

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE AGRONOMÍA
COMISIÓN DE ESTUDIOS DE POSTGRADO
POSTGRADO EN ESTADÍSTICA

**ESTIMACIÓN DE PARÁMETROS EN EL MODELO ESTADÍSTICO LINEAL DE
RANGO INCOMPLETO A TRAVÉS DE LA INVERSA GENERALIZADA:
VENTAJAS, DESVENTAJAS Y ANÁLISIS.**

Lic. Felvir Rivas.
Tutor: Dr. Miguel Balza.

TRABAJO DE GRADO PRESENTADO ANTE LOS MIEMBROS DEL COMITÉ ACADÉMICO DEL POSTGRADO EN ESTADÍSTICA DE LA FACULTAD DE AGRONOMÍA DE LA UNIVERSIDAD CENTRAL DE VENEZUELA COMO ÚLTIMO REQUISITO PARA OPTAR AL TÍTULO DE MAGISTER SCIENTIARUM EN ESTADÍSTICA.

Índice general.

Introducción	1
Resumen	4
Abstract	6
1. Planteamiento del Problema	
1.1. Objetivos	9
1.1.1. Objetivo general	9
1.1.2. Objetivos específicos	10
1.2. Justificación	10
1.3. Antecedentes	12
2. Conceptos básicos	
2.1. Sistema de ecuaciones lineales	17
2.2. Inversa generalizada de matrices	22
2.3. Inversa condicional de matrices	45
3. Modelo lineal	
3.1. Planteamiento del modelo lineal	57
3.2. Modelo lineal general de rango completo	62
3.3. Modelo lineal general de rango incompleto	73

4. Coeficiente de Determinación

4.1. Definiciones básicas	85
---------------------------------	----

5. Resultados

5.1. Problema	90
---------------------	----

5.2. Modelo reducido	95
----------------------------	----

5.3. Imponiendo restricciones	101
-------------------------------------	-----

5.3.1. Funciones estimables	107
-----------------------------------	-----

5.4. Inversa generalizada e Inversa condicional	114
---	-----

Ventajas y desventajas del uso de la inversa generalizada y/o condicional	127
--	------------

Conclusiones	129
---------------------------	------------

Bibliografía	130
---------------------------	------------

Introducción.

El problema de estimación en un modelo lineal de rango incompleto puede ser abordado mediante el concepto de estimabilidad de R. C. Bose (1959), concepto que es discutido con cierto detalle por F. 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 $p - \text{rang}(X)$.
- 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.

El siguiente trabajo está dividido en capítulos, estructurados de la siguiente manera:

El Capítulo 1 de este trabajo está dedicado a los preliminares como: objetivos, justificación y antecedentes.

El Capítulo 2 se trata todo lo relacionado con sistemas de ecuaciones lineales, inversa generalizada e inversa condicional, que serán de gran utilidad para la comprensión del trabajo.

El Capítulo 3 se define el modelo lineal, tanto de rango completo como de rango incompleto, tratando el proceso de estimación de parámetros.

El Capítulo 4 la teoría necesaria para obtener el coeficiente de determinación del modelo.

El Capítulo 5 la aplicación y los resultados con un conjunto de datos reales.

Resumen.

En un modelo de rango incompleto la matriz $X^t X$ 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 uniforms 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.

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

Capítulo 1.

Planteamiento del problema.

En las diferentes áreas de investigación, muchas veces surge la necesidad de hacer un estudio conjunto de varias características de una población, con el fin de intentar explicar el comportamiento de una de ellas en función de las otras. Este tipo de análisis generalmente se hace utilizando los modelos estadísticos lineales y en particular los modelos de regresión y de análisis de varianza.

Los modelos estadísticos lineales pueden clasificarse en dos tipos: modelos de rango completo y de rango incompleto o de rango deficiente (Kirk, 1982). En los modelos de rango completo las columnas de la matriz de coeficientes o de diseño son linealmente independientes y en los modelos de rango incompleto son linealmente dependientes.

Los modelos de rango completo, incluyen a los modelos formalizados en términos de medias, en éstos se asume que el valor esperado de la perturbación del i -ésimo sujeto en el j -ésimo tratamiento coincide con la media de su grupo. Es bien sabido, que la formalización de los diseños experimentales en términos de medias parece ser su representación más natural, sin embargo, la formalización más corriente la constituyen los

modelos de rango incompleto (Dunn & Clark, 1987), (Milliken & Johnson, 1984), (Searle, 1997), (Timm, 1975).

En los modelos de rango incompleto el número de parámetros a estimar es superior al número de ecuaciones distintas no colineales que pueden ser planteadas, es decir el modelo contiene más parámetros que los que pueden ser únicamente estimados a partir de los datos registrados. Esto se deriva a la falta de independencia lineal entre las columnas de la matriz de diseño. Técnicamente una inversa generalizada puede ser utilizada para resolver este problema, (Graybill, 1976). Pero el método más empleado es reparametrizar el modelo, agregando al sistema nuevas ecuaciones lineales de los parámetros, obteniéndose un modelo lineal de rango completo, sin embargo en este modelo las ecuaciones que se agregan tienen la particularidad de no ser estimables.

1.1. OBJETIVOS.

1.1.1. Objetivo general.

Comparar los parámetros estimados en el modelo estadístico lineal de rango incompleto por medio de la inversa generalizada con los estimados por vía mínimos cuadrados y máxima verosimilitud.

1.1.2. Objetivos específicos.

- i. Analizar la resolución de sistemas de ecuaciones lineales consistentes, vía inversa generalizada.
- ii. Estudiar las ventajas y desventajas de la estimación de los parámetros en el modelo estadístico lineal de rango incompleto a través de la inversa generalizada.
- iii. Comparar las soluciones obtenidas vía inversa generalizada con los métodos de estimación clásicos: método de mínimos cuadrados y método de máxima verosimilitud.
- iv. Analizar las inferencias estadísticas resultantes por ambas vías de estimación.

1.2. JUSTIFICACIÓN.

Los modelos lineales han constituido durante décadas la herramienta usada en identificación para control, validación y predicción. La aplicación de su éxito es doble: por un lado, existe una teoría matemática bien establecida que ha permitido el avance en la solución de los problemas prácticos y por otro lado, la aplicación a sistemas reales ha dado buenos resultados pues en muchos casos tienen un ajuste adecuado a estos modelos. Constituyen una de las metodologías estadísticas más ampliamente utilizadas en la modelización y el análisis de datos de todo tipo. Se introducen en campos tan diversos

como la experimentación industrial, la construcción y validación de tests psicológicos o el análisis de datos de chips de ADN de la moderna era post-genómica.

Cuando el modelo lineal corresponde al análisis de los datos a través de un Modelo de Diseño Experimental, la matriz de coeficientes X en el modelo, está formada por los valores 0 ó 1, representando el 0 la ausencia de un factor o su respectivo nivel y 1 la presencia, y sus columnas suelen ser linealmente dependientes. Se sabe que en este caso es posible hallar el estimador mínimo cuadrático de β , pero hay múltiples estimaciones de los parámetros β que más bien podemos considerar como soluciones $\hat{\beta}$ de las ecuaciones normales $X^t X \hat{\beta} = X^t Y$, debido a que la matriz $X^t X$ es de rango incompleto, en general.

Para dar solución a dicho problema se recurre a ciertas técnicas, las más comunes son aquellas donde se restringen ciertos parámetros o suma de parámetros del sistema $X^t X \hat{\beta} = X^t Y$ y se igualan a cero. Para ello, se agregan un conjunto de ecuaciones lineales de los parámetros “no estimables” para formar un sistema de ecuaciones con matriz de coeficientes de rango completo y así producir soluciones únicas. Por ejemplo, en el caso de modelos de diseño se agregan ecuaciones como $\tau_1 + \tau_2 + \dots + \tau_a = 0$ y $\alpha_1 + \alpha_2 + \dots + \alpha_b = 0$, en el caso de un modelo de diseño con dos factores y estas ecuaciones no son estimables ya que no son contrastes.

Es por ello que esta investigación persigue encontrar la estimación de parámetros en el modelo estadístico lineal de rango incompleto agregando dichas ecuaciones, también utilizando la inversa generalizada e inversa condicional para luego estudiar las ventajas y desventajas al comparar con los casos indicados arriba.

1.3. ANTECEDENTES.

Los modelos lineales de rango incompleto son muy comunes en econometría. En algunos modelos los parámetros de rango incompleto es una parte natural del modelo, en otros modelos se restringe el rango con el fin de reducir el número de parámetros. Es fácil convertir un modelo de rango incompleto en un modelo de rango completo, expresándolo como producto de dos matrices, pues en éste la estimación de los parámetros es un problema de lo que se conoce como regresión de rango reducido. La regresión de rango reducido está estrechamente relacionada con el análisis de covariables y las correlaciones canónicas de Hostelling (1935) y Anderson (1951, 1984). Los modelos de rango reducidos para variables fijas han sido analizados por Velu, Reinsel y Wichern (1986) y Velu y Reinsel (1987). Gudmundsson (1977) utilizó regresión de rango reducido para seleccionar las combinaciones lineales que mejor describen la variación media de un vector de variables dependientes.

La regresión de rango reducido tiene un impacto importante en el análisis de regresión lineal múltiple multivariante. Anderson (1951), estimó los coeficientes de regresión con restricciones lineales para matrices de rango menor, dando paso a la regresión multivariante con rango reducido, la cual ha sido aplicada en muchas disciplinas. Pero no fue sino hasta 1975 que Izenman (1975), discutió este método, regresión multivariante, para conjuntos de variables aleatorias, demostrando su relación con el análisis de componentes principales y dando los primeros pasos en la relación con el análisis de correlación canónica. También estudió, bajo el supuesto de normalidad, la distribución asintótica de varias matrices de coeficientes para la regresión con rango reducido. Davies y Tso (1982), describieron el procedimiento para el ajuste del modelo de regresión con rango reducido.

Anderson (1999), comparó la distribución asintótica de la matriz de coeficientes con rango completo y la matriz de coeficientes con rango reducido, concluyendo que en general, la estimación del modelo con rango reducido es mejor que la del modelo de rango completo. No obstante, el mismo autor tres años más tarde (2002) determina la influencia de la especificación del rango en ambos modelos, encontrando que el estimador del modelo con rango reducido es más eficiente cuando dicho rango es pequeño.

Álvarez Willin (2009), propone un procedimiento alternativo de estimación para los coeficientes en un modelo de regresión multivariante con rango reducido bajo la existencia de multicolinealidad, aprovechando las bondades del método de mínimos cuadrados parciales y la descomposición en valores singulares para la predicción de las variables dependientes

Por otro lado, cuando se trabaja con un modelo lineal de rango incompleto, se sabe que la inversa generalizada puede ser utilizada para resolver las ecuaciones normales y así reducir al mínimo la suma de cuadrados del error. Este concepto fue introducido por primera vez por Moore (1920). Extensiones de estas ideas las realizó Tseng (1949, 1956), pero ningún estudio sistemático del tema se hizo hasta 1955, cuando Penrose (1955), sin darse cuenta de la labor anterior, redefinió la inversa de Moore en una forma ligeramente diferente.

En el mismo año Rao (1955), dio un método de cálculo de lo que se conoce como pseudoinversa de la matriz singular, y lo aplicó para resolver las ecuaciones normales de una matriz singular en la teoría de mínimos cuadrados y para expresar las varianzas de los estimadores. La pseudoinversa definida por Rao no satisface todas las restricciones impuestas por Moore (1920) y Penrose (1955). Por lo tanto, difiere de la inversa de Moore-Penrose,

pero ha sido útil para proporcionar una teoría general de estimación de mínimos cuadrados sin cualquier restricción sobre el rango de la matriz de datos. En un artículo posterior, Rao (1962) mostró que una inversa con una definición mucho más débil que la de Moore y Penrose es suficiente para hacer frente a problemas de ecuaciones lineales. Desde 1955 se realizaron numerosos aportes en cuanto al tema, entre ellos Greville (1959), Bjerhammer (1958), Ben-Israel y Charnes (1963). Bose (1959) menciona el uso de la inversa generalizada en sus notas, "Análisis de la Varianza".

La inversa generalizada satisface la definición más débil propuesta por Rao y no es la única. Por lo tanto presenta un interesante estudio en el álgebra matricial. En una publicación en 1967, Rao (1967), mostró una gran variedad de inversas generalizadas, las cuales podrían adaptarse a diferentes propósitos y presentar una clasificación de inversas generalizadas.

Marquardt (1970), discute una clase de estimadores lineales sesgados, empleando inversa generalizada. Además muestra una parte de propiedades teóricas de estimadores por inversa generalizada, regresión ridge y procedimientos correspondientes a estimación no lineal.

Navarro, Youngquis y Compton (1992), estimaron las varianzas genéticas de maíz a partir de líneas S1 y S2, en dos localidades de E.E.U.U, utilizando el método de la matriz inversa generalizada, cuando las covarianzas usadas daban una matriz cuadrada no singular. Concluyeron que el uso de la matriz inversa generalizada de Moore-Penrose puede representar un método adecuado para la estimación de los parámetros de variabilidad genética, al menos de las varianzas aditivas y de dominancia, cuando se dispone de un juego de covarianzas de valores consistentes entre ellos mismos.

Capítulo 2.

Conceptos Básicos.

2.1. SISTEMAS DE ECUACIONES LINEALES

Un sistema de ecuaciones relaciona varias variables desconocidas denominadas incógnitas mediante una serie de condiciones expresadas por igualdades. Cada una de esas igualdades recibe el nombre de ecuación y el conjunto de todas ellas se conoce con el nombre de sistemas de ecuaciones. Cuando la ecuación es un polinomio de primer grado en sus incógnitas, se dice que la ecuación es lineal. Si todas las ecuaciones de un sistema son lineales, éste recibe el nombre de sistema de ecuaciones lineales.

Los sistemas de ecuaciones con solución se denominan consistentes o compatibles; los que no la tienen son inconsistentes o incompatibles. Un sistema consistente con un número finito de soluciones es un sistema determinado; si posee infinitas soluciones, el sistema es indeterminado.

En esta sección se estudian los sistemas de ecuaciones lineales desde la perspectiva del cálculo matricial (Graybill, 1976).

Si se hace

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Se puede escribir entonces el sistema anterior en forma matricial

$$\mathbf{Ax} = \mathbf{b}, \quad (2.1.1)$$

donde, \mathbf{A} es una matriz real $m \times n$, \mathbf{b} es un vector real $m \times 1$ y \mathbf{x} es un vector $n \times 1$. El objetivo es:

1. Saber cuándo el sistema tiene solución y cuándo no. Dicho de otra manera: para una matriz $\mathbf{A}_{m \times n}$ y un vector $\mathbf{b}_{m \times 1}$, existe un vector $\mathbf{x}_{n \times 1}$ que satisfaga la ecuación (2.1.1)?
2. Si tiene solución, saber cuántas soluciones tiene. Dicho de otra manera: ¿cuántas soluciones tiene el vector \mathbf{x} y cuántas de ellas dependen de \mathbf{A} y \mathbf{b} ?

Para resolver cada una de estas interrogantes se discutirán en la siguiente sección algunos teoremas concernientes a la existencia de soluciones de un sistema de ecuaciones lineales.

Soluciones de ecuaciones lineales

En esta sección, como se mencionó anteriormente, se discutirán algunos teoremas relacionados a la existencia de una solución del sistema $\mathbf{Ax} = \mathbf{b}$.

Teorema 2.1.1.

Si $\mathbf{A}_{n \times n}$ es una matriz no singular, entonces el sistema $\mathbf{Ax} = \mathbf{b}$ tiene una única solución.

Demostración.

Si $\mathbf{A}_{n \times n}$ es una matriz no singular, entonces existe \mathbf{A}^{-1} . Así, de la ecuación (2.1.1) se tiene

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b}$$

$$\mathbf{Ix} = \mathbf{A}^{-1}\mathbf{b}$$

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Es solución del sistema $\mathbf{Ax} = \mathbf{b}$.

□

En el sistema de ecuaciones $\mathbf{Ax} = \mathbf{b}$, la matriz \mathbf{A} es llamada matriz de coeficientes y la matriz $\mathbf{B} = [\mathbf{A}, \mathbf{b}]$ es llamada matriz aumentada, donde \mathbf{B} es la matriz que se forma de añadir a la matriz \mathbf{A} el vector columna \mathbf{b} . La matriz aumentada y de coeficiente pueden

ser usadas para determinar si existe solución en un sistema de ecuaciones lineales, tal como lo señala el siguiente teorema.

Teorema 2.1.2.

El sistema $\mathbf{Ax} = \mathbf{b}$ tiene solución si, y sólo si, el rango de \mathbf{A} es igual al rango de $\mathbf{B} = [\mathbf{A}, \mathbf{b}]$.

Demostración.

Llamemos \mathbf{A}_j a la columna j -ésima de la matriz de coeficientes, es decir,

$$\mathbf{A}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \quad \dots, \quad \mathbf{A}_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

El sistema lineal puede expresarse como

$$\mathbf{A}_1x_1 + \mathbf{A}_2x_2 + \dots + \mathbf{A}_nx_n = \mathbf{b},$$

Se expresa el vector de términos como una combinación lineal de las n columnas de matriz \mathbf{A} , cuyos coeficientes son las incógnitas. Así, se tiene que el sistema sólo admitirá solución cuando y sólo cuando \mathbf{b} sea combinación lineal de las columnas de \mathbf{A} .

Sea r el rango de la matriz de coeficientes A . Si el sistema tiene solución, entonces b puede escribirse como combinación lineal de A_1, A_2, \dots, A_n . Por lo tanto, el rango de B es r , así el rango de A es igual al rango de $B = [A, b]$.

Recíprocamente, si el rango de B , igual al rango de A , es r se tiene que b podrá escribirse como combinación lineal de las r columnas A_j linealmente independientes. Supóngase que sean A_1, A_2, \dots, A_r , luego existen $(s_1, s_2, \dots, s_r, 0, \dots, 0)$ soluciones tales que $A_1s_1 + A_2s_2 + \dots + A_rs_r + A_{r+1}0 + \dots + A_n0 = b$, y por lo tanto, el sistema tiene solución.

□

2.2. INVERSA GENERALIZADA DE MATRICES.

El concepto de inversa generalizada tiene sus principios en la teoría de ecuaciones lineales. Un conjunto de ecuaciones lineales consistente $Ax = b$ puede ser de solución única, si el rango de la matriz A es $r = m = n$. Sin embargo, cuando la matriz A es una matriz rectangular o singular, una representación simple de una solución en términos de A es más complicado. En esta sección se tratan esos sistemas de ecuaciones usando las inversas generalizadas de matrices.

Definición 2.2.1.

Sea \mathbf{A} una matriz de orden $m \times n$, se dice que \mathbf{A}^- es una inversa generalizada de \mathbf{A} , si satisface las siguientes condiciones:

- i. $\mathbf{AA}^- \mathbf{A} = \mathbf{A}$,
- ii. $\mathbf{A}^- \mathbf{AA}^- = \mathbf{A}^-$,
- iii. \mathbf{AA}^- es simétrica y
- iv. $\mathbf{A}^- \mathbf{A}$ es simétrica.

Si \mathbf{A} es una matriz no singular se tiene que \mathbf{A}^{-1} satisface las condiciones de la definición anterior. Además, para cada matriz \mathbf{A} existe \mathbf{A}^- y es única. Estos resultados se probarán más adelante primero se verán algunas propiedades de la matriz inversa generalizada.

Teorema 2.2.1.

Si la inversa generalizada de una matriz $\mathbf{A}_{m \times n}$ existe, entonces \mathbf{A}^- es de orden $n \times m$.

Demostración.

Si \mathbf{A}^- existe, entonces \mathbf{A}^- satisface las condiciones de la definición 2.2.1. Luego \mathbf{AA}^- es una matriz cuadrada, pues es simétrica. Así \mathbf{A}^- es de orden $n \times m$.

□

Teorema 2.2.2.

Si A es una matriz nula, entonces A^- es una matriz nula.

Demostración.

Si $A = 0$, de orden $m \times n$, es fácil ver que $A^- = 0$, de orden $n \times m$, satisface las condiciones de la definición 2.2.1.

□

Teorema 2.2.3.

Para cada matriz A existe una matriz A^- que satisface las condiciones de la definición 2.2.1. Esto es, cada matriz tiene una inversa generalizada.

Demostración.

Si $A = 0$, entonces por el teorema 2.2.2, se tiene que A^- existe y además $A^- = 0$.

Supóngase ahora que $A \neq 0$ y que tiene rango $r > 0$. Luego existen K y L de tamaño $m \times r$ y $r \times n$, respectivamente, ambas de rango r tales que

$$A = KL \tag{2.2.1}$$

Como K^-K y LL^- son matrices cuadradas de rango r se tiene que ambas son no singular, así se puede definir

$$A^- = L^t(LL^t)^{-1}(K^tK)^{-1}K^t \tag{2.2.2}$$

La cual es una inversa generalizada de A , ya que

- a. $AA^{-}A = KLL^t(LL^t)^{-1}(K^tK)^{-1}K^tKL = KL = A$
- b. $A^{-}AA^{-} = L^t(LL^t)^{-1}(K^tK)^{-1}K^tKLL^t(LL^t)^{-1}(K^tK)^{-1}K^t =$
 $L^t(LL^t)^{-1}(K^tK)^{-1}K^t = A^{-}$
- c. $AA^{-} = KLL^t(LL^t)^{-1}(K^tK)^{-1}K^t = K(K^tK)^{-1}K^t$, es simétrica, pues
 $(K(K^tK)^{-1}K^t)^t = K(K^{-1}(K^t)^{-1})^tK^t = KK^{-1}(K^t)^{-1}K^t = K(K^tK)^{-1}K^t$
- d. $A^{-}A = L^t(LL^t)^{-1}(K^tK)^{-1}K^tKL = L^t(LL^t)^{-1}L$, es simétrica, pues
 $(L^t(LL^t)^{-1}L)^t = L^t((L^t)^{-1}L^{-1})^tL = L^t(L^t)^{-1}L^{-1}L = L^t(LL^t)^{-1}L$

Así siempre existe una inversa generalizada de una matriz A .

□

Teorema 2.2.4.

Para cada matriz A existe una única matriz A^{-} que satisface las condiciones de la definición 2.2.1. Esto es, cada matriz tiene una única inversa generalizada.

Demostración.

Supóngase que A_1^{-} y A_2^{-} son dos matrices inversas generalizadas de la matriz A . Se probará que $A_1^{-} = A_2^{-}$. Cada una satisface las condiciones de la definición 2.2.1.

Por ser A_1^{-} inversa generalizada de A se tiene que

$$AA_1^{-}A = A \quad (2.2.3)$$

$$\therefore AA_1^{-}AA_2^{-} = AA_2^{-}$$

Ya que AA_2^{-} y AA_1^{-} son simétricas, se tiene que

$$\begin{aligned} \mathbf{AA}_2^- &= (\mathbf{AA}_2^-)^t = [(\mathbf{AA}_1^-)(\mathbf{AA}_2^-)]^t = (\mathbf{AA}_2^-)^t(\mathbf{AA}_1^-)^t = \mathbf{AA}_2^- \mathbf{AA}_1^- \\ &= \mathbf{AA}_1^- \end{aligned} \quad (2.2.4)$$

Ahora se multiplica la ecuación (2.2.3) por \mathbf{A}_2^- a la izquierda,

$$\mathbf{A}_2^- \mathbf{AA}_1^- \mathbf{A} = \mathbf{A}_2^- \mathbf{A}$$

Dada la simetría de $\mathbf{A}_2^- \mathbf{A}$ y $\mathbf{A}_1^- \mathbf{A}$, se tiene

$$\begin{aligned} \mathbf{A}_2^- \mathbf{A} &= (\mathbf{A}_2^- \mathbf{A})^t = [(\mathbf{A}_2^- \mathbf{A})(\mathbf{A}_1^- \mathbf{A})]^t = (\mathbf{A}_1^- \mathbf{A})^t (\mathbf{A}_2^- \mathbf{A})^t = \mathbf{A}_1^- \mathbf{AA}_2^- \mathbf{A} \\ &= \mathbf{A}_1^- \mathbf{A} \end{aligned} \quad (2.2.5)$$

Por último, multiplicando la ecuación (2.2.4) por \mathbf{A}_2^- a la izquierda, resulta

$$\mathbf{A}_2^- \mathbf{AA}_2^- = \mathbf{A}_2^- \mathbf{AA}_1^-$$

Así, de acuerdo con la ecuación (2.2.5) y la definición 2.2.1 se obtiene

$$\mathbf{A}_1^- = \mathbf{A}_1^- \mathbf{AA}_1^- = (\mathbf{A}_1^- \mathbf{A}) \mathbf{A}_1^- = (\mathbf{A}_2^- \mathbf{A}) \mathbf{A}_1^- = \mathbf{A}_2^- \mathbf{AA}_2^- = \mathbf{A}_2^-.$$

Por lo tanto, \mathbf{A}^- es única.

□

Teorema 2.2.5.

La matriz inversa generalizada de una matriz transpuesta es la transpuesta de la matriz inversa generalizada, es decir, $(\mathbf{A}^t)^- = (\mathbf{A}^-)^t$.

Demostración.

Se Probará que $(\mathbf{A}^-)^t$ es la matriz inversa generalizada de \mathbf{A}^t y como ésta es única se tiene que $(\mathbf{A}^t)^- = (\mathbf{A}^-)^t$.

Llámesese $A = KL$ como en la ecuación (2.2.1) y sea

$A^- = L^t(LL^t)^{-1}(K^tK)^{-1}K^t$ ya definida en (2.2.2). Luego,

$$(A^t)^- = (K^t)^t[K^t(K^t)^t]^{-1}[(L^t)^tL^t]^{-1}(L^t)^t = K(K^tK)^{-1}(LL^t)^{-1}L \quad (2.2.6)$$

pues, $A^t = L^tK^t$.

Por otra parte, de la ecuación (2.2.2) resulta

$$\begin{aligned} (A^-)^t &= [L^t(LL^t)^{-1}(K^tK)^{-1}K^t]^t = (K^t)^t[(K^tK)^t]^{-1}[(LL^t)^t]^{-1}(L^t)^t \\ &= K(K^tK)^{-1}(LL^t)^{-1}L \end{aligned} \quad (2.2.7)$$

Así, de las ecuaciones (2.2.6) y (2.2.7) se tiene que $(A^t)^- = (A^-)^t$.

□

Teorema 2.2.6.

La matriz inversa generalizada de A^- es igual a A , es decir, $(A^-)^- = A$.

Demostración.

Sea $(A^-)^-$ la matriz inversa generalizada de A^- , luego ésta satisface

- i. $A^-(A^-)^-A^- = A^-$
- ii. $(A^-)^-A^-(A^-)^- = (A^-)^-$
- iii. $A^-(A^-)^- = [A^-(A^-)^-]^t$
- iv. $(A^-)^-A^- = [(A^-)^-A^-]^t$

Ahora sustituyendo \mathbf{A} por $(\mathbf{A}^-)^-$, se tiene que las ecuaciones anteriores son iguales a las de la definición 2.2.1, y como la inversa generalizada de una matriz es única (teorema 2.2.4), se cumple que $(\mathbf{A}^-)^- = \mathbf{A}$.

□

Teorema 2.2.7.

El rango de una matriz inversa generalizada de \mathbf{A} es igual al rango de \mathbf{A} .

Demostración.

Sea \mathbf{A}^- la inversa generalizada de la matriz \mathbf{A} . Luego $\mathbf{AA}^-\mathbf{A} = \mathbf{A}$, así

$$\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{AA}^-\mathbf{A}) \leq \text{rang}(\mathbf{A}^-) \quad (2.2.8)$$

Por otra parte, ya que $\mathbf{A}^-\mathbf{AA}^- = \mathbf{A}^-$, se tiene

$$\text{rang}(\mathbf{A}^-) = \text{rang}(\mathbf{A}^-\mathbf{AA}^-) \leq \text{rang}(\mathbf{A}) \quad (2.2.9)$$

Por lo tanto, de las ecuaciones (2.2.8) y (2.2.9) se obtiene que $\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A}^-)$.

□

Corolario 2.2.1.

Si el rango de una matriz \mathbf{A} es r , entonces el rango de cada una de las matrices \mathbf{A}^- , \mathbf{AA}^- , $\mathbf{A}^-\mathbf{A}$, $\mathbf{AA}^-\mathbf{A}$ y $\mathbf{A}^-\mathbf{AA}^-$ es r .

Demostración.

Sea \mathbf{A} una matriz de rango r y \mathbf{A}^- la inversa generalizada de \mathbf{A} . Luego,

- i. Del teorema 2.2.7 se tiene que $\text{rang}(\mathbf{A}^-) = \text{rang}(\mathbf{A}) = r$
- ii. $\text{rang}(\mathbf{A}\mathbf{A}^-) \leq \text{rang}(\mathbf{A}^-) = r$ y $r = \text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A}\mathbf{A}^-\mathbf{A}) \leq \text{rang}(\mathbf{A}\mathbf{A}^-)$.
Así, $r \leq \text{rang}(\mathbf{A}\mathbf{A}^-) \leq r \therefore \text{rang}(\mathbf{A}\mathbf{A}^-) = r$.
- iii. $\text{rang}(\mathbf{A}^-\mathbf{A}) \leq \text{rang}(\mathbf{A}^-) = r$ y $r = \text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A}^-\mathbf{A}\mathbf{A}^-) \leq \text{rang}(\mathbf{A}^-\mathbf{A})$.
Así, $r \leq \text{rang}(\mathbf{A}^-\mathbf{A}) \leq r \therefore \text{rang}(\mathbf{A}^-\mathbf{A}) = r$.
- iv. $\text{rang}(\mathbf{A}\mathbf{A}^-\mathbf{A}) = \text{rang}(\mathbf{A}) = r$, por definición de \mathbf{A}^- .
- v. $\text{rang}(\mathbf{A}^-\mathbf{A}\mathbf{A}^-) = \text{rang}(\mathbf{A}^-) = r$, por definición de \mathbf{A}^- .

□

Teorema 2.2.8.

Si \mathbf{A} es una matriz simétrica, entonces \mathbf{A}^- también es una matriz simétrica.

Demostración.

Sea \mathbf{A} una matriz simétrica, luego $\mathbf{A} = \mathbf{A}^t$. Del teorema 2.2.5 se sabe que $(\mathbf{A}^t)^- = (\mathbf{A}^-)^t$. Por lo tanto, $(\mathbf{A}^-)^t = (\mathbf{A}^t)^- = \mathbf{A}^-$.

□

Teorema 2.2.9.

Si \mathbf{A} es una matriz no singular, entonces $\mathbf{A}^- = \mathbf{A}^{-1}$.

Demostración.

Supóngase que A es una matriz no singular, entonces existe A^{-1} la cual cumple con $AA^{-1} = A^{-1}A = I_d$, donde I_d es la matriz identidad.

Sea A^- la matriz generalizada de A , luego de la parte i. de la definición 2.2.1 se tiene:

$$\begin{aligned} AA^-A &= A \\ \therefore A^{-1}AA^-A &= A^{-1}A = I_d \\ \therefore A^-A &= I_d \end{aligned} \tag{2.2.10}$$

También se tiene que,

$$\begin{aligned} AA^-AA^{-1} &= AA^{-1} \\ \therefore AA^- &= I_d \end{aligned} \tag{2.2.11}$$

Por otra parte, como $AA^{-1} = I_d$ se tiene que $AA^{-1}A = A$, y de $A^{-1}A = I_d$ que $A^{-1}AA^{-1} = A^{-1}$. Además,

$$(A^{-1}A)^t = A^t(A^{-1})^t = A^t(A^t)^{-1} = I_d^t = (A^{-1}A)^t$$

$$(AA^{-1})^t = (A^{-1})A^{t^t} = (A^t)^{-1}A^t = I_d^t = (AA^{-1})^t$$

Por lo tanto, A^{-1} satisface las condiciones de la definición 2.2.1, así A^{-1} es una inversa generalizada de la matriz A .

Luego de las ecuaciones (2.2.10) y (2.2.11) se tiene que $A^- = A^{-1}$.

□

Teorema 2.2.10.

Si A es una matriz simétrica e idempotente, entonces $A^- = A$.

Demostración.

Sea A es una matriz simétrica idempotente, luego $A = A^t$ y $A = A^2$. Si A^- es la inversa generalizada de la matriz A , se sabe que ésta cumple con las condiciones de la definición 2.2.1.

Por otra parte, se tiene que A también satisface las condiciones, pues

- a. $AAA = AA = A$ (condición i y ii)
- b. $AA = A$ es simétrica (condición iii y iv)

Así, del teorema 2.2.4 se obtiene que $A^- = A$.

□

Teorema 2.2.11.

Sea D una matriz diagonal, cuyos elementos en la diagonal vienen dados por d_{ii} , $i = 1, 2, \dots, n$. La inversa generalizada D^- de D es una matriz diagonal cuyos elementos son d_{ii}^{-1} , si $d_{ii} \neq 0$ y son ceros si $d_{ii} = 0$.

Demostración.

Supóngase que \mathbf{D} sea una matriz diagonal y sea \mathbf{D}^- una matriz diagonal cuyos elementos son d_{ii}^{-1} , si $d_{ii} \neq 0$ y son ceros si $d_{ii} = 0$. Se verá que \mathbf{D}^- es la inversa generalizada de \mathbf{D} .

Si $\mathbf{D} = \text{diag}(d_{11}, d_{22}, \dots, d_{nn})$ y $\mathbf{D}^- = \text{diag}(d_{11}^{-1}, d_{22}^{-1}, \dots, d_{nn}^{-1})$, se tiene que $\mathbf{D}^- \mathbf{D} = \text{diag}(a_{11}, a_{22}, \dots, a_{nn})$ y $\mathbf{D} \mathbf{D}^- = \text{diag}(b_{11}, b_{22}, \dots, b_{nn})$ son matrices diagonales cuyos elementos son unos y ceros. Si $d_{ii} = 0$ se tiene que $d_{ii} = a_{ii} = b_{ii} = 0$ y si $d_{ii} \neq 0$, $d_{ii} d_{ii}^{-1} = d_{ii}^{-1} d_{ii} = a_{ii} = b_{ii} = 1$. Por lo tanto se cumple que:

- a. $\mathbf{D} \mathbf{D}^- \mathbf{D} = \mathbf{D}$
- b. $\mathbf{D}^- \mathbf{D} \mathbf{D}^- = \mathbf{D}^-$
- c. $\mathbf{D}^- \mathbf{D}$ sea simétrica, pues es una matriz diagonal
- d. $\mathbf{D} \mathbf{D}^-$ sea simétrica, pues es una matriz diagonal

Así \mathbf{D}^- es la inversa generalizada de \mathbf{D} .

□

Teorema 2.2.12.

Si \mathbf{A} es una matriz $m \times n$ de rango m , entonces $\mathbf{A}^- = \mathbf{A}^t (\mathbf{A} \mathbf{A}^t)^{-1}$ y $\mathbf{A} \mathbf{A}^- = \mathbf{I}_d$. Si \mathbf{A} es de rango n , entonces $\mathbf{A}^- = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t$ y $\mathbf{A}^- \mathbf{A} = \mathbf{I}_d$.

Demostración.

Supóngase que A es de rango m . Se verá que $A^- = A^t(AA^t)^{-1}$ cumple con las condiciones dadas en la definición 2.2.1.

$$a. AA^-A = AA^t(AA^t)^{-1}A = AA^t(A^t)^{-1}A^{-1}A = A$$

$$b. A^-AA^- = A^t(AA^t)^{-1}AA^t(AA^t)^{-1} = A^t(AA^t)^{-1}(AA^t)(AA^t)^{-1} = A^t(AA^t)^{-1} \\ = A^-$$

$$c. (AA^-)^t = (A^-)^tA^t = [A^t(AA^t)^{-1}]^tA^t = [(AA^t)^{-1}]^t(A^t)^tA^t = (AA^t)^{-1}AA^t \\ = I_d, \text{ pues } AA^t \text{ es una matriz cuadrada de rango completo } m$$

$$d. (A^-A)^t = A^t(A^-)^t = A^t[A^t(AA^t)^{-1}]^t = A^t[(AA^t)^{-1}]^t(A^t)^t = A^t(AA^t)^{-1}A \\ = A^-A, \text{ es simétrica}$$

Luego se cumple que $A^t(AA^t)^{-1}$ es una inversa generalizada de A si tiene rango m y además $AA^t(AA^t)^{-1} = I_d$.

Ahora si A es de rango n se probará que $A^- = (A^tA)^{-1}A^t$ verifica con las condiciones dadas en la definición 2.2.1.

$$a. AA^-A = A(A^tA)^{-1}A^tA = A(A^tA)^{-1}(A^tA) = A$$

$$b. A^-AA^- = (A^tA)^{-1}A^tA(A^tA)^{-1}A^t = (A^tA)^{-1}(A^tA)(A^tA)^{-1}A^t = (A^tA)^{-1}A^t \\ = A^-$$

$$\begin{aligned} \text{c. } (AA^-)^t &= (A^-)^t A^t = [(A^t A)^{-1} A^t]^t A^t = (A^t)^t [(A^t A)^{-1}]^t A^t = A(A^t A)^{-1} A^t \\ &= AA^-, \text{ es simétrica} \end{aligned}$$

$$\begin{aligned} \text{d. } (A^- A)^t &= A^t (A^-)^t = A^t [(A^t A)^{-1} A^t]^t = A^t (A^t)^t [(A^t A)^{-1}]^t = A^t A (AA^t)^{-1} \\ &= (A^t A)(AA^t)^{-1} = I_d, \text{ pues } A^t A \text{ es una matriz cuadrada de rango} \\ &\text{completo } n \end{aligned}$$

Así, $(A^t A)^{-1} A^t$ es una inversa generalizada de A si tiene rango n y también $(A^t A)^{-1} A^t A = I_d$.

□

Teorema 2.2.13.

Las matrices AA^- , A^-A , $I_d - AA^-$ y $I_d - A^-A$ son todas simétricas e idempotentes.

Demostración.

Si A^- es la inversa generalizada de la matriz A es claro que AA^- y A^-A son simétricas.

Se probará que son idempotentes,

$$(AA^-)(AA^-) = AA^- AA^- = AA^-$$

$$(A^-A)(A^-A) = A^- AA^- A = A^- A$$

Por otra parte,

$$(I_d - AA^-)^t = I_d^t - (AA^-)^t = I_d - AA^-$$

$$\begin{aligned}(I_d - AA^-)(I_d - AA^-) &= I_d - AA^- - AA^- + (AA^-)^2 = I_d - AA^- - AA^- + AA^- \\ &= I_d - AA^-\end{aligned}$$

$$(I_d - A^-A)^t = I_d^t - (A^-A)^t = I_d - A^-A$$

$$\begin{aligned}(I_d - A^-A)(I_d - A^-A) &= I_d - A^-A - A^-A + (A^-A)^2 = I_d - A^-A - A^-A + A^-A \\ &= I_d - A^-A\end{aligned}$$

Así, AA^- , A^-A , $I_d - AA^-$ y $I_d - A^-A$ son todas simétricas e idempotentes.

□

Teorema 2.2.14.

Sea \mathbf{B} una matriz $m \times r$ de rango r ($r > 0$), y \mathbf{C} una matriz $r \times m$ de rango r ; entonces $(\mathbf{BC})^- = \mathbf{C}^- \mathbf{B}^-$.

Demostración.

Si \mathbf{B} es una matriz $m \times r$ de rango r y \mathbf{C} una matriz $r \times m$ de rango r del teorema 2.2.12 se tiene que:

$$\mathbf{B}^- = (\mathbf{B}^t \mathbf{B})^{-1} \mathbf{B}^t, \quad \mathbf{C}^- = \mathbf{C}^t (\mathbf{C} \mathbf{C}^t)^{-1}$$

Así, $\mathbf{C}^- \mathbf{B}^- = \mathbf{C}^t (\mathbf{C} \mathbf{C}^t)^{-1} (\mathbf{B}^t \mathbf{B})^{-1} \mathbf{B}^t$. Ahora se verificará que $\mathbf{C}^- \mathbf{B}^-$ es la inversa generalizada de \mathbf{BC} .

$$\begin{aligned}\text{a. } (\mathbf{BC}) \mathbf{C}^- \mathbf{B}^- (\mathbf{BC}) &= (\mathbf{BC}) \mathbf{C}^t (\mathbf{C} \mathbf{C}^t)^{-1} (\mathbf{B}^t \mathbf{B})^{-1} \mathbf{B}^t (\mathbf{BC}) \\ &= \mathbf{B} (\mathbf{C} \mathbf{C}^t) (\mathbf{C} \mathbf{C}^t)^{-1} (\mathbf{B}^t \mathbf{B})^{-1} (\mathbf{B}^t \mathbf{B}) \mathbf{C} = \mathbf{BC}\end{aligned}$$

- b. $C^-B^-(BC)C^-B^- = C^t(CC^t)^{-1}(B^tB)^{-1}B^t(BC)C^t(CC^t)^{-1}(B^tB)^{-1}B^t$
 $= C^t(CC^t)^{-1}(B^tB)^{-1}(B^tB)(CC^t)(CC^t)^{-1}(B^tB)^{-1}B^t$
 $= C^t(CC^t)^{-1}(B^tB)^{-1}B^t = C^-B^-$
- c. $(BC)C^-B^- = (BC)C^t(CC^t)^{-1}(B^tB)^{-1}B^t = B(CC^t)(CC^t)^{-1}(B^tB)^{-1}B^t =$
 $B(B^tB)^{-1}B^t$
 $[B(B^tB)^{-1}B^t]^t = (B^t)^t[(B^tB)^{-1}]^tB^t = B[(B^tB)^t]^{-1}B^t = B(B^tB)^{-1}B^t,$ es
simétrica.
- d. $C^-B^-(BC) = C^t(CC^t)^{-1}(B^tB)^{-1}B^t(BC) = C^t(CC^t)^{-1}(B^tB)^{-1}(B^tB)C$
 $= C^t(CC^t)^{-1}C$
 $[C^t(CC^t)^{-1}C]^t = C^t[(CC^t)^{-1}]^t(C^t)^t = C^t[(CC^t)^t]^{-1}C = C^t(CC^t)^{-1}C,$ es
simétrica.

Por lo tanto, C^-B^- satisface las condiciones de la definición 2.2.1, luego $(BC)^- = C^-B^-$.

□

Teorema 2.2.15.

Sea A una matriz dada, luego $(A^tA)^- = A^-(A^t)^-$.

Demostración.

Sea $(A^tA)^-$ la inversa generalizada de la matriz A^tA . Se probará que $A^-(A^t)^-$ también es una inversa generalizada de A^tA y por ende $(A^tA)^- = A^-(A^t)^-$.

Ahora, se verificará que $A^-(A^t)^-$ satisfaga las condiciones:

$$\begin{aligned} \text{a. } (A^t A) A^-(A^t)^- (A^t A) &= A^t A A^-(A^t)^- A^t A = A^t A A^-(AA^-)^t A \\ &= A^t A A^- A A^- A = A^t A A^- A = A^t A \end{aligned}$$

$$\begin{aligned} \text{b. } A^-(A^t)^- (A^t A) A^-(A^t)^- &= A^-(A^t)^- A^t A A^-(A^t)^- = A^-(AA^-)^t A A^-(A^t)^- \\ &= A^- A A^- A A^-(A^t)^- = A^- A A^-(A^t)^- = A^-(A^t)^- \end{aligned}$$

$$\begin{aligned} \text{c. } (A^t A) A^-(A^t)^- &= A^t (A A^-) (A^t)^- = A^t (A A^-)^t (A^t)^- \\ &= A^t (A^t)^- A^t (A^t)^- = A^t (A^t)^- = (A^- A)^t = A^- A, \end{aligned}$$

es simétrica

$$\text{d. } A^-(A^t)^- (A^t A) = A^- [(A^t)^- A^t] A = A^- [AA^-]^t A = A^- A A^- A = A^- A,$$

es simétrica

Por lo tanto, $A^-(A^t)^-$ también es una inversa generalizada de $A^t A$ y como ésta es única se tiene que $(A^t A)^- = A^-(A^t)^-$.

□

Teorema 2.2.16.

Sea P una matriz ortogonal de orden $m \times m$, Q una matriz ortogonal de orden $n \times n$, y A una matriz de orden $m \times n$. Entonces $(PAQ)^- = Q^t A^- P^t$.

Demostración.

Se probará que $Q^t A^- P^t$ es la inversa generalizada de PAQ , donde P es una matriz ortogonal de $m \times m$, $P^t P = P P^t = I_m$, Q es una matriz ortogonal de $n \times n$, $Q^t Q =$

$QQ^t = I_d$, y A es una matriz de $m \times n$. Para ello se verificará si $Q^t A^{-1} P^t$ satisface las condiciones de la definición 2.2.1,

- a. $PAQ(PAQ)^{-1} PAQ = PAQQ^t A^{-1} P^t PAQ = PAA^{-1}AQ = PAQ$
- b. $(PAQ)^{-1} PAQ(PAQ)^{-1} = Q^t A^{-1} P^t PAQQ^t A^{-1} P^t = Q^t A^{-1} AA^{-1} P^t = Q^t A^{-1} P^t$
- c. $PAQ(PAQ)^{-1} = PAQQ^t A^{-1} P^t = PAA^{-1} P^t = P(AA^{-1})P^t$, es simétrica pues AA^{-1} es simétrica y $[P(AA^{-1})P^t]^t = P(AA^{-1})^t P^t = P(AA^{-1})P^t$
- d. $(PAQ)^{-1} PAQ = Q^t A^{-1} P^t PAQ = Q^t A^{-1} AQ = Q^t (A^{-1}A)Q$, es simétrica ya que $A^{-1}A$ es simétrica y también $[Q^t (A^{-1}A)Q]^t = Q^t (A^{-1}A)^t Q = Q^t (A^{-1}A)Q$

Así, se tiene que $(PAQ)^{-1} = Q^t A^{-1} P^t$.

□

Teorema 2.2.17.

Sea A una matriz de orden $m \times n$, entonces $AB = AA^{-1}$ si, y sólo si, B es tal que $ABA = A$ y AB es simétrica.

Demostración.

Sea A^{-1} la inversa generalizada de A .

Supóngase que $AB = AA^{-1}$. Luego:

$$ABA = AA^{-1}A = A$$

$AB = AA^{-}$ es simétrica pues AA^{-} es simétrica.

Ahora supóngase que $ABA = A$ y AB es simétrica. Así,

$$\begin{aligned} AB &= AA^{-}AB = (AA^{-})^t(AB)^t = (A^{-})^tA^tB^tA^t = (A^{-})^t(ABA)^t = (A^{-})^t(A)^t = (AA^{-})^t \\ &= AA^{-} \end{aligned}$$

□

Entre las múltiples aplicaciones que tiene el concepto de inversa generalizada cabe señalar el papel que desempeña en el análisis y solución de sistemas de ecuaciones lineales. En el caso de sistemas consistentes, permiten caracterizarlas y para sistemas inconsistentes posibilitan hallar soluciones aproximadas. Se analizará con ayuda de la inversa generalizada a determinar cuando existe una solución al sistema de ecuaciones lineales consistente $Ax = b$.

Teorema 2.2.18.

El sistema de ecuaciones lineales $Ax = b$ es consistente si, y sólo si, se verifica que $AA^{-}b = b$.

Demostración.

Supóngase que el sistema $Ax = b$ es consistente. Luego existe $x_0 \in \mathbb{R}^n$ tal que $Ax_0 = b$. Como A^{-} siempre existe, se tiene que

$$\begin{aligned}
 A^{-1}Ax_0 &= A^{-1}b \\
 \therefore AA^{-1}Ax_0 &= AA^{-1}b \\
 \therefore Ax_0 &= AA^{-1}b \\
 \therefore b &= AA^{-1}b
 \end{aligned}$$

Ahora bien, si $AA^{-1}b = b$ se probará que el sistema $Ax = b$ es consistente. Sea $x_0 = A^{-1}b$, entonces $Ax_0 = AA^{-1}b = b$. Por lo tanto el sistema es consistente, puesto que al menos $x_0 = A^{-1}b$ es solución del mismo.

□

Teorema 2.2.19.

Si A es una matriz $m \times n$ de rango m , entonces el sistema $Ax = b$ tiene una solución.

Demostración.

Sea A una matriz $m \times n$ de rango m , luego por el teorema 2.2.12 se tiene que $AA^{-1} = I_m$. Así, se tiene $AA^{-1}b = b$, y por el teorema anterior se concluye que el sistema $Ax = b$ es consistente, por lo tanto, tiene una solución.

□

Teorema 2.2.20.

Si A es una matriz $m \times n$. El sistema lineal homogéneo $Ax = 0$ tiene una solución que no sea $x = 0$ si, y sólo si, $\text{rang}(A) < n$.

Demostración.

Si $\mathbf{x} \neq 0$ es una solución, se probará que el $\text{rang}(\mathbf{A}) < n$. Supóngase que no es así, que el rango de \mathbf{A} no es menor que n . Luego, si $\mathbf{Ax} = 0$ se tiene que $\mathbf{A}^{-}\mathbf{Ax} = \mathbf{A}^{-}0 = 0$, pero $\mathbf{A}^{-}\mathbf{A} = \mathbf{I}_d$ por el teorema 2.2.12, así, $\mathbf{x} = \mathbf{A}^{-}0 = 0$ lo cual es una contradicción. Por lo tanto $\text{rang}(\mathbf{A}) < n$.

Ahora si $\text{rang}(\mathbf{A}) < n$ se tiene que las columnas de \mathbf{A} son linealmente dependientes. Denotando \mathbf{a}_i las columnas, se tiene que $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$, donde cada \mathbf{a}_i son linealmente dependientes, así existen constantes x_1, x_2, \dots, x_n no todas ceros tales que $\sum_{i=1}^n \mathbf{a}_i x_i = \mathbf{Ax} = 0$.

□

Teorema 2.2.21.

Si \mathbf{A} es una matriz de orden $m \times n$, \mathbf{A}^{-} la inversa generalizada de \mathbf{A} , y supóngase que el sistema $\mathbf{Ax} = \mathbf{b}$ tiene solución. Para cada vector $\mathbf{h} \in \mathbb{R}^n$ de orden $n \times 1$, el vector \mathbf{x}_0 es una solución, donde:

$$\mathbf{x}_0 = \mathbf{A}^{-}\mathbf{b} + (\mathbf{I}_d - \mathbf{A}^{-}\mathbf{A})\mathbf{h} \quad (2.2.12)$$

También, si \mathbf{x}_0 es una solución cualquiera del sistema, existe $\mathbf{h} \in \mathbb{R}^n$ tal que \mathbf{x}_0 puede expresarse en la forma dada en la ecuación (2.2.12).

Demostración.

Supóngase que el sistema tiene solución y se probará que el x_0 dado en la ecuación (2.2.12) es una solución del sistema, es decir, $Ax_0 = b$.

Por el teorema 2.2.19, se tiene $AA^{-1}b = b$. Así,

$$\begin{aligned} Ax_0 &= AA^{-1}b + A(I_d - A^{-1}A)h \\ \therefore Ax_0 &= AA^{-1}b + (A - AA^{-1}A)h \\ \therefore Ax_0 &= AA^{-1}b + (A - A)h \\ \therefore Ax_0 &= AA^{-1}b \\ \therefore Ax_0 &= b \end{aligned}$$

Por lo tanto, $x_0 = A^{-1}b + (I_d - A^{-1}A)h$ es una solución.

Ahora supóngase que x_0 es una solución cualquiera del sistema y se probará que existe $h \in \mathbb{R}^n$ tal que x_0 puede expresarse en la forma dada en la ecuación (2.2.12). Si x_0 es una solución del sistema se tiene,

$$\begin{aligned} A^{-1}Ax_0 &= A^{-1}b \\ \therefore A^{-1}b - A^{-1}Ax_0 &= 0 \end{aligned}$$

Sumando x_0 en ambos lados, se tiene:

$$x_0 = A^{-1}b + x_0 - A^{-1}Ax_0$$

$$\therefore \mathbf{x}_0 = \mathbf{A}^{-1}\mathbf{b} + (\mathbf{I}_d - \mathbf{A}^{-1}\mathbf{A})\mathbf{x}_0$$

Luego, tomando $\mathbf{h} = \mathbf{x}_0$ se obtiene lo deseado.

□

Corolario 2.2.2.

Si el sistema $\mathbf{Ax} = \mathbf{b}$ es consistente, el vector $\mathbf{x}_0 = \mathbf{A}^{-1}\mathbf{b}$ es solución del sistema.

Demostración.

Si el sistema $\mathbf{Ax} = \mathbf{b}$ es consistente, por el teorema anterior, se tiene que $\mathbf{x}_0 = \mathbf{A}^{-1}\mathbf{b} + (\mathbf{I}_d - \mathbf{A}^{-1}\mathbf{A})\mathbf{h}$ es solución. Así, haciendo $\mathbf{h} = \mathbf{0}$ se tiene que $\mathbf{x}_0 = \mathbf{A}^{-1}\mathbf{b}$ también lo es.

□

Corolario 2.2.3.

Si el sistema $\mathbf{Ax} = \mathbf{b}$ es consistente, entonces la solución $\mathbf{x}_0 = \mathbf{A}^{-1}\mathbf{b}$ es única si, y sólo si,

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_d$$

Demostración.

Supóngase que el sistema es consistente. Luego por el teorema 2.2.22, se tiene que $\mathbf{x}_0 = \mathbf{A}^{-1}\mathbf{b} + (\mathbf{I}_d - \mathbf{A}^{-1}\mathbf{A})\mathbf{h}$ es solución, para cada vector \mathbf{h} . Y desde luego, $\mathbf{x}_0 = \mathbf{A}^{-1}\mathbf{b}$ para cada vector \mathbf{h} , si, y sólo si, $\mathbf{I}_d - \mathbf{A}^{-1}\mathbf{A} = \mathbf{0}$, así, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_d$.

□

Corolario 2.2.4.

Si el sistema $\mathbf{Ax} = \mathbf{b}$ es consistente, donde \mathbf{A} es una matriz $m \times n$, entonces el sistema tiene una solución única si, y sólo si, el rango de \mathbf{A} es n .

Demostración.

Si el sistema $\mathbf{Ax} = \mathbf{b}$ es consistente, del corolario anterior se tiene que el sistema tiene solución única si, y sólo si, $\mathbf{A}^- \mathbf{A} = \mathbf{I}_d$, ya que \mathbf{A} es una matriz de $m \times n$, se tiene que el rango de $\mathbf{A}^- \mathbf{A}$ y de \mathbf{I}_d es n .

Ahora bien, por propiedades del rango de una matriz se tiene que

$$\text{rang}(\mathbf{A}) + \text{rang}(\mathbf{A}^-) - n \leq \text{rang}(\mathbf{A}^- \mathbf{A}) \leq \min(\text{rang}(\mathbf{A}), \text{rang}(\mathbf{A}^-))$$

Supóngase que el rango de \mathbf{A} es $r > 0$, así

$$2r - n \leq n \leq r, \text{ pues } \text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A}^-)$$

Por lo tanto, el rango de \mathbf{A} es n , ya que de la ecuación anterior se concluye que $r = n$.

Por otro lado, si el rango de \mathbf{A} es n , del teorema 2.2.12 se tiene que $\mathbf{A}^- \mathbf{A} = \mathbf{I}_d$, luego $\mathbf{A}^- \mathbf{A} \mathbf{b} = \mathbf{b}$ y por el teorema 2.2.19 el sistema de ecuaciones lineales $\mathbf{Ax} = \mathbf{b}$ es consistente.

□

2.3. INVERSA CONDICIONAL DE MATRICES.

Al igual que el concepto de inversa generalizada de una matriz, el concepto de inversa condicional es de gran utilidad en los cursos de modelos lineales y en la caracterización del conjunto solución de sistemas lineales de ecuaciones.

Como se sabe, toda matriz \mathbf{A} tiene una única inversa generalizada \mathbf{A}^- y como se verá más adelante ésta, a su vez, por definición, es una inversa condicional \mathbf{A}^c . Así que, toda matriz \mathbf{A} tiene al menos una inversa condicional, más aún, una matriz \mathbf{A} puede tener varias (incluso infinitas) inversas condicionales, salvo cuando la matriz \mathbf{A} es invertible, en cuyo caso \mathbf{A}^{-1} es la única inversa condicional.

Existen diferentes métodos para calcular la inversa condicional, uno de ellos es el siguiente:

- i. Tomar una submatriz de $(\mathbf{X}^t\mathbf{X})$ de rango completo y orden igual al rango de $\mathbf{X}^t\mathbf{X}$ (submatriz no singular). Llámese a esa submatriz \mathbf{B} de orden $k \times k$ donde $k = \text{rang}(\mathbf{X})$.
- ii. Invertir \mathbf{B} .
- iii. Transponer \mathbf{B} .

- iv. Reemplazar con los elementos de \mathbf{B} los elementos homólogos en $(\mathbf{X}^t\mathbf{X})$ y sustituir con ceros el resto para obtener la matriz \mathbf{C} de orden $p \times p$.
- v. Transponer \mathbf{C} para obtener una inversa condicional de $(\mathbf{X}^t\mathbf{X})$.

Para facilitar la comprensión de este procedimiento, se ilustrara aplicando los pasos a una matriz \mathbf{A} , de orden 5×5 y de rango 4, dada por:

$$\mathbf{A} = \begin{bmatrix} 16 & 4 & 4 & 4 & 4 \\ 4 & 4 & 0 & 0 & 0 \\ 4 & 0 & 4 & 0 & 0 \\ 4 & 0 & 0 & 4 & 0 \\ 4 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Se toma una submatriz de rango completo y de orden igual al rango de \mathbf{A} (sub matriz no singular). El rango de esta matriz \mathbf{A} es 4, por lo tanto el orden de la sub matriz será de 4×4 ,

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

Se invierte esta sub matriz para encontrar \mathbf{B}

$$\mathbf{B} = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}$$

Se transpone \mathbf{B} . En este caso $\mathbf{B} = \mathbf{B}^t$ y $\mathbf{B}^t = \mathbf{B}$ (simétrica)

$$\mathbf{B}^t = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}$$

Se reemplaza en \mathbf{A} y se completa con ceros para encontrar \mathbf{C}

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}$$

Se vuelve a trasponer C para obtener una inversa condicional A^c , por ser simétrica A , este último paso se puede obviar, ya que la traspuesta de una inversa condicional de una matriz simétrica, es también una inversa condicional de esa matriz. Luego, A^c es una inversa generalizada de A .

$$A^c = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}$$

Definición 2.3.1.

Sea A es una matriz de orden $m \times n$, se dice que A^c es una inversa condicional de A si, y sólo si, satisface que $AA^cA = A$.

Note que de la definición anterior se tiene que si A^- es una inversa generalizada de una matriz A , también se tiene que A^- es una inversa condicional de A , es decir, toda inversa generalizada es una inversa condicional de una matriz cualquiera, ya que la condición de que $AA^cA = A$ es la condición i de la inversa generalizada. Pero no se cumple que toda inversa condicional sea una inversa generalizada ya que necesariamente no se cumplen

las otras tres condiciones de la definición 2.2.1. Lo cual se expresa en el siguiente teorema.

Teorema 2.3.1.

La inversa generalizada de una matriz \mathbf{A} es también una inversa condicional de \mathbf{A} , pero no toda inversa condicional es necesariamente una inversa generalizada.

Los siguientes teoremas son una consecuencia de los teoremas 2.2.1 y 2.2.3 de inversas generalizadas respectivamente.

Teorema 2.3.2.

Si la inversa condicional de una matriz $\mathbf{A}_{m \times n}$ existe, entonces \mathbf{A}^c es de orden $n \times m$.

Teorema 2.3.3.

Cada matriz \mathbf{A} tiene una inversa condicional pero puede no ser única.

Dado que toda inversa generalizada es una inversa condicional de una matriz \mathbf{A} cualquiera se tiene que todos los teoremas expuestos en la sección anterior son aplicables en esta sección.

Teorema 2.3.4.

Sea \mathbf{A} una matriz de orden $m \times n$ de rango $r > 0$, entonces se cumple que:

- i. El rango de \mathbf{A}^c no es menor que r , es decir, $\text{rang}(\mathbf{A}^c) \geq \text{rang}(\mathbf{A})$.
- ii. $\mathbf{A}^c \mathbf{A}$ y $\mathbf{A} \mathbf{A}^c$ son matrices idempotentes.
- iii. $\text{rang}(\mathbf{A}^c \mathbf{A}) = \text{rang}(\mathbf{A} \mathbf{A}^c) = r$.
- iv. $\mathbf{A}^c \mathbf{A} = \mathbf{I}_d$ si, y sólo si, el rango de \mathbf{A} es n .
- v. $\mathbf{A} \mathbf{A}^c = \mathbf{I}_d$ si, y sólo si, el rango de \mathbf{A} es m .
- vi. Si \mathbf{A}^c es una inversa generalizada de \mathbf{A} , entonces $(\mathbf{A}^c)^t$ es la inversa condicional de \mathbf{A}^t .

Demostración.

Sea \mathbf{A}^c la inversa generalizada de la matriz \mathbf{A} . Luego $\mathbf{A} \mathbf{A}^c \mathbf{A} = \mathbf{A}$, así

- i. $\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A} \mathbf{A}^c \mathbf{A}) \leq \text{rang}(\mathbf{A}^c)$.
- ii. $(\mathbf{A}^c \mathbf{A})(\mathbf{A}^c \mathbf{A}) = \mathbf{A}^c \mathbf{A} \mathbf{A}^c \mathbf{A} = \mathbf{A}^c \mathbf{A}$
 $(\mathbf{A} \mathbf{A}^c)(\mathbf{A} \mathbf{A}^c) = \mathbf{A} \mathbf{A}^c \mathbf{A} \mathbf{A}^c = \mathbf{A} \mathbf{A}^c$
- iii. Por una parte se tiene que $\text{rang}(\mathbf{A}^c \mathbf{A}) \leq \text{rang}(\mathbf{A})$ y $\text{rang}(\mathbf{A} \mathbf{A}^c) \leq \text{rang}(\mathbf{A})$.

Además ya que $\mathbf{A} \mathbf{A}^c \mathbf{A} = \mathbf{A}$ se tiene

$$\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A} \mathbf{A}^c \mathbf{A}) \leq \text{rang}[\mathbf{A}(\mathbf{A}^c \mathbf{A})] \leq \text{rang}(\mathbf{A}^c \mathbf{A}) \quad \text{y}$$

$$\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A} \mathbf{A}^c \mathbf{A}) \leq \text{rang}[(\mathbf{A} \mathbf{A}^c)\mathbf{A}] \leq \text{rang}(\mathbf{A} \mathbf{A}^c)$$

$$\text{Así, } \text{rang}(\mathbf{A}^c \mathbf{A}) = \text{rang}(\mathbf{A} \mathbf{A}^c) = \text{rang}(\mathbf{A}) = r$$

- iv. Dado que A una matriz de orden $m \times n$ entonces $A^c A$ es una matriz de orden $n \times n$.

Si $A^c A = I_d$ entonces $\text{rang}(A^c A) = \text{rang}(I_d)$, pero I_d también es una matriz de orden $n \times n$ y así $\text{rang}(I_d) = n$, por lo tanto $\text{rang}(A^c A) = n$. Por otra parte se tiene que $\text{rang}(A^c A) = \text{rang}(A)$, luego $\text{rang}(A) = n$.

Si el rango de A es n , entonces $A^c A$ es invertible, pues el $\text{rang}(A^c A) = \text{rang}(A) = n$, así $(A^c A)^c = (A^c A)^{-1}$. Pero $(A^c A)^c = A^c A$ (teorema 2.2.14). Luego $(A^c A)^c A^c A = I_d \therefore A^c A A^c A = I_d \therefore A^c A = I_d$.

- v. Dado que A una matriz de orden $m \times n$ entonces AA^c es una matriz de orden $m \times m$.

Si $AA^c = I_d$ entonces $\text{rang}(AA^c) = \text{rang}(I_d)$, pero I_d es una matriz de orden $m \times m$ y así $\text{rang}(I_d) = m$, por lo tanto $\text{rang}(AA^c) = m$. Por otra parte se tiene que $\text{rang}(AA^c) = \text{rang}(A)$, luego $\text{rang}(A) = m$.

Si el rango de A es m , entonces AA^c es invertible, pues el $\text{rang}(AA^c) = \text{rang}(A) = m$, así $(AA^c)^c = (AA^c)^{-1}$. Pero $(AA^c)^c = AA^c$ (teorema 2.2.14). Luego $(AA^c)^c AA^c = I_d \therefore AA^c AA^c = I_d \therefore AA^c = I_d$.

- vi. Si $AA^c A = A$, se tiene que $[AA^c A]^t = A^t$, pero $[AA^c A]^t = [(AA^c)A]^t = A^t (AA^c)^t = A^t (A^c)^t A^t$. Luego $A^t (A^c)^t A^t = A^t$, es decir, $(A^c)^t$ es la inversa condicional de A^t .

□

Teorema 2.3.5.

Para una matriz A de orden $m \times n$ de rango r , se define K por $K = A(A^t A)^c A^t$.

Entonces K es invariante para una inversa condicional de $A^t A$.

Demostración.

Sean B y C dos inversas condicionales de $A^t A$. Luego $A^t A B A^t A = A^t A C A^t A$ y se tomará $A = A_L A_R$, donde A_L es de orden $m \times r$, A_R de orden $r \times n$ y ambas son de rango r (la descomposición de rango completo de la matriz A), así

$$\begin{aligned} (A_L A_R)^t A_L A_R B (A_L A_R)^t A_L A_R &= (A_L A_R)^t A_L A_R C (A_L A_R)^t A_L A_R \\ A_R^t A_L^t A_L A_R B A_R^t A_L^t A_L A_R &= A_R^t A_L^t A_L A_R C A_R^t A_L^t A_L A_R \end{aligned} \quad (2.3.1)$$

Multiplicando la ecuación anterior por $(A_R^t)^c$ por la izquierda y por la derecha por $(A_R)^c$ y dado que $(A_R^t)^c A_R^t = I_d$ y $A_R (A_R)^c = I_d$ (teorema 2.3.4) se tiene

$$A_L^t A_L A_R B A_R^t A_L^t A_L = A_L^t A_L A_R C A_R^t A_L^t A_L \quad (2.3.2)$$

Ya que A_L es de orden completo por columnas se tiene que existe $(A_L^t A_L)^{-1}$ así multiplicando la ecuación 2.3.2 por $(A_L^t A_L)^{-1}$ a la derecha e izquierda se tiene

$$\begin{aligned} A_R B A_R^t &= A_R C A_R^t \\ \therefore A_L A_R B A_R^t A_L^t &= A_L A_R C A_R^t A_L^t \\ \therefore A B A^t &= A C A^t \end{aligned}$$

Como B y C son inversas condicionales de $A^t A$ se obtiene que K es invariante para una inversa condicional de $A^t A$.

□

Teorema 2.3.6.

$A(A^t A)^c A^t = AA^-$, donde $(A^t A)^c$ es la inversa condicional de $A^t A$.

Demostración.

Del teorema anterior se tiene que $A(A^t A)^c A^t$ es invariante para una inversa condicional de $A^t A$. Como $(A^t A)^-$ es una inversa condicional de $A^t A$ se tiene

$$A(A^t A)^c A^t = A(A^t A)^- A^t = AA^-(A^t)^- A^t = AA^-(A^-)^t A^t = AA^-(AA^-)^t = AA^- AA^-$$

$$A(A^t A)^c A^t = AA^-.$$

□

Teorema 2.3.7.

Definamos K por $K = A(A^t A)^c A^t$ igual que en el teorema 2.3.5, entonces

- i. K es simétrica e idempotente.
- ii. $\text{rang}(K) = \text{rang}(A) = r$.
- iii. $KA = A$, $A^t K = A^t$

- iv. $(A^t A)^c A^t$ es una inversa condicional de A para cualquier inversa condicional de $A^t A$.
- v. $A(A^t A)^c$ es una inversa condicional de A^t para cualquier inversa condicional de $A^t A$.

Demostración.

Sea $K = A(A^t A)^c A^t$. Luego del teorema 2.3.6 se tiene que $K = AA^-$. Así,

- i. Se probará que $K = K^t$ y $K = K^2$.

$$K^t = [AA^-]^t = AA^-, \text{ ya que } AA^- \text{ es simétrica.}$$

$$K^2 = [AA^-][AA^-] = AA^-AA^- = AA^- = K$$

- ii. $\text{rang}(K) = \text{rang}(AA^-)$ pero $\text{rang}(A) = r$ y por el corolario 2.5.7 se tiene que $\text{rang}(AA^-) = r$ así, $\text{rang}(K) = \text{rang}(A) = r$.
- iii. $KA = AA^-A = A$ y $A^t K = A^t K^t = A^t (AA^-)^t = A^t (A^-)^t A^t = A^t (A^t)^- A^t = A^t$
- iv. Sea $(A^t A)^c$ una inversa condicional de $A^t A$. Luego $A[(A^t A)^c A^t]A = KA = AA^-A = A$, así $(A^t A)^c A^t$ es una inversa condicional de A .
- v. Sea $(A^t A)^c$ una inversa condicional de $A^t A$. Luego $A^t [A(A^t A)^c] A^t = A^t K^t = A^t (AA^-)^t = A^t (A^-)^t A^t = A^t (A^t)^- A^t = A^t$, así $A(A^t A)^c$ es una inversa condicional de A^t .

□

El siguiente resultado es una consecuencia de los Teoremas 2.3.1 y 2.2.22 respectivamente.

Corolario 2.3.1.

Suponga que el sistema de ecuaciones $\mathbf{Ax} = \mathbf{b}$, \mathbf{A} una matriz de orden $m \times n$, tiene solución. Entonces para cada vector $\mathbf{h} \in \mathbb{R}^n$ de orden $n \times 1$, el vector \mathbf{x}_0 es una solución, donde:

$$\mathbf{x}_0 = \mathbf{A}^c \mathbf{b} + (\mathbf{I}_d - \mathbf{A}^c \mathbf{A}) \mathbf{h} \quad (2.3.3)$$

También, si \mathbf{x}_0 es una solución cualquiera del sistema, existe $\mathbf{h} \in \mathbb{R}^n$ tal que \mathbf{x}_0 puede expresarse en la forma dada en la ecuación (2.3.3).

Capítulo 3.

Modelo Lineal.

El planteamiento y resolución de ecuaciones matemáticas tienen como objeto relacionar el comportamiento de una variable respuesta con el de una o varias variables explicativas. Se puede distinguir entre diversos tipos de ecuaciones: lineales, no lineales, diferenciales, etc. Se estudiará fundamentalmente las primeras, es decir, se considerará básicamente relaciones de tipo lineal entre la variable respuesta y las variables explicativas.

Este tipo de relación se observa con relativa frecuencia en la naturaleza, que su principal virtud es su fácil manejo, su excelente y natural comportamiento desde el punto de vista formal, la cual se puede considerar como relaciones lineales, asumiendo en consecuencia cierto error como tributo a la sencillez del modelo. Cuando este error resulta excesivo se buscan cambios apropiados en las variables que permitan establecer relaciones aproximadamente lineales entre las variables transformadas. Se admite una variación o error de carácter aleatorio, lo cual conduce a considerar un modelo de tipo probabilístico. Dado que las distribuciones de probabilidad no están especificadas por completo, se podría considerar un modelo estadístico, que se denomina Modelo Lineal.

Los modelos lineales es una herramienta muy utilizada para el análisis de datos que presentan una relación causa-efecto. El punto de partida en un modelo lineal son conjuntos de datos que se presentan simultáneamente, y que a priori pueden explicar el comportamiento de la variable que se quiere analizar (la que se denomina la variable respuesta o dependiente) a partir del resto.

Graybill (1976), define cuatro modelos que pueden ser utilizados para representar las situaciones en el mundo real. Estos modelos los clasifica como cuantitativos y cualitativos:

- I. Modelos cuantitativos: Modelo Lineal General y Modelo de Regresión lineal.
- II. Modelos cualitativos: Modelos de Diseños y Modelo de Componentes de Varianza.

Estos modelos están estrechamente relacionados y los procedimientos para la elaboración de las inferencias revelan una gran similitud entre ellos. Los modelos expuestos anteriormente pueden ser vistos como variaciones del Modelo Lineal General.

Basado en una revisión de Graybill (1976), se presentan a continuación algunos conceptos básicos de Modelo Lineal.

3.1. PLANTEAMIENTO DEL MODELO LINEAL.

Un Modelo Lineal consiste en expresar un variable Y como función de p variables

X_1, X_2, \dots, X_p de la forma:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon,$$

donde, Y es aleatoria debido a que ε , conocido como el término de error, es considerado aleatorio, las X_i no aleatorias, en general, y los β_i son parámetros desconocidos. Estas consideraciones pueden ser detalladas en la siguiente definición.

Definición 3.1.1

Considere las n ecuaciones

$$\left. \begin{array}{l} Y_i = \sum_{j=1}^p X_{ij} \beta_j + \varepsilon_i \\ E(\varepsilon_i) = 0 \end{array} \right\} \quad i = 1, 2, \dots, n \quad (3.1.1)$$

donde,

1. Y_i son variables aleatorias observables;
2. X_{ij} , $1 \leq j \leq p$ son variables no aleatorias (en general) observables en un dominio D ;
3. β_j son parámetros desconocidos definidos en un espacio Ω ;

4. ε_i son variables aleatorias no observables, tales que $E(\varepsilon_i) = 0$, $V(\varepsilon_i) = \sigma^2$ y $cov[\varepsilon_i, \varepsilon_j] = \sigma_{ij}$ para $i \neq j$.

Con estas especificaciones, el sistema (3.1.1) es definido como Modelo Estadístico Lineal.

Definición 3.1.2

Considere la ecuación matricial

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{n \times 1}, \quad (3.1.2)$$

donde,

1. $\mathbf{Y}_{n \times 1}$ es un vector aleatorio observable;
2. $\mathbf{X}_{n \times p}$ es una matriz de datos observables en un dominio D , no aleatorio, en general;
3. $\boldsymbol{\beta}_{p \times 1}$ es un vector de parámetros desconocidos definidos en un espacio Ω ;
4. $\boldsymbol{\varepsilon}_{n \times 1}$ es un vector aleatorio no observable, tal que $E(\boldsymbol{\varepsilon}_{n \times 1}) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_{n \times 1}) = \sigma^2 \mathbf{I}_d$ y $cov[\boldsymbol{\varepsilon}_{n \times 1}] = \boldsymbol{\Sigma}_{n \times n}$.

Con estas especificaciones, el sistema (3.1.2) es definido como Modelo Estadístico Lineal General.

Cuando se establece el supuesto de normalidad en la distribución de los errores, esto es, cada $\varepsilon_i \sim N(0, \sigma^2)$ y que son estocásticamente independientes, entonces se dice que el modelo definido es un modelo lineal normal. Así se tendrá que

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_d),$$

es decir, \mathbf{Y} sigue la distribución normal múltiple multivariante de vector de medias $\mathbf{X}\boldsymbol{\beta}$ y la matriz de covarianzas $\sigma^2 \mathbf{I}_d$. Pues sí, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_d)$ se tiene que el vector de medias de \mathbf{Y} viene dado por:

$$E(\mathbf{Y}_{n \times 1}) = E(\mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{n \times 1})$$

$$E(\mathbf{Y}_{n \times 1}) = E(\mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1}) + E(\boldsymbol{\varepsilon}_{n \times 1})$$

$$E(\mathbf{Y}_{n \times 1}) = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1}$$

Y la matriz de varianzas y covarianzas por $\mathbf{I}_d (\sigma^2 \mathbf{I}_d) \mathbf{I}_d^t = \sigma^2 \mathbf{I}_d$

Se encontrará la estimación puntual del vector de parámetros desconocido $\boldsymbol{\beta}_{p \times 1}$, el cual se denotará por $\hat{\boldsymbol{\beta}}_{p \times 1}$ y del parámetro σ^2 denotado por $\hat{\sigma}^2$, y también estimación de estos parámetros a través de intervalos de confianza y las respectivas pruebas de hipótesis. Al hallar $\hat{\boldsymbol{\beta}}_{p \times 1}$, se estaría estimando el vector respuesta $\mathbf{Y}_{n \times 1}$, es decir, $\hat{\mathbf{Y}}_{n \times 1}$ donde:

$$\hat{\mathbf{Y}}_{n \times 1} = \mathbf{X}_{n \times p} \hat{\boldsymbol{\beta}}_{p \times 1} \quad (3.1.3)$$

La estimación del vector de parámetros $\boldsymbol{\beta}_{p \times 1}$ puede hacerse por vía de máxima verosimilitud, bajo el supuesto que se conozca la distribución de los errores, maximizando

la respectiva función de verosimilitud siguiendo ésta la misma expresión de los errores, pero considerada como función dependiente de los parámetros involucrados en el modelo, es decir, $\beta_{p \times 1}$, y σ^2 . Por esta vía se logra estimar a $\beta_{p \times 1}$, y σ^2 .

Cuando la distribución de los errores en el modelo es desconocida, se puede obtener la estimación de $\beta_{p \times 1}$, por vía de los mínimos cuadrados, minimizando la correspondiente suma de cuadrados de los errores.

El proceso de estimación de parámetros, bien sea por vía de máxima verosimilitud o por mínimos cuadrados, conduce al sistema de ecuaciones normales

$$X^t X \hat{\beta} = X^t Y \quad (3.1.4)$$

Nótese que este es un sistema de ecuaciones lineales, no homogéneo.

Todo esto, teniendo en cuenta lo que se conoce como rango de diseño al rango de la matriz X , $\text{rang}(X) = r$. El valor de r es el número efectivo de parámetros del diseño, en el sentido de que si $r < p$ es posible reparametrizar el modelo para que r sea igual al número de parámetros. En muchos casos el diseño verifica directamente que $r = p$. Es

por ello que los modelos lineales pueden clasificarse en dos tipos: modelos de rango completo, ($r = p$), y de rango incompleto o de rango deficiente, ($r < p$), (Kirk, 1982).

3.2. MODELO LINEAL DE RANGO COMPLETO: ESTIMACIÓN DE PARÁMETROS.

Considerando el modelo expresado en (3.1.2) con las características dadas y que el rango de la matriz $X_{n \times p}$ es p , el modelo se denomina Modelo Estadístico Lineal de Rango Completo, ya que $X_{n \times p}$ es de rango completo por columnas, es decir, los p vectores columnas de X son linealmente independientes. Si la matriz X es de rango completo, se tiene que $X^t X$ también lo es, por lo tanto, el sistema expresado en (3.1.4) tiene solución única.

Estimación de los parámetros.

(a) Método de mínimos cuadrados ordinarios (MCO).

En el contexto habitual de trabajo e investigación de los científicos e ingenieros se puede utilizar las diferentes herramientas que nos facilitan los métodos matemáticos de ajuste. Entre estos procedimientos, uno de los más interesantes por la extensión de su aplicación y por su flexibilidad, es el conocido como ajuste por el método de mínimos cuadrados

ordinarios, un método de ajustes de curvas que a principios del siglo XIX, en 1805, lo publicó el matemático francés Adrien Legendre y Gauss lo utilizó en 1795 y lo publicó en 1809 (Stigler, 1999).

Bajo ciertos supuestos, el método tiene algunas propiedades estadísticas muy interesantes que lo han convertido en uno de los más eficaces y populares del análisis de regresión. Los supuestos para su aplicación son:

- i. El modelo es lineal en los parámetros.
- ii. Los valores de X son fijos en muestreo repetido.
- iii. El valor medio del error es cero, $E(\varepsilon) = 0$.
- iv. La varianza de los errores es constante, $V(\varepsilon) = \sigma^2$.
- v. No existe autocorrelación entre los errores.
- vi. La covarianza entre ε y X es cero, $Cov(\varepsilon, X) = 0$.
- vii. El número de observaciones debe ser mayor al número de parámetros a estimar.
- viii. Variabilidad en los valores de X .
- ix. El modelo está correctamente especificado tanto en la forma funcional como en las variables que están en el modelo.

El método de mínimos cuadrados ordinarios permite ajustar un conjunto de datos presentados en un diagrama de dispersión. El ejemplo más simple de una aproximación

por mínimos cuadrados es el ajuste a través del modelo $y = \beta_0 + \beta_1 x + \varepsilon$, a un conjunto de parejas de datos observados. Una estrategia que obtiene la “mejor” línea a través de los puntos es minimizar la suma de los errores residuales.

Se trata de hallar el conjunto de valores de los parámetros $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$ que minimicen la suma de cuadrados $S(\beta)$,

$$S(\beta_1, \beta_2, \dots, \beta_p) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_1 X_{i1} - \dots - \hat{\beta}_p X_{ip})^2, \quad (3.2.1)$$

Derivando respecto a los parámetros $\beta_1, \beta_2, \dots, \beta_p$ e igualando a cero, se tiene lo que se conoce como sistema de ecuaciones normales (para más detalles ver Graybill, 1976); expresado ya en (3.1.4)

$$(\mathbf{X}^t \mathbf{X}) \hat{\beta} = \mathbf{X}^t \mathbf{Y} \quad (3.2.2)$$

Las cuales pueden resolverse mediante cualquier método apropiado para resolver sistemas de ecuaciones lineales, para obtener de esta manera la estimación del modelo.

La solución única de $\hat{\beta}$ es dada por:

$$\hat{\beta} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y} \quad (3.2.3)$$

Ya que, en este caso, $(\mathbf{X}^t\mathbf{X})^{-1}$ existe.

Así, el vector estimado $\hat{\mathbf{Y}}$ viene dado por $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} = \mathbf{X}\mathbf{X}^{-}\mathbf{Y}$, el cual es el vector proyección de \mathbf{Y} sobre el espacio Θ generado por los vectores columnas de la matriz \mathbf{X} .

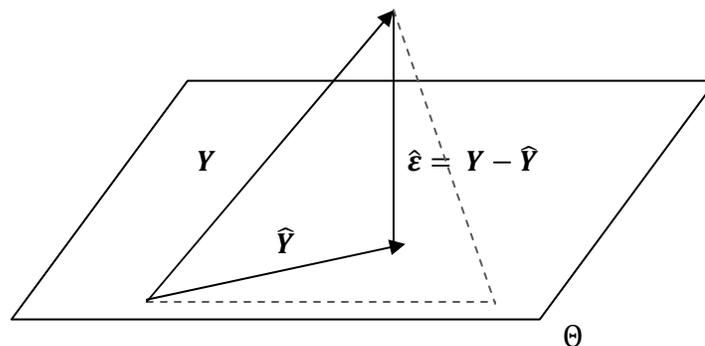


Figura 1. Proyección de \mathbf{Y} sobre el espacio Θ generado por los vectores columnas de la matriz \mathbf{X} . Fuente propia.

Como $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}}$ es ortogonal a Θ , se verifica que $\boldsymbol{\varepsilon}^t\hat{\mathbf{Y}} = \mathbf{Y}^t\mathbf{0}\mathbf{Y} = 0$.

La estimación MCO es equivalente hallar la proyección ortogonal \mathbf{Y} de $\hat{\mathbf{Y}}$ sobre Θ , es decir, que la norma euclídea de $\hat{\boldsymbol{\varepsilon}}$ sea mínima, ya que la norma del vector estimado $\hat{\boldsymbol{\varepsilon}}$ representa la distancia del vector \mathbf{Y} sobre el espacio Θ ,

$$\|\hat{\boldsymbol{\varepsilon}}\|^2 = \hat{\boldsymbol{\varepsilon}}^t \hat{\boldsymbol{\varepsilon}} = [(\mathbf{Y} - \hat{\mathbf{Y}})^t (\mathbf{Y} - \hat{\mathbf{Y}})] = [\mathbf{Y} - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}]^t [\mathbf{Y} - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}]$$

$$\|\hat{\boldsymbol{\varepsilon}}\|^2 = [\mathbf{Y}^t - \mathbf{Y}^t \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t] [\mathbf{Y} - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}]$$

$$\|\hat{\boldsymbol{\varepsilon}}\|^2 = \mathbf{Y}^t [\mathbf{I} - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t] [\mathbf{I} - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t] \mathbf{Y}$$

$$\|\hat{\boldsymbol{\varepsilon}}\|^2 = \mathbf{Y}^t [\mathbf{I} - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t] \mathbf{Y}$$

Dividiendo $\|\hat{\boldsymbol{\varepsilon}}\|^2$ por n se obtiene el estimador del parámetro σ^2 ;

$$\hat{\sigma}^2 = \frac{1}{n} \mathbf{Y}^t [\mathbf{I} - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t] \mathbf{Y} = \frac{1}{n} \hat{\boldsymbol{\varepsilon}}^t \hat{\boldsymbol{\varepsilon}} \quad (3.2.4)$$

(b) Método de máxima verosimilitud.

Hasta ahora se ha visto el método de estimación de mínimos cuadrados ordinarios, que consiste en asignar los parámetros desconocidos de manera que la suma cuadrada del error sea lo menos posible y que la única hipótesis necesaria para garantizar la existencia de los estimadores sea que la matriz $\mathbf{X}^t \mathbf{X}$ debe ser invertible. A continuación se presentará el método de máxima verosimilitud. Es un método de estimación que propone como estimador del parámetro, el valor que maximiza la probabilidad de obtener las observaciones muestrales disponibles. Fue recomendado, analizado y popularizado por R. A. Fisher entre 1912 y 1922 (Ferguson, 1996), aunque había sido utilizado antes por Gauss, Laplace, Thiele y F. Y. Edgeworth (Hald, 1998).

La diferencia con el método de estimación mínimos cuadrados ordinarios es que la estimación por máxima verosimilitud se basa en la hipótesis que se establecen para la distribución de las variables aleatorias que aparecen en el modelo. El modelo que se utilizará es el expresado en (3.1.2), bajo las hipótesis expresadas anteriormente se tiene:

$$Y_{n \times 1} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_d)$$

Por lo tanto, la función de densidad de probabilidad conjunta del vector aleatorio $Y_{n \times 1}$ es:

$$N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_d) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^t (\sigma^2 \mathbf{I}_d)^{-1} (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}$$

$$N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_d) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^t (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}$$

Así que, la función de verosimilitud correspondiente, denotada por $L(\boldsymbol{\beta}, \sigma^2)$ es:

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^t (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}$$

Luego se determinan las expresiones de $\boldsymbol{\beta}_{p \times 1}$, y σ^2 que optimizan en el máximo de $L(\boldsymbol{\beta}, \sigma^2)$. Para facilitar el proceso de derivación y resolución del sistema homogéneo que se ha de plantear, se optimiza al logaritmo neperiano de $L(\boldsymbol{\beta}, \sigma^2)$ (Ec. 3.2.5), la cual es una función monótona, continua y derivable y se optimiza en los mismos puntos de donde se ha de optimizar $L(\boldsymbol{\beta}, \sigma^2)$,

$$\ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad (3.2.5)$$

Luego derivando e igualando a cero la expresión anterior con respecto al vector $\boldsymbol{\beta}_{p \times 1}$, se obtiene (para más detalles ver Graybill, 1976):

$$(\mathbf{X}^t\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^t\mathbf{Y}$$

El mismo sistema de ecuaciones expresado en (3.1.4).

La solución de este sistema nos conduce al estimador de $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \quad (3.2.6)$$

Ahora derivando (3.2.5) con respecto a σ^2 e igualando a cero y despejando, se obtiene el estimador para σ^2 ,

$$\hat{\sigma}^2 = \frac{1}{n}[\mathbf{Y}^t\mathbf{Y} - 2\hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{Y} + \hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{X}\hat{\boldsymbol{\beta}}]$$

$$\hat{\sigma}^2 = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}^t\hat{\boldsymbol{\varepsilon}} \quad (3.2.7)$$

De las ecuaciones 3.2.3, 3.2.4, 3.2.6 y 3.2.7, se puede concluir que bajo el supuesto de normalidad la solución al sistema de ecuaciones $(\mathbf{X}^t\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^t\mathbf{Y}$, bien sea utilizando la inversa generalizada o la inversa condicional, va a coincidir con las soluciones vía mínimos

cuadrados y máxima verosimilitud, debido a que éstas son la misma inversa ordinaria de $(X^t X)$, ya que esta es una matriz de rango completo, el cual es uno de los objetivos.

(c) Otro método.

Si la matriz $X^t X$ es de rango completo, el sistema (3.1.4) tiene solución única. Se tomará $S = X^t X$ y $s = X^t Y$, así (3.1.4) puede ser denotado como:

$$S\hat{\beta} = s, \quad (3.2.8)$$

donde, S es una matriz definida positiva. El método para resolverlo es el siguiente, (Graybill, 1976):

Se factoriza S como producto de dos matrices $S = AB$, donde A es una matriz triangular inferior y B es una matriz triangular superior (lo que se conoce como descomposición LU), por lo tanto el sistema (3.2.8) puede ser escrito como:

$$(AB)\hat{\beta} = s, \quad \text{lo que es igual a}$$

$$A(B\hat{\beta}) = s$$

Haciendo $z = B\hat{\beta}$ se tiene

$$Az = s \quad (3.2.9)$$

Como A está en la forma triangular inferior, (3.2.9) se resuelve directamente para z y como B es triangular superior se resuelve el siguiente sistema

$$B\hat{\beta} = z \quad (3.2.10)$$

Claramente la resolución de ambos sistemas (3.2.9) y (3.2.10) es sencilla por tratarse de sistemas triangulares y requiere menos operaciones que resolver $X^t X \hat{\beta} = X^t Y$ cuya solución es $\hat{\beta} = (X^t X)^{-1} X^t Y$.

Distribución de los estimadores.

De acuerdo a que el modelo está correctamente especificado la expresión $\hat{\beta} = (X^t X)^{-1} X^t Y$ puede escribirse como:

$$\hat{\beta} = (X^t X)^{-1} X^t Y = (X^t X)^{-1} X^t (X\beta + \varepsilon) = \beta + (X^t X)^{-1} X^t \varepsilon \quad (3.2.11)$$

El cual cumple con:

- i. La distribución del estimador es normal, ya que $\hat{\beta}$ es una función lineal normal determinista de una variable aleatoria normal.
- ii. El estimador es insesgado. Aplicando el operador de esperanza a ambos lados en (3.2.11) se obtiene $E(\hat{\beta}) = \beta + (X^t X)^{-1} X^t E(\varepsilon) = \beta$

Por tanto, β es un vector determinista, pero su estimador por MCO es un vector de variables aleatorias normales, centradas en el valor que se quiere estimar. Para caracterizar completamente la distribución del estimador, es necesario obtener su matriz de covarianzas. A partir de (3.2.11) se tiene:

$$\begin{aligned} cov(\hat{\beta}) &= E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t\right] = E((X^t X)^{-1} X^t \varepsilon \varepsilon^t X (X^t X)^{-1}) \\ &= (X^t X)^{-1} X^t E(\varepsilon \varepsilon^t) X (X^t X)^{-1} \\ &= \sigma^2 (X^t X)^{-1} \end{aligned}$$

Así la matriz de covarianzas del estimador $\hat{\beta}$ puede estimarse usando la expresión:

$$cov(\hat{\beta}) = \hat{\sigma}^2 (X^t X)^{-1},$$

y si, $E(\hat{\sigma}^2) = \sigma^2$ y los valores de X son fijos, esta estimación será insesgada.

Así,

$$\hat{\beta}_{p \times 1} \sim N(\beta_{p \times 1}, \hat{\sigma}^2 (X^t X)^{-1})$$

Con respecto al estimador $\hat{\sigma}^2 = \frac{1}{n} Y^t [I - X(X^t X)^{-1} X^t] Y$, se tiene que sigue una distribución Ji-cuadrada siempre que $\frac{1}{n} [I - X(X^t X)^{-1} X^t] (\sigma^2 I_d)$ sea idempotente, de acuerdo con Graybill (1976), pero:

$$\left[\frac{1}{n} [I - X(X^t X)^{-1} X^t] (\sigma^2 I_d) \right] \left[\frac{1}{n} [I_d - X(X^t X)^{-1} X^t] (\sigma^2 I_d) \right]$$

$$= \frac{1}{n^2} (\sigma^2)^2 [I_d - X(X^t X)^{-1} X^t], \text{ ya que } [I_d - X(X^t X)^{-1} X^t] \text{ es una}$$

matriz idempotente. Es decir, no cumple la condición.

Sin embargo, la variable $\frac{n\hat{\sigma}^2}{\sigma^2}$ sigue una distribución Ji-cuadrada con grados de libertad $\text{rang}[I_d - X(X^t X)^{-1} X^t] = n - p$ y parámetro de centralidad dado por:

$$\lambda = \frac{1}{2\sigma^2} E(Y^t) [I_d - X(X^t X)^{-1} X^t] E(Y) = \frac{1}{2\sigma^2} \beta^t X^t [I_d - X(X^t X)^{-1} X^t] X \beta$$

$$\lambda = \frac{1}{2\sigma^2} \beta^t X^t [X \beta - X(X^t X)^{-1} X^t X \beta] = \frac{1}{2\sigma^2} \beta^t X^t [X \beta - X \beta] = 0$$

Así,

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p),0}$$

Por otra parte, ya que $E\left[\frac{n\hat{\sigma}^2}{\sigma^2}\right] = n - p$, se tiene que $E(\hat{\sigma}^2) = \frac{\sigma^2(n-p)}{n}$, lo que dice que el estimador $\hat{\sigma}^2 = \frac{1}{n} \hat{\boldsymbol{\varepsilon}}^t \hat{\boldsymbol{\varepsilon}}$ es sesgado. Por lo tanto, para estimar la varianza del término de error puede usarse la expresión:

$$\hat{\sigma}^2 = \frac{n}{(n-p)} \cdot \frac{\hat{\boldsymbol{\varepsilon}}^t \hat{\boldsymbol{\varepsilon}}}{n} = \frac{1}{n-p} \hat{\boldsymbol{\varepsilon}}^t \hat{\boldsymbol{\varepsilon}}$$

La cual produce estimaciones insesgadas. Además se tiene que $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes, pues:

$cov(\hat{\beta}, \hat{\sigma}^2) = E(\hat{\sigma}^2 \hat{\beta}) - E(\hat{\beta})E(\hat{\sigma}^2) = \hat{\sigma}^2 E(\hat{\beta}) - E(\hat{\beta})\hat{\sigma}^2 = 0$, ya que la no correlación entre normales multivariantes implica su independencia.

Con todo esto, se ha probado el siguiente teorema:

Teorema 3.2.1

Sea $Y_{n \times 1} \sim N(X\beta, \sigma^2 I_d)$ con $\text{rang}(X) = p$. Entonces se tiene:

1. $\hat{\beta} = (X^t X)^{-1} X^t Y$ es el estimador insesgado del vector de parámetros β .
2. $\hat{\sigma}^2 = \frac{1}{n-p} \hat{\epsilon}^t \hat{\epsilon}$ es el estimador insesgado del vector de parámetro σ^2 .
3. $\hat{\beta}_{p \times 1} \sim N(\beta_{p \times 1}, \hat{\sigma}^2 (X^t X)^{-1})$
4. $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p), 0}$
5. $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes.

3.3. MODELO LINEAL DE RANGO INCOMPLETO

El modelo lineal de rango incompleto corresponde al modelo expuesto en (3.1.2), con las mismas características, donde el rango de la matriz $X_{n \times p}$ es k , con $n > p > k$. En este caso los p vectores columnas de X son linealmente dependientes.

Cuando la matriz X no es de rango máximo, se tiene que $X^t X$ es singular y no es posible obtener una inversa ordinaria. Ya se sabe que la solución puede ser la utilización de la inversa generalizada, cuya solución no es única debido a que la matriz de coeficientes es de rango incompleto. En el caso de la regresión múltiple es difícil, aunque no imposible, que alguna columna sea linealmente dependiente de las demás. Si ocurriera esto se dice que existe colinealidad entre las columnas de X . Sin embargo, el término colinealidad o multicolinealidad se refiere al caso, mucho más frecuente, de que la dependencia entre las columnas no es exacta sino aproximada, es decir, a la quasi-dependencia lineal entre las variables regresoras.

La multicolinealidad es una condición intrínseca de los datos y esta puede provocar problemas de computación de los parámetros y en el cálculo de la precisión de los mismos, entre ellos, los estimadores por mínimos cuadrados son indeterminados y sus varianzas tienden a ser grandes (Gujarati, 1992). Debido al gran tamaño de los errores estándar, los intervalos de confianza para los parámetros tienden a ser imprecisos. Para casos con severa multicolinealidad, las cifras muestrales pueden ser compatibles con un conjunto de diversas hipótesis, por lo que la probabilidad de aceptar una hipótesis falsa aumenta. Por lo tanto, los estimadores de los coeficientes y sus errores estándar se vuelven muy sensibles incluso a pequeños cambios en los datos, (Mansfield & Helms, 1982).

Azocar, R (1986), estudió el problema de la multicolinealidad en casos de las investigaciones agrícolas. Encontró que los estimadores mínimos cuadráticos presentaron altas varianzas y una baja precisión en comparación con los estimadores obtenidos por medio del análisis de componentes principales. El análisis de componentes principales redujo la multicolinealidad presente en los datos y permitió una mejor interpretación de los resultados al hacer explícitas las dependencias lineales existentes entre las variables regresoras.

Flores, J (2001), aplicó el procedimiento de selección de variables Press-Ridge en la estimación de una variable repuesta en función de tres variables regresoras colineales. Las estimaciones de los parámetros por mínimos cuadrados ordinarios resultaron en valores enormemente alejados de los valores verdaderos, con errores estándar más de veinte veces mayor que la varianza del error y en todos los casos, para el modelo con todas las regresoras, los signos de las estimaciones resultaron contrarios a los correspondientes a los parámetros. Concluyó que el modelo seleccionado por el Press-Ridge tiene una suma de cuadrados de predicción inferior incluso al mejor de los modelos obtenidos mediante el método mínimos cuadrados ordinarios y sus coeficientes están menos alejados de los valores verdaderos.

Entre las múltiples formas de detección de la multicolinealidad se puede señalar:

1.- Examen de la matriz de correlación, si las variables regresoras X_i y X_j están correlacionadas entonces los elementos r_{ij} de la matriz $\mathbf{X}^t\mathbf{X}$ en forma de correlación están próximos a la unidad. También se ha planteado que si alguno de los coeficientes de correlación simples es mayor que el coeficiente de determinación del modelo, entonces existen razones para pensar que la multicolinealidad está presente.

2.- Factores infladores de la varianza, se estudian los elementos r_{ij}^{-1} de la diagonal principal de la matriz $(\mathbf{X}^t\mathbf{X})^{-1}$, que son los factores infladores de la varianza, (VIF). Algunos autores señalan que si dos o más VFI son mayores o iguales a 10, entonces se evidencia la existencia de multicolinealidad entre las variables regresoras.

3.- Valores propios de la matriz $\mathbf{X}^t\mathbf{X}$, si existen dependencias lineales entre las variables regresoras, entonces los autovalores de $\mathbf{X}^t\mathbf{X}$ estarán muy próximos a cero. También se puede estudiar la razón $K(x) = \lambda_{max}/\lambda_{min}$, que no es más que el valor propio máximo entre el valor propio mínimo. Se han propuestos rangos que permiten aproximar la severidad de la multicolinealidad: si $0 < K(x) < 100$ no existen problemas serios de multicolinealidad, si $100 < K(x) < 1000$ se considera de moderada a fuerte y si $K(x) > 1000$ existe severa multicolinealidad, (Montgomery & Peck, 1982).

Por lo tanto, en presencia de multicolinealidad se tiene que el proceso de estimación de parámetros, conduce al sistema de ecuaciones consistente $\mathbf{X}^t \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{Y}$, del cual se tiene que la matriz $\mathbf{X}^t \mathbf{X}$ es de rango $k < p$, pues \mathbf{X} es de rango $k < p$, e infinitas soluciones para el vector de parámetros $\boldsymbol{\beta}_{p \times 1}$. Si denotamos por $\hat{\boldsymbol{\beta}}$ a la estimación del parámetro $\boldsymbol{\beta}$ se tiene que la solución general al sistema de ecuaciones consistente señalado arriba, según el Teorema 2.2.22, viene dada por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-} \mathbf{X}^t \mathbf{Y} + [\mathbf{I}_d - (\mathbf{X}^t \mathbf{X})^{-} (\mathbf{X}^t \mathbf{X})] \mathbf{b}, \quad \mathbf{b} \in E_p, \quad (3.3.1)$$

O también (Corolario 2.3.1),

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^c \mathbf{X}^t \mathbf{Y} + [\mathbf{I}_d - (\mathbf{X}^t \mathbf{X})^c (\mathbf{X}^t \mathbf{X})] \mathbf{b}, \quad \mathbf{b} \in E_p, \quad (3.3.2)$$

Bien sea considerando la inversa generalizada de Penrose $(\mathbf{X}^t \mathbf{X})^{-}$ o la inversa condicional $(\mathbf{X}^t \mathbf{X})^c$ de la matriz $(\mathbf{X}^t \mathbf{X})$.

Siendo el estimador de σ^2 :

$$\hat{\sigma}^2 = (\mathbf{Y}^t \mathbf{Y} - \hat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{Y})$$

Sustituyendo $\hat{\boldsymbol{\beta}}^t$ por las soluciones señaladas arriba, se tiene que:

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{Y}^t \mathbf{Y} - \mathbf{Y}^t \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-} \mathbf{X}^t \mathbf{Y}) = \frac{1}{n} (\mathbf{Y}^t \mathbf{Y} - \hat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{Y}) \quad (3.3.3)$$

O también:

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{Y}^t \mathbf{Y} - \mathbf{Y}^t \mathbf{X} (\mathbf{X}^t \mathbf{X})^c \mathbf{X}^t \mathbf{Y}) = \frac{1}{n} \mathbf{Y}^t [\mathbf{I} - \mathbf{X} (\mathbf{X}^t \mathbf{X})^c \mathbf{X}^t] \mathbf{Y} \quad (3.3.4)$$

Así, se tienen infinitas soluciones de $\hat{\beta}$ y cada solución maximiza la función de verosimilitud, lo cual implica que no hay un único estimador de máxima verosimilitud para los parámetros de β y σ^2 en el modelo.

Este es el proceso de estimación que se va a realizar en este trabajo para el Modelo Lineal estadístico de Rango Incompleto, comparándolo con otros tres enfoques descritos abajo, a través del coeficiente de determinación y los intervalos de confianza obtenidos en cada caso.

a. Método del Modelo reducido.

Sea X_1 la matriz $n \times k$ con las $k = \text{rang}(X)$ columnas linealmente independientes de la matriz de diseño X , entonces $X = (X_1 | X_2)$, donde $X_2 = X_1 V$, ya que las columnas de X_2 son linealmente dependientes de las de X_1 , por lo que existe una matriz V que hace esto posible y por lo tanto, $X = X_1(I_d | V)$. Así, si $\beta = (\beta_1 \beta_2)^t$ se tiene

$$X\beta = X_1\beta_1 + X_2\beta_2 = X_1(\beta_1 + V\beta_2) = X_1\beta^*$$

El modelo puede ser reescrito como: $Y = X\beta + \varepsilon = X_1\beta^* + \varepsilon$ que es un modelo de rango completo sobre X_1 equivalente sobre el modelo de rango incompleto

$$\hat{\beta}^* = \hat{\beta}_1 + V\hat{\beta}_2 = (X_1^t X_1)^{-1} X_1^t Y$$

$$E(\hat{\beta}^*) = E((X_1^t X_1)^{-1} X_1^t Y) = E[(X_1^t X_1)^{-1} X_1^t (X_1\beta^* + \varepsilon)] = E(\beta^* + \varepsilon) = \beta^*$$

$$\text{cov}(\widehat{\boldsymbol{\beta}}^*) = E\left[(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*)^t\right] = E\left((\mathbf{X}_1^t \mathbf{X}_1)^{-1} \mathbf{X}_1^t \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^t \mathbf{X}_1 (\mathbf{X}_1^t \mathbf{X}_1)^{-1}\right)$$

$$\text{cov}(\widehat{\boldsymbol{\beta}}^*) = (\mathbf{X}_1^t \mathbf{X}_1)^{-1} \mathbf{X}_1^t E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^t) \mathbf{X}_1 (\mathbf{X}_1^t \mathbf{X}_1)^{-1}$$

$$\text{cov}(\widehat{\boldsymbol{\beta}}^*) = \sigma^2 (\mathbf{X}_1^t \mathbf{X}_1)^{-1}$$

b. Método de Funciones estimables.

Cuando el rango de \mathbf{X} es estrictamente menor que p , el sistema de ecuaciones lineales $\mathbf{X}^t \mathbf{X} \widehat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{Y}$ admite una infinidad de soluciones. En este caso no todo es estimable, es decir, hay ciertas funciones paramétricas lineales para las cuales no es posible construir un estimador lineal insesgado. Tal observación condujo, de acuerdo con Scheffé (1959), al matemático hindú R. C. Bose, a la siguiente definición:

Definición 3.3.1.

En el modelo de rango incompleto $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, una función lineal del vector de parámetros $\ell_1\beta_1 + \ell_2\beta_2 + \dots + \ell_p\beta_p = \boldsymbol{\ell}^t \boldsymbol{\beta}$, se dice que es estimable si, y solo si, existe una función lineal del vector aleatorio \mathbf{Y} , $c_1Y_1 + c_2Y_2 + \dots + c_nY_n = \mathbf{c}^t \mathbf{Y}$, el cual es un estimador insesgado de $\boldsymbol{\ell}^t \boldsymbol{\beta}$; es decir, $E(\mathbf{c}^t \mathbf{Y}) = \boldsymbol{\ell}^t \boldsymbol{\beta}$. Si tal vector \mathbf{c} no existe, entonces se dice que $\boldsymbol{\ell}^t \boldsymbol{\beta}$ es linealmente no estimable.

Con esta definición, se tiene entonces que todas y cada una de las ecuaciones lineales de los parámetros que intervienen en la parte izquierda del sistema de ecuaciones

$(\mathbf{X}^t\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^t\mathbf{Y}$, son todas “estimables”. Observe, $(\mathbf{X}^t\mathbf{X})\boldsymbol{\beta}$ contiene p ecuaciones lineales de los parámetros $\boldsymbol{\beta}$'s, si se le coloca a cada parámetro la “tilde” de “estimación”, esto es: “ $\hat{\boldsymbol{\beta}}$ ”, sería igual a una función lineal del vector aleatorio $\mathbf{Y}_{n \times 1}$, ya que $(\mathbf{X}^t\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^t\mathbf{Y}$.

Se probarán a continuación algunas condiciones para que $\boldsymbol{\ell}^t\boldsymbol{\beta}$ sea estimable, los detalles referentes a las demostraciones ver Graybill (Graybill, 1976).

Teorema 3.3.1.

Considere el modelo de rango incompleto $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Entonces una función lineal de $\boldsymbol{\beta}$, $\boldsymbol{\ell}^t\boldsymbol{\beta}$, es estimable si, y sólo si, existe un vector \mathbf{c} de orden $n \times 1$ tal que $\boldsymbol{\ell} = \mathbf{X}^t\mathbf{c}$.

Teorema 3.3.2.

Considere el modelo de rango incompleto $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Entonces una función lineal de $\boldsymbol{\beta}$, $\boldsymbol{\ell}^t\boldsymbol{\beta}$, es estimable si, y sólo si, se cumplen las siguientes condiciones:

- i. $\boldsymbol{\ell}$ es una combinación lineal de las columnas de \mathbf{X}^t .
- ii. $\text{rang}(\mathbf{X}^t, \boldsymbol{\ell}) = \text{rang}(\mathbf{X}^t)$
- iii. $\text{rang}(\mathbf{X}^t\mathbf{X}, \boldsymbol{\ell}) = \text{rang}(\mathbf{X}^t\mathbf{X})$
- iv. $\boldsymbol{\ell}^t\mathbf{X}^{-}\mathbf{X} = \boldsymbol{\ell}^t$ para una inversa generalizada de \mathbf{X}
- v. $\mathbf{X}^t(\mathbf{X}^t)^{-}\boldsymbol{\ell} = \boldsymbol{\ell}$ para una inversa generalizada de \mathbf{X}^t

Definición 3.3.2.

Un conjunto m de funciones lineales de β ; sea $\{\ell_1^t \beta, \ell_2^t \beta, \dots, \ell_m^t \beta\}$ se define como un conjunto de m funciones lineales de β estimables linealmente independientes si, y sólo si, se satisface que:

- i. $\ell_i^t \beta$ son estimables para cada $i = 1, 2, \dots, m$
- ii. $\{\ell_1^t, \ell_2^t, \dots, \ell_m^t\}$ son vectores $p \times 1$ linealmente independientes

Teorema 3.3.3.

El número de funciones estimables linealmente independientes de β es igual al rango de X , el cual es k .

Definición 3.3.3.

El conjunto $\{\ell_1^t \beta, \ell_2^t \beta, \dots, \ell_m^t \beta\}$ conforma un conjunto base de funciones lineales de β estimables si, y sólo si, se satisface que:

- i. $\ell_i^t \beta$ son linealmente independientes para cada $i = 1, 2, \dots, m$
- ii. $\text{rang}(X) = k = m$

Se sabe que cuando la matriz de diseño es de rango incompleto la estimación MC de los parámetros no es única pero la estimación de cualquier función paramétrica estimable

utilizando cualquiera de los estimadores MC si es única como lo formaliza el siguiente teorema.

Teorema 3.3.4.

Considere el modelo de rango incompleto $Y = X\beta + \varepsilon$ tal que $E(\varepsilon_{n \times 1}) = \mathbf{0}$, $V(\varepsilon_{n \times 1}) = \sigma^2 \mathbf{I}_d$ y el sistema de ecuaciones $X^t X \hat{\beta} = X^t Y$. Entonces:

- i. Si $\ell^t \beta$ es estimable entonces $\ell^t \hat{\beta}$ es invariante para alguna solución $\hat{\beta}$ de las ecuaciones dadas y $\ell^t \hat{\beta} = \ell^t X^{-} Y$.
- ii. Si $\ell^t \beta$ es estimable entonces $\ell^t \hat{\beta}$ es el mejor estimador de $\ell^t \beta$.
- iii. $\hat{\sigma}^2 = \frac{1}{n-k} (Y^t Y - \hat{\beta}^t X^t Y)$ es invariante para alguna solución $\hat{\beta}$ de las ecuaciones dadas.
- iv. $\hat{\sigma}^2$ es un estimador insesgado de σ^2

Entonces, a partir del sistema de ecuaciones $(X^t X) \hat{\beta} = X^t Y$, se tienen p ecuaciones lineales de los parámetros β 's estimables, pero todas ellas en conjunto no son linealmente independientes, se debe determinar un conjunto base de funciones lineales estimables, conjunto que debe tener exactamente k funciones lineales estimables, ya que el rango de X es k . A partir de aquí, hallar una solución al sistema.

c. Método de Imposición de restricciones.

Este método consiste en imponer un conjunto de restricciones sobre el vector de parámetros β del tipo $V\beta = \mathbf{0}$, agregando al sistema de ecuaciones original y así obtener un nuevo sistema de ecuaciones “ampliado” cuya matriz de coeficiente es de rango completo. Las restricciones apropiadas, llamadas identificables, son aquellas que, para cada X^tY , existe un único $\hat{\beta}$ que satisface $(X^tX)\beta = X^tY$ y $V\beta = \mathbf{0}$, es decir, que satisface:

$$\begin{bmatrix} X^tX & V^t \\ V & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} X^tY \\ \mathbf{0} \end{bmatrix}$$

La solución es simple. Se debe elegir como filas de V un conjunto de $p - k$ vectores v_i de dimensión $p \times 1$ linealmente independientes que sean también linealmente independientes de las filas de X . Así, la matriz $H = \begin{bmatrix} X^tX & V^t \\ V & \mathbf{0} \end{bmatrix}$ de orden $(2p - k) \times (2p - k)$ tendrá rango $(2p - k)$ por lo tanto tiene inversa ordinaria y produce soluciones únicas.

Dichas soluciones vienen dadas por:

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

Este método es muy útil en los modelos de análisis de la varianza para los que V se halla con mucha facilidad. La situación acá es que se imponen restricciones sobre los parámetros que no son funciones lineales estimables.

Ahora bien, se va a plantear un modelo de rango incompleto, en el cual las soluciones al sistema de ecuaciones normales $(X^t X)\hat{\beta} = X^t Y$ y al parámetro σ^2 , se van a realizar utilizando el método de modelo reducido, imponiendo restricciones, método de funciones estimables, la inversa generalizada y la inversa condicional. Dichos métodos se van a comparar entre sí por medio del coeficiente de determinación del modelo estimado (Capítulo 4) y por medio de los intervalos de confianza para cada β_j con un nivel de confianza del 95% ($\alpha = 0,05$) mediante la fórmula:

$$I(\beta)_j = \left(\hat{\beta}_j - t_{n-p}^{\alpha/2} \sqrt{V(\hat{\beta}_j)}, \hat{\beta}_j + t_{n-p}^{\alpha/2} \sqrt{V(\hat{\beta}_j)} \right), \text{ donde, } t_{n-p}^{\alpha/2} = t_{78}^{0,025} = 1,9886.$$

Capítulo 4.

Coeficiente de Determinación.

4.1. DEFINICIONES BÁSICAS.

La varianza de los términos de perturbación, σ^2 , es una medida de la importancia que tienen los factores no considerados en el modelo, en las variaciones de la variable dependiente. Un valor alto de la varianza indicaría que los valores poblacionales de la variable dependiente están alejados de los valores teóricos que supuestamente debería tener, como consecuencia de la influencia de las variables explicativas incluidas en el modelo. Esta situación se refleja en los valores muestrales lo que ocasiona que los datos muestrales no se ajusten bien al modelo teórico, los residuos serán grandes y s^2 (estimados de σ^2) también será grande.

La magnitud de σ^2 como medida de habilidad del modelo para explicar el comportamiento de la variable dependiente, y por ende, la de s^2 como medida de la “bondad del ajuste” no son completamente adecuadas. Cuando se califica s^2 como grande o pequeña, cabría preguntarse con respecto a que es grande o pequeña. Los valores de σ^2 y s^2 están afectados por la unidad de medida de la variable dependiente. Además, si se

quiere comparar el grado de variación o la bondad del ajuste en distintos modelos se necesita tener una medida *standard* que pueda usarse como indicador relativo de comparación.

Esta medida, para los datos poblacionales, la constituye el coeficiente de correlación múltiple o su cuadrado, el coeficiente de determinación múltiple. Está definido como el porcentaje de variación de la variable dependiente explicado por la influencia lineal de las variables independientes. Este coeficiente es un parámetro poblacional desconocido. Para determinar la bondad del ajuste a partir de los datos muestrales se utiliza un estimador del coeficiente, al considerar el modelo ajustado como un estimador de la verdadera influencia que las variables explicativas tienen en la variable dependiente, el estimador que se usa es R^2 , conocido como coeficiente de determinación múltiple muestral,

$$R^2 = 1 - \frac{SCE}{SCT}$$

donde,

SCE = Suma de los cuadrados del error o suma de los cuadrados no explicada por el modelo,

$$SCE = \mathbf{e} \mathbf{e}^t = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

SCT = Suma de cuadrados total,

$$SCT = \sum_{i=1}^n (y_i - \bar{Y})^2$$

R^2 es un número comprendido entre cero y uno, un R^2 igual a uno, es decir, $SCE = 0$ si $y_i = \hat{y}_i$, implicaría un ajuste perfecto de los datos al modelo, mientras que, en el extremo opuesto, un R^2 igual a cero, es decir, $SCE = SCT$, mostraría que no hay ninguna relación aparente entre la variable dependiente y las variables independientes, valores intermedios de R^2 indican una mayor o menor relación aparente entre la variable explicada y las variables explicativas.

Ahora bien, dado que $SCT = SCR + SCE$, donde SCR es la suma explicada de los cuadrados por el modelo o suma de cuadrados de regresión, se tiene que (para más detalles ver López, 2006):

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{Y})^2 &= \sum_{i=1}^n (\hat{y}_i - \bar{Y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ SCT &= \sum_{i=1}^n (y_i - \bar{Y})^2 = \mathbf{Y}^t \mathbf{Y} - \frac{(\sum_{i=1}^n y_i)^2}{n} = \mathbf{Y}^t \mathbf{Y} - n\bar{Y}^2 \\ SCR &= \sum_{i=1}^n (\hat{y}_i - \bar{Y})^2 = \hat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{Y} - \frac{(\sum_{i=1}^n y_i)^2}{n} = \hat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{Y} - n\bar{Y}^2 \\ SCE &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = SCT - SCR = \mathbf{Y}^t \mathbf{Y} - \hat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{Y} \end{aligned}$$

Luego, si

$$R^2 = 1 - \frac{SCE}{SCT} = \frac{SCR}{SCT}$$

Se tiene que:

$$R^2 = \frac{\hat{\beta}^t X^t Y - n\bar{Y}^2}{Y^t Y - n\bar{Y}^2}$$

Dado que R^2 mide el porcentaje de variación de la variable dependiente atribuible a las variables explicativas, su interpretación debe hacerse con precaución. Un coeficiente de determinación alto, incluso igual a uno, no implica necesariamente que exista una relación teórica íntima entre las variables, si la alta correlación se debe a que exista una relación teórica o a la casualidad debe determinarse a priori desde el punto de vista teórico; esto implica que antes de intentar medir la relación entre las variables se debe analizar, comprender y racionalizar las hipótesis teóricas sobre estas relaciones, no se debe hacer análisis estadísticos de medición de relaciones entre variables sin que exista una base teórica lógica que justifique esa medición.

Un coeficiente de determinación muy pequeño, no implica necesariamente la inexistencia de relación entre la variable dependiente e independientes. Un R^2 cercano a cero, indica que la función utilizada se ajusta mal a los datos reales de las variables, lo cual podría

deberse a que las variables explicativas utilizadas son deficientes en el sentido de que sus variaciones no afectan a la variable que se trata de explicar.

En el próximo capítulo se calculará el coeficiente de determinación para cada uno de los enfoques señalados en los capítulos anteriores.

Capítulo 5.

Aplicación y Resultados.

5.1. PROBLEMA.

Los datos que se van a analizar corresponden a un Trabajo Especial de Grado titulado “Cambios en 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”, para optar al título de Ingeniero Agrónomo de la Universidad Nacional Experimental Francisco de Miranda que tiene como autor Vanessa Ruiz. Los datos son los siguientes:

	b_1	b_2	b_3
a_1	0,0073	2,4503	1,3119
	0,6959	2,7643	0,9909
	0,1522	0,7254	1,5114
	0,1934	1,257	1,1112
	0,1175	0,7909	0,84
	0,1483	2,2933	0,9157
	0,0086	0,4089	0,8955
	0,1341	0,5828	0,8289
	0,7628	1,2047	1,4583
	0,0786	0,1754	0,4243
	0,2682	1,5321	0,4364
	0,6905	3,5568	0,7118
	0,1419	0,5689	0,8104
	0,2615	2,6797	1,313
	0,169	2,8261	0,5736

	b_1	b_2	b_3
a_2	1,4972	1,4972	0,4713
	1,6496	1,6496	1,9746
	1,8567	1,8567	1,1794
	0,944	0,944	1,315
	1,7569	1,7569	0,6114
	1,1508	1,1508	1,0597
	1,6514	1,6514	0,6378
	1,7064	1,7064	1,0885
	1,4995	1,4995	1,6085
	1,2134	1,2134	1,7516
	1,7455	1,7455	1,2016
	1,4052	1,4052	3,1487
	1,047	1,047	0,113
	1,1548	1,1548	1,7728
	1,5669	1,5669	0,4695

Tabla 1. Datos para el modelo factorial 2×3 .

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 a_1 y a_2 , tres niveles para el segundo factor b_1 , b_2 y b_3 .

El modelo viene dado por:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_d), \quad (5.1)$$

donde,

$$\mu = [\mu], \quad \beta^t = [\mu \quad \tau^t \quad \alpha^t \quad (\tau\alpha)^t], \quad \tau^t = [\tau_1 \quad \tau_2], \quad \alpha^t = [\alpha_1 \quad \alpha_2 \quad \alpha_3],$$

$(\tau\alpha)^t = [(\tau\alpha)_{11} \quad (\tau\alpha)_{12} \quad (\tau\alpha)_{13} \quad (\tau\alpha)_{21} \quad (\tau\alpha)_{22} \quad (\tau\alpha)_{23}]$, X es la matriz de diseño.

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 información de los datos suele resumirse en la siguiente tabla:

Niveles del Factor A	Niveles del Factor B			Media
	1	2	3	
1	y_{111}	y_{121}	y_{131}	$\bar{y}_{1..}$
	y_{112}	y_{122}	y_{132}	
	\vdots	\vdots	\vdots	
	y_{1115}	y_{1215}	y_{1315}	
	$\bar{y}_{11.}$	$\bar{y}_{12.}$	$\bar{y}_{13.}$	

2	y_{211}	y_{221}	y_{231}	$\bar{y}_{2..}$
	y_{212}	y_{222}	y_{232}	
	\vdots	\vdots	\vdots	
	y_{2115}	y_{2215}	y_{2315}	
	$\bar{y}_{21.}$	$\bar{y}_{22.}$	$\bar{y}_{23.}$	
Media	$\bar{y}_{.1.}$	$\bar{y}_{.2.}$	$\bar{y}_{.3.}$	$\bar{y}_{..}$

Tabla 2. Datos para el diseño factorial 2×3 , con interacción.

Así, utilizando los valores de la tabla 1, se tiene:

Niveles del Factor A	Niveles del Factor B			Media
	1	2	3	
1	0,0073	2,4503	1,3119	0,92843
	0,6959	2,7643	0,9909	
	\vdots	\vdots	\vdots	
	0,169	2,8261	0,5736	
	$\bar{y}_{11.}$	$\bar{y}_{12.}$	$\bar{y}_{13.}$	
	0,25532	1,58777	0,94222	
2	1,4972	1,4972	0,4713	1,37986
	1,6496	1,6496	1,9746	
	\vdots	\vdots	\vdots	
	1,5669	1,5669	0,4695	
	$\bar{y}_{21.}$	$\bar{y}_{22.}$	$\bar{y}_{23.}$	
	1,45635	1,45635	1,22689	
Media	0,85583	1,52206	1,08455	1,15415

Tabla 3. Medias para los datos de la Tabla 1.

La matriz de diseño es:

$$\mathbf{X} = \begin{bmatrix}
 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & . & . & 0 & . & . & . & . & . & . \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 . & . & . & 1 & 0 & . & 1 & 0 & 0 & 0 & 0 & 0 \\
 . & . & . & 0 & 1 & . & 0 & 1 & 0 & 0 & 0 & 0 \\
 . & . & . & 0 & . & . & 0 & . & . & . & . & . \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 . & . & . & . & 1 & . & . & 1 & 0 & . & . & . \\
 . & . & . & . & 0 & 1 & . & 0 & 1 & . & . & . \\
 . & . & . & . & 0 & . & . & 0 & . & . & . & . \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 . & 1 & 0 & 0 & . & . & 0 & . & . & . & . & . \\
 . & 1 & 0 & . & . & 1 & . & . & 1 & 0 & . & . \\
 . & 0 & 1 & 1 & . & 0 & 0 & . & 0 & 1 & . & . \\
 . & 0 & 1 & . & . & 0 & . & . & 0 & . & . & . \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 . & . & . & . & 0 & . & . & . & . & . & . & . \\
 . & . & . & . & 1 & 0 & . & . & . & 1 & 0 & . \\
 . & . & . & . & 0 & 1 & . & . & . & 0 & 1 & . \\
 . & . & . & . & 0 & . & . & . & . & 0 & . & . \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 . & . & . & . & . & 0 & . & . & . & . & . & . \\
 . & . & . & . & 1 & 0 & . & . & . & . & 1 & . \\
 . & . & . & . & 0 & 1 & . & . & . & . & 0 & 1 \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 . & . & . & . & . & . & . & . & . & . & . & . \\
 1 & 0 & 1 & 0 & . & . & 0 & 0 & 0 & 0 & . & . \\
 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
 \end{bmatrix}_{90 \times 12}$$

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 $(\mathbf{X}^t \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{Y}$.

Ahora bien, como se está trabajando con un modelo de rango incompleto $(\mathbf{X}^t \mathbf{X})^{-1}$ 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 .

5.2. MODELO REDUCIDO.

Se tomará \mathbf{X}_1 la matriz de orden (90×6) con las $6 = \text{rang}(\mathbf{X})$ columnas linealmente independientes de la matriz de diseño \mathbf{X} y \mathbf{X}_2 la matriz de orden (90×6) formada con las demás columnas de \mathbf{X} , es decir,

$$\mathbf{X}_1 = [X_{i1} \ X_{i2} \ X_{i4} \ X_{i5} \ X_{i7} \ X_{i8}], \text{ para } i = 1, 2, \dots, 90.$$

$$\mathbf{X}_2 = [X_{i3} \ X_{i6} \ X_{i9} \ X_{i10} \ X_{i11} \ X_{i12}], \text{ para } i = 1, 2, \dots, 90.$$

$$\mathbf{X}_1 = \begin{bmatrix}
 1 & 1 & 1 & 0 & 1 & 0 \\
 1 & 1 & . & . & . & . \\
 . & . & . & . & . & . \\
 . & . & . & . & . & . \\
 . & . & 1 & 0 & 1 & 0 \\
 . & . & 0 & 1 & 0 & 1 \\
 . & . & 0 & . & 0 & . \\
 . & . & . & . & . & . \\
 . & . & . & . & . & . \\
 . & . & . & 1 & . & 1 \\
 . & . & . & 0 & . & 0 \\
 . & . & . & 0 & . & 0 \\
 . & . & . & . & . & . \\
 . & 1 & 0 & . & 0 & . \\
 . & 1 & . & . & . & . \\
 . & 0 & 1 & . & 0 & . \\
 . & 0 & . & . & . & . \\
 . & . & . & . & . & . \\
 . & . & . & 0 & . & . \\
 . & . & 1 & 0 & 0 & . \\
 . & . & 0 & 1 & 0 & . \\
 . & . & 0 & . & 0 & . \\
 . & . & . & . & . & . \\
 . & . & . & . & . & . \\
 . & . & . & 1 & . & . \\
 . & . & . & 0 & . & . \\
 . & . & . & . & . & . \\
 . & . & . & . & . & . \\
 1 & 0 & 0 & . & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0
 \end{bmatrix}_{90 \times 6}$$

Luego, la matriz \mathbf{V} de orden (6×6) , que cumpla con la condición $\mathbf{X}_2 = \mathbf{X}_1 \mathbf{V}$ viene dada

por:

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

Así $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 = \mathbf{X}_1(\boldsymbol{\beta}_1 + \mathbf{V}\boldsymbol{\beta}_2) = \mathbf{X}_1\boldsymbol{\beta}^*$, donde $\boldsymbol{\beta}_1$ de orden (6×1) y $\boldsymbol{\beta}_2$ de orden (6×1) .

Por lo tanto, el modelo $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$ es de rango completo por columnas y podemos estimar $\boldsymbol{\beta}^*$, donde $\hat{\boldsymbol{\beta}}^* = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Y}$.

Las matrices $(\mathbf{X}_1^t\mathbf{X}_1)$ y $(\mathbf{X}_1^t\mathbf{Y})$ vienen dadas por:

$$\mathbf{X}_1^t\mathbf{X}_1 = \begin{bmatrix} 90 & 45 & 30 & 30 & 15 & 15 \\ 45 & 45 & 15 & 15 & 15 & 15 \\ 30 & 15 & 30 & 0 & 15 & 0 \\ 30 & 15 & 0 & 30 & 0 & 15 \\ 15 & 15 & 15 & 0 & 15 & 0 \\ 15 & 15 & 0 & 15 & 0 & 15 \end{bmatrix}$$

$$\mathbf{X}_1^t\mathbf{Y} = \begin{bmatrix} 103,8737 \\ 49,4917 \\ 25,6751 \\ 35,9786 \\ 3,8298 \\ 21,8453 \end{bmatrix}$$

Donde,

$$(\mathbf{X}_1^t\mathbf{X}_1)^{-1} = \begin{bmatrix} 0,0667 & -0,0667 & -0,0667 & -0,0667 & 0,0667 & 0,0667 \\ -0,0667 & 0,1333 & 0,0667 & 0,0667 & -0,1333 & -0,1333 \\ -0,0667 & 0,0667 & 0,1333 & 0,0667 & -0,1333 & -0,0667 \\ -0,0667 & 0,0667 & 0,0667 & 0,1333 & -0,0667 & -0,1333 \\ 0,0667 & -0,1333 & -0,1333 & -0,0667 & 0,2667 & 0,1333 \\ 0,0667 & -0,1333 & -0,0667 & -0,0667 & 0,1333 & 0,2667 \end{bmatrix}$$

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 $\hat{Y}^* = X_1 \hat{\beta}^*$ es

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

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:

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

Donde,

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

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

$$n = 90$$

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

Luego,

$$R^2 = \frac{138,3178 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = \frac{18,4218}{47,7160} = 0,3860$$

Y para los intervalos de confianza se tiene

$$\hat{\sigma}^2 = \frac{1}{n - k} (Y^t Y - (\hat{\beta}^*)^t X_1^t Y)$$

$$\hat{\sigma}^2 = \frac{1}{90 - 6} (167,6120 - 138,3178)$$

$$\hat{\sigma}^2 = 0,348740.$$

Ahora se necesitan las varianzas de cada $(\hat{\beta}^*)_j$, $j = 1, 2, \dots, 6$. Para ello se calculará la covarianza de $\hat{\beta}^*$, la cual viene dada por:

$$\text{Cov}(\hat{\beta}^*) = \hat{\sigma}^2 (\mathbf{X}_1^t \mathbf{X}_1)^{-1}$$

Donde,

$$(\mathbf{X}_1^t \mathbf{X}_1)^{-1} = \begin{bmatrix} 0,0667 & -0,0667 & -0,0667 & -0,0667 & 0,0667 & 0,0667 \\ -0,0667 & 0,1333 & 0,0667 & 0,0667 & -0,1333 & -0,1333 \\ -0,0667 & 0,0667 & 0,1333 & 0,0667 & -0,1333 & -0,0667 \\ -0,0667 & 0,0667 & 0,0667 & 0,1333 & -0,0667 & -0,1333 \\ 0,0667 & -0,1333 & -0,1333 & -0,0667 & 0,2667 & 0,1333 \\ 0,0667 & -0,1333 & -0,0667 & -0,0667 & 0,1333 & 0,2667 \end{bmatrix}$$

Así,

$$\text{Cov}(\hat{\beta}^*) = \begin{bmatrix} \mathbf{0,0232} & -0,0232 & -0,0232 & -0,0232 & 0,0232 & 0,0232 \\ -0,0232 & \mathbf{0,0465} & 0,0232 & 0,0232 & -0,0465 & -0,0465 \\ -0,0232 & 0,0232 & \mathbf{0,0465} & 0,0232 & -0,0465 & -0,0232 \\ -0,0232 & 0,0232 & 0,0232 & \mathbf{0,0465} & -0,0232 & -0,0465 \\ 0,0232 & -0,0465 & -0,0465 & -0,0232 & \mathbf{0,0930} & 0,0465 \\ 0,0232 & -0,0465 & -0,0232 & -0,0465 & 0,0465 & \mathbf{0,0930} \end{bmatrix}$$

Luego:

$$I(\beta^*)_1 = (1,2269 - (1,9886)\sqrt{0,0232} ; 1,2269 + (1,9886)\sqrt{0,0232})$$

$$I(\beta^*)_1 = (0,9240 ; 1,5297), \quad L(I(\beta^*)_1) = 0,6057$$

$$I(\beta^*)_2 = (0,3609 - (1,9886)\sqrt{0,0465} ; 0,3609 + (1,9886)\sqrt{0,0465})$$

$$I(\beta^*)_2 = (-0,0679 ; 0,7897), \quad L(I(\beta^*)_2) = 0,8576$$

$$I(\beta^*)_3 = (0,2295 - (1,9886)\sqrt{0,0465} ; 0,2295 + (1,9886)\sqrt{0,0465})$$

$$I(\beta^*)_3 = (-0,1993 ; 0,6583), \quad L(I(\beta^*)_3) = 0,8576$$

$$I(\beta^*)_4 = (-0,2847 - (1,9886)\sqrt{0,0465} ; -0,2847 + (1,9886)\sqrt{0,0465})$$

$$I(\beta^*)_4 = (-0,7135 ; 0,1441), \quad L(I(\beta^*)_4) = 0,8576$$

$$I(\beta^*)_5 = (-1,5619 - (1,9886)\sqrt{0,0930} ; -1,5619 + (1,9886)\sqrt{0,0930})$$

$$I(\beta^*)_5 = (-2,1683 ; -0,9554), \quad L(I(\beta^*)_5) = 1,2129$$

$$I(\beta^*)_6 = (0,1533 - (1,9886)\sqrt{0,0930} ; 0,1533 + (1,9886)\sqrt{0,0930})$$

$$I(\beta^*)_6 = (-0,4531 ; 0,7597), \quad L(I(\beta^*)_6) = 1,2128$$

5.3. 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, $H = \begin{bmatrix} X^t X & V^t \\ V & \mathbf{0} \end{bmatrix}$ dada por:

$$\mathbf{H} = \begin{bmatrix}
 90 & 45 & 45 & 30 & 30 & 30 & 15 & 15 & 15 & 15 & 15 & 15 \\
 45 & 45 & 0 & 15 & 15 & 15 & 15 & 15 & 15 & 0 & 0 & 0 \\
 45 & 0 & 45 & 15 & 15 & 15 & 0 & 0 & 0 & 15 & 15 & 15 \\
 30 & 15 & 15 & 30 & 0 & 0 & 15 & 0 & 0 & 15 & 0 & 0 \\
 30 & 15 & 15 & 0 & 30 & 0 & 0 & 15 & 0 & 0 & 15 & 0 \\
 30 & 15 & 15 & 0 & 0 & 30 & 0 & 0 & 15 & 0 & 0 & 15 \\
 15 & 15 & 0 & 15 & 0 & 0 & 15 & 0 & 0 & 0 & 0 & 0 \\
 15 & 15 & 0 & 0 & 15 & 0 & 0 & 15 & 0 & 0 & 0 & 0 \\
 15 & 15 & 0 & 0 & 0 & 15 & 0 & 0 & 15 & 0 & 0 & 0 \\
 15 & 0 & 15 & 15 & 0 & 0 & 0 & 0 & 0 & 15 & 0 & 0 \\
 15 & 0 & 15 & 0 & 15 & 0 & 0 & 0 & 0 & 0 & 15 & 0 \\
 15 & 0 & 15 & 0 & 0 & 15 & 0 & 0 & 0 & 0 & 0 & 15 \\
 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0
 \end{bmatrix}$$

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

Es una matriz cuadrada de orden (18×18) , de rango completo por columnas y por ende invertible. Luego,

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

Donde,

$$\begin{bmatrix} X^t Y \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} 9868/95 \\ 8958/181 \\ 4840/89 \\ 6085/237 \\ 6728/187 \\ 2111/50 \\ 6323/1651 \\ 3954/181 \\ 5454/229 \\ 3954/181 \\ 28309/2003 \\ 3239/176 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Así,

$$\begin{bmatrix} \hat{\beta} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} 1,1542 \\ -0,0543 \\ 0,0543 \\ 0,2479 \\ -0,7358 \\ 0,4880 \\ -1,0924 \\ 1,0924 \\ 1,1370 \times 10^{-16} \\ -1,1370 \times 10^{-16} \\ 0,4696 \\ -0,4696 \\ 0 \\ -1,0658 \times 10^{-14} \\ 0 \\ -1,7764 \times 10^{-15} \\ 3,5527 \times 10^{-15} \\ 0 \end{bmatrix}$$

Es decir,

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

Luego,

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

En la cual, $y_{11} = y_{21} = \dots = y_{151} = 0,2553$, $y_{161} = y_{171} = \dots = y_{301} = 1,4564$,

$y_{311} = y_{321} = \dots = y_{451} = 1,5878$, $y_{461} = y_{471} = \dots = y_{601} = 1,4564$,

$y_{611} = y_{621} = \dots = y_{751} = 0,9422$, $y_{761} = y_{771} = \dots = y_{901} = 1,2269$

El coeficiente de determinación es:

$$R^2 = \frac{138,3178 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = \frac{18,4218}{47,7160} = 0,3860$$

Donde,

$$(\hat{\beta}_r)^t X^t Y = 138,3178$$

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

$$n = 90$$

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

Por otro lado,

$$\hat{\sigma}^2 = \frac{1}{n - k} (Y^t Y - (\hat{\beta}_r)^t X^t Y)$$

$$\hat{\sigma}^2 = \frac{1}{90 - 6} (167,6120 - 138,3178)$$

$$\hat{\sigma}^2 = 0,348740.$$

Por lo tanto,

$$Cov\left(\begin{bmatrix} \hat{\beta}_r \\ \mathbf{0} \end{bmatrix}\right) = \hat{\sigma}^2 \mathbf{H}^{-1}$$

En este caso la matriz de covarianza es una matriz de orden (18×18) , en la cual las varianzas de los $(\hat{\beta}_r)_j$, $j = 1, 2, \dots, 12$ son:

$$V(\hat{\beta}_r)_1 = V(\hat{\beta}_r)_2 = V(\hat{\beta}_r)_3 = 0,0039, \quad V(\hat{\beta}_r)_4 = 0,0155, \quad V(\hat{\beta}_r)_5 = 0,0310,$$

$$V(\hat{\beta}_r)_6 = 0,0155, \quad V(\hat{\beta}_r)_7 = V(\hat{\beta}_r)_8 = 0,0310, \quad V(\hat{\beta}_r)_9 = 5,3732 \times 10^{-34},$$

$$V(\hat{\beta}_r)_{10} = 0, \quad V(\hat{\beta}_r)_{11} = V(\hat{\beta}_r)_{12} = 0,0310$$

Así los intervalos de confianza son:

$$I(\beta_r)_1 = (1,1542 - (1,9886)\sqrt{0,0039} ; 1,1542 + (1,9886)\sqrt{0,0039})$$

$$I(\beta_r)_1 = (1,0300 ; 1,2783), \quad L(I(\beta_r)_1) = 0,2483$$

$$I(\beta_r)_2 = (-0,0543 - (1,9886)\sqrt{0,0039} ; -0,0543 + (1,9886)\sqrt{0,0039})$$

$$I(\beta_r)_2 = (-0,1784 ; 0,0698), \quad L(I(\beta_r)_2) = 0,2482$$

$$I(\beta_r)_3 = (0,0543 - (1,9886)\sqrt{0,0039} ; 0,0543 + (1,9886)\sqrt{0,0039})$$

$$I(\beta_r)_3 = (-0,0698 ; 0,1784), \quad L(I(\beta_r)_3) = 0,2482$$

$$I(\beta_r)_4 = (0,2479 - (1,9886)\sqrt{0,0155} ; 0,2479 + (1,9886)\sqrt{0,0155})$$

$$I(\beta_r)_4 = (-3,2129 \times 10^{-4} ; 0,4954), \quad L(I(\beta_r)_4) = 0,4961$$

$$I(\beta_r)_5 = (-0,7358 - (1,9886)\sqrt{0,0310} ; -0,7358 + (1,9886)\sqrt{0,0310})$$

$$I(\beta_r)_5 = (-1,0859 ; -0,3856), \quad L(I(\beta_r)_5) = 0,7002$$

$$I(\beta_r)_6 = (0,4880 - (1,9886)\sqrt{0,0155} ; 0,4880 + (1,9886)\sqrt{0,0155})$$

$$I(\beta_r)_6 = (0,2404 ; 0,7355), \quad L(I(\beta_r)_6) = 0,4951$$

$$I(\beta_r)_7 = (-1,0924 - (1,9886)\sqrt{0,0310} ; -1,0924 + (1,9886)\sqrt{0,0310})$$

$$I(\beta_r)_7 = (-1,4425 ; -0,7422), \quad L(I(\beta_r)_7) = 0,7002$$

$$I(\beta_r)_8 = (1,0924 - (1,9886)\sqrt{0,0310} ; 1,0924 + (1,9886)\sqrt{0,0310})$$

$$I(\beta_r)_8 = (0,7422 ; 1,4425), \quad L(I(\beta_r)_8) = 0,7003$$

$$I(\beta_r)_9 = \left(1,1370 \times 10^{-16} - (1,9886)\sqrt{5,37 \times 10^{-34}} ; 1,1370 \times 10^{-16} \right. \\ \left. + (1,9886)\sqrt{5,37 \times 10^{-34}} \right)$$

$$I(\beta_r)_9 = (6,7617 \times 10^{-17} ; 1,5978 \times 10^{-16}), \quad L(I(\beta_r)_9) = 9,2165 \times 10^{-17}$$

$$I(\beta_r)_{10} = (-1,137 \times 10^{-16} - (1,9886)\sqrt{0} ; -1,137 \times 10^{-16} + (1,9886)\sqrt{0}) = 0$$

$$I(\beta_r)_{11} = (0,4696 - (1,9886)\sqrt{0,0310} ; 0,4696 + (1,9886)\sqrt{0,0310})$$

$$I(\beta_r)_{11} = (0,1194 ; 0,8197), \quad L(I(\beta_r)_{11}) = 0,7003$$

$$I(\beta_r)_{12} = (-0,4696 - (1,9886)\sqrt{0,0310} ; -0,4696 + (1,9886)\sqrt{0,0310})$$

$$I(\beta_r)_{12} = (-0,8197 ; -0,1194), \quad L(I(\beta_r)_{12}) = 0,7002$$

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.

5.3.1. FUNCIONES ESTIMABLES.

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

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

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

$$30\hat{\mu} + 15 \sum_{i=1}^2 \hat{\tau}_i + 30\alpha_j + 15 \sum_{i=1}^2 (\widehat{\tau\alpha})_{ij} = y_{.j}$$

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

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:

$$\mu + \tau_1 + \alpha_1 + (\tau\alpha)_{11} \quad (5.2)$$

$$\mu + \tau_1 + \alpha_2 + (\tau\alpha)_{12} \quad (5.3)$$

$$\mu + \tau_1 + \alpha_3 + (\tau\alpha)_{13} \quad (5.4)$$

$$\mu + \tau_2 + \alpha_1 + (\tau\alpha)_{21} \quad (5.5)$$

$$\mu + \tau_2 + \alpha_2 + (\tau\alpha)_{22} \quad (5.6)$$

$$\mu + \tau_2 + \alpha_3 + (\tau\alpha)_{23} \quad (5.7)$$

Y además se tiene, que el mejor estimador de τ_i es $\bar{y}_{i..} - \bar{y}_{...}$ y el mejor estimador de α_j es $\bar{y}_{.j} - \bar{y}_{...}$.

Así, utilizando los valores de la tabla 3, se tiene:

$$\hat{\tau}_1 = \bar{y}_{1..} - \bar{y}_{...} \quad \therefore \hat{\tau}_1 = -0,22572$$

$$\hat{\tau}_2 = \bar{y}_{2..} - \bar{y}_{...} \quad \therefore \hat{\tau}_2 = 0,22571$$

$$\hat{\alpha}_1 = \bar{y}_{.1} - \bar{y}_{..} \quad \therefore \hat{\alpha}_1 = -0,29832$$

$$\hat{\alpha}_2 = \bar{y}_{.2} - \bar{y}_{..} \quad \therefore \hat{\alpha}_2 = 0,36791$$

$$\hat{\alpha}_3 = \bar{y}_{.3} - \bar{y}_{..} \quad \therefore \hat{\alpha}_3 = -0,0696$$

Además, de las ecuaciones (5.2), (5.3), (5.4), (5.5), (5.6) y (5.7) se tiene que

$$(\widehat{\tau\alpha})_{11} = \bar{y}_{11} - \hat{\mu} - \hat{\tau}_1 - \hat{\alpha}_1 \quad \therefore (\widehat{\tau\alpha})_{11} = -0,37479$$

$$(\widehat{\tau\alpha})_{12} = \bar{y}_{12} - \hat{\mu} - \hat{\tau}_1 - \hat{\alpha}_2 \quad \therefore (\widehat{\tau\alpha})_{12} = 0,29143$$

$$(\widehat{\tau\alpha})_{13} = \bar{y}_{13} - \hat{\mu} - \hat{\tau}_1 - \hat{\alpha}_3 \quad \therefore (\widehat{\tau\alpha})_{13} = 0,08339$$

$$(\widehat{\tau\alpha})_{21} = \bar{y}_{21} - \hat{\mu} - \hat{\tau}_2 - \hat{\alpha}_1 \quad \therefore (\widehat{\tau\alpha})_{21} = 0,37481$$

$$(\widehat{\tau\alpha})_{22} = \bar{y}_{22} - \hat{\mu} - \hat{\tau}_2 - \hat{\alpha}_2 \quad \therefore (\widehat{\tau\alpha})_{22} = -0,29142$$

$$(\widehat{\tau\alpha})_{23} = \bar{y}_{23} - \hat{\mu} - \hat{\tau}_2 - \hat{\alpha}_3 \quad \therefore (\widehat{\tau\alpha})_{23} = -0,08337$$

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

El coeficiente de determinación es:

$$\hat{\beta}_e^t X^t Y_e = 133,0803$$

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

$$n = 90$$

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

$$R^2 = \frac{133,0803 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = \frac{13,1843}{47,7160} = 0,2763$$

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

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

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

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

donde,

$$\hat{\sigma}^2 = \frac{1}{(6 * 15 - 6)} \left[\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^{15} (y_{ijk})^2 - (\hat{\beta}_e)^t \mathbf{X}^t \mathbf{Y} \right]$$

$$\hat{\sigma}^2 = \frac{1}{84} \left[\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^{15} (y_{ijk})^2 - (\hat{\beta}_e)^t \mathbf{X}^t \mathbf{Y} \right]$$

$$\hat{\sigma}^2 = \frac{1}{84} [167,612003 - 133,0803]$$

$$\hat{\sigma}^2 = 0,411091$$

Así,

$$\mu: \bar{y}_{...} \pm t_{84}^{0,025} \sqrt{\frac{0,411091}{90}} = 1,15415 \pm (1,9886) \sqrt{0,004567} = (1,0197; 1,2885)$$

$$L(I(\mu)) = 0,2688$$

$$\tau_1: \bar{y}_{1..} - \bar{y}_{...} \pm t_{84}^{0,025} \sqrt{\frac{0,411091(2-1)}{90}} = -0,22572 \pm (1,9886)\sqrt{0,00456}$$

$$I(\tau_1) = (-0,3599; -0,0914) \quad L(I(\tau_1)) = 0,2685$$

$$\tau_2: \bar{y}_{2..} - \bar{y}_{...} \pm t_{84}^{0,025} \sqrt{\frac{0,411091(2-1)}{90}} = 0,22571 \pm (1,9886)\sqrt{0,00456}$$

$$I(\tau_2) = (0,0914; 0,3599) \quad L(I(\tau_2)) = 0,2685$$

$$\alpha_1: \bar{y}_{.1.} - \bar{y}_{...} \pm t_{84}^{0,025} \sqrt{\frac{0,411091(3-1)}{90}} = -0,29832 \pm (1,9886)\sqrt{0,00913}$$

$$I(\alpha_1) = (-0,4883; -0,1083) \quad L(I(\alpha_1)) = 0,3800$$

$$\alpha_2: \bar{y}_{.2.} - \bar{y}_{...} \pm t_{84}^{0,025} \sqrt{\frac{0,411091(3-1)}{90}} = 0,36791 \pm (1,9886)\sqrt{0,00913}$$

$$I(\alpha_2) = (0,1778; 0,5579) \quad L(I(\alpha_2)) = 0,3800$$

$$\alpha_3: \bar{y}_{.3.} - \bar{y}_{...} \pm t_{84}^{0,025} \sqrt{\frac{0,411091(3-1)}{90}} = -0,0696 \pm (1,9886)\sqrt{0,00913}$$

$$I(\alpha_3) = (-0,2596; 0,1204) \quad L(I(\alpha_3)) = 0,3800$$

$$(\tau\alpha)_{11}: (\bar{y}_{11.} - \hat{\mu} - \hat{t}_1 - \hat{\alpha}_1) \pm t_{84}^{0,025} \sqrt{\frac{0,411091(2-1)(3-1)}{90}}$$

$$(\tau\alpha)_{11}: = -0,37479 \pm (1,9886)\sqrt{0,00913}$$

$$I((\tau\alpha)_{11}) = (-0,5647; -0,1846) \quad L(I((\tau\alpha)_{11})) = 0,3800$$

$$(\tau\alpha)_{12}: (\bar{y}_{12} - \hat{\mu} - \hat{\tau}_1 - \hat{\alpha}_2) \pm t_{84}^{0,025} \sqrt{\frac{0,411091(2-1)(3-1)}{90}}$$

$$(\tau\alpha)_{12}: 0,29143 \pm (1,9886)\sqrt{0,00913}$$

$$I((\tau\alpha)_{12}) = (0,1014 ; 0,4814) \quad L(I((\tau\alpha)_{12})) = 0,3800$$

$$(\tau\alpha)_{13}: (\bar{y}_{13} - \hat{\mu} - \hat{\tau}_1 - \hat{\alpha}_3) \pm t_{84}^{0,025} \sqrt{\frac{0,411091(2-1)(3-1)}{90}}$$

$$(\tau\alpha)_{13}: 0,08339 \pm (1,9886)\sqrt{0,00913}$$

$$I((\tau\alpha)_{13}) = (-0,1066 ; 0,2734) \quad L(I((\tau\alpha)_{13})) = 0,3800$$

$$(\tau\alpha)_{21}: (\bar{y}_{21} - \hat{\mu} - \hat{\tau}_2 - \hat{\alpha}_1) \pm t_{84}^{0,025} \sqrt{\frac{0,411091(2-1)(3-1)}{90}}$$

$$(\tau\alpha)_{21}: 0,37481 \pm (1,9886)\sqrt{0,00913}$$

$$I((\tau\alpha)_{21}) = (0,1847 ; 0,5648) \quad L(I((\tau\alpha)_{21})) = 0,3800$$

$$(\tau\alpha)_{22}: (\bar{y}_{22} - \hat{\mu} - \hat{\tau}_2 - \hat{\alpha}_2) \pm t_{84}^{0,025} \sqrt{\frac{0,411091(2-1)(3-1)}{90}}$$

$$(\tau\alpha)_{22}: -0,29142 \pm (1,9886)\sqrt{0,00913}$$

$$I((\tau\alpha)_{22}) = (-0,4814 ; -0,1014) \quad L(I((\tau\alpha)_{22})) = 0,3800$$

$$(\tau\alpha)_{23}: (\bar{y}_{23} - \hat{\mu} - \hat{\tau}_2 - \hat{\alpha}_3) \pm t_{84}^{0,025} \sqrt{\frac{0,411091(2-1)(3-1)}{90}}$$

$$(\tau\alpha)_{23}: -0,08337 \pm (1,9886)\sqrt{0,00913}$$

$$I((\tau\alpha)_{23}) = (-0,2733 ; 0,1066) \quad L(I((\tau\alpha)_{23})) = 0,3800$$

5.4. ESTIMACIÓN A TRAVÉS DE LA INVERSA GENERALIZADA E INVERSA CONDICIONAL.

Se debe encontrar una estimación de β , haciendo uso de las ecuaciones 3.3.1 y 3.3.2, tomando el caso particular que $b = \mathbf{0}_{p \times 1}$, bien sea utilizando la inversa generalizada,

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

$$\mathbf{X}^t \mathbf{Y} = \begin{bmatrix} 9868/95 \\ 8958/181 \\ 4840/89 \\ 6085/237 \\ 6728/187 \\ 2111/50 \\ 6323/1651 \\ 3954/181 \\ 5454/229 \\ 3954/181 \\ 28309/2003 \\ 3239/176 \end{bmatrix}$$

Primero se calcula la inversa generalizada de la matriz $(\mathbf{X}^t \mathbf{X})$. 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. Así,

$$(\mathbf{X}^t \mathbf{X})^{-1} = \begin{bmatrix}
 1/360 & 1/720 & 1/720 & 1/1080 & 1/1080 & 1/1080 \\
 1/720 & 1/144 & -1/180 & 1/2160 & 1/2160 & 1/2160 \\
 1/720 & -1/180 & 1/144 & 1/2160 & 1/2160 & 1/2160 \\
 1/1080 & 1/2160 & 1/2160 & 11/1080 & -1/216 & -1/216 \\
 1/1080 & 1/2160 & 1/2160 & -1/216 & 11/1080 & -1/216 \\
 1/1080 & 1/2160 & 1/2160 & -1/216 & -1/216 & 11/1080 \\
 1/2160 & 1/432 & -1/540 & 11/2160 & -1/432 & -1/432 \\
 1/2160 & 1/432 & -1/540 & -1/432 & 11/2160 & -1/432 \\
 1/2160 & 1/432 & -1/540 & -1/432 & -1/432 & 11/2160 \\
 1/2160 & -1/540 & 1/432 & 11/2160 & -1/432 & -1/432 \\
 1/2160 & -1/540 & 1/432 & -1/432 & 11/2160 & -1/432 \\
 1/2160 & -1/540 & 1/432 & -1/432 & -1/432 & 11/2160 \\
 \\
 1/2160 & 1/2160 & 1/2160 & 1/2160 & 1/2160 & 1/2160 \\
 1/432 & 1/432 & 1/432 & -1/540 & -1/540 & -1/540 \\
 -1/540 & -1/540 & -1/540 & 1/432 & 1/432 & 1/432 \\
 11/2160 & -1/432 & -1/432 & 11/2160 & -1/432 & -1/432 \\
 -1/432 & 11/2160 & -1/432 & -1/432 & 11/2160 & -1/432 \\
 -1/432 & -1/432 & 11/2160 & -1/432 & -1/432 & 11/2160 \\
 11/432 & -5/432 & -5/432 & -11/540 & 1/108 & 1/108 \\
 -5/432 & 11/432 & -5/432 & 1/108 & -11/540 & 1/108 \\
 -5/432 & -5/432 & 11/432 & 1/108 & 1/108 & -11/540 \\
 -11/540 & 1/108 & 1/108 & 11/432 & -5/432 & -5/432 \\
 1/108 & -11/540 & 1/108 & -5/432 & 11/432 & -5/432 \\
 1/108 & 1/108 & -11/540 & -5/432 & -5/432 & 11/432
 \end{bmatrix}$$

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,

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

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

El coeficiente de determinación viene dado por:

$$R^2 = \frac{138,3178 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = \frac{18,4218}{47,7160} = 0,3860,$$

donde,

$$(\hat{\boldsymbol{\beta}}_g)^t \mathbf{X}^t \mathbf{Y} = 138,3178$$

$$\mathbf{Y}^t \mathbf{Y} = 167,6120$$

$$n = 90$$

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

Ahora se calcularán las varianzas de cada $(\hat{\beta}_g)_j$, $j = 1, 2, \dots, 12$; para ello se tiene que:

$$\hat{\sigma}^2 = \frac{1}{n - k} (\mathbf{Y}^t \mathbf{Y} - (\hat{\beta}_g)^t \mathbf{X}^t \mathbf{Y})$$

$$\hat{\sigma}^2 = \frac{1}{90 - 6} (167,6120 - 138,3178)$$

$$\hat{\sigma}^2 = 0,348740.$$

Luego,

$$Cov(\hat{\beta}_g) = \hat{\sigma}^2 (\mathbf{X}^t \mathbf{X})^{-1}$$

$$Cov(\hat{\beta}_g) = \begin{bmatrix} \mathbf{0,0010} & 0,0005 & 0,0005 & 0,0003 & 0,0003 & 0,0003 \\ 0,0005 & \mathbf{0,0024} & -0,0019 & 0,0002 & 0,0002 & 0,0002 \\ 0,0005 & -0,0019 & \mathbf{0,0024} & 0,0002 & 0,0002 & 0,0002 \\ 0,0003 & 0,0002 & 0,0002 & \mathbf{0,0036} & -0,0016 & -0,0016 \\ 0,0003 & 0,0002 & 0,0002 & -0,0016 & \mathbf{0,0036} & -0,0016 \\ 0,0003 & 0,0002 & 0,0002 & -0,0016 & -0,0016 & \mathbf{0,0036} \\ 0,0002 & 0,0008 & -0,0006 & 0,0018 & -0,0008 & -0,0008 \\ 0,0002 & 0,0008 & -0,0006 & -0,0008 & 0,0018 & -0,0008 \\ 0,0002 & 0,0008 & -0,0006 & -0,0008 & -0,0008 & 0,0018 \\ 0,0002 & -0,0006 & 0,0008 & 0,0018 & -0,0008 & -0,0008 \\ 0,0002 & -0,0006 & 0,0008 & -0,0008 & 0,0018 & -0,0008 \\ 0,0002 & -0,0006 & 0,0008 & -0,0008 & -0,0008 & 0,0018 \end{bmatrix}$$

0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
0,0008	0,0008	0,0008	-0,0006	-0,0006	-0,0006
-0,0006	-0,0006	-0,0006	0,0008	0,0008	0,0008
0,0018	-0,0008	-0,0008	0,0018	-0,0008	-0,0008
-0,0008	0,0018	-0,0008	-0,0008	0,0018	-0,0008
-0,0008	-0,0008	0,0018	-0,0008	-0,0008	0,0018
0,0089	-0,0040	-0,0040	-0,0071	0,0032	0,0032
-0,0040	0,0089	-0,0040	0,0032	-0,0071	0,0032
-0,0040	-0,0040	0,0089	0,0032	0,0032	-0,0071
-0,0071	0,0032	0,0032	0,0089	-0,0040	-0,0040
0,0032	-0,0071	0,0032	-0,0040	0,0089	-0,0040
0,0032	0,0032	-0,0071	-0,0040	-0,0040	0,0089

Por lo tanto, los intervalos de confianza vienen dados por:

$$I(\beta_g)_1 = (0,5771 - (1,9886)\sqrt{0,0010} ; 0,5771 + (1,9886)\sqrt{0,0010})$$

$$I(\beta_g)_1 = (0,5142 ; 0,6399), \quad L(I(\beta_g)_1) = 0,1257$$

$$I(\beta_g)_2 = (0,2478 - (1,9886)\sqrt{0,0024} ; 0,2478 + (1,9886)\sqrt{0,0024})$$

$$I(\beta_g)_2 = (0,1503 ; 0,3452), \quad L(I(\beta_g)_2) = 0,1949$$

$$I(\beta_g)_3 = (0,3293 - (1,9886)\sqrt{0,0024} ; 0,3293 + (1,9886)\sqrt{0,0024})$$

$$I(\beta_g)_3 = (0,2318 ; 0,4267), \quad L(I(\beta_g)_3) = 0,1949$$

$$I(\beta_g)_4 = (-0,0065 - (1,9886)\sqrt{0,0036} ; -0,0065 + (1,9886)\sqrt{0,0036})$$

$$I(\beta_g)_4 = (-0,1258 ; 0,1128), \quad L(I(\beta_g)_4) = 0,2386$$

$$I(\beta_g)_5 = (0,2224 - (1,9886)\sqrt{0,0036} ; 0,2224 + (1,9886)\sqrt{0,0036})$$

$$I(\beta_g)_5 = (0,1030 ; 0,3417), \quad L(I(\beta_g)_5) = 0,2387$$

$$I(\beta_g)_6 = (0,3611 - (1,9886)\sqrt{0,0036} ; 0,3611 + (1,9886)\sqrt{0,0036})$$

$$I(\beta_g)_6 = (0,2417 ; 0,4804), \quad L(I(\beta_g)_6) = 0,2387$$

$$I(\beta_g)_7 = (-0,5630 - (1,9886)\sqrt{0,0089} ; -0,5630 + (1,9886)\sqrt{0,0089})$$

$$I(\beta_g)_7 = (-0,7506 ; -0,3753), \quad L(I(\beta_g)_7) = 0,3752$$

$$I(\beta_g)_8 = (0,4090 - (1,9886)\sqrt{0,0089} ; 0,4090 + (1,9886)\sqrt{0,0089})$$

$$I(\beta_g)_8 = (0,2213 ; 0,5966), \quad L(I(\beta_g)_8) = 0,3753$$

$$I(\beta_g)_9 = (0,4018 - (1,9886)\sqrt{0,0089} ; 0,4018 + (1,9886)\sqrt{0,0089})$$

$$I(\beta_g)_9 = (0,2141 ; 0,5894), \quad L(I(\beta_g)_9) = 0,3753$$

$$I(\beta_g)_{10} = (0,5565 - (1,9886)\sqrt{0,0089} ; 0,5565 + (1,9886)\sqrt{0,0089})$$

$$I(\beta_g)_{10} = (0,3688 ; 0,7441), \quad L(I(\beta_g)_{10}) = 0,3753$$

$$I(\beta_g)_{11} = (-0,1866 - (1,9886)\sqrt{0,0089} ; -0,1866 + (1,9886)\sqrt{0,0089})$$

$$I(\beta_g)_{11} = (-0,3742 ; 0,001004), \quad L(I(\beta_g)_{11}) = 0,3752$$

$$I(\beta_g)_{12} = (-0,0406 - (1,9886)\sqrt{0,0089} ; -0,0406 + (1,9886)\sqrt{0,0089})$$

$$I(\beta_g)_{12} = (-0,2282 ; 0,1470), \quad L(I(\beta_g)_{12}) = 0,3752$$

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

- i. Tome una submatriz de $(\mathbf{X}^t \mathbf{X})$ de rango completo y orden igual al rango de $(\mathbf{X}^t \mathbf{X})$ (sub matriz no singular). Llame a esa submatriz \mathbf{B} de orden (6×6) donde $6 = \text{rang}(\mathbf{X})$.

$$\mathbf{B} = \begin{bmatrix} 90 & 45 & 30 & 30 & 15 & 15 \\ 45 & 45 & 15 & 15 & 15 & 15 \\ 30 & 15 & 30 & 0 & 15 & 0 \\ 30 & 15 & 0 & 30 & 0 & 15 \\ 15 & 15 & 15 & 0 & 15 & 0 \\ 15 & 15 & 0 & 15 & 0 & 15 \end{bmatrix}$$

ii. Invierta \mathbf{B}

$$\mathbf{B}^{-1} = \begin{bmatrix} 1/15 & -1/15 & -1/15 & -1/15 & 1/15 & 1/15 \\ -1/15 & 2/15 & 1/15 & 1/15 & -2/15 & -2/15 \\ -1/15 & 1/15 & 2/15 & 1/15 & -2/15 & -1/15 \\ -1/15 & 1/15 & 1/15 & 2/15 & -1/15 & -2/15 \\ 1/15 & -2/15 & -2/15 & -1/15 & 4/15 & 2/15 \\ 1/15 & -2/15 & -1/15 & -2/15 & 2/15 & 4/15 \end{bmatrix}$$

iii. Transponga \mathbf{B}^{-1}

$$\mathbf{B}^{-1} = \begin{bmatrix} 1/15 & -1/15 & -1/15 & -1/15 & 1/15 & 1/15 \\ -1/15 & 2/15 & 1/15 & 1/15 & -2/15 & -2/15 \\ -1/15 & 1/15 & 2/15 & 1/15 & -2/15 & -1/15 \\ -1/15 & 1/15 & 1/15 & 2/15 & -1/15 & -2/15 \\ 1/15 & -2/15 & -2/15 & -1/15 & 4/15 & 2/15 \\ 1/15 & -2/15 & -1/15 & -2/15 & 2/15 & 4/15 \end{bmatrix}$$

- iv. Reemplace con los elementos de \mathbf{B}^{-1} los elementos homólogos en $(\mathbf{X}^t\mathbf{X})$ y sustituir con ceros el resto para obtener la matriz \mathbf{C} de orden (12×12) .
- v. Trasponer \mathbf{C} para obtener la inversa condicional $(\mathbf{X}^t\mathbf{X})^c$ de $(\mathbf{X}^t\mathbf{X})$.

$$(\mathbf{X}^t \mathbf{X})^c = \begin{bmatrix} 1/15 & -1/15 & 0 & -1/15 & -1/15 & 0 & 1/15 & 1/15 & 0 & 0 & 0 & 0 \\ -1/15 & 2/15 & 0 & 1/15 & 1/15 & 0 & -2/15 & -2/15 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1/15 & 1/15 & 0 & 2/15 & 1/15 & 0 & -2/15 & -1/15 & 0 & 0 & 0 & 0 \\ -1/15 & 1/15 & 0 & 1/15 & 2/15 & 0 & -1/15 & -2/15 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/15 & -2/15 & 0 & -2/15 & -1/15 & 0 & 4/15 & 2/15 & 0 & 0 & 0 & 0 \\ 1/15 & -2/15 & 0 & -1/15 & -2/15 & 0 & 2/15 & 4/15 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Por lo tanto,

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

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

En este caso coeficiente de determinación viene dado por:

$$R^2 = \frac{138,3178 - 90(1,1542)^2}{167,6120 - 90(1,1542)^2} = \frac{18,4218}{47,7160} = 0,3860,$$

donde,

$$(\hat{\beta}_c)^t X^t Y = 138,3178$$

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

$$n = 90$$

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

Por otro lado,

$$\hat{\sigma}^2 = \frac{1}{n-k} (\mathbf{Y}^t \mathbf{Y} - (\hat{\boldsymbol{\beta}}_c)^t \mathbf{X}^t \mathbf{Y})$$

$$\hat{\sigma}^2 = \frac{1}{90-6} (167,6120 - 138,3178)$$

$$\hat{\sigma}^2 = 0,348740.$$

Así,

$$\text{Cov}(\hat{\boldsymbol{\beta}}_c) = \hat{\sigma}^2 (\mathbf{X}^t \mathbf{X})^c$$

$$\text{Cov}(\hat{\boldsymbol{\beta}}_c) = \begin{bmatrix} \mathbf{0,0232} & -0,0232 & 0 & -0,0232 & -0,0232 & 0 \\ -0,0232 & \mathbf{0,0465} & 0 & 0,0232 & 0,0232 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 & 0 \\ -0,0232 & 0,0232 & 0 & \mathbf{0,0465} & 0,0232 & 0 \\ -0,0232 & 0,0232 & 0 & 0,0232 & \mathbf{0,0465} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{0} \\ 0,0232 & -0,0465 & 0 & -0,0465 & -0,0232 & 0 \\ 0,0232 & -0,0465 & 0 & -0,0232 & -0,0465 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0,0232 & 0,0232 & 0 & 0 & 0 & 0 \\ -0,0465 & -0,0465 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0,0465 & -0,0232 & 0 & 0 & 0 & 0 \\ -0,0232 & -0,0465 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{0,0930} & 0,0465 & 0 & 0 & 0 & 0 \\ 0,0465 & \mathbf{0,0930} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{0} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix}$$

Por lo tanto los intervalos de confianza son:

$$I(\beta_c)_1 = (1,2269 - (1,9886)\sqrt{0,0232} ; 1,2269 + (1,9886)\sqrt{0,0232})$$

$$I(\beta_c)_1 = (0,9240 ; 1,5297), \quad L(I(\beta_c)_1) = 0,6057$$

$$I(\beta_c)_2 = (0,3609 - (1,9886)\sqrt{0,0465} ; 0,3609 + (1,9886)\sqrt{0,0465})$$

$$I(\beta_c)_2 = (-0,0679 ; 0,7897), \quad L(I(\beta_c)_2) = 0,8576$$

$$I(\beta_c)_3 = (0 - (1,9886)\sqrt{0} ; 0 + (1,9886)\sqrt{0}) = 0$$

$$I(\beta_c)_4 = (0,2295 - (1,9886)\sqrt{0,0465} ; 0,2295 + (1,9886)\sqrt{0,0465})$$

$$I(\beta_c)_4 = (-0,1993 ; 0,6583), \quad L(I(\beta_c)_4) = 0,8576$$

$$I(\beta_c)_5 = (-0,2847 - (1,9886)\sqrt{0,0465} ; -0,2847 + (1,9886)\sqrt{0,0465})$$

$$I(\beta_c)_5 = (-0,7135 ; 0,1441), \quad L(I(\beta_c)_5) = 0,8576$$

$$I(\beta_c)_6 = (0 - (1,9886)\sqrt{0} ; 0 + (1,9886)\sqrt{0}) = 0$$

$$I(\beta_c)_7 = (-1,5619 - (1,9886)\sqrt{0,0930} ; -1,5619 + (1,9886)\sqrt{0,0930})$$

$$I(\beta_c)_7 = (-2,1683 ; -0,9554), \quad L(I(\beta_c)_7) = 1,2128$$

$$I(\beta_c)_8 = (0,1533 - (1,9886)\sqrt{0,0930} ; 0,1533 + (1,9886)\sqrt{0,0930})$$

$$I(\beta_c)_8 = (-0,4531 ; 0,7597), \quad L(I(\beta_c)_8) = 1,2128$$

$$I(\beta_c)_9 = (0 - (1,9886)\sqrt{0} ; 0 + (1,9886)\sqrt{0}) = 0$$

$$I(\beta_c)_{10} = (0 - (1,9886)\sqrt{0} ; 0 + (1,9886)\sqrt{0}) = 0$$

$$I(\beta_c)_{11} = (0 - (1,9886)\sqrt{0} ; 0 + (1,9886)\sqrt{0}) = 0$$

$$I(\beta_c)_{12} = (0 - (1,9886)\sqrt{0} ; 0 + (1,9886)\sqrt{0}) = 0$$

En resumen, los resultados de los métodos empleados se pueden expresar en la siguiente tabla.

Métodos	Modelo Reducido	Imposición de restricciones	Funciones estimables
$\hat{\beta}$	$\hat{\beta}^* = \begin{bmatrix} 1,2269 \\ 0,3609 \\ 0,2295 \\ -0,2847 \\ -1,5619 \\ 0,1533 \end{bmatrix}$	$\hat{\beta}_r = \begin{bmatrix} 1,1542 \\ -0,0543 \\ 0,0543 \\ 0,2479 \\ -0,7358 \\ 0,4880 \\ -1,0924 \\ 1,0924 \\ 1,1370 \times 10^{-16} \\ -1,1370 \times 10^{-16} \\ 0,4696 \\ -0,4696 \end{bmatrix}$	$\hat{\beta}_e = \begin{bmatrix} 1,15415 \\ -0,22572 \\ 0,22571 \\ -0,29832 \\ 0,36791 \\ -0,0696 \\ -0,37470 \\ 0,29143 \\ 0,08339 \\ 0,37481 \\ -0,29142 \\ -0,08337 \end{bmatrix}$
\hat{Y}	$\hat{Y}^* = \begin{bmatrix} 0,2553 \\ \vdots \\ 1,4564 \\ \vdots \\ 1,5878 \\ \vdots \\ 1,4564 \\ \vdots \\ 0,9422 \\ \vdots \\ 1,2269 \\ \vdots \end{bmatrix}_{90 \times 12}$	$\hat{Y}_r = \begin{bmatrix} 0,2553 \\ \vdots \\ 1,4564 \\ \vdots \\ 1,5878 \\ \vdots \\ 1,4564 \\ \vdots \\ 0,9422 \\ \vdots \\ 1,2269 \\ \vdots \end{bmatrix}_{90 \times 12}$	$\hat{Y}_e = \begin{bmatrix} 0,2553 \\ \vdots \\ 1,4564 \\ \vdots \\ 1,5878 \\ \vdots \\ 1,4564 \\ \vdots \\ 0,9422 \\ \vdots \\ 1,2269 \\ \vdots \end{bmatrix}_{90 \times 12}$
R^2	$R^2 = 0,3860$	$R^2 = 0,3860$	$R^2 = 0,2763$
Longitud de los intervalos de confianza	5,6042	5,2371	4,2258

Métodos	Inversa Generalizada	Inversa Condicional
$\hat{\beta}$	$\hat{\beta}_g = \begin{bmatrix} 0,5771 \\ 0,2478 \\ 0,3293 \\ -0,0065 \\ 0,2224 \\ 0,3611 \\ -0,5630 \\ 0,4090 \\ 0,4018 \\ 0,5565 \\ -0,1866 \\ -0,0406 \end{bmatrix}$	$\hat{\beta}_c = \begin{bmatrix} 1,2269 \\ 0,3609 \\ 0 \\ 0,2295 \\ -0,2847 \\ 0 \\ -1,5619 \\ 0,1533 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
\hat{Y}	$\hat{Y}_g = \begin{bmatrix} 0,2553 \\ \vdots \\ 1,4564 \\ \vdots \\ 1,5878 \\ \vdots \\ 1,4564 \\ \vdots \\ 0,9422 \\ \vdots \\ 1,2269 \\ \vdots \end{bmatrix}_{90 \times 12}$	$\hat{Y}_c = \begin{bmatrix} 0,2553 \\ \vdots \\ 1,4564 \\ \vdots \\ 1,5878 \\ \vdots \\ 1,4564 \\ \vdots \\ 0,9422 \\ \vdots \\ 1,2269 \\ \vdots \end{bmatrix}_{90 \times 12}$
R^2	$R^2 = 0,3860$	$R^2 = 0,3860$
Longitud de los intervalos de confianza	3,483	5,6041

Tabla 4. Resumen de los resultados obtenidos de los métodos empleados.

Ventajas y desventajas del uso de la inversa generalizada y/o condicional.

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.

Conclusiones.

Los métodos de estimación empleados están uniformes en cuanto a resultados, ya que el vector de respuesta es el mismo a pesar de que los vectores del parámetro estimado $\hat{\beta}$ son diferentes. Así como también, las varianzas de los errores obtenidos fueron iguales excepto en el método de funciones estimables, el cuál arrojó una varianza de error más alta y por lo tanto, el coeficiente de determinación fue más bajo y diferente con respecto a los otros métodos estudiados.

Se puede decir, gracias a los intervalos de confianza que el método más eficiente es el del uso de la inversa generalizada, pues fue en este donde se obtuvo la longitud más pequeña de los intervalos y el método más deficiente fue el del uso del modelo reducido.

De acuerdo a los resultados, el empleo del coeficiente de determinación usual no es el adecuado para comparar el ajuste de modelos con criterios de estimación diferentes, ya que para todos los métodos empleados este fue relativamente el mismo, excepto en el método de funciones estimables donde fue más bajo.

Bibliografía.

Álvarez, W. (2009). *Contribución a la estimación en regresión multivariante con rango reducido utilizando mínimos cuadrados parciales*. Tesis Doctoral. Universidad Central de Venezuela. Venezuela.

Anderson, T. (1951). *Estimating Linear Restrictions on Regression for Multivariate Normal Distributions*. *Annals of Mathematical Statistics*, 22, 327-351.

Anderson, T. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, 2nd ed.

Anderson, T. (1999). *Asymptotic Distribution of the Reduced Rank Regression estimator under general conditions*. *The Annals of Statistics*, 7, 1141-1154.

Anderson, T. (2002). *Specification and Misspecification in Reduced Rank Regression*. *The Indian Journal of Statistics*, 64, 193-205.

Azocar, R. (1986). *El análisis de componentes principales y la multicolinealidad en problemas de la investigación agrícola*. Tesis de Maestría. Universidad Central de Venezuela. Venezuela.

Ben-Israel, A & Charnes, A. (1963). *Contributions to the theory of generalized inverses*. SIAM J. Appl. Math., Vol. 11 pp. 667-699.

Bjerhammer, A. (1958). *A generalized matrix algebra*. Kungl. Tekn. Hogsk. Handl. Stockholm. No. 124 pp. 1-32.

Bose, R. C. (1959). *Analysis of Variance*. Unpublished lecture notes, University of North Carolina.

Davies, P. & Tso, M. (1982). *Procedures for Reduced Rank Regression*. Applied Statistic. 31, 244-255.

Dunn, O.J., & Clark, V.A. (1987). *Applied Statistics: Analysis of Variance and Regression*. Second ed. John Wiley, New York.

Ferguson, Thomas S. (1996). *A course in large sample theory*. Chapman & Hall. ISBN 0412043718.

Flores, J. (2001). *Press-Ridge en la estimación de un modelo de regresión lineal múltiple en presencia de multicolinealidad severa*. Tesis de Maestría. Universidad Central de Venezuela. Venezuela.

Graybill, F. A. (1976). *Theory and application of the linear model*. Duxbury press. Massachusetts.

Greville, T. N. E. (1959). *The pseudo-inverse of a rectangular matrix and its application to the solution of systems of linear equations*. SIA M Rev., Vol. 1, pp. 38-43.

Gudmundsson, G. (1977). *Multivariate Analysis of Economic Variables*. Applied statistics, 26, 48-59.

Gujarati, D. (1992). *Econometría Básica*. México. McGraw-Hill. 436p

Hald, Anders. (1998). *A history of mathematical statistics from 1750 to 1930*. New York: Wiley. ISBN 0471179124.

Hostelling, H. (1935). *The Most Predictable Criterion*. Journal of education psychology, 26, 139-142.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. Journal of Multivariate Analysis, 5, 248-264.

Kirk, R. E. (1982). *Experimental design*. (2da. ed.). Belmont, CA: Brooks/Cole Publishing Company.

López, C; R. (2006). *Cálculo de Probabilidades e Inferencia Estadística con tópicos en Econometría*. Universidad Católica Andrés Bello. Editorial texto. 4e. Caracas- Venezuela.

Milliken, G. A. & Johnson, D. E. (1984). *Analysis of Messy Data: Designed Experiments*. New York: Van Nostrand Reinhold.

Mansfield, E. & Helms, B. (1982). *Detecting Multicollinearity*. The American Statistician 36: 158-160.

Marquard, D.W. (1970). *Generalized Inverse, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation*. Technometrics 12: 591-612.

Montgomery, D.C & Peck, E. A. (1982). *Introduction to Near Regression Analisis*. John Wiley and Sons. Inc. New York. USA.

Moore, E. H. (1920). *On the reciprocal of the general algebraic matrix (abstract)*. Bull. Amer. Math. Soc., Vol. 26 pp. 394-395.

Navarro, F., Youngquis, W., Compton, W. (1992). *Estimación de varianzas genéticas en maíz a partir de las líneas S1 y S2*. Revista Agronomía Mesoamericana 3: 9-15.

Penrose, R. (1955). *A generalized inverse for matrices*. Proc. Cambridge Philos. Soc., Vol. 51 pp. 406-413.

Rao, C. R. (1955). *Analysis of dispersion for multiply classified data with unequal numbers in cells*. Sankhy. Vol. 15 pp. 253-280.

Rao, C. R. (1962). *A note on a generalized inverse of a matrix with applications to problems in mathematical statistics*. J. Roy. Statist. Soc. Ser. B. Vol. 24 pp. 152-158.

Rao, C. R. (1967). *Least squares theory using an estimated dispersion matrix and its application to measurement of signals*. Proceedings of the Fifth Berkeley Symposium on

Statistics and Probability, Berkeley and Los Angeles, University of California Press, Vol. 1, pp. 355-372.

Searle, S. R. (1997). *Linear Models*. Wiley Classics Library. New York: John Wiley & Sons, Inc.

Scheffé, H. (1959). *The Analysis of Variance*. John Wiley. USA.

Stigler, Stephen M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press, Cambridge. Mass.

Tseng, Y. Y. (1949). *Generalized inverses of unbounded operators between two unitary spaces*. Dokl. Akad. Nauk. SSSR., Vol. 67 pp. 431-434.

Tseng, Y. Y. (1949). *Properties and classifications of generalized inverses of closed operators*. Dokl. Akad. Nauk. SSSR, Vol. 67 pp. 607-610.

Tseng, Y. Y. (1956). *Virtual solutions and general inversions*. Uspehi. Mat. Nauk., Vol. 11 pp.213-215.

Timm, N. H. (1975). *Multivariate analysis with application in education and psychology*.

Monterey: Brooks/Kole.

Velu, R. P & Reinsel, G. C. (1987). *Reduced Rank Regression with Autoregressive Errors*.

Journal of Econometrics, 35, 317-375

Velu, R. P; Reinsel, G. C & Wichern, D. W. (1986). *Reduced Rank Models for Multiple Time*

Series. Biometrika, 73, 105-118.