

Diseño de Bloques al azar

Introducción

En cualquier experimento, la variabilidad proveniente de un factor de ruido puede afectar los resultados.

Un factor de ruido es un factor que probablemente tiene un efecto en la respuesta pero que no nos interesa estudiar.

Si el factor de ruido es **desconocido y no controlable**, la solución es la aleatorización, que tiende a distribuir los niveles y efectos de este factor entre todas las u.e.

Si el factor de ruido es **conocido y no controlable**, pero por lo menos podemos medir su valor en cada corrida del experimento, entonces podemos compensarlo usando análisis de covarianza.

Si el factor de ruido es **conocido y controlable**, se utilizan bloques para eliminar su efecto en la comparación estadística de los tratamientos.

Bloques al azar

Nuestro objetivo es tener comparaciones precisas entre los tratamientos bajo estudio. Utilizar bloques es una forma de reducir y controlar la varianza del error experimental para tener mayor precisión.

En el diseño completamente al azar se supone que las u.e. son relativamente homogéneas con respecto a factores que afectan la variable de respuesta. Sin embargo, algunas veces no tenemos disponibles suficiente número de u.e. homogéneas.

Cualquier factor que afecte la variable de respuesta y que varíe entre u.e. aumentará la varianza del error experimental y disminuirá la precisión de las comparaciones.

Factores como la edad y el peso de los animales, diferentes lotes de material, sexo de las personas y parcelas alejadas son ejemplos de variables externas a los tratamientos que pueden incrementar la variación entre las observaciones de la variable de respuesta.

Bloques al azar

Usar bloques estratifica a las u.e. en grupos homogéneos. Una buena elección del criterio de bloqueo resulta en menor variación entre las u.e. dentro de los bloques comparada con la variación entre u.e. de diferentes bloques. Generalmente los criterios de bloqueo son:

- proximidad (parcelas vecinas)
- características físicas (edad, peso, sexo)
- tiempo
- manejo de las u.e. en el experimento

Bloques al azar

Suponga que se tienen t tratamientos que se quieren comparar en b bloques.

Bloque 1	Bloque 2	...	Bloque b
y_{11}	y_{12}	...	y_{1b}
y_{21}	y_{22}	...	y_{2b}
...
...
...
y_{t1}	y_{t2}	...	y_{tb}

El diseño de bloques (completos) al azar implica que en cada bloque hay una sola observación de cada tratamiento. El orden en que se “corren” los tratamientos dentro de cada bloque es aleatorio (restricción en la aleatorización).

Bloques al azar

El modelo estadístico para este diseño es:

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad i = 1, \dots, t \quad j = 1, \dots, b$$

μ media general

τ_i efecto del i-ésimo tratamiento

β_j efecto del j-ésimo bloque

ϵ_{ij} error experimental en la unidad j del tratamiento i

$\epsilon_{ij} \sim NID(0, \sigma^2)$.

Se supone que los efectos de tratamientos y bloques son aditivos. La aditividad significa que no hay interacción entre tratamientos y bloques. Es decir, la relación entre los tratamientos es la misma en cada uno de los bloques.

Bloques al azar

tratamiento	bloque				medias de tratamientos
	1	2	...	b	
1	y_{11}	y_{12}	...	y_{1b}	$\bar{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2b}	$\bar{y}_{2.}$
..
..
t	y_{t1}	y_{t2}	...	y_{tb}	$\bar{y}_{t.}$
medias de bloque	$\bar{y}_{.1}$	$\bar{y}_{.2}$...	$\bar{y}_{.b}$	$\bar{y}_{..}$

Bloques al azar

El Análisis de Varianza para este diseño se basa en una descomposición de la variabilidad de las observaciones.

$$y_{ij} - \bar{y}_{..} = (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})$$

desviación		desviación		desviación		desviación
total	=	debida	+	debida	+	debida
		al error		a tratamientos		a bloques

$$\sum_{i=1}^t \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = b \sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2 + t \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^t \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

$$SS_{total} = SS_{trat} + SS_{bloque} + SS_{error}$$

$$bt - 1 = (t - 1) + (b - 1) + (b - 1)(t - 1)$$

Bloques al azar

Suponiendo normalidad en los errores, se puede demostrar que

$$\frac{SS_{trat}}{\sigma^2}, \frac{SS_{bloques}}{\sigma^2}, \frac{SS_E}{\sigma^2}$$

son v.a. independientes con distribución χ^2 con sus correspondientes grados de libertad.

Tabla de Análisis de Varianza

F.V.	g.l.	SS	CM	F	$E(CM)$
trat's	$t - 1$	$b \sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2$	$SS_{trat}/(t - 1)$	$\frac{CM_{trat}}{CM_E}$	$\sigma^2 + b \sum_{i=1}^t \frac{\tau_i^2}{t-1}$
bloques	$b - 1$	$t \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2$	$SS_{bloques}/(b - 1)$		
error	$(t - 1)(b - 1)$	$\sum \sum (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$SS_E/(t - 1)(b - 1)$		σ^2
total	$bt - 1$	$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$			

Si $F_c > F_{t-1, (t-1)(b-1)}^\alpha$ se rechaza $H_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0$

Bloques al azar

No se deben probar bloques.

La aleatorización se aplicó solamente a tratamientos dentro de bloques, esto es, los bloques representan una **restricción** a la aleatorización.

Si se tiene un grupo de u.e. donde se supone que ocurre un efecto β en **todas** las u.e. simultáneamente, esto no equivale a la ocurrencia independiente de β en cada u.e.

$$\tau_2\beta_1 \quad \tau_3\beta_1 \quad \tau_1\beta_1$$

$$\tau_3\beta_2 \quad \tau_1\beta_2 \quad \tau_2\beta_2$$

$$\tau_1\beta_3 \quad \tau_2\beta_3 \quad \tau_3\beta_3$$

$$\tau_3\beta_4 \quad \tau_2\beta_4 \quad \tau_1\beta_4$$

$$\tau_2\beta_2 \quad \tau_1\beta_1 \quad \tau_3\beta_2 \quad \tau_1\beta_4$$

$$\tau_3\beta_4 \quad \tau_2\beta_3 \quad \tau_1\beta_2 \quad \tau_3\beta_1$$

$$\tau_1\beta_3 \quad \tau_2\beta_4 \quad \tau_3\beta_3 \quad \tau_2\beta_1$$

Bloques al azar

Los residuales en este caso son:

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

$$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$$

$$e_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$$

Los contrastes y comparaciones múltiples se hacen igual que antes, considerando este nuevo CME .

Diseños con bloques

Bloques al azar: bloques de tamaño t (número de tratamientos)

Bloques al azar generalizados: se repiten los tratamientos en cada bloque

Bloques incompletos: los bloques no contienen a todos los tratamientos

Ejemplo con bloques

Se realizó un experimento para estudiar el funcionamiento de cuatro diferentes detergentes quita manchas. Las lecturas de “blancura” (valor mayor=más limpio) se obtuvieron usando un aparato especial en tres diferentes tipos de manchas comunes. Hay diferencia significativa entre los detergentes?

	Mancha 1	Mancha 2	Mancha 3
Detergente 1	45	43	51
Detergente 2	47	46	52
Detergente 3	48	50	55
Detergente 4	42	37	49

Diseño en Cuadro Latino

El diseño de Cuadro Latino se usa para eliminar dos fuentes de variabilidad que no interesa estudiar por si mismas. Se hace un bloqueo en dos direcciones. Los renglones y las columnas representan dos restricciones en la aleatorización.

En general, un cuadro latino $p \times p$ es un cuadrado que contiene p renglones y p columnas. Cada una de las p^2 celdas contiene una de las p letras que corresponden a los tratamientos, y cada letra ocurre una sola vez en cada renglón y columna.

Ejemplo:

4×4

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

5×5

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

Diseño en cuadro latino

El modelo es:

$$y_{ij} = \mu + \rho_i + \gamma_j + \tau_k + \epsilon_{ij}$$

$$i = 1, \dots, t \quad j = 1, \dots, t \quad k = 1, \dots, t$$

y_{ij} medición de la observación del renglón i -ésimo
de la columna j -ésima

ρ_i efecto del renglón i

γ_j efecto de la columna j

τ_k efecto del tratamiento k

ϵ_{ij} error

El número de tratamientos debe ser igual de número de renglones y de columnas.

Diseño en cuadro latino

La partición de suma de cuadrados se deriva de la identidad algebraica:

$$\begin{aligned} y_{ij} - \bar{y}_{..} &= (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (\bar{y}_k - \bar{y}_{..}) \\ &+ (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} - \bar{y}_k + 2\bar{y}_{..}) \end{aligned}$$

La desviación de una observación de la media global ($y_{ij} - \bar{y}_{..}$) se expresa como la suma de:

- una desviación renglón ($\bar{y}_{i.} - \bar{y}_{..}$)
- una desviación columna ($\bar{y}_{.j} - \bar{y}_{..}$)
- una desviación de tratamientos ($\bar{y}_k - \bar{y}_{..}$)
- error experimental ($y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} - \bar{y}_k + 2\bar{y}_{..}$)

Tabla de Análisis de Varianza

F.V.	g.l.	SS	CM	F	$E(CM)$
trat's	$t - 1$	$t \sum_{k=1}^t (\bar{y}_k - \bar{y}_{..})^2$	$SS_{trat}/(t - 1)$	$\frac{CM_{trat}}{CM_E}$	$\sigma^2 + t\theta_t^2$
renglones	$t - 1$	$t \sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2$	$SS_{reng}/(t - 1)$		
columnas	$t - 1$	$t \sum_{j=1}^t (\bar{y}_{.j} - \bar{y}_{..})^2$	$SS_{col}/(t - 1)$		
error	$(t - 1)(t - 2)$	SSE	$SS_E/(t - 1)(t - 2)$		σ^2
total	$t^2 - 1$	$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$			

Si $F_c > F_{t-1, (t-1)(t-2)}^\alpha$ se rechaza $H_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0$

Ejemplo de cuadro latino

Un ingeniero industrial está investigando el efecto de cuatro métodos de ensamblado (A,B,C,D) en el tiempo de ensamblado de una componente de televisión a color.

Seleccionó a cuatro operadores para el estudio. Además, el ingeniero sabe que cada método de ensamblado produce cierta fatiga, de tal manera que el tiempo requerido para el último ensamblado puede ser mayor que el tiempo requerido para el primero, independientemente del método.

Para tomar en cuenta estas dos fuentes de variabilidad (operador, orden de ensamblado), el ingeniero decide usar un diseño en Cuadro Latino, cuyos resultados se presentan a continuación.

Ejemplo de cuadro latino

orden de ensamblado	Operador			
	1	2	3	4
1	C=10	D=14	A=7	B=8
2	B=7	C=18	D=11	A=8
3	A=5	B=10	C=11	D=9
4	D=10	A=10	B=12	C=14

Diseño en cuadro latino

La generación de un arreglo aleatorio en, por ejemplo, un cuadro 5×5 :

Se constuye un cuadro básico:

(1)	A	B	C	D	E
(2)	B	C	D	E	A
(3)	C	D	E	A	B
(4)	D	E	A	B	C
(5)	E	A	B	C	D

Se aleatoriza el orden de los renglones y después el de las columnas. Y se aleatorizan los tratamientos a las letras.

Diseño en cuadro latino

	(1)	(2)	(3)	(4)	(5)
(5)	E	A	B	C	D
(3)	C	D	E	A	B
(1)	A	B	C	D	E
(2)	B	C	D	E	A
(4)	D	E	A	B	C

	(3)	(2)	(4)	(1)	(5)
(5)	B	A	C	E	D
(3)	E	D	A	C	B
(1)	C	B	D	A	E
(2)	D	C	E	B	A
(4)	A	E	B	D	C

Residuales

El residual para la observación en el tratamiento k en el renglón i y columna j es

$$e_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} - \bar{y}_k + 2\bar{y}_{..}$$

Diseño Grecolatino

Una compañía que produce procesadores de alimentos desea determinar el efecto del diseño de empaque en las ventas de uno de sus productos. Se tienen 5 diseños para probar: A,B,C,D y E. Hay algunas fuentes de variación:

1. día de la semana
2. diferencia entre tiendas
3. efecto de la altura del entrepaño

Se realiza un experimento en diseño cuadro grecolatino con los cinco días de la semana como clasificación renglón, cinco diferentes tiendas como clasificación columna y cinco alturas del entrepaño como clasificación de letra griega.

La siguiente tabla contiene los resultados del experimento. La medición es el número de ventas en el día.

Diseño Grecolatino

Día	Tienda				
	1	2	3	4	5
Lunes	$E\alpha(238)$	$C\delta(228)$	$B\gamma(158)$	$D\epsilon(188)$	$A\beta(74)$
Martes	$D\delta(149)$	$B\beta(220)$	$A\alpha(92)$	$C\gamma(169)$	$E\epsilon(282)$
Miérc	$B\epsilon(222)$	$E\gamma(295)$	$D\beta(104)$	$A\delta(54)$	$C\alpha(213)$
Jueves	$C\beta(187)$	$A\epsilon(66)$	$E\delta(242)$	$B\alpha(122)$	$D\gamma(90)$
Viernes	$A\gamma(65)$	$D\alpha(118)$	$C\epsilon(279)$	$E\beta(278)$	$B\delta(176)$