Ejemplo 6.1 del libro de Montgomery "Análisis y Diseño de Experimento"

Considere una investigación llevada a cabo para estudiar el efecto que tienen la concentración de un reactivo y la presencia de un catalizador sobre el tiempo de reacción de un proceso químico. Sea la concentración del reactivo el factor *A* con dos niveles de interés, 15 y 25%. El catalizador constituye el factor *B;* el nivel alto (o superior) denota el uso de dos sacos de catalizador y el nivel bajo (o inferior) denota el uso de sólo un saco. El experimento se realiza ("réplica" o "repite") tres veces, y los datos son como sigue:

Factor		Combinación de	Réplica			
A	*B*	tratamientos	I	II	III	Total
–	–	*A* bajo, *B* bajo	28	25	27	80
+	–	*A* alto, *B* bajo	36	32	32	100
–	+	*A* bajo, *B* alto	18	19	23	60
+	+	*A* alto, *B* alto	31	30	29	90

En la Fig 1 se presentan gráficamente las combinaciones de tratamientos para este diseño. Por convención, el efecto de un factor se denota por la letra latina mayúscula. De este modo, *"A"* se refiere al efecto del factor *A*, *"B"* se refiere al efecto del factor *B*, y *"AB"* se refiere a la interacción *AB.* En el diseño 2^2, los niveles bajo y alto de *A* y *B* se denotan por "-" y "+ ", respectivamente, en los ejes *A* y *B*. Así, - en el eje *A* representa el nivel bajo de concentración (15%), mientras que + representa el nivel alto (25%), y - en el eje *B* representa el nivel bajo de catalizador mientras que + denota el nivel alto.

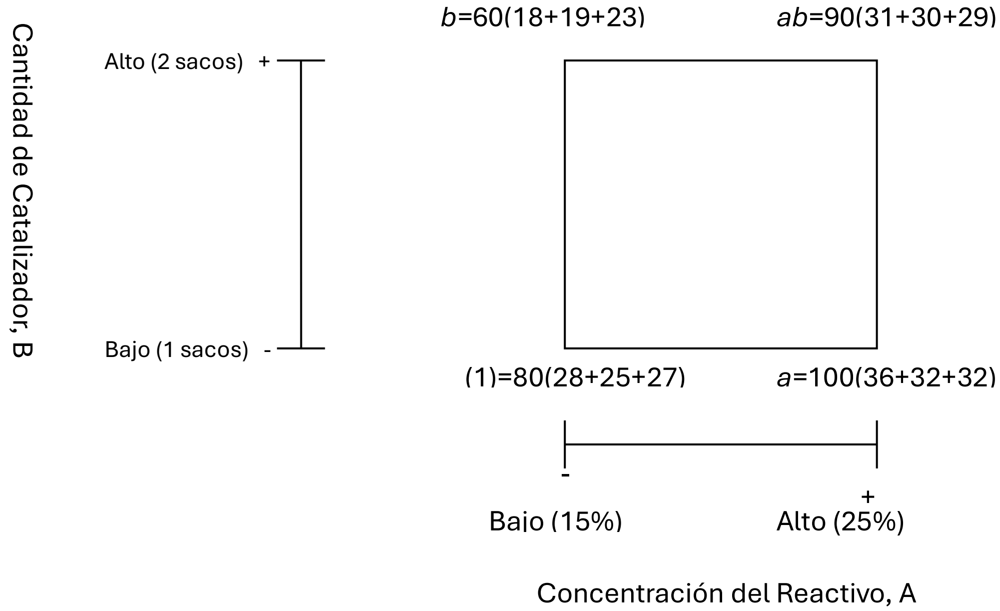

Figura N° 1 Combinaciones de tratamientos en el diseño 2^2

Las cuatro combinaciones de tratamientos en el diseño suelen representarse por letras minúsculas, como se muestra en la Fig. 1. En esta figura se aprecia que el nivel superior de cualquier factor de una combinación de tratamientos está representado por la presencia de la letra minúscula correspondiente, mientras que la ausencia de esta última representa el nivel inferior del factor. Así *a* representa la combinación de tratamientos, en la que *A* se encuentra en el nivel superior y *B* en el inferior; *b* representa aquella en la que *A* se halla en el nivel inferior y *B* en el superior, y *ab* representa a ambos factores en el nivel superior. Por convención (1) se usa para representar a ambos factores en el nivel inferior. Esta notación se usará a lo largo de toda la serie 2^k. Las letras minúsculas (1), *a, b* y *ab* también se usan para representar los totales de la variable respuesta de las *n* réplicas de las combinaciones de tratamientos correspondientes

Los efectos de los factores

$$A = \frac{1}{2(3)}(90 + 100 - 60 - 80) = 8.33$$

$$B = \frac{1}{2(3)}(90 + 60 - 100 - 80) = -5.00$$

$$AB = \frac{1}{2(3)}(90 + 80 - 100 - 60) = 1.67$$

Obteniendo las sumas de cuadrados y ANVA.

```
> # sumas de cuadrados y ANVA
> anva<-aov(mod)
> summary(anva)
            Df Sum Sq Mean Sq F value   Pr(>F)    
A            1 208.33  208.33  53.191 8.44e-05 ***
B            1  75.00   75.00  19.149  0.00236 ** 
A:B          1   8.33    8.33   2.128  0.18278    
Residuals    8  31.33    3.92                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.
```

Hipótesis sobre la interacción

H_o:no existe un efecto significativo de la interacción sobre el rendimiento

H_a: existe un efecto significativo de la interacción sobre el rendimiento

F=2.128, pvalor = 0.18278.

Conclusión: no existe un efecto significativo de la interacción sobre el rendimiento

Hipótesis sobre los efectos principales

H_o:no existe un efecto significativo del factor A sobre el rendimiento

H_a: existe un efecto significativo del factor A sobre el rendimiento

F=53.191, pvalor = 8.44e-5 < 0.05, se rechaza la Ho

Conclusión: Existe un efecto significativo del factor A sobre el rendimiento

H_o:no existe un efecto significativo del factor B sobre el rendimiento

H_a: existe un efecto significativo del factor B sobre el rendimiento

F=19.149, pvalor = 0.0023 < 0.05, se rechaza la Ho

Conclusión: Existe un efecto significativo del factor B sobre el rendimiento

Verificar supuestos

Modelo de Regresión para verificar supuestos

- El modelo de regresión que se considera para realizar el chequeo de supuesto está dado por:

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

donde :

$$x_1 = \frac{Conc - \left(Conc_{baja} + Conc_{alta}\right)/2}{\left(Conc_{alta} - Conc_{baja}\right)/2} = \frac{Conc - \left(15+25\right)/2}{\left(25-15\right)/2} = \frac{Conc - 20}{5}$$

$$x_2 = \frac{Catal - \left(Catal_{baja} + Catal_{alta}\right)/2}{\left(Catal_{alta} - Catal_{baja}\right)/2} = \frac{Catal - \left(1+2\right)/2}{\left(2-1\right)/2} = \frac{Catal - 1.5}{0.5}$$

En este caso, $x_1 = \begin{cases} -1, \text{ si conc=15} \\ 1, \text{ si conc=25} \end{cases}$ y $x_2 = \begin{cases} -1, \text{ si catal=1} \\ 1, \text{ si catal=2} \end{cases}$

Luego, con los datos del **ejemplo 6.1**, la ecuación de regresión estimada está dado por:

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)x_1 + \left(\frac{-5.00}{2}\right)x_2$$

```
H<-X%*%solve(t(X)%*%X)%*%t(X)
ri<-e/sqrt(CME*(1-diag(H)))
ri
            [,1]
[1,]  1.19170805
[2,] -0.45834925
[3,]  0.64168895
[4,]  1.00836835
[5,] -1.19170805
[6,] -1.19170805
[7,] -1.55838744
[8,] -1.00836835
[9,]  1.19170805
[10,]  1.00836835
[11,]  0.45834925
[12,] -0.09166985
```

```
> ri<-rstandard(mod1)
> ri
          1           2           3           4 
 1.19170805 -0.45834925  0.64168895  1.00836835 
          5           6           7           8 
-1.19170805 -1.19170805 -1.55838744 -1.00836835 
          9          10          11          12 
 1.19170805  1.00836835  0.45834925 -0.09166985 
> ri
          1           2           3           4 
 1.19170805 -0.45834925  0.64168895  1.00836835 
          5           6           7           8 
-1.19170805 -1.19170805 -1.55838744 -1.00836835 
          9          10          11          12 
 1.19170805  1.00836835  0.45834925 -0.09166985 
```

Se ha comprobado los supuestos, donde se verifico normalidad

```
> shapiro.test(ri)

        Shapiro-Wilk normality test

data:  ri
W = 0.88179, p-value = 0.09239
```

Conclusión: se cumple el supuesto de normalidad en los residuos

```
> library(nortest)
> ad.test(ri)

        Anderson-Darling normality test

data:  ri
A = 0.55037, p-value = 0.1216
```

Con la prueba de Anderson Darling, también se verifico que cumple el supuesto de normalidad.

Prueba de varianza constante

```
> # VARIANZA CONSTANTE-- BREUSCH PAGAN
> library(car)
Cargando paquete requerido: carData
> ncvTest(mod1)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.1148473, Df = 1, p = 0.73469
```

Como pvalor>0.05, no se rechaza la hipótesis nula, se cumple el supuesto de homogeneidad.

Gráfica de la superficie

LaeEcuación de regression en términos de las variables codificada está dada por

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)x_1 + \left(\frac{-5.00}{2}\right)x_2 = 27.5 + \left(\frac{8.33}{2}\right)\left(\frac{A-20}{5}\right) - \left(\frac{5}{2}\right)\left(\frac{B-1.5}{0.5}\right)$$

$$\hat{y} = 27.5 - 16.6666 + 7.5 + 0.83333A - 5B$$

$$\hat{y} = 16.3333 + 0.83333A - 5B$$