INVERSAS GENERALIZADAS Y SU APLICACION AL ANALISIS ESTADISTICO DE LOS MODELOS LINEALES: UN ENFOQUE GEOMETRICO.

JOSE REFUGIO REYES VALDES

TESIS

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS
EN ESTADISTICA EXPERIMENTAL

Universidad Autonoma Agraria
Antonio Narro

PROGRAMA DE GRADUADOS
Buenavista, Saltillo, Coah.
SEPTIEMBRE DE 1990

Tesis elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial para optar al grado de

MAESTRO EN CIENCIAS EN

ESTADISTICA EXPERIMENTAL

COMITE PARTICULAR

Asesor principal: ______________________________
Dr. Rolando Cavazos Cadena

Asesor: ______________________________
M C. Félix de Jesús Sánchez Pérez

Asesor: ______________________________
M C. Emilio Padrón Corral

____
Dr. Eleuterio López Pérez
Subdirector de Asuntos de Postgrado

Buenavista, Saltillo, Coahuila. Septiembre de 1990

Me causa placer recordar cuantos rodeos tuve que dar, cuantas paredes tuve que tocar buscando a tientas en la obscuridad de mi ignorancia hasta encontrar la puerta que se abre a la luz de la verdad.

Kepler.

AGRADECIMIENTO

Agradezco profundamente al Dr. Rolando Cavazos Cadena por el apoyo desinteresado que me brindó, sin el cual no hubiera sido posible la realización de este trabajo.

A mi hermano Manuel Humberto Reyes Valdés por la motivación para realizar mis estudios y el consejo que me brindó en todo momento.

A Miguel Ramos Parra, Guadalupe Fuentes del Bosque y a todos aquellos que siguieron mis estudios paso a paso y de quienes nunca faltó una palabra que me alentara a seguir adelante.

A mis estudiantes, de quienes tanto aprendí

A mis Padres:

José Refugio Reyes Rodriguez

Magdalena Valdés de Reyes

Quienes han sido para mi ejemplo de entrega y sacrificio, así como cimiento y soporte de mi formación como profesionista y como hombre.

COMPENDIO

Inversas Generalizadas y su Aplicación al análisis Estadístico de los Modelos Lineales: Un Enfoque Geométrico

por

JOSE REFUGIO REYES VALDES

MAESTRIA

ESTADISTICA EXPERIMENTAL

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. SEPTIEMBRE DE 1990

Dr. Rolando Cavazos Cadena — Asesor —

Palabras clave: Proyección, Proyección Ortogonal, Inversa Condicional, Inversa Generalizada, Modelos Lineales.

Este trabajo trata sobre los conceptos de inversa condicional e inversa generalizada de una matriz arbitraria, nociones que desempeñan un importante papel en el análisis estadístico de un modelo lineal. Nuestro principal objetivo es proporcionar una descripción geométrica de las propiedades de este tipo de inversas, así

como ilustrar su aplicación a los problemas de prueba de hipótesis y estimación de parámetros de un modelo lineal bajo condiciones de normalidad.

ABSTRACT

Generalized Inverses and their application

to the Statistical Analysis of Linear Models:

A Geometrical Approach

BY

JOSE REFUGIO REYES VALDES

MASTER OF SCIENCE

EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, SEPTEMBER 1990

Dr. Rolando Cavazos Cadena —— Advisor ——

Key words: Projection, Orthogonal Projection, Conditional Inverse, Generalized Inverse, Linear Models.

In this work we study the notions of conditional and generalized inverses of an arbitrary matrix. Our main objetive is to provide a geometrical description of the properties of those inverses, as well as to illustrate the

application of the results to the problems of parameter estimation and hipothesis testing in a linear statistical model under the usual normality assumption.

INDICE DE CONTENIDO

INTRODUCCION

El presente trabajo trata sobre algunas técnicas matriciales de interés en Estadística, en especial aquellas que giran en torno a las nociones de inversa condicional e inversa generalizada. Estudiaremos algunas aplicaciones de estos conceptos al análisis de un modelo lineal, las cuales pondrán de manifiesto la importancia de los mismos.

Nuestros objetivos principales son los siguientes:

1. Proporcionar una interpretación geométrica de las propiedades de las inversas condicional y generalizada de un matriz, así como presentar algoritmos prácticos para su cálculo.

2. Aplicar los resultados obtenidos a los siguientes problemas: (*i*) Estimación de parámetros y prueba de hipótesis en un modelo lineal general, y (*ii*) Obtención de la densidad condicional en una distribución normal arbitraria.

Con estos objetivos pretendemos ilustrar claramente la importancia de los conceptos de inversas condicional y generalizada en la solución de problemas que son de importancia capital en la Estadística, tanto teórica como aplicada.

La organización del trabajo es la siguiente: En el capítulo uno se estudian operadores de proyección definidos en espacios provistos con un producto interno, y se desarrolla el material que será el soporte teórico fundamental para el desarrollo e interpretación geométrica de los resultados posteriores. Entre los resultados mas importantes se encuentra el teorema que garantiza la diagonalización de matrices simétricas. El Capítulo dos, se refiere a la distribución normal multidimensional (no singular) y se incluye una demostración del llamado Teorema de Cochran. Por otro lado, la parte central del trabajo está contenida en el Capítulo tres, el cual está conformado con el desarrollo y método de cálculo de las inversas condicional y generalizada. También forman parte de este capítulo, la estimación de los parámetros de un modelo lineal y la distribución condicional para una distribución normal general, aplicaciones que ilustran la potencia de la noción de inversa condicional. Finalmente, en el Capítulo cuatro se deduce el estadístico de prueba de una hipótesis lineal general, donde se hará uso de la inversa generalizada.

CAPITULO I

Proyecciones y Producto Interno

En este capítulo se estudian algunos resultados fundamentales del álgebra lineal que serán básicos para el desarrollo del presente trabajo.

En la primera sección se introduce el concepto de operador de proyección, prestando especial interés a las proyecciones ortogonales. Estudiar este tipo de transformaciones lineales es indispensable para demostrar el así llamado Teorema Espectral. Este resultado esta intimamente relacionado con el problema de la diagonalización de matrices simétricas, al cual prestamos especial interés pués desempeña un papel central en el estudio de la distribución de formas cuadráticas asociadas a un modelo lineal bajo la hipótesis de normalidad. Además, en la Sección cuatro estableceremos las condiciones requeridas para la diagonalización simultánea de matrices simétricas.

1.1. Proyecciones

En esta sección se estudia una clase especial de operadores lineales conocidos como operadores de proyección. Analizaremos algunas de las características de estos operadores, asi como las propiedades de las matrices que representan una proyección con respecto a alguna base dada. Durante el resto del capítulo, V es un espacio vectorial de dimensión finita sobre el conjunto de números reales.

Definicion 1.1.1 Sea $T:V \longrightarrow V$ una transformación lineal. Entonces, T es una proyección si y sólo si $T^2 = T$, esto es,

$$T(T\alpha) = T\alpha, \ \alpha \in V;$$

en la literatura, a una proyección T también se le refiere como una tranformación u operador idempotente.

Dada una transformación lineal $T:V \longrightarrow V$ y una base $\mathcal{B} = \{\beta_1, \cdots, \beta_n\}$ de V, es posible representar a T mediante una matriz, la cual denotaremos por $[T]_{\mathcal{B}}$ Hoffman y Kunze (1973), Graybill (1976). Esta matriz es única y se determina de acuerdo a la siguiente condición: $[T]_{\mathcal{B}} = A$ si y solo si,

$$T\beta_j = \sum_{j=1}^{n} A_{ij}\beta_i, \ j = 1, \cdots, n.$$

La correspondencia $T \leftrightarrow [T]_{\mathcal{B}}$ es biunívoca y además se tiene que $[T^2]_{\mathcal{B}} = [T]^2_{\mathcal{B}}$. Por lo tanto, $T^2 = T$ si y sólo si $[T^2]_{\mathcal{B}} = [T]_{\mathcal{B}}$, o equivalentemente,

$$[T]^2_{\mathcal{B}} = [T]_{\mathcal{B}}. \tag{1.1}$$

Una matriz cuadrada **A** se dice idempotente si $\mathbf{A}^2 = \mathbf{A}$. Con esta terminología podemos resumir la discusión anterior de la siguiente manera: El operador lineal $\mathbf{T}:V \longrightarrow V$ es una proyección si y sólo si la matriz $\mathbf{A} = [T]_{\mathcal{B}}$ es idempotente.

Ejemplo 1.1.1 Sea **M** una matriz de orden n x n. Defina el operador $T_{\mathbf{M}}:\mathbb{E}_n \longrightarrow \mathbb{E}_n$ mediante

$$T_{\mathbf{M}}x := \mathbf{M}x, \quad x \in \mathbb{E}_n.$$

Sea $\mathcal{E}$ la base canónica de $\mathbb{E}_n$ Hoffman y Kunze (1973). Es fácil ver que $[T_{\mathbf{M}}]_{\mathcal{E}} = \mathbf{M}$ y entonces $T_{\mathbf{M}}$ es una proyección si y sólo si, **M** es idempotente. □

El Teorema 1.1.1 que aparece enseguida proporciona una interpretación geométrica de la manera en que una proyección T actúa sobre los vectores de V. Para establecer este resultado, recordemos que si W y U son subespacios de V, entonces V es la suma directa de W y U si las siguientes condiciones (*i*) y (*ii*) se satisfacen.

(*i*) Todo vector $\alpha \in V$ se puede representar como $\alpha = \mathbf{w} + \mathbf{u}$, donde $\mathbf{w} \in W$ y $\mathbf{u} \in U$

(*ii*) $W \cap U = \{0\}$.

Estas dos condiciones juntas son equivalentes a la siguiente: Para cada vector $\alpha \in V$, puede expresarse de manera única como $\alpha = \mathbf{w} + \mathbf{u}$, donde $\mathbf{w} \in W$ y $\mathbf{u} \in U$. Cuando V es la suma directa de los subespacios W y U escribiremos

$$V = W \oplus U.$$

Dada una transformación lineal $T:V \longrightarrow V$, la imágen de T y el núcleo de T se definen mediante $\mathcal{R}(T) := \{T\alpha \mid \alpha \in V\}$ y $\mathcal{N}(T) := \{\alpha \in V \mid T\alpha = 0\}$, respectivamente.

Teorema 1.1.1 Sea $T:V \longrightarrow V$ una proyección. Entonces,

(*i*) Para cada $\rho \in \mathcal{R}(T)$, $T\rho = \rho$.

(*ii*) $V = \mathcal{R}(T) \oplus \mathcal{N}(T)$.

(*ii*) Si $\alpha = \rho + \nu$, $\rho \in \mathcal{R}(T)$, $\nu \in \mathcal{N}(T)$, entonces $T\alpha = \rho$

Demostración. (*i*) Sea $\rho \in \mathcal{R}(T)$ un vector arbitrario. En este caso, $\rho = T\beta$ para algún $\beta \in V$ y entonces, $T\rho = T(T\beta) = T^2\beta = T\beta = \rho$, donde en la tercera igualdad hemos usado que T es una proyección.

(*ii*) Note que para $\alpha \in V$, $T\alpha \in \mathcal{R}(T)$ y $T(\alpha - T\alpha) = T\alpha - T^2\alpha = 0$, esto es, $\alpha - T\alpha \in \mathcal{N}(T)$. Además, es claro que $\alpha = T\alpha + (\alpha - T\alpha)$, y concluimos que $V = \mathcal{R}(T) + \mathcal{N}(T)$. Para terminar la demostración de la parte (*ii*) es suficiente ver que

$$\mathcal{R}(T) \cap \mathcal{N}(T) = \{0\}. \tag{1.2}$$

Suponga que $\rho \in \mathcal{R}(T) \cap \mathcal{N}(T)$. Entonces,

$$\begin{aligned} \rho &= T\rho \text{ (por la parte } (i)) \\ &= 0 \quad (\text{pues } \rho \in \mathcal{N}(T)), \end{aligned}$$

y como $\rho \in \mathcal{R}(T) \cap \mathcal{N}(T)$ es arbitrario (1.2) se sigue de inmediato.

(*iii*) Sea $\alpha = \rho + \nu$, $\rho \in \mathcal{R}(T)$, $\nu \in \mathcal{N}(T)$. Entonces,

$$\begin{aligned} T\alpha &= T\rho + T\nu \\ &= T\rho \ (\nu \in \mathcal{N}(T)) \\ &= \rho \text{ (por la parte } (i)). \qquad \blacksquare \end{aligned}$$

De acuerdo al teorema anterior, una proyección en V induce una descomposición del espacio V como suma directa de dos subespacios. Recíprocamente, veremos que si V se descompone como

$$V = W_1 \oplus \cdots \oplus W_K,$$

entonces existen operadores lineales T_i, $i = 1,\cdots,k$ tales que, para cada i, la imágen de T_i es el subespacio W_i.

Teorema 1.1.2 Sean $W_1,\cdots,W_k$ subespacios de V y suponga que

$$V = W_1 \oplus \cdots \oplus W_K.$$

Entonces existen operadores lineales $T_i : V \longrightarrow V$, $1 \le i \le k$, que satisfacen las siguientes condiciones.

(i) Para cada $i = 1,\cdots,k$, el operador T_i es una proyección.

(ii) $T_iT_j = 0$ si $1 \le i \ne j \le k$.

(iii) $\mathbf{I} = T_1 + \cdots + T_k$, donde $\mathbf{I}$ es el operador identidad en V.

(iv) Para cada $i = 1,2,\cdots,k$, la imagen de T_i es W_i.

Recíprocamente, si $T_1,\cdots,T_k$ son operadores lineales en V que satisfacen las condiciones (i)-(iii), entonces,

$$V = W_1 \oplus \cdots \oplus W_k,$$

donde $W_i := \mathcal{R}(T_i)$, $i = 1,2,\cdots,k$.

Demostración. Supongamos primero que $V = W_1 \oplus \cdots \oplus W_k$. Sea $\alpha \in V$ un vector arbitrario y expréselo como

$$\alpha = \alpha_1 + \cdots + \alpha_k, \quad \alpha_i \in W_i. \tag{1.3}$$

Ahora, defina $T_j : V \longrightarrow V$, $j = 1,\cdots,k$, mediante

$$T_j\alpha = \alpha_j, \quad \alpha \in V. \tag{1.4}$$

Es fácil ver que, para $1 \le j \le k$, (*a*) T_j es un operador lineal en V, (*b*) $\mathcal{R}(T_j) = W_j$ y (*c*) $T_j^2 = T_j$. Además, el espacio nulo de T_j es

$$\mathcal{N}(T_j) = W_1 \oplus \cdots \oplus W_{j-1} \oplus W_{j+1} \oplus \cdots \oplus W_k; \tag{1.5}$$

de hecho, la afirmación $T_j\alpha = 0$ equivale a $\alpha_j = 0$, es decir, α debe ser una suma de vectores en W_i, con $i \neq j$. Veremos ahora que los operadores T_j satisfacen las condiciones (*i*)-(*iv*). Observe que (*i*) se satisface pues, como señalamos antes, $T_j^2 = T_j$. Por otro lado, debido a (1.5), para $1 \le i \ne j \le k$, $\mathcal{R}(T_i) = W_i \subset \mathcal{N}(T_j)$ y entonces $(T_jT_i)\alpha = T_j(T_i\alpha) = 0$, $\alpha \in V$, es decir, $T_jT_i = 0$, lo cual establece la condición (*ii*). Combinando (1.3) y (1.4) vemos que

$$\alpha = T_1\alpha + \cdots + T_k\alpha, \quad \alpha \in V,$$

y entonces $I = T_1 + \cdots + T_k$, que es precisamente la condición (*iii*). Por último, hemos notado antes que para cada j, $\mathcal{R}(T_j) = W_j$, esto es, la condición (*iv*) también se satisface.

Para finalizar la demostración, suponga que $T_1, \cdots, T_k$ son operadores lineales en V satisfacen las condiciones (*i*)-(*iii*), y defina $W_i := \mathcal{R}(T_i)$, $i = 1,2 \cdots, k$. En este caso, por la condición (*iii*) se tiene que

$$\alpha = T_1\alpha + \cdots + T_k\alpha, \quad \alpha \in V; \tag{1.6}$$

claramente $T_i\alpha \in W_i = \mathcal{R}(T_i)$ para todo i, y entonces (1.6) implica que

$$V = W_1 + \cdots + W_k. \tag{1.7}$$

Por último, tenemos que demostrar que la suma en (1.7) es directa, es decir que cada $\alpha \in V$ puede expresarse de manera única como en (1.3). Para verificar esto, suponga que $\alpha \in V$ se expresa como

$$\alpha = \alpha_1 + \cdots + \alpha_k, \; \alpha_i \in W_i, \; i = 1,2,\cdots,k.$$

Veremos que en este caso, $\alpha_i = T_i\alpha$ para cada i. Para $i = 1,2,\cdots k$, seleccione $\beta_i \in W_i$ tal que $\alpha_i = T_i\beta_i$; esto es posible pues T_i es una proyección con imagen W_i. Entonces,

$$\begin{aligned} T_j\alpha &= T_j\alpha_1 + \cdots + T_j\alpha_k \\ &= T_jT_1\beta_1 + \cdots + T_jT_k\beta_k \\ &= T_j^2\beta_j \quad \text{(por la condición } (ii)) \\ &= T_j\beta_j \quad \text{(por la condición } (i)) \\ &= \alpha_j. \end{aligned}$$

Así, vemos que la descomposición de α en (1.3) es única, y por la tanto la suma en (1.7) es directa. Esto concluye la demostración del teorema. ■

1.2. Producto Interno y Proyecciones Ortogonales

En esta sección se introduce el concepto de producto interno y enseguida se estudia una clase especial de proyecciones, a saber, proyecciones ortogonales. Este tipo de operador lineal desempeña un papel de primera importancia en el análisis estadístico de un modelo lineal.

Definición 1.2.1 Sea V un espacio vectorial sobre el conjunto de los números reales. Un producto interno sobre V, es una función que a cada pareja de vectores $\alpha,\beta \in V$, asigna un escalar $\langle\alpha, \beta\rangle$ de tal modo que, para todos los vectores $\alpha,\beta,\gamma \in V$, y cualquier escalar $k \in \mathbb{R}$, se cumplen las siguientes condiciones:

(*i*) $\langle\alpha + \beta, \gamma\rangle = \langle\alpha, \gamma\rangle + \langle\beta, \gamma\rangle$;

(*ii*) $\langle k\alpha, \beta\rangle = k\langle\alpha, \beta\rangle$;

(*iii*) $\langle\beta, \alpha\rangle = \langle\alpha, \beta\rangle$;

(*iv*) $\langle\alpha, \alpha\rangle \geq 0$, y $\langle\alpha, \alpha\rangle > 0$ si $\alpha \neq \mathbf{0}$.

Ejemplo 1.2.1 Sea V un espacio vectorial de dimensión n y sea $\mathbf{Q}$ una matriz inversible de orden $n \times n$. Dada una base ordenada de V, digamos $\mathcal{B} = \{\beta_1, \cdots, \beta_n\}$, denotemos por $[\alpha]_{\mathcal{B}}$ al vector de coordenadas del $\alpha \in V$ respecto a la base $\mathcal{B}$. Mas explícitamente, $[\alpha]'_{\mathcal{B}} := (a_1, \cdots, a_n)$ esta determinado por

$$\alpha = a_1 \cdot \beta_1 + \cdots + a_n \cdot \beta_n .$$

En este caso, defina

$$\langle \alpha, \beta \rangle := [\alpha]_{\mathscr{B}}' \mathbf{Q}' \mathbf{Q} [\beta]_{\mathscr{B}}, \quad \alpha, \beta \in V.$$

No es difícil verificar que $\langle \cdot , \cdot \rangle$ definido de esta manera es un producto interno. En particular, cuando $\mathbf{Q} = \mathbf{I}$, $V = \mathbb{E}_n$ y $\mathscr{B}$ es la base canónica de $\mathbb{E}_n$, se obtiene el producto interno canónico en $\mathbb{E}_n$. □

El ejemplo anterior muestra que una matriz puede ser útil para describir un producto interno. En general, dado un espacio vectorial V y una base $\mathscr{B}$ de V, es posible representar a un producto interno en V mediante una matriz. Esto puede verse como sigue. Sea $\mathscr{B} = \{\beta_1, \cdots, \beta_n\}$ una base ordenada del espacio vectorial V el cual está provisto con el producto interno $\langle \cdot , \cdot \rangle$. Defina la matriz **H** de orden $n \times n$ mediante,

$$\mathbf{H}_{jk} := \langle \beta_k, \beta_j \rangle . \qquad (1.8)$$

Sean $\alpha = \sum_k x_k \beta_k$ y $\gamma = \sum_j y_j \beta_j$ dos vectores arbitrarios. Usando las propiedades lineales del producto interno se obtiene

$$\begin{aligned} \langle \alpha, \gamma \rangle &= \langle \sum_k x_k \beta_k, \sum_j y_j \beta_j \rangle \\ &= \sum_k x_k \langle \beta_k, \sum_j y_j \beta_j \rangle \\ &= \sum_k x_k \sum_j y_j \langle \beta_k, \beta_j \rangle \\ &= \sum_{j,k} x_k y_j \langle \beta_k, \beta_j \rangle \end{aligned}$$

$$= \sum_{j,k} y_j H_{jk} x_k \qquad \text{(de (1.8))}$$

$$= \mathbf{y}'\mathbf{H}\mathbf{x},$$

donde $\mathbf{x}$ y $\mathbf{y}$ son los vectores de coordenadas de α y γ respecto a la base $\mathscr{B}$, respectivamente, esto es, $\mathbf{x} = [\alpha]_{\mathscr{B}}$, y $\mathbf{y} = [\gamma]_{\mathscr{B}}$. La matriz $\mathbf{H}$ se llama matriz del producto interno $\langle\cdot,\cdot\rangle$ en base ordenada $\mathscr{B}$. Esta discusión se resume enseguida.

Teorema 1.2.1 Sea V un espacio vectorial de dimensión finita sobre el campo de los números reales. Suponga que $\langle\cdot,\cdot\rangle$ es un producto interno en V y sea $\mathscr{B} = \{\beta_1, \cdots, \beta_n\}$ una base ordenada de V. Defina la matriz $\mathbf{H}$ mediante

$$H_{jk} := \langle\beta_k, \beta_j\rangle, \ 1 \le i,j \le n.$$

Entonces,

(*i*) Para todo $\alpha, \gamma \in V$, $\langle\alpha, \gamma\rangle = [\gamma]_{\mathscr{B}}' \, \mathbf{H} \, [\alpha]_{\mathscr{B}}$.

(*ii*) $\mathbf{H}' = \mathbf{H}$.

(*iii*) Para todo $\mathbf{x} \in \mathbb{E}_n - \{0\}$, $\mathbf{x}'\mathbf{H}\mathbf{x} > 0$.

Demostración. La parte (*i*) fue establecida en la discusión que precede al teorema, mientras que (*ii*) y (*iii*) se obtienen combinando la definición de $\mathbf{H}$ con las propiedades (lineales) de un producto interno contenidas en la Definición 1.2.1. ■

A partir del Teorema 1.2.1(*iii*) se desprende que la matriz $\mathbf{H}$ en (1.6) es inversible. De hecho, $\mathbf{H}\mathbf{x} = \mathbf{0}$ implica que $x'Hx = 0$, lo cual implica $x = 0$, es decir, $Hx = 0$ sólo tiene la solución trivial.

El concepto de producto interno permite extender a espacios vectoriales generales la noción usual de perpendicularidad de vectores. Esta extensión se introduce a continuación.

Definicion 1.2.2 Sea V un espacio vectorial provisto con el producto interno $\langle\cdot,\cdot\rangle$.

(*i*) Los vectores α, $\beta \in V$ son ortogonales si $\langle\alpha, \beta\rangle = 0$.

(*ii*) Dos subconjuntos U, W $\subset$ V son ortogonales si $\langle\mathbf{u}, \mathbf{w}\rangle = 0$ para todos los vectores $\mathbf{u} \in U$ y $\mathbf{w} \in W$.

Para entender el significado de la noción de ortogonalidad de vectores, considere $\mathbb{E}_3$ dotado con el producto interno canónico. En este caso, $\langle\alpha, \alpha\rangle = \alpha'\alpha = \|\alpha\|^2$, donde $\|\cdot\|$ es la norma Euclideana usual, y es conocido que

$$\langle\alpha,\beta\rangle = \|\alpha\| \, \|\beta\| \cos\theta, \quad \alpha, \beta \in \mathbb{E}_3,$$

Florey (1980), donde θ es el ángulo entre los vectores α y β. En este caso es claro que $\langle\alpha, \beta\rangle = 0$ equivale a $\cos\theta = 0$, es decir, $\theta = \pi/2$ y entonces $\langle\alpha, \beta\rangle = 0$ significa que los vectores α y β son perpendiculares. De este modo, vemos que el concepto de ortogonalidad, introducido en la Definición 1.2.2 es una extensión a espacios vectoriales generales de la noción usual de perpendicularidad.

Definición 1.2.3 Un operador lineal $T:V \longrightarrow V$ es una proyección ortogonal, si y sólo si

(*i*) T es proyección, y

(*ii*) $\mathcal{R}(T) \perp \mathcal{N}(T)$.

Enseguida se introduce el concepto de operador autoadjunto, el cual se utilizará para establecer una caracterización de las proyecciones ortogonales.

Definición 1.2.4 Sea V un espacio con producto interno $\langle\cdot,\cdot\rangle$ y sea $T:V \longrightarrow V$ un operador lineal. Entonces, T es un operador autoadjunto si

$$\langle T\alpha, \beta\rangle = \langle\alpha, T\beta\rangle \text{ para todo } \alpha, \beta \in V.$$

A un operador autoadjunto también se le llama autoconjugado Golovina (1980) o simétrico Maltsev (1978). Los conceptos de proyección ortogonal y de operador autoadjunto se relacionan en el siguiente teorema.

Teorema 1.2.2 Un operador lineal $T:V \longrightarrow V$ es una proyección ortogonal, si y sólo si las siguientes condiciones (*i*) y (*ii*) son satisfechas:

(*i*) $T^2 = T$, es decir, T es una proyección.

(*ii*) T es autoadjunto.

Demostración. Suponga que $T:V \longrightarrow V$ es una proyección ortogonal. En particular T es una proyección y entonces T^2

= T, es decir, la condición (i) es satisfecha. Ahora veremos que T es autoadjunto. Sean $\mathbf{x},\mathbf{y} \in V$ dos vectores arbitrarios. Observe que

$$\mathbf{y} = \rho_y + \nu_y, \qquad \mathbf{x} = \rho_{\mathbf{x}} + \nu_y,$$

donde ρ_x, $\rho_y \in \mathcal{R}(T)$ y ν_x, $\nu_y \in \mathcal{N}(T)$. Entonces $T_x = T\rho_x + T\nu_x = T\rho_x = \rho_x$ y similarmente $T\rho_y = \rho_y$. Luego,

$$\begin{aligned} \langle T\mathbf{x},\ \mathbf{y}\rangle &= \langle \rho_x,\ \rho_y + \nu_y\rangle \\ &= \langle \rho_x,\ \rho_y\rangle, \text{ pues } \nu_y \in \mathcal{N}(T). \end{aligned}$$

Similarmente,

$$\begin{aligned} \langle \mathbf{x},\ T\mathbf{y}\rangle &= \langle \rho_x + \nu_x,\ \rho_y\rangle \\ &= \langle \rho_x,\ \rho_y\rangle, \end{aligned}$$

y concluímos que

$$\langle T\mathbf{x},\ \mathbf{y}\rangle = \langle \mathbf{x},\ T\mathbf{y}\rangle.$$

Recíprocamente, sea $T:V \longrightarrow V$ una transformación lineal que satisface (i) y (ii). De acuerdo a (i) T es proyección. Para finalizar debemos mostrar que $\mathcal{R}(T) \perp \mathcal{N}(T)$. Sean $\nu \in \mathcal{N}(T)$ y $\rho \in \mathcal{R}(T)$. Entonces,

$$\begin{aligned} \langle \nu,\ \rho\rangle &= \langle \nu,\ T\rho\rangle, \text{ (Por la condición (i))} \\ &= \langle T\nu,\ \rho\rangle, \text{ (Por la condición (ii))} \\ &= 0, \quad \text{pues } T\nu = \mathbf{0}. \end{aligned}$$

Como $\rho \in \mathcal{R}(T)$ y $\nu \in \mathcal{N}(T)$ son arbitrarios, concluimos que $\mathcal{R}(T)$ y $\mathcal{N}(T)$ son ortogonales. ■

Veremos ahora un resultado matricial que se desprende del teorema anterior. Primero, necesitamos introducir la siguiente definición.

Definición 1.2.5 Sea V un espacio vectorial de dimensión finita y $\mathcal{B} = \{\beta_1,\cdots,\beta_n\}$ una base (ordenada) de V. Dado un producto interno $\langle\cdot,\cdot\rangle$ en V, la base $\mathcal{B}$ se dice ortonormal (respecto a $\langle\cdot,\cdot\rangle$) si y sólo si satisface los siguientes condiciones:

(*i*) $\beta_i \perp \beta_j$, $i\neq j$.

(*ii*) $\|\beta_i\| = 1$, $i = 1,\cdots,n$.

Si la base $\mathcal{B}$ satisface la condición (*i*) se dice que $\mathcal{B}$ es una base ortogonal.

Dado un espacio vectorial V provisto con un producto interno, siempre es posible construir una base ortonormal. De hecho, si $\mathcal{B}_1 = \{\beta_1,\cdots,\beta_n\}$ es cualquier base de V, podemos aplicar el proceso de Gram-Schmidt Hoffman y Kunze (1973) ó Granero (1985) para construir una base ortogonal, digamos $\mathcal{B}_2= \{\alpha_1,\cdots,\alpha_n\}$. Definiendo

$$\alpha_i^* = \|\alpha_i\|^{-1}\alpha_i, \quad i = 1,\cdots,n,$$

es claro que $\|\alpha_i^*\| = 1$ y que $\mathcal{B} = \{\alpha_1^*,\cdots,\alpha_n^*\}$ es una base ortonormal.

Sea $T:V \longrightarrow V$ una proyección, $\mathcal{B} = \{\beta_1,\cdots,\beta_n\}$ una base ortonormal de V, y $A = [T]_{\mathcal{B}}$. Entonces, por el Teorema 1.1.1(*ii*), $A^2 = A$. Por otro lado, sabemos que $T\beta_j = \sum_j A_{ij}\beta_i$, y usando que $\mathcal{B}$ es una base ortonormal se desprende que

$$\langle T\beta_j,\ \beta_i\rangle = A_{ij}\beta_i,\ 1 \le i,j \le n.$$

Entonces, si T es una proyección ortogonal tenemos

$$\begin{aligned} A_{ij} &= \langle T\beta_j,\ \beta_i\rangle \\ &= \langle \beta_j,\ T\beta_i\rangle \\ &= A_{ji} \end{aligned}$$

y entonces **A** es simétrica. Recíprocamente, si $\mathbf{A} = [T]_{\mathscr{B}}$ es simétrica e idempotente, entonces T es una proyección ortogonal; una demostración de esta última afirmación puede encontrarse en Hoffman y Kunze (1973). Resumimos esta discusión en el siguiente teorema.

Teorema 1.2.3 Sea $T:V \longrightarrow V$ un operador lineal en el espacio V el cual esta provisto de un producto interno. Entonces, si T es una proyección ortogonal y $\mathscr{B}$ es una base ortonormal, $[T]_{\mathscr{B}}$ es simétrica e idempotente. Recíprocamente, si para alguna base ortonormal $\mathscr{B}$, la matriz $[T]_{\mathscr{B}}$ es simétrica e idempotente, entonces T es una proyección ortogonal.

Recordemos que una matriz cuadrada **P** es ortogonal, si y sólo si $\mathbf{P'P} = \mathbf{I}$.

Corolario 1.2.3 Sea **A** una matriz simétrica e idempotente de orden $n \times n$. Entonces, existe una matriz ortogonal **P** tal que $\mathbf{D} = \mathbf{P'AP}$ es una matriz diagonal. Mas aún, las componentes de **D** que se encuentran sobre la diagonal principal son uno o cero.

Demostración. Defina $T_{\mathbf{A}}:\mathbb{E}_n \longrightarrow \mathbb{E}_n$ mediante $T_{\mathbf{A}}\mathbf{x} = \mathbf{A}\mathbf{x}$, $\mathbf{x} \in \mathbb{E}_n$. Como $\mathbf{A}$ es simétrica e idempotente, el teorema anterior implica que $T_{\mathbf{A}}$ es una proyección ortogonal. Entonces, $\mathcal{R}(T_{\mathbf{A}})$ es ortogonal a $\mathcal{N}(T_{\mathbf{A}})$ y por el **Teorema 1.1.1**

$$\mathbb{E}_n = \mathcal{R}(T_{\mathbf{A}}) \oplus \mathcal{N}(T_{\mathbf{A}}).$$

Sean $\mathcal{B}_1 = \{\beta_1, \cdots, \beta_r\}$ y $\mathcal{B}_2 = \{\beta_{r+1}, \cdots, \beta_n\}$ bases ortonormales de $\mathcal{R}(T_{\mathbf{A}})$ y $\mathcal{N}(T_{\mathbf{A}})$, respectivamente. Entonces, la unión $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 = \{\beta_1, \cdots, \beta_r, \beta_{r+1}, \cdots, \beta_n\}$ es una base de $\mathbb{E}_n$ y satisface lo siguiente:

$$\mathbf{A}\beta_i = T_{\mathbf{A}}\beta_i = \beta_i,\ 1 \leq i \leq r,$$

y

$$\mathbf{A}\beta_i = T_{\mathbf{A}}\beta_i = \mathbf{0},\ r + 1 \leq i \leq n;$$

estas igualdades se deben a (*i*) $\beta_i \in \mathcal{R}(T_{\mathbf{A}})$ para $1 \leq i \leq r$, y (*ii*) $\beta_i \in \mathcal{N}(T_{\mathbf{A}})$ para $r + 1 \leq i \leq n$. De aquí se sigue inmediatamente que $\beta_j'\mathbf{A}\beta_i = 0$ si $i \neq j$. Mas aún,

$$\beta_i'\mathbf{A}\beta_i = 1 \text{ si } i \leq r,$$

mientras que

$$\beta_i'\mathbf{A}\beta_i = 0 \text{ para } i > r.$$

Sea $\mathbf{P}$ la matriz con columnas $\beta_1, \cdots, \beta_n$. La ortonormalidad de $\mathcal{B}$ implica que $\mathbf{P}$ es una matriz ortogonal, mientras que las dos últimas igualdades son equivalentes a

$$\mathbf{P}'\mathbf{A}\mathbf{P} = \mathrm{diag}\{1, \cdots, 1, 0, \cdots, 0\} =: \mathbf{D},$$

donde el número de unos en la diagonal principal de $\mathbf{D}$ es r, precisamente la dimensión de $\mathcal{R}(T_{\mathbf{A}})$. Esto completa la demostración del teorema. ■

Sea V un espacio vectorial de dimensión finita y W un subespacio de V. Veremos que existe un subespacio $W^{\perp}$ de V tal que, $V = W \oplus W^{\perp}$, donde cada vector en W es ortogonal a cada vector en $W^{\perp}$.

Definicion 1.2.6 Sea V un espacio vectorial provisto con producto interno y sea W un subespacio de V. El complemento ortogonal de W se denota por $W^{\perp}$ y se define como

$$W^{\perp} := \{\alpha \in V \mid \alpha \perp \mathbf{w} \text{ para todo } \mathbf{w} \in W\};$$

usando las propiedades lineales del producto interno, no es difícil ver que $W^{\perp}$ es un subespacio de V.

Teorema 1.2.4 Sea W un subespacio de V y $W^{\perp}$ el complemento ortogonal de W. Entonces V es la suma directa

$$V = W \oplus W^{\perp}$$

Demostracion. Sea $\mathcal{B} = \{\beta_1, \cdots, \beta_d\}$ una base ortonormal de W, y para cada $\alpha \in V$ defina

$$C_i := \langle \alpha, \beta_i \rangle, \ i = 1, \cdots, d.$$

Entonces,

$$\begin{aligned}\langle \alpha - \sum_{i=1}^{d} C_i \beta_i, \ \beta_j \rangle &= \langle \alpha, \beta_j \rangle - C_j \langle \beta_j, \beta_j \rangle \\ &= 0.\end{aligned}$$

Por tanto

$$\alpha = \sum_{i=1}^{d} C_i\beta_i + (\alpha - \sum_{i=1}^{d} C_i\beta_i >$$

$$= \rho + \nu,$$

donde $\rho \in W$ y $\nu \in W^{\perp}$. Como $\alpha \in V$ es arbitrario, concluímos que

$$V = W \oplus W^{\perp}. \tag{1.9}$$

Por otro lado, observe que para $\mathbf{w} \in W$ y $\mathbf{w}^{\perp} \in W^{\perp}$, $\mathbf{w} + \mathbf{w}^{\perp} = 0$ implica que $\mathbf{w} = -\mathbf{w}^{\perp}$, y entonces $\langle \mathbf{w}, \mathbf{w} \rangle = -\langle \mathbf{w}^{\perp}, \mathbf{w} \rangle = 0$, esto es $\mathbf{w} = -\mathbf{w}^{\perp} = 0$, de donde se obtiene que la suma en (1.9) es una suma directa. ■

Suponga ahora que W es un subespacio de $\mathbb{E}_n$. Queremos describir explícitamente la proyección ortogonal sobre W. El siguiente teorema nos permite encontrar la matriz de dicha proyección respecto a la base canónica.

Teorema 1.2.5 Sea W un subespacio de $\mathbb{E}_n$, y sea $\mathcal{B} = \{\beta_1, \cdots, \beta_d\}$ una base arbitraria de W. Entonces, la proyección ortogonal sobre W esta representada en la base canónica por la matriz

$$A = \mathbf{B}\,(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}',$$

donde $\mathbf{B}$ es la matriz de orden $n \times d$ con columnas $\beta_1, \cdots, \beta_d$.

Denotaremos a la proyección ortogonal en W mediante T_W. Observe que el teorema anterior establece que $T_W x = Ax$, $x \in \mathbb{E}_n$, donde $A = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$.

Demostracion. Sea $\alpha \in \mathbb{E}_n$ un vector arbitrario, y defina $\beta := \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\alpha$. Para ver que β es la proyección ortogonal de α en W, necesitamos demostrar que $(\alpha - \beta) \in W^{\perp}$. Con este fin, sea $\gamma \in W$ un vector arbitrario. Como $\mathscr{B}$ es una base de W y las columnas de **B** son precisamente los vectores de $\mathscr{B}$, se tiene que $\gamma = \mathbf{B}\delta$, para algún $\delta \in \mathbb{E}_d$. Luego,

$$
\begin{aligned}
\langle \alpha - \beta, \gamma \rangle &= \langle \alpha - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\alpha, \mathbf{B}\delta \rangle \\
&= \langle \alpha, \mathbf{B}\delta \rangle - \langle \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\alpha, \mathbf{B}\delta \rangle \\
&= \langle \alpha, \mathbf{B}\delta \rangle - \langle \alpha, \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{B}\delta \rangle \\
&= \langle \alpha, \mathbf{B}\delta \rangle - \langle \alpha, \mathbf{B}\delta \rangle \\
&= 0,
\end{aligned}
$$

donde en la tercera igualdad hemos usado el hecho de que la matriz $\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$ es simétrica, y por lo tanto representa un operador autoadjunto. Por otro lado debido a que $\gamma \in W$ es un vector arbitrario, concluimos que $\alpha - \beta \in W^{\perp}$ y, como mencionamos antes, esto implica el resultado. ■

Ahora sabemos como construir la matriz **A** tal que, $T_W x = \mathbf{A}x$ es la proyección ortogonal de x en W. Otra forma de obtener **A** geométricamente se ve a continuación (Graybill 1969):

Suponga que $\beta = T_W\alpha$ y además, sea **B** la matriz cuyas columnas forman una base de W, es claro que $\alpha - \beta \in W^{\perp}$. Se sigue entonces que,

$$(\alpha - \beta)'\mathbf{B} = 0,$$

o equivalentemente,

$$\beta'\mathbf{B} = \alpha'\mathbf{B}.$$

Debido a que $\beta \in W$ y las columnas de $\mathbf{B}$ forman una base de W, se tiene que $\beta = \mathbf{B}\delta$ para algún vector columna δ, y entonces

$$(\mathbf{B}\delta)'\mathbf{B} = \alpha'\mathbf{B},$$

y de aquí se obtiene que

$$\mathbf{B}'\mathbf{B}\delta = \mathbf{B}'\alpha.$$

Como $\mathbf{B}'\mathbf{B}$ es no singular, de la ecuación anterior se desprende que

$$\delta = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\alpha,$$

y entonces,

$$\beta = \mathbf{B}\delta = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\alpha.$$

Ejemplo 1.2.1 Sea $\mathbf{Q}$ la matriz de orden 3 x 3 definida como

$$\mathbf{Q} := \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 5 \\ 0 & 0 & 0 \end{bmatrix}.$$

Sea $\mathcal{L}\{\mathbf{Q}\}$ el subespacio de $\mathbb{E}_3$ generado por las columnas de $\mathbf{Q}$. Encontraremos la matriz $\mathbf{A}$ tal que, para cada $\mathbf{x} \in \mathbb{E}_3$, $\mathbf{A}\mathbf{x}$ es la proyección ortogonal del vector $\mathbf{x}$ en $\mathcal{L}\{\mathbf{Q}\}$, y mostraremos que la matriz $\mathbf{A}$ no depende de la elección de la base de $W = \mathcal{L}\{\mathbf{Q}\}$.

Solucion. Es claro que el rango de $\mathbf{Q}$ es 2, y dos bases de $W = \mathcal{L}\{\mathbf{Q}\}$ son $\mathcal{B}_1 := \{(1,2,0)', (1,5,0)'\}$ y $\mathcal{B}_2 := \{(2,4,0)', (1,5,0)'\}$.

Sean $\mathbf{B}_1$ y $\mathbf{B}_2$ las matrices

$$\mathbf{B}_1 = \begin{bmatrix} 1 & 1 \\ 2 & 5 \\ 0 & 0 \end{bmatrix} \quad y \quad \mathbf{B}_2 = \begin{bmatrix} 2 & 1 \\ 4 & 5 \\ 0 & 0 \end{bmatrix}.$$

Entonces, un cálculo directo muestra que

$$A = \mathbf{B}_1(\mathbf{B}_1'\mathbf{B}_1)^{-1}\mathbf{B}_1' = \mathbf{B}_2\,(\mathbf{B}_2'\mathbf{B}_2)^{-1}\mathbf{B}_2',$$

donde

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Geométricamente podemos ver a una proyección ortogonal $\mathbf{T}_W\mathbf{x} = \mathbf{A}\mathbf{x}$, como la "mejor aproximación" del vector $\mathbf{x} \in \mathbb{E}_n$ sobre el subespacio W; dicho de otra manera, $\|\mathbf{x} - \mathbf{T}_W\mathbf{x}\|$ es menor o igual que $\|\mathbf{x} - \mathbf{w}\|$, para todo $\mathbf{w} \in W$.

Teorema 1.2.6 Sea W un subespacio de V y para cada $\alpha \in \mathbb{E}_n$ sea $\beta = \mathbf{T}_W\alpha$ la proyección ortogonal de α en W. Entonces,

$$\|\alpha - \beta\|^2 \le \|\alpha - \mathbf{w}\|^2, \ \mathbf{w} \in W.$$

Demostración. Como $(\alpha - \mathbf{T}_W\alpha) \in W^{\perp}$ y $(\mathbf{T}_W\alpha - \mathbf{w}) \in W$, se tiene que $(\alpha - \mathbf{T}\,\alpha)$ y $(\mathbf{T}\,\alpha - \mathbf{w})$ son ortogonales. Entonces

$$\begin{aligned}\|\alpha - w\|^2 &= \|\alpha - T_W\alpha + T_W\alpha - w\|^2\\ &= \|\alpha - T_W\alpha\|^2 + \|T_W\alpha - w\|^2\\ &\geq \|\alpha - T_W\alpha\|^2\\ &= \|\alpha - \beta\|^2. \quad\blacksquare\end{aligned}$$

1.3. Teorema Espectral para Operadores Autoadjuntos

La diagonalización de matrices simétricas reviste un especial interés para nosotros, ya que dichas matrices son parte medular de este trabajo. El Teorema 1.3.1 que se presenta a continuación es conocido como el Teorema Espectral, y nos asegura que toda matriz simétrica tiene la propiedad, por demás interesante, de ser diagonalizable. Durante todo el resto de la sección, V es un espacio vectorial de dimensión finita dotado con un producto interno.

Teorema 1.3.1 Suponga que $T:V \longrightarrow V$ es un operador autoadjunto y sean $\lambda_1,\cdots,\lambda_k$ los valores propios distintos de **T**. Sea W_i el espacio propio de **T** correspondiente a λ_i, esto es $W_i := \mathcal{N}(T - \lambda_i I)$ y, finalmente, denote por E_i la proyección ortogonal sobre W_i, $i = 1,\cdots,k$. Entonces,

(*i*) Existe una base ortonormal de V que consiste de vectores propios de T.

(*ii*) $W_i \perp W_j$, para $1 \leq i \neq j \leq k$.

(*iii*) $V = W_1 \oplus \cdots \oplus W_k$, y

(*iv*) $T = \lambda_1 \cdot E_1 + \cdots + \lambda_K \cdot E_K$.

Para la demostración de este teorema necesitamos dos resultados preliminares los cuales se establecen en los siguientes lemas.

Lema 1.3.1 Sean $\mathbf{A}$ una matriz simétrica de orden $n \times n$ y defina

$$\mathbb{S} := \{\mathbf{x} \in \mathbb{E}_n \mid \|\mathbf{x}\| = 1\},$$

donde $\|\cdot\|$ es la norma usual en $\mathbb{E}_n$. Considere la función $\phi: \mathbb{S} \longrightarrow \mathbb{R}$ dada por

$$\phi(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}, \ \mathbf{x} \in \mathbb{S}.$$

Entonces,

(*i*) ϕ tiene un maximizador $\mathbf{x}^* \in \mathbb{S}$. Mas aún,

(*ii*) $\mathbf{x}^*$ es un vector propio de $\mathbf{A}$.

(*iii*) Si $T: V \longrightarrow V$ es autoadjunto, entonces T posee un vector propio.

Demostración. (*i*) Esta parte puede probarse usando propiedades básicas (y profundas) del sistema de números reales; el lector interesado puede encontrar una demostración en Fulks (1973), Rudin (1964), o Apostol (1974).

(*ii*) Sea $\mathbf{x}^* \in \mathbb{S}$ un máximizador de ϕ. Esto significa que $\phi(\mathbf{x}^*) \geq \phi(\mathbf{x})$, $\mathbf{x} \in \mathbb{S}$. Veremos que $\mathbf{x}^*$ es un vector propio de $\mathbf{A}$.

Sea $\mathbf{v} \perp \mathbf{x}^*$ y, sin pérdida de generalidad, suponga que $\|\mathbf{v}\| = 1$. Entonces $\|\alpha\mathbf{v} + (1 - \alpha^2)^{1/2}\mathbf{x}^*\| = 1$ para todo $\alpha \in [-1,1]$. Defina $\psi: [-1,1] \longrightarrow \mathbb{R}$ mediante

$$\psi(\alpha) = \phi(\alpha \mathbf{v} + (1 - \alpha^2)^{1/2}\mathbf{x}^*), \quad \alpha \in [-1,1].$$

Ahora observe que

$$\psi(0) \geq \psi(\alpha), \quad \alpha \in [-1,1],$$

pues el máximo de ϕ se alcanza en $\mathbf{x}^*$, que corresponde a α = 0. En particular, la desigualdad anterior dice que $\psi(\cdot)$ alcanza un máximo local en $\alpha = 0$, y entonces, ya que $\psi(\cdot)$ es una función derivable, se tiene que

$$\psi'(0) = 0,$$

donde ψ' denota la derivada de ψ. Por otro lado, un cálculo directo usando la regla de la cadena muestra que para $\alpha \in (-1,1)$,

$$\psi'(\alpha) = \text{grad } \phi(\alpha \mathbf{v} + (1 - \alpha^2)^{1/2}\mathbf{x}^*)\cdot(\mathbf{v} - [\alpha/(1 - \alpha^2)^{1/2}]\mathbf{x}^*),$$

y por lo tanto,

$$0 = \psi'(0) = \text{grad } \phi(\mathbf{x}^*)\cdot\mathbf{v},$$

y usando que grad $\phi(\mathbf{x}) = 2\cdot\mathbf{A}\mathbf{x}$, se sigue que

$$(2\mathbf{A}\mathbf{x}^*)\cdot\mathbf{v} = 0,$$

es decir, $\mathbf{A}\mathbf{x}^* \perp \mathbf{v}$. Como $\mathbf{v} \perp \mathbf{x}^*$ es arbitrario concluimos que, $\mathbf{A}\mathbf{x}^*$ pertenece al espacio lineal generado por $\mathbf{x}^*$, es decir, para algún escalar $\lambda \in \mathbb{R}$,

$$\mathbf{A}\mathbf{x}^* = \lambda \mathbf{x}^*,$$

lo cual significa que $\mathbf{x}^*$ es un vector propio de $\mathbf{A}$.

(*iii*) Sea $\mathbf{T}:V \longrightarrow V$ un operador autoadjunto y seleccione una base ortonormal de V, digamos $\mathcal{B}$. Entonces $\mathbf{A} := [\mathbf{T}]_{\mathcal{B}}$ es una matriz simétrica. Por la parte (*i*), existen $\mathbf{x} \in \mathbb{E}_n - \{\mathbf{0}\}$ y

$\lambda \in \mathbb{R}$ tales que, $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Esta igualdad es equivalente a $T\alpha = \lambda\alpha$, donde $\alpha \in V$ está determinado por $[\alpha]_{\mathcal{B}} = \mathbf{x}$. Además, $\alpha \neq \mathbf{0}$ pues $\mathbf{x}$ es no nulo, y entonces α es un vector propio de T. ∎

Una variante en la demostración de el teorema anterior puede encontrarse en Bugrov y Nikolski (1980).

Lema 1.3.2 Sea $T:V \longrightarrow V$ un operador autoadjunto. Si W es un subespacio de V invariante bajo T, entonces $W^{\perp}$ es invariante por T.

Demostración. Sea $\mathbf{v} \in W^{\perp}$ un vector arbitrario. Debemos demostrar que $T\mathbf{v} \in W^{\perp}$. Con este fin, seleccione un vector cualquiera $\mathbf{w} \in W$ y observe que

$$\begin{aligned} \langle T\mathbf{v}, \mathbf{w}\rangle &= \langle \mathbf{v}, T\mathbf{w}\rangle, \\ &= 0, \end{aligned}$$

donde hemos usado que (*a*) $T\mathbf{w} \in W$ (pues $\mathbf{w} \in W$ y W es invariante bajo T) y (*b*) $\mathbf{v} \in W^{\perp}$. En resumen, hemos demostrado que

$$\langle T\mathbf{v}, \mathbf{w}\rangle = 0, \quad \mathbf{v} \in W^{\perp}, \ \mathbf{w} \in W,$$

de donde se desprende que $\mathbf{v} \in W^{\perp} \Rightarrow T\mathbf{v} \in W^{\perp}$, es decir, $W^{\perp}$ es invariante bajo T. ∎

Demostraremos ahora el teorema 1.3.1.

Demostración (*i*). La demostración se hará por inducción en dim(V). El resultado es claro cuando dim(V) = 1. Suponga que el teorema se cumple para subespacios de dimensión menor que la de V, la cual denotaremos por n. Sea **w** un vector propio de T, el cual existe por el lema 1.3.1.(*ii*), y sea W el subespacio unidimensional generado por **w**. Como W es invariante bajo T, el lema anterior nos dice que $W^{\perp}$ también es invariante bajo T, con lo que además la restricción de T a W (denotada también por T) es un operador autoadjunto. Como $\dim(W^{\perp}) = n-1 < \dim(V)$, la hipótesis de inducción implica que existe una base ortonormal de $W^{\perp}$, digamos $\mathcal{B} = \{\beta_1, \cdots, \beta_k\}$ tal que $\mathcal{B}$ consiste de vectores propios de T, y tomando $\mathcal{B}^* = \{\beta_1, \cdots, \beta_k, \mathbf{w}\}$ vemos que $\mathcal{B}^*$ es una base de V que consiste sólo en vectores propios de T. Esto concluye la demostración de la parte (*i*).

(*ii*) Sean $\lambda_1, \cdots, \lambda_s$ los diferentes valores propios de T. Defina $W_k := \{\beta^{(k)}_{1_k}, \cdots, \beta^{(k)}_{r_k}\}$ donde los vectores $\beta^{(k)}_{1_k} \cdots \beta^{(k)}_{r_k}$ son todos los vectores propios de $\mathcal{B}^*$ que corresponden a λ_k, $k = 1, \cdots, s$. Se sigue que W_k es el espacio propio correspondiente a λ_k. Además, ya que vectores propios di diferentes de $\mathcal{B}^*$ son ortogonales se tiene que $W_k \perp W_t$ $1 \leq k \neq t \geq s$.

(*iii*) Sea W_k como en la demostración de la parte (*ii*). Como $\mathcal{B}^*$ es una base, se concluye que

$$V = W_1 \oplus \cdots \oplus W_k.$$

(*iv*) Sea E_i la proyección ortogonal sobre W_i $i = 1, \cdots, k$ y con núcleo $W_1 + \cdots + W_{j-1} + W_{j+1} + \cdots + W_k$. Por la parte (*iii*), cada $\alpha \in V$ se escribe como

$$\alpha = \alpha_1 + \cdots + \alpha_k, \ \alpha_i \in W_i, \ i = 1, \cdots, k$$

o equivalentemente

$$\alpha = E_1\alpha + \cdots + E_k\alpha.$$

Se sigue con esto que

$$\begin{aligned} T\alpha &= TE_1\alpha + \cdots + TE_k\alpha \\ &= \lambda_1 E_1\alpha + \cdots + \lambda_k E_k\alpha, \end{aligned}$$

donde hemos usado que (a) $T\mathbf{w} = \lambda_i \mathbf{w}$, $\mathbf{w} \in W_i$, y (b) $E_i\alpha \in W_i$. Como $\alpha \in V$ fue arbitrario, concluimos que

$$T = \lambda_1 E_1 + \cdots + \lambda_k E_k. \quad \blacksquare$$

Sea $T: V \rightarrow V$ un operador autoadjunto y sea $W_i = \mathcal{N}(T - \lambda_i I)$, $i = 1 \cdots k$. De acuerdo al teorema espectral, $V = W_1 + \cdots + W_k$, donde $W_i \perp W_j$. Sean

$$\mathcal{B}_1 = \{\rho_1, \cdots, \rho_{r_1}\}, \cdots, \mathcal{B}_k = \{\rho_{r_{k-1}+1}, \cdots, \rho_{r_k}\}$$

las bases ortonormales para $W_1 \cdots W_k$. Entonces, la matriz $[T]_\beta$ es de la forma

$$[T]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 I_{r_1} & & 0 \\ & & \\ 0 & & \lambda_k I_{r_k} \end{bmatrix},$$

donde $\mathcal{B} = \{\rho_1, \cdots, \rho_{r_k}\}$ es una base ortonormal de vectores propios para V.

Corolario 1.3.1. Sea **A** una matriz simétrica. Entonces, existe una matriz **P** ortogonal tal que **P'AP** es diagonal.

Demostracion Sea $T_A = \mathbb{E}_n \longrightarrow \mathbb{E}_n$ definida por $T_A x = Ax$. Entonces T_A es un operador autoadjunto. Sean $\lambda_1, \cdots, \lambda_k$ los distintos valores propios de T_A y $\mathcal{B}_1, \cdots, \mathcal{B}_k$ bases ortonormales de vectores propios de los λ_i, $i = 1, \cdots, k$ respectivamente. Considerando la base $\mathcal{B} = \{\mathcal{B}_1, \cdots, \mathcal{B}_k\}$ se tiene que $[T_A]_{\mathcal{B}} = \mathrm{diag}[\lambda_1 \cdots \lambda_1; \lambda_2 \cdots \lambda_2; \cdots; \lambda_k \cdots \lambda_k]$. Esto significa que si **P** es la matriz cuyas columnas son los vectores de $\mathcal{B}$, entonces $P^{-1}AP$ es una matriz diagonal; mas aun, como $\mathcal{B}$ es una base ortonormal, $P' = P^{-1}$, es decir, $P'AP$ es diagonal.

■

1.4. Diagonalizacion Simultanea de Matrices Simetricas

En la sección anterior vimos que toda matriz simétrica es semejante a una matriz diagonal. Recuerde que si $T:V \longrightarrow V$ es un operador autoadjunto, entonces $[T]_{\mathcal{B}}$ es simétrica. Por otro lado, sean $T:V \longrightarrow V$ y $U:V \longrightarrow V$ dos operadores autoadjuntos, veremos que existe una base $\mathcal{B}$ tal que, $[T]_{\mathcal{B}}$ y $[U]_{\mathcal{B}}$ son ambas diagonales si **T** y **U** conmutan.

Lema 1.4.1. Sean $T:V \longrightarrow V$ y $U:V \longrightarrow V$ operadores lineales que conmutan, esto es, $UT = TU$. Sean $\lambda_1, \cdots, \lambda_k$ los valores propios de **T**. Si $W_i = \mathcal{N}(T - \lambda_i I)$, $i = 1, \cdots, k$. Entonces, W_i

es invariante por **U**.

Demostracion Sea $\alpha \in W_i$. Entonces debemos ver que $U\alpha \in W_i$. Para esto observe que

$$\begin{aligned}(T - \lambda_i I)U\alpha &= (TU - \lambda_i U)\alpha \\ &= (UT - U\lambda_i)\alpha \\ &= U(T - \lambda_i I)\alpha \\ &= 0,\end{aligned}$$

pues α está en $N(T - \lambda_i I)$. Se sigue entonces que $U\alpha \in W$, y concluímos que W_i es invariante bajo **U**. ■

Teorema 1.4.1. Sean $T_1,\cdots,T_k$ operadores autoadjuntos que conmutan, es decir, $T_iT_j = T_jT_i$, $i,j = 1,\cdots,k$. Entonces, existe una base $\mathscr{B}$ tal que $[T_i]_{\mathscr{B}}$ es diagonal para cada T_i. En este caso $\mathscr{B}$ se puede escoger ortonormal.

Demostracion. Sin pérdida de generalidad suponemos que ninguno de los operadores $T_1,\cdots,T_k$ es la identidad. La demostración se hará por inducción. Si dim (V) = 1 el teorema se cumple. Supongamos que el teorema es válido para subespacios de dimensión menor que n. Sean $\lambda_1,\cdots,\lambda_k$ los valores propios de T_1. Si $W_i = \mathscr{N}(T_1 - \lambda_i I)$, por el lema anterior, W_i es invariante por cada operador que conmuta con T_1. Sean $T_{i_1},\cdots,T_{i_k}$ los operadores lineales en W_i que se obtienen por restricción de los T_j, $j = 1,\cdots,k$. Cada uno de estos operadores es autoadjunto. Como $\dim(W_i) < \dim(V)$, los operadores restringidos son diagonalizables en forma simultánea, es decir, existe una base ortonormal $\mathcal{B}_i$ de W_i, $i = 1,\cdots,k$, de vectores propios, donde los vectores en $\mathcal{B}_i$ son simultáneamente vectores propios para cada operador $T_{i_j} : W_i \longrightarrow W_i$, $j = 1,\cdots,k$. Poniendo

$$\mathcal{B} = \{\mathcal{B}_1,\cdots,\mathcal{B}_k\}$$

vemos que $\mathcal{B}$ es una base de V y que $[T_j]_{\mathcal{B}}$ es diagonal para todo j. Esta base es la base que necesitábamos para V. ■

Como producto de este teorema se tiene el siguiente resultado matricial análogo.

Corolario 1.4.1 Sean $A_1,\cdots,A_r$ matrices simétricas tal es que, $A_iA_j = A_jA_i$, $i,j = 1,\cdots,r$. Entonces, existe una matriz P ortogonal tal que, $P'A_iP$, $i=1,\cdots,r$ es una matriz diagonal.

Demostracion. Sean $T_{A_i}: \mathbb{E}_n \longrightarrow \mathbb{E}_n$, $1 \le i \le r$, operadores lineales tales que $T_{A_i}x = Ax$, y sean $A_1,\cdots,A_r$ matrices simétricas. Entonces $A_iA_j = A_jA_i$, o en forma equivalente $T_{A_i}T_{A_j} = T_{A_j}T_{A_i}$. Ahora recuerde que si A_i es simétrica, entonces T_{A_i} es autoadjunto (sección 1.2) Por el teorema anterior, existe una base $\mathcal{B}$ ortonormal tal que, la matriz $[T_{A_i}]_{\mathcal{B}}$, $i = 1,\cdots,r$, es una matriz diagonal. Sea P la matriz con columnas los vectores de $\mathcal{B}$. Se sigue entonces que $P'A_iP$, $i = 1,\cdots,r$ es una matriz diagonal.

Ilustraremos este corolario con un ejemplo.

Ejemplo 1.4.1 Considere las siguientes matrices:

$$\mathbf{A}:=\begin{bmatrix} \frac{13}{5} & \frac{2}{5} \\ -\frac{2}{5} & \frac{7}{5} \end{bmatrix},\ \mathbf{B}:=\begin{bmatrix} -1 & -2 \\ -2 & -4 \end{bmatrix}\ \text{y}\ \mathbf{C}:=\begin{bmatrix} 2 & -2 \\ -2 & -1 \end{bmatrix}.$$

Es fácil comprobar que **AB** = **BA**, **AC** = **CA** y **BC** = **CA**, es decir, cada par de matrices conmutan. Ahora considere la matriz

$$\mathbf{P}:=\begin{bmatrix} 5^{-1/2} & -2\cdot 5^{-1/2} \\ 2\cdot 5^{-1/2} & 5^{-1/2} \end{bmatrix}.$$

En este caso es fácil ver que **P'P** = **I**, es decir, **P** es ortogonal. Ahora, haciendo un cálculo directo tenemos,

$$\mathbf{P'AP}=\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix},\ \mathbf{P'BP}=\begin{bmatrix} -5 & 0 \\ 0 & 0 \end{bmatrix}\ \text{y}\ \mathbf{P'CP}=\begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix},$$

donde en cada caso, los elementos en la diagonal de cada matriz son precisamente los valores propios de **A**, **B** y **C**, respectivamente.

1.5. Matrices no Negativas

Las matrices no negativas son un tipo especial de matrices, que son ampliamente utilizadas en la representación y distribución de formas cuadráticas, asi como en modelos lineales. En particular, son de uso frecuente para representar una suma de cuadrados, y en la construcción de estadísticos de un modelo lineal.

Definicion 1.5.1 Sea A una matriz simétrica de orden $n \times n$. Entonces,

(*i*) La matriz A es positiva definida (o, simplemente positiva) si

$$x'Ax > 0, \quad x \in \mathbb{E}_n - \{0\}.$$

(*ii*) A es no negativa definida (o simplemente no negativa) si y solo si

$$x'Ax \geq 0, \quad x \in \mathbb{E}_n.$$

Denotaremos $A > 0$ si A es positiva definida y $A \geq 0$ si es no negativa. Es fácil demostrar que $A > 0$ si y solo si la función $x'Ay$, $x, y \in \mathbb{E}_n$ es un producto interno.

Teorema 1.5.1 Las siguientes afirmaciones son equivalentes

(*i*) $A > 0$;

(*ii*) La función definida por

$$\langle x, y\rangle := x'Ay, \quad x, \ y \in \mathbb{E}_n,$$

es un producto interno en $\mathbb{E}_n$;

(*iii*) Existe una matriz Q ortogonal tal que, $A = Q'Q$.

Demostracion (*i*) $\Rightarrow$ (*ii*) Para esta parte se define

$$g(x,y) := x'Ay.$$

Es fácil ver que $g(x, y)$ cumple con las tres condiciones de la Definición 1.2.1.

(*ii*) $\Rightarrow$ (*iii*) Como $\mathbf{x'Ay}$ es un producto interno, entonces existen vectores $\rho_1,\cdots,\rho_n$ tales que $\delta_{ij} = \langle \rho_i, \rho_j \rangle = \rho_i\mathbf{A}\rho_j$, donde δ_{ij} es el símbolo de Kronecker. Ahora, sea $\mathbf{R}$ la matriz con columnas $\rho_1,\cdots,\rho_n$. Entonces la componente i, j de $\mathbf{R'AR}$ es $\rho_i\mathbf{A}\rho_j$, lo cual es equivalente a $\mathbf{R'AR} = \mathbf{I}$. En particular, esta última ecuación implica que $\mathbf{R}$ es no singular, y ademas $\mathbf{A} = (\mathbf{R}')^{-1}\mathbf{R}^{-1} = (\mathbf{R}^{-1})'\mathbf{R}^{-1}$. Por lo tanto, la conclusión se sigue poniendo $\mathbf{Q} := \mathbf{R}^{-1}$.

(*iii*) $\Rightarrow$ (*i*) Supongamos que $\mathbf{A} = \mathbf{Q'Q}$. Entonces,

$$\begin{aligned} \mathbf{x'Ax} &= \mathbf{x'Q'Qx} \\ &= (\mathbf{Qx})'\mathbf{Qx} \\ &> 0, \ \mathbf{x} \neq 0. \end{aligned} \qquad \blacksquare$$

La parte (*iii*) de este teorema nos da una forma para caracterizar a una matriz positiva, aunque en la práctica no es fácil aplicar tal criterio. Existen otros criterios mas sencillos para saber si una matriz es positiva. A continuación veremos dos de ellos.

Teorema 1.5.2 La matriz simétrica $\mathbf{A}$ de orden $n \times n$ es positiva y definida si y sólo si, todos sus valores propios son mayores que cero.

Demostracion. Recuerde que si $\mathbf{A}$ es simétrica, existe una matriz $\mathbf{P}$ ortagonal tal que $\mathbf{P'AP} = \mathbf{D} = \mathrm{diag}\{\lambda_1,\cdots,\lambda_n\}$; (Teorema 1.3.1). Sean $\mathbf{x}$, $\mathbf{y} \in \mathbb{E}_n - \{0\}$. Luego tenemos que

$$x'Ax > 0 \Longleftrightarrow y'P'APy > 0 \qquad (x,\ y \neq 0)$$

$$\Longleftrightarrow y'Dy > 0$$

$$\Longleftrightarrow \lambda_i > 0,\ i = 1,\cdots,n. \qquad \blacksquare$$

El siguiente teorema determina otro criterio, Mardsen y Tromba (1987), también conocido como el criterio de Sylvester, Krutitskaya y Shiskin (1985), para saber si una matriz es positiva.

Teorema 1.1.3 Sea $A = [a_{ij}]$ una matriz de orden $n \times n$. Se Define la matriz A_k como la submatriz de A con elementos en su diagonal $a_{11},\cdots,a_{kk}$. La matriz A es definida positiva si

$$|A_k| > 0,\ k = 1,\cdots,n.$$

Una demostración de este teorema se puede encontrar en Hoffman (1973).

Hasta ahora hemos considerado solamente matrices positivas. Para matrices no negativas existe un resultado similar al Teorema 1.5.1.

Teorema 1.5.4 Sea A una matriz de orden $n \times n$. Las aseveraciones siguientes son equivalentes.

(*i*) $A \geq 0$;

(*ii*) $\lambda_i \geq 0,\ i = 1,\cdots,k$, donde los λ_i son los valores propios de A;

(*iii*) $A = Q'Q$ para alguna matriz Q.

Demostración. (*i*) $\Rightarrow$ (*ii*) Sean $\lambda_i, \cdots, \lambda_k$ los distintos valores propios de A. Sea $x_i \in \mathbb{E}_n - \{0\}$ tal que $Ax_i = \lambda_i x_i$, $i = 1, \cdots, k$. Como $A \geq 0$ se tiene que $0 \leq x_i' A x_i = \lambda_i \|x_i\|^2$, y como $x_i \neq 0$, esto inmediatamente implica que

$$\lambda_i \geq 0, \; i = 1, \cdots, k.$$

(*ii*) $\rightarrow$ (*iii*) Sea A una matriz simétrica, entonces existe una matriz P ortogonal tal que

$$P'AP = D = \mathrm{diag}\{\lambda_1, \cdots, \lambda_k, 0, \cdots, 0\}, \qquad (1.10)$$

donde los números λ_i, $i = 1, \cdots, k$ son los valores propios de A contando multiplicidades. Defina

$$D^{1/2} := \mathrm{diag}\{\lambda_1^{1/2}, \cdots, \lambda_k^{1/2}, 0, \cdots, 0\},$$

y observe que (*a*) $D = D^{1/2}D^{1/2}$, y (*b*) $D^{1/2}$ es una matriz simétrica. Además, usando (1.10) se tiene que $D^{1/2}D^{1/2} = P'AP$, lo cual es equivalente a

$$\begin{aligned} A &= (P')^{-1}D^{1/2}D^{1/2}P^{-1} \\ &= (P')^{-1}(D^{1/2})'D^{1/2}P^{-1} \\ &= Q'Q \end{aligned}$$

donde $Q := D^{1/2}P^{-1}$.

(*iii*) $\Rightarrow$ (*i*) Supongamos que $A = Q'Q$. Entonces, para $x \in \mathbb{E}_n$,

$$\begin{aligned} x'Ax &= x'Q'Qx \\ &= (Qx)'Qx \\ &\geq 0. \quad \blacksquare \end{aligned}$$

CAPITULO 2

La Distribucion Normal Multidimensional

Recuerde que si Z es una variable aleatoria unidimensional con distribución normal estándar, su función de densidad está dada por:

$$f_{z}(z) = (1/(2\pi)^{1/2} \exp(-z^{2}/2), \quad -\infty < z < \infty;$$

en este caso escribimos $Z \sim \mathcal{N}(0, 1)$. La notación está justificada , pues es bien sabido que, si Z tiene la distribución normal estándar, su media es cero y su varianza es 1. Sabemos también, que la función generadora de momentos de Z es

$$M_{z}(t) = \exp(-\ t^{2}/2), \quad t \in \mathbb{R}.$$

Suponga ahora que se tiene un conjunto de variables aleatorias con distribución normal estándar, y estamos interesados en definir explícitamente la distribución conjunta de las mismas. Cuando las variables aleatorias son independientes estamos hablando de la distribución normal estándar multidimensional.

2.1. El Caso Estandar

Sean $Z_1, \cdots, Z_p$ variables aleatorias , cada una con distribución normal estándar y formemos el vector $\mathbf{Z}$ con dichas variables. La distribución conjunta de las variables Z_i, $i = 1, \cdots, p$ — o equivalentemente, la distribución del vector aleatorio $\mathbf{Z}$ — recibe un nombre especial.

Definicion 2.1.1. El vector aleatorio $\mathbf{Z}' = (z_1, \cdots, z_p) \in \mathbb{E}_p$ tiene la distribución normal estándar p-dimensional si y sólo si,

(*i*) Z_i, $i = 1, \cdots, p$ son variables aleatorias independientes;

(*ii*) $Z_i \sim \mathcal{N}(0, 1)$, $i = 1, \cdots, p$.

Recuerde que la densidad conjunta de variables aleatorias independientes es el producto de las densidades dichas variables. Se tiene entonces que,

$$\begin{aligned} f_{\mathbf{z}}(\mathbf{z}) &= \prod_{i=1}^{p} f_{\mathbf{z}}(z_i) \\ &= \prod_{i=1}^{p} (1/(2\pi)^{1/2}\exp(-z_i^2/2) \\ &= \prod_{i=1}^{p} (1/(2\pi)^{1/2})\exp(-\tfrac{1}{2}\sum_{i=1}^{p} z_i^2) \\ &= (1/(2\pi)^{1/2})\exp(-\mathbf{z}'\mathbf{z}/2),\ \mathbf{z}' = (Z_1, \cdots, Z_p). \end{aligned}$$

Además $E(\mathbf{Z}) = [E(Z_1), \cdots, E(Z_p)]' = [0, \cdots, 0] = \mathbf{0} \in \mathbb{E}_p$.

En general, la matriz de covarianza de $\mathbf{Z}$, es aquella matriz en cuya diagonal tiene como elementos a $Var(Z_i)$, $i = 1,\cdots,p$, y fuera de su diagonal a $Cov(Z_i, Z_j)$. Se denotará a la matriz de covarianza de $\mathbf{Z}$ como $Cov(\mathbf{Z})$, y es claro que si $Z_i \sim \mathcal{N}(0, 1)$, $i = 1,\cdots,p$ entonces $Cov(\mathbf{Z}) = \mathbf{I}_p$. Esto se debe a que $Var(Z_j) = 1$, $j = 1,\cdots,p$, y como las Z_i son idependientes, $Cov(Z_i, Z_j) = 0$. Resumimos esta discusión como sigue:

Teorema 2.1.1 Si $\mathbf{Z} \in \mathbb{E}_p$ es un vector aleatorio con distribución normal estándar entonces,

(*i*) La densidad de $\mathbf{Z}$ es $f_{\mathbf{Z}}(z) = f(1/(2\pi)^{p/2})\exp(-z'z/2)$;

(*ii*) $E(\mathbf{Z}) = \mathbf{0} \in \mathbb{E}_p$ y $Cov(\mathbf{Z}) = \mathbf{I}_p$.

En adelante la notación $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p) = \mathcal{N}_p(\mathbf{0}, \mathbf{I})$ nos indicará que el vector aleatorio $\mathbf{Z}$ descrito en el teorema anterior tiene la distribución normal estándar p-dimensional.

Lema 2.1.1 Sea $\mathbf{Z}$ un vector aleatorio con $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$. Entonces su función generadora de momentos está dada por

$$M_{\mathbf{z}}(\mathbf{t}) = \exp(\tfrac{1}{2}\mathbf{t}'\mathbf{t}), \quad \mathbf{t} \in \mathbb{E}_p$$

Demostracion. Como $Z_i \sim \mathcal{N}(0, 1)$, $i = 1,\cdots,p$, entonces,

$$M_{z_i}(t_i) = \exp(\tfrac{1}{2}t_i^2), \quad i = 1,\cdots,p.$$

Además las Z_i son independientes y por consiguiente:

$$M_{\mathbf{z}}(\mathbf{t}) = \prod_{i=1}^{p} M_{z_i}(t_i)$$
$$= \prod_{i=1}^{p} \exp(\tfrac{1}{2}t_i^2)$$
$$= \exp(\tfrac{1}{2}\sum_{i=1}^{p} t_i^2),$$

esto es

$$M_{\mathbf{z}}(\mathbf{t}) = \exp(\tfrac{1}{2}\mathbf{t}'\mathbf{t}),$$

donde $\mathbf{t}' = (t_1, \cdots, t_p)$. ■

2.2. El Caso General no Singular

Si Z es una variable aleatoria con dstribución $\mathcal{N}(0, 1)$, sabemos que $x = \sigma z + \mu$ se distribuye normal con media μ y varianza σ^2. Además la función de densidad de x es

$$f_x(x) = (2\pi)^{-1/2}\exp(-(x - \mu)^2/2\sigma^2), \quad - \infty < x < \infty,$$

y su función generadora de momentos está dada por:

$$M_x(t) = \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2);$$

una desmostración de estos hechos puede encontrarse en Graybill (1976). Cómo vemos, en el caso unidimensional, una combinación lineal de una variable aleatoria con distribución normal estándar, también tiene distribución normal.

Suponga ahora que $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$. Es deseable determinar en la distribución del vector aleatorio $\mathbf{X} \in \mathbb{E}_p$, $x = \Gamma z + \gamma$, donde Γ es una matriz de orden $n \times p$ no singular, y $\gamma \in \mathbb{E}_p$, es decir, estamos interesados en la distribución de una combinación lineal de un vector con la distribución normal estándar p-dimensional.

Lema 2.2.1 Sea Γ una matriz de orden $n \times p$, γ un vector cualesquiera en $\mathbb{E}_p$ y $\mathbf{z} \in \mathbb{E}_p$ un vector aleatorio con distribución $\mathcal{N}(0, \mathrm{I})$. Defina el vector aleatorio $\mathbf{X}$ como

$$\mathbf{X} = \Gamma \mathbf{z} + \gamma$$

entonces,

(*i*) $E(\mathbf{X}) = \gamma$;

(*ii*) $\mathrm{Cov}(\mathbf{X}) = \Gamma\Gamma'$;

(*iii*) $M_x(\mathbf{t}) = \exp(\mathbf{t}'\gamma + \frac{1}{2}\mathbf{t}'\Gamma\Gamma'\mathbf{t})$, $\mathbf{t} \in \mathbb{E}_n$.

Demostracion.

$$\begin{aligned}
(i)\ E(\mathbf{X}) &= E(\Gamma z + \gamma) \\
&= \Gamma E(z) + \gamma \\
&= \gamma, \qquad (\text{pues } E(z) = \mathbf{0} \in \mathbb{E}_p).
\end{aligned}$$

$$\begin{aligned}
(ii)\ \mathrm{Cov}(\mathbf{X}) &= \mathrm{Cov}(\Gamma \mathbf{z} + \gamma) \\
&= E[(\Gamma \mathbf{z} + \gamma - \gamma)(\Gamma \mathbf{z} + \gamma - \gamma)' \\
&= E(\Gamma \mathbf{z})(\Gamma \mathbf{z})' \\
&= E(\Gamma \mathbf{z}\mathbf{z}'\Gamma') \\
&= \Gamma\Gamma', \qquad (\text{pues } E(\mathbf{z}\mathbf{z}') = \mathbf{I}_p).
\end{aligned}$$

$$\begin{aligned}
(iii)\ M_x(\mathbf{t}) &= E[\exp(\mathbf{t}'\mathbf{x})] \\
&= E[\exp(\mathbf{t}'\Gamma \mathbf{z} + \mathbf{t}'\gamma)] \\
&= \exp(\mathbf{t}'\gamma)E[\exp(\mathbf{t}'\Gamma \mathbf{z})]
\end{aligned}$$

$$= \exp(\mathbf{t}'\gamma)M_z(\Gamma'\mathbf{t})$$

$$= \exp(\mathbf{t}'\gamma)\exp[(\Gamma'\mathbf{t})'(\Gamma'\mathbf{t})/2]$$

$$= \exp(\mathbf{t}'\gamma + \mathbf{t}'\Gamma\Gamma'\mathbf{t}/2) \qquad \blacksquare$$

Definición 2.2.1 Sea $\mathbf{X}$ un vector aleatorio de dimensión p. Entonces, $\mathbf{X}$ tiene la distribución normal no singular si y sólo si, existe un vector $\mu \in \mathbb{E}_p$ y una matriz Σ de orden $p \times p$ no negativa definida tales que,

$$M_x(\mathbf{t}) = \exp[\mathbf{t}'\mu + \tfrac{1}{2}\mathbf{t}'\Sigma\mathbf{t}], \forall\mathbf{t} \in E_p.$$

En este caso escribamos $\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma)$.

Observe que dados $\mu \in \mathbb{E}_p$ y la matriz $\Sigma_{pxp} > 0$, existe un vector aleatorio con la función generadora de momentos dada; para ver esto seleccione una matriz $\mathbf{Q}$ de orden $p \times p$ tal que $\Sigma = \mathbf{QQ}'$. Luego, si $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$, por el lema anterior vemos que $\mathbf{X} = \mathbf{Q}\mathbf{z} + \mu$ tiene la función generadora de momentos deseada, y por lo tanto se tiene que $\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma)$.

Teorema 2.2.1 Si $\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma)$ entonces,

(*i*) $E(\mathbf{X}) = \mu$;

(*ii*) $\mathrm{Cov}(\mathbf{X}) = \Sigma$.

Este teorema es sólo una variante del lema 2.2.1. $\blacksquare$

Teorema 2.2.2 Sean $\mu \in \mathbb{E}_p$ y Σ_{pxp} una matriz positiva definida. Si $\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma)$, entonces $\mathbf{X}$ tiene la densidad dada por

$$f_{\mathbf{x}}(\mathbf{x}) = (2\pi)^{-p/2}|\Sigma|^{-1/2}\exp[-\tfrac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)], \mathbf{x} \in \mathbb{E}_p.$$

Demostración. Seleccione la matriz $\mathbf{Q}$ de orden $p \times n$, tal que $\mathbf{QQ}' = \Sigma$; note que $\mathbf{Q}$ es necesariamente inversible, pues Σ es positiva definida. Si $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$,

$$\mathbf{Y} = \mathbf{QZ} + \mu \sim \mathcal{N}_p(\mu, \Sigma), \tag{2.1}$$

entonces $\mathbf{X}$ e $\mathbf{Y}$ tienen la misma distribución, y por lo tanto la misma densidad. Por otro lado, la densidad de $\mathbf{Y}$ es

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{z}}(\mathbf{z})\,|\partial \mathbf{z}/\partial \mathbf{y}|, \ \mathbf{y} \in \mathbb{E}_p, \tag{2.2}$$

donde $|\partial \mathbf{z}/\partial \mathbf{y}|$ es el jacobiano de $\mathbf{z}$ respecto a $\mathbf{y}$. Usando la igualdad en (2.1), vemos que $\mathbf{Z} = \mathbf{Q}^{-1}(\mathbf{y} - \mu)$ y entonces $|\partial \mathbf{z}/\partial \mathbf{y}| = |\mathbf{Q}^{-1}|$. Sustituyendo este resultado en (2.2) se obtiene

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= (2\pi)^{-p/2}\exp(-\tfrac{1}{2}\mathbf{z}'\mathbf{z})\,|\mathbf{Q}^{-1}| \\ &= (2\pi)^{-p/2}\exp(-\tfrac{1}{2}[\mathbf{Q}^{-1}(\mathbf{y} - \mu)]'[\mathbf{Q}^{-1}(\mathbf{y} - \mu)])\,|\mathbf{Q}|^{-1} \\ &= (2\pi)^{-p/2}|\mathbf{Q}|^{-1}\exp[-\tfrac{1}{2}(\mathbf{y} - \mu)'\mathbf{Q}'^{-1}\,\mathbf{Q}^{-1}(\mathbf{y} - \mu)]. \end{aligned} \tag{2.3}$$

Por otro lado note que

$$\Sigma^{-1} = \mathbf{Q}'^{-1}\mathbf{Q}^{-1},$$

y

$$\det \Sigma = \det(\mathbf{QQ}')$$

$$= \det(\mathbf{Q})^2.$$

Usando esto, (2.3) es equivalente a

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-p/2}|\Sigma|^{-1/2}\exp[-\tfrac{1}{2}(\mathbf{y} - \mu)'\Sigma^{-1}(\mathbf{y} - \mu)].$$

Esto concluye la demostración pues $\mathbf{X}$ y $\mathbf{Y}$ tienen la misma densidad. ■

En Saxena y Surendran (1973) ó Arnold (1981) se encuentra una forma alternativa para obtener la densidad del vector $\mathbf{x}$. Para el caso general en que Σ es singular se puede recurrir a Rao (1973).

2.3. Teorema de Cochran

Sean $X_1, \cdots, X_n$ variables aleatorias independientes con $X_i \sim \mathcal{N}(0, 1)$, $i = 1, \cdots, n$. Recuerde que $U = x_i^2 + \cdots + x_n^2$ es una variable aleatoria con distribución χ^2 (en este caso con n grados de libertad) y con función generadora de momentos,

$$M_u(t) = (1 - 2t)^{-n/2}.$$

Suponga que U_1 y U_2 son variables aleatorias independientes, cada una con distribución χ^2 con n_1 y n_2 grados de libertad respectivamente; en este caso la variable aleatoria $U_1 + U_2$ también se distribuye χ^2 pero con $n_1 + n_2$ grados de libertad. Ahora generalizaremos este resultado para p variables aleatorias χ^2.

Lema 2.3.1 Si $U_1, \cdots, U_p$ son variables aleatorias con distribución χ^2 e independientes, con $n_1, \cdots, n_p$ grados de libertad respectivamente entonces $U_1 + \cdots + U_p$ se distribuye χ^2 con $n_1 + \cdots + n_p$ grados de libertad.

Demostracion. Defina $U = U_1 + \cdots + U_p$, y puesto que las U_i son independientes,

$$
\begin{aligned}
M_U(t) &= \prod_{i=1}^{p} M_{U_i}(t) \\
&= \prod_{i=1}^{p} (1-2t)^{-n_i/2} \\
&= (1 - 2t)^{-\sum_{i=1}^{p} n_i/2}
\end{aligned}
$$

y esta es la función generadora de momentos de una variable aleatoria χ^2 con $n_1 + \cdots + n_p$ grados de libertad. ■

En la demostración del teorema anterior la independencia juega un papel importante. El siguiente teorema, conocido como el Teorema de Cochran, se refiere a la independencia de variables aleatorias con distribución χ^2.

Teorema 2.3.1 Suponga que $\mathbf{X} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$ y defina Q como sigue:

$$
Q = \sum_{i=1}^{n} \mathbf{x}_i^2.
$$

Si

$$Q = Q_1 + \cdots + Q_p \tag{2.4}$$

donde Q_i, $i = 1,\cdots,p$ son formas cuadráticas no negativas en las x_i, con $r(Q_i) = n_i$, entonces la condición necesaria y suficiente para que $Q_1,\cdots,Q_p$ sean independientes con distribución χ^2 con $n_1,\cdots,n_p$ grados de libertad respectivamente es que

$$n = \sum_{i=1}^{p} n_i$$

Demostración. Suponga que Q_i tiene la distribución χ^2 con n_i grados de libertad, $i = 1,\cdots,p$, y que además las Q_i son independientes. Entonces $\sum_{i=1}^{p} Q_i$ es la suma de p variables aleatorias χ^2 independientes. Se sigue del lema anterior que $\sum_{i=1}^{p} Q_i$ también tiene distribución χ^2 con $\sum_{i=1}^{p} n_i$ grados de libertad. Por otro lado

$$\sum_{i=1}^{p} Q_i = \sum_{i=1}^{p} x_i^2,$$

y vemos que $\sum_{i=1}^{p} Q_i$ tiene distribución χ^2 con n grados de libertad. Por tanto,

$$n = \sum_{i=1}^{p} n_i .$$

Esto prueba la parte necesaria de la condición.

La suficiencia se demostrará por inducción. Para p = 1 es evidente que el teorema se cumple. Entonces, debemos demostrar que si el teorema es cierto para una descomposición de k-1 términos. lo es también para k términos; para esto, hagamos a (2.4) la transformación ortogonal $x = C_1 z$ a (2.4), la cual transforma a Q_1 en la suma

$$\sum_{i=1}^{n_1} \lambda_i z_i^2,$$

donde $\lambda_1, \cdots, \lambda_{n_1}$ son los valores propios de la matriz que caracteriza a Q_1. Asi tenemos:

$$\sum_{i=1}^{n_1} (1 - \lambda_i) z_i^2 + \sum_{i=n_1+1}^{n} z_i^2 = \tilde{Q}_2 + \cdots + \tilde{Q}_p, \qquad (2.5)$$

siendo $\tilde{Q}_2, \cdots, \tilde{Q}_p$ las transformaciones de $Q_2, \cdots, Q_p$ respectivamente. Veamos ahora que los λ_i son iguales a 1; para esto, supongamos que existen k de los λ_i diferentes a 1. Ambos miembros en (2.5) son formas cuadráticas en $z_1, \cdots, z_n$, y el rango del lado izquierdo es $n - n_1 + k$. Podemos ver también que el rango del lado derecho es $n_2 + \cdots + n_k = n - n_1$. Igualando los rangos resulta que $k = 0$ y, en consecuencia, los λ_i, $i = 1, \cdots, n_1$, de donde

$$\sum_{n_1+1}^{n} z_i^2 = \tilde{Q}_2 + \cdots + \tilde{Q}_p \qquad (2.6)$$

Aqui las variables $z_1, \cdots, z_{n_1}$ no aparecen en el primer miembro y vamos a probar que tampoco aparecen en ningún término del lado derecho de (2.6); para esto, supongamos por ejemplo, que $\tilde{Q}_2$ no es independiente de z_1, entonces $\tilde{Q}_2$ deberá contener un término az_1^2, con $a > 0$. Pero los coeficientes de z_1^2 en $Q_2, \cdots, Q_p$ son no negativos, lo cual implica una contradicción a (2.6).

Tenemos, de acuerdo a lo anterior, que (2.6) da una representación de $\sum_{n_1+1}^{n} z_i^2$ en una suma de k-1 formas no negativas de las variables $z_{n+1}, \cdots, z_n$. De acuerdo con la hipótesis , es cierto el teorema para esta descomposición, y por consiguiente, existe una transformación ortogonal con $n-n_1$ variables que transforma $z_{n+1}, \cdots, z_n$ en las variables $y_{n+1}, \cdots, y_n$, tal que

$$\tilde{Q}_2 = \sum_{i=n_1+1}^{n} y_i^2, \cdots, \tilde{Q}_k = \sum_{i=n-n_k+1}^{n} y_i^2 \qquad (2.7)$$

Si agregamos a esta tranformación las n_1 ecuaciones $z_1 = y_1, \cdots, z_{n_1} = y_{n_1}$, obtenemos una transformación ortogonal de n variables $z = \mathbf{C_2 y}$ tal que se verifica (2.7).

El resultado de efectuar $\mathbf{x} = \mathbf{C_1}z$ y $z = \mathbf{C_2 y}$ y, esta transformación es equivalente a $\mathbf{x} = \mathbf{C_1 C_2 y}$, la cual también es ortogonal pues $\mathbf{C_1}$ y $\mathbf{C_2}$ lo son. Esta transformación reune las condiciones requeridas, quedando demostrado el teorema.

■

2.4. Distribución Condicional

Sea $\mathbf{X} \in \mathbb{E}_p$ un vector aleatorio con distribución $\mathcal{N}(\mu, \Sigma)$, Σ no singular. Además consideramos la partición

$$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix}, \ \mathbf{X_1} \in \mathbb{E}_q, \ \mathbf{X_2} \in \mathbb{E}_{p-q} \text{ y } q < p.$$

Supongamos que se observó X_2, es decir, ya conocemos una parte del vector $\mathbf{X}$. Deseamos, de acuerdo a las circunstancias descritas encontrar la distribución de X_1; esto es, queremos determinar la distribución de X_1 dado X_2. La densidad condicional de X_1 dado X_2 se representa por $f_{x_1|x_2}(X_1|X_2)$, $X \in \mathbb{E}_q$.

En $\mathbb{E}_2$ por ejemplo, si $X' = (X_1, X_2)$ es un vector aleatorio. Entonces, la distribución de X_1 dado que $X_2 = x_2$ es constante está concentrada en el plano $X_2 = x_2$; para $\mathbb{E}_n$, con $n > 2$ nos referimos a un hiperplano.

En esta sección nos concretaremos a la distribución condicional para el caso no singular de la distribución normal.

Definicion 2.4.1 Sea Σ una matriz no singular particionada como

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \ \Sigma_{11} \text{ y } \Sigma_{22} \text{ cuadrados.}$$

Suponga que Σ_{22} es no singular y defina la matriz

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Teorema 2.4.1 Sea $\mathbf{X}$ un vector aleatorio con distribución

$\mathcal{N}(\mu, \Sigma)$. Considere la partición

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \\ \mathbf{X}_2 \end{bmatrix}, \ \mathbf{X}_1 \in \mathbb{E}_q, \ \mathbf{X}_2 \in \mathbb{E}_{p-q} \text{ y } q < p.$$

Sea Σ particionada como en la definición anterior, donde la partición es compatible con la de $\mathbf{X}$. Además, sean μ_1, Σ_{11}, μ_2 y Σ_{22} la media y varianza de $\mathbf{X}_1$ y $\mathbf{X}_2$ respectivamente. Entonces, la distribución condicional de $\mathbf{X}_1$ dado que $\mathbf{X}_2 = x_2$ es normal con media $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ y matriz de covarianza $\Sigma_{11.2}$.

Demostracion. Suponga que $\Sigma_{12} = \mathbf{0}$; con esto la distribución de $\mathbf{X}$ se escribe como

$$f_x(x) = \{[1/((2\pi)^{p/2}|\Sigma_{11}|^{1/2})]\exp[-\tfrac{1}{2}(\mathbf{X}_1 - \mu_1)'\Sigma_{11}^{-1}[\mathbf{X}_1 - \mu_1]\}$$

$$\{[1/((2\pi)^{(p-q)/2}|\Sigma_{22}|^{1/2})]\exp[-\tfrac{1}{2}(\mathbf{X}_2-\mu_2)'\Sigma_{11}^{-1}(\mathbf{X}_2-\mu_2)]\};$$

esto muestra que $\mathbf{X}_1$ y $\mathbf{X}_2$ son independientes y,

$$\mathbf{X}_1 \sim \mathcal{N}(\mu_1, \Sigma_{11}) \text{ y } \mathbf{X}_2 \sim \mathcal{N}(\mu_2, \Sigma_{22}).$$

Recuerde que la no correlación entre dos variables aleatorias normales, implica ausencia de dependencia lineal entre las mismas.

Definimos ahora

$$U_1 = X_1 + MX_2 \text{ y } U_2 = X_2,$$

y escogemos **M** de tal manera que U_1 y U_2 sean independientes, entonces tenemos que $E(U_1) = \mu_1 + M\mu_2$ y $E(U_2) = \mu_2$, y puesto que U_1 y U_2 son independientes, se tiene también que

$$E\{[X_1 - \mu_1 + M(X_2 - \mu_2)][X_2 - \mu_2]'\} = 0.$$

Usando la ecuación anterior, un cálculo directo muestra que $\Sigma_{12} + M\Sigma_{22} = 0$, lo cual implica

$$M = -\Sigma_{12}\Sigma_{22}^{-1}.$$

De acuerdo al valor de **M**,

$$E(U_1) = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,$$

mas aún, la matriz de covarianza de U_1 está dada por

$$E\{[X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}[X_2 - \mu_2]\}\{[X_1 - \mu_1] - \Sigma_{12}\Sigma_{22}^{-1}[X_2 - \mu_2]\}'$$

$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$= \Sigma_{11.2}.$$

Puesto que U_1 y U_2 son independientes su distribución conjunta es

$$f_u(u_1)f_u(u_2) =$$

$$\{[(2\pi)^{-q/2}|\Sigma_{11.2}|^{-1/2}]\exp[-\tfrac{1}{2}(u_1 - \mu_1)'\Sigma_{11.2}^{-1}(u_1 - \mu_1)\}$$

$$\{[(2\pi)^{-(p-q)/2}|\Sigma_{22}|^{-1/2}]\exp[-\tfrac{1}{2}(u_2 - \mu_2)'\Sigma_{22}^{-1}(u_2 - \mu_2)]\} \quad (2.8)$$

El jacobiano de la transformación $U_1 = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$ y $U_2 = X_2$ es $|I_p| = 1$. Entonces, usando (2.8) se sigue

$$f_{x_1,x_2}(x_1,x_2) =$$

$$\{[(2\pi)^{-q/2}|\Sigma_{11.2}|^{-1/2}]\exp[-\tfrac{1}{2}(x_1 - \mu_1) - \Sigma_{12}^{-1}\Sigma_{22}(x_2 - \mu_2')]$$

$$\Sigma_{11.2}^{-1}(x_1 - \mu_1) - \Sigma_{12}^{-1}\Sigma_{22}(X_2 - \mu_2)]\}\{[(2\pi)^{-(p-q)/2}|\Sigma_{22}|^{-1/2}]$$

$$\exp[-\tfrac{1}{2}(x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2)]\} \tag{2.9}$$

y esto es equivalente a

$$[(2\pi)^{-p/2}|\Sigma|^{-1/2}]\exp[-\tfrac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)$$

De (2.9) obtenemos la distribución marginal

$$f_{x_2}(X_2) = [(2\pi)^{-(p-q)/2}|\Sigma_{22}|^{-1/2}]\exp[-\tfrac{1}{2}(X_2 - \mu_2)'\Sigma_{22}^{-1}(X_2 - \mu_2)]$$

y recordando que $f_{x_1|x_2}(x_1|x_2) = f_{x_1|x_2}(x_1|x_2)/f_{x_2}(x_2)$

entonces tenemos

$$f_{x|x2}(x_1|x_2) = [(2\pi)^{-q/2}|\Sigma_{11.2}|^{-1/2}]$$

$$\exp\{-\tfrac{1}{2}[x_1 - \mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)]'\Sigma_{11.2}^{-1}[x_1 - \mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)]\}$$

Con esto se completa la demostración. ■

Ejemplo 2.4.1 Suponga que

$$\begin{bmatrix} X_1 \\ X_2 \\ X \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 3 & 1 & 200 \end{bmatrix}\right),$$

y queremos encontrar la distribución condicional de X_1 dado que

$$\begin{bmatrix} X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Aplicando las fórmulas anteriores vemos que la esperanza de X_1 dado $(X_2, X_3) = x_2' = (2, 0)$.

$$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2) = 1 + [2,3] \begin{bmatrix} 5 & 1 \\ 1 & 200 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

$$= 205/333$$

mientras que la correspondiente varianza condicional es

$$\Sigma_{11.2} = \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 1-[2,3] \begin{bmatrix} 5 & 1 \\ 1 & 200 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ 3 \end{bmatrix}$$

$$= 166/999$$

Se concluye que

$$X_1 | x_2' = (2, 0) \sim \mathcal{N}(205/333, 166/999).$$

CAPITULO 3

INVERSAS CONDICIONAL Y GENERALIZADA

La estimación de los parámetros de un modelo lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ está estrechamente relacionada con el problema de determinar la solución a un sistema de ecuaciones lineales, digamos $\mathbf{A}\mathbf{x} = \mathbf{g}$. En este sistema si $\mathbf{A}$ es inversible, entonces la solución es $\mathbf{x} = \mathbf{A}^{-1}\mathbf{g}$, sin embargo no siempre el sistema tiene solución. En este capítulo estudiaremos algunas matrices relacionadas con sistemas que no tienen solución, que no tienen solución única o en general, sistemas cuya matriz de coeficientes es arbitraria.

3.1. Inversa Condicional

Si el sistema $\mathbf{A}\mathbf{x} = \mathbf{g}$ tiene solución, entonces $\mathbf{g} \in \mathscr{L}\{\mathbf{A}\}$, y $\mathbf{x}$ es el vector de coordenadas que expresa a $\mathbf{g}$ como una combinación lineal de las columnas de $\mathbf{A}$. Ahora suponga que $\mathbf{g} \not\in \mathscr{L}\{\mathbf{A}\}$. Esto implica que para todo vector $\mathbf{x}$, se tiene que $\mathbf{A}\mathbf{x} = \mathbf{g} + \mathbf{h}$ para algún vector $\mathbf{h}$; en este caso es deseable encontrar $\mathbf{x}$ de modo que $\|\mathbf{h}\| = \|\mathbf{A}\mathbf{x} - \mathbf{g}\|$ sea mínima. La noción de inversa condicional desempeña un papel determinante en la solución de este problema.

Definicion 3.1.1 Sea **A** una matriz de orden $m \times n$. Una matriz **C** de orden $n \times m$ es una inversa condicional de **A** si sólo si

$$ACA=A. \tag{3.1}$$

Una inversa condicional de **A** se denota usualmente mediante A^c, y también se le conoce como c-inversa.

Recuerde que para una matriz **A** de orden $m \times n$ existe una matriz E no singular, tal que **EA** es una matriz escalón reducida por filas. Además, por medio de permutaciones de las columnas de **EA**, obtenidas postmultiplicando **EA** por alguna matriz no singular **P** se tiene que

$$EAP = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix},$$

donde r es el rango de **A** y **K** es cierta matriz. Note que **P** se obtiene permutando en I_n las mismas columnas que se requieren permutar a **EA** para obtener la última igualdad.

Lema 3.1.1 Sea **A** un matriz de orden $m \times n$, y sean **E** y **P** matrices no singulares de orden adecuado tales que

$$EAP = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix}, \tag{3.2}$$

donde r es el rango de **A**. Entonces para cualquier matriz **L** de orden $(n - r)\times(m - r)$ la matriz

$$C = P \begin{bmatrix} I_r & K \\ O & L \end{bmatrix} E$$

es una inversa condicional de **A**.

Demostración. Note que de (3.2) **A** se expresa como

$$A = E^{-1} \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} P^{-1},$$

y entonces

$$\begin{aligned} ACA &= E^{-1} \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} P^{-1}P \begin{bmatrix} I_r & K \\ O & L \end{bmatrix} E\, E^{-1} \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} P^{-1} \\ &= E^{-1} \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} P^{-1} \\ &= A. \end{aligned}$$

Por tanto **C** satisface 3.1. ■

Observe que el lema anterior nos dice que una inversa condicional no es única, y proporciona una forma de construir inversas condicionales. Note también que si **A** es no singular se sigue de (3.1) que $A^{-1}ACAA^{-1} = A^{-1}AA^{-1}$; y esto es equivalente a decir que $C = A^{-1}$. Veremos ahora algunos resultados que se desprenden de la definición de inversa condicional.

Lema 3.1.2 Sea $\mathbf{A}$ una matriz arbitraria

(*i*) $(\mathbf{A}^c)' = (\mathbf{A}')^c$;

(*ii*) $r(\mathbf{A}') \geq r(\mathbf{A})$;

(*iii*) las matrices $\mathbf{A}\mathbf{A}^c$ y $\mathbf{A}^c\mathbf{A}$ son indempotentes y de rango igual al de $\mathbf{A}$.

Demostración. (*i*) A partir de la definición 3.1.1 se sigue que $\mathbf{A}' = (\mathbf{A}\mathbf{A}^c\mathbf{A})' = \mathbf{A}'(\mathbf{A}^c)'\mathbf{A}'$, por lo que $(\mathbf{A}^c)'$ es una c-inversa de $\mathbf{A}'$; esto es $(\mathbf{A}^c)' = (\mathbf{A}')^c$.

(*ii*) en el lema 3.1.1, de la expresión (3.2) es claro que

$$\begin{aligned} r(\mathbf{A}^c) &= r + r(\mathbf{L}) \\ &\geq r(\mathbf{A}) \end{aligned}$$

(*iii*) Recordemos que para dos matrices $\mathbf{L}$ y $\mathbf{M}$ para las que $\mathbf{LM}$ está definida, $r(\mathbf{LM}) \leq r(\mathbf{L})$, y $r(\mathbf{LM}) \leq r(\mathbf{M})$.
Usando esto se obtiene

$$\begin{aligned} r(\mathbf{A}) &= r(\mathbf{A}\mathbf{A}^c\mathbf{A}) \\ &\geq r(\mathbf{A}^c\mathbf{A}) \\ &\geq r(\mathbf{A}), \end{aligned}$$

es decir $r(\mathbf{A}) = r(\mathbf{A}^c\mathbf{A})$. La parte $r(\mathbf{A}\mathbf{A}^c) = r(\mathbf{A})$ se demuestra en forma análoga. Para finalizar veremos $\mathbf{A}\mathbf{A}^c$ y $\mathbf{A}^c\mathbf{A}$ son idempotentes:

$$\begin{aligned} (\mathbf{A}\mathbf{A}^c)(\mathbf{A}\mathbf{A}^c) &= \mathbf{A}\mathbf{A}^c\mathbf{A}\mathbf{A}^c \\ &= \mathbf{A}\mathbf{A}^c, \end{aligned}$$

y también

$$(A^cA)(A^cA) = A^cAA^cA$$
$$= A^cA. \qquad \blacksquare$$

En particular si AA^c es simétrica, A^c es llamada inversa de mínimos cuadrados, pues para el sistema $Ax = g$, el vector

$$x = A^cg$$

hace que $\|h\| = \|Ax - g\|$ sea la mínima posible. Esto se sigue de que $Ax = AA^cg$ es la proyección ortogonal de g sobre $\mathscr{L}\{A\}$, pues AA^c es simétrica e idempotente.

Teorema 3.1.1 Sea A una matriz de orden $n \times n$, y sean $p \in \mathscr{L}\{A'\}$, $q \in \mathscr{L}\{A\}$. Entonces $p'A^cq$ es invariante por cualquier c-inversa de A.

Demostración. Sea A_1^c una c-inversa de A. Si $q \in \mathscr{L}\{A\}$ y $p \in \mathscr{L}\{A'\}$, entonces p y q se expresan como

$$p = A'v \text{ y } q = Aw,$$

para algunos vectores $v, w \in \mathbb{E}_n$. Tenemos ahora que

$$\begin{aligned} p'A_1^cq &= (A'v)'A_1^c(Aw) \\ &= v'AA_1^cAw \\ &= v'Aw \\ &= v'AA^cAw \\ &= p'A^cq, \end{aligned}$$

donde $\mathbf{A}^c$ es cualquier c-inversa de $\mathbf{A}$. ■

Corolario 3.1.1 Sea $\mathbf{X}$ una matriz de orden $m \times n$, con $r(\mathbf{x}) > 0$. Entonces $\mathbf{X}(\mathbf{X}'\mathbf{X})^c\mathbf{X}'$ es invariante por cualquier $(\mathbf{X}'\mathbf{X})^c$.

Demostración. Sean $\mathbf{x}_1, \cdots, \mathbf{x}_m$ las filas de $\mathbf{X}$, $\mathbf{x}_i \in \mathbb{E}_n$, $i = 1, \cdots, m$. Es claro que $\mathbf{x}_i' \in \mathscr{L}\{\mathbf{X}'\mathbf{X}\}$, entonces por el teorema anterior

$$k_{ij} = \mathbf{x}_i(\mathbf{X}'\mathbf{X})^c\mathbf{x}_j', \quad i,\ j = 1, \cdots, m$$

es invariante por cualquier c-inversa de $\mathbf{X}'\mathbf{X}$, luego se tiene

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^c\mathbf{X}' = \mathbf{K},$$

donde $\mathbf{K}$ es una matriz de orden $m \times m$. ■

El corolario anterior será parte fundamental en la estimación de parámetros de un modelo lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$.

Para concluir esta sección veremos una forma para calcular inversas condicionales de un matriz dada. Esta forma esta fundamentada en el lema 3.1.1.El procedimiento es como sigue:

Sea A una matriz de orden $m \times n$. Primero formaremos la matriz aumentada $[\mathbf{I}_m|\mathbf{A}]$. Luego por medio de operaciones de filas, esa matriz se lleva a la forma $[\mathbf{E}|\mathbf{R}]$, donde $\mathbf{R}$ es una matriz escalón reducida por filas y $\mathbf{EA} = \mathbf{R}$. Finalmente, se obtiene la matriz $\mathbf{P}$ tal que

$$EAP = \begin{bmatrix} I_r & k \\ 0 & 0 \end{bmatrix}, \tag{3.3}$$

para obtener **P** basta empezar con la matriz I_n y hacer las mismas permutaciones que se requieren en **R** para obtener (3.3), la matriz resultante es precisamente P. Asi, la matriz

$$C = P \begin{bmatrix} I_r & k \\ 0 & L \end{bmatrix} E \tag{3.4}$$

es una inversa condicional de **A**; distintas inversas condicionales se obtienen al variar la matriz **L**.

Ejemplo 3.1.1 Consideremos la matriz

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 4 \\ 3 & 0 & 6 \\ -1 & 4 & -2 \end{bmatrix}.$$

Queremos calcular dos inversas condicionales de **A**. Procediendo como se indicó formamos la matriz $[I_4|A]$, y reduciendo por filas obtenemos

$$\left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 1 & 0 & 0 & 2 & 1 & 4 \\ 0 & 0 & 0 & 1 & 3 & 0 & 6 \\ 0 & 0 & 0 & 0 & -1 & 4 & -2 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|ccc} 1/3 & -2/3 & 2/3 & 0 & 1 & 0 & 2 \\ 1/3 & 1/3 & -1/3 & 0 & 0 & 1 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -2 & 2 & 1 & 0 & 0 & 0 \end{array}\right].$$

Entonces, tenemos que

$$E = \begin{bmatrix} 1/3 & -2/3 & 2/3 & 0 \\ 1/3 & 1/3 & -1/3 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -2 & 2 & 1 \end{bmatrix},$$

y la matriz $\mathbf{P} = \mathbf{I}_3$ ya que no se requieren permutaciones en las columnas de **R**, la cual particionamos como:

$$R = \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0. \end{array}\right],$$

Tomando $L' = L'_1 = [0,\ 1]$, y sustituyendo en (3.4) con un cálculo directo tenemos que

$$A_1^{\,c} = P \begin{bmatrix} I_r & k \\ 0 & L \end{bmatrix} E$$

$$= \begin{bmatrix} 1/3 & -2/3 & 2/3 & 0 \\ 1/3 & 1/3 & -1/3 & 0 \\ -1 & 2 & -1 & 0 \end{bmatrix},$$

y similarmente, seleccionando $L' = L'_2 = [1,\ 0]$, se obtiene

$$A_2^{\,c} = P \begin{bmatrix} I_r & k \\ 0 & L_2 \end{bmatrix} E$$

$$= \begin{bmatrix} 1/3 & -2/3 & 2/3 & 0 \\ 1/3 & 1/3 & -1/3 & 0 \\ -1 & -2 & 2 & 1 \end{bmatrix},$$

donde es fácil ver que

$$AA_1^cA = AA_2^cA = A.$$

Distintas técnicas de cálculo de inversas condicionales se encuentran en Pringle y Rayner (1969) ó Searle (1971).

3.2. Estimación de un Modelo Lineal

Consideremos el modelo lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, donde $\mathbf{X}$ es una matriz de orden $m \times n$ con rango $r < m, n$, y suponga que $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$; en este caso la matriz es de rango incompleto. Deseamos estimar β y σ^2 asi como funciones lineales de β. Antes de ver el caso general estudiaremos el caso en que $\mathbf{X}$ es de rango completo por columnas.

Teorema 3.2.1 Sea $\varepsilon \in \mathbb{E}_n$ un vector tal que $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ y considere el modelo lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, donde $\mathbf{X}$ es una matriz de orden $n \times p$ de rango p. Entonces los estimadores de verosimilitud máxima de β y σ^2 son:

$$\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad \text{y} \quad \tilde{\sigma}^2 = \mathbf{y}'[\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}/n.$$

Demostración. Si $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ entonces $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$. Aplicando el método de verosimilitud máxima tenemos

$$L(\beta, \sigma^2 : \mathbf{y}) = [(2\pi\sigma^2)^{-n/2}]\mathrm{EXP}[-(1/2\sigma^2)(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)']$$

de aqui se sigue que

$$\begin{aligned}\mathcal{L}(\beta, \sigma^2 : \mathbf{y}) &= \log L(\beta, \sigma^2 : \mathbf{y}) \\ &= -(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)/(2\sigma^2) - \mathrm{Log}(2\pi\sigma^2)^{n/2}\end{aligned}$$

Derivando la última igualdad con respecto a σ^2 y β, se obtiene

$$(d/d\sigma^2)\mathcal{L}(\beta,\sigma^2:\mathbf{y}) = (\sigma^2)^{-2}(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)/2 - n/(2\sigma^2)$$

$$(d/d\beta)\mathcal{L}(\beta,\sigma^2:\mathbf{y}) = (2\sigma^2)^{-1}[2\mathbf{X}'\mathbf{X}\beta - 2\mathbf{X}'\mathbf{y}].$$

Ahora hay que resolver el sistema

$$\frac{d}{d\sigma^2}\mathcal{L}(\beta,\sigma^2:\mathbf{y}) = 0$$

$$\frac{d}{d\beta}\mathcal{L}(\beta,\sigma^2:\mathbf{y}) = 0,$$

el cual es equivalente a resolver

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y}$$

$$\tilde{\sigma}^2 = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta),$$

donde $\tilde{\sigma}^2$ es el estimador de σ^2. Como $\mathbf{X}'\mathbf{X}$ es no singular, los estimadores de β y σ^2 son

$$\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \text{ y}$$

$$\begin{aligned}\tilde{\sigma}^2 &= (\mathbf{y} - \mathbf{X}\tilde{\beta})'(\mathbf{y} - \mathbf{X}\tilde{\beta})/n \\ &= \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}.\end{aligned}$$

observe que $\tilde{\sigma}^2$ también se expresa como $\tilde{\sigma}^2 = \|\mathbf{y} - \mathbf{X}\tilde{\beta}\|^2/2$

Recuerde que si $\mathbf{A}$ es simétrica e idempotente y $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$, entonces $U = \mathbf{x}'\mathbf{A}\mathbf{x}$ tiene distribución χ^2 con p grados de libertad, donde p es el rango de $\mathbf{A}$. Además, $E(U) = n + 2\lambda$, donde $\lambda = \mu' A\mu/2$ es el parámetro de no centralidad (Graybill, 1976).

Lema 3.2.1 Sea $\mathbf{A}$ una matriz idempotente de orden $n \times n$. Entonces, $r(\mathbf{A}) = tr(\mathbf{A})$.

Demostración. Del álgebra lineal sabemos que $r(\mathbf{A}) = r(\mathbf{P}^{-1}\mathbf{A}\mathbf{P})$, donde $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathrm{diag}\,\{\lambda_1, \cdots, \lambda_n\}$. Además, como $\mathbf{A}$ es idempotente $\lambda_i \in \{0,1\}$, $i = 1, \cdots, n$ y con esto

$$\begin{aligned} r(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}) &= tr(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}) \\ &= r(\mathbf{A}\mathbf{P}^{-1}\mathbf{P}) \\ &= tr(\mathbf{A}). \quad \blacksquare \end{aligned}$$

Para encontrar estimadores insesgados de β y σ^2 necesitamos conocer la distribución de $\tilde{\beta}$ y $n\tilde{\sigma}^2/\sigma^2$.

Teorema 3.2.2

(*i*) $\tilde{\beta} \sim (\beta,\ \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$,

(*ii*) $\dfrac{n\tilde{\sigma}^2}{\sigma^2} \sim \chi^2$ con n-p grados de libertad.

Demostración. (*i*) $\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ es una función lineal de $\mathbf{y}$, $\tilde{\beta}$ tiene la distribución normal con esperanza

$$
\begin{aligned}
E(\tilde{\beta}) &= E[(X'X)^{-1}X'y] \\
&= (X'X)^{-1}X'E(y) \\
&= (X'X)^{-1}X'X\beta \\
&= \beta,
\end{aligned}
$$

y con covarianza

$$
\begin{aligned}
Cov(\tilde{\beta}) &= Cov[(X'X)^{-1}X'y] \\
&= (X'X)^{-1}X'Cov(y)[(X'X)^{-1}X']' \\
&= (X'X)^{-1}X'\sigma^2X'(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}.
\end{aligned}
$$

(*ii*) observe que

$$
\begin{aligned}
n\tilde{\sigma}^2/\sigma^2 &= (y/\sigma)'[I - X(X'X)^{-1}X'](y/\sigma) \\
&= z'[I-X(X'X)^{-1}X']z,
\end{aligned}
$$

donde $z = y/\sigma \sim \mathcal{N}(X\beta/\sigma, I)$. Luego como la matriz $I-X(X'X)^{-1}X'$ es idempotente y simétrica y $Cov(z) = I$,

$$
z'[I - X(X'X)^{-1}X']z \sim \chi^2
$$

con n-p grados de libertad, pues

$$
\begin{aligned}
r[I - X(X'X)^{-1}X'] &= tr[I - X(X'X)^{-1}X'] \\
&= tr(I) - tr[X(X'X)^{-1}X'] \\
&= n - r[X(X'X)^{-1}X'], \\
&= n - p.
\end{aligned}
$$

donde en la tercera igualdad utilizamos el resultado obtenido en el lema 3.2.1. Para finalizar, veremos que el parámetro λ de no centralidad es cero. De hecho,

$$
\begin{aligned}
\lambda &= \mu_z'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mu_z/2 \\
&= (\mathbf{X}\beta/\sigma)'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'](\mathbf{X}\beta/\sigma)/2 \\
&= \beta'\mathbf{X}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X}\beta/2 \\
&= \beta'[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}]\beta \\
&= 0. \qquad \blacksquare
\end{aligned}
$$

Con base en la distribución de $n\tilde{\sigma}^2/\sigma^2$ construiremos un estimador insesgado de σ^2.

Teorema 3.2.3 En el modelo lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ es posible construir estimadores insesgados de β y σ^2, los cuales están dados de la siguiente manera:

$$
\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \text{ y } \hat{\sigma}^2 = \mathbf{Y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}/(n - p)
$$

Demostración. La primera parte se sigue de que $E(\tilde{\beta}) = \beta$, y notando que $\hat{\beta} = \tilde{\beta}$. Usando el resultado obtenido en el teorema 3.2.2 podemos escribir

$$
\hat{\sigma}^2 = n\tilde{\sigma}^2/(n - p), \tag{3.5}
$$

donde $E(n\tilde{\sigma}^2) = (n - p)\sigma^2$; de aqui se desprende que

$$
\begin{aligned}
E(\hat{\sigma}^2) &= E[n\tilde{\sigma}^2/(n - p)] \\
&= \sigma^2. \qquad \blacksquare
\end{aligned}
$$

Hasta ahora hemos hablado del modelo lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ cuando $\mathbf{X}$ es de rango completo, resultados análogos se obtienen cuando $\mathbf{X}$ es de rango incompleto.

Supongamos que en el modelo $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, ε es como antes, es decir, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. Sabemos que $(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)$ se minimiza cuando β se selecciona de modo que se satisface el siguiente sistema de ecuaciones normales:

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y}.$$

Es claro que este sistema es consistente y tiene una solución dada por

$$\tilde{\beta} = (\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{y}. \qquad (3.6)$$

Con esto obtenemos,

$$\mathbf{X}\tilde{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{y} \qquad (3.7)$$

para cualquier $(\mathbf{X}'\mathbf{X})^c$.

Veamos ahora algunos resultados necesarios para encontrar los estimadores de β y σ^2 en el modelo de rango incompleto.

Lema 3.2.2 (*i*) sean $\mathbf{A}_1^c$ y $\mathbf{A}_2^c$ dos inversas condicionales de $\mathbf{A}$. Entonces la matriz $\mathbf{A}_1^c\mathbf{A}\mathbf{A}_2^c$ es también inversa condicional de $\mathbf{A}$;

(*ii*) la matriz $(\mathbf{A}'\mathbf{A})^c\mathbf{A}'$ es una inversa condicional de $\mathbf{A}$ para toda $(\mathbf{A}'\mathbf{A})^c$.

Demostración. (*i*) observe que

$$AA_1^cAA_2^cA = AA_2^cA \qquad \text{(definición 3.1.1)}$$
$$= A,$$

y entonces $A_1^cAA_2^c$ es una inverna condicional de A.

(*ii*) Note que $A(A'A)^cA'$ es simétrica y además,

$$A(A'A)^cA'A(A'A)^cA' = A(A'A)_1^cA', \quad \text{(parte (}i\text{))}$$
$$= A(A'A)^cA', \quad \text{(corolario 3.1.1)}$$

es decir , $A(A'A)^cA'$ es idempotente, y se concluye que representa a una proyección ortogonal sobre $\mathcal{L}\{A\}$, la cual denotaremos como

$$P_A = A(A'A)^cA';$$

con esto

$$A(A'A)^cA'A = P_AA$$
$$= A.$$

Por tanto $(A'A)^cA'$ es una c-inversa de A, independientemente de la matriz $(A'A)^c$ que se utilice. ∎

Por el resultado obtenido en el lema anterior, el vector (3.7)

$$X\tilde{\beta} = X(X'X)^cX'y$$

es la proyección ortogonal de y en $\mathcal{L}\{X\}$ y como sabemos, $X(X'X)^cX'$ es invariante por cualquier $(X'X)^c$; por esto, el vector $\tilde{\beta} = (X'X)^cX'y$, minimiza $\|\varepsilon\| = \|y - X\beta\|$.

Teorema 3.2.4 Considere el modelo lineal $\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$, con $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, $\mathbf{X}$ es una matriz de orden $m \times n$, con $r(\mathbf{X}) = p < m,n$ y $\beta \in \mathbb{E}_n$. Entonces, los estimadores de máxima verosimilitud de β y σ^2 están dados por

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{y} \quad \text{y} \quad \hat{\sigma}^2 = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^c\mathbf{X}']\mathbf{y}/(n - p)$$

Demostracion. La demostración se sigue de los teoremas 3.2.2 y 3.2.3, y del lema 3.2.2.

Observe que, aunque $\tilde{\beta}$ depende de la inversa condicional de $(\mathbf{X}'\mathbf{X})$ que se utilice, $\mathbf{X}\tilde{\beta}$ es constante ya que $\mathbf{X}(\mathbf{X}'\mathbf{X})^c\mathbf{X}'$ es invariante por cualquier $(\mathbf{X}'\mathbf{X})^c$.

Ahora volvemos la atención al problema de determinar si existen o no, estimadores insesgados de un función lineal de β, digamos $\ell'\beta$. Si es posible encontrar una función lineal del vector de observaciones, digamos $\mathbf{a}'\mathbf{y}$ tal que $E(\mathbf{a}'\mathbf{y}) = \ell'\beta$ para todo β, la función $\ell'\beta$ se dice que es estimable.

Teorema 3.2.5 Sea $\ell'\beta$ una función lineal del vector de parámetros β del modelo lineal $\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$.Entonces, las siguientes afirmaciones son equivalentes.

(*i*) Existe un vector constante $\mathbf{a} \in \mathbb{E}_n$ tal que $E(\mathbf{a}'\mathbf{y}) = \ell'\beta$ para todo β,

(*ii*) Existe $\mathbf{a} \in \mathbb{E}_n$ tal que $\mathbf{a}'\mathbf{X} = \ell'$.

Note que este resultado establece que $\ell'\beta$ es una función estimable si y sólo si, ℓ' pertenece al espacio de filas de $\mathbf{X}$.

Demostración. (i) $\Rightarrow$ (ii). Suponga que $E(\mathbf{a}'\mathbf{y}) = \ell'\beta$, para todo β. Entonces, ya que $E(\mathbf{y}) = \mathbf{X}\beta$, se tiene que $\mathbf{a}'\mathbf{X}\beta = \ell'\beta$ para todo $\beta \in \mathbb{E}_p$, lo cual implica que $\mathbf{a}'\mathbf{X} = \ell'$, es decir, la condición (ii) es válida.

(ii) $\Rightarrow$ (i). Si $\mathbf{a}'\mathbf{X} = \ell'$, para algún $\mathbf{a} \in \mathbb{E}_n$, se tiene que $E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\mathbf{X}\beta = \ell'\beta$ para todo β y entonces (i) ocurre. ■

Corolario 3.2.1 La función $\ell'\beta$, lineal de β es estimable si y solo si ocurre alguna de las condiciones (i) ó (ii),en seguida

(i) $\ell'\mathbf{X}^c\mathbf{X} = \ell'$.

(ii) $\ell'(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{X} = \ell'$.

Demostración. Por el teorema anterior sabemos que $\ell'\beta$ es estimable si y sólo si $\ell' = \mathbf{a}'\mathbf{X}$ para algún $\mathbf{a} \in \mathbb{E}_n$. Ahora observe que

$$\begin{aligned} \ell = \mathbf{X}'\mathbf{a}, \text{ para algún } \mathbf{X} &\Rightarrow \ell'\mathbf{X}^c\mathbf{X} = \mathbf{a}\mathbf{X}\mathbf{X}^c\mathbf{X} \\ &\Rightarrow \ell'\mathbf{X}^c\mathbf{X} = \mathbf{a}'\mathbf{X} \\ &\Rightarrow \ell'\mathbf{X}^c\mathbf{X} = \ell' \end{aligned}$$

Recíprocamente,

$\ell'\mathbf{X}^c\mathbf{X} = \ell \Rightarrow \mathbf{a}'\mathbf{X} = \ell'$ para $\mathbf{a} = \ell'\mathbf{X}^c$. Luego, vemos que $\ell'\beta$ es estimable si y sólo si (i) ocurre. Para concluir observe que

$$\ell = \mathbf{a}'\mathbf{X} \Rightarrow \ell'(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{X} = \mathbf{a}'\mathbf{X}(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{X}$$

$$= \mathbf{a}'\mathbf{X} \text{ (por el lema 3.2.2)}$$

$$= \ell'.$$

Recíprocamente,

$$\ell'(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{X} = \ell \Rightarrow \mathbf{a}'\mathbf{X} = \ell' \text{ con } \mathbf{a} = \ell'(\mathbf{X}'\mathbf{X})^c\mathbf{X}'. \qquad \blacksquare$$

Para concluir esta sección, obtendremos la esperanza y varianza del estimador $\ell'\hat{\beta}$ de la función estimable $\ell'\beta$.

Teorema 3.2.6 Sea $\ell'\beta$ una función estimable. Entonces la esperanza y la varianza de $\ell'\beta$ están dadas como sigue:

$$E(\ell'\hat{\beta}) = \ell'\beta, \text{ y } \operatorname{var}(\ell'\hat{\beta}) = \sigma^2\ell'(\mathbf{X}'\mathbf{X})^c\ell,$$

donde $(\mathbf{X}'\mathbf{X})^c$ es cualquier c-inversa de $\mathbf{X}'\mathbf{X}$.

Demostración. Note que, para todo $\beta \in \mathbb{E}_p$

$$E(\ell'\beta) = E[\ell'(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{y}]$$

$$= \ell'(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{X}\beta,$$

$$= \ell'\beta,$$

donde hemos usado la parte (*ii*) del corolario anterior para obtener la última igualdad.

Ahora demostraremos la fórmula para $\operatorname{Var}(\ell'\hat{\beta})$.

$$\operatorname{Var}(\ell'\hat{\beta}) = \ell'\operatorname{Var}(\hat{\beta})\ell$$

$$= \ell'\operatorname{Var}[(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\mathbf{y}]\ell$$

$$= \ell'(\mathbf{X}'\mathbf{X})^c\mathbf{X}'\operatorname{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^c\ell$$

$$= \sigma^2 \ell'(X'X)^c X'X(X'X)^c \ell$$

$$= \sigma^2 \ell'(X'X)^c \ell$$

$$= \sigma^2 \ell'(X'X)^c \ell,$$

donde hemos usamos la parte (*ii*) del corolario 3.2.1. para obtener la tercera igualdad. ■

3.3. Inversa Generalizada

En la sección 3.2 vimos la aplicación de la inversa condicional al problema de la estimación de los parámetros de un modelo lineal de rango completo $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, donde encontramos que $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ y que la matriz $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ es una c-inversa de $\mathbf{X}$. Esta matriz $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ es una inversa condicional de $\mathbf{X}$ con propiedades muy especiales, y recibe el nombre de inversa generalizada. El propósito de esta Sección es introducir y estudiar este concepto.

Definición 3.1.1 La matriz **G** es la inversa generalizada (o g-inversa) de **A** si cumple las condiciones (*i*) - (*iv*) que aparecen enseguida.

(*i*) $\mathbf{AGA} = \mathbf{A}$

(*ii*) $(\mathbf{AG})' = \mathbf{AG}$

(*iii*) $\mathbf{GAG} = \mathbf{G}$

(*iv*) $(\mathbf{GA})' = \mathbf{GA}$.

No es difícil demostrar que una matriz **A** posee exactamente una g-inversa, la cual denotaremos por A^-; esto se demuestra en el teorema 3.3.2. Note que si A es de orden $m \times n$, entonces A^- es de orden $n \times m$.

De la definición 3.3.1 se desprende que las matrices AG y GA son simétricas e idempotentes. Veremos enseguida algunas propiedades sencillas de A^-.

Teorema 3.3.1 Para toda matriz A las matrices A, AA^-A, A^-A y A^- tienen el mismo rango.

Demostración. Como $A = AA^-A$ es claro que $r(A) = r(AA^-A)$. Entonces, $r(A) = r(AA^-A) \leq r(A^-A) \leq r(A)$, esto es $r(A) = r(A^-A)$. Similarmente, $r(A) = r(AA^-A) \leq r(AA^-) \leq r(A)$ y entonces $r(A) = r(AA^-)$. Por último, $A = AA^-A$ implica que $r(A) \leq r(A^-)$, y además, $A^- = A^-AA^-$ implica $r(A^-) \leq r(A)$ de donde se desprende que $r(A) = r(A^-)$. ■

Corolario 3.3.1. Para toda matriz A las siguientes condiciones (i) - (iv) se satisfacen.

(i) $\mathcal{R}(AA^-) = \mathcal{R}(A)$;

(ii) $\mathcal{N}(AA^-) = \mathcal{R}(A)^{\perp}$;

(iii) $\mathcal{R}(A^-A) = \mathcal{R}(A^-)$;

(iv) $\mathcal{N}(A^-A) = \mathcal{N}(A)$.

Demostración. (i) Es claro que $\mathcal{R}(A) \supset \mathcal{R}(AA^-)$. Además, por el teorema anterior, se tiene que $r(AA^-) = r(A)$, y entonces, $\mathcal{R}(AA^-) = \mathcal{R}(A)$ se desprende de inmediato.

(ii) recuerde que AA^- es una proyección ortogonal, entonces $\mathcal{N}(AA^-) = \mathcal{R}(AA^-)^\perp = \mathcal{R}(A)^\perp$.

(iii) como $\mathcal{R}(A^-) \supset \mathcal{R}(A^-A)$ y $r(A^-) = r(A^-A)$, las imágenes de A^- y A^-A tienen la misma dimensión, con esto se tiene el resultado $\mathcal{R}(A^-A) = \mathcal{R}(A^-)$.

(iv) observe que $A^-Ax = 0 \Rightarrow AA^-Ax = 0 \Rightarrow Ax = 0 \Rightarrow A^-Ax = 0$. Además, es claro que $Ax = 0 \Rightarrow AA^-x = 0$. Luego, se tiene que $A^-Ax = 0 \Leftrightarrow Ax = 0$, y esto significa que $\mathcal{N}(A^-A) = \mathcal{N}(A)$.

■

El siguiente corolario nos proporciona una interpretación geométrica la manera en que A^- actúa en relación a la matriz A.

Corolario 3.3.2. Para toda matriz A,

(i) $\mathcal{R}(A^-) = \mathcal{N}(A)^\perp$;

(ii) $\mathcal{N}(A^-) = \mathcal{R}(A)^\perp$.

Demostración. (i) Usando que A^-A es una proyección ortogonal, junto con la parte (iii) del corolario 3.3.1, vemos que $\mathcal{R}(A^-) = \mathcal{R}(A^-A) = \mathcal{N}(A^-)^\perp = \mathcal{N}(A)^\perp$.

(ii) Usando la parte (i) con A reemplazada por A^-, vemos que $\mathcal{R}[(A^-)^-] = \mathcal{N}(A^-)^\perp$, es decir, $\mathcal{R}(A) = \mathcal{N}(A^-)^\perp$. ■

Como mencionábamos al principio, la inversa generalizada tiene la propiedad de ser única.

Teorema 3.3.2. Sea A una matriz arbitraria. Suponga que A_1^- y A^- son inversas generalizadas de A. Entonces $A_1^- = A_2^-$.

Demostración. Multiplicando $A = AA_1^-A$ por A_2^- en ambos lados de la igualdad tenemos que

$$AA_2^- = AA_1^-AA_2^-$$

y usando la definición de g-inversa tenemos $AA_2^- = AA_1^-AA_2^- = [(AA_1^-)(AA_2^-)]' = (AA_2^-)'(AA_1^-) = (AA_2^-)(AA_1^-) = AA_1^-$. En resumen,

$$AA_2^- = AA_1^-. \tag{3.8}$$

Análogamente se concluye que

$$A_1^-A = A_2^-A. \tag{3.9}$$

Utilizando (3.8) y (3.9) se sigue que

$$\begin{aligned} A_1^- &= A_1^-AA_1^- \\ &= (A_1^-A)A_1^- \\ &= (A_2^-A)A_1^- \\ &= A_2^-(AA_1^-) \\ &= A_2^-AA_2^- \\ &= A_2^-. \quad \blacksquare \end{aligned}$$

Como lo constatamos en la sección 3.2, la solución al sistema $Ax = g$ de ecuaciones lineales tiene repercusiones significativas que son de interés en estadística. Ahora veremos como caracterizar un sistema que tiene solución, y la manera de determinar las soluciones utilizando el concepto de inversa generalizada.

Lema 3.3.1 El sistema de ecuaciones lineales $\mathbf{Ax} = \mathbf{g}$ tiene solución si y sólo si, $\mathbf{AA^{-}g} = \mathbf{g}$.

Demostración. Supóngase primero que el sistema tiene solución, entonces $\mathbf{g} \in \mathscr{L}\{\mathbf{A}\}$, y como $\mathbf{AA^{-}}$ es la proyección ortogonal sobre $\mathscr{L}\{\mathbf{A}\}$ ($\mathbf{AA^{-}} = \mathbf{P_A}$) es claro que

$$\mathbf{P_A g} = \mathbf{g},$$

es decir

$$\mathbf{AA^{-}g} = \mathbf{g}.$$

Inversamente si $\mathbf{AA^{-}g} = \mathbf{g}$, entonces poniendo $\mathbf{x_o} = \mathbf{A^{-}g}$ se tiene que $\mathbf{Ax_o} = \mathbf{g}$, y $\mathbf{x} = \mathbf{x_o}$ es solución de $\mathbf{Ax} = \mathbf{g}$.

■

Ya tenemos una forma de caracterizar a un sistema consistente de ecuaciones lineales, nos resta encontrar una expresión que nos determine todas las soluciones.

Teorema 3.3.3 Sea $\mathbf{A}$ una matriz de orden $m \times n$, y suponga que el sistema $\mathbf{Ax} = \mathbf{g}$ es consistente. Entonces, todas las soluciones $\mathbf{x_o}$ son de la forma

$$\mathbf{x_o} = \mathbf{A^{-}g} + (\mathbf{I} - \mathbf{AA^{-}})\mathbf{h},$$

con $\mathbf{h} \in \mathbb{E}_n$ un vector arbitrario.

Demostración. Suponga que $\mathbf{Ax} = \mathbf{g}$ es un sistema consistente y sea $x_o = A^-g + (I - AA^-)h$ para algún $\mathbf{h}$. Entonces,

$$\begin{aligned} Ax_o &= AA^-g + A(I - AA^-)h \\ &= AA^-g + Ah - AA^-Ah \\ &= g + 0 \\ &= g \end{aligned}$$

Recíprocamente si x_o es solución a $Ax = g$ veremos que x_o se expresa en la forma $x_o = A^-g + (I - AA^-)h$ para algún vector $\mathbf{h}$. De hecho $x_o = A^-g + (I - AA^-)x_o$ pues el lado derecho de esta igualdad es

$$\begin{aligned} A^-g + x_o - AA^-x_o &= A^-g + x_o - AA^-Ag \\ &= A^-g + x_o - A^-g \\ &= x_o. \quad \blacksquare \end{aligned}$$

Suponga que el sistema $Ax = g$ es inconsistente, como consecuencia de esto $g \notin \mathscr{L}(A)$ y entonces $Ax = g + h$ para algún vector $\mathbf{h}$.

Definición 3.3.2. El vector x_o se define como la mejor solución aproximada al sistema de acuaciones $Ax = g$ si y sólo si:

(i) $\|Ax - g\|^2 \geq \|Ax_o - g\|^2$, para todo vector $\mathbf{x}$, y

(ii) si $\mathbf{x}$ es tal que $\|Ax - g\|^2 = \|Ax_o - g\|^2$, entonces $\|\mathbf{x}\| \geq \|x_o\|$.

La mejor solución x_o al sistema $Ax = g$ ocurre precisamente cuando $x_o = A^- g$; esta es una propiedad por demás interesante de la inversa generalizada.

Teorema 3.3.4 La mejor solución aproximada x_o al sistema de ecuaciones lineales $Ax = g$ es $x_o = A^- g$.

Demostración. Primero observe que $A(x - A^- g) \in \mathcal{R}(A)$, y note además $(AA^- - I)A = AA^- A - A = 0$. Como $AA^- - I$ esto significa que las columnas de $AA^- - I$ son ortogonales a las columnas de A. Equivalentemente, para todo vector h, ocurre que $(AA^- - I)h \in \mathcal{R}(A)^\perp$. Entonces para todo $x \in \mathbb{E}_n$, se tiene que $A(x - A^- g)$ es ortogonal a $(AA^- - I)g$, lo cual implica que para $x \in \mathbb{E}_n$,

$$
\begin{aligned}
\|Ax - g\|^2 &= \|Ax - AA^- g + AA^- g - g\|^2 \\
&= \|A(x - A^- g) + (AA^- + I)g\|^2 \\
&= \|A(x - A^- g)\|^2 + \|(AA^- + I)g\|^2 \\
&\geq \|(AA^- + I)g\|^2 \\
&= \|Ax_o - g\|^2,
\end{aligned}
$$

donde $x_o = A^- g$.

Para la segunda parte, suponga que x_1 es un vector tal que $Ax_1 = AA^- g$. Hay que ver que $\|x_1\| \geq \|x_o\|$. Primero observe que $Ax_1 = AA^- g \Rightarrow A(x_1 - A^- g) = 0 \Rightarrow x_1 - A^- g \in \mathcal{N}(A)$. Además, sabemos que $A^- g \in \mathcal{R}(A^-) = \mathcal{N}(A)^\perp$. Luego $A^- g$ es ortogonal a $x_1 - A^- g$. Entonces

$$\|x_1\|^2 = \|x_1 - A^-g + A^-g\|^2$$
$$= \|x_1 - A^-g\|^2 + \|A^-g\|^2$$
$$\geq \|A^-g\|^2 = \|x_o\|^2. \quad \blacksquare$$

Si $\mathbf{X}$ es una matriz de rango completo por columnas, es fácil ver que $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}^-$. Ahora suponga que $\mathbf{X}$ es de rango completo por filas. Definimos $\mathbf{A} = \mathbf{X}'$ entonces, $\mathbf{A}'$ es de rango completo por columnas, y entonces

$$(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \mathbf{A}^-, \tag{3.10}$$

Reemplazando $\mathbf{A}$ por $\mathbf{X}'$ en (3.10), se obtiene

$$(\mathbf{X}')^- = (\mathbf{X}''\mathbf{X}')^{-1}\mathbf{X}''$$
$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}''$$
$$= [\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}]'.$$

Por otro lado, la g-inversa de una matriz es también una inversa condicional, y entonces $(\mathbf{X}')^- = (\mathbf{X}^-)'$. Esto implica que $(\mathbf{X}^-)' = [\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}]'$,o equivalentemente,

$$\mathbf{X}^- = \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}.$$

Nos apoyaremos en los resultados anteriores para encontrar una expresión para la inversa generalizada de una matriz arbitraria.

Teorema 3.3.5. Sea $\mathbf{A}$ una matriz de orden $m \times n$, con $r(\mathbf{A})=p < m, n$. Suponga que $\mathbf{A} = \mathbf{BC}$ donde $\mathbf{B}$ es de orden $m \times p$ y $\mathbf{C}$ es de orden $p \times n$; note que en este caso $r(\mathbf{B}) = r(\mathbf{C}) = p$. Entonces,

$$A^- = C^- B^-.$$

Demostración. Veremos que $\mathbf{C}^-\mathbf{B}^-$ cumple con las cuatro condiciones en la definición de inversa generalizada. (*i*) Como observamos en el enunciado del teorema, $\mathbf{B}$ es de rango completo por columnas y $\mathbf{C}$ de rango completo por filas. Luego,

$$\mathbf{B}^- = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}', \text{ y } \mathbf{C}^- = \mathbf{C}(\mathbf{C}\mathbf{C}')^{-1}.$$

Además, es fácil verificar que $\mathbf{B}^-\mathbf{B} = \mathbf{C}\mathbf{C}^- = \mathbf{I}_p$. Entonces

$$\begin{aligned} A(\mathbf{C}^-\mathbf{B}^-)A &= \mathbf{B}\mathbf{C}\mathbf{C}^-\mathbf{B}\mathbf{B}^-\mathbf{C} \\ &= \mathbf{B}(\mathbf{I}\cdot\mathbf{I})\mathbf{C} \\ &= \mathbf{B}\mathbf{C} \\ &= A. \end{aligned}$$

(*ii*) AA^- se expresa como $A\mathbf{C}^-\mathbf{B}^-$, de aqui que $A(\mathbf{C}^-\mathbf{B}^-) = \mathbf{B}\mathbf{C}(\mathbf{C}\mathbf{B})^- = \mathbf{B}\mathbf{C}\mathbf{C}^-\mathbf{B}^- = \mathbf{B}\mathbf{B}^-$ es una matriz simétrica.

La verificación de que $\mathbf{G} = \mathbf{C}^-\mathbf{B}^-$ satisface las condiciones (*iii*) y (*iv*) en la definición de g-inversa se hace de modo similar. ■

Bajo las condiciones del teorema 3.3.5, podemos usar el teorema 3.3.4 para concluir que

$$A^- = \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'; \qquad (3.11)$$

A = **BC** es una factorización de rango completo de **A**. En conclusión: el cálculo de la inversa generalizada se reduce al problema de encontrar una factorización del tipo mencionado. Para obtener una factorización de rango completo se procede como sigue:

Sea **A** una matriz de orden $m \times n$, con $r(\mathbf{A}) = p$. Primero se reduce por filas la matriz **A** obteniendo **R**, las columnas cuyo único elemento sea 1 corresponden columnas independientes de **A**; formamos con estas columnas la matriz **B** y con las p filas no nulas de **R** la matriz **C**. Entonces, **A** = **BC** es una factorización de rango completo.

Utilizando el método anterior junto con (3.11) podemos calcular la inversa generalizada de una matriz.

Ejemplo 3.3.1 Sea A la matriz

$$\begin{bmatrix} 1 & 2 & 0 \\ 3 & 6 & -1 \\ 2 & 4 & 5 \\ 1 & 2 & 8 \end{bmatrix}.$$

Utilizaremos el método descrito para encontrar una factorización de rango completo de **A**.

Realizando operaciones de filas se obtiene

$$\begin{bmatrix} 1 & 2 & 0 \\ 3 & 6 & -1 \\ 2 & 4 & 5 \\ 1 & 2 & 8 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Luego

$$\mathbf{R} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

y las columnas independientes de **A** son la primera y la tercera. Asi, **B** y **C** están dadas por

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 3 & -1 \\ 2 & 5 \\ 1 & 8 \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

Con un cálculo directo muestra que $\mathbf{A} = \mathbf{BC}$ y que

$$\mathbf{A}^{-} = \mathbf{C}'(\mathbf{CC}')^{-1}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$$

$$= \frac{1}{375} \begin{bmatrix} 6 & 19 & 7 & -2 \\ 12 & 38 & 14 & -4 \\ -5 & -20 & 15 & 35 \end{bmatrix}.$$

Se puede comprobar que directamente $\mathbf{A}^{-}$ dada arriba satisface las condiciones de la definición 3.3.1. □

Para un desarrollo detallado y profundo de los conceptos de inversa condicional e inversa generalizada, el lector puede recurrir a Den-Israel y Greville (1974).

3.4. Distribución Condicional para la Distribución Normal General

Suponga que el vector aleatorio $\mathbf{X} = [X_1, \cdots, X_n]'$ tiene la distribución normal multivariada, donde sólo conjuntos de tamaño $p < n$ de las X_i, $i = 1, \cdots, n$ son independientes. En este caso la distribución de $\mathbf{X}$ es de rango p, y cada una de las variables $X_1, \cdots, X_n$ se expresa como combinación lineal de p variables independientes. En un sentido geométrico podemos decir que la masa total de la distribución está contenida en cierto subespacio de $\mathbb{E}_n$. Conjuntamos este comentario en el siguiente resultado:

Definición 3.4.1. Sea $\mathbf{X} = [X_1, \cdots, X_n]'$ un vector aleatorio con distribución normal. El rango de la distribución es $p \leq n$ si y sólo si, p es la máxima dimensión de un conjunto de las X_i, $i = 1, \cdots, n$, que son independientes. Cuando $p < n$ se tiene la distribución normal singular.

Generalizaremos en este sección el resultado obtenido en la sección 2.4 para la distribución condicional de la distribución normal multivariada.

Lema 3.4.1 Sea Σ una matriz simétrica no negativa definida particionada como sigue.

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Entonces $\mathcal{R}(\Sigma_{21}) \subset \mathcal{R}(\Sigma_{22})$. Equivalentemente, existe una matriz $\mathbf{Q}$ tal que

$$\Sigma_{21} = \Sigma_{22}\mathbf{Q}.$$

Demostración. Dividiremos la demostración en dos partes.

(a) Suponga que $\mathbf{A}$ es una matriz cuadrada no negativa definida de orden $n \times n$. Entonces, si $\mathbf{x} \in \mathbb{E}_n$ satisface $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$, se tiene que $\mathbf{x}'\mathbf{A} = \mathbf{0}$. Para demostrar esta afirmación, sea $\mathbf{x} \in \mathbb{E}_n$ tal que $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$, y recuerde que existe una matriz $\mathbf{P}$ tal que $\mathbf{A} = \mathbf{P}'\mathbf{P}'$. Entonces, $0 = \mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{P}'\mathbf{P}\mathbf{x} = (\mathbf{P}\mathbf{x})'(\mathbf{P}\mathbf{x})$ y entonces, $(\mathbf{P}\mathbf{x})' = \mathbf{0}$, de donde se desprende que $\mathbf{x}'\mathbf{A} = \mathbf{x}'\mathbf{P}'\mathbf{P} = (\mathbf{P}\mathbf{x})'\mathbf{P} = \mathbf{0}$, como se deseaba demostrar.

(b) Observe que $\mathcal{R}(\Sigma_{21})^{\perp} \subset \mathcal{R}(\Sigma_{22})$ es equivalente a $\mathcal{R}(\Sigma_{22})^{\perp} \subset \mathcal{R}(\Sigma_{21})^{\perp}$. Asi, el lema estará demostrado si logramos ver que

$$\alpha'\Sigma_{22} = \mathbf{0} \Rightarrow \alpha'\Sigma_{21} = \mathbf{0}, \tag{3.12}$$

donde α es un vector columna tal que el producto $\alpha'\Sigma_{22}$ está definido. Para verificar (3.12) observe que

$$\alpha'\Sigma_{22} = \mathbf{0} \Rightarrow \alpha'\Sigma_{22}\alpha = 0$$

$$\Rightarrow [\mathbf{0}', \alpha'] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{0} \\ \alpha \end{bmatrix} = 0$$

$$\Rightarrow [\mathbf{0}', \alpha'] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \mathbf{0} \quad \text{(por la parte (a))}$$

$$\Rightarrow [\alpha'\Sigma_{12}, \alpha'\Sigma_{22}] = (\mathbf{0}, \mathbf{0})$$

$$\Rightarrow \alpha' \Sigma_{12} = 0.$$

Asi, hemos verificado que (3.12) ocurre y esto concluye la demostración del teorema. ■

Teorema 3.4.1 Sea **X** un vector aleatorio n-dimensional tal que $\mathbf{X} \sim \mathcal{N}_n(\mu, \Sigma)$.

Particione **X** como $\mathbf{X} = (\chi_1', \chi_2')$, donde

$$\chi_1 = (X_1, \cdots, X_p), \text{ y } \chi_2 = (X_{p+1}, \cdots, X_n).$$

Particione Σ como

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

donde Σ_{11} es de orden $p \times p$. Entonces, la distribución condicional de χ_1 dado χ_2 es normal con media

$$\mu_1 + \Sigma_{12}\Sigma_{22}^{-} (\chi_2 - \mu_2),$$

y matriz de covarianza

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-} \Sigma_{21},$$

donde μ_1 y μ_2 es la media de χ_1 y χ_2, respectivamente.

Demostración. Usando el lema anterior , vemos que $\Sigma_{21} = \Sigma_{22}\mathbf{Q}$ para alguna matriz **Q**, o equivalentemente, $\Sigma_{12} = \mathbf{K}\Sigma_{22}$ donde $\mathbf{K} = \mathbf{Q}'$. Entonces,

$$\begin{aligned} \Sigma_{12}\Sigma_{22}^{-} \Sigma_{22} &= \mathbf{K}\Sigma_{22}\Sigma_{22}^{-} \Sigma_{22} \\ &= \mathbf{K}\Sigma_{22} \\ &= \Sigma_{12}. \end{aligned}$$

Ahora observe que

$$\mathrm{Cov}[\chi_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-} (\chi_2 - \mu_2),\ \chi_2 - \mu_2]$$

$$= \mathrm{Cov}(\chi_1 - \mu_1,\ \chi_2 - \mu_2) - \Sigma_{12}\Sigma_{22}^{-}\ \mathrm{Cov}(\chi_2 - \mu_2,\ \chi_2 - \mu_2)$$

$$= \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-}\ \Sigma_{22}$$

$$= \Sigma_{12} - \Sigma_{12}$$

$$= \mathbf{0}, \qquad (3.13)$$

Entonces, $\chi_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-} (\chi_2 - \mu_2)$ y $\chi_2 - \mu_2$ son independientes. Además la matriz de covarianza de

$$\chi_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-} (\chi_2 - \mu_2)$$

es

$$\mathrm{Cov}\ [\chi_1 - \Sigma_{12}\Sigma_{22}^{-}\ \chi_2,\ \chi_1 - \Sigma_{12}\Sigma_{22}^{-}\ \chi_2]$$

$$= \mathrm{Cov}\ (\chi_1,\ \chi_1) - \mathrm{Cov}(\chi_1,\ \chi_2)\Sigma_{22}^{-}\ \Sigma_{21}$$

$$- \Sigma_{12}\Sigma_{22}^{-}\ \mathrm{Cov}(\chi_2,\ \chi_1) + \Sigma_{12}\Sigma_{22}^{-}\ \mathrm{Cov}(\chi_2,\ \chi_2)\Sigma_{22}^{-}\ \Sigma_{21}$$

$$= \Sigma_{11} - 2\Sigma_{11}\Sigma_{22}^{-}\ \Sigma_{21} + \Sigma_{12}\Sigma_{22}^{-}\ \Sigma_{22}\Sigma_{22}^{-}\ \Sigma_{21}$$

$$= \Sigma_{11} - 2\Sigma_{12}\Sigma_{22}^{-}\ \Sigma_{21} + \Sigma_{12}\Sigma_{22}^{-}\ \Sigma_{21}$$

$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\ \Sigma_{21}. \qquad (3.14)$$

Puesto que $E[\chi_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-} (\chi_2 - \mu_2)] = \mathbf{0}$, de (3.14) se sigue que

$$\chi_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-} (\chi_2 - \mu_2) \sim \mathcal{N}(0,\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}). \qquad (3.15)$$

Debido a la independencia de

$$\chi_1 - \mu_1 - \Sigma_{21}\Sigma_{22}^{-}(\chi_2 - \mu_2)\ \text{y}\ \chi_2 - \mu_2$$

obtenida en (3.13), (3.15) implica que la distribución condicional de $\chi_1 - \mu_1 - \Sigma_{21}\Sigma_{22}^{-}(\chi_2 - \mu_2)$ dado χ_2 es $\mathcal{N}_p(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21})$. Luego la distribución de χ_1 dado χ_2 es $\mathcal{N}_p(\mu_1 + \Sigma_{12}\Sigma_{22}^{-}(\chi_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21})$. ■

CAPITULO 4

Prueba de una Hipótesis Lineal General

En este capítulo se presenta la prueba de razón de verosimilitud para una hipótesis lineal en el modelo $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ descrito en el capítulo anterior. El principal objetivo es resaltar el uso del concepto de inversa generalizada en la construcción de dicha prueba. Algunos detalles se omiten y pueden encontrarse, por ejemplo, en Graybill (1976) y Searle (1971).

4.1. Estadístico de Prueba para la Hipótesis Lineal $\mathbf{H}\beta=\mathbf{h}$

Considere el modelo $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$, donde $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, $\beta \in \mathbb{E}_p$, y $\mathbf{X}$ es una matriz conocida de orden $n \times p$. Considere la hipótesis $\mathbf{H}\beta = \mathbf{h}$, donde la matriz $\mathbf{H}$ y el vector $\mathbf{h}$ satisfacen las siguientes condiciones.

(*i*) $\mathbf{H}$ es una matriz de orden $q \times p$ de rango q, y $\mathbf{h} \in \mathbb{E}_q$ y tanto $\mathbf{H}$ como $\mathbf{h}$ son conocidos.
(*ii*) $\mathbf{H}\beta = \mathbf{h}$ es un sistema consistente.

Bajo estas condiciones, una vez observado el valor $\mathbf{Y} = \mathbf{y}$ del vector de observaciones, deseamos tomar una decisión sobre aceptar o rechazar la hipótesis $\mathbf{H}\beta = \mathbf{h}$. En forma intuitiva, se rechazará $\mathbf{H}\beta = \mathbf{h}$ si $\widehat{\mathbf{H}\beta}$ y $\mathbf{h}$ son "significativamente diferentes", y nuestro trabajo consiste en establecer en forma precisa el significado de esta expresión.

En seguida describimos la prueba de razón de verosimilitud para la hipótesis $\mathbf{H}\beta = \mathbf{h}$. Sean ω y Ω los subconjuntos de $E_p \times (0, \infty)$ dados por

$$\omega = \{(\beta, \sigma^2) \mid \mathbf{H}\beta = \mathbf{h}, \sigma^2 > 0\};$$
$$\Omega = \{(\beta, \sigma^2) \mid \beta \in \mathbb{E}_n, \sigma^2 > 0\}.$$

Ahora, para cada $\mathbf{y} \in \mathbb{E}_n$, defina

$$\hat{L}_\omega(\mathbf{y}) = \max_{\theta \in \omega} L(\mathbf{y}, \theta),$$

y

$$\hat{L}_\Omega(\mathbf{y}) = \max_{\theta \in \Omega} L(\mathbf{y}, \theta),$$

donde $L(\mathbf{y}, \theta) = (2\pi)^{-n/2}\exp[-\|\mathbf{y} - \mathbf{X}\beta\|^2/(2\sigma^2)]$, $\mathbf{y} \in \mathbb{E}_n$. Defina

$$\lambda(\mathbf{y}) := \hat{L}_\omega(\mathbf{y})/\hat{L}_\Omega(\mathbf{y}), \quad \mathbf{y} \in \mathbb{E}_n;$$

en general, se tiene que $\lambda(\cdot) \leq 1$. Sea $\alpha \in (0, 1)$ el nivel de significancia deseado. Entonces, la prueba de razón de verosimilitud para $\mathscr{H}_o: \mathbf{H}\beta = \mathbf{h}$ al nivel α es la siguiente:

$$\text{Rechace } \mathbf{H}\beta = \mathbf{h} \text{ si y sólo si } \lambda(\mathbf{y}) \leq C_\alpha, \tag{4.1}$$

donde C_α se elige de modo que, cuando $\mathcal{H}_o$ es cierta, la probabilidad máxima de rechazar $\mathcal{H}_o$ sea α.

El problema que tenemos ahora es dar en forma más explícita, una descripción de esta prueba. Con este fin, observe que

$$\hat{L}_\Omega(\mathbf{y}) = (2\pi\hat{\sigma}^2_\Omega)^{-n/2}\exp(-\|\mathbf{y} - \mathbf{X}\hat{\beta}_\Omega\|^2/(2\hat{\sigma}^2_\Omega),$$

donde

$$\hat{\beta}_\Omega = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^-\mathbf{y},$$

y

$$\hat{\sigma}^2_\Omega = \|\mathbf{y} - \mathbf{X}\hat{\beta}_\Omega\|^2/n = \mathbf{y}'(\mathbf{I} - \mathbf{X}\mathbf{X}^-)\mathbf{y}/n.$$

son los estimadores de verosimilitud máxima. Entonces, se sigue que

$$\hat{L}_\Omega(\mathbf{y}) = (2\pi\hat{\sigma}^2_\Omega)^{-n/2}\exp(-n/2).$$

Ahora necesitamos encontrar una expresión similar para $\hat{L}_\omega(\mathbf{y})$, y esto puede hacerse usando la técnica del modelo reducido. Dadas las q filas independientes de $\mathbf{H}$, encontramos p − q filas independientes entre si y ortogonales a las filas de $\mathbf{H}$, y con esas filas formamos la matriz $\mathbf{G}$ de orden $(p - q) \times q$. En este caso es claro que

$$\mathbf{H}\mathbf{G}' = \mathbf{0}.$$

Por las propiedades geométricas de la inversa generalizada descritas en el Capítulo 3, tenemos que, para cada $\beta \in \mathbb{E}_p$, $\mathbf{H}^-\mathbf{H}\beta$ es una proyección ortogonal sobre $\mathcal{L}\{\mathbf{H}'\}$ y $\mathbf{G}^-\mathbf{G}\beta$ es una proyección ortogonal sobre $\mathscr{L}\{G'\} = \mathscr{L}\{H'\}^{\perp}$. Luego, $G^{-}G = I - H^{-}H$, es decir,

$$H^{-}H + G^{-}G = I.$$

Entonces para un vector arbitrario $\beta \in E_p$ se tiene que

$$\begin{aligned}\beta &= (H^{-}H + G^{-}G)\beta \\ &= H^{-}H\beta + G^{-}G\beta,\end{aligned}$$

y en consecuencia, el modelo $\mathbf{y} = X\beta + \varepsilon$ puede presentarse como $\mathbf{y} = XH^{-}H\beta + XG^{-}G\beta + \varepsilon$, esto es

$$\mathbf{y} - XH^{-}H\beta = XG^{-}G\beta + \varepsilon. \tag{4.2}$$

Ahora suponga que la hipótesis $H\beta = \mathbf{h}$ es válida. Sustituyendo $H\beta = \mathbf{h}$ en (4.2), se obtiene

$$\mathbf{y} - XH^{-}\mathbf{h} = XG^{-}(G\beta) + \varepsilon. \tag{4.3}$$

Defina ahora $\gamma \in \mathbb{E}_{p-q}$, la matriz B de orden $n \times (p - q)$ y el vector aleatorio Z mediante

$$\begin{aligned}\gamma &:= G\beta; \\ B &:= XG^{-}; \\ Z &:= Y - XH^{-}\mathbf{h}.\end{aligned} \tag{4.4}$$

Entonces, (4.4) es equivalente a

$$Z = B\gamma + \varepsilon \tag{4.5}$$

y de esta manera hemos construído un nuevo modelo lineal que es equivalente al original cuando $H\beta = \mathbf{h}$ es válida. El modelo original $\mathbf{y} = X\beta + \varepsilon$ se denomina el modelo completo; su espacio de parámetros es $\Omega = \{(\beta, \sigma^2) \mid \beta \in \mathbb{E}_p,$

$\sigma^2 > 0\}$. El modelo dado en (4.5) se conoce como el modelo reducido asociado a la hipótesis $\mathbf{H}\beta = \mathbf{h}$; note que el espacio de parámetros de este modelo es $\omega = \{(\gamma, \sigma^2) | \gamma \in \mathbb{E}_{p-q}, \sigma^2 > 0\}$.

Como ya mencionamos para el modelo completo se tiene que

$$\hat{\beta}_\Omega = \mathbf{X}^-\mathbf{y}, \text{ y}$$

$$\hat{\sigma}^2_\Omega = \mathbf{y}'(\mathbf{I} - \mathbf{X}\mathbf{X}^-)\mathbf{y}/n, \quad \text{lo cual implica}$$

$$\hat{L}_\Omega = (2\pi\hat{\sigma}^2_\Omega)^{-n/2}\exp(-\,n/2).$$

Similarmente para el modelo reducido se encuentra que

$$\hat{\gamma}_\omega = \mathbf{B}^-\mathbf{z},$$

$$\hat{\sigma}^2_\omega = \mathbf{z}'(\mathbf{I} - \mathbf{B}\mathbf{B}^-)\mathbf{z}/n, \text{ y entonces,}$$

$$\hat{L}_\omega = (2\hat{\pi}^2{}_\omega)^{-n/2}\exp(-\,n/2).$$

Como consecuencia obtenemos que

$$\lambda(\mathbf{y}) = (\sigma^2_\Omega/\sigma^2_\omega)^{n/2}, \ \mathbf{y} \in \mathbb{E}_n.$$

La prueba de razón de verosimilitud en (4.1) rechazará $\mathbf{H}\beta = \mathbf{h}$ si y sólo si $\sigma^2_\Omega/\sigma^2_\omega \le C_\alpha^{n/2}$, o equivalentemente, cuando

$$(\hat{\sigma}^2_\omega - \hat{\sigma}^2_\Omega)/\hat{\sigma}^2_\Omega \ge k =: C_\alpha^{n/2} - 1. \tag{4.6}$$

Enseguida encontraremos una expresión sencilla para el cociente que aparece a la izquierda de (4.6). primero observa que

$$(\hat{\sigma}^2_\omega-\hat{\sigma}^2_\Omega)/\hat{\sigma}^2_\Omega = [z'(I-BB^-)z \quad - \quad y'(I-XX^-)y]/[y'(I-XX^-)y]. \quad (4.7)$$

Por otro lado,

$$z'(I - BB^-)z = z'[(XX^- - BB^-) + (I - XX^-)]z$$

$$= z'(XX^- - BB^-)z + z'(I - XX^-)]z.$$

$$(4.8)$$

Además, observe que $(I - XX^-)z = (I - XX^-)(y - XH^-h) =$ $(I - XX^-)y = ((I - XX^-)XH^-h = (X - XX^-X)H^-h = 0.H^-h = 0.$ Esto es

$$(I - XX^-)z = (I - XX^-)y. \quad (4.9)$$

Además, tomando transpuesta en ambos lados de esta igualdad se obtiene

$$z'(I - XX^-) = y'(I - XX^-).$$

Usando esta igualdad y (4.9) concluímos que

$$z'(I - XX^-)z = y'(I - XX^-)y,$$

y entonces (4.8) es equivalente a

$$z'(I - BB^-)z = z'(XX^- - BB^-)z + y'(I - XX^-)y.$$

combinando esta igualdad con (4.7) se obtiene que

$$(\hat{\sigma}^2_\omega - \hat{\sigma}^2_\Omega)/\hat{\sigma}^2_\Omega = z'(XX^- - BB^-)z\ /y'(I - \quad XX^-)y. \quad (4.10)$$

Enseguida simplificaremos el numerador del lado derecho de esta expresión. De hecho, nuestro objetivo es demostrar que

$$z'(XX^- - BB^-)z = (H\hat{\beta} - h)'[H(X'X)^{-1}H']^{-1}(H\hat{\beta} - h).$$

Para esto observe que

$$\begin{aligned}\mathcal{R}[(\mathbf{HX}^-)^-(\mathbf{HX}^-)] &\subset \mathcal{R}[(\mathbf{HX}^-)^-] \\ &= \mathcal{N}(\mathbf{HX}^-)^{\perp} \\ &= \mathcal{R}(\mathbf{X}'^-\mathbf{H}') \subset \mathcal{R}(\mathbf{X}^{-\prime}) \\ &= \mathcal{N}(\mathbf{X}^-)^{\perp} \\ &= \mathcal{R}(\mathbf{X})\end{aligned}$$

y además, dim $\mathcal{R}[(\mathbf{HX}^-)^-(\mathbf{HX}^-)] = q$. Por otro lado,

$$\begin{aligned}\mathcal{R}[(\mathbf{XG}^-)(\mathbf{XG}^-)^-] &= \mathcal{R}(\mathbf{XG}^-) \\ &\subset \mathcal{R}(\mathbf{X}),\end{aligned}$$

donde dim $\mathcal{R}[(\mathbf{XG}^-)(\mathbf{XG}^-)^-] = p-q$. También se puede ver que $(\mathbf{HX}^-)^-(\mathbf{HX}^-) + (\mathbf{XG}^-)(\mathbf{XG}^-)^-$ es una proyección en $\mathcal{R}(\mathbf{X})$, con rango $q + p - q = p = r(\mathbf{X})$, y por tanto coincide con $\mathbf{XX}^-$. De aqui se concluye que

$$\mathbf{XX}^- = (\mathbf{HX}^-)^-(\mathbf{HX}^-) + (\mathbf{XG}^-)(\mathbf{XG}^-)^-$$

es decir,

$$\mathbf{XX}^- - \mathbf{BB}^- = (\mathbf{HX}^-)^-(\mathbf{HX}^-). \tag{4.11}$$

Aplicando las propiedades de g-inversa un cálculo directo muestra que

$$(\mathbf{HX}^-)^-(\mathbf{HX}^-) = (\mathbf{HX}^-)'[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}\mathbf{HX}^-, \tag{4.12}$$

y combinando (4.11) y (4.12) se tiene

$$\mathbf{XX}^- - \mathbf{BB}^- = (\mathbf{HX}^-)'[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}\mathbf{HX}^-.$$

Usando esta igualdad junto con (4.10) obtenemos

$$(\hat{\sigma}^2_\omega - \hat{\sigma}^2_\Omega)/\hat{\sigma}^2_\Omega = \mathbf{z}'(\mathbf{HX}^-)'[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}\mathbf{HX}^-\mathbf{z} \;/\; \mathbf{y}'(\mathbf{I} - \mathbf{XX}^-)\mathbf{y}$$

$$= \frac{(y - XH^{-}h)'(HX^{-})'[H(X'X)^{-1}H]^{-1}(HX^{-})(y - XH^{-}h)}{y'(I - XX^{-})y},$$

$$= \frac{(H\hat{\beta} - h)'[H(X'X)^{-1}H']^{-1}(H\hat{\beta} - h)}{y'(I - XX^{-})y}.$$

Asi, la prueba de razón de verosimilitud para $\mathscr{H}_o: H\beta - h$ se expresa como sigue: rechace $\mathscr{H}_o$ si

$$W = \frac{(H\hat{\beta} - h)'[H(X'X)^{-1}H']^{-1}(H\hat{\beta} - h)}{y'(I - XX^{-})y} \geq k,$$

donde $k \in \mathbb{R}$ se selecciona de modo que el nivel de significancia sea el valor dado de α. Para ver como se selecciona k, observe que, cuando $\mathscr{H}_o$ es cierta,

$$z'(XX^{-} - BB^{-})z \sim \chi^2_{q,}$$

y además

$$y'(I - XX^{-})y \sim \chi^2_{n-p}.$$

Por otro lado es fácil ver que $(XX^{-} - BB^{-})(I - XX^{-}) = 0$, lo cual implica que $z'[XX^{-} - BB^{-}]z$ y $y'(I - XX^{-})y$ son variables aleatorias independientes. Entonces,

$$W: = \frac{z'[XX^{-} - BB^{-}]z/q}{y'(I - XX^{-})y/(n - p)} \sim F_{q,n-p.}$$

Luego, la prueba de nivel α para $\mathscr{H}_o$: $H\beta = h$ puede describirse como sigue:

Rechace $\mathcal{H}_o$ si y sólo si $W > F_{q,n-p,\alpha,}$ donde $F_{q,n-p,\alpha,}$ es el percentil de orden α para la distribución $F_{q,n-p,}$.

LITERATURA CITADA

Apostol, T. M. (1974) Mathematical Analysis, Addison Wesley, Reading, Massachusetts, 352 p.

Arnold, S. F. (1981) The Theory of Linear Models and Multivariate Analysis, John Wiley and Sons, New York, 475 p.

Bugrov Ya. S. y S. M. Nikolsky (1980) Matemáticas Superiores, Elementos de Algebra Lineal y Geometría Analítica, Mir, Moscú, 168 p.

Den-Israel A., and T. N. E. Greville, (1974) Generalized Inverses: Theory and Applications, Wiley, New York, 391 p.

Florey, F.G. (1980) Fundamentos de Algebra Lineal y Aplicaciones, Prentice-Hall, México D. F., México, 366 p.

Fulks, W. (1973) Cálculo Avanzado, Limusa, México D. F. México, 551 p.

Granero, R. F. (1985) Algebra y Geometría Analítica, McGraw Hill, México D. F., México, 568 p.

Graybill, F. A. (1969) Introduction to Matrices with Aplications in Statistics, Wadsworth, Belmont CA, 371 p.

Graybill, F. A. (1976) Theory and Application of the Linear Model, Wadsworth, Belmont CA, 704 p.

Golovina, Li. (1980) Algebra Lineal y Algunas de sus Aplicaciones, Mir, Moscú, 277 p.

Hoffman, K. y R. Kunze, (1973) Algebra Lineal, Prentice-Hall Hispanoamericana, México D. F. México, 400 p.

Krutitskaya, N. Ch. y A. A. Shishkin, (1985), Algebra Lineal, Mir, Moscú, 159 p.

Maltsev, A. I. (1978) Fundamentos de Algebra Lineal, Mir, Moscú, 400 p.

Mardsen, J. E. y A. J. Tromba (1987) Cálculo Vectorial, Addison Wesley, Wilmington, Delaware, 454 p.

Pringle R. M. and A. A. Rayner (1971) Generalized Inverse Matrices with Aplications to Statistics, Hafner Publishing Company, New York, 127 p.

Rao, C. R. (1973) Linear Statistical, Inference its Applications, Wiley, New York, 625 p.

Rudin, W.(1964) Principles of Mathematical Analysis, McGraw Hill, New York, 251 p.

Saxena, H. C. and P. V. Surendran, (1973) Statistical Inference, S. Chand & Co. Ram Negar, New Delhi, India, 403 p.

Searle, S. R. (1971) Linear Models, Wiley, New York, 531 p.