

Diseños y análisis de experimentos: Taller 3

Andrés Felipe Palomino - David Stiven Rojas - Mateo Trochez
Códigos:1922297-1924615-1931043

Universidad del Valle

26 de mayo de 2023



En un experimento agrícola se desea evaluar el rendimiento en t/ha de 4 cultivos, dado que se sospecha que existe un gradiente por la manera en que se aplica el riego, el experimento fue diseñado como un diseño en bloques completos al azar (3 bloques). Tenemos un total de 12 parcelas.

Diseño en bloques completos al azar (DBCA)

C4	C2	C4
C2	C3	C3
C1	C4	C1
C3	C1	C2

Cultivar	Bloque	Rendimiento
C1	B1	7.4
C1	B2	6.5
C1	B3	5.6
C2	B1	9.8
C2	B2	6.8
C2	B3	6.2
C3	B1	7.3
C3	B2	6.1
C3	B3	6.4
C4	B1	9.5
C4	B2	8.0
C4	B3	7.4

Ejercicio 1:

Objetivos del estudio:

- Evaluar el rendimiento en t/ha de 4 cultivos.

Hipótesis del estudio:

- H_0 : El rendimiento de los 4 cultivos es igual.
- H_1 : Por lo menos uno de los 4 cultivos presenta un rendimiento distinto.

Modelo estadístico DCBA:

$$y_{ij} = \mu + \tau_i + \alpha_j + \epsilon_{ij} \text{ con } i=1,2,3,4 \text{ y } j=1,2,3$$

y_{ij} : Rendimiento en t/ha del cultivar i y bloque j
 μ : Promedio global del rendimiento en t/ha
 τ_i : Efecto debido al i_esimo cultivar
 α_j : Efecto debido a j_esimo bloque
 ϵ_{ij} : Error aleatorio debido a factores no observados.

Factores:

- Tipo de cultivar

Niveles:

- Cultivar 1 (C1), Cultivar 2 (C2), Cultivar 3 (C3), Cultivar 4 (C4).

Tratamientos:

- Cultivar 1 (C1), Cultivar 2 (C2), Cultivar 3 (C3), Cultivar 4 (C4)

Hipótesis Estadística para DCA Y DBCA

- $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$
- $H_1 : \tau_i \neq 0$ para algún i donde i=1,2,3,4

Ejercicio 2:

Análisis de Varianza DCA

Para la construcción de la tabla ANOVA necesitamos los siguientes cálculos:

Tabla de análisis de varianza para el modelo con un solo factor				
Fuente de Variabilidad	Grados de Libertad	Suma de Cuadrados	Cuadrado Medio	F ₀
τ	$t-1$	$\sum_{i=1}^t \frac{y_{i.}^2}{r_i} - \frac{y_{..}^2}{n}$	$\frac{SC_{\tau}}{t-1}$	$\frac{CM_{\tau}}{CM_{Error}}$
Error	$n-t$	$YY - \sum_{i=1}^t \frac{y_{i.}^2}{r_i}$	$\frac{SC_{Error}}{n-t}$	
Total	$n-1$	$YY - \frac{y_{..}^2}{n}$		

$$\sum_{i=1}^t \frac{y_{i.}^2}{r_i} = \frac{19,5^2}{3} + \frac{22,8^2}{3} + \frac{19,8^2}{3} + \frac{24,9^2}{3} = 637,38 \quad \frac{y_{..}^2}{n} = \frac{87^2}{12} = 630,75 \quad Y^T Y = 649,56$$

$$SC_{\tau} = 637,38 - 630,75 = 6,63 \quad SC_{Error} = 649,56 - 637,38 = 12,18$$

$$SC_{Total} = 649,56 - 630,75 = 18,81$$

Dónde al reemplazar obtenemos la siguiente tabla de ANOVA:

Fuente de Variabilidad	Grados de Libertad	Suma de Cuadrados	Cuadrado Medio	F ₀
t	3	6.63	2.21	1.45155
Error	8	12.18	1.5225	
Total	11	18.81		

- Region de Rechazo y Valor P:

Rc: $\{F_0 | F_0 > F_{0,05,3,8}\} : \{F_0 | F_0 > 4,066181\}$ y Valor p= 0,2577425

Observando la región de rechazo y el valor del p concluimos que no rechazamos la hipótesis nula, lo que nos indica que el efecto de los cultivares es igual 0, lo que indicaría que el rendimiento en t/ha no cambia según el cultivar. Sin embargo, hay que evaluar si esto se debe debido al gradiente que se presenta en la manera que se aplica el riego. lo que perjudicaría en la homogeneidad en las unidades experimentales.

Análisis de Varianza DBCA

Para la construcción de la tabla ANOVA necesitamos los siguientes cálculos:

Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrado Medio	F ₀
Tratamientos (τ)	$t - 1$	$SC_{Tratamientos}$	$CM_{Trat} = \frac{SC_{Trat}}{t - 1}$	$F_0 = \frac{CM_{Trat}}{CM_{Error}}$
Bloques (α)	$b - 1$	$SC_{Bloques}$	$CM_{Bloque} = \frac{SC_{Bloque}}{b - 1}$	
Error	$(t - 1)(b - 1)$	SC_{Error}	$CM_{Error} = \frac{SC_{Error}}{(t - 1)(b - 1)}$	
Total	$tb - 1$	SC_{Total}		

Donde: $SC_{Trat} = \sum_{i=1}^t \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{tb}$ $SC_{Bloque} = \sum_{j=1}^b \frac{y_{.j}^2}{t} - \frac{y_{..}^2}{tb}$

$SC_{Error} = \sum_{i=1}^t \sum_{j=1}^b y_{ij}^2 - \frac{y_{i.}^2}{b} - \frac{y_{.j}^2}{t} + \frac{y_{..}^2}{tb}$ $SC_{Total} = \sum_{i=1}^t \sum_{j=1}^b y_{ij}^2 - \frac{y_{..}^2}{tb}$

Realizando los respectivos cálculos obtenemos:

$$SC_{Trat} = \frac{19,5^2}{3} + \frac{22,8^2}{3} + \frac{19,8^2}{3} + \frac{24,9^2}{3} - \frac{87^2}{12} = 637,38 - 630,75 = 6,63$$

$$SC_{Bloque} = \frac{34^2}{4} + \frac{27,4^2}{4} + \frac{25,6^2}{4} - \frac{87^2}{12} = 640,53 - 630,75 = 9,78$$

$$SC_{Total} = 649,56 - 630,75 = 18,81$$

$$SC_{Error} = 2,4$$

Fuente de Variabilidad	Grados de Libertad	Suma de Cuadrados	Cuadrado Medio	F ₀
t	3	6.63	2.21	5.525
α	2	9.78	4.89	
Error	6	2.4	0.4	
Total	11	18.81		

■ Region de Rechazo y Valor P:

Rc: $\{F_0 | F_0 > F_{0,05,3,6}\} : \{F_0 | F_0 > 4,757063\}$ y Valor p= 0,03673

Observando la región de rechazo y el valor del p concluimos que rechazamos la hipótesis nula, lo que nos indica que por lo menos el efecto de uno de los cultivares es distinto de cero 0, por ende el rendimiento en t/ha varía según el tipo de cultivar.

Se evidencia una clara diferencia entre los el diseño DCA y el diseño DBCA, en primer lugar, al realizar el diseño por bloques se observa una disminución significativa en la suma de cuadrados del error, que a su vez disminuye el cuadrado medio, lo que nos genera un aumento en el valor del estadístico de prueba lo que nos hace rechazar la hipótesis de investigación, dónde podemos concluir que en general para este diseño la opción de bloques nos presenta más información valiosa para toma de decisiones.

Ejercicio 3:

Matrices

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}_{12 \times 8} \quad \mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{41} \\ Y_{42} \\ Y_{43} \end{bmatrix}_{12 \times 1} = \begin{bmatrix} 7,4 \\ 6,5 \\ 5,6 \\ 9,8 \\ 6,8 \\ 6,2 \\ 7,3 \\ 6,1 \\ 6,4 \\ 9,5 \\ 8 \\ 7,4 \end{bmatrix}_{12 \times 1} \quad \beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}_{8 \times 1}$$

$$\mathbf{X}^T \mathbf{Y} = \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \\ Y_{4.} \\ Y_{.1} \\ Y_{.2} \\ Y_{.3} \end{bmatrix}_{8 \times 1} = \begin{bmatrix} 87 \\ 19,5 \\ 22,8 \\ 19,8 \\ 24,9 \\ 34 \\ 27,4 \\ 25,6 \end{bmatrix}_{8 \times 1} \quad \mathbf{X}^T \mathbf{X} = \begin{bmatrix} 12 & 3 & 3 & 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 0 & 0 & 0 & 1 & 1 & 1 \\ 3 & 0 & 3 & 0 & 0 & 1 & 1 & 1 \\ 3 & 0 & 0 & 3 & 0 & 1 & 1 & 1 \\ 3 & 0 & 0 & 0 & 3 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 & 1 & 4 & 0 & 0 \\ 4 & 1 & 1 & 1 & 1 & 0 & 4 & 0 \\ 4 & 1 & 1 & 1 & 1 & 0 & 0 & 4 \end{bmatrix}_{8 \times 8}$$

Estimaciones

Sabemos que el vector de estimaciones para cada elemento del vector β viene dado por la siguiente expresión:

$$\hat{\beta} = \begin{bmatrix} \bar{Y}_{..} \\ \bar{Y}_{1.} - \bar{Y}_{..} \\ \vdots \\ \bar{Y}_{t.} - \bar{Y}_{..} \\ \bar{Y}_{.1} - \bar{Y}_{..} \\ \vdots \\ \bar{Y}_{.b} - \bar{Y}_{..} \end{bmatrix}$$

Ahora de esta forma obtenemos:

$$\hat{\beta} = \begin{bmatrix} 7,25 \\ 6,5 - 7,25 \\ 7,6 - 7,25 \\ 6,6 - 7,25 \\ 8,3 - 7,25 \\ 8,5 - 7,25 \\ 6,85 - 7,25 \\ 6,4 - 7,25 \end{bmatrix}_{8 \times 1} = \begin{bmatrix} 7,25 \\ -0,75 \\ 0,34 \\ -0,65 \\ 1,05 \\ 1,25 \\ -0,4 \\ -0,85 \end{bmatrix}_{8 \times 1}$$

$$\hat{Y} = X\hat{\beta} = \begin{bmatrix} 7,75 \\ 6,1 \\ 5,65 \\ 8,84 \\ 7,19 \\ 6,74 \\ 7,85 \\ 6,2 \\ 5,75 \\ 9,55 \\ 7,9 \\ 7,45 \end{bmatrix}_{12 \times 1} \quad \hat{\epsilon} = \begin{bmatrix} -0,35 \\ 0,4 \\ -0,05 \\ 0,96 \\ -0,39 \\ -0,54 \\ -0,55 \\ -0,1 \\ 0,65 \\ -0,05 \\ 0,1 \\ -0,05 \end{bmatrix}_{12 \times 1}$$

Ejercicio 4:

LSD

Se desea probar $H_0 : \mu_i = \mu_j$ vs $H_1 : \mu_i \neq \mu_j$ para todo $i \neq j$.

Estadístico de prueba:

$$|\bar{y}_i - \bar{y}_j|$$

$$\text{LSD} = t_{\frac{\alpha}{2}, (t-1)(b-1)} * \sqrt{\frac{2CM_{Error}}{b}}$$

$$LSD = 2,446912 * \sqrt{\frac{2*0,4}{3}} = 1,2635$$

Donde $\mu_i = \mu_j$ son significativamente diferentes a nivel α si $|\bar{y}_i. - \bar{y}_j.| > LSD$

Comparación	$ \bar{y}_i. - \bar{y}_j. $	LSD	Conclusión
$\bar{y}_{1.} - \bar{y}_{2.}$	1.1	1.2635	No rechazo H_0
$\bar{y}_{1.} - \bar{y}_{3.}$	0.1	1.2635	No rechazo H_0
$\bar{y}_{1.} - \bar{y}_{4.}$	1.8	1.2635	Rechazo H_0
$\bar{y}_{2.} - \bar{y}_{3.}$	1	1.2635	No rechazo H_0
$\bar{y}_{2.} - \bar{y}_{4.}$	0.7	1.2635	No rechazo H_0
$\bar{y}_{3.} - \bar{y}_{4.}$	1.7	1.2635	Rechazo H_0

Se observa que el cultivar 4 presenta diferencias significativas entre el cultivar 1 y el cultivar 3, además este presenta mayor media, por lo que se podría decir que frente al cultivar 1 y el cultivar 3, el cultivar 4 presenta un mayor rendimiento en t/ha. No se evidencian mejores rendimientos entre los otros.

Ejercicio 5:

Contrastes

Contraste/tratamiento	τ_1	τ_2	τ_3	τ_4
K_1	1	-1	0	0
K_2	1	1	-2	0
K_3	1	1	1	-3

Hipótesis según el contraste K:

K1:

$$H_0 : \tau_1 - \tau_2 = 0 \text{ vs } H_1 : \tau_1 - \tau_2 \neq 0$$

K2:

$$H_0 : \tau_1 + \tau_2 - 2\tau_3 = 0 \text{ vs } H_1 : \tau_1 + \tau_2 - 2\tau_3 \neq 0$$

K3:

$$H_0 : \tau_1 + \tau_2 + \tau_3 - 3\tau_4 = 0 \text{ vs } H_1 : \tau_1 + \tau_2 + \tau_3 - 3\tau_4 \neq 0$$

De los contrastes anteriores tenemos lo siguiente:

	τ_1	τ_2	τ_3	τ_4	GL	$(\sum_i C_i Y_{i.})^2$	$b_i \sum_i C_i^2$	CM	F	F_0
C ₁	1.00	-1.00	0.00	0.00	1.00	10.89	6.00	1.81	4.53	5.99
C ₂	1.00	1.00	-2.00	0.00	1.00	7.29	18.00	0.40	1.01	
C ₃	1.00	1.00	1.00	3.00	1.00	158.7	36.00	4.41	11.025	

Se uso el cuadrado medio y los grados de libertad del error, los calculos fueron:

$$(\sum_i C_i Y_{i.}) = 19,5(1) + 22,8(-1) + 19,8(0) + 24,9(0) = (-3,3)^2 = 10,89$$

$$(\sum_i C_i Y_{i.}) = 19,5(1) + 22,8(1) + 19,8(-2) + 24,9(0) = (2,7)^2 = 7,29$$

$$(\sum_i C_i Y_{i.}) = 19,5(1) + 22,8(1) + 19,8(1) + 24,9(3) = (-12,6)^2 = 158,76$$

$$\begin{aligned}
b_i \sum_i C_i^2 &= 3(1 - 1 + 0 + 0)^2 = 6 & b_i \sum_i C_i^2 &= 3(1 + 1 - 2 + 0)^2 = 18 \\
b_i \sum_i C_i^2 &= 3(1 + 1 + 1 + 3)^2 = 36 & (\sum_i C_i Y_i.)^2 / (b_i \sum_i C_i^2) &= 1,815 \\
(\sum_i C_i Y_i.)^2 / (b_i \sum_i C_i^2) &= 0,405 & (\sum_i C_i Y_i.)^2 / (b_i \sum_i C_i^2) &= 519,84 \\
F_i = CM_c / CM_{error} &\rightarrow F_1 = 4,53 | F_2 = 1,01 | F_3 = 11,02
\end{aligned}$$

Dónde podemos concluir que si existen diferencias para la combinación lineal en el contraste número 3, es decir entre todos los tratamientos y en general vemos que dónde se presenta el tratamiento 4 es dónde más evidencias significativas se notan referente a los tratamientos como en el LSD. Por último no vemos diferencias significativas en k_1 es decir entre el tratamiento 1 y 2, y en k_2 que nos describe que no hay diferencias en general entre los tratamientos 1, 2 y 3.