Diseño Cuadrado latino

Universidad Autonoma de Guerrero

Diseño y análisis de experimentos

27 de noviembre de 2023

contenido

Ejemplo 1

Suponga que un experimentador estudia los efectos que tienen cinco formulaciones diferentes de la carga propulsora utilizada en los sistemas de expulsión de la tripulación de un avión basado en la rapidez de combustión. Cada formulación se hace con un lote de materia prima que sólo alcanza para probar cinco formulaciones. Además, las formulaciones son preparadas por varios operadores, y puede haber diferencias sustanciales en las habilidades y experiencia de los operadores.

Ejemplo 1

Cuadro: Diseño del cuadrado latino para el problema de la carga propulsora

Lotes de materia prima	Operadores				
	1	2	3	4	5
1	A=24	B=20	C=19	D=24	E=24
2	B=17	C=24	D=30	E=27	A=36
3	C=18	D=38	E=26	A=27	B=21
4	D=26	E=31	A=26	B=23	C=22
5	E=22	A=30	B=20	C=29	D=31

Ejemplo 2

Supóngase que un ingeniero desea realizar 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 cuadrado latino, donde el factor principal es el tipo de semilla y las variables de ruido son los tipos de abono e insecticida

Ejemplo 2

Cuadro: Diseño del cuadrado latino para el problema de la la producción de trigo

Abonos	Insecticidas			
	1	2	3	4
1	C=24	D=20	B=19	A=24
2	B=17	A=24	C=30	D=27
3	D=18	C=38	A=26	B=27
4	A=26	B=31	D=26	C=23

¿Cuándo es apropiado usar un DCL?

El diseño de cuadrado latino se usa para eliminar dos fuentes de variabilidad perturbadora; es decir, permite hacer la formación de bloques sistemáticamente en dos direcciones.

¿Qué es un gradiente de variación?

Un gradiente de variacion puede definirse como un factor del diseño que probablemente tenga un efecto sobre la respuesta, pero en el que no existe un interés específico.

¿Qué se puede hacer para su estudio ?

La aleatorización es la técnica de diseño que se utiliza para protegerse contra estos factores perturbadores

¿Cuales son los supuestos de un modelo de DCL?

- Normalidad
- Independencia
- Homogeneidad de varianza

Un cuadrado latino para p factores, o cuadrado latino de pxp, es un cuadrado que contiene p renglones y p columnas. Cada uno de las p^2 celdas contiene una de las p letras que corresponden a los tratamientos y cada letra ocurre únicamente una vez en cada renglón y columna del diseño.

A	B	C	D	A	D	B	E	C	A	D	C	E	B	F
B	C	D	A	D	A	C	B	E	B	A	E	C	F	D
C	D	A	B	C	B	E	D	A	C	E	D	F	A	B
D	A	B	C	B	E	A	C	D	D	C	F	B	E	A
				E	C	D	A	B	F	B	A	D	C	E
									E	F	B	A	D	C
		3x3				5x5						6x6		

SUDOKU

1		7			6	4	5	
	2	5	3	4				8
	6				1		7	
	5	3					2	9
6	1				9	8		
			6		2			7
		1		9	3	2		
		8						
	4			7	8	5	9	1

ANSWER

1	3	7	9	8	6	4	5	2
9	2	5	3	4	7	1	6	8
8	6	4	5	2	1	9	7	3
7	5	3	8	1	4	6	2	9
6	1	2	7	3	9	8	4	5
4	8	9	6	5	2	3	1	7
5	7	1	4	9	3	2	8	6
2	9	8	1	6	5	7	3	4
3	4	6	2	7	8	5	9	1

Figura: Juego Sudoku

Características del DCL

1. Se utiliza cuando la variabilidad del material experimental ocurre en dos direcciones simultáneamente, es decir, cuando se presentan dos posibles fuentes de variabilidad.
2. Se construye distribuyendo los tratamientos en un arreglo de filas y columnas.
3. Tiene el mismo número de filas y columnas.
4. Cada fila o columna constituye una repetición completa de los tratamientos.
5. Cuando el número de tratamientos a probar es grande, puede volverse poco práctico el DCL.

Modelo de efectos

$$\begin{aligned} Y_{ijk} &= \mu + \tau_i + \alpha_j + \beta_k + \varepsilon_{ijk} \quad \varepsilon_{ijk} \sim IIDN(0, \sigma^2) \\ i &= 1, 2, \ldots, p; \quad j = 1, 2, \ldots, p; \quad k = 1, 2, \ldots, p \end{aligned}$$

Parámetros: $\underbrace{\mu}_{1}, \underbrace{\tau_1, \tau_2, \ldots, \tau_p}_{p}, \underbrace{\alpha_1, \alpha_2, \ldots \alpha_p}_{p}, \underbrace{\beta_1, \beta_2, \ldots \beta_p}_{p}$ y $\underbrace{\sigma^2}_{1}$.

Y_{ijk}	VR del tratamiento i en el renglón j y columna k
μ	Media general
τ_i	Efecto del tratamiento i
α_j	Efecto del renglón j
β_k	Efecto de la columna k
ε_{ijk}	Error experimental del tratamiento i en el renglón j y columna k

Supuestos

Independencia	$\varepsilon_{111}, \ldots, \varepsilon_{ppp}$ son independientes
Distribución Normal	$\varepsilon_{ijk} \sim IIDN(0, \sigma^2)$
Homogeneidad de varianzas	$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_p^2$

Notación

Suma		Media	
General	$y_{\cdots} = \sum_{i=1}^{p}\sum_{k=1}^{p}\sum_{j=1}^{p} Y_{ijk}$	General	$\overline{y}_{\cdots} = \frac{y_{\cdots}}{p^2}$
del tratamiento i	$y_{i\cdot\cdot} = \sum_{j=1}^{p}\sum_{k=1}^{p} Y_{ijk}$	del tratamiento	$\overline{y}_{i\cdot\cdot} = \frac{y_{i\cdot\cdot}}{p}$
del renglón j	$y_{\cdot j\cdot} = \sum_{i=1}^{p}\sum_{k=1}^{p} Y_{ijk}$	del renglón j	$\overline{y}_{\cdot j\cdot} = \frac{y_{\cdot j\cdot}}{p}$
de la columna k	$y_{\cdot\cdot k} = \sum_{i=1}^{p}\sum_{j=1}^{p} Y_{ijk}$	de la columna k	$\overline{y}_{\cdot\cdot k} = \frac{y_{\cdot\cdot k}}{p}$

Los parámetros se estiman con el Método de Mínimos Cuadrados (Máxima Verosimilitud produce los mismos resultados).

$$\begin{aligned}
\varepsilon_{ijk} &= Y_{ijk} - \mu - \tau_i - \alpha_j - \beta_k \\
SCE &= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\varepsilon_{ijk}^2 = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\left\{Y_{ijk} - \mu - \tau_i - \alpha_j - \beta_k\right\}^2
\end{aligned}$$

Derivando SCE respecto a

$$\mu, \tau_1, \tau_2, ..., \tau_p, \alpha_1, \alpha_2, ..., \alpha_p, \beta_1, \beta_2, ..., \beta_p,$$

$$\underbrace{SCE}_{Trat} = \sum_{j=1}^{p}\sum_{k=1}^{p}\left\{Y_{1jk} - \mu - \tau_1 - \alpha_j - \beta_k\right\}^2 + \ldots + \sum_{j=1}^{p}\sum_{k=1}^{p}\left\{Y_{pjk} - \mu - \tau_p - \alpha_j - \beta_k\right\}^2$$

$$\underbrace{SCE}_{Reng} = \sum_{i=1}^{p}\sum_{k=1}^{p}\{Y_{i1k} - \mu - \tau_i - \alpha_1 - \beta_k\}^2 + \ldots + \sum_{i=1}^{p}\sum_{k=1}^{p}\left\{Y_{ipk} - \mu - \tau_i - \alpha_p - \beta_k\right\}^2$$

$$\underbrace{SCE}_{Col} = \sum_{i=1}^{p}\sum_{j=1}^{p}\left\{Y_{ij1} - \mu - \tau_i - \alpha_j - \beta_1\right\}^2 + \ldots + \sum_{i=1}^{p}\sum_{j=1}^{p}\left\{Y_{ijp} - \mu - \tau_i - \alpha_j - \beta_p\right\}^2$$

$$
\begin{aligned}
\frac{\partial SCE}{\partial \mu} &= \frac{\partial}{\partial \mu}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\{Y_{ij}-\mu-\tau_i-\alpha_j-\beta_k\}^2 \\
\frac{\partial SCE}{\partial \tau_i} &= \frac{\partial}{\partial \tau_i}\sum_{j=1}^{p}\sum_{k=1}^{p}\{Y_{ij}-\mu-\tau_i-\alpha_j-\beta_k\}^2 \\
\frac{\partial SCE}{\partial \alpha_j} &= \frac{\partial}{\partial \alpha_j}\sum_{i=1}^{p}\sum_{k=1}^{p}\{Y_{ij}-\mu-\tau_i-\alpha_j-\beta_k\}^2 \\
\frac{\partial SCE}{\partial \beta_k} &= \frac{\partial}{\partial \beta_k}\sum_{i=1}^{p}\sum_{j=1}^{p}\{Y_{ij}-\mu-\tau_i-\alpha_j-\beta_k\}^2
\end{aligned}
$$

Igualando a cero cada una de las $3p+1$ ecuaciones, agregando las restricciones de que las sumas de los efectos de los tratamientos (renglones, columnas) son cero y resolviendo el sistema de ecuaciones, se obtienen los siguientes estimadores

μ	Media general	$\widehat{\mu} = \overline{y}_{...}$
μ_i	Media del tratamiento i	$\widehat{\mu}_i = \overline{y}_{i..}$
τ_i	Efecto del tratamiento i	$\widehat{\tau}_i = \overline{y}_{i..} - \overline{y}_{...}$
α_j	Efecto del renglón j	$\widehat{\alpha}_j = \overline{y}_{.j.} - \overline{y}_{...}$
β_k	Efecto de la columna k	$\widehat{\beta}_k = \overline{y}_{..k} - \overline{y}_{...}$
σ^2	Varianza de los errores	$\widehat{\sigma}^2 = CME$
Y_{ijk}	Variable respuesta	$\widehat{Y}_{ijk} = \overline{y}_{i..} + \overline{y}_{.j.} + \overline{y}_{..k} - 2\overline{y}_{..k}$
ε_{ijk}	Error	$\widehat{\varepsilon}_{ijk} = Y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} - \overline{y}_{..k} + 2\overline{y}_{...}$

Contraste de hipótesis

La hipótesis de interés es

Las medias de la VR de cada nivel del factor son iguales

o equivalentemente

El efecto de cada nivel del factor en la VR es cero

Que se escriben

H_0:$\mu_1 = \mu_2 = \ldots = \mu_p$	equivalente	H_0: $\tau_1 = \tau_2 = \ldots = \tau_p = 0$
H_1: Al menos una μ_i es diferente	equivalente	H_1:Al menos una $\tau_i \neq 0$

Partición de las SC de las $N = p^2$ observaciones en SC para los tratamientos, renglones, columnas y error. Partición de la SC Total

$$
\begin{aligned}
SCT &= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(Y_{ijk} - \overline{y}_{...})^2 \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\left[(\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...}) + (\overline{y}_{..k} - \overline{y}_{...}) + (Y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} - \overline{y}_{..k} + 2\overline{y}_{...})\right]^2 \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(\overline{y}_{i..} - \overline{y}_{...})^2 + \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(\overline{y}_{.j.} - \overline{y}_{...})^2 \\
&+ \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(\overline{y}_{..k} - \overline{y}_{...})^2 + \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(Y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} - \overline{y}_{..k} + 2\overline{y}_{...})^2 \\
SCT &= SCTratamientos + SCRenglones + SCColumnas + SCError \\
p^2 - 1 &= (p-1) + (p-1) + (p-1) + (p-2)(p-1)
\end{aligned}
$$

$SCTr = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(\overline{y}_{i..} - \overline{y}_{...})^2$	$= \frac{1}{p}\sum_{i=1}^{p} y_{i..}^2 - \frac{Y_{...}^2}{p^2}$
$SCR = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(\overline{y}_{.j.} - \overline{y}_{...})^2$	$= \frac{1}{p}\sum_{j=1}^{p} y_{.j.}^2 - \frac{Y_{...}^2}{p^2}$
$SCC = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(\overline{y}_{..k} - \overline{y}_{...})^2$	$= \frac{1}{p}\sum_{k=1}^{p} y_{..k}^2 - \frac{Y_{...}^2}{p^2}$
$SCE = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} - \overline{y}_{..k} + 2\overline{y}_{...})^2$	Por diferencia
$SCT = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk} - \overline{y}_{...})^2$	$= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}^2 - \frac{Y_{...}^2}{p^2}$

Estadístico

- Si $\varepsilon_{ijk} \sim IIDN(0, \sigma^2)$ entonces,
  $\frac{SCTr}{\sigma^2} \sim \chi^2(p-1)$,
  $\frac{SCR}{\sigma^2} \sim \chi^2(p-1)$,
  $\frac{SCC}{\sigma^2} \sim \chi^2(p-1)$
  $\frac{SCE}{\sigma^2} \sim \chi^2(p-1)(p-2)$
  son independientes.
- Si $H_0 : \mu_1 = \mu_2 = \ldots, = \mu_p$ es cierta entonces

$$F = \frac{CMTrat}{CME} \sim F(p-1, (p-1)(p-2))$$

.

- Hipótesis

H_0:$\mu_1 = \mu_2 = \ldots = \mu_p$	equivalente	H_0: $\tau_1 = \tau_2 = \ldots = \tau_p = 0$
H_1: Al menos una μ_i es diferente	equivalente	H_1:Al menos una $\tau_i \neq 0$

- Estadístico de prueba: $F = \frac{CMTrat}{CME}$
  tiene distribución F con $p - 1$ y $(p - 1)(p - 2)$ gl.
- Regla de decisión: Rechazar H_0 si
  - $F_c > F_t$ donde $F_t = F_{1-\alpha}(p - 1, (p - 1)(p - 2))$
  - $p - value < \alpha$
- Conclusión: “(No) Existe evidencia que indica que las medias de los tratamientos sean iguales”

ANOVA

Fuentes de Variación	Grados de libertad	Suma de cuadrados	Cuadrado Medio	F	Pr> F
Tratamientos	$p-1$	SCTr	$CMTr = \frac{SCTr}{p-1}$	$\frac{CMTr}{CME}$	α^*
Renglón	$p-1$	SCR	$CMR = \frac{SCR}{p-1}$		
Columna	$p-1$	SCC	$CMC = \frac{SCC}{p-1}$		
Error	$(p-1)(p-2)$	SCE	$CME = \frac{SCE}{(p-1)(p-2)}$		
Total	p^2-1	SCT			

Verificación de los supuestos

- Normalidad
  - Gráfica QQ-Plot
  - Pruebas de normalidad
- Independencia
  - Gráfica de los residuos contra los valores predichos
- Varianzas homogéneas
  - Prueba de Bartlett
  - Prueba de Levene

- Si no se rechaza la hipótesis nula entonces todos los tratamientos producen el mismo efecto en la VR y el análisis estadístico termina.
- Si se rechaza la hipótesis nula entonces no todos los tratamientos producen el mismo efecto en la VR.
  Interesa conocer cual nivel o combinaciones de los niveles es el mejor.
- Existen dos casos importantes:
  - Todos los contrastes (combinaciones lineales): Prueba de Scheffé.
  - Comparaciones por pares: Tukey, DMS, Duncan, SNK, Dunnett.

Referencias

Montgomery C. D. (2010).
Diseño y analisis of experimentos.
Editorial Limusa. Mexico

Leal, J. G., & Porras, A. M. L. (1998).
Diseño estadístico de experimentos: análisis de la varianza.
Grupo Editorial Universitario.

Gracias