caso 3. Diseño factorial

Arturo Zuñiga

2025-09-20

Caso 3

Un proceso de producción química consiste de una primera reacción con un alcohol y una segunda reacción con una base. Se realizó un experimento factorial 3x2 con tres alcoholes y dos bases, con cuatro reacciones réplica en un diseño totalmente aleatorizado (completamente al azar). Los datos se reunieron como porcentaje de la reacción.

¿Qué significa "porcentaje de reacción"?

Cuando se hace una reacción química, no siempre toda la materia prima se convierte en producto.

Por ejemplo, si mezclas un alcohol con una base para obtener un compuesto, en teoría podrías obtener el **100% del producto esperado**.

Pero en la práctica, siempre hay pérdidas: parte del alcohol o de la base puede no reaccionar, la reacción puede detenerse antes de completarse, o pueden formarse subproductos.

Por eso, los químicos miden el **rendimiento de la reacción**, que se expresa en forma de porcentaje:

$$\text{Porcentaje de reacción} = \frac{\text{cantidad de producto obtenido}}{\text{cantidad de producto esperado (teórico)}} \times 100$$

Ejemplo sencillo

Imagina que, según los cálculos teóricos, al mezclar un alcohol con una base deberías obtener **10 gramos de producto**.

Si en la práctica obtienes **8 gramos**, entonces el porcentaje de reacción es:

$$\frac{8}{10} \times 100 = 80\%$$

Ese **80%** significa que la reacción fue bastante eficiente, aunque no perfecta.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

Variable respuesta: … porcentaje de reaccion

Factores en estudio: … A: tipo de Alcohol B: Base

Tratamientos: … 6 tratat a1b1 a2b1 a3b1 a1b2 a2b2 a3b2

Lectura y presentación de datos

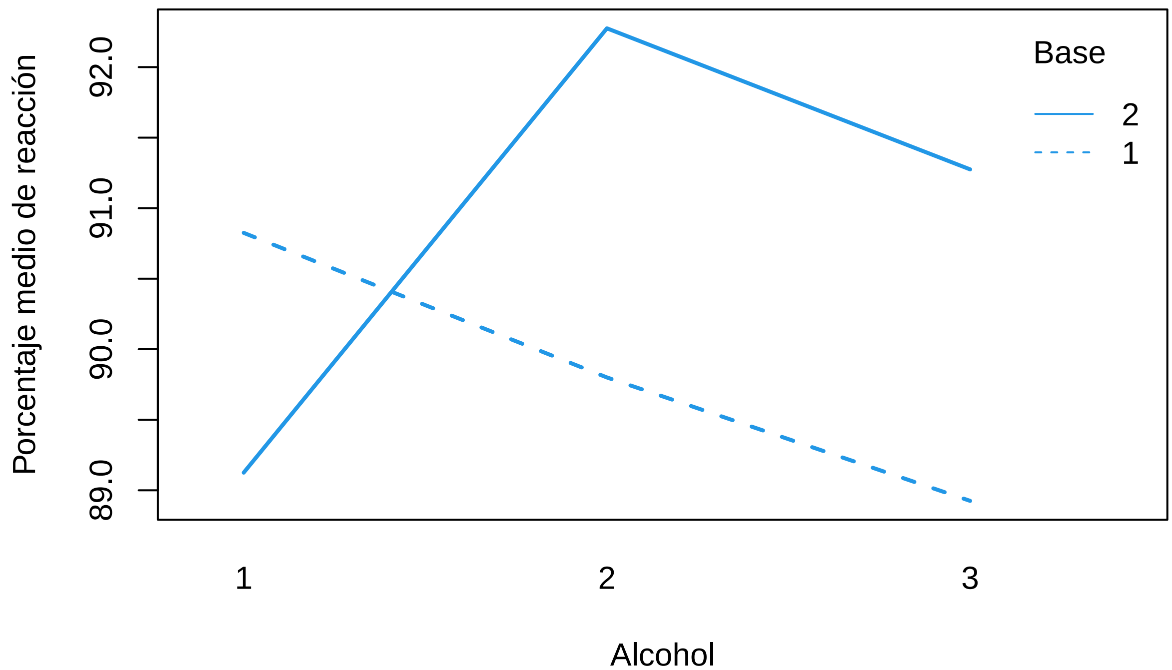
```r
rm(list=ls())
datos.3 = read.table("datos3.txt",T)
datos.3
```

```
##    Base Alcohol Porcentaje
## 1     1       1       91.3
## 2     1       1       90.7
## 3     2       1       87.3
## 4     2       1       91.5
## 5     1       1       89.9
## 6     1       1       91.4
## 7     2       1       89.4
## 8     2       1       88.3
## 9     1       2       89.3
## 10    1       2       90.4
## 11    2       2       92.3
## 12    2       2       90.6
## 13    1       2       88.1
## 14    1       2       91.4
## 15    2       2       91.5
## 16    2       2       94.7
## 17    1       3       89.5
## 18    1       3       88.3
## 19    2       3       93.1
## 20    2       3       91.5
## 21    1       3       87.6
## 22    1       3       90.3
## 23    2       3       90.7
## 24    2       3       89.8
```

```r
attach(datos.3)
Alcohol = as.factor(Alcohol)
Base    = as.factor(Base)
```

Interacción

```r
interaction.plot(Alcohol, Base, Porcentaje, lwd = 2, col = 4, ylab = "Porcentaje medio de reacción")
```



Modelo

```
mod.ef3 = lm(Porcentaje ~ Alcohol * Base)
summary(aov(mod.ef3))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Alcohol        2   5.40   2.698   1.321 0.2915
## Base           1   6.51   6.510   3.188 0.0910 .
## Alcohol:Base   2  22.57  11.283   5.525 0.0135 *
## Residuals     18  36.76   2.042
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$H_0 : (\alpha\beta)_{ij} = 0$ no existe efecto significativo de la interaccion de los factores sobre el porcentaje de reaccion

$H_1 : (\alpha\beta)_{ij} \neq 0$ para al menos un i, j.

$\alpha = 0.10$

$p - valor = 0.0135$ <0.10, se rechaza la Ho como verdadera.

Entonces la interacción entre Alcohol y Base es significativa. Debe realizarse el análisis de EFECTOS SIMPLES, sin embargo antes de ello conviene revisar los supuestos del modelo.

Verificación de supuestos

```
library(nortest)
shapiro.test(residuals(mod.ef3))
```

```
##
##  Shapiro-Wilk normality test
```

Ho: los residuos tienen distribucion normal
Ha: los residuos no tienen distribucion normal

```
##
## data:  residuals(mod.ef3)
## W = 0.95049, p-value = 0.2777
```

>0.10, presenta normalidad

```
ad.test(residuals(mod.ef3))
```

```
##
##  Anderson-Darling normality test
##
## data:  residuals(mod.ef3)
## A = 0.33394, p-value = 0.4837
```

```
library(car)
```

```
## Cargando paquete requerido: carData
```

```
ncvTest(mod.ef3)
```

Ho: la varianza de los residuos son contantes

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.01923634, Df = 1, p = 0.88969
```

```
library(lmtest)
```

```
## Cargando paquete requerido: zoo
```

```
##
## Adjuntando el paquete: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
dwtest(mod.ef3,alternative = c("two.sided"))
```

Ho: los errores son independientes
(no existe autocorrelacion de los residuos).

```
##
##  Durbin-Watson test
##
## data:  mod.ef3
## DW = 2.5673, p-value = 0.4686
## alternative hypothesis: true autocorrelation is not 0
```

>0.10, los errores son independientes

ANÁLISIS DE EFECTOS SIMPLES

Fijamos el nivel de Alcohol y comparamos las Bases:

En el primer nivel del alcohol:

$H_0 : \mu_{11} = \mu_{12}$

$H_1 : \mu_{11} \neq \mu_{12}$

En el segundo nivel del alcohol:

$H_0 : \mu_{21} = \mu_{22}$

$H_1 : \mu_{21} \neq \mu_{22}$

En el tercer nivel del alcohol:

$H_0 : \mu_{31} = \mu_{32}$

$H_1 : \mu_{31} \neq \mu_{32}$

```r
# install.packages(phia)
library(phia)
testInteractions(mod.ef3, fixed="Alcohol", across="Base")
```

```
## F Test:
## P-value adjustment method: holm
##            Value     SE Df Sum of Sq      F  Pr(>F)
## 1          1.700 1.0105  1     5.780 2.8304 0.10976
## 2         -2.475 1.0105  1    12.251 5.9994 0.07433 .
## 3         -2.350 1.0105  1    11.045 5.4087 0.07433 .
## Residuals               18    36.758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- En el primer nivel del alcohol, no existen diferencias significativas en las medias al usar Base 1 y Base 2.
- En el segundo nivel del alcohol, existe diferencias significativas en las medias al usar Base 1 y Base 2.
- En el tercer nivel del alcohol, existe diferencias significativas en las medias al usar Base 1 y Base 2.

```r
 aggregate(Porcentaje ~ Base*Alcohol, FUN = mean)
```

```
##   Base Alcohol Porcentaje
## 1    1       1     90.825
## 2    2       1     89.125
## 3    1       2     89.800
## 4    2       2     92.275
## 5    1       3     88.925
## 6    2       3     91.275
```

- En el primer nivel del alcohol, el porcentaje medio de reacci?n es el mismo al usar la Base 1 o Base 2.
- En el segundo nivel del alcohol, el porcentaje medio de reacci?n es mayor al usar la Base 2.
- En el tercer nivel del alcohol, el porcentaje medio de reacci?n tambi?n es mayor al usar la Base 2.

Fijamos el nivel de las Bases y comparamos los alcoholes

En el primer nivel de la base:

$H_0 : \mu_{11} = \mu_{21} = \mu_{31}$

H_1 : Al menos un μ_{i1} es distinto

En el segundo nivel de la base:

$H_0 : \mu_{12} = \mu_{22} = \mu_{32}$

H_1 : Al menos un μ_{i2} es distinto

```r
testInteractions(mod.ef3, fixed="Base", across="Alcohol")
```

```
## F Test:
## P-value adjustment method: holm
##           Alcohol1 Alcohol2    SE1    SE2 Df Sum of Sq      F  Pr(>F)
## 1             1.90    0.875 1.0105 1.0105  2     7.235 1.7715 0.19848
## 2            -2.15    1.000 1.0105 1.0105  2    20.727 5.0749 0.03574 *
## Residuals                                 18    36.758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- En el primer nivel de la base, no existen diferencias significativas en las medias al usar Alcohol 1, Alcohol 2 y Alcohol 3.
- En el segundo nivel de la base, existe diferencias significativas en al menos una de las medias al usar Alcohol 1, Alcohol 2, Alcohol 3.

```r
aggregate(Porcentaje ~ Alcohol*Base, FUN = mean)
```

```
##   Alcohol Base Porcentaje
## 1       1    1     90.825
## 2       2    1     89.800
## 3       3    1     88.925
## 4       1    2     89.125
## 5       2    2     92.275
## 6       3    2     91.275
```

Vemos que, efectivamente, los porcentajes medios cuando se usa la Base 1 son muy similares (90.825, 89.8, 88.925), sin embargo no sucede lo mismo al usar la Base 2.

```r
# install.packages("lsmeans")
library(lsmeans)
lsmeans(mod.ef3, pairwise ~ Alcohol|Base)
```

```
## $lsmeans
## Base = 1:
##  Alcohol lsmean    SE df lower.CL upper.CL
##  1         90.8 0.715 18     89.3     92.3
##  2         89.8 0.715 18     88.3     91.3
##  3         88.9 0.715 18     87.4     90.4
##
## Base = 2:
##  Alcohol lsmean    SE df lower.CL upper.CL
##  1         89.1 0.715 18     87.6     90.6
##  2         92.3 0.715 18     90.8     93.8
##  3         91.3 0.715 18     89.8     92.8
##
## Confidence level used: 0.95
##
## $contrasts
## Base = 1:
##  contrast            estimate   SE df t.ratio p.value
##  Alcohol1 - Alcohol2    1.025 1.01 18   1.014  0.5776
##  Alcohol1 - Alcohol3    1.900 1.01 18   1.880  0.1731
##  Alcohol2 - Alcohol3    0.875 1.01 18   0.866  0.6680
##
## Base = 2:
##  contrast            estimate   SE df t.ratio p.value
##  Alcohol1 - Alcohol2   -3.150 1.01 18  -3.117  0.0156
##  Alcohol1 - Alcohol3   -2.150 1.01 18  -2.128  0.1122
##  Alcohol2 - Alcohol3    1.000 1.01 18   0.990  0.5926
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Vemos que todos los p-valores en la Base 1 son altos, lo cual no contradice lo ya mencionado:

"En el primer nivel de la base, no existen diferencias significativas en las medias al usar Alcohol 1, Alcohol 2 y Alcohol 3."

Sin embargo para la Base 2: el porcentaje medio es significativamente distinto entre los Alcoholes 1 y 2 (pvalor = 0.0156); en todos los dem?s casos, Alcohol 1 versus Alcohol 3 y Alcohol 2 versus Alcohol 3, no se encuentran medias significativamente distintas.