Capítulo 7

Diseños en cuadrados latinos

7.1. Introducción

En el modelo en bloques aleatorizados, que estudiamos en el capítulo anterior, considerábamos un factor principal y un factor de control o variable de bloque que introduciamos con el objeto de eliminar su influencia en la variable respuesta y así reducir el error experimental. En este capítulo estudiaremos diseños que utilizan más de una variable de bloque para reducir el error experimental.

Así, si se consideran simultáneamente dos variables de bloque, un diseño completo en bloques aleatorizados consistiría en formar un bloque para cada combinación de niveles de dichas variables y después aplicar todos los niveles del factor principal en cada uno de los bloques obtenidos. Por ejemplo, supongamos un experimento en el que se quiere estudiar el efecto de distintos tipos de semilla en el rendimiento del trigo y se considera que en dicho rendimiento también pueden influir los tipos de abonos e insecticidas empleados. Para realizar dicho estudio, es posible utilizar un diseño completo en bloques aleatorizados, donde el factor principal es el tipo de semilla y las variables de bloque los tipos de abono e insecticida.

Un incoveniente que presentan a veces estos diseños es el de requerir excesivas unidades experimentales para su realización. Un diseño en bloques completos con un factor principal y dos factores de bloque, con K_1, K_2 y K_3 niveles en cada uno de los factores, requiere $K_1 \times K_2 \times K_3$ unidades experimentales. En nuestro ejemplo, si el factor principal, tipo de semilla, tiene 4 niveles, la primera variable de bloque, tipo de abono, 5 niveles, la segunda, tipo de insecticida, 3 niveles, se necesitarían $4 \times 5 \times 3 = 60$ unidades experimentales.

En un experimento puede haber diferentes causas, por ejemplo de índole económico, que no permitan emplear demasiadas unidades experimentales, ante esta situación se puede recurrir a un tipo especial de diseños en bloques incompletos aleatorizados. La idea básica de estos diseños es la de *fracción*; es decir, seleccionar una parte del diseño completo de forma que, bajo ciertas hipótesis generales, permita estimar los efectos que interesan.

Uno de los diseños en bloques incompletos aleatorizados más importante con dos factores de control es el modelo en *cuadrado latino*, dicho modelo requiere el mismo número de niveles para los tres factores.

En general, para K niveles en cada uno de los factores, el diseño completo en bloques aleatorizados utiliza K^2 bloques, aplicándose en cada bloque los K niveles del factor principal, resultando un total de K^3 unidades experimentales. Los diseños en cuadrado latino reducen el número de unidades experimentales a K^2 utilizando los K^2 bloques del experimento, pero aplicando sólo un tratamiento en cada bloque con una disposición especial. De esta forma, si K fuese 4, el diseño en bloques completos necesitaría $4^3 = 64$ observaciones, mientras que el diseño en cuadrado latino sólo necesitaría $4^2 = 16$ observaciones.

En este capítulo se van estudiar además del diseño en cuadrado latino otros diseños relacionados con él, como son los *cuadrados greco-latinos* que utilizan tres variables de bloque y los *cuadrados de Youden* que se pueden considerar como una variante de los cuadrados latinos, si bien también se pueden estudiar como bloques incompletos. En primer lugar, estudiaremos el diseño en cuadrado latino.

7.2. Diseños en cuadrados latinos

7.2.1. Descripción del modelo

Los diseños en *cuadrados latinos* son apropiados cuando es necesario controlar dos fuentes de variabilidad. En dichos diseños el número de niveles del *factor principal* tiene que coincidir con el número de niveles de las dos variables de bloque o *factores secundarios* y además hay que suponer que no existe interacción entre ninguna pareja de factores.

Supongamos que el número de niveles de cada uno de los factores es K. El diseño en cuadrado latino utiliza K^2 bloques, cada uno de estos bloques corresponde a una de las posibles combinaciones de niveles de los dos factores de control. En cada bloque se aplica un solo tratamiento de manera que cada tratamiento debe aparecer con cada uno de los niveles de los dos factores de control.

Si consideramos una tabla de doble entrada donde las filas y las columnas representan cada uno de los dos factores de bloque y las celdillas los niveles del factor principal o tratamientos, el requerimiento anterior supone que cada tratamiento debe aparecer una vez y sólo una en cada fila y en cada columna.

Recibe el nombre de *cuadrado latino de orden* K a una disposición en filas y columnas de K letras latinas, de tal forma que cada letra aparece una sola vez en cada fila y en cada columna.

A continuación vamos a dar una forma simple de construcción de cuadrados latinos. Se parte de una primera fila con las letras latinas ordenadas alfabéticamente

	Columna 1	Columna 2	Columna 3	$\cdots$	Columna k
Fila 1	A	B	C	$\cdots$	K

Las sucesivas filas se obtienen moviendo la primera letra de la fila anterior a la última posición (construcción por permutación cíclica), el cuadrado así obtenido es un *cuadrado latino estándar.* Un cuadrado latino se denomina *estándar* cuando las letras de la primera fila y la primera columna están ordenadas alfabéticamente. A parte de los cuadrados latinos así obtenidos existen otros cuadrados latinos diferentes, estándares y no estándares. En el Apéndice B se muestran algunos cuadrados latinos estándares para los órdenes 3, 4, 5, 6, 7, 8 y 9.

El procedimiento para construir un diseño en cuadrado latino es el siguiente:

1) Se elige aleatoriamente un cuadrado latino de los disponibles.

2) Se asigna aleatoriamente el orden de las filas y columnas.

3) Se asignan aleatoriamente los tres factores a las filas, columnas y letras, respectivamente.

Ilustremos este procedimiento con el ejemplo del rendimiento de la semilla de trigo. Al plantear este experimento se pensó que podría conseguirse mayor precisión si se controlaba la variabilidad introducida por los tipos de abono e insecticida. El instituto de experimentación agrícola está interesado en estudiar 4 tipos de semilla de trigo, $(s1, s2, s3, s4)$ y decide realizar el experimento utilizando un diseño en cuadrado latino. Para ello selecciona 4 niveles para cada una de las variables de bloque: abono, $(a1, a2, a3, a4)$, e insecticida, $(i1, i2, i3, i4)$.

La selección de uno de los cuadrados se hace al azar. Supongamos que el cuadrado latino elegido es el siguiente

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

A continuación, se asigna también al azar, el orden de las filas y las columnas. Supongamos que el orden seleccionado para las filas sea (2, 3, 1, 4), entonces el cuadrado latino anterior se convierte en

B	A	D	C
C	D	A	B
A	B	C	D
D	C	B	A

Se vuelven a generar otros 4 números aleatorios que se identifican con el orden de las columnas de este último cuadrado. Supongamos que los números obtenidos son (4, 3, 1, 2), obteniéndose el siguiente cuadrado latino

C	D	B	A
B	A	C	D
D	C	A	B
A	B	D	C

Por último, se asignan al azar las filas, las columnas y las letras latinas a los tres factores. Por ejemplo, supongamos que las filas, las columnas y las letras se asignan, respectivamente, a los tipos de insecticidas, semillas y abonos, de tal forma que el diseño resultante es

Tabla 5-1.

	Semillas			
Insecticidas	s1	s2	s3	s4
i1	a3	a4	a2	a1
i2	a2	a1	a3	a4
i3	a4	a3	a1	a2
i4	a1	a2	a4	a3

Por convenio, se suele situar el factor principal, en este caso el tipo de semilla, en las celdillas. Reordenando el diseño anterior se obtiene la siguiente tabla:

Tabla 5-2.

	Abonos			
Insecticidas	a1	a2	a3	a4
i1	s4	s3	s1	s2
i2	s2	s1	s3	s4
i3	s3	s4	s2	s1
i4	s1	s2	s4	s3

En resumen, podemos decir que un diseño en cuadrado latino tiene las siguientes características:

1^o) Se controlan tres fuentes de variabilidad, un factor principal y dos factores de bloque.

2^o) Cada uno de los factores tiene el mismo número de niveles, K.

3^o) Cada nivel del factor principal aparece una vez en cada fila y una vez en cada columna.

4^o) No hay interacción entre los factores.

7.2.2. Planteamiento del modelo

En un diseño en cuadrado latino intervienen los siguientes factores: un factor principal y dos factores secundarios o variables de bloque. Se supone que no existe interacción entre esos tres factores. Así el modelo empleado es un modelo adtivo.

Si consideramos que los tres factores son de efectos fijos, el modelo estadístico para este diseño es:

$$y_{ij(h)} = \mu + \tau_i + \beta_j + \gamma_h + +u_{ij(h)} \quad \left\{ \begin{array}{l} i = 1, 2 \ldots, K \\ j = 1, 2 \ldots, K \\ h = 1, 2 \ldots, K \end{array} \right. , \tag{7.1}$$

donde

- $y_{ij(h)}$ representa la observación correspondiente a la i-ésima fila, j-ésima columna y h-ésima letra latina.
- μ es la media global.
- τ_i es el efecto producido por el i-ésimo nivel del factor fila. Dichos efectos están sujetos a la restricción $\sum_i \tau_i = 0$.

- β_j es el efecto producido por el j-ésimo nivel del factor columna. Dichos efectos están sujetos a la restricción $\sum_j \beta_j = 0$.
- γ_h es el efecto producido por la h-ésima letra latina. Dichos efectos están sujetos a la restricción $\sum_h \gamma_h = 0$.
- $u_{ij(h)}$ son variables aleatorias independientes con distribución $N(0, \sigma)$

La notación $y_{ij(h)}$ indica que los niveles i y j determinan la letra latina para un cuadrado latino especificado. Así, en la situación de referencia, Tabla 5-2, si $i = 2$ y $j = 1$, automáticamente $h = 2$ (semilla 2). Es decir, el subíndice h toma valores que dependen de la celdilla (i, j). De esta forma, utilizando la notación $y_{ij(h)}$, el cuadrado latino correspondiente a la situación de referencia es

$$\begin{array}{cccc} y_{11(4)} & y_{12(3)} & y_{13(1)} & y_{14(2)} \\ y_{21(2)} & y_{22(1)} & y_{23(3)} & y_{24(4)} \\ y_{31(3)} & y_{32(4)} & y_{33(2)} & y_{34(1)} \\ y_{41(1)} & y_{42(2)} & y_{43(4)} & y_{44(3)} \end{array}$$

A continuación se muestra la notación que se va a utilizar en estos diseños

- $N = K^2$ es el número total de observaciones.
- El total y la media global

$$y_{...} = \sum_{i,j} y_{ij(.)} \qquad \bar{y}_{...} = \frac{y_{...}}{K^2} \tag{7.2}$$

- El total y la media por fila

$$y_{i..} = \sum_{j=1}^{K} y_{ij(.)} \qquad \bar{y}_{i..} = \frac{y_{i..}}{K} \tag{7.3}$$

- El total y la media por columna

$$y_{.j.} = \sum_{i=1}^{K} y_{ij(.)} \qquad \bar{y}_{.j.} = \frac{y_{.j.}}{K} \tag{7.4}$$

  donde la notación $y_{ij(.)}$ indica que consideramos la observación correspondiente a la celdilla (i, j) independientemente de la letra latina que le corresponda.

- El total y la media para cada letra latina.

$$y_{..h} = \sum_{i,j} y_{ij(h)} \qquad \bar{y}_{..h} = \frac{y_{..h}}{K} \tag{7.5}$$

donde $y_{..h}$ se obtiene sumando las K observaciones en las que el tratamiento se ha fijado al nivel h.

7.2.3. Estimación de los parámetros del modelo

2.4.1 Estimación por máxima verosimilitud

Se construye la función de verosimilitud asociada a la muestra para las $N = K^2$ observaciones

$$\mathbb{L}(\mu, \tau_i, \beta_j, \gamma_h, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left[y_{ij(h)} - \mu - \tau_i - \beta_j - \gamma_h\right]^2\right) \quad , \tag{7.6}$$

se determina el logaritmo de dicha función

$$\ln(\mathbb{L}(\mu, \tau_i, \beta_j, \gamma_h, \sigma^2)) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left[y_{ij(h)} - \mu - \tau_i - \beta_j - \gamma_h\right]^2 \quad , \tag{7.7}$$

se calculan las correspondientes derivadas parciales

$$\frac{\partial \ln \mathbb{L}}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left[y_{ij(.)} - \mu - \tau_i - \beta_j - \gamma_h\right] \tag{7.8}$$

$$\frac{\partial \ln \mathbb{L}}{\partial \tau_i} = \frac{1}{\sigma^2}\sum_{j=1}^{K}\left[y_{ij(.)} - \mu - \tau_i - \beta_j - \gamma_h\right] \qquad i = 1, \cdots, K \tag{7.9}$$

$$\frac{\partial \ln \mathbb{L}}{\partial \beta_j} = \frac{1}{\sigma^2}\sum_{i=1}^{K}\left[y_{ij(.)} - \mu - \tau_i - \beta_j - \gamma_h\right] \qquad j = 1, \cdots, K \tag{7.10}$$

$$\frac{\partial \ln \mathbb{L}}{\partial \gamma_h} = \frac{1}{\sigma^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left[y_{ij(h)} - \mu - \tau_i - \beta_j - \gamma_h\right] \qquad h = 1, \cdots, K \tag{7.11}$$

$$\frac{\partial \ln \mathbb{L}}{\partial \sigma^2} = \quad -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \left[y_{ij(.)} - \mu - \tau_i - \beta_j - \gamma_h \right]^2 \quad .$$

Igualando a cero estas derivadas parciales, se obtiene un sistema de ecuaciones que proporciona los estimadores máximo verosímiles. Dichos estimadores vienen dados por las expresiones

$$\widehat{\mu} = \frac{\sum_i \sum_j y_{ij(.)}}{K^2} = \bar{y}_{...} \quad , \tag{7.12}$$

$$\widehat{\tau}_i = \frac{1}{K} \sum_{j=1}^{K} y_{ij(.)} - \widehat{\mu} = \bar{y}_{i..} - \bar{y}_{...} \quad , \tag{7.13}$$

$$\widehat{\beta}_j = \frac{1}{K} \sum_{i=1}^{K} y_{ij(.)} - \widehat{\mu} = \bar{y}_{.j.} - \bar{y}_{...} \quad , \tag{7.14}$$

$$\widehat{\gamma}_h = \frac{1}{K} \sum_{i,j} y_{ij(h)} - \widehat{\mu} = \bar{y}_{..h} - \bar{y}_{...} \quad . \tag{7.15}$$

Se puede comprobar fácilmente que estos estimadores verifican

$$\sum_i \widehat{\tau}_i = \sum_j \widehat{\beta}_j = \sum_h \widehat{\gamma}_h = 0 \quad .$$

Finalmente, sustituyendo $\widehat{\mu}$, $\widehat{\tau}_i$, $\widehat{\beta}_j$ y $\widehat{\gamma}_h$ en la última ecuación de (7.11) e igualando a cero, obtenemos el estimador de máxima verosimilitud para la varianza

$$\widehat{\sigma}^2 = \frac{1}{N} \sum_{i,j} \left[y_{ij(.)} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j - \widehat{\gamma}_h \right]^2 \quad . \tag{7.16}$$

Residuos

Los residuos en este modelo adoptan la expresión

$$\begin{aligned} e_{ij(h)} = \quad & y_{ij(h)} - \widehat{y}_{ij(h)} = y_{ij(h)} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j - \widehat{\gamma}_h = \\ & y_{ij(h)} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..h} + 2\bar{y}_{..} \quad . \end{aligned} \tag{7.17}$$

Se verifica que los residuos suman cero por filas, columnas y para cada letra

$$\sum_j e_{ij(.)} = 0 \qquad i = 1, \cdots, K$$

$$\sum_i e_{ij(.)} = 0 \qquad j = 1, \cdots, K$$

$$\sum_i \sum_j e_{ij(h)} = 0 \qquad h = 1, \cdots, K$$

Por tanto el número de grados de libertad de los residuos es $(K-1)(K-2)$. En efecto

$$K^2 - (K + (K-1) + (K-1)) = K^2 - 3K + 2 = (K-1)(K-2) \quad .$$

Se comprueba que el estimador insesgado de la varianza poblacional es la suma de los cuadrados de los residuos dividida por sus grados de libertad, es decir

$$\widetilde{\sigma^2} = \widehat{S}_R^2 = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K}\left[y_{ij(h)} - \widehat{y}_{ij(h)}\right]^2}{(K-1)(K-2)} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K} e_{ij(h)}^2}{(K-1)(K-2)} \quad . \tag{7.18}$$

7.2.4. Descomposición de la variabilidad

Siguiendo el mismo procedimiento que en modelo en bloques aleatorizados, se comprueba que la ecuación básica del análisis de la varianza es

$$\begin{aligned}\sum_{i=1}^{K}\sum_{j=1}^{K}(y_{ij(h)} - \bar{y}_{...})^2 = \;& K\sum_{i=1}^{K}(\bar{y}_{i..} - \bar{y}_{...})^2 + K\sum_{j=1}^{K}(\bar{y}_{.j.} - \bar{y}_{...})^2 + K\sum_{h=1}^{K}(\bar{y}_{..h} - \bar{y}_{...})^2 + \\ & \sum_{i=1}^{K}\sum_{j=1}^{K}(y_{ij(h)} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..h} + 2\bar{y}_{...})^2\end{aligned} \tag{7.19}$$

que simbólicamente se puede escribir

$$SCT = SCF + SCC + SCL + SCR \quad ,$$

denominando por esas siglas los términos en el orden en que figuran en la ecuación (7.19) y que reciben los siguientes nombres

1) SCT, *suma total de cuadrados.*

2) SCF, *suma de cuadrados debida al efecto fila.*

3) SCC, *suma de cuadrados debida al efecto columna.*

4) SCL, *suma de cuadrados debida a la letra latina.*

5) SCR, *suma de cuadrados del error.*

Basándonos en estas sumas de cuadrados se construyen los correspondientes cuadrados medios que denotamos por $\widehat{S}_T^2$, $\widehat{S}_F^2$, $\widehat{S}_C^2$, $\widehat{S}_L^2$ y $\widehat{S}_R^2$ o bien por CMT, CMF, CMC, CML y CMR o CME.

Siguiendo el mismo razonamiento que en la subsección **??** del Capítulo 1, se demuestra que los valores esperados de los cuadrados medios correspondientes a las filas, columnas, letras latinas y residual son, respectivamente:

$$
\begin{aligned}
E(CMF) = \quad & E(\widehat{S}_F^2) = \quad \sigma^2 + \frac{K\sum_{i=1}^{K}\tau_i^2}{K-1} \\
E(CMC) = \quad & E(\widehat{S}_C^2) = \quad \sigma^2 + \frac{K\sum_{j=1}^{K}\beta_j^2}{K-1} \\
E(CML) = \quad & E(\widehat{S}_L^2) = \quad \sigma^2 + \frac{K\sum_{h=1}^{K}\gamma_h^2}{K-1} \\
E(CMR) = \quad & E(\widehat{S}_R^2) = \quad \sigma^2
\end{aligned} \tag{7.20}
$$

Por lo tanto,

a) Se verifica que $\widehat{S}_R^2 = SCR/(K-1)(K-2)$ es un estimador insesgado de la varianza poblacional σ^2.

b) Puede observarse que bajo el supuesto de que cada una de las hipótesis

$$
\begin{aligned}
H_{0\tau}: &\quad \tau_i = 0, \quad \forall i \\
H_{0\beta}: &\quad \beta_j = 0, \quad \forall j \\
H_{0\gamma}: &\quad \gamma_h = 0, \quad \forall h
\end{aligned} \tag{7.21}
$$

sea cierta, el correspondiente sumando de $E(CMF)$ es nulo y por tanto $\widehat{S}_F^2$, $\widehat{S}_C^2$ y $\widehat{S}_L^2$ son estimadores insesgados de σ^2.

Sin embargo, hay que notar que si existen diferencias entre las medias de las filas, las columnas o letras latinas, el respectivo valor esperado del cuadrado medio es mayor que σ^2.

De todo esto podemos deducir que:

1^o) Un contraste para verificar la hipótesis nula de igualdad de medias de las filas puede efectuarse comparando $\widehat{S}_F^2$ y $\widehat{S}_R^2$.

2^o) Un contraste para verificar la hipótesis nula de igualdad de medias de las columnas puede efectuarse comparando $\widehat{S}_C^2$ y $\widehat{S}_R^2$.

3^o) Un contraste para verificar la hipótesis nula de igualdad de medias de las letras latinas puede efectuarse comparando $\widehat{S}_L^2$ y $\widehat{S}_R^2$.

Siguiendo el mismo razonamiento que en los modelos anteriores se puede demostrar que bajo las hipótesis nulas (7.21), se verifica que

$$\frac{SCF}{\sigma^2} \rightsquigarrow \chi^2_{K-1}$$

$$\frac{SCC}{\sigma^2} \rightsquigarrow \chi^2_{K-1}$$

$$\frac{SCL}{\sigma^2} \rightsquigarrow \chi^2_{K-1}$$

$$\frac{SCR}{\sigma^2} \rightsquigarrow \chi^2_{(K-1)(K-2)}$$

y estas distribuciones son independientes entre sí.

Notamos que $\dfrac{SCR}{\sigma^2} \rightsquigarrow \chi^2_{(K-1)(K-2)}$ sean o no ciertas las hipótesis nulas dadas en (7.21).

Los estadísticos de contrastes:

$$F_\tau = \frac{\dfrac{SCF/\sigma^2}{K-1}}{\dfrac{SCR/\sigma^2}{(K-1)(K-2)}} = \frac{\widehat{S}_F^2}{\widehat{S}_R^2} \tag{7.22}$$

$$F_\beta = \frac{\dfrac{SCC/\sigma^2}{K-1}}{\dfrac{SCR/\sigma^2}{(K-1)(K-2)}} = \frac{\widehat{S}_C^2}{\widehat{S}_R^2} \tag{7.23}$$

y

$$F_\gamma = \frac{\dfrac{SCL/\sigma^2}{K-1}}{\dfrac{SCR/\sigma^2}{(K-1)(K-2)}} = \frac{\widehat{S}_L^2}{\widehat{S}_R^2} \quad , \tag{7.24}$$

siguen distribuciones F de Snedecor con $K-1$ y $(K-1)(K-2)$ grados de libertad cada uno. Y se rechazá H_0, al nivel de significación α, cuando el valor experimental del respectivo estadístico sea mayor que el valor crítico de la distribución F con $K-1$ y $(K-1)(K-2)$ grados de libertad.

Para una mayor sencillez en el cálculo se utilizan las expresiones abreviadas de SCT, SCF, SCC, SCL, y SCR, que se dan a continuación

$$SCT = \sum_{i=1}^{K}\sum_{j=1}^{K} y_{ij(.)}^2 - \frac{y_{...}^2}{K^2} \tag{7.25}$$

$$SCF = \frac{1}{K}\sum_{i=1}^{K} y_{i..}^2 - \frac{y_{...}^2}{K^2} \tag{7.26}$$

$$SCC = \frac{1}{K}\sum_{j=1}^{K} y_{.j.}^2 - \frac{y_{...}^2}{K^2} \tag{7.27}$$

$$SCL = \frac{1}{K}\sum_{h=1}^{K} y_{..h}^2 - \frac{y_{...}^2}{K^2} \quad . \tag{7.28}$$

La suma de cuadrados del error se obtiene por diferencia

$$SCR = SCT - SCF - SCC - SCL \quad . \tag{7.29}$$

La tabla ANOVA correspondiente a este modelo es:

Tabla 5-3. Tabla ANOVA para el modelo de cuadrado latino

Fuentes de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F_{exp}
E. fila	$K\sum_{i}^{K}(\bar{y}_{i..}-\bar{y}_{...})^2$	$K-1$	$\widehat{S}_F^2$	$\widehat{S}_F^2/\widehat{S}_R^2$
E. columna	$K\sum_{j}^{K}(\bar{y}_{.j.}-\bar{y}_{...})^2$	$K-1$	$\widehat{S}_C^2$	$\widehat{S}_C^2/\widehat{S}_R^2$
E. letra lat.	$K\sum_{h=1}^{K}(\bar{y}_{..h}-\bar{y}_{...})^2$	$K-1$	$\widehat{S}_L^2$	$\widehat{S}_L^2/\widehat{S}_R^2$
Residual	$SCT-SCF$ $SCC-SCL$	$(K-1)(K-2)$	$\widehat{S}_R^2$	
TOTAL	$\sum_{i}^{K}\sum_{j}^{K}\left(y_{ij(.)}-\bar{y}_{...}\right)^2$	K^2-1	$\widehat{S}_T^2$	

Alternativamente, utilizando las expresiones abreviadas de SCT, SCF, SCC, SCL y SCR, dadas en (7.25)-(7.29), la Tabla ANOVA se puede presentar como

Tabla 5-4. Forma práctica de la tabla ANOVA para el modelo de cuadrado latino

Fuentes de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F_{exp}
E. fila	$\frac{1}{K}\sum_{i}y_{i..}^2-\frac{y_{...}^2}{K^2}$	$K-1$	$\widehat{S}_F^2$	$\widehat{S}_F^2/\widehat{S}_R^2$
E. columna	$\frac{1}{K}\sum_{j}y_{.j.}^2-\frac{y_{...}^2}{K^2}$	$K-1$	$\widehat{S}_C^2$	$\widehat{S}_C^2/\widehat{S}_R^2$
E. letra lat.	$\frac{1}{K}\sum_{h}y_{..h}^2-\frac{y_{...}^2}{K^2}$	$K-1$	$\widehat{S}_L^2$	$\widehat{S}_L^2/\widehat{S}_R^2$
Residual	$SCT-SCF$ $SCC-SCL$	$(K-1)(K-2)$	$\widehat{S}_R^2$	
TOTAL	$\sum_{i}^{K}\sum_{j}^{K}y_{ij(.)}^2-\frac{y_{...}^2}{K^2}$	K^2-1	S_T^2	

Coeficiente de determinación

A continuación definimos el coeficiente de determinación R^2

$$R^2 = \frac{SCF + SCC + SCL}{SCT} \quad ,$$

donde, denotamos:

- R^2_τ al cociente entre la variabilidad explicada por las filas y la variabilidad total, es decir: $R^2_\tau = \dfrac{SCF}{SCT}$

- R^2_β al cociente entre la variabilidad explicada por las columnas y la variabilidad total, es decir: $R^2_\beta = \dfrac{SCC}{SCT}$

- R^2_γ al cociente entre la variabilidad explicada por las letras latinas y la variabilidad total, es decir: $R^2_\gamma = \dfrac{SCL}{SCT}$.

  Entonces el coeficiente de determinación también lo podemos expresar como

$$R^2 = R^2_\tau + R^2_\beta + R^2_\gamma \quad ,$$

siendo R^2_τ, R^2_β y R^2_γ los *coeficientes de determinación parciales* asociados a las filas, columnas y letras latinas, respectivamente.

7.2.5. Ejemplo numérico

A fin de ilustrar el análisis de la varianza de los diseños en cuadrado latino, consideremos la situación de referencia, en la que se ha realizado el experimento con la aleatorización correspondiente y hemos designado por las letras (A, B, C, D) a los tratamientos. Así, el cuadrado latino resultante junto con las observaciones obtenidas, dan lugar al Ejemplo 5-1, que se muestra en la siguiente tabla, a la que se han añadido las filas y columnas necesarias para su resolución.

Tabla 5-5 Datos para el ejemplo 5-1

Abonos	Insecticidas i1	i2	i3	i4	$y_{i..}$	$y_{i..}^2$
a1	C 7	D 8	B 4	A 3	22	484
a2	B 15	A 16	C 18	D 23	72	5184
a3	D 18	C 12	A 12	B 10	52	2704
a4	A 14	B 13	D 16	C 14	57	3249
$y_{.j.}$	54	49	50	50	203	11621
$y_{.j.}^2$	2916	2401	2500	2500	10317	
$\sum y_{ij(.)}^2$	794	633	740	834	3001	

Tabla 5-6

Letra latina	Observaciones				$y_{..h}$	$y_{..h}^2$
A	3	16	12	14	45	2025
B	4	15	10	13	42	1764
C	7	18	12	14	51	2601
D	8	23	18	16	65	4225
					203	10615

Las sumas de cuadrados necesarias para el análisis de la varianza se calculan como sigue:

$$SCT = \sum_{i=1}^{4}\sum_{j=1}^{4} y_{ij(.)}^2 - \frac{y_{...}^2}{K^2} = 3001 - \frac{203^2}{4^2} = 425{,}4375$$

$$SCF = \frac{1}{K}\sum_{i=1}^{4} y_{i..}^2 - \frac{y_{...}^2}{K^2} = \frac{11621}{4} - \frac{203^2}{4^2} = 329{,}6875$$

$$SCC = \frac{1}{K}\sum_{j=1}^{4} y_{.j.}^2 - \frac{y_{...}^2}{K^2} = \frac{10317}{4} - \frac{203^2}{4^2} = 3{,}6875$$

$$SCL = \frac{1}{K}\sum_{h=1}^{4} y^2_{..h} - \frac{y^2_{...}}{K^2} = \frac{10615}{4} - \frac{203^2}{4^2} = 78{,}1875$$

y la suma de cuadrados del error se obtiene por diferencia

$$SCR = SCT - SCF - SCC - SCL = 13{,}875 \quad .$$

La tabla ANOVA correspondiente a este modelo es

Tabla 5-7. Análisis de la varianza para los datos del Ejemplo 5-1

Fuentes de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F_{exp}	% Explicado
E. fila (Abonos)	329.6875	3	109.89583	47.523	77.49 %
E. colum. (Insect.)	3.6875	3	1.22917	0.532	0.87 %
E. trat. (Semillas)	78.1875	3	26.06250	11.270	18.38 %
Residual	13.8750	6	2.31250		
TOTAL	425.4375	15			96.74 %

Si realizamos el contraste al 5 % y comparamos los valores de las F_{exp} con el valor de la F teórica ($F_{0,05;3,6} = 4{,}7571$), se concluye que son significativos los efectos de los abonos y semillas, pero no lo son los efectos de los insecticidas.

Observamos, en la columna correspondiente al % explicado, que el coeficiente de determinación del modelo es $R^2 = 0{,}9674$, siendo el efecto más importante el referente al tipo de abono que explica un 77.49 % de la variabilidad presente en el experimento.

Bibliografía utilizada

∗ **García Leal, J. & Lara Porras, A.M.** (1998). *"Diseño Estadístico de Experimentos. Análisis de la Varianza."* Grupo Editorial Universitario.

∗ **Lara Porras, A.M.** (2000). *"Diseño Estadístico de Experimentos, Análisis de la Varianza y Temas Relacionados: Tratamiento Informático mediante SPSS"* Proyecto Sur de Ediciones.