EL TEOREMA DE COCHRAN Y ALGUNOS DE SUS EQUIVALENTES

MIGUEL PINEDO VARGAS

T E S I S

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.

MAYO DE 1998

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

SUBDIRECCIÓN DE POSTGRADO

EL TEOREMA DE COCHRAN Y ALGUNOS DE SUS EQUIVALENTES

TESIS

POR

MIGUEL PINEDO VARGAS

Elaborada bajo la supervisión del Comité Particular de Asesoría y aprobada como requisito parcial para optar al grado de:

MAESTRO EN CIENCIAS
EN ESTADÍSTICA EXPERIMENTAL

COMITE PARTICULAR

Asesor Principal: ______________________
M.C. EMILIO PADRÓN CORRAL

Asesor: ______________________
M.C. FÉLIX DE JESÚS SÁNCHEZ PÉREZ

Asesor: ______________________
M.C. ROBERTO CORONADO NIÑO

Asesor: ______________________
DR. HERIBERTO DÍAZ SOLÍS

____
DR. JESÚS M. FUENTES RODRÍGUEZ
SUBDIRECTOR DE POSTGRADO

Buenavista, Saltillo, Coahuila Mayo de 1998

AGRADECIMIENTOS

- A la Universidad Autónoma Agraria "Antonio Narro" por su oportunidad que me dio para realizar este postgrado, y a la Universidad Autónoma de Zacatecas, por su apoyo económico.

- Al M.C. Emilio Padrón Corral, por sus consejos, su paciencia y su estimable ayuda en la elaboración de esta tesis.

- A los miembros del Comité Particular, M.C. Félix de Jesús Sánchez Pérez, M.C. Roberto Coronado Niño y Dr. Heriberto Díaz Solis, por su preciosa aportación en cada una de las etapas en que se desarrolló el presente trabajo.

- A todos mis maestros, por su valiosa orientación en mis estudios.

- A todos mis compañeros de estudio, en especial a Salomón González y Dino U. González, por su apoyo en los momentos dificiles.

- Al Sr. Enrique de León Aguilar, por su paciente labor de capturar esta tesis.

DEDICATORIA

A mis padres: Roque –R.I.P.– y Rosa María. Con amor, admiración y respecto, por sus sabios consejos, por su comprensión y apoyo en los momentos difíciles.

A mis hermanos: Elena, Chagua, Nicolás, Marín, Emerita e Isidro. Por el cariño que nos une y el apoyo que siempre me han brindado.

A mis sobrinos: Luis, Pepe, Roberto, Mary, Mónica, Toña, Cuquito, Miguel, etc. a quienes deseo que logren sus anhelos.

COMPENDIO

El Teorema de Cochran y Algunos de sus Equivalentes.

POR

MIGUEL PINEDO VARGAS

MAESTRÍA

ESTADÍSTICA EXPERIMENTAL

UNIVERSIDAD AUTÓNOMA AGRARIA "ANTONIO NARRO"

BUENAVISTA, SALTILLO, COAHUILA, MAYO DE 1998

M.C. Emilio Padrón Corral –Asesor–

Palabras claves: forma cuadrática, distribución normal multivariada, distribución normal singular, χ^2 no central, modelo lineal, distribución de una forma cuadrática, independencia de formas cuadráticas.

El aspecto económico es uno de los más esenciales del bienestar mundial, nacional, estatal y local; se mide por el progreso de las actividades primarias; la más

importante de éstas es la agricultura; su avance se guía por la investigación agropecuaria; para optimizarla se diseñan experimentos, los cuales se interpretan mediante el análisis de la varianza; cuando en éste, la suma de cuadrados total se parte en sus componentes, la distribución conjunta de éstos, como sus respectivos grados de libertad, a menudo se deducen del Teorema de Cochran.

En el presente trabajo se demuestra el Teorema de Cochran desde dos puntos de vista. En un caso se deriva como un corolario, de una proposición más general; en un segundo caso, se infiere por inducción. Se concluye, que ambos métodos son equivalentes. Además, se prueba la equivalencia del Teorema de Cochran con otras proposiciones, y por último, se ilustra su aplicación a la estadística experimental.

ABSTRACT

Cocharn's Theorem and Some of its Equivalents.

BY

MIGUEL PINEDO VARGAS

MASTER OF SCIENCES

EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA "ANTONIO NARRO"

BUENAVISTA, SALTILLO, COAHUILA, MAY 1998.

M.S. Emilio Padrón-Corral –Adviser–

Key Words: quadratic form, multivariate normal, singular normal distribution, noncentral χ^2, linear model, distribution of a quadratic form, independence of quadratic forms.

The economic aspect is one of the more chief of the world, national, state, and local well-being; it is measured by the progress of the primary activities; the agriculture is the more important from them; its looting is led by the farming research; in order to

accomplish it of a optimum form the experiments are designed, they are interpreted through the analysis of variance; when in this one, the sums of squares (SS) total is partitioned into a sum of other SS's, with their respective degrees of freedom, often is deduced from Cochran's Theorem.

In the present paper Cochran's Theorem is shown from two viewpoints. At a case it is derived as a corollary, of a theorem more general; at a second case it is infered by induction. It is concluded that both methods are equivalent. In addition, it is shown the equivalence from Cochran's Theorem with other propositions; and lastly, it is illustrated its application to the experimental statistics.

ÍNDICE DE CONTENIDO

CAPÍTULO I

INTRODUCCIÓN

Uno de los aspectos fundamentales del bienestar social mundial, es el económico; el cual se cuantifica de manera crucial por el desarrollo de las actividades primarias. La más importante de éstas es la agricultura, cuyo avance de ella se basa en la investigación agropecuaria; para eficientizar u optimizar ésta, se diseñan experimentos. Según Mann (1949) fue R.A. Fisher el primero en fomentar la idea para planear experimentos sistemáticos con fines a su análisis estadístico. Tales diseños son interpretados por medio del análisis de la varianza (ANOVA); éste, como comúnmente se entiende y se practica hoy, ha sido desarrollado en forma substancial por Fisher (1918, 1925, 1935), quien introdujo las dicciones varianza y ANOVA en la estadística. Dos trabajos de Irwin (193 , 1934) reverificaron la base matemática de este análisis.

El ANOVA se define como una colección de métodos estadísticos usados para estudiar observaciones que satisfacen las dos hipótesis siguientes:

$$y_i = w_{1i}\, a_1 + w_{2i}\, a_2 + \ldots + w_{pi}\, a_p + e_i,$$

$$e_i \text{ son de forma independiente } N(0,\ \sigma^2)$$

donde $w_{1i}, w_{2i}, \ldots, w_{pi}$ son 0 ó 1 y las $a_1, a_2, \ldots, a_p$ son cantidades desconocidas, y N(0, σ^2) indica una distribución normal con media 0 y varianza σ^2 (Guenther, 1964; Scheffe, 1959).

El término ANOVA se refiere a dividir o analizar, la definición clásica de esta voz es división en partes es el antónimo de sintetizar, la varianza total en dos varianzas separadas, que representan la variación correspondiente a, dentro y entre grupos, con su división respectiva de los grados de libertad (Lindman, 1992). Sin embargo, lo que se divide no es la varianza total, sino la variación, o suma de cuadrados (SC) como también se le conoce, total; de aquí que, sería más correcto se diga análisis de la variación para referirse al ANOVA. No es lo mismo variación que varianza, pues esta última es la primera dividida por sus grados de libertad. Según Méndez (1976) los grados de libertad, o grados de independencia como también se les nombra, son el rango de la matriz **A** en la forma cuadrática **y'Ay.**

Cuando en el ANOVA la SC total se parte en sus componentes, la distribución conjunta de éstas con sus grados de libertad respectivos, a menudo se deduce del Teorema de Cochran (TC); el cual de acuerdo con Scheffé (1959) es de suficiente importancia en la literatura del ANOVA para que se justifique su inclusión en este trabajo. La demostración original de esta proposición se debe a Cochran (1934), de ahí su denominación, en el caso central; el caso no central se debe a Madow (1940); estuvo implícito en el primer uso del ANOVA de Fisher (1922, 1925), pero no recibió atención formal hasta 1934; según John (1971) es en lo esencial una consecuencia algebraica que se expresa en un contexto estadístico; y de acuerdo con Brownlee (1984), Chakrabarti (1962) y Kempthorne (1973) es el resultado teórico más poderoso para el tratamiento de los modelos I, II y III del ANOVA.

El objetivo de este trabajo consiste de forma básica en que se desarrollen algunas pruebas del TC, que se hallan dispersas en la literatura teórica estadística, desde diferentes perspectivas; quizás, su originalidad consiste en lo detallado de las verificaciones, lo que lo hace más accesible a un mayor público; también, se demuestra la equivalencia de este teorema con otras proposiciones, y se ilustra su aplicación a la estadística experimental.

Para cumplir el propósito del trabajo el material se organizó como se expresa enseguida:

El capítulo II se refiere a conceptos y proposiciones de álgebra matricial, dadas en lo fundamental sin demostraciones, desde las cuestiones elementales hasta los elementos esenciales para la comprobación del TC; en el capítulo III se tratan las distribuciones teóricas de mayor importancia como la normal, χ^2, F, t y F doble centrales y no centrales; en el capítulo IV se estudia el modelo lineal y la distribución de formas cuadráticas del tipo $\mathbf{x}'\mathbf{A}\mathbf{x}$ cuando $\mathbf{x}$ es $N(\mu, \mathbf{V})$, tanto para V singular como no singular, se dan varias pruebas del TC y se clarifica que éstas son en lo esencial las mismas; además, en este capítulo se ejemplifica el TC en los diseños experimentales.

CAPÍTULO II

MATRICES Y FORMAS CUADRÁTICAS

Como se mencionó en la Introducción, el TC es una conclusión de álgebra en un contexto estadístico; por lo que es natural, se desarrollen aquí los principales resultados de esta ciencia.

Conceptos Fundamentales

Definición 1: Matriz. Una matriz **A** de orden m×n es un arreglo, o esquema, rectangular de números o elementos de un campo dispuestos en m filas, hileras o renglones, y n columnas como

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots a_{1n} \\ a_{21} & a_{22} & \dots a_{2n} \\ \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots a_{mn} \end{bmatrix} \tag{1.1}$$

donde las cantidades a_{ij}, i = 1,2...m, j = 1, 2, ... , n, son números reales (Ayres, 1983; Graybill, 1983; Hadley, 1969; Hoffman y Kunze, 1992; Kleiman y de Kleiman, 1987; Kurosch, 1975; Lipschutz, 1971; Máltsev, 1972; Martínez y Castillo, 1987; Nering, 1977; Searle, 1966).

En las siete proposiciones siguientes se dan las propiedades básicas de las operaciones matriciales elementales.

Teorema 1. Sean **A** = (a_{ij}) y **B** = (b_{ij}) matrices cualesquiera del mismo orden mxn; entonces

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}. \tag{1.2}$$

Demostración. Sean **D**= (d_{ij}) = **A** + **B** y **E** = (e_{ij}) = **B** + **A** ; así, por definición de adición se tiene $d_{ij} = a_{ij} + b_{ij}$ y $e_{ij} = b_{ij} + a_{ij}$; como a_{ij} y b_{ij} son números reales, su adición es conmutativa; de aquí, $d_{ij} = a_{ij} + b_{ij} = b_{ij} + a_{ij} = e_{ij}$, expresión cierta para cualesquier i = 1,2, ... , *m*, *j* = 1,2, ... , *n; ergo* **D** = **E**, o sea, **A+B=B+A.**

Teorema 2. Si **A** = (a_{ij}), **B** = (b_{ij}) y **C** = (c_{ij}) son todas matrices de orden mxn; siendo así.

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A}+(\mathbf{B}+\mathbf{C}) \tag{1.3}$$

Teorema 3. Si **A** = (a_{ij}) y **B** = (b_{ij}) son matrices del mismo orden pxn, y **C** una matriz de orden mxp; en tal caso

$$\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB} \tag{1.4}$$

Teorema 4. Sean **A** = (a_{ij}) y **B** = (b_{ij}) matrices del mismo orden mxp, **C** matriz de orden pxn; entonces

$$(\mathbf{A} +\mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \tag{1.5}$$

Teorema 5. Si $\mathbf{A} = (a_{ij})$ y $\mathbf{B} = (b_{ij})$ son matrices del mismo orden mxn, y k una constante, se tiene

$$k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}. \tag{1.6}$$

Teorema 6. Sean las matrices $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$ y $\mathbf{C} = (c_{ij})$ de orden mxp, pxn y nxq, respectivamente; siendo así

$$\mathbf{(AB)C = A(BC)}\,. \tag{1.7}$$

Teorema 7. En general,

$$\mathbf{AB} \neq \mathbf{BA}, \tag{1.8}$$

aunque los productos **AB** y **BA** estén definidos.

Se llama combinación lineal de x_1 y x_2 a la adición de estas dos variables, cada una multiplicada por una constante diferente de cero, k_1 y k_2; esto es, $k_1x_1 + k_2x_2$; el calificativo lineal es porque k_1x_1 graficado contra x_2 es una línea recta (Searle, 1966). De los teoremas 1 a 7 se deduce, que una combinación lineal $k_1\mathbf{A}_1 + k_2\mathbf{A}_2 + \ldots + k_p\mathbf{A}_p$ estará definida únicamente cuando todas las $\mathbf{A}_i$ sean del mismo orden, y que los términos puedan ser cambiados de orden en forma arbitraria.

Definición 2: Transposición. La matriz transpuesta de $\mathbf{A} = (a_{ij})$ de orden mxn, es $\mathbf{A}' = (a'_{ij})$ de orden nxm, tal que

$$a'_{ij} = a_{ji}$$

Si se suponen bien definidas las operaciones indicadas, las propiedades básicas de la matriz traspuesta, se dan en los tres teoremas siguientes.

Teorema 8. $$(\mathbf{A}')' = \mathbf{A} \tag{1.9}$$

Demostración. Sean $(\mathbf{A}')' = (a''_{ij})$, $\mathbf{A}' = (a'_{ij})$ y $\mathbf{A} = (a_{ij})$, donde esta última es cualquier matriz real de orden mxn; al aplicar la definición de transposición: $a''_{ij} = a'_{ji}$ y $a'_{ji} = a_{ij}$; o sea, $a''_{ij} = a_{ij}$ con independencia de los valores de i y j; pero si $a''_{ij} = a_{ij}$ para cualesquier i y j, lo que implica que $(\mathbf{A}')' = \mathbf{A}$.

Teorema 9. $$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}' \tag{1.10}$$

Teorema 10. $$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \tag{1.11}$$

Definición 3: Matriz Simétrica. Si en una matriz $\mathbf{A} = (a_{ij})$ de orden n, resulta que

$$\mathbf{A} = \mathbf{A}';$$

o sea, $a_{ij} = a_{ji}$ para toda i y j; en tal caso se dice que $\mathbf{A}$ es una matriz simétrica.

Teorema 11. Sea $\mathbf{A} = (a_{ij})$ cualquier matriz de orden mxn; siendo así

$$\mathbf{AA}' \text{ y } \mathbf{A}'\mathbf{A}$$

son matrices simétricas.

Demostración. Sea **AA'** = **B** = (b_{ij}); de acuerdo con la definición de producto y sus propiedades se tiene

$$b_{ij} = \sum_{k=1}^{n} a_{ik}\, a'_{kj} = \sum_{k=1}^{n} a_{ik}\, a_{jk};$$

y también

$$\sum_{k=1}^{n} a_{jk}\, a'_{ki} = \sum_{k=1}^{n} a_{jk}\, a_{ik} = \sum_{k=1}^{n} a_{ik}\, a_{jk} = b_{ij},$$

lo que significa que **B** = **AA'** es simétrica. En forma análoga, **A'A** es también simétrica. Nótese que **AA'** y **A'A** son matrices cuadradas de orden m y n, respectivamente.

Definición 4: Matriz Diagonal. Una matriz simétrica que tenga nulos todos sus elementos no diagonales se llama matriz diagonal.

Teorema 12. Si **A** de orden m×n es una matriz cualquiera y $\mathbf{D}_1$ de orden m y $\mathbf{D}_2$ de orden n son matrices diagonales; el producto $\mathbf{D}_1$ **A** se obtiene al multiplicar las filas de **A** por los elementos respectivos de $\mathbf{D}_1$ y; de la misma forma, el producto $\mathbf{AD}_2$ se obtiene al multiplicar las columnas de **A** por los respectivos elementos diagonales de $\mathbf{D}_2$.

Definición 5: Matriz Unidad. Ésta es la matriz n×n definida por

$$\mathbf{I}_{ij} = \delta_{ij} = \begin{cases} 1, & \text{si} \quad i = j; \\ 0, & \text{si} \quad i \neq j; \end{cases} \tag{1.12}$$

donde δ es la delta de Kronecker (Hoffman y Kunze, 1992).

Teorema 13. Si **A** es de orden mxn, se tiene

$$\mathbf{IA} = \mathbf{AI} = \mathbf{I}, \tag{1.13}$$

donde **I** designa la matriz identidad de orden m, en el primer producto, n en el segundo.

Matriz Inversa

Sea **A** = (a_{ij}) una matriz cuadrada nxn; se desea asociar con ésta un escalar que mida, en cierto sentido, el "tamaño" de **A** y nos indique si **A** es singular o no.

Definición 6: Determinante. El determinante de la matriz **A** = (a_{ij}) se define como el escalar det **A** = $|a_{ij}|$ = A, calculado según la regla

$$\det \mathbf{A} = \sum \pm a_{1r_1} a_{2r_2} \ldots a_{nr_n}, \tag{1.14}$$

donde los segundos subíndices forman todas las n! permutaciones posibles de los números 1,2,...,n; en tanto que el signo de cada término de la suma es + o −, según que la permutación correspondiente sea par o impar; el número n se llama orden del determinante (Kurosch, 1975; Máltsev, 1972; Nering, 1977). Cada término de la suma es un producto de n elementos, cada uno tomado de un renglón y una columna diferentes de **A**, con su signo.

En general, un determinante de orden n será la suma de n! productos; conforme n crece, la cantidad de cálculos crece de forma astronómica; por lo que es necesario desarrollar formas más eficientes de manejar los determinantes.

Teorema 14. $$\det \mathbf{A}' = \det \mathbf{A}. \tag{1.15}$$

Teorema 15. Si **A''** es la matriz que se obtiene de **A**, al multiplicar un renglón (o columna) por un escalar c; en tal caso

$$\det \mathbf{A}'' = c \det \mathbf{A}.$$

Demostración. Cada término del desarrollo de det **A** contiene de modo preciso un elemento de cada fila de **A**; así que, al multiplicar una fila de **A** por c se introduce el factor c en cada término de det **A**; de donde, det **A''** = c det **A**.

Teorema 16. Si **A''** es la matriz que se obtiene de **A**, al intercambiar dos filas (o columnas) cualesquiera de **A**; siendo así

$$\det \mathbf{A}'' = -\det \mathbf{A}. \tag{1.16}$$

Teorema 17. Si **A** tiene dos filas iguales, det **A** = 0.

Teorema 18. Si **A''** es la matriz que se obtiene de **A**, al adicionar un múltiplo de una hilera (o columna) a otra; entonces,

$$\det \mathbf{A}'' = \det \mathbf{A}.$$

Definición 7: Matrices Elementales. Se definen tres tipos de operaciones elementales sobre las filas de una matriz **A**, a saber:

Tipo I: Multiplicación de una fila de **A** por un escalar diferente de cero.

Tipo II. Suma de un múltiplo de una fila, a otra fila.

Tipo III. Intercambio de dos filas.

Las operaciones elementales sobre las columnas se definen en forma similar. Las matrices que se obtienen mediante estos tres tipos de operaciones se llaman matrices elementales.

Teorema 19. Si **E** es una matriz elemental y **A** es cualquier matriz; en tal caso

$$\det \mathbf{EA} = \det \mathbf{E} \cdot \det \mathbf{A} = \det \mathbf{AE}. \tag{1.17}$$

Definición 8: Matriz no Singular. La matriz cuadrada **A** se dice singular si $|\mathbf{A}| = 0$, y regular, o no singular, si

$$|\mathbf{A}| \neq 0.$$

Teorema 20. La condición necesaria y suficiente sí y sólo sí para que **A** sea singular es $\det \mathbf{A} = 0$.

Teorema 21. Si **A** y **B** son cualesquiera dos matrices n×n, siendo así

$$\det \mathbf{AB} = \det \mathbf{A} \cdot \det \mathbf{B} = \det \mathbf{BA}.$$

Demostración. Si **A** y **B** son no singulares, el teorema se sigue a través de aplicaciones repetidas del Teorema 19. Si cualquiera de las matrices es singular, entonces **AB** y **BA** también son no regulares y todos los términos son cero.

Para una pareja dada i, j considérese en el desarrollo de det **A** aquellos términos que contienen a a_{ij} como factor; entonces,

$$\det \mathbf{A} = a_{ij}\, A_{ij} + (\text{términos que no contienen a } a_{ij} \text{ como factor}).$$

El escalar A_{ij} se llama cofactor, también se nombra adjunto o complemento algebraico, de a_{ij}.

Cada término en el desarrollo de det **A** contiene de modo preciso un factor de cada fila y cada columna de **A**; de donde, para cualquier fila dada de **A**, cada término de det **A** contiene de modo exacto un factor de esta fila; de aquí, para cualquier i dado,

$$\det \mathbf{A} = \sum_{j} a_{ij} A_{ij} \tag{1.18}$$

De igual modo, para cualquier k dado,

$$\det \mathbf{A} = \sum_{j} a_{jk} A_{jk} \tag{1.19}$$

Estos desarrollos de un determinante en términos de los cofactores de una fila o columna, reducen el cálculo de un determinante de n-ésimo orden, al de calcular n determinantes de orden n-1. Sin embargo, es aún una tarea formidable; por lo que al aplicar estos desarrollos en conjunción con el Teorema 18, se observa que esta técnica es más simple.

Si se multiplican los elementos en la fila i por los cofactores de los elementos en la fila $k \neq i$, se obtiene el mismo resultado que se tendría si los elementos en la fila k fueran iguales a los de la fila i; de aquí,

$$\sum_{j=1}^{n} a_{ij}\, A_{kj} = \begin{cases} A & \text{para } i = k \\ 0 & \text{para } i \neq k \end{cases}$$

$$\sum_{j=1}^{n} a_{ji}\, A_{jk} = \begin{cases} A & \text{para } i = k \\ 0 & \text{para } i \neq k \end{cases} \qquad (1.20)$$

y además,

$$A = a_{11}\; A_{11} \;-\; \sum_{i,k=2}^{n} a_{i1}\, a_{1k}\, A_{11.ik} \qquad (1.21)$$

en donde $A_{11.ik}$ es el adjunto de a_{ik} en A_{11}.

Definición 9: Matriz Adjunta Si $\mathbf{A} = (a_{ij})$ es una matriz cuadrada cualquiera y A_{ij} es el cofactor de a_{ij}, la matriz que

$$(A_{ij})' = \text{adj } \mathbf{A},$$

se llama agregada de **A**.

Teorema 22. $$\mathbf{A} \cdot \text{adj } \mathbf{A} = (\text{adj } \mathbf{A})\mathbf{A} = (\det \mathbf{A}) \cdot \mathbf{I} \qquad (1.22)$$

Demostración.

$$\mathbf{A} \cdot \text{adj. } \mathbf{A} = (a_{ij}).(A_{ij})' = \left(\sum_{j} a_{ij} A_{ij} \right) = (\det\ \mathbf{A})\mathbf{I}\,; \qquad (1.23)$$

$$(\text{adj. } \mathbf{A})\mathbf{A} = (A_{ij})'\,(a_{ij}) = \left(\sum_{i} A_{ij}\, a_{ij} \right) = (\det\ \mathbf{A})\mathbf{I}\,. \qquad (1.23')$$

Como es fácil ver las expresiones (1.23) y (1.23') son iguales, lo que completa la demostración de esta proposición.

Los métodos más comunes para calcular determinantes de orden n son:

— El método de reducción a la forma triangular;

— El método de variación de los elementos de un determinante;

— El método de representación del determinante en forma de adición de determinantes;

— El método de separación de los factores líneales

— El método de relaciones recurrentes, como el de Ókunev.

Aunque éstos son muy complejos, en seguida se da una breve explicación y se ejemplariza el primero de ellos.

Este método consiste en transformar el determinante de tal manera que todos los elementos, a un lado de las diagonales, sean ceros; el caso de la diagonal secundaria, se reduce al caso de la diagonal principal, al invertir el orden de filas (o columnas); el determinante así obtenido es igual al producto de los elementos de la diagonal principal (Proskuriakov, 1986).

Ejemplo 1. Calcúlese el determinante de orde n:

$$d = \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & 1 & \dots & 1 \\ 1 & 1 & 0 & \dots & 1 \\ \multicolumn{5}{c}{\dots\dots\dots\dots\dots\dots} \\ 1 & 1 & 1 & \dots & 0 \end{vmatrix} .$$

Réstese la primera fila de todas las demás:

$$d = \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & -1 & 0 & \dots & 0 \\ 0 & 0 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 \end{vmatrix} = 1(-1)^{n-1} = (-1)^{n-1}$$

Definición 10: Inversión. Si **A** es una matriz cuadrada; y si existe una matriz cuadrada $\mathbf{A}^{-1}$ que satisface la relación

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}; \qquad (1.24)$$

en tal caso $\mathbf{A}^{-1}$ se llama la inversa de **A**.

Teorema 23. Si una matriz tiene inversa, ésta es única.

Teorema 24. Si **A** tiene inversa, entonces $\mathbf{A}^{-1}$ tiene inversa y

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

Teorema 25. Si **A** y **B** son matrices no singulares; entonces

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Demostración. Nótese que

$$\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B}^{-1}\,\mathbf{IB} = \mathbf{B}^{-1}\,\mathbf{B} = \mathbf{I},$$

y

$$\mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{AA}^{-1} = \mathbf{I},$$

lo que completa la prueba.

Teorema 26. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$,

siempre que **A** sea no singular.

Teorema 27. $$\left|\mathbf{A}^{-1}\right| = \frac{1}{|\mathbf{A}|}$$

Teorema 28. Cuando **A** es simétrica, la inversa $\mathbf{A}^{-1}$, también lo es.

Teorema 29. La inversa de una matriz diagonal **D,** cuyos elementos diagonales son $d_1, d_2, \ldots, d_n$, es otra matriz diagonal con elementos diagonales $d_1^{-1}, d_2^{-1}, \ldots, d_n^{-1}$.

En la actualidad se dispone de varios métodos para la inversión de matrices, algunos de éstos son algebraicos, directos o exactos, y los demás son numéricos, o de aproximaciones sucesivas. El más conocido es el método de la matriz adjunta, el cual se basa en la teoría de los determinantes. Una matriz es no singular si sólo si det $\mathbf{A} \neq 0$; en este caso, se puede ver a partir del Teorema 22, que

$$\mathbf{A}^{-1} = \frac{1}{det\,\mathbf{A}} \operatorname{adj} \mathbf{A}$$

Sin embargo, el cálculo de la inversa, se puede hacer sin recurrir a los determinantes, dos métodos de éstos son: el método de transformaciones elementales y el método de la matriz orlada.

Independencia Lineal de Vectores

Una matriz $1\times n$ contiene una sola fila de elementos y por lo común se llama vector fila; en forma análoga, una matriz $m\times 1$ consta de una sola columna de elementos y por lo general se llama vector columna; es usual denominar de forma simple escalar a una matriz 1×1. La mayor parte de los vectores que aparecen en las aplicaciones son del tipo vector columna. El transpuesto de un vector columna $\mathbf{x} = (x_1, x_2, \dots, x_n)$ es el vector fila $\mathbf{x}' = (x_1, x_2, \dots, x_n)$ y viceversa Sean $\mathbf{x} = (x_1, x_2, \dots, x_n)$ y $\mathbf{y} = (y_1, y_2, \dots, y_n)$ dos vectores columna, el producto **x'y** es un escalar $1\mathrm{x}1$; o sea,

$$\mathbf{x}'\mathbf{y} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n; \tag{1.25}$$

para $\mathbf{x} = \mathbf{y}$, se tiene

$$\mathbf{x}'\mathbf{x} = x_1^2 + x_2^2 + \dots + x_n^2, \tag{1.26}$$

que se denomina longitud del vector $\mathbf{x}$. Dos vectores, **x** y **y**, se dicen ortogonales si $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = 0$; esto implica que **x** y **y** tienen el mismo número de elementos. Un conjunto de vectores de longitud unitaria que son en forma mutua ortogonales se denomina conjunto ortonormal.

Por otra parte, los productos **xy'** y **xx'** no son escalares, sino matrices $n\times n$.

Definición 11: Dependencia Lineal de Vectores. Un conjunto de m vectores de orden n definidos sobre un campo F,

$$\mathbf{x}_1 = [x_{11}, x_{12}, \ldots, x_{1n}]$$

$$\mathbf{x}_2 = [x_{21}, x_{22}, \ldots, x_{2n}]$$

$$\ldots\ldots\ldots\ldots\ldots\ldots \qquad (1.27)$$

$$\mathbf{x}_m = [x_{m1}, x_{m2}, \ldots, x_{mn}]$$

son linealmente dependientes cuando existen m elementos de F, no todos ceros, $k_1, k_2, \ldots, k_m$, de forma que

$$k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + \ldots + k_m\mathbf{x}_m = 0$$

En caso contrario, los vectores se llaman linealmente independientes.

Teorema 30. Cualquier vector $\mathbf{x}$ de orden n y el vector nulo $\mathbf{0}$ son linealmente dependientes.

Teorema 31. Si m vectores son linealmente dependientes, cualquiera de éstos se puede expresar como combinación lineal de los otros.

Teorema 32. Si m vectores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ son linealmente independientes, mientras que el conjunto obtenido al añadir otro vector $\mathbf{x}_{m+1}$ es linealmente dependiente, éste se puede expresar como combinación lineal de los m vectores.

Teorema 33. Si de m vectores $\mathbf{x}_1, \ldots, \mathbf{x}_m$ hay $r < m$ vectores linealmente dependientes, los vectores del conjunto total son linealmente dependientes.

Definición 12: Rango. La característica r de una matriz **A**, cuadrada o no, es el mayor número entero posible r tal que **A** contenga por lo menos algún menor de orden r que no sea nulo; si **A** es la matriz nula se pondrá r = 0. O sea, r (**A**) representa el número máximo de vectores columna (o hilera) de **A** que son linealmente independientes.

Teorema 34. Si el r(**A**), $m \leq n$, asociado a los m vectores de la definición 11, es r < m, existen en el conjunto r vectores que son linealmente independientes, y cada uno de los m-r restantes se puede expresar como combinación lineal de estos r vectores.

Teorema 35. La condición necesaria y suficiente para que los vectores (1.27) sean linealmente dependientes es que el r(**A**) sea r < m. Si r(**A**) = m, los vectores son linealmente independientes. El conjunto de vectores (1.27) es linealmente independiente si m>n; y si éste es así, también lo es cualquier subconjunto del mismo.

Teorema 36. Sean **A** y **B** matrices mxn. Entonces $\mathrm{r}(\mathbf{A}+\mathbf{B}) \leq \mathrm{r}(\mathbf{A})+\mathrm{r}(\mathbf{B})$.

Demostración. Sea un conjunto de r(**A**) columnas linealmente independientes de **A**, y β un conjunto de r(**B**) columnas linealmente independientes de **B**; sea γ la unión de los dos conjuntos y β . Cada columna de **C=A+B** es una combinación lineal de las columnas de γ. Entonces r(**C**) no es mayor que el número de columnas en γ, y así $\mathrm{r}(\mathbf{A}+\mathbf{B}) \leq \mathrm{r}(\mathbf{A})+\mathrm{r}(\mathbf{B})$. Este teorema se puede extender a un número finito de matrices.

Teorema 37. Sean **A** y **B** matrices tales que **AB** existe. Entonces $r(\mathbf{AB}) \leq r(\mathbf{A})$ y $r(\mathbf{AB}) \leq r(\mathbf{B})$.

Demostración. Las columnas de **AB** son combinaciones lineales de las columnas de **A**; así, pertenecen al espacio generado por las columnas de **A;** y así $r(\mathbf{AB}) \leq r(\mathbf{A})$. De forma similar, $r(\mathbf{B}'\mathbf{A}') \leq r(\mathbf{B}')$; pero $r(\mathbf{B}'\mathbf{A}') = r(\mathbf{AB})$ y $r(\mathbf{B}') \leq r(\mathbf{B})$; así que $r(\mathbf{AB}) \leq r(\mathbf{B})$.

Corolario 1. Si **A** es no singular, y **AB** existe; entonces, $r(\mathbf{AB}) = r(\mathbf{B})$.

Demostración. Por el teorema $r(\mathbf{AB}) \leq r(\mathbf{B})$; pero $\mathbf{B} = (\mathbf{A}^{-1})(\mathbf{AB})$, así que $r(\mathbf{B}) \leq r(\mathbf{AB})$. Por lo tanto, $r(\mathbf{AB}) \leq r(\mathbf{B}) \leq r(\mathbf{AB})$, y así $r(\mathbf{AB}) = r(\mathbf{B})$. Se puede de manera similar verificar que, si **CA** existe, $r(\mathbf{CA}) = r(\mathbf{C})$.

Corolario 2. El rango de una matriz no cambia por pre y postmultiplicarla por matrices no singulares; o sea, en símbolos, si **A** y **C** son no singulares $r(\mathbf{ABC}) = r(\mathbf{B})$. De aquí que, en particular si una matriz nxn es tal que det $\mathbf{A} \neq 0$, entonces es de rango n; por el contrario, una matriz cuadrada con det $\mathbf{A} = 0$ es de rango $r < n$.

Corolario 3. Si **A** y **B** son matrices cuadradas y **AB**=0, entonces **A**=0, o **B**=0, sino **A** y **B** son ambas singulares.

Demostración. Supóngase, en contradicción, que **A** es regular y $\mathbf{B}\neq 0$. Entonces $r(\mathbf{AB}) = r(\mathbf{B})>0$ y $\mathbf{AB} \neq 0$. En forma análoga, si **B** es regular y $\mathbf{A} \neq 0$, $\mathbf{AB} = 0$.

Definición 13: Dependencia Lineal de Funciones. De modo semejante a la definición 11, p funciones $f_1,\dots,f_p$ de una o más variables se dicen linealmente dependientes, si existe una relación $c_1f_1+\dots+c_pf_p = 0$, válida para todo valor de las variables sin que todas las c_i sean nulas. Cuando haya varias relaciones de esta forma entre las funciones f_i, dichas relaciones se dirán independientes si los vectores correspondientes $\mathbf{c}=(c_1,\dots,c_p)$ son linealmente independientes.

Transformaciones Lineales.

Una transformación lineal

$$\begin{aligned} x_1 &= a_{11}y_1 + a_{12}y_2+\dots+a_{1n}y_n \\ x_2 &= a_{21}y_1 + a_{22}y_2+\dots+a_{2n}y_n \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ x_m &= a_{m1}y_1 + a_{m2}y_2+\dots+a_{mn}y_n \end{aligned} \tag{1.28}$$

pone una relación entre dos conjuntos de variables x_i y y_j . La matriz $\mathbf{A} =(a_{ij})$ se llama la matriz de la transformación.

Cuando $\mathbf{x} = (x_1,\dots,x_m)$ y $\mathbf{y} = (y_1,\dots,y_n)$ son vectores columna, los miembros derechos de las ecuaciones (1.28) son los elementos de la matriz producto **Ay**, que es de orden $m\times 1$, o sea, un vector columna; así, pues, (1.28) expresa que los elementos respectivos de los vectores columna **x** y **Ay** son iguales, por lo que en notación matricial, la transformación (1.28) llega a ser **x=Ay**.

Definición 14: Forma Bilineal. Al considerar las componentes x_i y y_j de los vectores columna **x** y **y**, como dos conjuntos de variables independientes, se observa que la matriz producto **x'Ay**, es un escalar, y se tiene:

$$\mathbf{x}'\mathbf{A}\mathbf{y} = \sum_{i,j} a_{ij} x_i y_j \tag{1.29}$$

donde $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$. Esta expresión se nombra una forma bilineal en las variables x_i y y_j, se ve que en notación matricial tiene una expresión simple.

Definición 15: Forma Cuadrática. Si en la definición anterior se tiene m=n, **x=y**, la forma bilineal (1.29) llega a ser

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i,j}^{n} a_{ij}\, x_i\, x_j = \mathrm{Q}(\mathbf{x}) = \mathrm{Q}, \tag{1.30}$$

expresión que se llama forma cuadrática en las variables $x_1, x_2, \ldots, x_n$. La matriz **A** se llama matriz de la forma Q; si, en particular, **A=I** se tendrá $\mathrm{Q} = \mathbf{x}'\mathbf{I}\mathbf{x} = \mathbf{x}'\mathbf{x} = x_1^2 + \ldots + x_n^2$.

Teorema 38. Sin pérdida de generalidad, se puede suponer que la matriz de una forma cuadrática es simétrica.

Demostración. Puesto que Q es una matriz 1x1, $Q'=Q$; por lo tanto, $\mathbf{x}'\mathbf{A}'\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x}$, o $Q = \frac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{A}'\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}$, donde $\mathbf{B} = \frac{1}{2}\left(\mathbf{A} + \mathbf{A}'\right)$. Así, si se remplaza la matriz de la forma cuadrática Q por **B** se obtendrá la misma Q y $\mathbf{B}'=\mathbf{B}$.

Ejemplo 2. Si $Q = 5x_1^2 + 12x_1x_2 + 7x_2^2$, se escribe

$$Q = 5x_1^2 + 6x_1x_2 + 6x_1x_2 + 7x_2^2 = \left(x_1, x_2\right)\begin{bmatrix} 5 & 6 \\ 6 & 7 \end{bmatrix}\left(x_1, x_2\right)$$

y la matriz de la forma es simétrica.

Cuando la matriz de una transformación lineal sea o no singular, la transformación se designará con el mismo nombre. En muchas situaciones donde una forma cuadrática se conoce, es interesante hacer una transformación lineal no singular de las variables, tal que la forma de Q es más simple cuando se expresa en las nuevas variables. Se verá que siempre es posible eliminar los términos de productos cruzados en Q, así que en las nuevas variables $y_1, y_2, \ldots, y_n$ la forma de Q es simplemente $Q = \sum_1^n \lambda_i y_i^2$, y que esto puede ser hecho con una clase especial y simple de transformación lineal, que se llama ortogonal. Los coeficientes resultantes $\{\lambda_i\}$ son, entonces de gran interés.

Ahora bien, supóngase que se transforma de $\mathbf{x} = (x_1,\ldots,x_n)$ a $\mathbf{y} = (y_1,\ldots,y_n)$ mediante una transformación lineal $\mathbf{y} = \mathbf{P}^{-1}\mathbf{x}$,

$$\mathbf{x} = \mathbf{P}\mathbf{y}. \tag{1.31}$$

Entonces, **x'** = **y'P'**, por lo tanto, Q = **x'Ax** = **y'P'APy,** y la matriz de la forma cuadrática cuando se expresa en términos de **y** es **P'AP.** La formulación matricial del problema de transformación es así, dada una matriz simétrica **A**, hallar una matriz **P** no singular tal que **P'AP** tiene una forma simple.

Teorema 39. Si **A** es simétrica y de rango r, uno al menos de sus menores de orden r es distinto de cero. Por consiguiente, al particularizar, el rango de una matriz diagonal es igual al número de sus elementos diagonales diferentes de cero.

Por definición, el rango de una forma cuadrática Q = **x'Ax** $= \sum_{i,j} a_{ij}\, x_i\, x_j$ es igual al rango de la matriz **A** de la forma; según que **A** sea o no singular, lo mismo se dice de la forma Q. Una transformación lineal no singular no afecta el rango de la forma; así, si mediante aquella, Q se cambia en $\sum_{1}^{r} \Delta_i y_i^2$ (con $\Delta_i \neq 0$), para i = 1, 2,..., r, se deduce que el rango de Q es r. El rango es, pues, el menor número de variables independientes por medio del cual se expresa Q cuando se la cambia por una transformación lineal regular.

La reducción de una forma cuadrática Q, de rango r, a la forma diagonal,

$$h_1\, y_1^2 + h_2\, y_2^2 + \ldots + h_r\, y_r^2, \qquad h_i \neq 0 \tag{1.32}$$

se puede realizar por el algoritmo de Lagrange, que consiste, en esencia, en formar de manera repetida cuadrados perfectos. Máltsev (1972) lo ilustra así:

Ejemplo 3. Redúzcase a la forma diagonal, la forma

$$Q = x_1^2 + 4x_2^2 + 8x_3^2 - x_4^2 - 4x_1x_2 + 6x_1x_3 - 12x_2x_3 + 2x_3x_4 + x_2x_5 - x_4x_5 .$$

Se tiene

$$Q = (x_1 - 2x_2 + 3x_3)^2 - x_2^2 - x_4^2 + 2x_3x_4 + x_2x_5 - x_4x_5.$$

Al transformar

$$x_1' = x_1 - 2x_2 + 3x_3, \quad x_i' = x_i \quad (i > 1),$$

se obtiene la forma

$$Q_1 = x_1'^2 - x_3'^2 + 2x_3'x_4' - x_4'^2 + x_2'x_5' - x_4'x_5' = x_1'^2 - (x_3' - x_4')^2 + x_2'x_5' - x_4'x_5',$$

que por medio de la transformación

$$y_3 = x_3' - x_4', \quad y_i = x_i' \qquad (i \neq 3),$$

se reduce a la forma

$$Q_2 = y_1^2 - y_3^2 + y_2y_5 - y_4y_5.$$

Ahora, se realiza la transformación

$$y_4' = y_2 - y_4 - y_5, \quad y_i' = y_i \quad (i \neq 4),$$

obteniendo así la forma

$$Q_3 = y_1'^2 - y_3'^2 + (y_4' + y_5')y_5'$$

$$= y_1'^2 - y_3'^2 + (y_5' + \frac{1}{2}y_4')^2 - \frac{1}{4}y_4'.$$

Esta forma mediante la transformación

$$z_2 = y_5' + \frac{1}{2}y_4', \quad z_i = y_i' \quad (i = 1,3,4), \qquad z_5 = y_2'$$

se reduce a la forma (1.32)

$$Q_4 = z_1^2 + z_2^2 - z_3^2 - \frac{1}{4}z_4^2.$$

Para encontrar la matriz **T** de la transformación de las variables z_i en las variables iniciales x_i es suficiente, multiplicar las matrices de las transformaciones intermedias.

Teorema 40. Cuando Q se expresa como $Q = L_1^2+...+L_p^2$, donde las L_i son funciones lineales de las $x_1,...,x_n$, y si existen de modo exacto h relaciones lineales independientes entre las L_i; entonces el rango de Q es p-h. De aquí que, si se conoce la existencia de por lo menos h relaciones lineales, se puede aseverar que el rango de Q es $\leq$ p-h.

Cuando $Q = \mathbf{x'Ax}$ es una forma cuadrática no singular, la forma $Q^{-1} = \mathbf{x'A^{-1}x}$ se nombra forma inversa de Q, y se ve en seguida que $(\mathbf{Q}^{-1})^{-1} = \mathbf{Q}$.

Definición 16: Variables Contragradientes. Sean $\mathbf{x} = (x_1, ..., x_n)$ y $\mathbf{t} = (t_1, ... ,t_n)$ dos vectores columna variables. Si se introducen nuevas variables $\mathbf{y} = (y_1, ... , _m)$ y $\mathbf{u} = (u_1, ... ,u_m)$, por medio de las transformaciones

$$\mathbf{y = Cx}, \qquad \mathbf{t = C'u}, \tag{1.33}$$

donde **C** es de orden mxn, se tiene

$$\mathbf{t'x = u'Cx = u'y}. \tag{1.34}$$

Es decir, la forma bilineal $\mathbf{t'x} = t_1x_1 + ... + t_nx_n$ se cambiará en la forma análoga $\mathbf{u'y} = u_1 y_1 + ... + u_my_m$ en las nuevas variables. Dos conjuntos de variables x_i y t_i que se transforman según (1.33) se denominan conjunto de variables contragradientes. Para m=n , y $\mathbf{C}$ no singular, las relaciones (1.33) llegan a ser

$$\mathbf{y = Cx}, \qquad \mathbf{u = (C')^{-1} t}. \tag{1.35}$$

Ecuaciones Lineales

Supóngase que F es un campo; se considera la cuestión de hallar n escalares, elementos de F, $x_1, ..., x_n$ que satisfagan las condiciones

$$\begin{aligned} a_{11} x_1 + a_{12} x_2 + ... + a_{1n} x_n &= h_1 \\ a_{21} x_1 + a_{22} x_2 + ... + a_{2n} x_n &= h_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1} x_1 + a_{m2} x_2 + ... + a_{mn} x_n &= h_m \end{aligned} \tag{1.36}$$

donde $h_1, h_2, ..., h_m$ y a_{ik}, $1 \leq i \leq m$, $1 \leq k \leq n$, son elementos de F; a (1.36) se le denomina un sistema (o juego) de m ecuaciones lineales con n incógnitas; todo n-tuple $(x_1,...,x_n)$ de elementos de F que satisface cada ecuación de (1.36) se le nombra una solución del sistema; cuando $h_1 = h_2 = ... = h_m = 0$, se dice que el sistema es homogéneo, o que cada ecuación es homogénea (Hoffman y Kunze, 1992; Lipschutz, 1971).

Se dirá que dos sistemas de ecuaciones lineales, son equivalentes si cada ecuacion de cada sistema es combinación lineal de las ecuaciones del otro sistema (Kurosch, 1975; Lipschutz, 1971).

Las formas matriciales equivalentes del sistema (1.36) son

$$\mathbf{Ax} = \mathbf{h} \tag{1.37}$$

$$\mathbf{Ax} = \mathbf{0} \tag{1.38}$$

donde $\mathbf{A} = (a_{ij})$, la matriz de los coeficientes, $\mathbf{x} = (x_1,\ldots,x_n)$ y $\mathbf{h} = (h_1, \ldots, h_m)$ o $\mathbf{h} = 0$, según se trate de (1.37) o (1.38). Si $\mathbf{A}$ es no singular en (1.37) y m = n se puede multiplicar a la izquierda ambos lados por $\mathbf{A}^{-1}$, y así se obtiene las solución única $\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{h}$ o, en forma explícita,

$$x_k = \frac{1}{|\mathbf{A}|}\sum_{i=1}^{n} h_i A_{ik} \qquad (k = 1,2, \ldots, n). \tag{1.39}$$

Así que, x_k está expresada por una fracción cuyo denominador es $|\mathbf{A}|$, y numerador igual al determinante obtenido de $\mathbf{A}$, sustituyendo los elementos de la columna k por los segundos miembros $h_1,\ldots,h_n$. Esta solución clásica, porque se halló en 1750, se llama Regla de Cramer. Para el sistema homogéneo, $\mathbf{A}$ tiene rango $r \leq n$; si $r = n$ este sistema tiene la solución trivial única $\mathbf{x} = \mathbf{0}$; cuando $r < n$ es posible hallar $n - r$ vectores $\mathbf{c}_1,\ldots,\mathbf{c}_{n-r}$ linealmente independientes, tales que la solución general de (1.36) se puede escribir en la forma $\mathbf{x} = t_1\,\mathbf{c}_1 + \ldots + t_{n-r}\,\mathbf{c}_{n-r}$, donde las t_i son cualesquier constantes.

Clases de Matrices Cuadradas. Raíces Características .

Una matrtiz $\mathbf{C}$ de orden n que tiene en sus columnas un conjunto de vectores ortonormales se designa como matriz ortogonal; más precisamente se tiene el concepto que sigue.

Definición 17: Matriz Ortogonal. Se entiende por ella toda matriz cuadrada **C** = (c_{ik}) de orden n, tal que

$$\mathbf{CC'} = \mathbf{C'C} = \mathbf{I}. \tag{1.40}$$

Las formas cuadráticas con matrices idempotentes se usan de modo extenso en teoría estadística.

Definición 18: Matrices Idempotentes. Sea **B** una matriz n×n tal que (i) **B** = **B'** e (ii) **B** = $\mathbf{B}^2$. En tal caso, **B** se define como una matriz idempotente si (ii) se satisface; y como matriz simétrica idempotente si ambas condiciones se cumplen de forma simultánea (Graybill, 1983).

Teorema 41. Sea **A** una matriz n×n (simétrica) idempotente. Entonces

(1) **I-A** es matriz (simétrica) idempotente;

(2) **A'** es matriz (simétirca) idempotente;

(3) $\mathbf{PAP}^{-1}$ es una matriz idempotente, si **P** es una matriz n×n no singular;

(4) **P'AP** es una matriz (simétrica) idempotente, donde **P** es una matriz nxn ortogonal.

Teorema 42. El determinante de una matriz ortogonal vale ± 1.

Teorema 43. La transpuesta **C'** de una matriz ortogonal **C,** es también ortogonal.

Teorema 44. La matriz inversa de una matriz ortogonal, es también ortogonal.

Demostración. Se deduce de (1.40) que $\mathbf{C}^{-1} = \mathbf{C}'$, al pasar en ésta a las matrices traspuestas, resulta:

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

De (1.40), $\mathbf{C}^{-1} = \mathbf{C}'$, y así, por definición de matriz inversa, resulta que $C_{ik} = Cc_{ik}$ para todos los valores de i y k, donde C es el determinante de la matriz $\mathbf{C}$, c_{ik} es el elemento generador de $\mathbf{C}$ y C_{ik} es el cofactor correspondiente a c_{ik}; de aquí mediante las relaciones (1.20), se obtiene

$$\sum_{j=1}^{n} c_{ij}\, c_{kj} = \begin{cases} 1 & \text{para } i = k, \\ 0 & \text{para } i \neq k, \end{cases} \tag{1.41}$$

$$\sum_{j=1}^{n} c_{ji}\, c_{jk} = \begin{cases} 1 & \text{para } i = k, \\ 0 & \text{para } i \neq k, \end{cases} \tag{1.42}$$

Teorema 45. El producto $\mathbf{C}_1\mathbf{C}_2 \ldots \mathbf{C}_n$ de n matrices ortogonales $\{\mathbf{C}_i\}$ del mismo orden, es también una matriz ortogonal.

Si se da un cierto número $p < n$ de filas $c_{i1}, c_{i2}, \ldots, c_{in}$ $(i = 1, 2, \ldots, p)$, que satisfagan a las relaciones (1.41), siempre es posible hallar otras n - p filas tales que la matriz resultante, de orden n, sea ortogonal. También puede aplicarse este resultado a las columnas utilizando las identidades (1.42).

Es común, que en muchas proposiciones sobre la distribución de formas cuadráticas, se hace uso considerable de la traza de una matriz.

Definición 19: Traza. La traza de una matriz **A** n×n, que se escribe como tr (**A**), se define como la adición de los elementos de la diagonal de **A**; o sea,

$$\mathrm{tr}\,(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}, \tag{1.43}$$

Teorema 46. Si **A** y **B** son matrices n×n, y a y b son escalares, siendo así

$$\mathrm{tr}\,(a\mathbf{A} + b\mathbf{B}) = a\,\mathrm{tr}\,(\mathbf{A}) + b\,\mathrm{tr}\,(\mathbf{B}) \tag{1.44}$$

Teorema 47. Si **A** es una matriz n×n; en tal caso

$$\mathrm{tr}\,(\mathbf{A}') = \mathrm{tr}\,(\mathbf{A}) \tag{1.45}$$

Teorema 48. Sea **A** una matriz mxn; siendo así

$$\mathrm{tr}\,(\mathbf{A}'\mathbf{A}) = \mathrm{tr}\,(\mathbf{A}\mathbf{A}') = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^{2} \quad y \quad \mathrm{tr}\,(\mathbf{A}'\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{A}') = \mathbf{0} \text{ si y sólo si } \mathbf{A} = \mathbf{0}.$$

Teorema 49. Si **A** es una matriz n×n y $\mathbf{A}^2 = m\mathbf{A}$; entonces,

$$\mathrm{tr}\,(\mathbf{A}) = m\, r(\mathbf{A})\,. \tag{1.46}$$

Corolario. Si **A** es idempotente, siendo así m = 1 y tr (**A**) = r (**A**).

Teorema 50. Sean **A** y **B** matrices m×n y n×m, respectivamente; entonces

$$\mathrm{tr}\,(\mathbf{AB}) = \mathrm{tr}\,(\mathbf{BA}) \tag{1.47}$$

Demostración. Sea **AB** = **C**; entonces $c_{pq} = \sum_{j=1}^{n} a_{pj}\, b_{jq}$. Sea **G** = **BA**; entonces

$g_{rs} = \sum_{i=1}^{m} b_{ri}\, a_{is}$. Pero tr (**AB**) = tr (**C**) = $\sum_{p=1}^{m} c_{pp} = \sum_{p=1}^{m} \sum_{j=1}^{n} a_{pj}\, b_{jp}$. También,

$$\mathrm{tr}(\mathbf{BA}) = \mathrm{tr}(\mathbf{G}) = \sum_{r=1}^{n} g_{rr} = \sum_{r=1}^{n} \sum_{i=1}^{m} b_{ri}\, a_{ir}\,.$$

Así, tr(**AB**) = tr(**BA**). Este teorema se puede inducir a un número finito de matrices.

Teorema 51. Sea **A** cualquier matriz n×n y sea **P** cualquier matriz n×n regular; en tal caso

$$\mathrm{tr}\,(\mathbf{A}) = \mathrm{tr}\,(\mathbf{P}^{-1}\,\mathbf{AP}) \qquad (1.48)$$

Si **P** es una matriz ortogonal, siendo así

$$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}\,(\mathbf{P}'\mathbf{AP}) \qquad (1.49)$$

Demostración. Por el Teorema 49,

$$\mathrm{tr}\,(\mathbf{P}^{-1}\mathbf{AP}) = \mathrm{tr}\left[\mathbf{P}^{-1}(\mathbf{AP})\right] = \mathrm{tr}\left[(\mathbf{AP})\mathbf{P}^{-1}\right] = tr\,(\mathbf{AI}) = tr\,(\mathbf{A}).$$

Ya que una forma cuadrática es un escalar, es igual a su propia traza y, por lo tanto

$$\mathbf{x}'\mathbf{Ax} = tr\,(\mathbf{x}'\mathbf{Ax}) = tr\,(\mathbf{Axx}') \qquad (1.50)$$

Definición 20: Transformación Ortogonal. Una transformación lineal **x = Cy,** se llama ortogonal si su matriz **C** es ortogonal.

$$|\mathbf{A}-\lambda\mathbf{I}| = \begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix} = 0. \tag{1.52}$$

Ejemplo 4. Ilustrar la definición 21, para $\mathbf{A} = \begin{bmatrix} 7 & -2 & 1 \\ -2 & 10 & -2 \\ 1 & -2 & 7 \end{bmatrix}$.

Al aplicar (1.52) la ecuación secular resulta:

$$|\mathbf{A}-\lambda\mathbf{I}| = \begin{vmatrix} \lambda-7 & 2 & -1 \\ 2 & \lambda-10 & 2 \\ -1 & 2 & \lambda-7 \end{vmatrix} = \lambda^3 - 24\lambda^2 + 180\lambda - 432 = 0,$$

que al resolver da como raíces características 6, 6, 12.

$$\text{Para } \lambda = 6, \text{ se tiene } \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix} (x_1, x_2, x_3) = 0,$$

es decir $x_1 - 2x_2 + x_3 = 0$, y se escogen como vectores propios el par de vectores de forma mutua ortogonales $\mathbf{x}_1 = (1, 0, -1)$ y $\mathbf{x}_2 = (1, 1, 1)$. Para $\lambda = 12$, sea $\mathbf{x}_3 = (1, -2, 1)$ el vector propio asociado. Al tomar la forma normalizada de estos vectores como columnas de $\mathbf{C}$ se tiene

$$\mathbf{C} = \begin{Bmatrix} 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & 1/\sqrt{3} & -2/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{Bmatrix}.$$

Así $\mathbf{C}^{-1}\mathbf{A}\mathbf{C} = \mathbf{D}\,(6, 6, 12)$. También se ha ilustrado la proposición de enseguida:

Teorema 52. Las raíces características de una matriz simétrica real son todas reales.

Teorema 53. Sea **B** cualquier matriz n×n de rango p.

1). Si **B** es idempotente, entonces **B** tiene p raíces características diferentes de cero y son igual cada una a +1;

2). Si **B** es simétrica, en tal caso una condicion necesaria y suficiente para que **B** sea idempotente es que tenga p raíces características diferentes de cero y cada una igual a +1.

Teorema 54. Sea **A** una matriz n×n con raíces características $\lambda_1,\lambda_2,\ldots,\lambda_n$; entonces tr (**A**) = $\sum_{i=1}^{n}\lambda_i$.

Puesto que **C** es regular, en (1.51), **A** y **L** tienen el mismo rango; de aquí que, el rango de **A** sea igual al número de raíces λ_i que no sean nulas. También de (1.51) se obtiene, al tomar determinantes en ambos lados, y al tener presente que $C^2 = 1$, así como aplicar el método de reducción a la forma triangular

$$A = \lambda_1 \lambda_2 \ldots \lambda_n \tag{1.53}$$

Si **A** es regular, la identidad

$$\left|\mathbf{A}^{-1} - \lambda\mathbf{I}\right| = (-\lambda)^n \mathrm{A}^{-1} \left|\mathbf{A} - \tfrac{1}{\lambda}\,\mathbf{I}\right| \tag{1.54}$$

indica que las raíces características de $\mathbf{A}^{-1}$ son las recíprocas de las respectivas de **A.**

Si **B** es una matriz de orden m×n, m ≤ n, de rango m, la matriz simétrica **BB'**, de orden m, tiene positivas todas sus raíces características. En particular, se deduce que **BB'** es regular; lo que se verifica así: si, en (1.51) se pone **A** = **BB'**, y se expresa una raíz característica cualquiera λ_i mediante (1.53).

Formas Cuadráticas no Negativas.

Definición 22: Forma Cuadrática no Negativa. Cuando para todo valor real de las variables x_i, se tiene:

$$Q(x_1, \ldots, x_n) = \sum_{i,k=1}^{n} a_{ik}\, x_i\, x_k \geq 0, \tag{1.55}$$

Q se llamará una forma cuadrática no negativa.

Definición 23: Forma Cuadrática Definida Positiva. Se dice que Q es definida positiva, si el signo de igualdad se verifica sólo cuando todas las variables sean nulas en (1.55); en seguida se da una ilustración de este concepto.

Ejemplo 5. Sea la forma

$$\mathbf{x'Ax} = [x_1, x_2, x_3]' \begin{bmatrix} 3 & 5 & 1 \\ 5 & 13 & 0 \\ 1 & 0 & 1 \end{bmatrix} (x_1, x_2, x_3)$$

$$= 3x_1^2 + 13x_2^2 + x_3^2 + 10x_1x_2 + 2x_1x_3$$

$$= (x_1 + 2x_2)^2 + (x_1 + 3x_2)^2 + (x_1 + x_3)^2$$

que es positiva para todo valor real de las x_i, excepto $x_1 = x_2 = x_3 = 0$; esto es, $\mathbf{x} = \mathbf{0}$, en este caso **x'Ax** siempre es cero. Por lo tanto, **x'Ax** es definida positiva.

Definición 24: Forma Cuadrática Semidefinida Positiva. Se llamará semidefinida positiva a una forma Q que sea no negativa, y no sea definida positiva.

Ejemplo 6. Sea la forma

$$\mathbf{x'Ax} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix} (x_1 \quad x_2 \quad x_3)$$

$$= (x_1 - 2x_2)^2 + (6x_1 - 4x_3)^2 + (3x_2 - x_3)^2$$

es cero cuando $\mathbf{x} = (2, 1, 3)$. Por consiguiente, **x'Ax** es semidefinida positiva.

Ejemplo 7. Sea la forma

$$\mathbf{y'y} = \mathbf{y'Iy} = \sum_{i=1}^{n} y_i^2$$

es definida positiva porque es cero sólo cuando $\mathbf{y} = \mathbf{0}$. Sin embargo, la forma

$$\mathbf{y'y} - n\bar{y}^2 = \mathbf{y}'(\mathbf{I} - n^{-1}\mathbf{J}_n)\mathbf{y} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2;$$

donde $\mathbf{J}_n$ es una matriz cuadrada de orden n con cada elemento igual a 1, es una forma cuadrática semidefinida positiva porque es cero no sólo si $\mathbf{y} = \mathbf{0}$, sino también si cada elemento de **y** es el mismo; esto es, si $\mathbf{y} = \alpha\mathbf{1}$ para cualquier α.

Definición 25: Menor Principal. Un menor de una matriz **A** se dice principal si, se obtiene al suprimir en **A** las filas y columnas correspondientes del mismo subíndice.

Teorema 55. Una forma cuadrática Q con coeficientes reales es definida positiva; si, y sólo si, todos sus menores principales son de modo estricto positivos.

Ejemplo 8. La forma

$$Q = 5x_1^2 + x_2^2 + 5x_3^2 + 4x_1x_2 - 8x_1x_3 - 4x_2x_3$$

es definida positiva, ya que sus menores principales

$$5, \quad \begin{vmatrix} 5 & 2 \\ 2 & 1 \end{vmatrix} = 1, \quad \begin{vmatrix} 5 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 5 \end{vmatrix} = 1$$

son todos positivos.Sin embargo, la forma

$$Q_1 = 3x_1^2 + x_2^2 + 5x_3^2 + 4x_1x_2 - 8x_1x_3 - 4x_2x_3$$

no es definida positiva, puesto que su segundo menor principal es negativo $\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} = -1$.

Teorema 56. Si **B** es una matriz n×n idempotente de rango n, siendo así **B = I.** Si **B** es simétrica idempotente de rango menor que n, en tal caso **B** es una matriz semidefinida positiva.

Cada una de las propiedades de una forma cuadrática Q = **x'Ax** de ser no negativa, definida positiva o semidefinida positiva es, de forma evidente, invariante para toda transformación lineal regular; el mismo nombre de estas propiedades recibe la matriz simétrica $\mathbf{A} = (a_{ik})$ correspondiente a la forma Q.

Teorema 57. La matriz simétrica **A** es definida positiva si, y sólo si, todos sus menores principales tienen determinantes positivos.

Corolario. Las matrices definidas positivas son regulares.

Nótese que el recíproco de este corolario no es cierto: las matrices no singulares no son, en general, definidas positivas.

Teorema 58. Para **P** regular, **P'AP** es o no (semi-)definida positiva, según tenga o no **A** estas propiedades.

La transformación ortogonal $\mathbf{x} = \mathbf{Cy}$, en donde **C** es la matriz ortogonal de (1.51), cambia la forma Q en otra forma que sólo consta de términos de segundo grado, a saber:

$$Q(x_1, \ldots, x_n) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2, \tag{1.56}$$

o también, $\mathbf{x'Ax} = \mathbf{y'Ly}$ en notación matricial, en donde las λ_i son las raíces características de **A**, mientras que **L** es la matriz diagonal. Por esta transformación ortogonal, la forma

$$Q - \lambda(x_1^2 + \ldots + x_n^2) \tag{1.57}$$

se cambia en la forma

$$(\lambda_1 - \lambda)y_1^2 + \ldots + (\lambda_n - \lambda)y_n^2. \tag{1.58}$$

Si λ es menor o igual que la menor de las raíces características de **A**, (1.58) es no negativa, y por lo tanto, (1.57) –cuya matriz es **A** - λ**I**– también lo es.

Teorema 59. Los valores propios de una matriz (semi-)definida positiva son todos no negativos; o sea, cero o positivos.

Corolario. Las matrices semidefinidas positivas son singulares.

El corolario recíproco no es verdadero: en general, las matrices singulares no son semidefinidas positivas.

En efecto, cuando la forma Q sea definida positiva, también lo será la forma del lado derecho de (1.56), esto implica que, todas las raíces características $\{\lambda_i\}$ son positivas; de aquí, y al aplicar (1.53), se tiene A > 0, lo que significa que **A** es regular. Por otra parte, si Q es semidefinida positiva, un razonamiento análogo indica que, al menos una de las raíces características es cero, luego A = 0. Cuando Q sea de característica r, hay r valores propios positivos, siendo nulos los demás n-r; también hay n-r vectores $\mathbf{x}_p = (x_1^{(p)}, \ldots, x_n^{(p)})$ linealmente independientes, tales que Q $(\mathbf{x}_p) = 0$.

Teorema 60. Una matriz simétrica es definida positiva si, y sólo si, puede ser escrita como **P'P** para una **P** regular.

Teorema 61. **A'A** es definida positiva, cuando **A** tiene el máximo rango columna; y, semidefinida positiva de otra forma.

Corolario. **AA'** es definida positiva, cuando **A** tiene el máximo rango fila; y, semidefinida positiva de otro modo.

La significancia geométrica de esta transformación ortogonal es que, por medio de una rotación adecuada del sistema de coordenadas, la cuádrica Q = constante, está referida a sus ejes trascendentales; así, cuando Q sea definida positiva, esta ecuación representa un elipsoide de n dimensiones cuyos semiejes son $\lambda_i^{-\frac{1}{2}}$; para las formas semidefinidas Q, se obtienen varios tipos de cilindros elípticos.

La transformación $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ cambia la forma $Q = \mathbf{x}'\mathbf{A}\mathbf{x}$ en la forma inversa $Q^{-1} = \mathbf{y}'\mathbf{A}^{-1}\mathbf{y}$; así que, cuando Q sea definida positiva, también lo es Q^{-1}, y viceversa; lo que se puede ver de forma directa a partir de (1.54).

Teorema 62. Una suma de matrices (semi-)definidas positivas es (semi-)definida positiva.

Teorema 63. Una matriz simétrica **A**, de orden n y rango r, puede ser escrita como **LL'** donde **L** es n×r de rango r; es decir, **L** tiene el máximo rango columna.

Considérese ahora la identidad (1.21), para el caso de una matriz **A** simétrica y definida positiva: ya que toda matriz principal menor de **A** también es definida positiva por el Teorema 55 (Ejemplo 8), se infiere que el último término del lado derecho de (1.21) es una forma cuadrática definida positiva en las variables $a_{12},\ldots,a_{1n}$, de modo que se tendrá $0 < A \leq a_{11}A_{11}$, y al inducir resulta

$$0 < A \leq a_{ii}\, A_{ii}, \qquad (i = 1, 2,\ldots,n) \tag{1.59}$$

Al repetir n veces este razonamiento se obtiene:

$$0 < A \leq a_{11}a_{22}\ldots a_{nn}. \tag{1.60}$$

Para toda matriz no negativa esta identidad se verifica, poniendo el doble signo ≤ en vez de < ; cuando **A** sea una matriz diagonal se presenta la igualdad.

Teorema 64. Una matriz simétrica que tiene las raíces características iguales a 0 y 1, es idempotente.

Demostración. Una matriz simétrica **X** siempre se puede expresar en forma canónica, bajo la identidad ortogonal **U'XU** = **D** donde **D** es diagonal, los elementos diagonales, la raíces características de **X**. Cuando estas raíces son 0 ó 1

$$\mathbf{U'XU} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

de la cual es trivial demostrar que $\mathbf{X}^2 = \mathbf{X}$ (Searle, 1966).

En ciertas aplicaciones estadísticas se presentan relaciones de la clase:

$$\sum_{1}^{n} x_1^2 = Q_1 + \ldots + Q_k, \tag{1.61}$$

donde Q_i es, para i = 1, 2,...,k, una forma cuadrática de las variables $x_1,\ldots,x_n$, y característica r_i.

Para k = 2 se supone que existe una transformación ortogonal que convierte Q_1 en una adición de r_1 cuadrados; es decir, $Q_1 = \sum_{1}^{r_1} y_i^2$. Al aplicar esta transformación a ambos miembros de (1.61), el izquierdo se convierte en $\sum_{1}^{n} y_i^2$; y en consecuencia, se observa que Q_2 se convertirá en $\sum_{r_1+1}^{n} y_i^2$. Así, pues, la característica de Q_2 es $r_2 = n - r_1$, y todos sus valores propios son 0 ó 1.

Ejemplo 9. Considérese la relación:

$$\sum_{1}^{n} x_i^2 = n\bar{x}^2 + \sum_{1}^{n} (x_i - \bar{x})^2, \tag{1.62}$$

en la que $\bar{x}$ es la media. Toda transformación ortogonal **y = Cx** tal que la fila primera de **C** sea $1/\sqrt{n}, 1/\sqrt{n},\ldots, 1/\sqrt{n}$, convertirá la forma $n\bar{x}^2 = \left(\frac{x_1}{\sqrt{n}} + \frac{x_2}{\sqrt{n}} + \ldots \frac{x_n}{\sqrt{n}}\right)^2$ en y_1^2.

También esta transformación convierte $\sum_{1}^{n}(x_i - \overline{x})^2$ en $\sum_{2}^{n} y_i^2$. En el análisis de $\sum_{1}^{n} x_i^2$, según (1.62), los rangos de los dos términos del miembro derecho son 1 y n - 1, respectivamente.

Teorema 65. Si **A** y **V** son simétricas y **V** es definida positiva; entonces **AV** tiene las raíces características 0 y 1, implica que **AV** es idempotente.

Demostración. $|\mathbf{AV} - \lambda\mathbf{I}| = 0$ tiene las raíces 0 y 1. Por el Teroema 60, **V = P'P** para alguna matriz regular **P**. Por lo tanto,

$$|\mathbf{P}||\mathbf{AV} - \lambda\mathbf{I}||\mathbf{P}^{-1}| = 0 \qquad \text{tiene raíces 0 y 1;}$$

esto es,

$$|\mathbf{PAP}' - \lambda\mathbf{I}| = 0 \qquad \text{tiene raíces 0 y 1.}$$

Así, **PAP'** tiene las raíces características 0 y 1; pero **PAP'** es simétrica porque **A** lo es. Por lo tanto, por el Teorema 64, **PAP'** es idempotente; esto es, **PAP'PAP' = PAP'.** En consecuencia, porque **P** es regular, **AP'PAP'P=APP'** ; es decir, **AVAV = AV,** muestra la idempotencia de **AV** (Searle, 1971).

CAPÍTULO III

LA NORMAL Y SUS DISTRIBUCIONES DERIVADAS.

En este capítulo se resumen, las propiedades sobresalientes de la distribución normal y sus derivadas; no se intenta llegar al pleno rigor; ya que casi cualquier texto de teoría estadística, como Borovkov (1988), Cramér (1970), Dudewicz y Mishra (1988), Graybill (1976), Hogg y Craig (1970), Lindgren (1976), Mood y Graybill (1972), Mood *et al.* (1974), Searle (1971) y Wilks (1962), da los detalles pertinentes.

Funciones de Densidad Multivariada

Sean n variables aleatorias $X_1, X_2, \ldots, X_n$, para las cuales $x_1, x_2 \ldots, x_n$ representa un conjunto de valores realizados, u observaciones, la función de densidad acumulativa se escribe como

$$\Pr(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = F(x_1, x_2, \ldots . x_n) \tag{2.1}$$

(Cramér, 1970; Dudewicz y Mishra, 1988; Lindgren, 1976). Entonces, la función de densidad se define por

$$f(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F(x_1, \ldots, x_n). \tag{2.2}$$

(Hogg y Craig, 1970; Lindgren, 1976); al integrar (2.2) se llega a

$$F(x_1,\ldots,x_n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f(u_1,\ldots,u_n)\, du_n \ldots du_1 . \tag{2.3}$$

Dos son las condiciones que $f(x_1, \ldots, x_n)$ debe satisfacer, a saber:

$f(x_1, x_2, \ldots, x_n) \geq 0$ para $-\infty < x_i < \infty$ y para todo i,

y

$$\int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n = 1,$$

(Hogg y Craig, 1970; Koroliuk, 1981; Mood y Graybill, 1972; Mood *et al.*, 1974).

Ejemplo 1. La función

$$f(x_1, x_2, x_3, x_4) = \begin{cases} 16x_1x_2x_3x_4 & 0 < x_i < 1 \\ 0 & \text{de otra forma} \end{cases}$$

es una función de densidad, porque satisface las dos condiciones anteriores; la probabilidad de que un punto caiga en la región $x_1 < \frac{1}{2}$ y $x_4 > \frac{1}{3}$, se calcula

$$\Pr(x_1 < \tfrac{1}{2}, x_4 > \tfrac{1}{3}) = \int_{1/3}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{1/2} f(x_1, x_2, x_3, x_4)\, dx_1\, dx_2\, dx_3, dx_4$$

$$= \int_{1/3}^{1}\int_{0}^{1}\int_{0}^{1}\int_{0}^{1/2} 16x_1x_2x_3x_4\, dx_1\, dx_2\, dx_3\, dx_4$$

$$= \frac{2}{9}.$$

La función de densidad marginal de $x_{k+1}, \ldots, x_n$, que se podría llamar de las "últimas n - k x's" consiste en $f(x_1, x_2, \ldots, x_n)$ luego de integrar fuera de las primeras k x's; o sea,

$$g\,(x_{k+1},\,\dots\,,\,x_n\,) = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} f\;(x_1,\,\dots\,,x_k\,,\,x_{k+1},\,\dots\,,\,x_n\,)\;dx_1\,\dots\,dx_k \qquad (2.4)$$

(Cramér, 1970; Searle, 1971). La distribución condicional para las "primeras k x's" dadas las "últimas n - k x's" consiste en la razón geométrica de f $(x_1,x_2, \dots , x_n)$ a la marginal para las "últimas n - k x' s"; esto es

$$f\;(x_1,\dots,x_k \,|x_{k+1},\dots,x_n\,) = \frac{f(x_1,x_2,\dots,x_n)}{g(x_{k+1},\dots,x_n)}, \qquad (2.5)$$

expresión que se obtiene al aplicar (2.2) y(2.4), según lo detallan Cramér (1970) y Searle (1971).

Momentos de una Distribución.

Cuando existe, la cantidad $E\left[(x-b)^k\right]$ se llama el momento k de la variable aleatoria X alrededor del punto x = b, o el momento k de la distribución de X; el momento absoluto k alrededor de x = b se define como la cantidad $E\left[|x-b|^k\right]$. Algunas veces, el término momento por sí mismo se usa para denotar un momento alrededor del punto particular x = 0; y se define como el valor esperado de la potencia k de x_i; esto es,

$$\mu_{x_i}^{(k)} = E\,(x_i^k) = \int_{-\infty}^{\infty} x_i^k\, g(x_i)dx_i$$

$$= \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} x_i^k\, f\,(x\,,\,x_2\,,\dots\,,x_n\,)\;dx_1\;dx_2\,\dots\,dx_n\,, \qquad (2.6)$$

relación obtenida al aplicar (2.4).

La covarianza entre la i-ésima y la j-ésima variables, para $i \neq j$ se define como

$$\sigma_{ij} = E(x_i - \mu_i)(x_j - \mu_j)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i - \mu_i)(x_j - \mu_j)g(x_i, x_j)dx_i dx_j$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_i - \mu_i)(x_j - \mu_j)f(x_1,\ldots,x_n)dx_1\ldots dx_n, \qquad (2.7)$$

y en forma análoga, la varianza de la i-ésima variable se define por

$$\sigma_{ii} = \sigma_i^2 = E(x_i - \mu_i)^2$$

$$= \int_{-\infty}^{\infty}(x_i - \mu_i)^2 g(x_i)dx_i$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_i - \mu_i)^2 f(x_1,\ldots,x_n)dx_1\ldots dx_n. \qquad (2.8)$$

Las varianzas y las covarianzas entre las variables en el vector

$$\mathbf{x}' = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}$$

se dan en (2.7) y (2.8), respectivamente. Al arreglar estas varianzas y covarianzas como los elementos de una matriz, resulta la matriz de varianzas y covarianzas de las x_i; en otras palabras,

$$\text{var}(\mathbf{x}) = \mathbf{V} = \{\sigma_{ij}\} \qquad \text{para } i, j = 1, 2, \ldots, n.$$

Los elementos diagonales de **V** son las varianzas, y los no diagonales, las covarianzas.

La varianza de una variable aleatoria escalar x se escribirá como v(x), mientras la matriz de varianzas y covarianzas de un vector de variables aleatorias **x** será denotado por var (**x**).

El vector de medias correspondiente a **x**' es

$$E(\mathbf{x}') = \mu' = [\mu_1 \ \mu_2 \cdots \mu_n]$$

así, por la definición de la varianza y la covarianza,

$$\text{var}(\mathbf{x}) = E\left[(\mathbf{x} - \mu)(\mathbf{x} - \mu)'\right] = \mathbf{V} \quad (2.9)$$

simismo, ya que la correlación entre la i-ésima y j-ésima variables es $\sigma_{ij}/\sigma_i\sigma_j$, la matriz ː correlaciones se expresa por

$$\mathbf{R} = \left\{\frac{\sigma_{ij}}{\sigma_i\sigma_j}\right\} = \mathbf{D}\{1/\sigma_i\}\mathbf{V}\mathbf{D}\{1/\sigma_i\} \quad \text{para } i, j = 1,2,\ldots,n, \quad (2.10)$$

ɔnde, al usar (1.51), las **D** son matrices diagonales con elementos $1/\sigma_i$ para i = 1, 2,...,n. 'e modo evidente los elementos diagonales de **R** son todos la unidad, es simétrica y se ɔnoce como la matriz de corrrelaciones .

Teorema 1. La matriz **V** es definida no negativa.

Demostración. Considérese **t'Vt** para algún vector no nulo **t**; entonces,

$$\mathbf{t'Vt} = \sum_i \sum_j t_i t_j \sigma_{ij} = v\left(\sum_i t_i x_i\right) = v(\mathbf{t'x})$$

ue, por definición de varianza, es positiva; a menos que **t'x** es de forma idéntica cero, ı el caso v (**t'x**) = **t'Vt** = 0. Por lo tanto, **V** es una matriz definida no negativa; además, ː también lo es, porque en (2.10) todas las σ son positivas (Searle, 1971).

Transformación Lineal de Variables

Si las variables **x** se convierten a variables **y** por la transformación lineal **y** = **Tx**, los momentos de **y** se derivan con facilidad; *e. g.*,

$$\mu_y = T\mu_x \qquad \text{y} \qquad \text{var}(\mathbf{y}) = \mathbf{TVT'} \qquad 2.11)$$

Cuando se hace una transformación de esta naturaleza que comprende una **T** regular, una integral que incluye las diferenciales $dx_1, dx_2, \ldots, dx_n$ se convierte al sustituir las x en términos de las y, y remplazar las diferenciales por $\|\mathbf{J}\|$ $dy_1\, dy_2 \ldots dy_n$, donde $\|\mathbf{J}\|$ es el jacobiano de las x con respecto a las y. La matriz jacobiana se define como $\mathbf{J} = \{\partial x_i / \partial y_i\}$ para i, j = 1, 2,...,n y $\|\mathbf{J}\|$ es el valor absoluto del determinante de **J**. Porque $\mathbf{x} = \mathbf{T}^{-1}\mathbf{y}$, esto significa $\mathbf{J} = \mathbf{T}^{-1'}$, y así $\|\mathbf{J}\| = 1/\|\mathbf{T}\|$.De aquí que, cuando la transformación de **x** a **y** sea **y** = **Tx**, el producto de las diferenciales

$$dx_1 \ldots dx_n \text{ se reemplaza por } (dy_1 \ldots dy_n)/\ \|\mathbf{T}\|. \qquad (2.12)$$

Este es el procedimiento, *e. g*, al derivar la función de densidad de **y** = **Tx** de **x.** Se sustituye $\mathbf{x} = \mathbf{T}^{-1}\mathbf{y}$ para cada x_i en $f(x_1, \ldots, x_n)$. Supóngase que la función resultante de las y se escribe como $f(\mathbf{T}^{-1}\mathbf{y})$. Entonces, por la primera de las condiciones que $f(x_1, \ldots x_n)$ debe satisfacer, la transformación resulta

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f\left(\mathbf{T}^{-1}\mathbf{y}\right) (1/\|\mathbf{T}\|) dy_1 \text{L}\, dy_n = 1,$$

como se debería esperar. Ahora, supóngase que $h(y_1,\ldots,y_n)$ es la función de densidad de las y. Entonces,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_n) dy_1 \ldots dy_n = 1.$$

Por lo tanto, por comparación se halla

$$h\,(y_1, \ldots, y_n) = \frac{f\left(\mathbf{T}^{-1}\mathbf{y}\right)}{\|\mathbf{T}\|},$$

así la deriva Searle (1971).

Ejemplo 2. Cuando

$$y_1 = 3x_1 - 2x_2$$

$$y_2 = 5x_1 - 4x_2$$

sea la transformación $\mathbf{y} = \mathbf{Tx}$, entonces $\|\mathbf{T}\| = 2$; y

$$h\,(y_1, y_2) = \frac{1}{2}\, f\left[x_1 = 2y_1 - y_2, x_2 = \frac{1}{2}(5y_1 - 3y_2)\right].$$

Funciones Generatrices de Momentos

Los momentos y las relaciones entre las distribuciones, a menudo son derivados por medio de las funciones generatrices de momentos. En el caso univariado la función generatriz de momentos (fgm) de la variable aleatoria x, escrita como una función de t es de acuerdo con Mood y Graybill (1972),

$$M_x^{(t)} = E\,(e^{tx})$$

$$= \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

$$= \int_{-\infty}^{\infty} \left(1 + tx + t^2x^2 / 2 + t^3x^3 / 3! + \ldots\right) f(x) dx \qquad (2.13)$$

$$= \left[1 + t\mu_x^{(1)} + \left(t^2/2\right)\mu_x^{(2)} + \left(t^3/3!\right)\mu_x^{(3)} + \ldots\right].$$

Por consiguiente,

$$\mu_x^{(k)} = \left.\frac{\partial^k M_x(t)}{\partial t^k}\right|_{t=0} \tag{2.14}$$

Del mismo modo, para alguna función de x, h (x), su fgm se expresa

$$M_{h(x)}(t) = E\left(e^{t\,h(x)}\right) = \int_{-\infty}^{\infty} e^{t\,h(x)} f(x)dx \tag{2.15}$$

y el k-ésimo momento alrededor de cero de (2.15) es

$$\mu_{h(x)}^{(k)} = \left.\frac{\partial^k M_{h(x)}(t)}{\partial t^k}\right|_{t=0} \tag{2.16}$$

La fgm de la distribución conjunta de n variables utiliza un vector de parámetros **t'** = $[t_1\ t_2,\ldots,t_n]$:

$$M_x(t) = E\left(e^{t_1x_1 + t_2x_2 + \Lambda + t_nx_n}\right)$$

$$= E(e^{t'x})$$

$$= \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} e^{\mathbf{t'x}} f\left(x_1,\ldots,x_n\right)dx_1\ldots dx_n. \tag{2.17}$$

Y, por un método análogo, la fgm de la forma cuadrática **x'Ax,** llega a ser

$$M_{x'Ax}(t) = E(e^{t\,x'Ax})$$

$$= \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} e^{t\,\mathbf{x}'A\mathbf{x}} f\,(x_1,\ldots,x_n)dx_1\ldots dx_n$$

(2.18)

Otros dos usos importantes de la fgm son:

1). Si dos variables aleatorias tienen la misma fgm, entonces tienen también igual función de densidad;

2). Si $M_{(x_1 x_2)}(t_1, t_2) = M_{x_1}(t_1)M_{x_2}(t_2)$, entonces x_1 y x_2 son independientes.

Funciones Características

Se llama función característica (fc) de la magnitud aleatoria X, con la función de distribución $F(x) = P(X < x)$, una función de valor complejo

$$f(t) = Me^{itX} = \int_{-\infty}^{\infty} e^{itx}\, dF(x) \qquad (2.19)$$

$$= \int \cos tx\, dF(x) + i\int \text{sen } tx\, dF(x) \qquad t \varepsilon R$$

(Koroliuk, 1981; Loéve, 1976). En particular, si existe la densidad de distribución $p(x) = F'(x)$, la fc será la transformada de Fourier de la densidad de distribución:

$$f(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx \qquad (2.20)$$

Para una magnitud aleatoria discreta, que toma los valores x_k con la probabilidad p_k, la fc se representa por medio de la serie

$$f(t) = \sum_k e^{itxk} kp_k, \qquad t \varepsilon R \qquad (2.21)$$

(Koroliuk, 1981). De acuerdo con éste y Loéve (1976), las propiedades esenciales de la fc son:

1. $f(0) = 1$, $|f(t)| \leq 1$, $-\infty < t < \infty$.

2. f(t) es uniformemente continua en el eje numérico.

3. Cuando todo $n > 0$ sea entero, para cualesquiera números complejos $z_1, z_2, \ldots, z_n$ y todo número real $t_1, t_2, \ldots, t_n$

$$\sum_{k,r=1}^{n} f(t_k - t_r) z_k \bar{z}_r \geq 0. \tag{2.22}$$

Esta propiedad implica la definición positiva de la fc.

4. $f(-t) = \overline{f(t)}$; o sea, forma hermitiana.

5. La fc de una suma de magnitudes aleatorias independientes es igual al producto de las fc de los sumandos:

$$f_{x_1+x_2}(t) = f_{x_1}(t) f_{x_2}(t) \tag{2.23}$$

6. Si $Y = aX + b$, donde a y b son constantes,

$$f_Y(t) = f_X(at)\, e^{ibt} \tag{2.23'}$$

Teorema 2: Teorema de Bohner-Jinchin. Para que una función continua f(t), definida en un eje real y que satisface la condición f(0) = 1, sea característica, es necesario y suficiente que esté en lo positivo definida.

Teorema 3: Teorema de Inversión. Una función de distribución se determina en forma unívoca mediante su fc, f(t). Tiene lugar la siguiente fórmula de inversión

$$F(x) - F(y) = \frac{1}{2\pi} \lim_{c \to \infty} \int_{-c}^{c} \frac{e^{-itx} - e^{-ity}}{it} f(t)\, dt, \tag{2.24}$$

válida para cualesquiera puntos x y y de continuidad de la distribución F (x).

En particular, si $|f(t)/t|$ es integrable en el infinito, entonces

$$F(x) - F(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} f(t)\,dt \tag{2.25}$$

Si, en cambio, la fc f(t) es sumable en el eje real, entonces la función de distribución F(x) tiene la densidad continua acotada p (x) = F'(x), la cual se determina por la fórmula

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} f(t)\,dt \tag{2.26}$$

Para la distribución en retículo

$$p_k = P(X = a + kh) = \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-it(a+kh)} f(t)\,dt \tag{2.27}$$

Diversas fórmulas de inversión se pueden obtener haciendo uso de la Igualdad de Parseval.

Teorema 4: Teorema de Continuidad. Una sucesión de las funciones de distribución $F_n(x)$ converge en forma débil hacia la distribución de probabilidades F(x), cuando, y sólo cuando, la sucesión de sus funciones características f_n (t) converge hacia la función límite continua f(t); en este caso f(t) es una función característica de la distribución límite F(x) y la convergencia de f_n (t) hacia f(t) es uniforme en todo intervalo finito; si una sucesión de funciones características integrables f_n (t) converge en media hacia la fc límite f(t), o sea, $\int_{-\infty}^{\infty} |f_n(t) - f(t)|\,dt \to 0$ para $n \to \infty$, entonces la sucesión de las correspondientes densidades de distribución $p_n(x)$ converge de manera uniforme hacia la densidad límite de distribución, p(x), (Cramér, 1970; Koroliuk, 1981).

Sea $\mu = (\mu_1, \mu_2, \ldots, \mu_m)$ un vector aleatorio cuyos valores están definidos en el espacio euclidiano R_m con la fc $F(x) = P\{\mu_1 < x_1, \mu_2 < x_2, \ldots, \mu_n < x_n\}$,

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \varepsilon R_n.$$

La fc del vector aleatorio μ se determinará mediante la igualdad

$$f(t) = E\, \mathbf{e}^{it\mu} = \int_{R_n} e^{it\mu}\, dF(\mathbf{x}), \tag{2.28}$$

donde $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, $\mathbf{tx} = \sum_1^n t_k x_k$ es el producto escalar de los vectores $\mathbf{t}$ y $\mathbf{x}$.

Las propiedades de las funciones características de las distribuciones multidimensionales, son análogas a las de las funciones características de las magnitudes aleatorias. En seguida se indican algunas diferencias. Se llaman momentos del vector aleatorio μ los números

$$m_{k_1 k_2 \ldots k_n} = E\left(\mu_1^{k_1} \mu_2^{k_2} \ldots \mu_n^{k_n}\right) \tag{2.29}$$

El número $k = k_1 + k_2 + \ldots + k_n$ recibe el nombre de orden del momento. Los momentos de índices enteros se pueden determinar por derivación de la fc

$$m_{k_1 k_2 \ldots k_n} = \left. \frac{(-i)^k \partial^k f(t)}{\partial t_1^{k_1} \partial t_2^{k_2} \ldots \partial t_n^{k_n}} \right|_{t=0} \tag{2.30}$$

La Distribución Normal Univariada

Cuando la variable aleatoria X tenga una distribución normal con media μ y varianza σ^2, se escribirá $x \sim N(\mu, \sigma^2)$ o "x es $N(\mu, \sigma^2)$"; la función de densidad de x es entonces

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}, \qquad \text{para } -\infty < x < \infty,$$

en la cual la aplicación de (2.6) y (2.8) mostrará que $E(x) = \mu$, y $E(x - \mu)^2 = \sigma^2$, respectivamente. De acuerdo con (2.13), la fgm de x es

$$M_x(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left[tx - \frac{1}{2}(x-\mu)^2/\sigma^2\right]dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} exp - \frac{1}{2}\left\{ \left[x - (\mu + t\sigma^2)\right]^2 - (2\mu t\sigma^2 + t\sigma^4) \right\}/\sigma^2 dx$$

$$= e^{\mu t + \frac{1}{2}t^2\sigma^2}\left(1/\sigma\sqrt{2\pi}\right)\int_{-\infty}^{\infty} exp\left[-\left(x - \mu - t\sigma^2\right)^2/2\sigma^2\right]dx$$

$$= e^{\mu t + \frac{1}{2}t^2\sigma^2}$$

(Searle, 1971; Spiegel, 1994). De (2.18) se establece que $\mu_x^{(1)} = \mu$, y $\mu_x^{(2)} = \sigma^2 + \mu^2$, de donde se infiere que $\sigma^2 = \mu_x^{(2)} - \mu^2 = E(x - \mu)^2$.

Distribución Normal Multivariada

Si las variables aleatorias en $\mathbf{x}' = [x_1\ x_2 \ldots x_n]$ tienen una distribución normal multivariada, con vector de medias μ y matriz de varianzas y covarianzas $\mathbf{V}$, esto se escribe "$\mathbf{x} \sim N(\mu, \mathbf{V})$" o "$\mathbf{x}$ es $N(\mu, \mathbf{V})$". Si $E(x_i) = \mu$ para todo i entonces $\mu = \mu\mathbf{1}$; y cuando las x_i son de manera mutua independientes, todas con la misma varianza σ^2, entonces $\mathbf{V} = \sigma^2\mathbf{I}$ y, esto se resume "$\mathbf{x}$ es $N(\mu\mathbf{1}, \sigma^2\mathbf{I})$" ; la cual equivale a la notación mas usual NID (μ, σ^2).

Cuando **V** es definida positiva, la función de densidad normal multivariada se expresa

$$f(x_1,\ldots,x_n) = \frac{e^{-\frac{1}{2}(\mathbf{x}-\mu)'\mathbf{V}^{-1}(\mathbf{x}-\mu)}}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} \qquad (2.31)$$

Si **A** es una matriz simétrica definida positiva de orden n, entonces

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-\frac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x}}\,dx_1\ldots dx_n = (2\pi)^{n/2}|\mathbf{A}|^{-1/2} \qquad (2.31)$$

resultado que se llama integral de Aitken; se puede probar al aplicar en lo esencial (2.12). Al aplicar (2.32) a (2.31), se tiene

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\ldots,x_n)dx_1\ldots dx_n = (2\pi)^{n/2}|\mathbf{V}^{-1}|^{-1/2}(\sqrt{2\pi})^n|\mathbf{V}|^{\frac{1}{2}} = 1,$$

como se debería esperar.

Teorema 5. Para cualesquier vector **g** y matriz simétrica definida positiva **W**, se tiene

$$(2\pi)^{\frac{1}{2}n}|\mathbf{W}|^{\frac{1}{2}}e^{\frac{1}{2}\mathbf{g}'\mathbf{W}\mathbf{g}} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left(-\tfrac{1}{2}\mathbf{x}'\mathbf{W}\mathbf{x}+\mathbf{g}'\mathbf{x}\right)dx_1\ldots dx_n. \qquad (2.33)$$

Demostración. De (2.31), se tiene

que al sustituir **g**' por μ'$\mathbf{W}^{-1}$, ésta da (2.33) (Searle, 1971). Como en (2.17) la fgm para la distribución normal multivariada consiste en

$$M_x(t) = (2\pi)^{\frac{1}{2}n}\left|\mathbf{V}\right|^{-\frac{1}{2}}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\exp\left[t'\mathbf{x}-\tfrac{1}{2}(\mathbf{x}-\mu)'\mathbf{V}^{-1}(\mathbf{x}-\mu)\right]dx_1\ldots dx_n\,.$$

Al reordenar el exponente, ésta llega a ser

$$M_x(t)=(2\pi)^{-\frac{1}{2}n}\left|\mathbf{V}\right|^{-\frac{1}{2}}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\exp\left[-\tfrac{1}{2}(\mathbf{x}-\mu-\mathbf{Vt})'\mathbf{V}^{-1}(\mathbf{x}-\mu-\mathbf{Vt})+\mathbf{t}'\mu+\tfrac{1}{2}\mathbf{t}'\mathbf{Vt}\right]dx_1\ldots dx_n$$

$$=\frac{e^{t'\mu+\frac{1}{2}t'\mathrm{vt}}}{(2\pi)^{\frac{1}{2}n}\left|\mathbf{V}\right|^{\frac{1}{2}}}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}(\mathbf{x}-\mu-\mathbf{Vt})'\mathbf{V}^{-1}(\mathbf{x}-\mu-\mathbf{Vt})\right]dx_1\ldots dx_n$$

Al hacer la transformación **y** = **x** - μ - **Vt** de **x** a **y**, para que el jacobiano sea la unidad, la integral entonces se reduce a la integral de Aitken con matriz $\mathbf{V}^{-1}$. Por lo tanto,

$$M_x(t) = \frac{e^{\mathbf{t}'\mu+\frac{1}{2}\mathbf{t}'\mathbf{Vt}}(2\pi)^{\frac{1}{2}n}\left|\mathbf{V}\right|^{-1/2}}{(2\pi)^{\frac{1}{2}n}\left|\mathbf{V}\right|^{1/2}} = e^{\mathbf{t}'\mu+\frac{1}{2}\mathbf{t}'\mathbf{Vt}} \tag{2.34}$$

Al diferenciar (2.34) en la forma de (2.14), se ve que el vector de medias es μ y la matriz de varianzas y covarianzas, **V**.

El p-vector aleatorio **X** se dice ser normal multivariado cuando, y sólo cuando la combinación lineal $\boldsymbol{a}'\mathbf{x} = a_1x_1 + \ldots + a_px_p$ es normal para cualquier a. Que tales distribuciones existen es claro del hecho que, cualquier combinación lineal de variables normales independientes es normal.

La función característica de una distribución normal multivariada, se puede derivar de la función característica normal univariada. Así, si μ y $\mathbf{M}$ son, respectivamente, el vector de medias y la matriz de covarianzas de $\mathbf{X}$, entonces

$$\mu_{a'\mathbf{x}} = a\mu, \qquad \sigma^2_{a'\mathbf{x}} = a'\mathbf{M}a$$

y

$$f_{a'\mathbf{x}}(t) = E\left[e^{ia'\mathbf{x}tt}\right] = f_{\mathbf{x}}(ta) = exp\left[i(a'\mu)t - \tfrac{1}{2}(a'\mathbf{M}a)\right]t^2,$$

así que, con $s \equiv ta$,

$$f_{\mathbf{x}}(s) = exp\left[is'\mu - \frac{1}{2}s'Ms\right].$$

Las comprobaciones de las tres proposiciones de en seguida, las desarrollan desde diferentes perspectivas técnicas Cramér (1970), Lindgren (1976), Mood y Graybill (1972) y Wilks (1962).

Teorema 6. Si $\mathbf{X}$ es normal multivariada y su segunda matriz de momentos es diagonal –todas las covarianzas iguales a 0– entonces las componentes de $\mathbf{X}$ son independientes.

Teorema 7. Las distribuciones marginales de una distribución normal multivariada, son todas normales multivariadas.

Teorema 8. Si $\mathbf{X}$ es normal multivariada, entonces $\mathbf{AX}$ es normal multivariada para cualquier matriz $\mathbf{A}$ (compatible).

Ejemplo 3. Sean Y_1, Y_2 y Y_3 variables aleatorias tipificadas independientes.

Considérese **X** definida por
$$\begin{cases} X_1 = 2Y_1 - Y_2 + Y_3, \\ X_2 = Y_1 - 3Y_2 \end{cases}$$

o

$$\mathbf{X} = \begin{pmatrix} 2 & -1 & 1 \\ 1 & -3 & 0 \end{pmatrix} \mathbf{Y}.$$

La matriz de covarianzas de **X** es

$$\mathbf{M} = \begin{pmatrix} 2 & -1 & 1 \\ 1 & -3 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & -3 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 5 \\ 5 & 10 \end{pmatrix},$$

con det $\mathbf{M} = 35$ y $\mathbf{M}^{-1} = \dfrac{1}{35}\begin{pmatrix} 10 & -5 \\ -5 & 6 \end{pmatrix}$.

Con éstas se escribe la fc de **X**: $\exp\left[-\frac{1}{2}\left(6t_1^2 + 10t_1t_2 + 10t_2^2\right)\right]$; y la función de densidad:

$$\frac{1}{2\pi\sqrt{35}} exp\left[-\frac{1}{70}\left(10x_1^2 - 10x_1x_2 + 6x_2^2\right)\right].$$

Supóngase en seguida que **Z** se define como

$$\mathbf{Z} = \begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 1 & 2 \end{pmatrix} \mathbf{X}.$$

La distribución de **Z** debe ser singular, puesto que hay más **Z**'s que **X**'s; en efecto, $\mathbf{Z}_3 = \mathbf{Z}_2 - \mathbf{Z}_1$, así la distribución de **Z** está concentrada en el plano $z_1 - z_2 + z_3 = 0$. La matriz de covarianzas de **Z** es

$$\mathbf{M}_z = \begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 6 & 5 \\ 5 & 10 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 6 & -3 & -9 \\ -3 & 54 & 57 \\ -9 & 57 & 66 \end{pmatrix}$$

que es singular.

Otra vez, en el caso multivariado, la definición de "normal" comprende distribuciones singulares. Cuando se escribió "$\mathbf{x}$ es N(μ,$\mathbf{V}$)", aunque no se clarificó, se dejó implícito que $\mathbf{V}$ era no singular; ahora se considerará la situación cuando $\mathbf{V}$ sea singular; una ilustración sencilla de esto es el ejemplo anterior. Para que tales variables estén distribuidas de modo normal, se enfatiza la singularidad de $\mathbf{V}$ por escribir, en general, $\mathbf{x} \sim$ NS(μ, $\mathbf{V}$).Porque $\mathbf{V}^{-1}$ no existe, la función de densidad de la distribución NS(μ,$\mathbf{V}$) no se puede escribir; sin embargo, su fc, en la fgm se remplaza it en vez de t, existe; ya que es $e^{it'\mu - \frac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t}}$. En consecuencia, por el Teorema de Continuidad para funciones características, se garantiza que la función de densidad existe, aún cuando no se puede escribir de manera explícita. La caracterización general de la distribución NS(μ,$\mathbf{V}$) dada por Anderson (1958) es útil aquí. Supóngase que $\mathbf{y}$ es un vector que tiene la distribución N($\mathbf{0}$,$\mathbf{1}$); entonces, las variables obtenidas por la transformación $\mathbf{x} = \mu + \mathbf{L}\mathbf{y}$ tienen la distribución NS(μ, $\mathbf{LL'}$), cuando $\mathbf{LL'}$ no es de rango completo. Las situaciones surgen en los modelos lineales que son similares a éste, cuando se desarrollan ecuaciones $\mathbf{X'Xb^o} = \mathbf{X'y}$ que tienen una solución $\mathbf{b}^0 = \mathbf{GX'y}$ donde $\mathbf{X'X}$ es singular; entonces, si $\mathbf{y}$ tiene distribución normal, $\mathbf{b}^0$ también; mas su matriz de varianzas y covarianzas será singular. Por lo tanto, la discusión de la distribución normal singular es pertinente. Se puede considerar que la distribución N(μ,$\mathbf{V}$), es un caso particular de la distribución NS(μ, **V**); o sea, en el Capítulo IV, **V** se verá como singular o no; en el caso que **V** sea no singular, los corolarios inductivos se reducen a los teoremas 2 - 6, respectivamente.

Cramér (1970), Searle (1971) y Wilks (1962), aunque por diferentes técnicas, verifican que la distribución condicional es también normal ; o sea,

$$\mathbf{x}_1 \mid \mathbf{x}_2 \sim N\left[\mu_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{x}_2 - \mu_2), \mathbf{W}_{11}^{-1}\right].$$

Supóngase que el vector $\mathbf{x}' = [x_1 x_2 \ldots x_n]$ se particiona en p subvectores $\mathbf{x}' = \left[\mathbf{x}_1' \mathbf{x}_2' \ldots \mathbf{x}_p'\right]$; entonces, una condición necesaria y suficiente para que los vectores sean de modo mutuo independientes es, en la partición correspondiente de $\mathbf{V} = \{\mathbf{V}_{ij}\}$ para i, j = 1, 2,...,p, que **V** = **0**, para i $\neq$ j. Searle (1971) demuestra esta proposición como sigue: La fgm de **x** es, por (2.34),

$$M_X(t) = e^{t'\mu + \frac{1}{2}t'\mathbf{V}t} = exp\left(\sum_{i=1}^{p} \mathbf{t}_i' \mu_i + \frac{1}{2}\sum_{i,j=1}^{p} \mathbf{t}_i' \mathbf{V}_{ij}\mathbf{t}_i\right).$$

y si $\mathbf{V}_{ij} = \mathbf{0}$ para i $\neq$ j la expresión anterior se reduce a

$$M_X(t) = exp\sum_{i=1}^{p}\left(\mathbf{t}_i' \mu_i + \frac{1}{2}\mathbf{t}_i' \mathbf{V}_{ii}\, \mathbf{t}_i\right) = \prod_{i=1}^{p} exp\left(\mathbf{t}_i' \mu_i + \frac{1}{2}\mathbf{t}_i' \mathbf{V}_{ii}\, \mathbf{t}_i\right)$$

Porque la fgm de la distribución conjunta de conjuntos de variables independientes, es el producto de sus fgm (segundo uso antes del tema fc), se concluye que las $\mathbf{x}_i$ son independientes; y viceversa, si ellas son independientes, cada una con su $\mathbf{K}_{ii}$ de varianzas y covarianzas; entonces, la fgm de la distribución se expresa

$$\prod_{i=1}^{p} exp(t_i' \mu_i + (1/2)\, t_i' \mathbf{K}_{ii}\, t_i) = exp\sum_{i=1}^{p}\left(t_i'\mu_i + (1/2)\, t_i' \mathbf{K}_{ii}\mathbf{t}_i\right)$$

$$= \exp\,(\mathbf{t}'\mu + \mathbf{t}'\mathbf{V}\mathbf{t}),$$

donde

$$\mathbf{V} = \text{diag}\left\{\mathbf{K}_{11}\mathbf{K}_{22}, \ldots, \mathbf{K}_{pp}\right\}. \text{ Luego } \mathbf{V}_{ij} = 0 \text{ para } i \neq j.$$

Distribución χ^2

La aplicación más común de la distribución χ^2 es que, cuando $\mathbf{x}$ sea $N(\mu\mathbf{1}, \sigma^2\mathbf{I})$, entonces $\sum_{i=1}^{n}(x_i - \bar{x})^2 / \sigma^2$ es χ^2_{n-1}; proposición, que se deducirá como un corolario del Teorema 3 Capítulo IV; sin embargo, también se puede comprobar por usar la transformación $\mathbf{y} = \mathbf{H}_0\mathbf{x}$ donde $\mathbf{H}_0$ consiste en las últimas n -1 filas de la matriz de Helmert; las propiedades de ésta se encuentran en Lancaster (1965).

χ^2 Central

Si $\mathbf{x}$ es $N(\mathbf{0}, \mathbf{I})$, entonces $x_1^2 + x_2^2 + \ldots + x_n^2$ tiene la distribución χ^2 central, u ordinaria, con n grados de libertad; de esta manera, cuando

$\mathbf{x}$ sea $N(\mathbf{0}, \mathbf{I})$ y $u = \sum_{1}^{n} x_i^2 = \mathbf{x}'\mathbf{x}$ entonces $u \sim \chi^2_n$.

La función de densidad se puede expresar

$$f(u) = \frac{u^{\frac{1}{2}n-1}e^{-\frac{1}{2}u}}{2^{\frac{1}{2}n}\Gamma\left(\frac{1}{2}n\right)} \qquad \text{para } u > 0 \tag{2.35}$$

donde $\Gamma\left(\frac{1}{2}n\right)$ es la función gamma. La fgm que corresponde a (2.35) resulta

$$M_u(t) = (1-2t)^{-(1/2)n}, \tag{2.36}$$

la cual se puede obtener como $M_{\mathbf{x'x}}(t)$ usando la función de densidad $N(\mathbf{0}, \mathbf{I})$ en (2.18), o bien directamente de (2.13) al usar (2.35). La media y la varianza de u son n y 2n, respectivamente; otras expresiones que se utilizan mucho son

$$E(u^2) = n(n+2), \qquad E(u^{1/2}) = \frac{2^{1/2}\Gamma((n+1)/2)}{\Gamma(n/2)}, \qquad E(1/u) = 1/(n-2) \qquad n > 2.$$

su fc es $\qquad f(t) = (1-2it)^{-n/2}$.

χ^2 no Central

Ya se vio que cuando $\mathbf{x}$ sea $N(\mathbf{0}, \mathbf{I}_n)$ la distribución de $\mathbf{x'x} = \sum_1^n x_i^2$, se conoce como distribución χ^2 ordinaria. Ahora considérese la distribución de $\mathbf{u} = \mathbf{x'x}$ cuando $\mathbf{x}$ sea $N(\mu, \mathbf{I})$; nótese que la única diferencia es la media de $\mathbf{x}$, μ, y no $\mathbf{0}$. La distribución resultante de $\mathbf{x'x}$ se conoce como χ^2 no central; la cual incluye los grados de libertad n, y también $\lambda = \frac{1}{2}\mu'\mu = \frac{1}{2}\sum\mu_i^2$, conocido como el parámetro de no centralidad; esto equivale a $\chi^{2'}(n, \lambda)$; si $\mu = 0$, entonces $\lambda = 0$ y se reduce a la χ^2 ordinaria. En la literatura el parámetro de no centralidad es común que se defina de forma diferente, algunos autores usan $\lambda = \delta^2$, mientras otros $\lambda = \frac{1}{2}\delta^2$, ambos usan el mismo símbolo λ. Su función de densidad es

$$f(u) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\frac{u^{\frac{1}{2}n+k-1}\; e^{-(1/2)u}}{2^{(1/2)n+k}\Gamma((1/2)n+k)}, \tag{2.37}$$

una suma ponderada infinita de funciones de densidad de χ^2 ordinarias; o sea, la función de densidad de la distribución $\chi^2_{\frac{1}{2}n+k}$. Su fgm es

$$M_u(t) = (1-2t)^{-(1/2)n} e^{-\lambda\left[1-(1-2t)^{-1}\right]}, \tag{2.38}$$

la cual se deriva por los métodos análogos a los indicados para la χ^2 ordinaria; los detalles se dan en Graybill (1976) y Searle (1971), por ejemplo. Su fc se obtiene de (2.38) con sólo remplazar t por it.

Las expresiones $n+2\lambda$ *y* $2n+8\lambda$, son la media y la varianza de la χ^2 no central, respectivamente; las cuales se pueden deducir mediante la diferenciación de la fgm, o de forma directa de la independencia de las x_i; de esta manera, por sumación sobre i, se tiene

$$\begin{aligned} E\sum x_i^2 &= \sum E\left(x_i^2\right) = \sum\left(\sigma_x^2 + \mu_i^2\right) \\ &= \sum 1 + \sum \mu_i^2 \lambda \\ &= n + 2\lambda; \end{aligned}$$

y

$$
\begin{aligned}
v\left(\sum x_i^2\right) &= \sum v\left(x_i^2\right) \\
&= \sum v\left[(x_i-\mu_i)^2 + 2\mu_i(x_i-\mu_i) + \mu_i^2\right] \\
&= \sum v(x_i-\mu_i)^2 + 4\sum \mu_i^2 v(x_i-\mu_i) \\
&= \sum\left\{E(x_i-\mu_i)^4 - \left[E(x_i-\mu_i)^2\right]^2\right\} + 4\sum \mu_i^2\sigma_x^2 \\
&= \sum\left\{3\sigma_x^4 - \left(\sigma_x^2\right)^2\right\} + 8\lambda \\
&= 2n + 8\lambda
\end{aligned}
$$

Las propiedades de χ^2 no central, se reducen a las correspondientes de la χ^2 ordinaria, como se debería esperar. Si, para $i = 1,2\ldots,k$, las u_i son $\chi^{2'}(n_i,\lambda_i)$ e independientes, entonces $\sum u_i$ es $\chi^{2'}\left(\sum n_i, \sum \lambda_i\right)$; este resultado se llama propiedad reproductiva o Teorema de la Adición; la verificación de esta proposición se establece a través de la fgm y la independencia de las u_i; los pormenores se hallan en Graybill (1976), Scheffé (1959) y Searle (1971).

Distribución F

Se distinguirán tres casos, la F central, la F no central y la F doble no central.

F Central.

La base de la distribución F, se forma de dos variables independientes, cada una al tener χ^2 central; así, si

$$u_1 \text{ es } \chi^2_{n_1} \quad \text{y} \quad u_2 \quad \text{es } \chi^2_{n_2}; \text{ entonces } v = \frac{u_1/n_1}{u_2/n_2} \sim F_{n_1,n_2},$$

la distribución F con n_1 y n_2 grados de libertad. La función de densidad resulta ser

$$f(v) = \frac{\Gamma\left(\frac{1}{2}n_1 + \frac{1}{2}n_2\right) n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2} v^{\frac{1}{2}n_1 - 1}}{\Gamma\left(\frac{1}{2}n_1\right)\Gamma\left(\frac{1}{2}n_2\right)(n_2 + n_1 v)^{\frac{1}{2}n_1 + \frac{1}{2}n_2}} \quad \text{para} \quad v > 0, \tag{2.39}$$

la cual se deriva en Borovkov (1988). La media de la distribución es $n_2/(n_2-2)$ para $n_2>2$, y la varianza es $2n_2^2\left[1+(n_2-2)/n_1\right]/(n_2-2)^2(n_2-4)$ para $n_2 > 4$. La fgm de la distribución F no existe. La distribución de Fisher también a veces se llama distribución de Snedecor; esto se debe al hecho de que Fisher propuso utilizar y tabuló, en realidad, no la distribución de v, sino la de la variable $\frac{1}{2} \ln v$. En cuanto a la distribución de v, ésta fue tabulada un poco más tarde por Snedecor.

F no Central.

Si u_1 y u_2 son independientes, y u_1 es $\chi^{2'}(n_1,\lambda)$ y u_2 es $\chi^2_{n_2}$; entonces $v = \dfrac{u_1/n_1}{u_2/n_2}$ se distribuye como $F'(n_1,n_2,\lambda)$, la distribución F no central con n_1 y n_2 grados de libertad y parámetro de no centralidad λ. Su función de densidad se escribe

$$f(v) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{n_1^{\frac{1}{2}n_1+k} n_2^{\frac{1}{2}n_2} \Gamma(\frac{1}{2}n_1 + \frac{1}{2}n_2 + k)}{\Gamma(\frac{1}{2}n_1 + k)\Gamma\left(\frac{1}{2}n_2\right)} \cdot \frac{v^{\frac{1}{2}n_1+k-1}}{(n_2 + n_1 v)^{\frac{1}{2}n_1 + \frac{1}{2}n_2 + k}},$$

cuya deducción se basa en la independencia de u_1 y u_2; así como en la equivalencia de las transformaciones, $v = n_2u_1/n_1u_2$ y $z = u_1 + u_2$, con $u_1 = n_1vz/(n_1v + n_2)$ y $u_2 = n_2z/(n_1v + n_2)$; los pormenores se pueden ver en Graybill (1976) y Searle (1971).

Cuando $\lambda = 0$ f(v) se reduce a (2.39), la función de densidad de la F central, para la cual $\lambda = 0$, $k = 0$. La media de la distribución es

$$E(v) = \frac{n_2}{n_2 - 2}\left(1 + \frac{2\lambda}{n_1}\right),$$

y su varianza

$$\frac{2n_2^2}{n_1^2(n_2 - 2)}\left[\frac{(n_1 + 2\lambda)^2}{(n_2 - 2)(n_2 - 4)} + \frac{n_1 + 4\lambda}{n_2 - 4}\right].$$

Cuando $\lambda = 0$, éstas se reducen, claro, a la media y varianza de F_{n_1,n_2}.

F no Central Doble

La distribución F no central doble se basa en la relación geométrica de dos variables χ^2 no centrales independientes; así, si u_1 es $\chi^{2'}(n_1,\lambda_1)$ y u_2 es $\chi^{2'}(n_2,\lambda_2)$, entonces $v = n_2u_1/n_1u_2$ se distribuye como $F''(n_1,n_2,\lambda_1,\lambda_2)$, la distribución F no central doble con grados de libertad n_1 y n_2 y parámetros de no centralidad λ_1 y λ_2; su función de densidad se deduce exactamente de la misma manera que la de la F no central, y da

f (v) =

$$\sum_{k_1,k_2=0}^{\infty} \frac{[exp(-\lambda_1 - \lambda_2)]\lambda_1^{k_1}\lambda_2^{k_2}\Gamma\left(\frac{1}{2}n_1 + \frac{1}{2}n_2 + k_1 + k_2\right)n_1^{\frac{1}{2}n_1+k_1}n_2^{\frac{1}{2}n_2+k_2}v^{\frac{1}{2}n_1+k_1-1}}{k_1!k_2!\Gamma\left(\frac{1}{2}n_1 + k_1\right)\Gamma\left(\frac{1}{2}n_2 + k_2\right)(n_1v + n_2)^{\frac{1}{2}n_1+\frac{1}{2}n_2+k_1+k_2}}$$

Distribución t

Se distinguirán dos casos, como en la χ^2, t central y t no central.

t Central.

La distribución t de Student, se basa en la razón de una variable distribuida de modo normal a una que tiene una distribución χ^2; así, cuando x es N(0, 1) y u es χ^2_n, independiente de x; entonces

$z = x/\sqrt{u/n}$ se distribuye como t_n,

la distribución t con n grados de libertad. Se ve de manera fácil que $-t$ tiene la misma distribución y, por lo tanto, la distribución de Student es simétrica con respecto al origen de coordenadas. Su función de densidad es

$$f(z) = \frac{\Gamma\left(\frac{1}{2}n + \frac{1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{1}{2}n\right)}\left(1 + \frac{z^2}{n}\right)^{-\frac{1}{2}(n+1)} \qquad \text{para } -\infty < z < \infty, \tag{2.40}$$

la cual se deduce en Borovkov (1988); su media es 0 y su varianza n/(n-2) para n > 2; su fgm no existe.

Una aplicación frecuente de esta distribución, es que si **x** es N($\mu\mathbf{1}$, $\sigma^2\mathbf{I}$), entonces

$$\frac{\bar{x} - \mu}{1/\sqrt{n}}\sqrt{\frac{n-1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \text{ tiene la distribución } t_{n-1}.$$

CAPÍTULO IV

MODELO LINEAL Y FORMAS CUADRÁTICAS

Introducción

En este Capítulo se introduce el modelo lineal, se prueba el TC desde varios puntos de vista equivalentes de modo esencial; así también, se dan otros resultados importantes relacionados con él.

Por sencillez, se escribirá el modelo como

$$y_{ij} = m + t_i + e_{ij} \,,$$

donde y_{ij} es la j-ésima observación sobre el i-ésimo tratamiento (i = 1, 2; j = 1, 2,..., n), m es una gran media, y t_i el efecto del i-ésimo tratamiento. Si se arreglan las observaciones en un vector **Y**, las n primeras observaciones son aquellas sobre el primer tratamiento; entonces,

$$\mathbf{Y} = \mathbf{X}\begin{bmatrix} m \\ t_1 \\ t_2 \end{bmatrix} + \mathbf{e},$$

donde **e** es el vector de errores y **X** es una matriz con tres columnas y 2n = N filas. La primera columna de **X** es $\mathbf{1}_N$, una columna de N elementos unidad que corresponde a la media. La segunda columna corresponde a t_1 y consiste de n unos seguidos por n ceros. La tercera columna tiene n ceros seguidos por n unos. El rango de **X** es 2, ya que la suma de las dos últimas columnas es igual a la primera.

El problema consiste en buscar una vector (m, t_1, t_2) de estimaciones, tal que $S_e = \sum_{i,j}(y_{ij} - y_{ij})^2$ sea un mínimo, donde $y_{ij} = m + t_i$. Si se diferencia S_e con respecto a m, t_1, t_2 , se llega a las tres ecuaciones que siguen

$$\sum y_{ij} = Nm + nt_1 + nt_2, \quad n\bar{y}_1 = \sum y_{1j} = nm + nt,$$

$$n\bar{y}_2 = \sum y_{2j} = nm + nt_2 \quad o \quad \mathbf{X'Y} = \mathbf{X'X}(\mathrm{m}, \mathrm{t}_1, \mathrm{t}_2).$$

No se puede invertir **X'X** porque es singular; sin embargo, $(m + t_1) - (m + t_2) = t_1 - t_2$; al restar la tercera ecuación normal de la segunda, $t_1 - t_2 = \bar{y}_1 - \bar{y}_2$; otra vez, $y_{ij} = \bar{y}_i$.

Ahora, considérense las cuatro formas cuadráticas. Una de ellas es la suma de cuadrados de filas $\sum_i \sum_j y_{ij}^2 = \mathbf{Y'Y} = \mathbf{Y'IY}$. La segunda es $N\bar{y}^2$, donde $\bar{y}$ es la media de las N observaciones, se puede escribir como $\mathbf{Y'J_NY}/N$, donde $\mathbf{J}_N$ denota una matriz cuadrada de orden N con cada elemento la unidad. La identidad

$$\sum_i \sum_j y_{ij}^2 = N\bar{y}^2 + \sum_i \sum_j (y_{ij} - \bar{y})^2 = N\bar{y}^2 + n(\bar{y}_1 - \bar{y}_2)^2/2 + \sum_{i,j} (y_{ij} - \bar{y}_i)^2,$$

se puede escribir

$$\mathbf{Y'Y} = \mathbf{Y'A_1Y} + \mathbf{Y'A_2Y} + \mathbf{Y'A_3Y},$$

donde

$$\mathbf{A}_1 = N^{-1}\,\mathbf{J}_N,$$

$$\mathbf{A}_2 = (2n)^{-1}\begin{bmatrix} \mathbf{J}_n & -\mathbf{J}_n \\ -\mathbf{J}_n & \mathbf{J}_n \end{bmatrix},$$

$$A_3 = \begin{bmatrix} \mathbf{I}_n - n^{-1}\mathbf{J}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n - n^{-n-1}\mathbf{J}_n \end{bmatrix}.$$

Se ve que, $\mathbf{A}_i^2 = \mathbf{A}_i$, por lo que estas tres matrices son idempotentes; lo que implica que las formas cuadráticas, divididas por σ^2, tienen distribuciones χ^2; además, $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_1\mathbf{A}_3 = \mathbf{A}_2\,\mathbf{A}_3 = \mathbf{0}$, lo que significa que las fomas están distribuidas de modo independiente. De aquí que, se pueda probar la hipótesis $m_1 = m_2$, o $t_1 = t_2$ por la estadística $F_o = (\mathbf{Y'A_2Y})/(\mathbf{Y'A_3Y}/(N\text{-}2))$. Nótese que los rangos de las formas son iguales a los números de grados de libertad. Bajo la hipótesis nula esta estadística tiene la distribución F(1, N-2), y la prueba es la misma como la prueba t ordinaria. En efecto, se tiene

$$F_o = \frac{n(\bar{y}_1 - \bar{y}_2)^2}{2} \div \frac{S_e}{N-2} = \frac{n(\bar{y}_1 - \bar{y}_2)^2}{2s^2} = t^2,$$

como se debería esperar.

Cuando se arreglan los datos, u observaciones, de un experimento en un vector $\mathbf{Y}$, el procedimiento del ANOVA comprende separar las SC de las observaciones $\sum y^2$ o $\mathbf{Y'Y}$, en un conjunto de formas cuadráticas

$$\mathbf{Y'Y} = \mathbf{Y'A_1Y} + \mathbf{Y'A_2Y} + \ldots$$

Si las observaciones $\mathbf{Y}$ están de modo normal distribuidas, las pruebas de las hipótesis son hechas por examinar algunas de las razones de estas formas; así se tiene

i) los cumulantes de formas cuadráticas;

ii) $\mathbf{Y'A_iY}$ tiene con los divisores apropiados, una distribución χ^2 con el número de grados de libertad igual al rango de $\mathbf{A}_i$;

iii) las formas son independientes, y de aquí,

iv) las razones correspondientes tienen distribuciones F.

Modelos Lineales

En los modelos lineales de interés, una observación y puede ser representada por

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + e\ ,$$

donde $x_1,\ldots,x_p$ son las coordenadas, $\beta_1,\ldots,\beta_p$ son los coeficientes, y e denota el error aleatorio. Si hay más de un punto de los datos bajo consideración, se puede denotar el i-ésimo punto de éstos por y_i y escribir

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i\ ,$$

donde x_{ij} denota el valor de la j-ésima coordenada en el i-ésimo punto de los datos. Se acostumbra el uso de la notación matricial y se escribe

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e},$$

donde **Y** es un vector de N observaciones, **X** es una matriz de constantes conocidas, β es un vector de p parámetros y **e** es el vector de los errores aleatorios.

La matriz $\mathbf{X} = (x_{ij})$ algunas veces se llama matriz diseño; tiene N filas y p columnas. En algunas situaciones del diseño experimental, x_{ij} toma sólo los valores de cero y la unidad; sin embargo, esta restricción no es necesaria para los resultados de este capítulo. Los elementos del vector de los parámetros β serán de forma usual constantes desconocidas; a veces se consideran variables aleatorias. El propósito principal es obtener estimaciones de estos parámetros β_j, o algunas funciones de ellos, y probar algunas hipótesis sobre los mismos; usualmente, las más comunes es algún subconjunto de los parámetros $\beta_1, \ldots, \beta_p$ son todos ceros.

El modelo se dice lineal porque expresa, y, como una combinación lineal de los parámetros β_j; no hay restricción que el modelo tiene que ser una función lineal de las x_i o que éstas sean independientes; se puede poner el polinomio $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ en esta forma por fijar $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$. Algunos modelos que no son lineales se pueden hacer así por transformaciones; *e.g.*, $y = Ae^{Bx}$ llega a ser $z = \beta_0 + \beta_1 x$ si se toma $z = \ln y$, entonces $\beta_0 = \ln A$ y $\beta_1 = B$. Por otra parte, los dos modelos $y = \beta_1 + 1/(\beta_2+\beta_3 x)$ y $y = Ae^{Bx} + Ce^{Dx}$ no son lineales, y no se pueden linealizar por transformar y.

En el modelo lineal original, la i-ésima observación consta de dos componentes; a saber: $\sum_j x_{ij}\, \beta_j$ y e_i, el "error" en la i-ésima observación. Se hacen las hipótesis siguientes sobre los e_i:

i) Tienen esperanzas cero; o sea, $E(e_i) = 0$ para todo i;

ii) Son incorrelacionados; es decir, $E\left(e_i e_{i'}\right) = 0$ si $i \neq i'$;

iii) Tienen varianza homogénea; esto es, $E\,(e_i^2) = \sigma^2$ para todo i;

iv) El supuesto de normalidad; *i.e.*, que los errores e tienen una distribución normal multivariada con $E(\mathbf{e}) = \mathbf{0}$, $\mathrm{cov}\,(\mathbf{e}) = \mathbf{I}\sigma^2$, lo que será escrito $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$; o de forma equivalente se puede decir que $\mathbf{Y}$ tiene una distribución normal multivariada, $\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{Y}\sigma^2)$.

Estas suposiciones se pueden escribir E(**e**) = **0**, E(**ee'**) = **I**σ^2, y se sigue de la suposición i) que E (**Y**) = **X**β.

En general, si **Z** es un vector de variables aleatorias, E (**Z**) es el vector cuyo i-ésimo elemento es E (z_i). La matriz de varianzas y covarianzas se denotará por cov (**Z**); si cov (**Z**) = **V**σ^2, la covarianza de z_i y z_j es $v_{ij}\ \sigma^2$, a veces se omite el escalar σ^2. Si **A** es una matriz tal que **AZ** existe, se tiene E(**AZ**) = **A**E(**Z**), y cov(**AZ**) = **AVA'**, donde cov (**Z**)=**V**.

La suposición **Y** ~ N(**X**β, **I**σ^2) es importante, cuando se realizan pruebas de significancia o en estimación de intervalos; mas no es esencial para la subdivisión de sumas de cuadrados o para obtener estimaciones de cuadrados mínimos; en efecto, el método de verosimilitud máxima, bajo la hipótesis de normalidad, da las mismas estimaciones de funciones estimables como el método de cuadrados mínimos.

La proposicion siguiente será usada más adelante en la verificación del TC; s.s pruebas alternativas y otros resultados relevantes son dados por Banerjee (1964) –cuya prueba es mejorada por Loynes (1966)– Graybill (1983), Graybill y Marsaglia (1957), Searle (1971) y Zyskind (1975).

Teorema 1. Sean $\mathbf{A}_1,\mathbf{A}_2,\ldots,\mathbf{A}_k$ matrices reales y simétricas de orden n, tales que

$$\mathbf{I} = \sum_{i=1}^{k} \mathbf{A}_i \qquad (3.1)$$

Considérense las cuatro condiciones siguientes:

c_1: cada $\mathbf{A}_i$ es idempotente;

c_2: $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}$, para todos los $i \neq j$;

c_3: $\sum_{i=1}^{k} n_i = n$, donde $n_i = r(\mathbf{A}_i)$;

c_4: Existe una matriz ortogonal $\mathbf{R}$ que diagonaliza, de forma simultánea las matrices dadas, de tal forma que cada matriz diagonal $\mathbf{RA}_i\mathbf{R}'$, $i = 1, 2, ..., k$, tiene r_i elementos diagonales consecutivos iguales a la unidad, con los restantes $(n - r_i)$ elementos diagonales iguales a cero; de una forma necesaria, entonces ningún par de elementos diagonales diferentes de cero resultan en la misma posición de las diversas matrices diagonales resultantes.

Cada una de las cuatro condiciones anteriores es, de una forma simultánea, necesaria y suficiente para que cada una de las tres restantes sea cierta.

Demostración. Hay una matriz ortogonal $\mathbf{P}$ tal que

$$\mathbf{P'IP} = \mathbf{I} = \mathbf{P'A_1P} + \sum_{j=2}^{k}\mathbf{P'A_jP},$$

donde
$$\mathbf{P'A_1P} = \begin{bmatrix}\mathbf{D_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \quad \text{y} \quad \mathbf{P'A_jP} = \begin{bmatrix}\mathbf{B_j} & \mathbf{C_j}\\ \mathbf{C_j'} & \mathbf{E_j}\end{bmatrix},$$

donde $\mathbf{D}_1$ es la matriz diagonal cuyos elementos diagonales son las n_i raíces características diferentes de cero de $\mathbf{A}_1$, y $\mathbf{B}_j$ es una matriz cuadrada de orden n_1; j toma los valores $2 \leq j \leq k$. Supóngase que todas las $\mathbf{A}_i$ son idempotentes; para establecer c_2, es suficiente demostrar que $\mathbf{A}_1\mathbf{A}_j = \mathbf{0}$ para cualesquier j; se tiene $\mathbf{D}_1 = \mathbf{I}$ y $\sum_j \text{tr}\,(\mathbf{B}_j) = 0$;

pero $\mathbf{P'A_j\ P}$ es idempotente, y así los elementos diagonales de $\mathbf{B_j}$ no pueden ser negativos. Por lo tanto, $\mathbf{B_j} = \mathbf{0}$ y $\mathbf{C_j} = \mathbf{0}$ para cada j; se sigue que $\mathbf{A_1 A_j} = \mathbf{P'A_1PP'A_jP} = \mathbf{0}$. Ya que $\mathbf{P'A_jP}$ es idempotente, $\mathbf{E_j}$ también; entonces, $n = tr\ (\mathbf{I_n}) = n_1 + \sum tr(\mathbf{E_j}) = n_1 + \sum n_j$ y c_3 se verifica. Supóngase que $\mathbf{A_i\ A_j} = \mathbf{0}$ para todos los pares i, j donde $i \neq j$; entonces, para cada i, $\mathbf{A_i} = \mathbf{A_iI} = \mathbf{A_i}(\mathbf{A_1} + \ldots + \mathbf{A_i} + \ldots + \mathbf{A_k}) = \mathbf{A}_i^2$. Si c_3 es cierta, se puede demostrar que $\mathbf{A_1}$ es idempotente: se escribe **B, C, E, Q** para $\sum\mathbf{B_j}$, $\sum\mathbf{C_j}$, $\sum\mathbf{E_j}$, $\sum\mathbf{A_j}$; entonces, ya que $\mathbf{P'A_1P} + \mathbf{P'QP} = \mathbf{I}$, se tiene $\mathbf{E} = \mathbf{I}_{n-n_1}$, $\mathbf{C} = \mathbf{0}$; pero $r(\mathbf{P'QP}) = r\ (\mathbf{Q}) = n - n_1 = r(\mathbf{E})$, y así $\mathbf{B} = \mathbf{0}$ y $\mathbf{D_1} = \mathbf{I}$, así que $\mathbf{A_1}$ es idempotente (John, 1971; Zyskind, 1975). Se ha demostrado la equivalencia de las tres primeras condiciones; sólo falta comprobar que una cualquiera de éstas implica c_4, lo cual se realiza en el penúltimo subcapítulo.

Una extensión del Teorema 1, se da en la proposición siguiente, que se demuestra al aplicar: Si **B** es simétrica e idempotente, **Q** es simétrica y definida no negativa, e **I- B-Q** es definida no negativa; entonces $\mathbf{BQ} = \mathbf{QB} = \mathbf{0}$. Este resultado se llama Lema de Loynes. Los detalles se encuentran en Searle (1971).

Corolario 1.1. Si $\mathbf{X_i}$, n×n, simétrica, rango k_i, $i = 1,2,\ldots,p$; $\mathbf{X} = \sum_{i=1}^{p}\mathbf{X_i}$, que es simétrica con rango k. Entonces, de las condiciones:

(a) $\mathbf{X_i}$, es idempotente para todo i;

(b) $\mathbf{X_i X_j} = \mathbf{0}$ para $i \neq j$;

(c) $\mathbf{X}$ es idempotente;

(d) $k = \sum_{i=1}^{p} k_i$;

es cierto que:

I: cualesquier dos de (a), (b) y (c) implican (a), (b), (c) y (d);

II: (c) y (d) implican (a) y (b),

III: (c) y $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{p-1}$ son idempotentes con $\mathbf{X}_p$, que es definida no negativa; implica que $\mathbf{X}_p$ es también idempotente y por tanto (a), (b) y (d).

Distribución de Formas Cuadráticas

Si **x** es un vector de N variables aleatorias, la forma cuadrática **x'Ax** es también una variable aleatoria: La mayoría de los resultados en este subcapítulo, se sostienen sólo cuando **x** tiene la distribución normal multivariada.

El logaritmo de la fgm de **x** es la función generatriz de acumulantes, o semiinvariantes; si la fgm existe, así también la función generatriz de acumulantes. El coeficiente de $t^r/r!$ de la expansión en la serie de Taylor de la función (2.13), es el r-ésimo acumulante de **x**, y se denota por $\mathbf{K}_r$ (**x**), o cuando no hay confusión simplemente por K_r; de forma evidente, K_r es una función de los momentos de **x**. El Teorema que sigue se demuestra en Searle (1971).

Teorema 2. Cuando **x** sea $N(\mu, \mathbf{V})$, entonces

i). $E(\mathbf{x}'\mathbf{A}\mathbf{x}) = \mathrm{tr}\,(\mathbf{AV}) + \mu'\mathbf{A}\,\mu$, (3.2) también verdadera cuando **x** sea no normal.

ii) el r-ésimo acumulante de **x'Ax** es

$$K_r(\mathbf{x'Ax}) = 2^{r-1}(r-1)!\left[\mathrm{tr}(\mathbf{AV})^{\mathbf{r}} + r\mu'\mathbf{A}(\mathbf{VA})^{\mathbf{r-1}}\mu\right]$$

iii) la covarianza de **x** con **x'Ax** es

$$\mathrm{cov}(\mathbf{x}, \mathbf{x'Ax}) = 2\mathbf{VA}\mu$$

Demostración. i) Si $E(\mathbf{x}) = \mu$ y $\mathrm{var}(\mathbf{x}) = \mathbf{V}$, se tiene

$$E(\mathbf{xx'}) = \mathbf{V} + \mu\mu'.$$

De aquí,

$$\begin{aligned} E(\mathbf{x'Ax}) &= E\,\mathrm{tr}(\mathbf{Axx'}) = \mathrm{tr}\left[\mathbf{A}\,E(\mathbf{xx'})\right] \\ &= \mathrm{tr}(\mathbf{AV} + \mathbf{A}\mu\mu') \\ &= \mathrm{tr}(\mathbf{AV}) + \mu'\mathbf{A}\mu. \end{aligned}$$

Una prueba alternativa la da John (1971).

ii) La fgm de **x'Ax** es

$$M_{\mathbf{x'Ax}}(t) = (2\pi)^{-n/2}|\mathbf{V}|^{-1/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[t\mathbf{x'Ax} - \tfrac{1}{2}(\mathbf{x}-\mu)'\mathbf{V}^{-1}(\mathbf{x}-\mu)\right]dx_1\ldots dx_n$$

y al rearreglar el exponente ésta resulta

$$M_{\mathbf{x'Ax}}(t) = \frac{e^{-\frac{1}{2}\mu'\mathbf{V}^{-1}\mu}}{(2\pi)^{n/2}|\mathbf{V}|^{\frac{1}{2}}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\mathbf{x}'(\mathbf{I}-2t\mathbf{AV})\mathbf{V}^{-1}\mathbf{x} + \mu'\mathbf{V}^{-1}\mathbf{x}\right]dx_1\ldots dx_n \quad (3.3)$$

Ahora, si se pone $\mathbf{g}' = \mu'\mathbf{V}^{-1}$ y $\mathbf{W} = \left[(\mathbf{I}-2t\,\mathbf{AV})\mathbf{V}^{-1}\right]^{-1} = \mathbf{V}(\mathbf{I}-2t\mathbf{AV})^{-1}$ en (2.33);

el miembro derecho de ésta se iguala a la integral múltiple de (3.3), y así resulta

$$M_{\mathbf{x'Ax}}(t) = e^{-\frac{1}{2}\mu'\mathbf{V}^{-1}\mu}\,|\mathbf{V}|^{-\frac{1}{2}}\left|\mathbf{V}(\mathbf{I}-2t\mathbf{AV})^{-1}\right|^{\frac{1}{2}}\exp\left[\tfrac{1}{2}\mu'\mathbf{V}^{-1}\mathbf{V}(\mathbf{I}-2t\mathbf{AV})^{-1}\mathbf{V}^{-1}\mu\right]$$

que se simplifica a

$$M_{\mathbf{x'Ax}}(t) = \left|\mathbf{I} - 2t\mathbf{AV}\right|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}\mu'\left[\mathbf{I} - (\mathbf{I} - 2t\mathbf{AV})^{-1}\right]\mathbf{V}^{-1}\mu\right\}. \tag{3.4}$$

Ya que la función generatriz de acumulantes es el logaritmo de la fgm; se tiene

$$\sum_{r=1}^{\infty} K_r t^r / r! = \log\left[M_{x'Ax}(t)\right]$$

$$= -\tfrac{1}{2} log\left|\mathbf{I} - 2t\mathbf{AV}\right| - \tfrac{1}{2}\mu'\left[\mathbf{I} - (\mathbf{I} - 2t\mathbf{AV})^{-1}\right]\mathbf{V}^{-1}\mu. \tag{3.5}$$

Los dos términos de (3.5) se evalúan como sigue: para $|t|$ suficientemente pequeño, "λ_i de $\mathbf{X}$" ≡ "i-ésima raíz característica de $\mathbf{X}$"

$$-\tfrac{1}{2}\log\left|\mathbf{I} - 2t\mathbf{AV}\right| = -\tfrac{1}{2}\sum_{i=1}^{n} log\left[\lambda_i \text{ de } (\mathbf{I} - t\mathbf{AV})\right]$$

$$= -\tfrac{1}{2}\sum_{i=1}^{n} log\left[1 - 2t(\lambda_i \ de\ \mathbf{AV})\right]$$

$$= \tfrac{1}{2}\sum_{i=1}^{n}\sum_{r=1}^{\infty} -\left[2t(\lambda_i \ de\ \mathbf{AV})\right]^r / r$$

$$= \sum_{r=1}^{\infty} 2^{r-1} t^r / r \sum_{i=1}^{n} (\lambda_i \ de\ \mathbf{AV})^r$$

$$= \sum_{r=1}^{\infty} (2^{r-1} t^r / r)\operatorname{tr}(\mathbf{AV})^r.$$

Y, por la expansión binomial directa, para $|t|$ suficientemente pequeña

$$\mathbf{I} - (\mathbf{I} - 2t\mathbf{AV})^{-1} = -\sum_{r=1}^{\infty} 2^r t^r (\mathbf{AV})^r$$

Al sustituir los dos resultados anteriores, en (3.5) e igualar los coeficientes de t^r da

$$K_r\,(\mathbf{x'Ax}) = 2^{r-1}\,(r-1)!\left[tr(\mathbf{AV})^{\mathbf{r}} + r\mu'\mathbf{A}(\mathbf{VA})^{\mathbf{r-1}}\mu\right]. \tag{3.6}$$

Una prueba alternativa se halla en John (1971).

iii) La covarianza entre **x** y **x'Ax** es por definición

$$\text{cov}\,(\mathbf{x}, \mathbf{x'Ax}) = E(\mathbf{x}-\mu)[\mathbf{x'Ax} - E(\mathbf{x'Ax})]$$

$$= E(\mathbf{x}-\mu)[\mathbf{x'Ax} - \mu'\mathbf{A}\mu - tr(\mathbf{AV})]$$

$$= E(\mathbf{x}-\mu)[(\mathbf{x}-\mu)'\mathbf{A}(\mathbf{x}-\mu) + 2(\mathbf{x}-\mu)'\mathbf{A}\mu - tr(\mathbf{AV})\ ]$$

$$= 0 + 2\mathbf{VA}\mu - 0$$

$$= 2\ \mathbf{VA}\mu,$$

ya que el primero y tercer momentos de ($\mathbf{x} - \mu$) son cero. De interés empírico son los tres siguientes casos particulares de este teorema:

Corolario 2.1. Si $\mu = \mathbf{0}$, entonces

$$E(\mathbf{x'Ax}) = \text{tr}\,(\mathbf{AV}),$$

y bajo normalidad

$$K_r\,(\mathbf{x'Ax}) = 2^{r-1}\,(r-1)!\ \text{tr}(\mathbf{AV})^r$$

y

$$\text{cov}\,(\mathbf{x},\mathbf{x'Ax}) = \mathbf{0}$$

Corolario 2.2. Cuando r = 2, resulta la varianza de **x'Ax**; o sea

$$K_2\,(\mathbf{x'Ax}) = 2^{2-1}(2-1)!\left[\text{tr}(\mathbf{AV})^{\mathbf{2}} + 2\mu'\mathbf{A}(\mathbf{VA})^{\mathbf{2-1}}\mu\right]$$

$$= 2\left[tr(\mathbf{AV})^{\mathbf{2}} + 2\mu'\mathbf{AVA}\mu\right]$$

$$= 2\ \mathrm{tr}\ (\mathbf{AV})^2 + 4\ \boldsymbol{\mu}'\mathbf{AVA}\boldsymbol{\mu}$$

$$= \mathrm{v}\ (\mathbf{x'Ax}).$$

De aquí que, $$\mathrm{v}(\mathbf{x'Ax}) = 2\ \mathrm{tr}\ (\mathbf{AV})^2 + 4\ \boldsymbol{\mu}'\mathbf{AVA}\boldsymbol{\mu} \qquad (3.7)$$

Corolario 2.3. Cuando $\mathbf{x} \sim N(\mathbf{0}, \mathbf{V})$, (3.7) llega a ser

$$\mathrm{v}(\mathbf{x'Ax}) = 2\mathrm{tr}(\mathbf{AV})^2.$$

La proposición que sigue, de interés teórico, es una extensión del Teorema 2.

Corolario 2.4.Cuando $\mathbf{x}$ sea $NS(\boldsymbol{\mu}, \mathbf{V})$, entonces ... Los resultados de esta proposición son idénticos a aquellos del Teorema 2. Las pruebas de los incisos i) e iii) son también las mismas. Como en Rohde y Tallis (1969) y Searle (1971), la comprobación de ii) procede como sigue: Cuando $\mathbf{x}$ sea $NS(\boldsymbol{\mu}, \mathbf{V})$ con $\mathbf{V}$ singular, no hay pérdida de generalidad en suponer que $\mathbf{x} = \boldsymbol{\mu} + \mathbf{Ly}$ donde $\mathbf{y}$ es $N(\mathbf{0}, \mathbf{I}_k)$, y $\mathbf{V} = \mathbf{LL'}$ con k como máximo rango columna de $\mathbf{L}$; entonces, la fgm de $\mathbf{x'Ax}$ resulta

$M_{\mathbf{x'Ax}}(t) =$

$$(2\pi)^{-\frac{1}{2}k}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\exp\left(\mathbf{ty'L'ALy} - \tfrac{1}{2}\mathbf{y'y} + 2\mathbf{t}\mu'\mathbf{ALy} + \mathbf{t}\mu'\mathbf{A}\mu\right)dy_1\ldots dy_k$$

y la aplicación de (2.33), reduce ésta a

$$M_{\mathbf{x'Ax}}(t) = \left|\mathbf{I} - 2\mathbf{tL'AL}\right|^{-\frac{1}{2}} exp\left[\mathbf{t}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + 2t^2\boldsymbol{\mu}'\mathbf{AL}(\mathbf{I} - 2t\mathbf{L'AL})^{-1}\mathbf{L'A}\boldsymbol{\mu}\right].$$

Si se usan sumas infinitas para - $\frac{1}{2}\log\left|\mathbf{I} - 2\mathbf{tL'AL}\right|$ y $(\mathbf{I} - 2\mathbf{tL'AL})^{-1}$ análogas a como se utilizaron al derivar (3.6), el r-ésimo semiinvariante llega a ser

$$K_{\mathbf{x}'\mathbf{A}\mathbf{x}}(t) = \sum_{r=1}^{\infty} (2^{r-1} t^r) tr(\mathbf{L}'\mathbf{A}\mathbf{L})^r + t\mu'\mathbf{A}\mu + 2t^2\mu'\mathbf{A}\mathbf{L}\sum_{r=0}^{\infty} 2^r t^r (\mathbf{L}'\mathbf{A}\mathbf{L})^r \mathbf{L}'\mathbf{A}\mathbf{L}$$

$$= t\left[tr(\mathbf{L}'\mathbf{A}\mathbf{L}) + \mu'\mathbf{A}\mu\right] + \sum_{r=2}^{\infty} t^r 2^{r-1}\left[\mu'\mathbf{A}\mathbf{L}(\mathbf{L}'\mathbf{A}\mathbf{L})^{r-2}\mathbf{L}'\mathbf{A}\mu + tr(\mathbf{L}'\mathbf{A}\mathbf{L})^r / r\right]$$

..(3.8)

Luego $\mathbf{V} = \mathbf{LL'}$ y así

$$\text{tr }(\mathbf{L'AL})^r = \text{tr }(\mathbf{VA})^r \text{ para todos los enteros positivos r;} \qquad (3.9)$$

también, por inducción, se puede demostrar que

$$\mathbf{AL}(\mathbf{L'AL})^{r-2}\,\mathbf{L'A} = \mathbf{A}(\mathbf{VA})^{r-1} \;. \qquad (3.10)$$

Por consiguiente,

$$K_{\mathbf{x}'\mathbf{A}\mathbf{x}}(t) = t\left[\mu'\mathbf{A}\mu + tr(\mathbf{AV})\right] + \sum_{r=2}^{\infty} t^r 2^{r-1}\left[\mu'\mathbf{A}(\mathbf{VA})^{r-1}\mu + tr(\mathbf{VA})^r / r\right]$$

$$= \sum_{r=1}^{\infty} t^r 2^{r-1}\left[\mu'\mathbf{A}(\mathbf{VA})^{\mathbf{r}-1}\mu + tr(\mathbf{VA})^r / r\right].$$

Por lo tanto, el r-ésimo acumulante de **x'Ax**, el coeficiente de $t^r/r!$ en $K_{x'Ax}$ (t), es como el dado en (3.6).

La proposición siguiente señala la importancia de las matrices idempotentes, en la teoría de la distribución de formas cuadráticas; se prueba en John (1971) y Searle (1971).

Teorema 3. Si $\mathbf{x} \sim N(\mu, \mathbf{V})$, entonces $\mathbf{x'Ax} \sim \chi^{2'}\left[r(\mathbf{A}), \tfrac{1}{2}\mu'\mathbf{A}\mu\right] \Leftrightarrow \mathbf{AV}$ es idempotente.

Demostración. Desde el punto de vista operacional, la parte más importante de esta proposición es la condición de suficiencia, *i. e.*, que si **AV** es idempotente, entonces **x'Ax** tiene una distribución χ^2 no central; sin embargo, también hay ocasiones cuando la condición de necesidad sea útil.

Suficiencia: Dado que **AV** es idempotente demostrar que :

$$\mathbf{x'Ax} \text{ es } \chi^{2'}\left[r(\mathbf{A}),\tfrac{1}{2}\mu'\mathbf{A}\mu\right].$$

Según (3.4) la fgm de **x'Ax** resulta

$$M_{x'Ax}(t) = \prod_{i=1}^{n}(1-2t\lambda_i)^{-\frac{1}{2}}\; exp\left\{-\frac{1}{2}\mu'\left[-\sum_{\mathbf{k}=1}^{\infty}(2t)^k(\mathbf{AV})^{\mathbf{k}}\right]\mathbf{V}^{-1}\mu\right\}.$$

donde las λ_i para i = 1, 2, ... , n, son las raíces características de **AV**; ahora bien, si ésta es idempotente y r es su rango, r valores de las λ_i son la unidad y n -r son cero; y $(\mathbf{AV})^r = \mathbf{AV}$, así que

$$M_{\mathbf{x'Ax}}(t) = \prod_{i=1}^{r}(1-2t)^{-\frac{1}{2}}exp\left\{-\frac{1}{2}\mu'\left[-\sum_{k=1}^{\infty}(2t)^k\right]\mathbf{AVV}^{-1}\mu\right\}$$

$$= (1-2t)^{-\frac{1}{2}r}exp\left\{-\frac{1}{2}\mu'\left[1-(1-2t)^{-1}\right]\mathbf{A}\mu\right\}$$

$$=(1-2t)^{-\frac{1}{2}r}\;exp\left\{-\frac{1}{2}\mu'\mathbf{A}\mu\left[1-(1-2t)^{-1}\right]\right\} \tag{3.11}$$

Si se comparan (2.38) y (3.11) se nota que **x'Ax** es $\chi^{2'}(r,\frac{1}{2}\mu'\mathbf{A}\mu)$, donde r = r(**AV**)= r(**A**), ya que **V** es no singular. En consecuencia, **x'Ax** es $\chi^{2'}\left[r(\mathbf{A}),\frac{1}{2}\mu'\mathbf{A}\mu\right]$.

Necesidad: Dado que **x'Ax** es $\chi^{2'}$ (r, $\frac{1}{2}\mu'\mathbf{A}\mu$) demostrar que **AV** es idempotente de rango r. Las dos formas (3.4) y (3.11) deben ser iguales –e iguales para todos los valores de μ, en particular para $\mu = \mathbf{0}$– si se remplaza este valor en las dos formas anteriores, y se igualan resulta

$$(1-2t)^{-\frac{r}{2}} = |\mathbf{I} - 2t\mathbf{AV}|^{-1/2}.$$

Si se escribe u por 2t y se rearregla da

$$(1-u)^{r} = |\mathbf{I} - u\mathbf{AV}|.$$

Si $\lambda_1, \lambda_2, \dots, \lambda_n$ son las raíces características de **AV**, entonces se tiene

$$(1-u)^{r} = \prod_{i=1}^{n}(1-u\lambda_i),$$

la cual es una identidad en u, cuyo miembro derecho no tiene potencias de u que excedan r; luego al menos una λ_i es cero; el uso repetido de este razonamiento muestra que (n - r) de las λ_i son cero, y así se puede escribir

$$(1-u)^{r} = \prod_{i=1}^{r}(1-u\lambda_i),$$

al tomar logaritmos de ambos lados e igualar coeficientes, resultan r ecuaciones con r incógnitas λ, *i. e.*, todas las sumas de potencias de las λ iguales a r; este sistema tiene una solución $\lambda_i = 1$ para i = 1, 2 ,... , r; de esta manera n-r raíces características de **AV** son cero y r de éstas son la unidad; *ergo* por el Teorema 65 Capítulo II, **AV** es idempotente, y el teorema está verificado.

En la práctica la importancia de esta proposición es que, proporciona la distribución de las SC atribuibles a ciertas hipótesis planteadas en el modelo, las cuales siempre se pueden representar por formas cuadráticas del tipo $\mathbf{x'Ax}$; asimismo, determina los grados de libertad asociados a cada SC.

Como es natural, dependiendo de los valores de $\boldsymbol{\mu}$, $\mathbf{V}$ y la elección de $\mathbf{A}$, el teorema da una infinidad de corolarios; de interés especial son los siguientes cinco:

Corolario 3.1. Cuando $\mathbf{x}$ sea $N(\mu\mathbf{1}, \sigma^2\mathbf{I})$, entonces $\sum_{i=1}^{n}(x_i - \bar{x})^2/\sigma^2$ es $\chi^2(n\text{-}1,0)$.

Demostración. Considérese $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \mathbf{x'H'_0H_0x}$, donde $\mathbf{H_0}$ son las últimas n-1 filas de la matriz de Helmert de orden n; entonces, $\mathbf{H_0H'_0} = \mathbf{I}$ y $\mathbf{H'_0H_0}$ es idempotente; por consiguiente, si $\mathbf{x}$ es $N(\mu\mathbf{1}, \sigma^2\mathbf{I})$, la proposición dice que $\sum_{i=1}^{n}(x_i - \bar{x})^2/\sigma^2$ es $\chi^2(n, \frac{1}{2}\mu\mathbf{1'H'_0H_0 1}\mu/\sigma^2) = \chi^2(n-1, 0)$ porque $\mathbf{1'H'_0H_0 1} = 0$.

Corolario 3.2. Cuando $\mathbf{x}$ sea $N(\mathbf{0}, \mathbf{I})$, entonces $\mathbf{x'Ax}$ es χ^2_r, sí y sólo sí $\mathbf{A}$ es idempotente de rango r.

Corolario 3.3. Si $\mathbf{x}$ es $N(\mathbf{0}, \mathbf{V})$, entonces $\mathbf{x'Ax}$ es χ^2_r cuando, y sólo cuando, $\mathbf{AV}$ sea idempotente de rango r.

Corolario 3.4. Cuando **x** sea N(μ, σ^2**I**), entonces **x'x**/σ^2 es $\chi^{2'}$ $(n, \frac{1}{2}\mu'\mu / \sigma^2)$.

Corolario 3.5. Cuando **x** sea N(μ, **I**), entonces **x'Ax** es $\chi^{2'}$ $(r, \frac{1}{2}\mu'\mathbf{A}\mu)$ si, y sólo si, **A** es idempotente de rango r.

Como la prueba del Teorema 3, se basó sobre las fgm, la proposición de en seguida se deduce de éste y del inciso ii) del Teorema 2.

Colorario 3.6. Cuando **x'Ax** tenga una distribución χ^2 no central, *i. e.*, si **AV** es idempotente de rango r, entonces el k-ésimo semiinvariante de **x'Ax**, con **A** simétrica, es

$$K_k(\mathbf{x}'\mathbf{A}\mathbf{x}) = 2^{k-1}(\mathrm{k}-1)!\left[r(\mathbf{A}) + \mathrm{k}\mu'\mathbf{A}\mu\right] \tag{3.12}$$

Una extensión de este teorema es de interés: se declara sin prueba, la cual se encuentra en Rao (1966) y Rayner y Livingstone (1965), como un corolario.

Corolario 3.7. Si **x** ~ NS(μ, **V**), entonces la forma no homogénea **x'Ax** + **m'x** + d tiene una distribución χ^2 no central, con grados de libertad tr(**AV**) y parámetro de no centralidad $\frac{1}{2}(\mathbf{A}\mu + \frac{1}{2}\mathbf{m})'\mathbf{V}(\mathbf{A}\mu + \frac{1}{2}\mathbf{m})$ si, sólo si,

(i) $\mathbf{VAVAV} = \mathbf{VAV}$

(ii) $(\mathbf{A}\mu + \frac{1}{2}\mathbf{m})'\mathbf{V} = (\mathbf{A}\mu + \frac{1}{2}\mathbf{m})'\mathbf{VAV}$

(iii) $\mu'\mathbf{A}\mu + \mathbf{m}'\mu + d = (\mathbf{A}\mu + \frac{1}{2}\mathbf{m})'\mathbf{V}(\mathbf{A}\mu + \frac{1}{2}\mathbf{m})$

De las muchas proposiciones que se puede deducir de ésta, se mencionan sólo dos.

Corolario 3.7.1. ($\mathbf{m} = \mathbf{0}$, $d = 0$, $\mu = \mathbf{0}$). Cuando $\mathbf{x}$ sea $N(\mathbf{0}, \mathbf{V})$, $\mathbf{V}$ singular o no, $\mathbf{x'Ax}$ es $\chi^2_{tr(\mathbf{AV})}$, sii $\mathbf{VAVAV} = \mathbf{VAV}$.

Corolario 3.7.2. ($\mathbf{m} = \mathbf{0}$, $d = 0$). Si $\mathbf{x} \sim N(\mu, \mathbf{V})$, $\mathbf{V}$ es singular o no, $\mathbf{x'Ax}$ se distribuye como $\chi^{2'}\left[tr(\mathbf{AV}), \frac{1}{2}\mu'\mathbf{A}\mu\right]$, sii (i) $\mathbf{VAVAV} = \mathbf{VAV}$, (ii) $\mu'\mathbf{AV} = \mu'\mathbf{AVAV}$ e (iii) $\mu'\mathbf{A}\mu = \mu'\mathbf{AVA}\mu$.

El Teorema 3 es un caso especial de este último corolario; se declara así:

Corolario 3.7.3. Si $\mathbf{V}$ es regular las condiciones del Corolario 3.7.2, se reducen a la idempotencia de $\mathbf{AV}$.

El ejemplo subsecuente ilustra, que a pesar, de la condición de idempotencia de este corolario, no se debe concluir en el Teorema 3 o en los otros dos corolarios; *i. e.*, que $\mathbf{AV}$ es idempotente; ya que se sabe por el Teorema 65 Capítulo II, que sólo si $\mathbf{V}$ es regular esta condición implica la idempotencia de $\mathbf{AV}$.

Ejemplo.

$$\text{Sean}\mathbf{V}= \begin{bmatrix} 2 & 0 & -2 \\ 0 & 2 & -2 \\ -2 & -2 & 4 \end{bmatrix} \text{ y } \mathbf{A} = \frac{1}{16}\begin{bmatrix} 16 & 6 & 5 \\ 6 & 4 & 3 \\ 5 & 3 & 2 \end{bmatrix} \Rightarrow \mathbf{VA} = \frac{1}{16}\begin{bmatrix} 22 & 6 & 6 \\ 2 & 2 & 2 \\ -24 & -8 & -8 \end{bmatrix}$$

y $$(\mathbf{VA})^2 = \frac{1}{16}\begin{bmatrix} 22 & 6 & 6 \\ 0 & 0 & 0 \\ -22 & -6 & -6 \end{bmatrix},$$

resultados que indican que **VA** no es idempotente; nótese que la condición i) del Corolario 3.7 sí se satisface.

El resultado subsiguiente se refiere a la independencia de dos formas, una lineal y la otra cuadrática, se asevera bajo las hipótesis de la normalidad; además, **x'Ax** no tiene que tener una distribución χ^2 no central para emplearlo, y no incluye **AVB**, un producto que en lo necesario no existe.

Teorema 4. Cuando $\mathbf{x} \sim N(\mu, \mathbf{V})$, la condición necesaria y suficiente para que **x'Ax** y **Bx** se distribuyan de un modo independiente es $\mathbf{BVA} = \mathbf{0}$.

Demostración. Si se sigue a Searle (1971), se tiene:

Suficiencia: que $\mathbf{BVA} = \mathbf{0} \Rightarrow$ la independencia . Ya que **A** es simétrica, el Teorema 63 Capítulo II da $\mathbf{A} = \mathbf{LL'}$ para alguna **L** de máximo rango columna; luego, si $\mathbf{BVA} = \mathbf{0}$, entonces $\mathbf{BVLL'} = \mathbf{0}$; y puesto que **L** tiene máximo rango columna, $(\mathbf{L'L})^{-1}$ existe y de esta forma

$$\mathbf{BVLL'} = \mathbf{0} \Rightarrow \mathbf{BVLL'}(\mathbf{L'L})^{-1} = \mathbf{0},\ i.\ e.\ ,\ \mathbf{BVL} = \mathbf{0};$$

ergo cov($\mathbf{Bx}$, $\mathbf{x'L}$) = $\mathbf{BVL} = \mathbf{0}$; por lo tanto, porque $\mathbf{x}$ es un vector de variables distribuidas de forma normal, $\mathbf{Bx}$ y $\mathbf{x'L}$ se distribuyen de modo independiente; así que, $\mathbf{Bx}$ y $\mathbf{x'Ax} = \mathbf{x'LL'x}$ se distribuyen de la misma manera.

Necesidad: la independencia de $\mathbf{x'Ax}$ y $\mathbf{Bx} \Rightarrow \mathbf{BVA} = \mathbf{0}$. Cuando dos variables aleatorias se distribuyen de forma independiente su covarianza siempre es cero; de aquí, cov ($\mathbf{Bx}$, $\mathbf{x'Ax}$) = $\mathbf{0}$; y el Teorema 2 iii) da cov ($\mathbf{Bx}$, $\mathbf{x'Ax}$) = $2\mathbf{BVA}\mu$; luego $2\mathbf{BVA}\mu = \mathbf{0}$, y puesto que ésta es cierta para toda μ, $\mathbf{BVA} = \mathbf{0}$, lo que completa la prueba de la necesidad.

Una generalización de esta proposición, al caso normal no singular se da en seguida.

Corolario 4.1. Si $\mathbf{x}$ *es* $NS(\mu, \mathbf{V})$, *entonces* $\mathbf{x'Ax}$ y $\mathbf{Bx}$ *son independientes, sii* $\mathbf{BVAV} = \mathbf{0}$ y $\mathbf{BVA}\mu = \mathbf{0}$.

El resultado posterior concierne a la independencia, de dos formas cuadráticas en variables normales; existen por lo menos seis pruebas en la literatura, la mayoría para $\mu = \mathbf{0}$ y $\mathbf{V} = \mathbf{I}$, son las dadas por Aitken (1950), Craig (1943), Hotelling (1944), Lancaster (1954), Ogawa (1949) y Searle (1971); la validez de la prueba de Craig fue cuestionada por Hotelling, y la de éste por Ogawa; la que aquí se da es la de Searle, por corta y más general, no se basa en las fgm, las que así lo hacen son muy tediosas.

Teorema 5. Si $\mathbf{x}$ es $N(\mu, \mathbf{V})$, entonces las formas $\mathbf{x'Ax}$ y $\mathbf{x'Bx}$ se distribuyen de forma independiente, sii $\mathbf{AVB} = \mathbf{0}$.

Demostración. Cuando $\mathbf{AVB} = \mathbf{0}$, $\Rightarrow$ $\mathbf{BVA} = \mathbf{0}$, ya que éstas son equivalentes, porque $\mathbf{A}$, $\mathbf{B}$ y $\mathbf{V}$ son simétricas.

Suficiencia: que $\mathbf{AVB} = \mathbf{0}$ $\Rightarrow$ la independencia. Del Teorema 63 Capítulo II se puede escribir $\mathbf{A} = \mathbf{LL'}$ y $\mathbf{B} = \mathbf{MM'}$, donde $\mathbf{L}$ y $\mathbf{M}$ tienen cada una máximo rango columna; *ergo* cuando $\mathbf{AVB} = \mathbf{0}$, $\mathbf{LL'VMM'} = \mathbf{0}$, y porque $(\mathbf{LL'})^{-1}$ y $(\mathbf{MM'})^{-1}$ existen, lo que significa que $\mathbf{L'VM} = \mathbf{0}$; de aquí, cov $(\mathbf{L'x}, \mathbf{x'M}) = \mathbf{L'VM} = \mathbf{0}$; así, porque $\mathbf{x}$ es un vector de variables que se distribuyen en forma normal, $\mathbf{L'x}$ y $\mathbf{x'M}$ se distribuyen de modo independiente; por lo tanto, $\mathbf{x'Ax} = \mathbf{x'LL'x}$ y $\mathbf{x'Bx} = \mathbf{x'MM'x}$ se distribuyen de manera independiente.

Necesidad: la independencia $\Rightarrow$ $\mathbf{AVB} = \mathbf{0}$. Si $\mathbf{x'Ax}$ y $\mathbf{x'Bx}$ se distribuyen de forma independiente, cov $(\mathbf{x'Ax}, \mathbf{x'Bx}) = \mathbf{0}$, así que

$$\mathrm{v}\,(\mathbf{x'Ax} + \mathbf{x'Bx}) = \mathrm{v}\,(\mathbf{x'Ax}) + \mathrm{v}\,(\mathbf{x'Bx})$$

i. e.,
$$\mathrm{v}\left[\mathbf{x'}(\mathbf{A} + \mathbf{B})\mathbf{x}\right] = \mathrm{v}(\mathbf{x'Ax}) + \mathrm{v}(\mathbf{x'Bx}).$$

Por la ecuación (3.7) aplicada a estas expresiones, resulta

$$\mathrm{tr}(\mathbf{VAVB}) + 2\mu'\mathbf{AVB}\mu = 0,$$

la cual es cierta para cualquier μ, incluyendo $\mu = \mathbf{0}$, por lo que $\mathrm{tr}(\mathbf{VAVB}) = 0$, y de aquí $2\mu'\mathbf{AVB}\mu = \mathbf{0}$, la que es cierta para toda μ, y de esta forma $\mathbf{AVB} = \mathbf{0}$.

En seguida se da una extensión del Teorema 5.

Corolario 5.1. Si $\mathbf{x} \sim NS(\mu, \mathbf{V})$, entonces las formas $\mathbf{x'Ax}$ y $\mathbf{x'Bx}$ son independientes, sii $\mathbf{VAVBV} = \mathbf{0}$, $\mathbf{VAVB}\mu = \mathbf{VBVA}\mu = \mathbf{0}$ y $\mu'\mathbf{AVB}\mu = \mathbf{0}$.

La proposición subsecuente, se refiere a la independencia de conjuntos de formas cuadráticas; su importancia empírica consiste, en que es fundamental para el análisis de los modelos lineales; puesto que para las pruebas de las hipótesis se generan formas cuadráticas del tipo $\mathbf{x'A_ix}$ que son independientes; de modo que se pueden formar valores de F para probar diferentes hipótesis, lo que en sí mismo justifica la construcción de tablas de ANOVA con descomposición aditiva de los grados de libertad y las SC; sin embargo, también se usa para tablas de ANOVA no aditivas, en las que de hecho se mezclan varios tipos de descomposiciones aditivas de grados de libertad y SC; también, se puede emplear en la descomposición sucesiva de las SC, *i. e.*, que puede haber dos o más niveles de tales descomposiciones.

Teorema 6. Sean

$\mathbf{x}$, de orden nx1, se distribuye como $N(\mu, \mathbf{V})$;

$\mathbf{A}_i$, nxn, simétrica de rango k_i, para $i = 1, 2, \ldots, p$;

$\mathbf{A} = \sum_{i=1}^{p} \mathbf{A_i}$, que es simétrica de rango k.

Entonces:

$$\mathbf{x'A_i\, x} \sim \chi^{2'}\,(k_i,\ \tfrac{1}{2}\mu'\mathbf{A_i}\mu),$$

las $\mathbf{x'Ax}$ son independientes par a par,

y $$\mathbf{x'Ax} \sim \chi^{2'}(k,\ \tfrac{1}{2}\mu'\mathbf{A}\mu).$$

Si, y sólo si:

I. Cualesquiera dos de (*a*) $\mathbf{A_iV}$ es idempotente para todo i,

(b) $\mathbf{A_iVA_j} = \mathbf{0}$ para todo $i < j$,

(c) $\mathbf{AV}$ es idempotente,

son ciertas;

o II. (c) es cierta y (d) $k = \sum_{i=1}^{p} k_i$,

o III. (c) es cierta y (e) $\mathbf{A_1V}, \ldots, \mathbf{A_{(p-1)}V}$ son idempotentes y $\mathbf{A_pV}$ es definida no negativa.

Demostración. Ya que $\mathbf{V}$ es definida positiva y simétrica, $\mathbf{V} = \mathbf{T'T}$ por el Teorema 59 Capítulo II, para alguna $\mathbf{T}$ no singular; entonces, puesto que $\mathbf{A_i}$ es simétrica, así es $\mathbf{TA_iT'}$ y $r(\mathbf{A_i}) = r(\mathbf{TA_iT'})$; y $\mathbf{A_iV}$ es idempotente, sii $\mathbf{TA_iT'}$ lo es; $\mathbf{A_iV_j} = \mathbf{0}$ sii $\mathbf{TA_iT'TA_jT'} = \mathbf{0}$. Entonces el Corolario 1.1 se sostiene cierto usando $\mathbf{TA_iT'}$ en vez de $\mathbf{X_i}$, y $\mathbf{TAT'}$ en vez de $\mathbf{X}$; entonces las partes I, II y III del Corolario 1.1 aplicadas a $\mathbf{TA_iT'}$ y $\mathbf{TAT'}$ demuestran que cuando las secciones I, II o III del Teorema 6 existen, implica que las condiciones (a), (b) y (c) siempre existen; mas por el Teorema 3,

$\mathbf{x'A_ix} \sim \chi^{2'}(k_i, \frac{1}{2}\mu'\mathbf{A_i}\mu)$ sii (a) es cierta; además, $\mathbf{x'Ax} \sim \chi^{2'}(k, \frac{1}{2}\mu'\mathbf{A}\mu)$ sii *(c)* es cierta; y por el Teorema 5, $\mathbf{x'A_ix}$ y $\mathbf{x'A_jx}$ son independientes, sii la condición (b) es cierta.

Si se extiende este teorema al caso normal singular, resulta:

Corolario 6.1. Si

$\mathbf{x}$, de orden nx1, $\sim NS(\boldsymbol{\mu}, \mathbf{V})$;

$\mathbf{A_i}$, nxn, simétrica, $r(\mathbf{VA_iV}) = r_i$, i = 1, 2,..., p.,

y
$$\mathbf{A} = \sum_{i=1}^{p} \mathbf{A_i}, \quad r(\mathbf{VAV}) = r.$$

Si (i)$\mathbf{V}$ es no singular, (ii) $\mathbf{V}$ es singular y $\mu = \mathbf{0}$ o si (iii) $\mathbf{V}$ es singular, μ no es en lo necesario nula y $\mathbf{A_i}$ es semidefinida positiva para i = 1, 2,..., r; entonces las cuatro proposiciones:

(a) $\mathbf{x'A_ix} \sim \chi^{2'}(r_i, \frac{1}{2}\mu'\mathbf{A_i}\mu)$;

(b) las $\mathbf{x'A_ix}$ son de modo mutuo independientes;

(c) $\mathbf{x'A_ix} \sim \chi^{2'}(r, \frac{1}{2}\mu'\mathbf{A}\mu)$

(d) $r = \sum_{i=1}^{p} r_i$

son implicadas por cualesquier dos de (a), (b) y (c) y por (a) y (d).

Teorema de Cochran

Un caso especial del Teorema 6, ocurre cuando $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I_n})$, este importante resultado se llama Teorema de Cochran.

Corolario 6.2 Si $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_n)$ y $\mathbf{A}_i$ es simétrica de rango r_i para i = 1, 2,..., p con

$\mathbf{I}_n = \sum_{i=1}^{p} \mathbf{A}_i$; entonces las $\mathbf{x'A}_i\mathbf{x}$ se distribuyen de forma independiente como $\chi^2_{r_i}$, o χ^2 no central, sii $n = \sum_{i=1}^{p} r_i$.

Demostración. Se pone $\mu = \mathbf{0}$ y $\mathbf{V} = \mathbf{I}_n = \mathbf{A}$ y el Teorema 6, da el resultado de un modo directo.

Otra forma equivalente del Corolario 6.2, la dan Brownlee (1984), Cramér (1970), Graybill (1983), John (1971), Lindgren (1976), Mann (1949), Scheffé (1959) y Zyskind (1975) por ejemplo; Cramér (1970) lo expresa y prueba por inducción como sigue:

Teorema 7: Teorema de Cochran. Si $n = \sum_{1}^{k} r_i$, existe una transformación ortogonal $\mathbf{x} = \mathbf{Cy}$ que cambia cada Q_i en una SC, según las relaciones:

$$Q_1 = \sum_{1}^{r_1} y_i^2, \quad Q_2 = \sum_{r_1+1}^{r_1+r_2} y_i^2, \; \ldots, \; Q_k = \sum_{r_1+\ldots+r_{k-1}+1}^{r_1+\ldots+r_k} y_i^2$$

i. e. , que ningún par de las Q_i tengan una variable y_i en común.

Demostración. Para k = 1, la certeza de la proposición es muy clara; se debe pues, probar que si el teorema es cierto para una descomposición en k - 1 términos, también es cierto para k términos. Para probar esto, primero se aplica una sustitución ortogonal

$\mathbf{x} = \mathbf{C}_1\mathbf{z}$ a 1.61, la cual convierte Q_1 en $\sum_{1}^{r_1} \lambda_i z_i^2$; así, se tiene

$$\sum_{1}^{r_1}(1-\lambda_i)z_i + \sum_{r_1+1}^{n} z_i^2 = Q_2' + \ldots + Q_k', \tag{3.13}$$

donde $Q_2', \ldots, Q_k'$ son las transformadas de $Q_2, \ldots, Q_k$. Ahora nótese que todos los λ_i son iguales a 1. Supóngase que haya p de los λ_i diferentes de 1, mientras el resto sean iguales a 1. Los dos lados de (3.13) son formas cuadráticas en z_i $(i = 1, 2, \ldots, n)$, el rango del primero es $n - r_1 + p$, mientras que el rango del segundo es a lo más $r_2 + \ldots + r_k = n - r_1$; resulta así que $p = 0$ y, por lo tanto, todos los $\lambda_i = 1$, de donde (3.13) se reduce

$$\sum_{r_1+1}^{n} z_i^2 = Q_2' + \ldots + Q_k', \tag{3.14}$$

en la cual las variables z_i $(i = 1, 2, \ldots, r)$ no existen en el lado izquierdo, ni en ningún término del derecho. Para ver esto, por ejemplo si, Q_2' no fuere independiente de z_1, entonces de acuerdo con lo dicho antes, Q_2' debería contener un término cz_1^2, con $c > 0$; mas dado que los coeficientes de z_1^2 en Q_i^2 $(i = 3, 4, \ldots, k)$ son de un modo cierto nulos o positivos, lo que implica una antinomia con (3.14).

Así, pues (3.14) da una representación de $\sum_{r_1+1}^{n} z_i^2$ en una adición de k - 1 formas no negativas de las variables z_i $(i = r_1 + 1, r_1 + 2, \ldots, n)$; según la hipótesis, es cierto el TC para esta descomposición y en consecuencia, existe una transformación ortogonal con n - r_1 variables que sustituye las z_i ($i = r_1 + 1, \dots, n$) por otras nuevas variables y_i ($i = r_1 + 1, r_1 + 2, \dots, n$), y tal que

$$Q_2' = \sum_{r_1+1}^{r_1+r_2} y_i^2, \quad Q_3' = \sum_{r_1+r_2+1}^{r_1+r_2+r_3} y_i^2, \dots, \quad Q_k' = \sum_{n-r_k+1}^{n} y_i^2 . \tag{3.15}$$

Si se agrega a esta transformación las r_1 ecuaciones $z_1 = y_1$, $z_2 = y_2$, ... , $z_{r_1} = y_{r_1}$, se obtiene una transformación ortogonal, de n variables $\mathbf{z} = \mathbf{C_2 y}$, tal que se cumplen las identidades (3.15).

El resultado de realizar de forma sucesiva las transformaciones $\mathbf{x} = \mathbf{C_1 z}$ y $\mathbf{z} = \mathbf{C_2 y}$ será una transformación compuesta $\mathbf{x} = \mathbf{C_1 C_2 y}$, la cual es ortogonal por ser el producto de dos transformaciones ortogonales; esta transformación reúne todas las condiciones requeridas, y queda así probada la proposición.

Las dos observaciones subsecuentes son importantes para un entendimiento más completo del TC.

Escolio 7.1. Si en (1.61) se sabe sólo que cada una de las Q_i es no negativa y que su rango es a lo más r_i, donde $n = \sum_1^k r_i$, se puede asegurar en seguida que Q_i tiene en efecto rango r_i, de manera que se satisfacen las condiciones del TC. En efecto, dado que el rango de una adición de formas cuadráticas es a lo más igual a la suma de los rangos de los sumandos, si r_i es el de Q_i, entonces n $\leq \sum_1^k r_i' \leq \sum_1^k r_i = n$; así, pues, será $\sum r_i' = \sum r_i$, y como es $r_i' \leq r_i$, esto implica que ha de ser $r_i' = r_i$, para todo i .

Escolio 7.2. El TC se verifica si el miembro izquierdo de (1.61), se remplaza por una forma cuadrática Q de cualquier número de variables, la cual, por medio de una sustitución ortogonal, se puede transformar en $\sum_1^n x_i^2$.

Ya se ha demostrado que $\sum n_i = n$, sii cada $\mathbf{A}_i$ es idempotente, y sii $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}$ donde $i \neq j$; mas $\mathbf{A}_i$ idempotente es iquivalente a $\mathbf{y}'\mathbf{A}_i\mathbf{y} \sim \chi^2(n_i)$, y $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}$ equivale a que $\mathbf{y}'\mathbf{A}_i\mathbf{y}$ y $\mathbf{y}'\mathbf{A}_j\mathbf{y}$ sean independientes. Ahora se puede reparafrasear el TC así: sean $\mathbf{y} \sim N(\mu, \mathbf{I})$ y $\mathbf{y}'\mathbf{y} = \sum \mathbf{y}'\mathbf{A}_i\mathbf{y}$ donde $r(\mathbf{A}_i) = n_i$; entonces cualquiera de las cinco condiciones siguientes implica las otras cuatro:

(i) $\sum n_i = n$;

(ii) cada $\mathbf{A}_i$ es idempotente;

(iii) $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}$ para todos los pares i, j donde $i \neq j$; (3.16)

(iv) $\mathbf{y}'\mathbf{A}_i\mathbf{y} \sim \chi^2(n_i)$ para cada i;

(v) $\mathbf{y}'\mathbf{A}_i\mathbf{y}$, $\mathbf{y}'\mathbf{A}_j\mathbf{y}$ son independientes para todo i, j donde $i \neq j$.

Cada una de las formas $\mathbf{y'A_iy}$ se puede escribir como $\mathbf{y'B_i'B_iy}$, es una matriz de n_i filas y n columnas tal que $\mathbf{B_iB_i'} = \mathbf{I_{n_i}}$, *i. e.*, tal que las columnas de $\mathbf{B_i'}$ forman un conjunto ortonormal de n_i vectores; se sigue del hecho $\mathbf{A_iA_j} = \mathbf{0}$ que tambien $\mathbf{B_iB_j'} = \mathbf{0}$, por lo que se tiene $\mathbf{A_iA_j} = \mathbf{B_i'B_iB_j'B_j} = \mathbf{0}$, y al multiplicar a la izquierda por $\mathbf{B_i}$ y a la derecha por $\mathbf{B_j'}$, $\mathbf{IB_iB_j'I} = \mathbf{B_iB_j'} = 0$. Sea $\mathbf{B}$ la matriz definida por $\mathbf{B'} = (\mathbf{B_1'}, \mathbf{B_2'}, \ldots)$; entonces $\mathbf{B}$ es una matriz ortogonal, y la transformación $\mathbf{z} = \mathbf{By}$ es una transformación ortogonal si se toma $\mathbf{y'y}$ en vez de $\mathbf{z'z}$.

Según Scheffé (1959) es notable que la condición $\sum n_j = n$, también asegure que cada una de las formas cuadráticas Q_i es positiva, y tenga todos sus valores característicos iguales a 0 ó 1; ya que es equivalente bajo la transformación ortogonal a una suma de n_j de las $\{z_i\}$.

En el Teorema 1, se dejó sin demostrar que c_4 era equivalente a una cualquiera de las tres primeras condiciones; el método de probar –*e. g.*, que c_1 equivale a c_4– es análogo a como se probó el TC; los pormenores algebraicos se hallan en Zyskind (1975); además, agrega las dos notas que siguen:

Escolio 7.3. El hecho de que la condición de rango c_3, implica la posibilidad de una diagonalización simultánea, c_4 es la esencia de la proposición conocida entre los estadísticos como el TC.

Escolio 7.4. En el enunciado del TC se supone con frecuencia, por ejemplo Cramér en su edición de 1951, que las matrices $\mathbf{A}_i$ son positivas semidefinidas; empero, ya que c_3 implica c_1, y puesto que todas las matrices simétricas idempotentes son positivas semidefinidas, esta hipótesis se ve que es, de hecho, superflua.

Kempthorne (1973) y Mann (1949), dan la proposición subsiguiente como un corolario del TC.

Corolario 7.1. Sean $x_1, \ldots, x_n$ variables distribuidas normal e independientemente con medias 0 y varianza σ^2, y sean Q_i $(i = 1, 2, \ldots, s)$ s formas cuadráticas en x_i $(i = 1, 2, \ldots, n)$ con rangos n_i $(i = 1, 2, \ldots, s)$ y

$$Q_1 + Q_2 + \ldots + Q_s = Q = \sum_{i=1}^{n} x_i^2 ,$$

entonces $(n_j/n_i)(Q_i/Q_j)$ tiene la distribución F con n_i y n_j grados de libertad, respectivamente.

Si se sigue a Lindgren (1976), el TC es una clase de recíproco del teorema de la adición para las distribuciones χ^2; *i. e.*, de la propiedad reproductiva de las χ^2 no centrales; su demostración –excepto detalles algebraicos– es la misma que la dada por otros autores como Brownlee (1984), Cramér (1970), Graybill (1983), Hogg y Craig (1970), Mann (1949) y Scheffé (1959), para mencionar unos pocos.

Scheffé (1959) deriva el TC como un corolario de la proposición subsecuente.

Teorema 8. Supóngase que

$$\sum_{i=1}^{n} y_i^2 = Q_1 + Q_2 + \ldots + Q_s$$

donde Q_j ($j = 1, 2, \ldots, s$) es una forma cuadrática en las variables y_i ($i = 1, 2, \ldots, n$) de rango n_j. Una condición suficiente y necesaria que exista una transformación ortogonal $\mathbf{z} = \mathbf{A}\mathbf{y}$ del vector $\mathbf{y} = (y_1. y_2, \ldots, y_n)$ a un vector $\mathbf{z} = (z_1, z_2, \ldots, z_n)$, tal que

$$Q_1 = \sum_{i=1}^{n_1} z_i^2, \quad Q_2 = \sum_{n_1+1}^{n_1+n_2} z_i^2, \ \ldots, \quad Q_s = \sum_{n_1+\ldots+n_{s-1}+1}^{n_1+\ldots+n_s} z_i^2$$

es

$$n_1 + n_2 + \ldots + n_s = n.$$

Para probar esta proposición, Scheffé (1959) sigue en forma esencial a Cochran (1934). El corolario siguiente a los dos teoremas mencionados arriba –el Teorema 8 y el TC– es útil para establecer las relaciones de ortogonalidad, si se usa el método del TC, para la teoría de la distribución de las SC componentes en una partición de la SC total.

Corolario 8.1. Supóngase que las hipótesis del TC se satisfacen y $\sum_1^s n_i = n$; además, que cada Q_j se escribe en cualquier forma como una SC de formas lineales $\{L_{jt}\}$ en las $\{y_i\}$,

$$Q_j = \sum_{t=1}^{\tau_j} L_{jt}^2 \qquad (j = 1, 2, \dots, s).$$

Entonces, para $j \neq j'$ y toda t, t' ($t = 1, 2, \dots, \tau_j$; $t' = 1, 2, \dots, \tau_{j'}$), las formas L_{jt} y $L_{j't'}$ son ortogonales.

Otra consecuencia importante es:

Corolario 8.2. Sean $\sum_1^n y_i^2 = Q_1 + Q_2 + \dots + Q_s$, donde Q_j es una forma cuadrática de rango $\leq m_j$ ($j = 1, 2, \dots, s$), y $m_1 + m_2 + \dots + m_s = n$; entonces, rango $Q_j = m_j$.

Ilustración del Teorema de Cochran

Los dos lemas subsiguientes se usarán en la ejemplificación del TC.

Lema A. El rango de una suma de formas cuadráticas es $\leq$ a la adición de sus rangos.

Demostración. Bastará probar que si $\mathbf{A}_1$ y $\mathbf{A}_2$ son matrices del mismo tamaño con rango $\mathbf{A}_i = r_i$; entonces, $(\mathbf{A}_1 + \mathbf{A}_2) \leq r_1 + r_2$. Para cada $\mathbf{A}_i$ elegir una base de r_i vectores para el espacio vectorial generado por las columnas; entonces, ya que las columnas de $\mathbf{A}_1 + \mathbf{A}_2$ son las adiciones de las columnas correspondientes de $\mathbf{A}_1$ y $\mathbf{A}_2$, éstas son combinaciones lineales de los $r_1 + r_2$ vectores en las dos bases; en consecuencia, el número de columnas linealmente independientes en $\mathbf{A}_1 + \mathbf{A}_2$ no puede exceder a $r_1 + r_2$.

Lema B. Sea Q una forma cuadrática en las x_i ($i = 1, 2, \ldots ,n$), que se puede expresar como una forma cuadrática en las z_i ($i= 1, 2, \ldots , p$), donde las z_i son formas lineales en las x_i; entonces, rango $Q \leq p$.

Demostración. Supóngase que $Q = \mathbf{x}'\mathbf{A}^{nxn}\mathbf{x} = \mathbf{z}'\mathbf{B}^{pxp}\mathbf{z}$ y $\mathbf{z} = \mathbf{C}^{pxn}\mathbf{x}$, donde $\mathbf{A}$ y $\mathbf{B}$ son simétricas; entonces, $Q = \mathbf{x}'\mathbf{C}'\mathbf{B}\mathbf{C}\mathbf{x}$ implica que $\mathbf{A} = \mathbf{C}'\mathbf{B}\mathbf{C}$, $r(Q) = r(\mathbf{A}) \leq r(\mathbf{C})$, porque rango $\mathbf{AB} \leq$ min (rango $\mathbf{A}$, rango $\mathbf{B}$); y rango $\mathbf{C} \leq p$, ya que $\mathbf{C}$ es pxn.

Para ilustrar el uso del TC, considérese la distribución de dos clasificaciones con una observación por celda, caso muy frecuente en la práctica; para obtener pruebas exactas e intervalos de confianza relativos a efectos principales es en lo general necesario con el modelo de efectos fijos, mas no el de efectos aleatorios o el mixto, suponer que no hay interacciones; lo que implica que las medias de celda verdaderas tienen la estructura $\eta_{ij} = \mu + \alpha_i + \beta_j$, donde $\alpha. = \beta. = 0$; si se agrega a ésta las hipótesis de normalidad, independencia estadística e igualdad de varianzas de celda, y denotándose por y_{ij} la observación en la i, j celda, se tiene

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij},$$

$$\alpha. = \beta. = 0,$$

$\{e_{ij}\}$ son de forma independiente $N(0, \sigma^2)$.

Las hipótesis de interés principal son

H_A: todas las $\alpha_i = 0$ y H_B: todas las $\beta_j = 0$.

La hipótesis H_A dice que las medias $\{\mu + \alpha_i\}$ para los diferentes niveles de A son todas iguales; en otros términos, que los diferentes niveles de A todos tienen el mismo efecto; de forma análoga para H_B; los rangos de los subíndices son respectivamente $i = 1, \dots, I$; $j = 1, \dots, J$. Por aplicar la teoría general se llega a las SC:

$$SC_A = J\sum_i (y_{i.} - y_{..})^2,$$

$$SC_B = I\sum_j (y_{.j} - y_{..})^2,$$

$$SC_e = \sum_i \sum_j (y_{ij} - y_{i.} - y_{.j} + y_{..})^2$$

También, resultan estas expresiones, por cavilar que es una estadística atrayente de forma intuitiva, para medir las diferencias entre filas, etc.; la SC_e podría ser sugerida por la identidad

$$\sum\sum y_{ij}^2 = IJy_{..}^2 + SC_A + SC_B + SC_e, \tag{3.17}$$

la cual se puede verificar de forma directa, por preservar los paréntesis, elevar al cuadrado y sumar la expresión

$$y_{ij} = y_{..} + (y_{i.} - y_{..}) + (y_{.j} - y_{..}) + (y_{ij} - y_{i.} - y_{.j} + y_{..}).$$

Para aplicar el TC a (3.17), se necesitan los rangos de las cuatro formas cuadráticas del lado derecho; para esto se utiliza el Lema A.

Ahora bien, la SC_A es una forma cuadrática en las $\{z_i = y_{i.} - y_{..}\}$, las cuales son formas lineales *en las observaciones* $\{y_{ij}\}$; *en efecto*, $SC_A = J\sum_1^I z_i^2$; *ya que* $\sum_1^I z_i = 0$, se puede remplazar $z_I = -\sum_{1}^{I-1} z_i$ en SC_A para expresarla como una forma cuadrática en las I - 1 formas lineales $\{z_1, z_2, \ldots, z_{I-1}\}$; luego rango $SC_A \leq I - 1$ por el Lema B; asimismo, rango $SC_B \leq J - 1$. De forma similar, la SC_e es de rango $\leq (I - 1)(J-1)$, porque es una forma cuadrática en las IJ formas lineales $\{z_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..}\}$, donde $\sum_i z_{ij} = 0$, $\sum_j z_{ij} = 0$; *ergo* pueden expresarse como una forma cuadrática en las $(I-1)(J-1)$ de las $\{z_{ij}\}$ con $i < I$, $j < J$, ya que $z_{iJ} = -\sum_{j=1}^{J-1} z_{ij}$ para $i < I$, $z_{Ij} = -\sum_{i=1}^{I-1} z_{ij}$ para $j < J$, y $z_{IJ} = -\sum_{j=1}^{J-1} z_{Ij} = \sum_{j=1}^{J-1}\sum_{i=1}^{I-1} z_{ij}$. Finalmente, el rango de $IJy_{..}$ es ≤ 1 por el Lema B. Ahora bien, se sigue por el Corolario 8.2 que el rango $(IJy_{..}) = 1$, rango $SC_A = I - 1$, rango $SC_B = J - 1$ y rango $SC_e = (I- 1)(J - 1)$.

Si se reescribe (3.17) como

$$\sum_i \sum_j (y_{ij} / \sigma)^2 = \sigma^{-2} IJ y_{..}^2 + \sigma^{-2} SC_A + \sigma^{-2} SC_B + \sigma^{-2} SCe,$$

y se le aplica el TC; el cual dice que las cuatro SC en el miembro derecho tienen distribuciones χ^2 no centrales independientes, con los números y valores familiares de grados de libertad y parámetros de no centralidad, respectivamente; además, el Corolario 8.1 produce la propiedad de ortogonalidad familiar, que toda forma lineal en cualesquiera de las cuatro SC, tal como $y_{i.} - y_{..}$, es ortogonal a toda forma lineal en cualesquier de las otras SC, tal como $y_{i'j} - y_{i'.} - y_{.j} + y_{..}$.

En conclusión, en la revisión bibliográfica no se encontró ninguna otra demostración del TC, diferente a la dada por su autor (Cochran, 1934); todas las comprobaciones son iguales –excepto los detalles algebraicos– en un caso se probó el TC, como un corolario de otro más general (Searle, 1971); en otro caso, se verificó por inducción (Cramér, 1970); a veces a éste último, se le califica de moderno, *e. g.*, en Brownlee (1984); lo que significa que es diferente a la prueba original de 1934; mas el mismo Cramér (1970) reconoce que se basó en ésta. Con respecto a las demostraciones de Cochran (1934), para el caso central, y de Madow (1940), para el caso no central; se puede considerar la última como una generalización de la primera; ya que cuando las Q_j ($j = 1, 2, \ldots, s$) tienen distribuciones χ^2 centrales, la prueba del segundo, se reduce a la del primero. También, se concluye que el TC tiene varios equivalentes; esto se ve muy claro en la condición c_4 del Teorema 1 y las 5 condiciones (3.16), las cuales son equivalentes; en el sentido que cualesquiera de ellas, implica a todas las demás.

LITERATURA CITADA

Aitken, A. C. 1950."On the Statistical Independence of Quadratic Forms in Normal Variates". *Biometrika,* **37**, pp. 93-96.

Anderson, T W. 1958. An introduction to Multivariate Statistical Analysis. Wiley, New York.

Ayres, F. J. 1983. Teoría y Problemas de Matrices. McGraw-Hill, Serie Shaum, México.

Banerjee, K. S. 1964. "A Note on Idempotent Matrices". *Ann. Math. Statist.,* **35**, pp. 880-882.

Borovkov, A. A. 1988. Estadística Matemática. Editorial Mir Moscú, URSS.

Brownlee, K. A. 1984. Statistical Theory and Methodology in Science and Engineering. Robert E. Krieger Publishing Company, Inc., Malabar, Florida, USA.

Cochran, W. G. 1934. "The Distribution of Quadratic Forms in a Normal System with Applications to the Analysis of Covarance" *Proceedings of the Cambridge Philosophical Society, Vol.* **30**, pp. 178-191.

Craig, A. T. 1943. "Note on the Independence of Certain Quadratic Forms". *Ann. Math. Statist.,* **14**, pp. 195-197.

Cramér, H. 1970. Métodos Matemáticos de Estadística. Aguilar, S.A., de Ediciones, Madrid.

Chakrabarti, M.C. 1962. Mathematics of Design and Analysis of Experiments. Asia Publishing House, New Delhi, India.

Dudewicz, E. J. and S. N. Mishra. 1988. Modern Mathematical Statistics. Wiley, New York.

Fisher, R. A. 1918. "The Correlation Between Relatives on the Supposition of Mendelian Inheritance". *Trans. Roy. Soc. Edinburgh, Vol.* **52**, pp. 399-433.

———— 1922. "The Goodness of Fit and Regression Formulae, and the Distribution of Regression Coefficients". *Journal of the Royal Statistical Society,* **85**, pp. 597-612.

———— 1925. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

———— 1935. The Design of Experiments. Oliver & Boyd, Edinburgh.

Graybill, F. A. 1976. Theory and Application of the Linear Model. Wadsworth Publishing Company, Pacific Grove, California, USA.

———— 1983. Matrices with Applications in Statistics. Wadsworth & Brooks/Cole, Pacific Grove, California, USA.

———— and G. Marsaglia. 1957. "Idempotent Matrices and Quadratic Forms in the General Linear Hypothesis". *Ann. Math. Statist.,* **28**, pp. 678-686.

Guenther, W. C. 1964. Analysis of Variance. Prentice-Hall, Inc., Englewood, N. J., USA.

Hadley, G. 1969. Linear Algebra. Fondo Educativo Interamericano, S. A., Bogotá.

Hoffman, K. y R. Kunze. 1992. Álgebra Lineal. Editorial Melo, S. A., México.

Hogg, R. V. and A. T. Craig. 1970. Introduction to Mathematical Statistics. Macmillan Publishing Co., Inc., New York.

Hotelling, H. 1944. "On a Matrix Theorem of A. T. Craig". *Ann. Math. Statist.,* **15**, pp. 427-429.

Irwin, J. O. 1931. "Mathematical Theorems Involved in the Analysis Variance", *Journal of Royal Statistical Society,* **94**, pp. 284-300.

———— 1934. "On the Independence of the Constituent Items in the Analysis of Variance". *Supplement to the Journal of the Royal Statistical Society,* **1,** pp. 236-252

John, P. W. M. 1971. Statistical Design and Analysis of Experiments. The Macmillan Company, New York.

Kempthorne, O. 1973. The Design and Analysis of Experiments. Robert E. Krieger Publishing Company, Huntington, N. Y.

Kleiman, A., y E. K. de Kleiman. 1987. Matrices. Aplicaciones Matemáticas en Economía y Administración. Editorial Limusa, México.

Koroliuk, V. S. 1981. Manual de la Teoría de las Probabilidades y Estadística Matemática. Editorial Mir Moscú, URSS.

Kurosch, A. G. 1975. Curso de Álgebra Superior. Editorial Mir Moscú, URSS.

Lancaster, H. O. 1954. "Traces and Cumulants of Quadratic Forms in Normal Variables". *J. Roy. Statist. Soc., Ser. B.,* **16**, pp. 247-254.

———— 1965. The Helmert Matrices. *Am. Math. Monthly,* **72,** pp. 4-12.

Lindgren, B. W. 1976. Statistical Theory. Collier-Macmillan Limited, London.

Lindman, H. R. 1992. Analysis of Variance in Experimental Design. Springer-Verlag New York, Inc., USA.

Lipschutz, S. 1971. Teoría y Problemas de Álgebra Lineal. McGraw-Hill, Serie Schaum, México.

Loève, M. 1976. Teoría de la Probabilidad. Editorial Tecnos, S. A., Madrid.

Loynes, R. M. 1966. "On Idempotent Matrices". *Ann. Math. Statist.*, **37**, pp. 295-296.

Madow, W. G. 1940. "The Distribution of Quadratic Forms in Non-Central Normal Random Variables", *Annals Mathematical Statistics,* Vol. **11**, pp. 100-103.

Máltsev, A. I. 1972. Fundamentos de Álgebra Lineal. Editorial Mir Moscú, URSS.

Mann, H. B. 1949. Analysis and Design of Experiments. Dover Publications, Inc., New York.

Martínez G., A. y A. Castillo M. 1987. La Teoría de la Regresión con Aplicaciones Agronómicas. Colegio de Postgraduados, Chapingo, Edo. de México.

Méndez R., I. 1976. Modelos Estadísticos Lineales. Interpretación y Aplicaciones. foccavi/conacyt, México.

Mood, A. M. y F. A. Graybill. 1972. Introducción a la Teoría de la Estadística. Aguilar, S. A. de Ediciones, Madrid.

—————— and D. C. Boes. 1974. Introduction to the Theory of Statistics. McGraw-Hill, Inc., New York.

Nering, E. D. 1977. Álgebra Lineal y Teoría de Matrices. Editorial Limusa, México.

Ogawa, J. 1949. "On the Independence of Bilinear and Quadratic Forms of a Random Sample from a Normal Population". *Ann. Inst. Statist. Math. Tokyo,* **1**, pp. 83-108.

Proskuriakov, I. 1986. Problemas de Álgebra Lineal. Editorial Mir Moscú, URSS.

Rao, C. R. 1966. "Generalized Inverse for Matrices and its Applications in Mathematical Statistics". *Research Papers in Statistics.* Festschrift for J. Neyman, Ed., F. N. David. Wiley, New York.

Rayner, A. A., and D. Livingstone. 1965. "On the Distribution of Quadratic Forms in Singular Normal Variables". *S. African J. Agric. Sci.,* **8**, pp. 357-370.

Rohde, C. A., and G. M. Tallis. 1969. "Exact First- and Second-Order Moments of Estimates of Components of Covariance" . *Biometrika,* **56,** pp. 517-526

Scheffé, H. 1959. The Analysis of Variance. Wiley, New York.

Searle, S. R. 1966. Matrix Algebra for the Biological Sciences . Wiley, New York.

———— 1971. Linear Models. Wiley, New York

Spiegel, M. R. 1994. Teoría de Probabilidad y Estadística. McGraw-Hill, Serie Schaum, México.

Wilks, S. S. 1962. Mathematical Statistics. Wiley, New York.

Zyskind, G. 1975. Teoría General de las Hipótesis Lineales. Colegio de Postgraduados, Chapingo, México.