

Estimación de parámetros en el modelo estadístico lineal de rango incompleto a través de la inversa generalizada: ventajas, desventajas y análisis.

Estimation of parameters in the linear statistical model of incomplete range through the generalized inverse: advantages, disadvantages and analysis.

Rivas, Felvir

Universidad Nacional Experimental de los Llanos Occidentales Ezequiel Zamora
Programa de Ingeniería, Arquitectura y Tecnología
Barinas, Venezuela
felvir11@yahoo.com

Resumen

En un modelo de rango incompleto la matriz X^tX del sistema de ecuaciones normales no tiene inversa ordinaria y por lo tanto se origina un problema para encontrar los estimadores del parámetro β , ya que, a pesar de ser un sistema de ecuaciones consistente, admite infinitas soluciones. Para dar solución a dicho problema se recurre a ciertas técnicas, las más comunes son aquellas donde se hacen restricciones sobre los parámetros, a pesar de que al utilizar estas restricciones se pueden obtener diferentes resultados. Por tal razón, cuando se trabaja con un modelo de rango incompleto deben de considerarse solo parámetros o funciones de parámetros que tengan estimadores idénticos. La inversa generalizada o condicional es una alternativa para dar solución al sistema de ecuaciones normales en un modelo de diseño de rango incompleto sin realizar restricciones sobre los parámetros. En este trabajo se comparan los parámetros estimados en el modelo estadístico lineal de rango incompleto a través de la inversa generalizada (o condicional) con los estimados por vía mínimos cuadrados y máxima verosimilitud, por medio del coeficiente de determinación y los intervalos de confianza. Se encontró que los métodos de estimación empleados están uniformes en cuanto a resultados, ya que el vector de respuesta es el mismo a pesar de que el vector de parámetros estimado $\hat{\beta}$ es diferente por cada vía. El método de estimación por inversa generalizada da un coeficiente de determinación del modelo similar a los otros métodos de estimación utilizados, sin embargo fue en éste donde se obtuvo la longitud más pequeña de los intervalos de confianza de los parámetros, y el método más deficiente fue el modelo reducido.

Palabras claves: Modelo estadístico lineal de rango incompleto, inversa generalizada, método de mínimos cuadrados, método de máxima verosimilitud, funciones estimables.

Abstract

In an incomplete model range the X^tX matrix of normal equations system has no ordinary inverse and therefore it causes a problem to find the estimators of the parameter β , since, although being a consistent equations system, admits infinite solutions. To solve this problem resorts to certain techniques, the most common are those where restrictions are made about parameters, although using these restrictions can get different results. For this reason, when working with an incomplete model range should be considered only parameters or functions of parameters having identical estimators. The generalized inverse or conditional is an alternative to solve the system of normal equations in a model of incomplete design range without making restrictions on the parameters. In this work compares the estimated parameters in the statistical model range line incomplete through the generalized inverse (or conditional) with estimated minimal squares route and maximal likelihood, using the coefficient of determination and confidence intervals. It found that the estimation methods used are uniform in terms of results, as the response vector is the same even though the estimated parameter $\hat{\beta}$ vector is different for each route. The estimation method for generalized inverse gives a determination coefficient of the similar model to the others estimation methods used, however this was where it got the shorter length of the confidence intervals of the parameters, and the most deficient method was the reduced model.

Keywords: Statistical lineal model of incomplete range, generalized inverse, minimal square method, maximal likelihood method, estimable functions.

1 Introducción

El problema de estimación en un modelo lineal de rango incompleto puede ser abordado mediante el concepto de estimabilidad de Bose (Bose 1959), concepto que es discutido con cierto detalle Graybill (Graybill 1976). En este modelo estadístico, se presta menor atención a la presentación y desarrollo de la hipótesis lineal general; concretamente al problema de las condiciones que deben cumplir las funciones que componen dicha hipótesis. Cuando se trata el problema se suele suponer la condición de estimabilidad de las funciones componentes (Searle 1997) ya que el numerador del test F que se espera, por analogía al caso de rango completo, sería invariante frente a las infinitas soluciones de las ecuaciones normales.

La estimación de parámetros se realiza mediante el de Análisis de Varianza, y este a su vez es una consecuencia de los métodos de mínimos cuadrados o máxima verosimilitud, los cuales garantizan estimaciones insesgadas. El problema radica cuando la matriz de diseño es de rango incompleto, ya que el número de parámetros a estimar es superior al número de ecuaciones distintas no colineales que pueden ser planteadas.

Existen y se estudiarán tres formas de solucionar el problema:

- i. Reparametrizar.
- ii. Imponer restricciones a las soluciones para aumentar el rango columna de X . Serán tantas restricciones como p -rang (X).
- iii. Utilizar la inversa generalizada o la inversa condicional.

Aunque cualquiera de estas alternativas funciona, en el sentido que se obtienen “buenas” estimaciones de los parámetros, la última forma permite un tratamiento teórico más consistente y general, debido a que considera todas las ecuaciones del sistema de ecuaciones normales y no usa funciones lineales de los parámetros no estimables, por lo que se hará especial énfasis en este caso, y comparar con los otros, desde el punto de vista del coeficiente de determinación y la longitud de los intervalos de confianza.

2 Marco Teórico

Los datos corresponden a un Trabajo de Investigación “Cambios en la actividad biológica del suelo e influencia del tipo de uso de la tierra sobre la fertilidad del suelo” de la Universidad Nacional Experimental Francisco de Miranda del Estado falcón.

Se plantea un diseño factorial 2×3 con efectos fijos e interacción, es decir, un diseño con dos factores **A** y **B**, con dos niveles el primer factor \mathbf{a}_1 y \mathbf{a}_2 , tres niveles para el segundo factor \mathbf{b}_1 , \mathbf{b}_2 y \mathbf{b}_3 .

El modelo viene dado por:

$$Y = X\beta + \varepsilon, \varepsilon \sim N(\mathbf{0}, \sigma^2 I_d) \quad (1)$$

donde,

$$\begin{aligned} \mu &= [\mu], \\ \beta^t &= [\mu \tau^t \alpha^t (\tau\alpha)^t], \\ \tau^t &= [\tau_1 \tau_2], \quad \alpha^t = [\alpha_1 \alpha_2 \alpha_3], \\ (\tau\alpha)^t &= [(\tau\alpha)_{11} (\tau\alpha)_{12} (\tau\alpha)_{13} (\tau\alpha)_{21} (\tau\alpha)_{22} (\tau\alpha)_{23}], \\ X &\text{ es la matriz de diseño.} \end{aligned}$$

El modelo también puede ser escrito como:

$$y_{ijk} = \mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk} \\ i = 1, 2; j = 1, 2, 3; k = 1, \dots, 15$$

donde,

y_{ijk} = Cambio de la actividad biológica del suelo e influencia del tipo de uso de la tierra sobre la fertilidad del suelo en la escuela técnica San Luis, estado Falcón.

μ = Media global del modelo.

τ_i = Efecto del i –ésimo implante la media global.

α_j = Efecto del j –ésimo balance de nitrógeno sobre la media global.

$(\tau\alpha)_{ij}$ = Efecto combinado del i –ésimo implante y el j –ésimo balance de nitrógeno sobre la media global.

ε_{ijk} = Error experimental aleatorio asociado al i –ésimo implante, el j –ésimo balance de nitrógeno en la k –ésima réplica.

La matriz de diseño, es una matriz de orden (90×12) de rango 6, por lo tanto se tiene un modelo de diseño de rango incompleto por columnas.

Se sabe que el método mínimos cuadrados y el método de máxima verosimilitud (bajo el supuesto de normalidad) conducen al mismo sistema de ecuaciones $(X^t X)\hat{\beta} = X^t Y$. Ahora bien, como se está trabajando con un modelo de rango incompleto $(X^t X)^{-1}$ no existe, es por ello que para resolver el problema se analizarán los enfoques señalados en este trabajo que son: mediante un modelo reducido, a partir de funciones estimables o mediante restricciones identificables sobre los coeficientes del modelo. También se analizará utilizando la inversa generalizada y condicional para así comparar dichos métodos por medio del coeficiente de determinación y los intervalos de confianza de cada β .

3 Resultados y Discusión

3.1. Modelo reducido.

Se tomará X_1 la matriz de orden (90×6) con las $6 = \text{rang}(X)$ columnas linealmente independientes de la matriz de diseño X y X_2 la matriz de orden (90×6) forma-

da con las demás columnas de X , es decir,

$$X_1 = [X_{i1} X_{i2} X_{i4} X_{i5} X_{i7} X_{i8}], \text{ para } i = 1, 2, \dots, 90.$$

$$X_2 = [X_{i3} X_{i6} X_{i9} X_{i10} X_{i11} X_{i12}], \text{ para } i = 1, 2, \dots, 90.$$

Luego, la matriz V de orden (6×6) , que cumpla con la condición $X_2 = X_1 V$ viene dada por:

$$V = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & 1 \end{bmatrix}$$

Así $X\beta = X_1\beta_1 + X_2\beta_2 = X_1(\beta_1 + V\beta_2) = X_1\beta^*$, donde β_1 de orden (6×1) y β_2 de orden (6×1) .

Por lo tanto, el modelo $Y = X_1\beta^* + \varepsilon$ es de rango completo por columnas y podemos estimar β^* , donde $\hat{\beta}^* = (X_1^t X_1)^{-1} X_1^t Y$.

Las matrices $(X_1^t X_1)$ y $(X_1^t Y)$ vienen dadas por:

$$X_1^t X_1 = \begin{bmatrix} 904530301515 \\ 454515151515 \\ 3015300150 \\ 3015030015 \\ 1515150150 \\ 1515015015 \end{bmatrix}$$

$$X_1^t Y = \begin{bmatrix} 103,8737 \\ 49,4917 \\ 25,6751 \\ 35,9786 \\ 3,8298 \\ 21,8453 \end{bmatrix}$$

Así, el vector de parámetros estimado para el modelo de rango completo reducido es:

$$\hat{\beta}^* = \begin{bmatrix} 1,2269 \\ 0,3609 \\ 0,2295 \\ -0,2847 \\ -1,5619 \\ 0,1533 \end{bmatrix}$$

Y la respuesta estimada $\hat{Y}^* = X_1 \hat{\beta}^*$ viene dada por:

$$\hat{Y}^* = \begin{bmatrix} 0,2553 \\ \vdots \\ 0,2553 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 1,5878 \\ \vdots \\ 1,5878 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 0,9422 \\ \vdots \\ 0,9422 \\ 1,2269 \\ \vdots \\ 1,2269 \end{bmatrix}_{90 \times 12}$$

En la cual,

$$y_{11}^* = y_{21}^* = \dots = y_{151}^* = 0,2553,$$

$$y_{161}^* = y_{171}^* = \dots = y_{301}^* = 1,4564,$$

$$y_{311}^* = y_{321}^* = \dots = y_{451}^* = 1,5878,$$

$$y_{461}^* = y_{471}^* = \dots = y_{601}^* = 1,4564,$$

$$y_{611}^* = y_{621}^* = \dots = y_{751}^* = 0,9422,$$

$$y_{761}^* = y_{771}^* = \dots = y_{901}^* = 1,2269$$

El coeficiente de determinación viene dado por:

$$R^2 = \frac{(\hat{\beta}^*)^t X_1^t Y - n(\bar{Y})^2}{Y^t Y - n(\bar{Y})^2} \quad (2)$$

donde,

$$(\hat{\beta}^*)^t X_1^t Y = 138,3178$$

$$Y^t Y = 167,6120$$

$$n = 90$$

$$\bar{Y} = 1,1542.$$

Luego,

$$R^2 = \frac{138,3178 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = 0,3860$$

Y los intervalos de confianza:

$$I(\beta^*)_1 = (0,9240; 1,5297),$$

$$L(I(\beta^*)_1) = 0,6057$$

$$I(\beta^*)_2 = (-0,0679; 0,7897),$$

$$L(I(\beta^*)_2) = 0,8576$$

$$I(\beta^*)_3 = (-0,1993; 0,6583),$$

$$L(I(\beta^*)_3) = 0,8576$$

$$I(\beta^*)_4 = (-0,7135; 0,1441),$$

$$L(I(\beta^*)_4) = 0,8576$$

$$I(\beta^*)_5 = (-2,1683; -0,9554),$$

$$L(I(\beta^*)_5) = 1,2129$$

$$I(\beta^*)_6 = (-0,4531; 0,7597),$$

$$L(I(\beta^*)_6) = 1,2128$$

3.2. *Imponiendo restricciones.*

Se debe encontrar un conjunto de $12 - 6 = 6$ vectores v_i de dimensión 12×1 linealmente independientes que sean también linealmente independientes de las filas de X . Se tomará:

$$V = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Así, $H = \begin{bmatrix} X^t X & V^t \\ V 0 & \end{bmatrix}$ es una matriz cuadrada de orden (18×18) , de rango completo por columnas y por ende invertible. Luego,

$$\begin{bmatrix} \hat{\beta} \\ 0 \end{bmatrix} = \begin{bmatrix} X^t X & V^t \\ V 0 & \end{bmatrix}^{-1} \begin{bmatrix} X^t Y \\ 0 \end{bmatrix}$$

Donde,

$$\hat{\beta}_r = \begin{bmatrix} 1,1542 \\ -0,0543 \\ 0,0543 \\ 0,2479 \\ -0,7358 \\ 0,4880 \\ -1,0924 \\ 1,0924 \\ 1,1370 \times 10^{-16} \\ -1,1370 \times 10^{-16} \\ 0,4696 \\ -0,4696 \end{bmatrix}$$

$$\hat{Y}_r = \begin{bmatrix} 0,2553 \\ \vdots \\ 0,2553 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 1,5878 \\ \vdots \\ 1,5878 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 0,9422 \\ \vdots \\ 0,9422 \\ 1,2269 \\ \vdots \\ 1,2269 \end{bmatrix}_{90 \times 12}$$

En la cual,

$$\begin{aligned} y_{11} &= y_{21} = \dots = y_{151} = 0,2553, \\ y_{161} &= y_{171} = \dots = y_{301} = 1,4564, \\ y_{311} &= y_{321} = \dots = y_{451} = 1,5878, \\ y_{461} &= y_{471} = \dots = y_{601} = 1,4564, \\ y_{611} &= y_{621} = \dots = y_{751} = 0,9422, \\ y_{761} &= y_{771} = \dots = y_{901} = 1,2269 \end{aligned}$$

El coeficiente de determinación es:

$$R^2 = \frac{138,3178 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = 0,3860$$

Los intervalos de confianza son:

$$\begin{aligned} I(\beta_r)_1 &= (1,0300 ; 1,2783), \\ L(I(\beta_r)_1) &= 0,2483 \\ I(\beta_r)_2 &= (-0,1784 ; 0,0698), \\ L(I(\beta_r)_2) &= 0,2482 \\ I(\beta_r)_3 &= (-0,0698 ; 0,1784), \\ L(I(\beta_r)_3) &= 0,2482 \\ I(\beta_r)_4 &= (-3,2129 \times 10^{-4} ; 0,4954), \\ L(I(\beta_r)_4) &= 0,4961 \\ I(\beta_r)_5 &= (-1,0859 ; -0,3856), \\ L(I(\beta_r)_5) &= 0,7002 \\ I(\beta_r)_6 &= (0,2404 ; 0,7355), \\ L(I(\beta_r)_6) &= 0,4951 \\ I(\beta_r)_7 &= (-1,4425 ; -0,7422), \\ L(I(\beta_r)_7) &= 0,7002 \\ I(\beta_r)_8 &= (0,7422 ; 1,4425), \\ L(I(\beta_r)_8) &= 0,7003 \\ I(\beta_r)_9 &= (6,7617 \times 10^{-17} ; 1,5978 \times 10^{-16}), \\ L(I(\beta_r)_9) &= 9,2165 \times 10^{-17} \\ I(\beta_r)_{10} &= 0 \\ I(\beta_r)_{11} &= (0,1194 ; 0,8197), \\ L(I(\beta_r)_{11}) &= 0,7003 \\ I(\beta_r)_{12} &= (-0,8197 ; -0,1194), \\ L(I(\beta_r)_{12}) &= 0,7002 \end{aligned}$$

Este método, presenta el siguiente inconveniente: “las ecuaciones lineales de los parámetros que se agregan tienen la particularidad de no ser estimables ya que no son contrastes”. Es por ello, que un problema interesante para el investigador es establecer si las funciones de interés son o no estimables.

4.3. *Funciones estimables.*

Para encontrar dichas funciones se utilizarán las ecuaciones normales las cuales vienen dadas por:

$$90\hat{\mu} + 45 \sum_{i=1}^2 \hat{\tau}_i + 30 \sum_{j=1}^3 \hat{\alpha}_j + 15 \sum_{i=1}^2 \sum_{j=1}^3 (\hat{\tau}\hat{\alpha})_{ij} = y_{\dots} \quad (3)$$

$$45\hat{\mu} + 45\hat{\tau}_i + 15 \sum_{j=1}^3 \hat{\alpha}_j + 15 \sum_{j=1}^3 (\hat{\tau}\hat{\alpha})_{ij} = y_{i..} \quad (4)$$

$$30\hat{\mu} + 15 \sum_{i=1}^2 \hat{\tau}_i + 30\hat{\alpha}_j + 15 \sum_{i=1}^2 (\hat{\tau\alpha})_{ij} = y_j. \quad (5)$$

$$15\hat{\mu} + 15\hat{\tau}_i + 15\hat{\alpha}_j + 15(\hat{\tau\alpha})_{ij} = y_{ij}. \quad (6)$$

De donde se tiene que, $\mu + \tau_i + \alpha_j + (\tau\alpha)_{ij}$ para $i = 1, 2$ y $j = 1, 2, 3$ forma un conjunto base de funciones estimables que corresponde a las expresiones:

$$\begin{aligned} &\mu + \tau_1 + \alpha_1 + (\tau\alpha)_{11} \\ &\mu + \tau_1 + \alpha_2 + (\tau\alpha)_{12} \\ &\mu + \tau_1 + \alpha_3 + (\tau\alpha)_{13} \\ &\mu + \tau_2 + \alpha_1 + (\tau\alpha)_{21} \\ &\mu + \tau_2 + \alpha_2 + (\tau\alpha)_{22} \\ &\mu + \tau_2 + \alpha_3 + (\tau\alpha)_{23} \end{aligned}$$

El mejor estimador de τ_i es $\bar{y}_{i.} - \bar{y}_{...}$ y el mejor estimador de α_j es $\bar{y}_{.j} - \bar{y}_{...}$.

El vector de parámetros utilizando funciones estimables es:

$$\hat{\beta}_e = \begin{bmatrix} 1,15415 \\ -0,22572 \\ 0,22571 \\ -0,29832 \\ 0,36791 \\ -0,0696 \\ -0,37470 \\ 0,29143 \\ 0,08339 \\ 0,37481 \\ -0,29142 \\ -0,08337 \end{bmatrix}$$

Y la respuesta estimada:

$$\hat{Y}_e = \begin{bmatrix} 0,2553 \\ \vdots \\ 0,2553 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 1,5878 \\ \vdots \\ 1,5878 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 0,9422 \\ \vdots \\ 0,9422 \\ 1,2269 \\ \vdots \\ 1,2269 \end{bmatrix}_{90 \times 12}$$

En la cual,

$$\begin{aligned} y_{11} = y_{21} = \dots = y_{151} &= 0,2554, \\ y_{161} = y_{171} = \dots = y_{301} &= 1,5878, \\ y_{311} = y_{321} = \dots = y_{451} &= 0,9422, \\ y_{461} = y_{471} = \dots = y_{751} &= 1,4564, \\ y_{761} = y_{771} = \dots = y_{901} &= 1,2269 \end{aligned}$$

El coeficiente de determinación es:

$$R^2 = \frac{133,0803 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = 0,2763$$

En este caso, los intervalos de confianza vienen dados por:

$$\tau_i: \bar{y}_{i.} - \bar{y}_{...} \pm t_{ab(m-1)}^{\frac{\alpha}{2}} \sqrt{\frac{\hat{\sigma}^2(a-1)}{abm}}, \quad i = 1, 2. \quad (7)$$

$$\begin{aligned} \alpha_j: \bar{y}_{.j} - \bar{y}_{...} \pm t_{ab(m-1)}^{\frac{\alpha}{2}} \sqrt{\frac{\hat{\sigma}^2(b-1)}{abm}}, \quad j \\ = 1, 2, 3. \quad (8) \end{aligned}$$

$$(\tau\alpha)_{ij}: \bar{y}_{ij} - \bar{y}_{...} \pm t_{ab(m-1)}^{\frac{\alpha}{2}} \sqrt{\frac{\hat{\sigma}^2(a-1)(b-1)}{abm}} \quad (9)$$

Así,

$$\begin{aligned} I(\mu) &= (1,0197; 1,2885) \\ L(I(\mu)) &= 0,2688 \\ I(\tau_1) &= (-0,3599; -0,0914) \\ L(I(\tau_1)) &= 0,2685 \\ I(\tau_2) &= (0,0914; 0,3599) \\ L(I(\tau_2)) &= 0,2685 \\ I(\alpha_1) &= (-0,4883; -0,1083) \\ L(I(\alpha_1)) &= 0,3800 \\ I(\alpha_2) &= (0,1778; 0,5579) \\ L(I(\alpha_2)) &= 0,3800 \\ I(\alpha_3) &= (-0,2596; 0,1204) \\ L(I(\alpha_3)) &= 0,3800 \\ I((\tau\alpha)_{11}) &= (-0,5647; -0,1846) \\ L(I((\tau\alpha)_{11})) &= 0,3800 \\ I((\tau\alpha)_{12}) &= (0,1014; 0,4814) \\ L(I((\tau\alpha)_{12})) &= 0,3800 \\ I((\tau\alpha)_{13}) &= (-0,1066; 0,2734) \\ L(I((\tau\alpha)_{13})) &= 0,3800 \\ I((\tau\alpha)_{21}) &= (0,1847; 0,5648) \\ L(I((\tau\alpha)_{21})) &= 0,3800 \\ I((\tau\alpha)_{22}) &= (-0,4814; -0,1014) \\ L(I((\tau\alpha)_{22})) &= 0,3800 \\ I((\tau\alpha)_{23}) &= (-0,2733; 0,1066) \\ L(I((\tau\alpha)_{23})) &= 0,3800 \end{aligned}$$

3.4. Inversa generalizada.

Se debe encontrar una estimación de β , bien sea utili-

zandola inversa generalizada, $\hat{\beta} = (X^t X)^{-1} X^t Y$ ó la inversa condicional, $\hat{\beta} = (X^t X)^c X^t Y$ de $(X^t X)$

Se calcula la inversa generalizada de la matriz $(X^t X)$. Para ello se utiliza la función "pinv" del software matemático MATLAB 7.5.0, la cual retorna la inversa generalizada de Moore-Penrose de una matriz dada. Luego:

$$\hat{\beta}_g = \begin{bmatrix} 0,5771 \\ 0,2478 \\ 0,3293 \\ -0,0065 \\ 0,2224 \\ 0,3611 \\ -0,5630 \\ 0,4090 \\ 0,4018 \\ 0,5565 \\ -0,1866 \\ -0,0406 \end{bmatrix}$$

Así,

$$\hat{Y}_g = \begin{bmatrix} 0,2553 \\ \vdots \\ 0,2553 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 1,5878 \\ \vdots \\ 1,5878 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 0,9422 \\ \vdots \\ 0,9422 \\ 1,2269 \\ \vdots \\ 1,2269 \end{bmatrix}_{90 \times 12}$$

En la cual,

$$\begin{aligned} y_{11} = y_{21} = \dots = y_{151} &= 0,2553, \\ y_{161} = y_{171} = \dots = y_{301} &= 1,4564, \\ y_{311} = y_{321} = \dots = y_{451} &= 1,5878, \\ y_{461} = y_{471} = \dots = y_{601} &= 1,4564, \\ y_{611} = y_{621} = \dots = y_{751} &= 0,9422, \\ y_{761} = y_{771} = \dots = y_{901} &= 1,2269 \end{aligned}$$

El coeficiente de determinación viene dado por:

$$R^2 = \frac{138,3178 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = 0,3860,$$

Los intervalos de confianza vienen dados por:

$$I(\beta_g)_1 = (0,5142 ; 0,6399),$$

$$\begin{aligned} L(I(\beta_g)_1) &= 0,1257 \\ I(\beta_g)_2 &= (0,1503 ; 0,3452), \\ L(I(\beta_g)_2) &= 0,1949 \\ I(\beta_g)_3 &= (0,2318 ; 0,4267), \\ L(I(\beta_g)_3) &= 0,1949 \\ I(\beta_g)_4 &= (-0,1258 ; 0,1128), \\ L(I(\beta_g)_4) &= 0,2386 \\ I(\beta_g)_5 &= (0,1030 ; 0,3417), \\ L(I(\beta_g)_5) &= 0,2387 \\ I(\beta_g)_6 &= (0,2417 ; 0,4804), \\ L(I(\beta_g)_6) &= 0,2387 \\ I(\beta_g)_7 &= (-0,7506 ; -0,3753), \\ L(I(\beta_g)_7) &= 0,3752 \\ I(\beta_g)_8 &= (0,2213 ; 0,5966), \\ L(I(\beta_g)_8) &= 0,3753 \\ I(\beta_g)_9 &= (0,2141 ; 0,5894), \\ L(I(\beta_g)_9) &= 0,3753 \\ I(\beta_g)_{10} &= (0,3688 ; 0,7441), \\ L(I(\beta_g)_{10}) &= 0,3753 \\ I(\beta_g)_{11} &= (-0,3742 ; 0,001004), \\ L(I(\beta_g)_{11}) &= 0,3752 \\ I(\beta_g)_{12} &= (-0,2282 ; 0,1470), \\ L(I(\beta_g)_{12}) &= 0,3752 \end{aligned}$$

Ahora bien, se calculará $\hat{\beta}$ utilizando la inversa condicional de $(X^t X)$. El método para encontrar esta inversa es el siguiente:

- i. Tome una submatriz de $(X^t X)$ de rango completo y orden igual al rango de $(X^t X)$ (sub matriz no singular). Llame a esa submatriz B de orden (6×6) donde $6 = \text{rang}(X)$.
- ii. Invierta B
- iii. Transponga B^{-1}
- iv. Reemplace con los elementos de B^{-1} los elementos homólogos en $(X^t X)$ y sustituir con ceros el resto para obtener la matriz C de orden (12×12) .
- v. Trasponer C para obtener la inversa condicional $(X^t X)^c$ de $(X^t X)$.

Por lo tanto,

$$\hat{\beta}_c = \begin{bmatrix} 1,2269 \\ 0,3609 \\ 0 \\ 0,2295 \\ -0,2847 \\ 0 \\ -1,5619 \\ 0,1533 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

En este caso,

$$\hat{Y}_c = \begin{bmatrix} 0,2553 \\ \vdots \\ 0,2553 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 1,5878 \\ \vdots \\ 1,5878 \\ 1,4564 \\ \vdots \\ 1,4564 \\ 0,9422 \\ \vdots \\ 0,9422 \\ 1,2269 \\ \vdots \\ 1,2269 \end{bmatrix}_{90 \times 12}$$

En la cual,

$$\begin{aligned} y_{11} = y_{21} = \dots = y_{151} &= 0,2553, \\ y_{161} = y_{171} = \dots = y_{301} &= 1,4564, \\ y_{311} = y_{321} = \dots = y_{451} &= 1,5878, \\ y_{461} = y_{471} = \dots = y_{601} &= 1,4564, \\ y_{611} = y_{621} = \dots = y_{751} &= 0,9422, \\ y_{761} = y_{771} = \dots = y_{901} &= 1,2269 \end{aligned}$$

El coeficiente de determinación viene dado por:

$$R^2 = \frac{138,3178 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = 0,3860,$$

Los intervalos de confianza son:

$$\begin{aligned} I(\beta_c)_1 &= (0,9240 ; 1,5297), \\ L(I(\beta_c)_1) &= 0,6057 \\ I(\beta_c)_2 &= (-0,0679 ; 0,7897), \\ L(I(\beta_c)_2) &= 0,8576 \\ I(\beta_c)_3 &= 0 \\ I(\beta_c)_4 &= (-0,1993 ; 0,6583), \\ L(I(\beta_c)_4) &= 0,8576 \\ I(\beta_c)_5 &= (-0,7135 ; 0,1441), \end{aligned}$$

$$\begin{aligned} L(I(\beta_c)_5) &= 0,8576 \\ I(\beta_c)_6 &= 0 \\ I(\beta_c)_7 &= (-2,1683 ; -0,9554), \\ L(I(\beta_c)_7) &= 1,2128 \\ I(\beta_c)_8 &= (-0,4531 ; 0,7597), \\ L(I(\beta_c)_8) &= 1,2128 \\ I(\beta_c)_9 &= 0 \\ I(\beta_c)_{10} &= 0 \\ I(\beta_c)_{11} &= 0 \\ I(\beta_c)_{12} &= 0 \end{aligned}$$

4 Conclusiones

Las ventajas que se pueden señalar a lo largo del trabajo son las siguientes:

1. Los métodos mínimos cuadrados y máxima verosimilitud no garantizan estimaciones insesgadas para el estimador σ^2 .
2. Consideran todas las ecuaciones del sistema de ecuaciones normales y no usa funciones lineales de los parámetros no estimables.
3. Se pueden utilizar para dar solución al sistema de ecuaciones normales de un modelo de diseño de experimentos sin realizar restricciones sobre los parámetros, ya que al utilizar restricciones sobre los parámetros se pueden obtener diferentes resultados, es decir, si dos investigadores aplican técnicas diferentes de restricción de parámetros para un mismo problema, estos pueden llegar a diferentes conclusiones.
4. Reducen al mínimo la suma de cuadrados del error.
5. Permiten caracterizar las soluciones de los sistemas consistentes y en los sistemas inconsistentes permite hallar soluciones aproximadas.
6. La longitud de los intervalos de confianza de los parámetros estimados es más pequeña, es decir, permite una estimación más precisa.

Como desventaja del uso de la inversa generalizada y/o condicional se puede mencionar en el caso de la condicional que no es única y en el caso de la generalizada sería más que todo el cálculo que fue su problema original, sin embargo hoy en día con ayuda de los software Matlab, NumPy y SciPy se puede resolver este problema.

Referencias

Bose RC, 1959, Analysis of Variance. Unpublished lecture notes, University of North Carolina.
 GraybillFA, 1976, Theory and application of the linear model. Duxbury press. Massachusetts.
 Searle SR, 1997, --Linear Models. Wiley Classics Library. New York: John Wiley & Sons, Inc.

Recibido: 30 de julio de 2017

Aceptado: 01 de marzo de 2018

Rivas, Felvir: MSc en Estadística (UCV). Licenciada en Matemáticas (ULA). Docente Dedicación Exclusiva UNE-LLEZ-Barinas.