Diseños y análisis de experimentos: Taller 2

Andrés Felipe Palomino - David Stiven Rojas - Mateo Trochéz
Códigos:1922297-1924615-1931043
Universidad del Valle

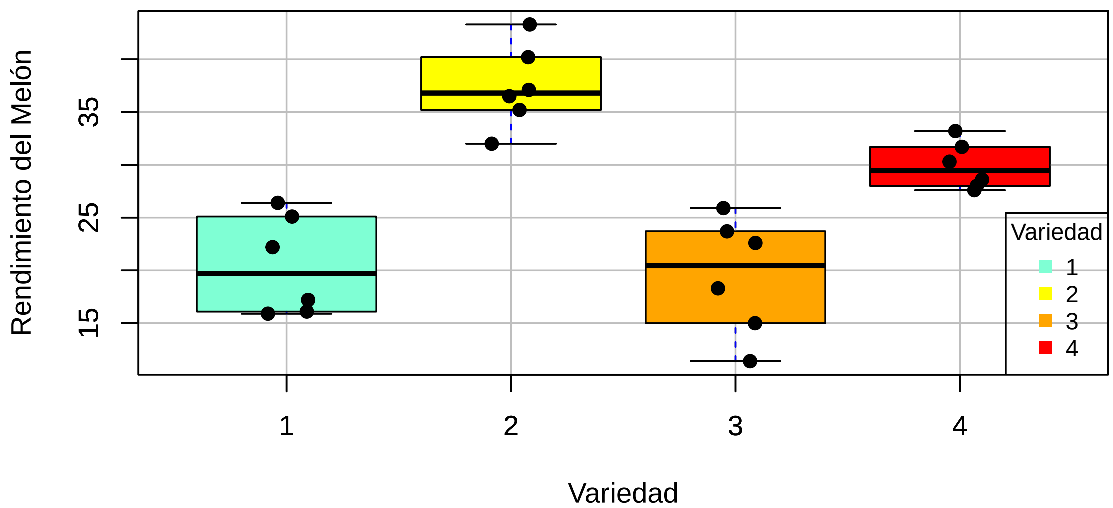
26 de julio de 2023

Universidad del Valle

Ejercicio 1: En un centro de investigación para el mejoramiento de melones se realizó un experimento con el interés de evaluar el rendimiento de 4 variedades. Cada variedad se probó en seis parcelas de campo. La asignación de los tratamientos a las unidades experimentales fue completamente al azar. Por lo tanto, el experimento se planteó como un diseño completamente al azar -(DCA).

Variedad	Replicaciones					
	1	2	3	4	5	6
V1	25.1	17.2	26.4	16.1	22.2	15.9
V2	40.2	35.2	32.0	36.5	43.3	37.1
V3	18.3	22.6	25.9	15.0	11.4	23.7
V4	28.6	28.0	33.2	31.7	30.3	27.6

Figura 1: Diagrama de caja y bigotes según el tipo de variedad de melón.



Objetivos del estudio:

- Evaluar el rendimiento de los 4 tipos de variedades de melón.

Hipótesis del estudio:

- El rendimiento de las 4 variedades de melón es igual.
- Por lo menos una de las 4 variedades presenta un rendimiento distinto.

Factores:

- Variedad

Niveles:

- Variedad 1 (V1), Variedad 2 (V2), Variedad 3 (V3), Variedad 4 (V4)

Tratamientos:

- Variedad 1 (V1), Variedad 2 (V2), Variedad 3 (V3), Variedad 4 (V4)

Unidad experimental:

- Parcela de melón.

Matrices

$$\mathbf{X}=\begin{bmatrix}1&1&0&0&0\\1&1&0&0&0\\1&1&0&0&0\\1&1&0&0&0\\1&1&0&0&0\\1&1&0&0&0\\.&.&.&.&.\\.&.&.&.&.\\.&.&.&.&.\\1&0&0&0&1\\1&0&0&0&1\\1&0&0&0&1\\1&0&0&0&1\\1&0&0&0&1\\1&0&0&0&1\end{bmatrix}_{24x5} \quad \mathbf{Y}=\begin{bmatrix}Y_{11}\\Y_{12}\\Y_{13}\\Y_{14}\\Y_{15}\\Y_{16}\\Y_{..}\\Y_{..}\\Y_{..}\\Y_{..}\\Y_{41}\\Y_{42}\\Y_{43}\\Y_{44}\\Y_{45}\\Y_{46}\end{bmatrix}_{24x1}=\begin{bmatrix}25{,}1\\17{,}2\\26{,}4\\16{,}1\\22{,}2\\15{,}9\\..\\..\\..\\..\\28{,}6\\28\\33{,}2\\31{,}7\\30{,}3\\27{,}6\end{bmatrix}_{24x1} \quad X^T\mathbf{Y}=\begin{bmatrix}Y..\\Y_1.\\Y_2.\\Y_3.\\Y_4.\end{bmatrix}_{5x1}=\begin{bmatrix}643{,}5\\122{,}9\\224{,}3\\116{,}9\\179{,}4\end{bmatrix}_{5x1} \quad \beta=\begin{bmatrix}\mu\\\tau_1\\\tau_2\\\tau_3\\\tau_4\end{bmatrix}$$

$$X^T\mathbf{X}=\begin{bmatrix}24&6&6&6&6\\6&6&0&0&0\\6&0&6&0&0\\6&0&0&6&0\\6&0&0&0&6\end{bmatrix}_{5x5}$$

Estimaciones

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

$$\begin{aligned}\hat{\mu} &= \bar{Y}_{..}\\ \hat{\tau}_1 &= \bar{Y}_{1.}-\hat{\mu} \quad = \bar{Y}_{1.}-\bar{Y}_{..}\\ &\vdots\\ \hat{\tau}_t &= \bar{Y}_{t.}-\bar{Y}_{..}\end{aligned}$$

Ahora de esta forma obtenemos:

$$\hat{\beta}=\begin{bmatrix}26.8125\\20.48333\text{-}26.8125\\37.38333\text{-}26.8125\\19.48333\text{-}26.8125\\29.9\text{-}26.8125\end{bmatrix}_{5x1}=\begin{bmatrix}26{,}8125\\-6{,}329167\\10{,}57083\\-7{,}329167\\3{,}0875\end{bmatrix}_{5x1}$$

$$\hat{Y} = X\hat{\beta} = \begin{bmatrix} 20{,}48333 \\ 20{,}48333 \\ 20{,}48333 \\ 20{,}48333 \\ 20{,}48333 \\ 20{,}48333 \\ 37{,}38333 \\ 37{,}38333 \\ 37{,}38333 \\ 37{,}38333 \\ 37{,}38333 \\ 37{,}38333 \\ 19{,}48333 \\ 19{,}48333 \\ 19{,}48333 \\ 19{,}48333 \\ 19{,}48333 \\ 19{,}48333 \\ 29{,}90000 \\ 29{,}90000 \\ 29{,}90000 \\ 29{,}90000 \\ 29{,}90000 \\ 29{,}90000 \end{bmatrix}_{24x1}$$

Para realizar el calculo de la estimacion de la varianza se procede a desarrollar la siguiente expresión:

- $\hat{\sigma^2} = \frac{Y^TY - \sum_{i=1}^{t} \frac{y_{i.}^2}{r_i}}{n-t}$ Esto porque la estimación se calcula a partir de la suma de cuadrados media de los residuos.

  $Y^TY = 18912{,}35$; $\quad \sum_{i=1}^{t} \frac{y_{i.}^2}{r_i} = 2517{,}401 + 8385{,}081 + 2277{,}6 + 5364{,}06 = 18544{,}142$

  $\hat{\sigma^2} = \frac{18912{,}35 - 18544{,}142}{20} = 18{,}4104$

- $\hat{\tau}_1 - \hat{\tau}_2 = \lambda_1^T \hat{\beta}$ donde $\lambda_1^T = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}$

  $$\hat{\tau}_1 - \hat{\tau}_2 = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 26{,}8125 \\ -6{,}329167 \\ 10{,}57083 \\ -7{,}329167 \\ 3{,}0875 \end{bmatrix} = -16{,}899$$

- $\hat{\tau}_1 + \hat{\tau}_2 - 2\hat{\tau}_3 = \lambda_1^T \hat{\beta}$ donde $\lambda_1^T = \begin{bmatrix} 0 & 1 & 1 & -2 \end{bmatrix}$

$$\hat{\tau}_1 + \hat{\tau}_2 - 2\hat{\tau}_3 = \begin{bmatrix} 0 & 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} 26{,}8125 \\ -6{,}329167 \\ 10{,}57083 \\ -7{,}329167 \\ 3{,}0875 \end{bmatrix} = 18{,}899$$

- $R(\tau/\mu) = \sum_{i=1}^{t} (\bar{y}_{i.} - \bar{y}_{..}) * y_{i.} = -777{,}8586 + 2371{,}0371 - 856{,}7799 + 553{,}8975 = 1290{,}29602$

Análisis de Varianza e Hipótesis Estadistica

- $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$

- $H_1 : \tau_i \neq \tau_j$ para algún par i, j donde i=1,2,3,4 ; j=1,2,3,4

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	$n-1$	$YY - \frac{y_{..}^2}{n}$		

Dónde al reemplazar obtenemos la siguiente tabla de ANOVA:

Fuente de Variabilidad	Grados de Libertad	Suma de Cuadrados	Cuadrado Medio	F_0
t	3	1290.30	430.10	23.36
Error	20	368.20	18.41	
Total	23	1658.50		

- Region de Rechazo y Valor P:

Rc: $\{F_0 | F_0 > F_{0{,}05,3,20}\} : \{F_0 | F_0 > 3{,}098391\}$
Observando la region de rechazo y dado a que al valor del estadístico de prueba F_0 le corresponde un valor p de 0.00000096, concluimos que rechazamos la hipótesis nula lo que nos indica que por lo menos un par τ_i y τ_j son diferentes de 0. Concluimos que el rendimiento de las 4 variedades de melón no es igual, entonces es pertinente evaluar cuál produce el mayor rendimiento para obtener el respectivo mejoramiento de este. Situación que observábamos inicialmente en la Figura 1 dónde se evidencia que las variedades 1 y 3 presentan rendimientos inferiores respecto a 2 y 4, dónde incluso

observamos gráficamente que el mayor de estos fue presentado por la variedad 2. Estas descripciones iniciales también se pueden evidenciar en la siguiente tabla con un resumen de las descriptivas según el tipo de variedad del melón, dónde los rangos, medias, medias y demás son respaldas por el ANOVA.

	V1	V2	V3	V4
Min.	15.90	32.00	11.40	27.60
1st Qu.	16.38	35.53	15.82	28.15
Median	19.70	36.80	20.45	29.45
Mean	20.48	37.38	19.48	29.90
3rd Qu.	24.38	39.43	23.42	31.35
Max.	26.40	43.30	25.90	33.20
Sd	4.70	3.94	5.58	2.23

Tabla 1: Estadísticas descriptivas según la variedad del melón.