Nota 3.1

Observe la notación de Yates en cada tratamiento. Es claro que el código (1) se usa cuando todos los factores están en su nivel bajo, y para los demás casos, en la cadena de caracteres se colocan en orden alfabético, las letras minúsculas de los factores que en el tratamiento se encuentran en el nivel alto.

Nota 3.2

Podemos escribir el modelo para el factorial 2^k en un DCA, bien sea como modelo Anova (media global más efectos principales más efectos de interacciones dobles, triples etc, más error) con sus supuestos y restricciones lineales sobre los distintos tipos de efectos, o a través del MRLM equivalente en el cual definimos las variables codificadas $X_j = -1$ para el nivel bajo y $X_j = +1$, para el nivel alto, con $j = 1, 2, \ldots, k$ indicando el j-ésimo factor, donde las interacciones dobles son representadas por los productos cruzados $X_i X_j$, $i \neq j$, las interacciones triples por los productos cruzados $X_i X_j X_l$, con $i \neq j \neq l$, etc. Pero con muchos factores resulta engorrosa la escritura de los modelos.

Estimación

Usando contrastes de totales de tratamientos. Siguiendo a Gutiérrez y de la Vara Salazar (2012),

$$\widehat{\alpha}_{ABC\cdots K} = \frac{(\text{Contraste totales})_{ABC\cdots K}}{n2^{k-1}}. \tag{15}$$

La suma de cuadrados de un contraste de totales (ver de nuevo ec. (14)),

$$SS(ABC\cdots K) = \frac{[(\text{Contraste totales})_{ABC\cdots K}]^2}{n2^k}. \tag{16}$$

También podemos escribir la sumas de cuadrados en función de los efectos,

$$SS(ABC\cdots K) = n2^{k-2}(\widehat{\alpha}_{ABC\cdots K})^2 \tag{17}$$

Para la suma SST cambiamos un poco la notación para simplificar la escritura,

$$SST = \sum_{i=1}^{n2^k} Y_i^2 - \frac{Y_\bullet^2}{n2^k}, \tag{18}$$

con $n2^k - 1$ grados de libertad, Y_i cada respuesta observada en el experimento, $Y_\bullet = \sum_{i=1}^{n2^k} Y_i$ la suma de las $N = n2^k$ observaciones del experimento.

Para la estimación de los coeficientes de regresión del modelo de RLM equivalente usando variables codificadas, tenemos la siguiente relación general con los efectos estimados,

$$\widehat{\beta} = \frac{\widehat{\alpha}}{2}, \tag{19}$$

mientras que la estimación del intercepto es simplemente el promedio de todas las observaciones,

$$\widehat{\beta}_0 = \bar{Y}_\bullet = \frac{Y_\bullet}{n2^k}. \tag{20}$$

Ejemplo 9.7 de Notas de Clase, página 331-343

En un experimento para estudiar un sistema particular de filtración de carbón, se agrega un coagulante a una solución en un tanque que contiene carbón y lodo, el cual se coloca entonces en un sistema de recirculación a fin de que se pueda lavar el carbón. Se hacen variar tres factores en el proceso experimental:

- FACTOR A: porcentaje de sólidos que circulan inicialmente en el sobre flujo; niveles: 20 % y 40 %.
- FACTOR B: Tasa de flujo del polímero; niveles: 5 lb/s y 10 lb/s
- FACTOR C: PH del tanque; niveles: 5 y 5.5

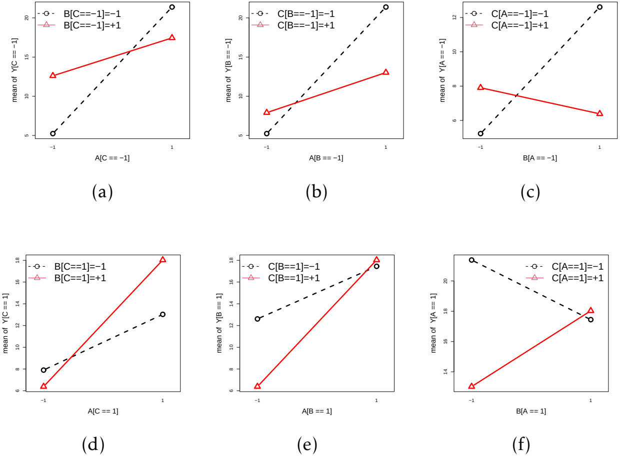
La cantidad de sólidos en el flujo inferior del sistema de purificación determina qué tan limpio queda el carbón. Se usan dos niveles de cada factor y se realizan dos corridas experimentales para cada una de las 8 combinaciones. Las respuestas, porcentaje de sólidos por peso, en el flujo inferior del sistema de circulación se especifican en la siguiente tabla:

Combinación tratamientos	réplica 1	réplica 2
(1)	4.65	5.81
a	21.42	21.35
b	12.66	12.56
ab	18.27	16.62
c	7.93	7.88
ac	13.18	12.87
bc	6.51	6.26
abc	18.23	17.83

ir a archivo Tablacalculosejemplo9-7notasdeclase.pdf.

```
#----------------------------------------
#Con ejemplo 9.7 de Notas de Clase pp. 331
#----------------------------------------
rm(list=ls(all=TRUE))
library(daewr)
library(rsm)
library(pid)
library(FrF2)
datos=data.frame(scan(what=list(A=0,B=0,C=0,Y=0)))
-1 -1 -1 4.65
-1 -1 -1 5.81
 1 -1 -1 21.42
 1 -1 -1 21.35
-1  1 -1 12.66
-1  1 -1 12.56
 1  1 -1 18.27
 1  1 -1 16.62
-1 -1  1 7.93
-1 -1  1 7.88
 1 -1  1 13.18
 1 -1  1 12.87
-1  1  1 6.51
-1  1  1 6.26
 1  1  1 18.23
 1  1  1 17.83

attach(datos)

modelo=lm(Y~A*B*C)
summary(modelo)
anova(modelo)

detach(datos)
```

```
> summary(modelo)
Call: lm.default(formula = Y ~ A * B * C)
Residuals:
    Min      1Q  Median      3Q     Max
-0.8250 -0.1325  0.0000  0.1325  0.8250
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.751875   0.131162  97.222 1.40e-13 ***
A            4.719375   0.131162  35.981 3.90e-10 ***
B            0.865625   0.131162   6.600 0.000169 ***
C           -1.415625   0.131162 -10.793 4.79e-06 ***
A:B         -0.599375   0.131162  -4.570 0.001826 **
A:C         -0.528125   0.131162  -4.027 0.003807 **
B:C          0.005625   0.131162   0.043 0.966844
A:B:C        2.230625   0.131162  17.007 1.45e-07 ***
---
Residual standard error: 0.5246 on 8 degrees of freedom
Multiple R-squared:  0.9955,    Adjusted R-squared:  0.9916
F-statistic: 254.4 on 7 and 8 DF,  p-value: 9.297e-09

> anova(modelo)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq   F value    Pr(>F)
A          1 356.36  356.36 1294.6482 3.899e-10 ***
B          1  11.99   11.99   43.5554 0.0001694 ***
C          1  32.06   32.06  116.4875 4.788e-06 ***
A:B        1   5.75    5.75   20.8824 0.0018264 **
A:C        1   4.46    4.46   16.2127 0.0038065 **
B:C        1   0.00    0.00    0.0018 0.9668436
A:B:C      1  79.61   79.61  289.2251 1.451e-07 ***
Residuals  8   2.20    0.28
---
```



(a) (b) (c)

(d) (e) (f)

Figura 1: Gráficas para evaluar la interacción triple. (a) y (d) Interacción AB en niveles bajo y alto de C, respectivamente; (b) y (e) Interacción AC en niveles bajo y alto de B, respectivamente; (c) y (f) Interacción BC en niveles bajo y alto de A, respectivamente. Desde que las interacciones entre dos factores cambian según niveles del tercero de los factores, eso nos indica posible interacción triple ABC significativa.

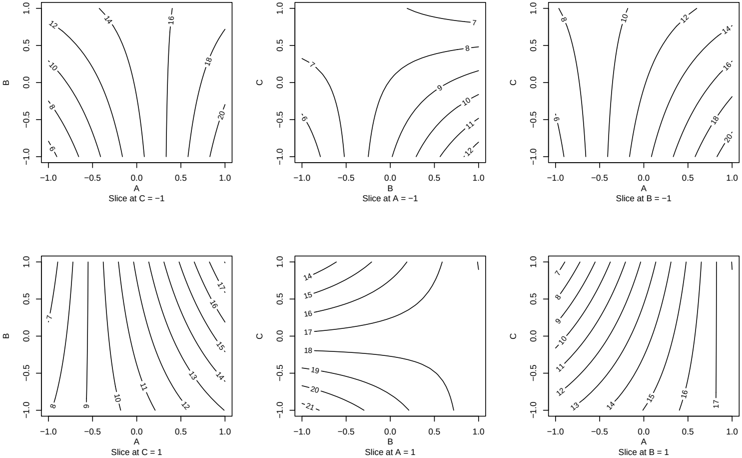


Figura 2: Contornos de respuesta o curvas de nivel. La respuesta con máximo valor cuando $A = 1$ es de aprox. 21 cuando $B = C = -1$, en tanto que cuando $A = -1$, la respuesta con máximo valor es de aprox. 12 y se obtiene cuando $B = 1$ y $C = -1$, entonces la condición que maximiza la respuesta es con $A = +1$, $B = -1$ y $C = -1$.

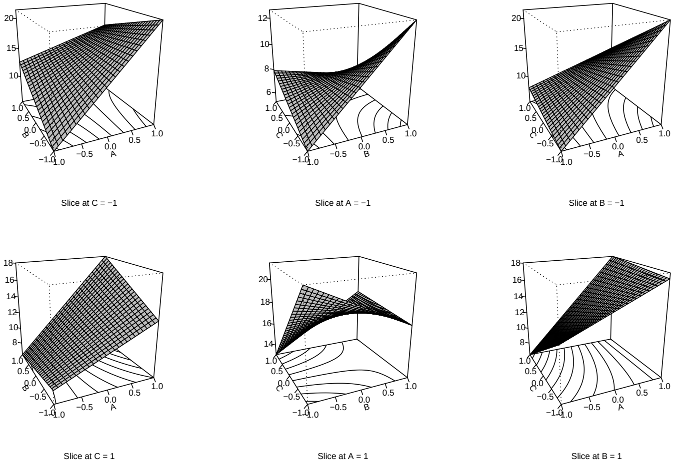


Figura 3: Superficies de respuesta. Con estas gráficas se llega a la misma conclusión alcanzada evaluando los contornos de respuesta, con relación al mejor tratamiento experimental, es decir, $A = +1$, $B = -1$ y $C = -1$ da la máxima respuesta media.

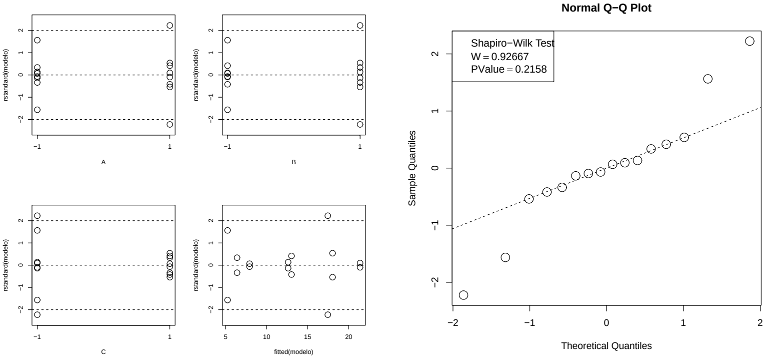


Figura 4: Gráficos de residuos estudentizados internamente y de probabilidad normal. Hay problemas con supuesto errores con distribución normal y varianza constante.

Dean, A., Voss, D., and Draguljić, D. (2017). *Design and Analysis of Experiments*, 2nd Edition. Springer.

Gutiérrez Pulido, H. y de la Vara Salazar, R. (2012). *Análisis y Diseño de Experimentos*, 3ª Edición. McGraw-Hill.

Kuehl, R. O. (2001). *Diseño de Experimentos. Principios Estadísticos de Diseño y Análisis de Investigación*, 2ª Edición. Thomson Learning.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*, 5th Edition. McGraw-Hill Irwin.

Montgomery, D. C. (2020). *Design and Analysis of Experiments*, 10th Edition. John Wiley & Sons, Inc.