Modelos de Rango Incompleto
Modelos de Clasificación Simple
Diseño Completamente Aleatorizado

Johan Steven Aparicio

Escuela de Estadística
Facultad de Ingeniería - Universidad del Valle
johan.aparicio@correounivalle.edu.co

Abril, 2022

Tabla de Contenido

Los modelos de rango incompleto se presentan cuando en el experimento al menos uno de los factores es cualitativo, es decir, que los niveles de algún factor no se especifican en términos cuantitativos sino cualitativos.

Los modelos de rango incompleto se presentan cuando en el experimento al menos uno de los factores es cualitativo, es decir, que los niveles de algún factor no se especifican en términos cuantitativos sino cualitativos.

Ejemplo:

Se desea evaluar el efecto de la cantidad de fertilizante sobre el rendimiento productivo de una planta, para ello se establecen 3 cantidades: 20, 50, 100 gr/planta.

Los anteriores niveles también se pueden especificar cualitativamente como: 20=Bajo, 50=Medio, 100=Alto.

Considere el experimento donde se estudia un sólo factor de tratamientos con niveles especificados cualitativamente.

Específicamente se estudiará el rendimiento de un tipo de planta el cual depende del factor riego. Considérense tres niveles o clases de riego:

- Riego por goteo (G)
- Riego por gravedad (F)
- Riego por aspersión (A)

El objetivo es comparar los efectos sobre el rendimiento de estas tres clases de riego.

El experimento se ha planeado de tal forma que el número de replicaciones por clase de riego es el siguiente:

- 3 replicas de riego por goteo
- 2 replicas de riego por gravedad
- 4 replicas de riego por aspersión

Se necesitan entonces 9 unidades experimentales (parcelas) para los tres tratamientos que serán asignados al azar en esas unidades experimentales.

Sea y_{ij} el rendimiento de la planta, con i indicando el nivel de riego (1:G, 2:F, 3:A) y j indicando la replicación (1,2,3,4).
Los datos se pueden organizar en la siguiente tabla:

Nivel	G	F	A	
	y_{11}	y_{21}	y_{31}	
	y_{12}	y_{22}	y_{32}	
	y_{13}		y_{33}	
			y_{34}	
Totales	$y_{1.}$	$y_{2.}$	$y_{3.}$	$y_{..}$
Medias $\bar{y}_{i.}$	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$	$\bar{y}_{..}$
repeticiones	r_1	r_2	r_3	n

El modelo puede estructurarse de la siguiente manera:

$$y_{ij} = \beta_0 + \beta_1 G + \beta_2 F + \beta_3 A + \varepsilon_{ij}$$

con $G, F, A = \{0, 1\}$

En los diseños con factores cualitativos habrán tantos parámetros como niveles se tengan (más β_0).

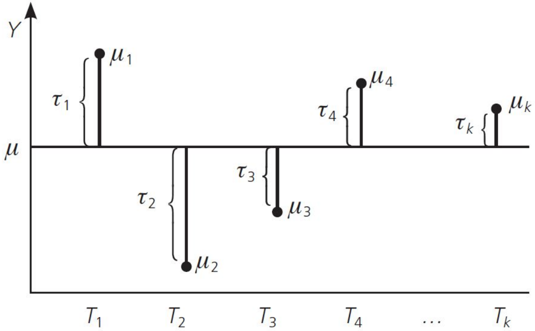
Dado que las variables asociadas a los niveles del factor cualitativo deben ser 1 o 0, el modelo se puede reexpresar como:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \text{ con } i = 1, ..., t \text{ y } j = 1, ..., r_i$$

donde,

- μ :
- τ_i :
- $\mu + \tau_i$:
- $(\mu + \tau_i) - (\mu + \tau_i)$:
- ε_{ij} :

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \text{ con } i = 1, ..., t \text{ y } j = 1, ..., r_i$$



Ejercicios 1

1. Especifique de manera general las matrices X (matriz de diseño), $X'X$ y los vectores Y, $X'Y$, β.
2. Halle el rango de X y de $X'X$.

Como se puede corroborar, el rango de X no es completo porque la suma de las columnas correspondientes a $\tau_1, \tau_2, ..., \tau_t$ es igual a la columna de unos correspondiente a μ.

Como el rango de X es t; entonces el rango de $X'X$ es también igual a t, lo que implica que $X'X$ es una matriz singular y no posee inversa. Por lo tanto, el SEN no tendrá solución o tiene infinitas soluciones, siendo esta la primer desventaja.

La segunda desventaja que se tiene, la constituye el hecho de que la estimación de β no es insesgada.

En el modelo de rango completo se estima con:

$$\hat{\beta} = (X'X)^{-1}X'Y = CY$$

$$\text{con, } E[CY] = E[\hat{\beta}] = \beta$$

Donde cada $\hat{\beta}_i$ (estimador lineal del parámetro β_i) es una combinación lineal de Y .

Probar que C no existe calculando $E[CY]$.

Probemos que C no existe:

$$E[CY] = E[C(X\tilde{\beta} + \varepsilon)] = CX\beta + E(\varepsilon) = CX\beta$$

Si el estimador fuese insesgado, entonces $CX\beta = \beta$

Lo que implicaría que $CX = I$, donde $r(I) = t + 1$

El rango de I debe ser igual al menor de los dos rangos de X y C lo cual no se cumple, ya que $r(X) = t$

Según lo anterior, C no existe. Luego β no tiene estimador lineal insesgado.

Estas desventajas son superables si en vez de parámetros individuales, se consideran combinaciones lineales de ellos.

Sea $\widetilde{\lambda'\beta}$ una combinación lineal de las componentes β.

Definición

$\widetilde{\lambda'\beta}$ es estimable insesgadamente si y solo si existe una combinación lineal de Y, tal que el valor esperado de esa combinación es $\lambda'\beta$

Supongamos que existe un vector b tal que $E[b'Y] = \lambda'\beta$, b de dimensiones $n \times 1$, λ de dimensiones $p \times 1$

Supongamos que existe un vector b tal que $E[b'Y] = \lambda'\beta$, b de dimensiones $n \times 1$, λ de dimensiones $p \times 1$

Si existe ese b entonces:

$$\begin{aligned} E[b'Y] &= E[b'(X\tilde{\beta} + \varepsilon)] \\ &= E[b'X\tilde{\beta}] + E[\varepsilon] \\ &= b'XE[\tilde{\beta}] \\ &= \lambda'\beta \end{aligned}$$

Existe una estimación insesgada de una combinación lineal de parámetros, si el vector λ' resulta como una combinación lineal de las filas de la matriz de diseño.

$b'X = \lambda'$

$$\begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_{t+1} \end{bmatrix}$$

Es decir que:

$$b'X = \sum_{i=1}^{n} b_i [X_{i1} \ldots X_{i(t+1)}]$$

Ejercicios 2

Para el caso ilustrativo planteado inicialmente:

1. Especifique las matrices (o vectores) Y, matriz de diseño, β , $X'X$ y $X'Y$.
2. Es $\tau_1 - \tau_2$ estimable insesgadamente?
3. Es $\tau_1 + \tau_2 - 2\tau_3$ estimable insesgadamente?
4. Es $\mu + \tau_1$ estimable insesgadamente?
5. Es $\tau_1 + \tau_2$ estimable insesgadamente?