03061
2ej.
6

UNIVERSIDAD NACIONAL AUTONOMA

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Unidad Académica de los Ciclos Profesional y Posgrado del Colegio de Ciencias y Humanidades

UNA TEORIA GENERAL PARA EL ANALISIS DE MODELOS LINEALES MULTIVARIADOS DE ANALISIS DE VARIANZA

T E S I S

Que para obtener el título de

Maestro en Estadística e Investigacion de Operaciones

P r e s e n t a

José Rodolfo Mendoza Blanco

Mexico, D. F. Abril 1989

UNA TEORIA GENERAL PARA EL ANALISIS DE MODELOS LINEALES MULTIVARIADOS DE ANALISIS DE VARIANZA

PREFACIO

Prefacio

En años recientes el desarrollo de técnicas para el análisis estadístico de modelos lineales multivariados ha tenido un fuerte impulso. Potthoff & Roy (1964) propusieron un modelo lineal general de análisis de varianza multivariado que comprende los modelos de regresión lineal multivariada, modelos multivariados de diseños experimentales y algunos modelos de crecimiento entre otros, mostrando de esta manera que dichos modelos no son sino casos particulares de un modelo general. De esta manera aparece atractivo el desarrollar las técnicas de análisis estadístico para este modelo general, permitiendo así un análisis unificado de los modelos lineales que históricamente han sido presentados por separado.

Varios autores han considerado el problema de contrastar hipótesis generales sobre parámetros de modelos multivariados de análisis de varianza, entre los que se pueden mencionar Bargmann & Posten (1964), Glesser & Olkin (1970), Grizzle & Allen (1969), Kabe (1975), Lee (1974), Mudholkar, Davidson & Subbaiah (1974), Potthoff & Roy (1964), Rao (1959), Rao (1966), Smith, Gnanadesikan & Hughes (1962), Srivastava (1966) y Tubbs, Lewis & Duran (1975) entre otros. En este trabajo se señalan incovenientes en algunos métodos de análisis propuestos, se presenta una propuesta de análisis estadístico del modelo general que involucra como casos particulares a los modelos de rango incompleto, proponiéndose una manera de identificar las hipótesis contrastables en este modelo.

El propósito de este estudio es, por una parte, derivar de manera exacta el cociente de verosimilitudes generalizado asociado a la prueba de una hipótesis lineal general en el modelo generalizado de Análisis de Varianza Multivariado de Potthoff y Roy y por otra derivar la distribución exacta de la estadística de prueba asociada al contraste.

El contenido de los capítulos está distribuído de la manera siguiente:

En el capítulo 1 se sitúa el "Análisis de Modelos Lineales" como una de las diferentes áreas de estudio del ánalisis multivariado, justificándose el uso de la distribución Normal en el análisis de este tipo de modelos. Se exponen también en este capítulo los propósitos generales de este trabajo.

El capítulo 2 tiene como propósito la formulación de un modelo generalizado de análisis de varianza multivariado, la proposición de una hipótesis lineal general sobre los parámetros del modelo y la presentación de algunas contribuciones previas al análisis de este tipo de modelos.

En el capítulo 3 se formulan y demuestran los teoremas necesarios tanto para la obtención de la estadística de prueba de cociente de verosimilitudes asociada al contraste de interés en este trabajo, como para la determinación de la distribución asociada a esta estadística de prueba. Cabe señalar que en la determinación de la distribución asociada a la estadística de prueba, sólo se utiliza el teorema 3.17; sin embargo, para la demostración de este teorema se requirió de la elaboración de los teoremas 3.1 a 3.16.

En el capítulo 4 se define el concepto de matrices estimables que resulta de utilidad en la formulación de hipótesis en modelos de rango incompleto. Se formula también una hipótesis lineal general sobre los parámetros estimables del modelo y se determina el cociente de verosimilitudes asociado a esta hipótesis bajo el modelo de análisis de varianza multivariado de Potthoff & Roy.

El capítulo 5 tiene como objeto determinar, mediante la aplicación del teorema 17 del capítulo 3, la distribución de la estadística de prueba asociada al contraste de la hipótesis lineal general definida en el capítulo 4.

En el capítulo 6 se presentan resúmenes en los que se expone por separado el análisis del modelo generalizado de Pothoff & Roy dependiendo del rango de las matrices que definen el modelo.

En el capítulo 7 se desarrollan como casos particulares del modelo de Potthoff & Roy algunos resultados de uso común en modelos de análisis de regresión, modelos de crecimiento y modelos de diseño experimental.

Se presentan por último en el capítulo 8 las conclusiones de este trabajo, así como la proposición de algunas lineas de investigación que pueden seguirse de este material.

Como puede observarse, el material de este trabajo se ha limitado a un nivel teórico dado que no se han incluído ejemplos numéricos; sin embargo, se describe detalladamente el proceso que debe seguirse en una situación práctica.

Se ha hecho un esfuerzo por mantener en las demostraciones de los teoremas procedimientos matemáticos simples; sin embargo, la lectura del material supone un buen conocimiento del álgebra de matrices, conocimientos básicos de estadística matemática y análisis multivariado. También se ha hecho un esfuerzo por incluir en este trabajo la demostración de todos los resultados que se utilizan, dando de esta manera autocontenido al material. Este trabajo puede ser utilizado como material de apoyo en cursos de estadística tanto a nivel licenciatura como posgrado.

El método utilizado en este trabajo para realizar el contraste hace uso extensivo de matrices particionadas, por lo que con objeto de facilitar la lectura se han incluído en la mayoría de los casos, escritas como subíndices, las dimensiones de las matrices. La mayor parte de los resultados sobre álgebra de matrices utilizados en este trabajo se incluyen en el ápendice como referencia. Las referencias a los teoremas que aparecen en el ápendice se anteponen mediante la letra A. Probablemente resulte de utilidad al lector revisar primero el ápendice con objeto de aclarar cuestiones relacionadas con la notación utilizada en este trabajo.

INDICE

INDICE

1

INTRODUCCION

1 Introducción

De la misma manera que ocurre con muchas otras disciplinas, la estadística tiene una estrecha relación con lo que genéricamente se conoce como computación. Una característica común, aunque no siempre presente, en un problema estadístico es el grado de complejidad en la estructura de los datos, no solamente por su volumen sino también por las interrelaciones que comúnmente involucran y la complejidad de los cálculos asociados a una aplicación determinada. Esta situación hace atractiva la aparición del uso de la computadora en el análisis de este tipo de datos. Sin embargo, no solamente las aplicaciones a colecciones complejas de datos requieren de un soporte computacional adecuado. El natural desarrollo de las técnicas estadísticas ha tenido como consecuencia la aparición de procedimientos cada vez más eficaces y confiables que, en contrapartida, aún para conjuntos de datos relativamente pequeños y simples involucran un nivel de dificultad numérica que solamente se puede abordar con los instrumentos apropiados de cálculo.

La consecuencia más importante del proceso de desarrollo de las computadoras en el ámbito estadístico, es que cada vez un mayor número de investigadores contemplan en sus proyectos la implementación de las técnicas que desarrollan, sin considerar la dificultad numérica como un obstáculo insalvable. De hecho, existen técnicas de implementación relativamente reciente que permanecieron sin aplicación durante un período importante de tiempo después de haber sido propuestas debido a la falta de investigadores con esta inclinación.

El análisis estadístico multivariado es un área de la estadística cuyo desarrollo ha estado estrechamente ligado al uso de las computadoras, debido a la complejidad de los cálculos involucrados en una particular aplicación de cualesquiera de las técnicas que lo componen. El análisis multivariado trata con el análisis de observaciones en más de una variable, las cuales en general presentan un esquema de asociación. Sin embargo y dado que las técnicas estadísticas desarrolladas en esta área suponen una dimensión general p de las observaciones, es posible resolver problemas de aplicación univariada substituyendo $p=1$ en estos procedimientos. Existen diferentes técnicas del análisis multivariado que responden a diferentes necesidades de los usuarios, entre las cuales se pueden mencionar: Econometría, Análisis de Componentes Principales, Análisis de Factores, Análisis de Correlación Canónica, Análisis Discriminante, Escalamiento Multidimensional, Análisis de Conglomerados, Análisis de Regresión, Análisis de Diseños Experimentales y Modelos de Crecimiento.

El rápido avance de la estadística multivariada en los últimos años ha repercutido en la continua aparición de nuevos textos en el área, cuyo contenido principal se centra en las técnicas mencionadas anteriormente y más aún, el amplio desarrollo que en la actualidad presenta cada una de estas áreas, en ocasiones se traduce en la aparición de libros especializados en una sola de estas áreas.

El contenido de este trabajo se centra en el estudio de los modelos lineales o modelos de análisis de varianza. Existen diferentes conceptos de lo que es un modelo lineal o un modelo de análisis de varianza. En un número importante de textos, se asocian los modelos de análisis de varianza con el análisis de modelos de diseño experimental, en los que la característica común reside en que existe una matriz diseño asociada a las observaciones provenientes del experimento, cuyas entradas son ceros y unos y resulta de rango incompleto. Ejemplos de estos modelos son el modelo completamente al azar con un criterio de clasificación y el modelo en bloques al azar. En otros textos, el análisis de modelos que tienen asociada una matriz diseño de rango incompleto aparece con el título simple de modelos lineales. Con este último título algunos textos asocian los modelos que tienen una matriz diseño de rango completo, mientras que en otros textos estos modelos aparecen con el nombre de modelos de regresión. Por último, existe otra categoría importante de modelos lineales que se utilizan para modelar curvas de crecimiento y que aparecen con el nombre de modelos de crecimiento.

Dado que los modelos de análisis de regresión, los modelos de diseño experimental y algunos modelos de crecimiento comparten la característica de que el valor esperado de cualquier observación es una forma lineal en los parámetros desconocidos del modelo, se agruparán éstos indistintamente con los nombres de modelos lineales o modelos de análisis de varianza.

En algunos textos se sugiere la posibilidad de analizar modelos de diseño experimental utilizando las técnicas desarrolladas para el análisis de modelos de rango completo, mediante una reparametrización, ya que las técnicas de análisis para este último tipo de modelos brindan facilidad en el estudio del modelo, debido a la simplificación algebráica en varios de los cálculos involucrados y en los que incluso puede hacerse un diagnóstico del modelo, a través de los residuos, situación que no resulta común en el uso de los modelos de diseño experimental. Sin embargo y en general, la elección y/o conveniencia de una particular reparametrización no siempre resulta clara, por lo que a la fecha este esquema de análisis no resulta comúnmente utilizado. Como se discutirá más adelante, puede definirse de manera simple lo que es una reparametrización en el caso multivariado, a partir de la cual el análisis del modelo puede manejarse con relativa facilidad utilizando instrumentos apropiados de cálculo.

Un área de particular importancia en la estadística multivariada está constituida por los modelos lineales multivariados. Tradicionalmente tres tipos de modelos lineales han recibido una amplia atención en el ámbito estadístico, éstos son los modelos de regresión lineal multivariada, modelos multivariados de diseño experimental y los modelos de crecimiento. Estas clases de modelos que históricamente han sido presentadas por separado pueden plantearse como casos particulares de una forma canónica de un modelo simple, a partir del cual problemas como son la estimación de parámetros y prueba de hipótesis puedan ser fácilmente manejados. La proposición de esta forma canónica puede consultarse en Potthoff & Roy (1964). Cabe señalar que el desarrollo aislado que han tenido estas clases de modelos, ha repercutido incluso en que los paquetes de cómputo estadístico que existen actualmente contienen rutinas separadas de análisis para cada uno de estos modelos, siendo que comparten una serie de características comunes.

En relación a los contrastes de hipótesis sobre los parámetros de distribuciones multivariadas pueden señalarse dos problemas comunes. El primero se deriva del extenso número de hipótesis que pueden ser formuladas sobre los parámetros de la distribución y el segundo se encuentra relacionado con la selección de entre varias estadísticas de prueba asociadas a un mismo constraste de hipótesis.

La distribución Normal Multivariada ha jugado un papel fundamental en el desarrollo de diferentes técnicas del Análisis Multivariado. Este hecho se debe por una parte, a que muchos fenómenos producen observaciones cuyo comportamiento es similar al que se esperaría de una variable aleatoria Normal (Unimodalidad, Simetría, etc.). Por otra parte, el modelo Normal resulta atractivo desde un punto de vista técnico, por que su empleo, a diferencia de lo que ocurre con otras distribuciones, generalmente requiere de procedimientos matemáticos simples.

En relación al número de hipótesis que pueden ser formuladas sobre los parámetros de esta distribución, Mardia et al (1979, pag 120) señalan que en los $p(p+3)/2$ parámetros de la distribución p-variada de esta familia, existen $2^{p(p+3)/2}$ distintas hipótesis que especifican valores para diferentes subconjuntos de los parámetros. Este número resulta elevado aún para valores de p tan pequeños como 3. Por supuesto, en la práctica, el número de hipótesis en las que el investigador resulta interesado resulta ser mucho menor.

Respecto a la selección de una estadística de prueba para una hipótesis particular, Plackett (1946) señala que, en general, la dificultad no se encuentra en la derivación de estadísticas de prueba, sino en la búsqueda de la distribución asociada a éstas cuando la hipótesis nula es verdadera y en seleccionar el mejor criterio. El método de análisis estadístico asociado a la prueba de la hipótesis lineal general bajo el modelo general de análisis de varianza multivariado resulta confuso, ya que existe una serie de estadísticas de prueba propuestas en la literatura, sobre las cuales no resulta claro cual es la más conveniente y aún más, no se conoce la distribución de algunas de las estadísticas de prueba propuestas para realizar el contraste.

De esta manera resulta atractivo, por una parte, desarrollar una hipótesis lineal general a partir de la cual diversos contrastes de interés puedan ser planteados como casos particulares y, por otro lado, derivar un criterio de prueba con un método comúnmente empleado como es el cociente de verosimilitudes generalizado, el cual poseé una serie de propiedades que hacen atractiva su utilización. Cabe mencionar que en la derivación de estadísticas de prueba en el análisis multivariado existen dos enfoques generales de uso común. El primero se basa en la prueba de cociente de verosimilitudes generalizado y el segundo en la prueba asociada al principio de unión e intersección (Mardia et al, 1979). En algunas ocasiones, los dos enfoques conducen a la misma estadística de prueba pero en otras producen diferentes criterios de decisión.

Dado lo anterior, los propósitos de este trabajo son proponer una manera de identificar y manejar las matrices estimables, presentar una derivación de la prueba de cociente de verosimilitudes generalizado de una hipótesis lineal general, sobre parámetros estimables, bajo el modelo general de análisis de varianza multivariado de Potthoff y Roy (1964) y derivar por último la distribución exacta de la estadística de prueba obtenida mediante este procedimiento.

2

ANTECEDENTES DEL ANALISIS ESTADISTICO DE MODELOS DE ANALISIS DE VARIANZA

2 Antecedentes

A partir de la proposición de un modelo generalizado de Análisis de Varianza Multivariado sugerido por Potthoff & Roy (1964) varios autores han realizado trabajos relacionados con el análisis de este modelo, entre los cuales pueden mencionarse Bargmann & Posten (1964), Fujikoshi (1974), Geisser (1980), Gleser & Olkin (1970), Grizzle & Allen (1969), Kabe (1975), Kenward (1986), Khatri (1966), Krishnaiah (1969), Lee J.(1982), Lee Y. (1974), Mudholkar, Davidson & Subbaiah (1974), Potthoff & Roy (1964), Rao (1959, 1965, 1966), Smith, Gnanadesikan & Hughes (1962), Srivastava (1966), Timm (1980) y Tubbs, Lewis & Duran (1975).

Esta sección tiene el propósito de exponer un breve resumen sobre la formulación de modelos de Análisis de Varianza Multivariado planteando una hipótesis lineal general. Se presentan también algunas de las estadísticas de prueba que han sido propuestas para realizar el contraste.

2.1. Formulación de un Modelo Usual de Análisis de Varianza Multivariado

Como señalan Potthoff y Roy (1964), el modelo usual de análisis de varianza multivariado es

$$E(X_{nxp})=D_{nxg}\,\mu_{gxp} \tag{2.1}$$

donde $r(D)=r\leq g\leq n$, los n renglones de X tienen distribución independiente y cada renglón sigue una distribución Normal p-variada con matriz de covarianzas Σ y vectores de medias definidos por la relación (2.1). La matriz D es una matriz de constantes conocidas y μ es una matriz de parámetros desconocidos. Como puede observarse, este modelo incluye como casos particulares a los modelos multivariados de Análisis de Regresión y Diseños Experimentales.

2.2. Formulación de un Modelo General de Análisis de Varianza Multivariado

El modelo usual de Análisis de Varianza Multivariado (2.1) puede ser generalizado mediante la adición de una matriz post-multiplicando el valor esperado de la matriz X y su definición es la siguiente:

Sea X_{nxp} una matriz cuyos vectores renglón son independientes y siguen una distribución Normal con matriz de covarianzas Σ y vectores de medias definidos por la relación

$$E(X_{nxp})=D_{nxg}\,\mu_{gxq}\,E_{qxp}\,, \tag{2.2}$$

donde $r(D)=r\leq g\leq n$ y $r(E)=s\leq q\leq p$. Las matrices D y E son matrices de constantes conocidas y μ es una matriz de parámetros desconocidos.

Como puede observarse, el modelo (2.1) se obtiene como caso particular de (2.2) cuando $q=p$ y $E=I_p$. Una de las utilidades de esta generalización, radica en que cierto tipo de modelos de crecimiento pueden ser analizados con este modelo y no pueden ser planteados como casos particulares del modelo usual de Análisis de Varianza Multivariado (Potthoff y Roy, 1964).

2.3. Formulación de la Hipótesis Lineal General

Bajo el modelo (2.1), una hipótesis lineal general sobre la matriz μ de parámetros desconocidos del modelo está dada por

$$H_0: A_{txg}\, \mu_{gxp} B_{pxu} = C_{txu} \,, \tag{2.3}$$

donde A, B y C son matrices de constantes conocidas tales que

$$r(A)= t\leq r \;,\quad r(B)= u\leq p \quad y \quad u\leq n-r \,. \tag{2.4}$$

Bajo el modelo (2.2), una hipótesis lineal general sobre la matriz de parámetros desconocidos del modelo puede definirse como

$$H_0: A_{txg}\, \mu_{gxq} B_{qxu} = C_{txu} \,, \tag{2.5}$$

donde A, B y C son matrices de constantes conocidas tales que

$$r(A)= t\leq r \;,\quad r(B)= u\leq s \quad y \quad u\leq n-r-p+s \,. \tag{2.6}$$

Las hipótesis (2.3) y (2.5) no pueden ser contrastadas contra sus correspondientes negaciones para cualesquier matrices A y B. Este hecho, como se discutirá mas adelante, depende de la relación que guardan las matrices A y B con las matrices que definen el valor esperado de la matriz X. Esta situación es análoga a la que se presenta en los modelos de diseños experimentales, en los cuales sólo se pueden probar hipótesis sobre funciones lineales de los parámetros linealmente estimables (Searle, 1971; Montgomery, 1976).

2.4. Contribuciones Previas

En esta sección se presenta un breve resúmen sobre los métodos de análisis propuestos para realizar el constraste de la hipótesis lineal general bajo los modelos usual y general de Análisis de Varianza Multivariado.

2.4.1. Contribuciones al Análisis del Modelo Usual de Análisis de Varianza Multivariado

Las contribuciones previas al análisis del modelo (2.1), como puede consultarse en Potthoff y Roy (1964), estan asociadas a la hipótesis (2.3) en el caso particular C=0 y pueden resumirse de la manera siguiente:

Sea $D_{nxg} = (D_{1_{nxr}}, D_{2_{nx(g-r)}})$ donde D_1 es de rango r; es decir, se supone que las columnas de D estan ordenadas de tal forma que las primeras r son linealmente independientes. Esta suposición no es restrictiva, ya que si las primeras r columnas de D no son linealmente independientes, se pueden reordenar los términos del modelo, de tal forma que D_1 sea de rango completo.

Defínase también

$$A_{txg} = (A_{1_{txr}}\ A_{2_{tx(g-r)}}).$$

La hipótesis (2.3) en el caso particular C=0 ha sido considerada por Potthoff y Roy (1964). Estos autores señalan que para que la hipótesis pueda ser contrastada se debe tener

$$A_2 = A_1(D'_1D_1)^{-1}D'_1D_2\ , \tag{2.7}$$

de donde se sigue que $r(A_1) = r(A) = t$. Esta condición, como se discutirá mas adelante, garantiza que la estimación máximo verosímil de la matriz de parámetros AμB es única, ya que corresponde a una condición de estimabilidad cuya definición se formaliza en la sección 4.1.1. Sean las matrices S_h debida a la hipótesis y S_e debida al error definidas por

$$S_h = B'X'D_1(D'_1D_1)^{-1}A'_1\left[A_1(D'_1D_1)^{-1}A'_1\right]^{-1}A_1(D'_1D_1)^{-1}D'_1XB$$

y

$$S_e = B'X'\left[I - D_1(D'_1D_1)^{-1}D'_1\right]XB$$

Las 3 pruebas propuestas de la hipótesis (2.3) en el caso C=0, bajo el modelo (2.1) son:

a) Prueba de Roy. La estadística de prueba está dada por el valor propio mayor de $S_hS_e^{-1}$.

b) Prueba de Suma de Valores Propios. Esta prueba constituye una extensión de un concepto desarrollado por Hotelling (1951) y produce una prueba basada en la traza de $S_h\ S_e^{-1}$.

c) Criterio Λ. Esta prueba es una generalización de un concepto desarrollado por Wilks (1932) y utiliza como estadística de prueba a

$$\Lambda = \frac{|S_e|}{|S_e + S_h|} .$$

Cabe señalar en este punto algunos comentarios en relación a las suposiciones realizadas por Potthoff y Roy respecto al rango de la matriz D. Consideran también las restricciones $r \leq g$ y $r < n$, eliminando en las condiciones (2.4) la restricción $u \leq n-r$. Un análisis de la matriz S_e revela que su rango coincide con $k=\min(u,n-r)$ y dado que las condiciones $r \leq g$ y $r<n$ no implican que la matriz S_e tenga rango u, se tiene por tanto que su inversa no siempre existe por lo que en estos casos las estadísticas de prueba no resultan estar bien definidas. Naturalmente, al agregar la condición $n-r \geq u$ se tiene $k=u$ y por tanto la matriz S_e resulta ser de rango completo. Al respecto, Smith, Gnanadesikan y Hughes (1962) consideran la restricción $n-r \geq p$ en lugar de la restricción $n-r \geq u$ señalada en (2.4). Utilizando esta restricción se deduce que el rango de S_e es u, ya que $n-r \geq p \geq u$; sin embargo, la condición $n-r \geq u$ es suficiente y menos restrictiva que la condición $n-r \geq p$, por lo cual se adoptó en este trabajo. A su vez, el rango de la matriz S_h coincide con $l=\min(r(A),r(B))=\min(t,u)$ y dado que para cualesquier hipótesis se tiene $l \geq 1$, la matriz S_h posee al menos un valor propio distinto de cero. Cabe señalar también que para $l=1$ los tres criterios coinciden, pero para $l \geq 2$ los criterios son distintos y pueden llevar a diferentes conclusiones. La forma de las regiones de rechazo para cada prueba, así como información adicional sobre tablas de cuantiles asociados a estos criterios puede consultarse en Smith, Gnanadesikan y Hughes (1962).

Potthoff y Roy (1964) señalan que se conoce muy poco sobre las potencias relativas de estas pruebas, a excepción de que ninguna es uniformemente mejor que las otras. Señalan también que la prueba de Roy tiene la ventaja de que la distribución de la estadística de prueba bajo la hipótesis nula se conoce y se encuentra tabulada (Heck, 1960). Puede demostrarse que las matrices S_h y S_e son independientes con distribuciones Wishart, por lo que la variable Λ posee distribución Λ de Wilks. Actualmente existen tablas para la extensión del criterio de Hotelling y la distribución Λ de Wilks (Kress, 1983), por lo que puede decirse que en este sentido, estos criterios no se encuentran a la fecha en desventaja respecto a la prueba de Roy.

En relación a los tres criterios de prueba definidos anteriormente, Smith, Gnanadesikan y Hughes (1962) señalan que los tres criterios de prueba más comunes coinciden con los tres anteriores con la siguiente modificación al inciso **a)**

a) Prueba de Roy: La estadística de prueba está dada por el valor propio mayor de $S_h(S_e+S_h)^{-1}$.

Denotando por λ el valor propio mayor de $S_h S_e^{-1}$ y por θ el valor propio mayor de $S_h(S_e+S_h)^{-1}$ se sigue que $\theta=\lambda(1+\lambda)^{-1}$ de donde se deduce que ambos procedimientos de prueba son equivalentes.

2.4.2. Contribuciones al Análisis del Modelo General de Análisis de Varianza Multivariado

Potthoff y Roy (1964) señalan que el análisis de este modelo puede realizarse como sigue:

En el modelo general (2.2), suponen E de rango completo e indican que esta suposición no es restrictiva, ya que si E no es de rango completo, es posible reescribir el modelo redefiniendo μ y E, de tal forma que la nueva μ sea de rango completo. Partiendo de esta suposición, sea H definida por $H=G^{-1}E'(EG^{-1}E')^{-1}$ donde G es cualquier matriz de rango completo y sea $Y_{n\times q}$ una matriz aleatoria, dada por Y=XH. La matriz Y satisface la ecuación

$$E(Y_{n\times q})= D_{n\times g}\,\mu_{g\times q}\,.$$

Dado que este modelo es análogo al modelo usual definido en (2.1) se puede utilizar el análisis descrito en la sección anterior para este tipo de modelos. Cuando p=q, la matriz Y es igual a XE^{-1} y no existe necesidad de escoger la matriz G. Cuando q<p, la selección de la matriz G afecta la potencia de las pruebas, el ancho de los intervalos de confianza y la varianza de los estimadores. Potthoff y Roy sugieren respecto a la selección de la matriz G que cuando Σ sea conocida se seleccione $G=\Sigma$ y cuando Σ sea desconocida se reemplaze la matriz G por un estimador de la matriz Σ basado en información independiente a X.

Utilizando el método sugerido por Potthoff y Roy, el estimador de la matriz μ puede derivarse del Análisis de Regresión multivariado y está dado por

$$\begin{aligned}\hat{\mu}&=(D'D)^{-1}D'Y\\ &=(D'D)^{-1}D'XG^{-1}E'(EG^{-1}E')^{-1}\,. \end{aligned}\tag{2.8}$$

Respecto al proceso de estimación sugerido por Potthoff y Roy cabe señalar que cuando Σ es conocida y se selecciona $G=\Sigma$, el estimador $\hat{\mu}$ coincide con el mejor estimador lineal insesgado de μ. Este hecho puede demostrarse utilizando el teorema de Gauss-Markov y considerando particiones de las matrices X y μ definidas por

$$X_{n\times p}=\begin{bmatrix}X'_1\\ \vdots\\ X'_n\end{bmatrix} \qquad y \qquad \mu_{g\times q}=\begin{bmatrix}\mu'_1\\ \vdots\\ \mu'_g\end{bmatrix}.$$

Definiendo el operador v sobre una matriz A_{mxn} de manera que A^v sea un vector columna que resulta de apilar los m renglones de A se obtiene que

$$X^v_{npx1} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \qquad y \qquad \mu^v_{gqx1} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_g \end{bmatrix},$$

de donde se sigue que

$$E(X^v) = (D \otimes E')\mu^* , \qquad V(X^v) = I_n \otimes \Sigma$$

y por tanto la matriz $Y=(I_n \otimes \Sigma^{-\frac{1}{2}})X^v$ es tal que

$$E(Y^v) = (I_n \otimes \Sigma^{-\frac{1}{2}})(D \otimes E')\mu^* = (D \otimes \Sigma^{-\frac{1}{2}} E')\mu^v$$

y

$$V(Y^v) = (I_n \otimes \Sigma^{-\frac{1}{2}})(I_n \otimes \Sigma)(I_n \otimes \Sigma^{-\frac{1}{2}}) = I_{np}.$$

Por una aplicación del teorema de Gauss Markov, el estimador lineal insesgado de varianza mínima del vector μ^v está dado por

$$\hat{\mu}^v = [(D \otimes \Sigma^{-\frac{1}{2}} E')'(D \otimes \Sigma^{-\frac{1}{2}} E')]^{-1}(D \otimes \Sigma^{-\frac{1}{2}} E')'Y$$

$$= [D'D \otimes E\Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} E']^{-1}(D' \otimes E\Sigma^{-\frac{1}{2}})(I \otimes \Sigma^{-\frac{1}{2}})X^v$$

$$= [D'D \otimes E\Sigma^{-1}E']^{-1}(D' \otimes E\Sigma^{-1})X^v$$

$$= [(D'D)^{-1} \otimes (E\Sigma^{-1}E')^{-1}][D' \otimes E\Sigma^{-1}]X^v$$

$$= \{[(D'D)^{-1}D'] \otimes [(E\Sigma^{-1}E')^{-1}E\Sigma^{-1}]\}X^v$$

$$= [(D'D)^{-1}D'X\Sigma^{-1}E'(E\Sigma^{-1}E')^{-1}]^v ,$$

de donde se deduce que

$$\hat{\mu} = (D'D)^{-1}D'X\Sigma^{-1}E'(E\Sigma^{-1}E')^{-1} \tag{2.9}$$

y de aquí que la alternativa natural en la selección de la matriz G, en el estimador de μ definido en (2.8), sea Σ cuando ésta es conocida y una estimación de ésta en base a información independiente cuando sea desconocida.

Rao (1966) señala que el procedimiento de Potthoff y Roy es insatisfactorio por dos razones. Primero, la matriz G es arbitraria y segundo, la matriz de observaciones X_{nxp} se reduce a una matriz Y_{nxq} y si $q<p$, existe una pérdida de información a menos que la matriz de covarianzas Σ sea conocida y se seleccione $G=\Sigma$. Respecto a la elección de la matriz G, sugerida por Potthoff y Roy, Rao señala que también esta sugerencia está sujeta a crítica.

Respecto a la suposición de que la matriz E sea de rango completo hecha por Potthoff y Roy, en efecto, si E no es de rango q, es posible reparametrizar el modelo obteniendo matrices μ^*_{gxt} y E^*_{txp} tales que $E(X)=D\mu^*E^*$ y $r(E^*)=t$. Cabe señalar en este punto, que en la reducción del modelo general (2.3) a un modelo de la forma $E(Y)=D_{nxg}\ \mu_{gxq}$ propuesta por Potthoff y Roy se puede, de manera análoga a la suposición hecha para la matriz E, suponer que D es una matriz de rango completo, ya que si D no es de rango completo, es posible determinar D^*_{nxu} y μ^*_{uxq} donde $r(D^*)=u$, de manera que $E(Y)=D^*\mu^*$ y emplear técnicas de Análisis de Regresión Multivariado (Mardia et al, 1979) para realizar inferencias sobre μ^*. En modelos de diseños experimentales el empleo de esta técnica se conoce como reparametrización del modelo.

Por otra parte, y en relación al procedimiento de análisis sugerido por Potthoff y Roy, no resulta clara la posibilidad de realizar un contraste particular que involucre a la matriz μ y el procedimiento a seguir en caso de ser una hipótesis contrastable.

La elección o conveniencia de una particular reparametrización del modelo respecto a la matriz D ó E no es clara, por lo que resulta conveniente trabajar el modelo sin la suposiciones de que D y E son matrices de rango completo y determinar de manera general que tipo de hipótesis son las que se pueden contrastar.

Otra técnica de análisis del modelo general de análisis de varianza se debe a Rao (1966) y consiste en lo siguiente:

Sea $s=r(E)$, $H_{pxp}=(H_1,H_2)$ una matriz no singular tal que $EH_2=0$ y las s columnas de H_1 forman una base del espacio generado por los renglones E. Postmultiplicando ambos lados de (2.2) por H se tiene

$$E(XH_1)=D\mu EH_1 \quad \text{y} \quad E(XH_2)=0.$$

EL rango de EH_1 es s, de manera que μEH_1 puede ser substituida por la matriz μ^*_{gxs} de parámetros independientes, por lo que el modelo (2.2) puede escribirse como

$$E(Y)=D\mu^* \quad \text{y} \quad E(Z)=0,$$

donde $Y=XH_1$ y $Z=XH_2$. La esperanza condicional de Y dada Z está dada por

$$E(Y|Z)= D\mu+Z\beta ,$$

donde β es una matriz de parámetros de regresión desconocidos. Con esta transformación del modelo, Rao (1966) propone un análisis condicional sobre hipótesis de la forma $L(\mu^*,\beta)=0$. Cuando el rango de E es q, Rao señala que puede seleccionarse $H_1=G^{-1}E'(EG^{-1}E')^{-1}$ y H_2 tal que $EH_2=0$, donde G es cualquier matriz positiva definida. El modelo condicional en este caso resulta ser

$$E(Y|Z)=D\mu+Z\beta \tag{2.8}$$

en el que los parámetros originales son retenidos. Rao señala que en el método de análisis de Potthoff y Roy la información contenida en Z es ignorada.

Un estudio realizado por Lee (1974) revela que el estimador de μ obtenido a partir del modelo condicional (2.10) tomando G=I está dado por

$$\hat{\mu}=(D'D)^{-1}D'XS^{-1}E'(ES^{-1}E')^{-1} ,$$

donde S es la matriz de residuales de los datos analizados; es decir,

$$S=X'(I-D(D'D)^{-1}D')X .$$

El estimador anterior coincide con el estimador de Potthoff y Roy (2.6) substituyendo G=S y con el estimador (2.9) substituyendo Σ=S, por lo que Lee concluye que el modelo condicional de Rao es escencialmente el mismo que el de Potthoff y Roy para propósitos de estimación.

El inconveniente del procedimiento de Rao radica en que las inferencias resultan obtenidas de un modelo condicional y en que de nuevo, no resulta clara la posibilidad de realizar un contraste particular que involucre a los parámetros originales en la matriz μ y el procedimiento de prueba en caso de ser la hipótesis contrastable.

Por otro lado, también resulta de interés determinar la posibilidad de contrastar hipótesis mas generales que las propuestas por Potthoff & Roy (1964) y Rao (1966) que sean de la forma (2.5), bajo el modelo general de Análisis de Varianza Multivariado.

Otra técnica de análisis, basada en el criterio del cociente de verosimilitudes, es propuesta por Gleser y Olkin (1970) y puede resumirse de la siguiente manera: Sean D y E en el modelo (2.2), matrices de rango completo. Si D y E no son matrices de rango completo, siempre es posible reparametrizar el modelo de manera que las nuevas matrices D y E sean de rango completo. De los teoremas A.19 y A.20 se obtiene que las matrices D y E pueden escribirse como

$$D_{n\times g}=\Gamma_{1_{n\times n}}\begin{pmatrix} I_g \\ 0 \end{pmatrix}T_{1_{g\times g}} \quad \text{y} \quad E_{q\times p}=T_{2_{q\times q}}(I_q,0)\,\Gamma_{2_{p\times p}},$$

donde Γ_1 y Γ_2 son matrices ortogonales y T_1 y T_2 son matrices de rango completo. Sea $X^*=\Gamma_1' X\Gamma_2$ y sea una partición de X^* definida por

$$X^*_{n\times p}=\begin{pmatrix} X^*_{1\,g\times p} \\ X^*_{2\,(n-g)\times p} \end{pmatrix}.$$

Utilizando los teoremas A.20 y A.21 puede escribirse

$$A_{t\times g}\,T^{-1}_{1_{g\times g}}=T_{3_{t\times t}}(I_t,0)\,\Gamma_{3_{g\times g}} \quad \text{y} \quad T^{-1}_{2_{q\times q}}B_{q\times u}=\Gamma_{4_{q\times q}}\begin{pmatrix} 0 \\ I_u \end{pmatrix}T_{4_{u\times u}},$$

donde Γ_3 y Γ_4 son matrices ortogonales y T_3 y T_4 son matrices de rango completo. Utilizando estas descomposiciones se definen las matrices

$$Z_{g\times p}=\Gamma_{3_{g\times g}}X^*_{1\,g\times p}\begin{pmatrix} \Gamma_{4_{q\times q}} & 0 \\ 0 & I_{p-q} \end{pmatrix}$$

y

$$V_{p\times p}=\begin{pmatrix} \Gamma'_{4_{q\times q}} & 0 \\ 0 & I_{p-q} \end{pmatrix}S_{p\times p}\begin{pmatrix} \Gamma_{4_{q\times q}} & 0 \\ 0 & I_{p-q} \end{pmatrix},$$

donde $S=X_2^{*\prime}X_2^*$. Particionando las matrices Z y V como

$$Z_{g\times p}=\begin{pmatrix} Z_{11_{t\times(q-u)}} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22_{(g-t)\times u}} & Z_{23} \end{pmatrix}$$

y

$$V_{p\times p}=\begin{bmatrix} V_{11_{(q-u)\times(q-u)}} & V_{12} & V_{13} \\ V_{21} & V_{22_{u\times u}} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{bmatrix},$$

el cociente de verosimilitudes generalizado (Λ) para la hipótesis (2.5) en el caso particular C=0, bajo el modelo (2.2), está dado por

$$\Lambda^{2/n}=\frac{|I+Z_{13}V_{33}^{-1}Z_{13}'|}{\left|I+(Z_{12},Z_{13})\begin{pmatrix} V_{22} & V_{23} \\ V_{32} & V_{33} \end{pmatrix}^{-1}(Z_{12},Z_{13})'\right|}$$

y la estadística $T=\Lambda^{2/n}$ puede ser utilizada para realizar el contraste. Gleser & Olkin (1970) señalan que la prueba de cociente de verosimilitudes condicional propuesta por Rao es también la prueba no condicional.

Con objeto de determinar los cuantiles de la distribución de T, Gleser & Olkin (1970) señalan que la distribución de la variable T es igual a la distribución de $\prod_{i=1}^{a_1} \beta_i$ donde $\beta_1,\ldots\ldots,\beta_{a_1}$ son variables independientes tales que

$$\beta_i \sim \text{Beta}[-\tfrac{1}{2}(n-g-p+q-u+a_1-i+1),\ \tfrac{1}{2}a_2] \qquad i=1,\ldots\ldots,a_1\ ,$$

donde $a_1=\min(u,t)$ y $a_2=\max(u,t)$. Como los momentos de la estadística T son conocidos puede utilizarse para realizar la prueba la siguiente aproximación

$$P(-\tfrac{2h}{n}\ ln\Lambda<x)=\ c_1P(\chi^2_{tu}<x)+c_2P(\chi^2_{tu+4}<x)+c_3P(\chi^2_{tu+8}<x)+O(n^{-6})\ ,$$

donde

$$h=\ n-g-p+q+\tfrac{1}{2}(t-u+1)\ ,$$

$$c_1=\ (1-\frac{\delta_1}{h^2})^2+\frac{\delta_1}{h^2}-\frac{\delta_2}{h^4}\ ,$$

$$c_2=\frac{\delta_1}{h^2}\ (1-\frac{\delta_1}{h^2})\ ,$$

$$c_3=\frac{\delta_1^2}{2h^4}+\frac{\delta_2}{h^4}\ ,$$

$$\delta_1=\ \frac{tu(u^2+t^2-5)}{48}\ ,$$

y

$$\delta_2=\ \frac{tu}{1920}\ [3(u^4+t^4)+10t^2u^2-50(u^2+t^2)+159].$$

Como comentarios generales a la exposición de Gleser y Olkin puede mencionarse que no se analiza el hecho de que una particular hipótesis pueda contrastarse; es decir, que involucre sólo funciones lineales estimables de los parámetros. Los autores señalan también que tomar C=0 en la hipótesis (2.5) no es una pérdida de generalidad; sin embargo, resulta de interés proporcionar la forma explícita de la estadística de prueba cuando C es diferente de cero. Resulta también conveniente, obtener una expresión general de la estadística de prueba asociada a una hipótesis lineal general sobre los parámetros linealmente estimables del modelo, es decir, cuando D y E no necesariamente sean de rango completo. Por otro lado, aparece atractivo un estudio más detallado sobre la distribución de la estadística de prueba y realizar pruebas con cuantiles obtenidos de manera más precisa.

3

RESULTADOS GENERALES SOBRE DISTRIBUCION DE FORMAS CUADRATICAS

3 Resultados Generales sobre Distribución de Formas Cuadráticas

Con objeto de presentar de manera simple la obtención y distribución de la estadística de prueba obtenida vía cociente de verosimilitudes generalizado asociada a una hipótesis lineal general en el modelo generalizado de análisis de varianza multivariado (2.2), en esta sección se definen las distribuciones Normal Multivariada, Wishart y Λ de Wilks, se presentan resultados sobre estas distribuciones y algunos teoremas que resultan de utilidad en la obtención del cociente de verosimilitudes generalizado.

3.1 Definiciones

En esta sección se definen densidades que resultan involucradas en la formulación de teoremas sobre distribución de matrices de formas cuadráticas y se presenta una extensión de la definición de función característica al caso de matrices aleatorias.

3.1.1 Definición de las Densidades Normal Multivariada, Wishart y Λ de Wilks

Las siguientes definiciones proporcionan caracterizaciones de las densidades Normal Multivariada, Wishart y Λ de Wilks.

DEFINICION 3.1 Se dice que un vector aleatorio $X_{p\times 1}$ tiene distribución Normal Multivariada de dimensión p con vector de medias $\mu_{p\times 1}$ y matriz de covarianzas $\Sigma_{p\times p}$ definida positiva, si su función de densidad está dada por

$$f_x(x;\, \mu,\, \Sigma)= |2\pi\Sigma|^{-\frac{1}{2}}\, exp\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\} \quad ; \quad x\varepsilon\mathbb{R}^p .$$

Para simplificar se denota por $X\sim N_p(\mu,\Sigma)$.

DEFINICION 3.2. Se dice que una matriz aleatoria $M_{p\times p}$ tiene distribución Wishart de dimensión p con matriz de escala Σ definida positiva y n grados de libertad si se puede representar como $M=X'X$ donde las columnas de $X'=(X_1,\ldots,X_n)$ son independientes y tienen distribución Normal de dimensión p con media cero y matriz de covarianzas Σ. Para simplificar se emplea la notación $M\sim W_p(\Sigma,n)$. Cuando $n\geq p$ se dice que M tiene distribución no singular Si $n<p$ se dice que M tiene distribución Wishart singular.

El valor esperado de una variable con distribución $W_p(\Sigma,n)$ es $n\Sigma$. Esto puede verificarse tomando $X'=(X_1,\ldots,X_n)$ y observando que

$$E(X'X)=\sum_{i=1}^{n} E(X_i\, X_i')$$

$$=\sum_{i=1}^{n}\{E(X_i X_i')-E(X_i)E(X_i')\}$$

$$=\sum_{i=1}^{n} V(X_i)$$

$$=\sum_{i=1}^{n}\Sigma$$

$$=n\Sigma.$$

DEFINICION 3.3. Sean W_1 y W_2 matrices aleatorias con distribuciones independientes $W_p(I,m)$ y $W_p(I,n)$ respectivamente con $m>p$. Si la variable X puede escribirse como

$$X=\frac{|W_1|}{|W_1+W_2|}=|I+W_1^{-1}W_2|^{-1},$$

se dice que tiene distribución Λ de Wilks con parámetros p,m y n. Para simplificar se utiliza la notación $X\sim\Lambda(p,m,n)$.

3.1.2 Definición de la Función Característica de una Matriz Aleatoria.

DEFINICION 3.4. Sea X_{pxq} una matriz aleatoria y θ una matriz de dimensión pxq. La función característica de X denotada por $\varphi_x(\theta)$ se define como

$$\varphi_x(\theta)=E[exp\{\iota\, tr(\theta' X)\}].$$

Es interesante señalar en este punto algunos comentarios en relación a la definición de la función característica de una matriz aleatoria. La traza de la matriz $\theta'X$ está dada por

$$tr\,\theta' X=\sum_{i=1}^{q}\sum_{j=}^{p}\theta_{ji}X_{ji},$$

de donde puede observarse que cada elemento de la matriz X se asocia con un elemento de la matriz θ. Naturalmente esta definición reproduce como casos particulares las definiciones de función característica en el caso de variables aleatorias reales y vectoriales. Los momentos conjuntos de la distribución pueden obtenerse mediante la relación

$$E\left[X_{11}^{k_{11}}\,X_{12}^{k_{12}}\cdots X_{qp}^{k_{qp}}\right]=\frac{1}{\iota^{k_{11}+k_{12}+\cdots+k_{qp}}}\left\{\frac{\partial^{k_{11}+k_{12}+\cdots+k_{qp}}}{\partial\theta_{11}^{k_{11}}\cdots\partial\theta_{qp}^{k_{qp}}}\,\varphi_x(\theta)\right\}\Bigg|_{\theta=0.} \quad (3.1)$$

Esta igualdad puede demostrarse derivando los miembros de la ecuación

$$\varphi_x(\theta)= \int_{-\infty}^{\infty} exp\{\lambda tr\,\theta' x\}\, f_x(x)dx.$$

Otro aspecto de interés en relación a la definición de la función característica consiste en señalar que si la matriz X satisface algunas restricciones, como es el caso en que algunos elementos de la matriz X son por construcción iguales a algunos otros, la matriz θ se puede definir con las correspondientes restricciones o equivalentemente haciendo que algunas de sus entradas sean cero, ya que en realidad en este caso el número de variables aleatorias diferentes es menor que pq. Por ejemplo, si X es una matriz simétrica, la matriz θ se puede escoger simétrica, triangular superior ó triangular inferior. Particularmente, si se desea determinar momentos conjuntos de la distribución de X, es conveniente la elección de θ triangular para que (3.1) siga siendo válida, ya que si X es simétrica y se selecciona θ también simétrica, los momentos conjuntos satisfacen la relación

$$E\left[X_{11}^{k_{11}}\, X_{12}^{k_{12}} \cdots X_{pp}^{k_{pp}}\right] = \frac{1}{(2\lambda)^{k_{11}+k_{12}+\cdots+k_{pp}}}\left\{ \frac{\partial^{k_{11}+k_{12}+\cdots+k_{pp}}}{\partial\theta_{11}^{k_{11}}\cdots\partial\theta_{pp}^{k_{pp}}}\ \varphi_x(\theta)\right\}\Bigg|_{\theta=0}.$$

Como último punto respecto a la definición de la función característica cabe señalar que si X_{pxq} y Y_{rxs} son matrices aleatorias, la función característica conjunta de X y Y está definida por

$$\varphi_{x,y}(\theta_1,\theta_2) = E[exp\{\lambda tr\,\theta_1' X+\lambda tr\,\theta_2' Y\}],$$

donde θ_1 y θ_2 son matrices de dimensión pxq y rxs respectivamente. Esta definición es consistente con la definición 3.4, ya que si se consideran las matrices

$$Z=\begin{pmatrix} X & 0\\ 0 & Y\end{pmatrix} \quad y \quad \theta=\begin{pmatrix} \theta_1 & 0\\ 0 & \theta_2\end{pmatrix},$$

la función característica de Z, que coincide con la función característica conjunta de X y Y, resulta utilizando la definición 3.4

$$\begin{aligned}\varphi_{x,y}(\theta_1,\theta_2) &= \varphi_z(\theta)\\ &= E[\,exp\{\,\lambda tr\,\theta' Z\}]\\ &= E[\,exp\{\,\lambda tr\,\theta_1' X+\lambda tr\,\theta_2' Y\}].\end{aligned}$$

3.2 Teoremas sobre Matrices Aleatorias

En esta sección se generaliza el teorema de unicidad y el teorema de inversión para vectores aleatorios, al caso de matrices aleatorias. Se presenta también un teorema de utilidad en la caracterización de independencia en matrices aleatorias.

3.2.1 El Teorema de Unicidad para Vectores y Matrices Aleatorias.

TEOREMA 3.1. (Teorema de unicidad para vectores aleatorios). Dos vectores aleatorios tienen la misma función característica si y sólo si tienen la misma función de densidad salvo por un conjunto de medida cero.

DEMOSTRACION. Ver Cramér (1946, pág. 101).

El teorema anterior puede generalizarse fácilmente el caso de matrices aleatorias. El resultado se establece en el siguiente teorema.

TEOREMA 3.2. (Teorema de unicidad para matrices aleatorias). Dos matrices aleatorias tienen la misma función característica si y sólo si tienen la misma función de densidad.

DEMOSTRACION. Sea $M_{p\times q}=(M_{(1)},\dots,M_{(q)})$ una matriz aleatoria y sea $N'=(M'_{(1)},\dots,M'_{(q)})$; es decir, N es un vector de dimensión pq que se obtiene apilando las q columnas de matriz M. La función característica de M puede expresarse como

$$\begin{aligned}\varphi_M(\theta) &= E[exp\{\lambda tr(\theta' M)\}]\\ &= E[exp\{\lambda \delta' N\}]\\ &= \varphi_N(\delta),\end{aligned}$$

donde δ es un vector de dimensión pq obtenido apilando las columnas de θ; esto es $\delta'=(\theta'_{(1)},\dots,\theta'_{(q)})$. Del teorema 3.1 se sigue la unicidad de la función de densidad asociada a $\varphi_N(\delta)$ y por tanto se sigue la unicidad de la densidad asociada a $\varphi_M(\theta)$.

3.2.2 El Teorema de Inversión para Vectores y Matrices Aleatorias.

TEOREMA 3.3. (Teorema de inversión para vectores aleatorios). Sea $X_{p\times 1}$ un vector aleatorio. Si la función característica $\varphi_x(t)$ es absolutamente integrable, entonces X tiene función de densidad dada por

$$f_x(x)= (2\pi)^{-p}\int_{-\infty}^{\infty} exp\,(\lambda t'x)\ \varphi_x(t)\,dt.$$

DEMOSTRACION. Ver Cramér (1946, pág. 101).

TEOREMA 3.4. (Teorema de inversión para matrices aleatorias). Sea $X_{p\times q}$ una matriz aleatoria. Si la función característica $\varphi_x(\theta)$ es absolutamente integrable, entonces X tiene función de densidad dada por

$$f_x(x)= (2\pi)^{-pq}\int_{-\infty}^{\infty} exp(\imath\, tr\, \theta' x)\varphi_x(\theta)d\theta.$$

DEMOSTRACION. Sea $X=(X_{(1)},\ldots,X_{(q)})$ y sea Y un vector aleatorio de dimensión pq definido por $Y'=(X'_{(1)},\ldots,X'_{(q)})$. Definiendo $t'=(\theta'_{(1)},\ldots,\theta'_{(q)})$ donde $\theta_{(1)},\ldots,\theta_{(q)}$ corresponden a las columnas de la matriz θ se obtiene que

$$\begin{aligned}\varphi_x(\theta) &= E[exp\{\imath\, tr\, \theta' X\}] \\ &= E[exp\{\imath t' Y\}] \\ &= \varphi_y(t).\end{aligned}$$

Del teorema 3.3 se sigue que

$$\begin{aligned}f_x(x) &= f_x(y) \\ &= (2\pi)^{-pq}\int_{-\infty}^{\infty} exp(\imath t' y)\varphi_y(t)dt \\ &= (2\pi)^{-pq}\int_{-\infty}^{\infty} exp(\imath\, tr\, \theta' x)\,\varphi_x(\theta)d\theta,\end{aligned}$$

lo cual demuestra el resultado.

3.2.3 Una Caracterización de Independencia en Matrices Aleatorias.

Una forma de caracterizar la independencia entre dos matrices aleatorias se formula en el siguiente teorema.

TEOREMA 3.5. Las matrices aleatorias $X_{p\times q}$ y $Y_{r\times s}$ son independientes si y sólo si la función característica conjunta de X y Y se factoriza como el producto de las funciones características marginales de X y Y.

DEMOSTRACION. Sean X y Y independientes. La función característica conjunta de X y Y está dada por

$$\varphi_{x,y}(\theta,\Lambda)= E[exp\{\imath\, tr\, \theta' X + \imath\, tr\, \Lambda' Y\}]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} exp\{\imath tr\theta' x+\imath tr \Lambda' y\} f_{x,y}(x,y)dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} exp\{\imath tr\theta' x+\imath tr\Lambda' y\} f_x(x) f_y(y)\,dx\,dy$$

$$= \varphi_x(\theta)\ \varphi_y(\Lambda).$$

La demostración en sentido inverso resulta de utilidad para caracterizar la independencia entre variables. Sean X y Y dos matrices aleatorias tales que $\varphi_{x,y}(\theta,\Lambda)= \varphi_x(\theta)\ \varphi_y(\Lambda)$. Sean Z y U vectores columna definidos por $Z'=(X'_{(1)},\ldots,X'_{(q)},\ Y'_{(1)},\ldots,Y_{(s)})$ y $U'=(\theta'_{(1)},\ldots,\theta'_{(q)},\ \Lambda'_{(1)},\ldots,\Lambda'_{(s)})$ donde $X=(X_{(1)},\ldots,X_{(q)})$, $Y=(Y_{(1)},\ldots,Y_{(s)})$, $\theta=(\theta_{(1)},\ldots,\theta_{(q)})$ y $\Lambda=(\Lambda_{(1)},\ldots,\Lambda_{(s)})$. Del teorema 3.3 se obtiene que

$$f_{x,y}(x,y)= f_z(z)$$

$$= (2\pi)^{-(pq+rs)} \int_{-\infty}^{\infty} exp(\imath\, u'z)\,\varphi_z(u)du$$

$$= (2\pi)^{-(pq+rs)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} exp(\imath tr\,\theta'x+\imath tr\,\Lambda'y)\,\varphi_{x,y}(\theta,\Lambda)\,d\theta d\Lambda$$

$$= (2\pi)^{-(pq+rs)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} exp(\imath tr\,\theta'x+\imath tr\,\Lambda'y)\,\varphi_x(\theta)\ \varphi_y(\Lambda)\,d\theta d\Lambda$$

$$=\left[(2\pi)^{-pq} \int_{-\infty}^{\infty} exp(\imath tr\,\theta'x)\ \varphi_x\ \theta)\,d\theta\right]\left[(2\pi)^{-rs}\int_{-\infty}^{\infty} exp(\imath tr\Lambda'y)\,\varphi_y(\Lambda)d\Lambda\right]$$

$$= f_x(x)\ f_y(y),$$

lo cual implica la independencia entre X y Y.

3.3 La Función Característica de la Densidad Normal Multivariada

En el siguiente teorema se obtiene la función característica de una variable aleatoria con distribución normal multivariada.

TEOREMA 3.6. Si $X\sim N_p(\mu,\Sigma)$ y $t'=(t_1,\ldots,t_p)$ entonces

$$\varphi_x(t)= exp\,\{\imath\, t'\mu - \tfrac{1}{2} t'\Sigma t\}.$$

DEMOSTRACION. Definiendo $Y=\Sigma^{-\frac{1}{2}}(X-\mu)$, la función característica de X puede calcularse como

$$\varphi_x(t)= E\{ exp(\imath t'X)\}$$

$$= E\{exp(\imath t'[\Sigma^{\frac{1}{2}}Y+\mu]\}$$

$$= exp(\imath t'\mu)\, E\{exp(\imath t'\Sigma^{\frac{1}{2}}Y\}$$

$$= exp(\imath t'\mu)\, \varphi_y(\Sigma^{\frac{1}{2}}t).$$

La función característica de Y valuada en $\delta'=(\delta_1,\ldots,\delta_p)$ puede obtenerse como

$$\varphi_y(\delta) = \int_{-\infty}^{\infty} exp(\imath\delta'y)\, f_y(y)dy$$

$$= \int_{-\infty}^{\infty} exp(\imath\delta'y)\,|2\pi\Sigma|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}(\Sigma^{\frac{1}{2}}y+\mu-\mu)'\Sigma^{-1}(\Sigma^{\frac{1}{2}}y+\mu-\mu)\right\}|\Sigma^{\frac{1}{2}}|\,dy$$

$$= \int_{-\infty}^{\infty} (2\pi)^{-p/2} exp\left\{\imath\delta'y-\tfrac{1}{2}y'y\right\}dy$$

$$= \prod_{i=1}^{p}\left[(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} exp\{(\imath\delta_i y_i-\tfrac{1}{2}y_i^2\}dy_i\right]$$

$$= \prod_{i=1}^{p}\left[exp\left\{-\tfrac{1}{2}\delta_i^2\right\}\int_{-\infty}^{\infty}(2\pi)^{\frac{1}{2}}\; exp\left\{-\tfrac{1}{2}(y_i-\imath\delta_i)^2\right\}dy_i\right]$$

$$= \prod_{i=1}^{p} exp\left\{-\tfrac{1}{2}\delta_i^2\right\}$$

$$= exp\left\{-\tfrac{1}{2}\delta'\delta\right\}.$$

de donde se sigue finalmente que

$$\varphi_x(t) = exp(\imath t'\mu)\varphi_y(\Sigma^{\frac{1}{2}}t)$$

$$= exp\left\{\imath t'\mu-\tfrac{1}{2}t'\Sigma t\right\}.$$

lo cual demuestra el teorema.

3.4 Resultados sobre la Distribución Normal Multivariada en el Caso de Vectores.

En esta sección se obtiene la densidad de transformaciones lineales de variables normales multivariadas y se deduce la función característica de una forma cuadrática general con distribución normal multivariada.

3.4.1 Distribución de Transformaciones Lineales de Variables con Densidad Normal Multivariada

TEOREMA 3.7. Sea $X \sim N_p(\mu, \Sigma)$, $Y = AX + b$ donde $A_{q \times p}$ es una matriz de constantes con $r(A) = q \leq p$ y $b \in \mathbb{R}^q$ un vector de constantes. La distribución de Y es Normal Multivariada de media $A\mu + b$ y matriz de covarianzas $A\Sigma A'$.

DEMOSTRACION. La función característica de Y puede calcularse como

$$\begin{aligned}\varphi_y(t) &= E\{exp(\imath t'Y)\}\\ &= E\{exp[\imath t'(AX+b)]\}\\ &= exp(\imath t'b)\, E\{exp(\imath t'AX)\}\\ &= exp(\imath t'b)\,\varphi_x(A't)\end{aligned}$$

y por una aplicación del teorema 3.6 se sigue que

$$\begin{aligned}\varphi_y(t) &= exp(\imath t'b + \imath t'A\mu - \tfrac{1}{2}\, t'A\Sigma A't)\\ &= exp(\imath t'\delta - \tfrac{1}{2}\, t'\theta t),\end{aligned}$$

donde

$$\delta = A\mu + b$$

y

$$\theta = A\Sigma A'.$$

Como A es de rango completo y Σ definida positiva, se sigue que θ es definida positiva. Del teorema 3.6 se deduce que $\varphi_y(t)$ es la función característica de una variable con distribución $N_q(\delta, \theta)$ y del teorema 3.2 se sigue el resultado.

3.4.2 La Función Característica de una Transformación Cuadrática General de una Variable con Distribución Normal Multivariada

TEOREMA 3.8. Si $X \sim N_p(\mu, \Sigma)$ y $Z = X'AX + b'X$ donde A es simétrica y $b \in \mathbb{R}^p$ entonces

$$\varphi_z(t) = |I - 2\imath t A\Sigma|^{-\frac{1}{2}}\, exp\left\{\imath t\,(\mu'A\mu + b'\mu) - \tfrac{1}{2} t^2 (2A\mu + b)'(\Sigma^{-1} - 2\imath t A)^{-1}(2A\mu + b)\right\}.$$

DEMOSTRACION. Como Σ^{-1} es definida positiva y A es simétrica, del teorema A.18 se tiene que existe una matriz $F_{p \times p}$ de rango completo tal que $F'\Sigma^{-1}F = I$ y $F'AF = D$, donde $D = diag(d_1, \ldots, d_p)$. Definiendo $Y = F^{-1}(X - \mu)$ se sigue del teorema 3.7 que $Y \sim N_p(0, I)$. Utilizando estas relaciones puede escribirse

$$\varphi_z(t) = E\{exp(\lambda tZ)\}$$

$$= E\{exp[\lambda t(X'AX+b'X)]\}$$

$$= exp\{\lambda t(\mu'A\mu+b'\mu)\}\ E\{exp[\lambda t(Y'DY+m'Y)]\}, \qquad (3.2)$$

donde $m' = (m_1,...,m_p) = 2\mu'AF + b'F$. Utilizando la independencia de los componentes del vector Y se sigue que

$$E\{exp[\lambda t(Y'DY+m'Y)]\} = E\left\{exp\left[\lambda t\sum_{j=1}^{p}(d_j Y_j^2+m_j Y_j)\right]\right\}$$

$$= \prod_{j=1}^{p} E\left\{exp\left[\lambda t(d_j Y_j^2+m_j Y_j)\right]\right\}$$

$$= \prod_{j=1}^{p} \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}(1-2\lambda td_j)y_j^2+\lambda tm_j y_j\right\} dy_j$$

$$= \prod_{j=1}^{p} exp\left\{-\tfrac{1}{2}t^2m_j^2(1-2\lambda td_j)^{-1}\right\}\int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}(1-2\lambda td_j)\left[y_j-\lambda tm_j(1-2\lambda td_j)^{-1}\right]^2\right\}dy_j$$

$$= \prod_{j=1}^{p} exp\left\{-\tfrac{1}{2}t^2m_j^2(1-2\lambda td_j)^{-1}\right\}(1-2\lambda td_j)^{-\frac{1}{2}}$$

$$= \left[\prod_{j=1}^{p}(1-2\lambda td_j)^{-\frac{1}{2}}\right] exp\left\{-\tfrac{1}{2}t^2\sum_{j=1}^{p} m_j^2(1-2\lambda td_j)^{-1}\right\}$$

$$= |I-2\lambda tD|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}t^2m'(I-2\lambda tD)^{-1}m\right\}$$

$$= |I-2\lambda tF'AF|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}t^2(2A\mu+b)'F(I-2\lambda tF'AF)^{-1}F'(2A\mu+b)\right\}$$

$$= |(F')^{-1}|\ |I-2\lambda tF'AF|^{-\frac{1}{2}}\ |F|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}t^2(2A\mu+b)'\left[(F')^{-1}(I-2\lambda tF'AF)F^{-1}\right]^{-1}(2A\mu+b)\right\}$$

$$= |I-2\lambda tA\Sigma|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}t^2(2A\mu+b)'(\Sigma^{-1}-2\lambda tA)^{-1}(2A\mu+b)\right\}.$$

Substituyendo este resultado en (3.2) se obtiene

$$\varphi_z(t) = |I-2\lambda tA\Sigma|^{-\frac{1}{2}} exp\left\{\lambda t(\mu'A\mu+b'\mu)-\tfrac{1}{2}t^2(2A\mu+b)'(\Sigma^{-1}-2\lambda tA)^{-1}(2A\mu+b)\right\},$$

lo cual demuestra el teorema.

COLORARIO 3.8. Si $X\sim N_p(\mu,\Sigma)$ y A_{pxp} es una matriz simétrica entonces

$$\varphi_{x'Ax}(t)= |I-2\lambda tA\Sigma|^{-\frac{1}{2}}\, exp\{\lambda t\mu'(I-2\lambda tA\Sigma)^{-1}A\mu\}.$$

DEMOSTRACION. Del teorema 3.8, en el caso particular b=0 se tiene que

$$\begin{aligned}\varphi_{x'Ax}(t)&= |I-2\lambda tA\Sigma|^{-\frac{1}{2}}\, exp\{\lambda t\mu'A\mu-2t^2\mu'A(\Sigma^{-1}-2\lambda tA)^{-1}A\mu\}\\ &= |I-2\lambda tA\Sigma|^{-\frac{1}{2}}\, exp\{\lambda t\mu'[\,I+2\lambda tA(\Sigma^{-1}-2\lambda tA)^{-1}]\,A\mu\}\\ &= |I-2\lambda tA\Sigma|^{-\frac{1}{2}}\, exp\,\{\lambda t\mu'[(\Sigma^{-1}-2\lambda tA)+2\lambda tA]\,(\Sigma^{-1}-2\lambda tA)^{-1}A\mu\}\\ &= |I-2\lambda tA\Sigma|^{-\frac{1}{2}}\, exp\{\lambda t\mu'\Sigma^{-1}(\Sigma^{-1}-2\lambda tA)^{-1}A\mu\}\\ &= |I-2\lambda tA\Sigma|^{-\frac{1}{2}}\, exp\{\lambda t\mu'(I-2\lambda tA\Sigma)^{-1}A\mu\},\end{aligned}$$

lo que demuestra el corolario.

3.5 Caracterización de la Densidad Wishart y de una Matriz de Datos Normales

El propósito de este apartado es determinar la función característica de una variable con distribución Wishart y de una matriz de observaciones normales independientes.

3.5.1 La Función Característica de una Variable con Densidad Wishart

TEOREMA 3.9. Si $M_{pxp}\sim W_p(\Sigma,n)$ y $\theta_{pxp}=\{\theta_{ij}\}$ es una matriz simétrica entonces la función característica de M está dada por

$$\varphi_M(\theta)= |I-2\lambda\theta\Sigma|^{-n/2}.$$

DEMOSTRACION. De la definición 3.2 puede escribirse $M=\sum_{j=1}^{n} Y_j\, Y_j'$ donde $Y_1,\ldots,Y_n$ son independientes con distribución $N_p(0,\Sigma)$. Utilizando esta igualdad se obtiene

$$\begin{aligned}\varphi_M(\theta)&= E\{exp(\lambda\, tr\,\theta' \sum_{j=1}^{n} Y_j\, Y_j')\}\\ &= \prod_{j=1}^{n} E\{exp(\lambda Y_j'\,\theta\, Y_j)\}\\ &= [E\{exp(\lambda Y'\theta Y)\}]^n,\end{aligned}$$

donde $Y\sim N_p(0,\Sigma)$. Utilizando el corolario 3.8 se deduce que

$$\varphi_M(\theta)=\left[\varphi_{y'\theta y}(\imath)\right]^n$$

$$= |I-2\imath\theta\Sigma|^{-n/2},$$

lo cual demuestra el resultado.

3.5.2 La Función Característica de una Matriz de Variables con Distribución Normal Multivariada

TEOREMA 3.10. Sea X_{nxp} una matriz aleatoria donde las columnas de $X'=(X_1,\dots,X_n)$ son independientes y tales que

$$X_i \sim N_p(\mu_i, \Sigma) \qquad i=1,\dots,n.$$

La función característica de X está dada por

$$\varphi_x(\theta)= exp\,\{\imath tr\, \theta'\mu-\tfrac{1}{2}tr\theta\,\Sigma\,\theta'\},$$

donde $\mu'=(\mu_1\ \dots,\mu_n)$ y θ es una matriz de dimensión nxp.

DEMOSTRACION. Definiendo $\theta'=(\theta_1,\dots,\theta_n)$ se obtiene que la función característica de X puede expresarse como

$$\varphi_x(\theta) = E\,\{exp(\imath tr\,\theta' X)\}$$

$$= E\,\{exp(\imath tr \sum_{i=1}^{n} \theta_i X_i')\}$$

$$= E\,\left\{exp(\imath \sum_{i=1}^{n} \theta_i X_i')\right\}$$

$$= \prod_{i=1}^{n} E\left\{exp\ (\imath\,\theta_i' X_i)\right\}$$

$$= \prod_{i=1}^{n}\ \varphi_{x_i}(\theta_i).$$

Por una aplicación del teorema 3.6 se obtiene

$$\varphi_x(\theta)=\prod_{i=1}^{n}\ \exp\{\,\imath\,\theta_i'\mu_i-\tfrac{1}{2}\theta_i'\Sigma\,\theta_i\}$$

$$= exp\left\{\imath tr \sum_{i=1}^{n}\theta_i\mu_i-\tfrac{1}{2}\sum_{i=1}^{n}\ \theta_i'\,\Sigma\,\theta_i\right\}$$

$$= exp\,\left\{\imath tr\,\theta'\mu-\tfrac{1}{2}\,tr\,\Sigma \sum_{i=1}^{n}\,\theta_i\,\theta_i'\right\}$$

$$= exp\left\{\imath tr\,\theta'\mu-\tfrac{1}{2}\,tr\,\Sigma\ \theta'\ \theta\right\}$$

$$= exp\{ \imath tr\, \theta'\mu - \tfrac{1}{2}\, tr\, \theta\Sigma\theta' \},$$

lo cual finaliza la demostración.

3.6 Resultados sobre la Distribución Normal Multivariada en el Caso de Matrices, la Distribución Wishart y la distribución Λ de Wilks

Esta sección tiene como propósito establecer una serie de resultados asociados con matrices de observaciones normales multivariadas. En estos teoremas se determina la distribución de un tipo de transformaciones lineales de variables Wishart y se establecen las condiciones necesarias y suficientes para que una transformación lineal de una matriz de observaciones independientes de la distribución normal resulte en una matriz de observaciones independientes de esta distribución. Se determinan también las condiciones bajo las cuales dos transformaciones lineales de una matriz de observaciones normales resultan ser independientes obteniéndose la función característica de una forma cuadrática general de una matriz de datos normales y estableciéndose las condiciones bajo las cuales esta forma cuadrática poseé distribución Wishart.

3.6.1 Distribución de Transformaciones Lineales de Variables Wishart

TEOREMA 3.11. Si $M \sim W_p(\Sigma, n)$ y $B_{p\times q}$ con $q<p$ es una matriz de entradas reales entonces $B'MB \sim W_q(B'\Sigma B, n)$.

DEMOSTRACION. Dado que M se puede escribir como X'X donde los renglones de X son independientes y tienen distribución $N_p(0,\Sigma)$, se tiene que

$$B'MB = B'X'XB = Y'Y,$$

donde $Y' = B'X' = (B'X_1, \ldots, B'X_n)$ y por ser $X_1, \ldots, X_n$ independientes se obtiene que $B'X_1, \ldots, B'X_n$ son independientes. Por el teorema 3.7 $B'X_i \sim N_q(0, B'\Sigma B)$ $i=1,\ldots,n$. Utilizando la definición 3.2 se concluye que

$$B'MB \sim W_q(B'\Sigma B, n),$$

lo cual demuestra la afirmación.

3.6.2 Transformaciones Lineales de Matrices de Observaciones Independientes de la Distribución Normal

En esta sección se determinan las condiciones necesarias y suficientes que se requieren para que una transformación lineal de una matriz de observaciones independientes de la distribución normal multivariada resulte de nuevo en uma matriz de observaciones independientes de esta distribución.

TEOREMA 3.12. Sea X_{nxp} una matriz aleatoria donde las columnas de $X'=(X_1,\ldots,X_n)$ son independientes con distribución $N_p(\mu,\Sigma)$ y sea $Y_{mxq}=AXB$ donde A y B son matrices de constantes tales que $mq \geq np$. Las columnas de $Y'=(Y_1,\ldots,Y_m)$ son independientes con distribución $N_q(\delta B'\mu, \eta B'\Sigma B)$ si y sólo si

a) (Media Común) $A1=\delta 1$ para algún $\delta \in \mathbb{R}$ ó $B'\mu=0$.

b) (Independencia) $AA'=\eta I$ para algún $\eta \in \mathbb{R}$.

donde 1 es el vector de dimensión n con todas sus entradas iguales a uno.

DEMOSTRACION. Sea Y^{a}_{mq} un vector columna obtenido apilando las columnas de Y, esto es

$$Y^{a'}_{mq} = (Y'_{(1)},\ldots,Y'_{(q)}).$$

Del teorema A.5h se obtiene que

$$Y^{a}_{mq} = (AXB)^{a} = (B' \otimes A)X^{a},$$

donde $\otimes$ denota el producto Kronecker.

La distribución de X^{a} es Normal multivariada de dimensión np. Esto puede demostrarse observando que los renglones de X tienen distribución independiente y por tanto la distribución de X puede escribirse como el producto de las densidades de cada vector X_i; es decir

$$f_X(X;\mu,\Sigma) = \prod_{i=1}^{n} f_{x_i}(x_i;\mu,\Sigma) = N_{np}(\mu \otimes 1, \Sigma \otimes I),$$

de donde se sigue que la distribución de X^{a} es $N_{np}(\mu \otimes 1, \Sigma \otimes I)$. Si $mq \leq np$ del teorema 3.7 se obtiene que

$$Y^{a} \sim N_{mq}(B'\mu \otimes A1,\ B'\Sigma B \otimes AA').$$

Los vectores $Y_{(1)},\ldots,Y_{(m)}$ tienen la misma media si y sólo si

$$B'\mu \otimes A1 = 0$$

ó

$$B'\mu \otimes A1 = M \otimes 1,$$

donde M es cualquier vector de dimensión mq. Estas dos condiciones equivalen a $B'\mu=0$ ó $A1=\delta 1$ para algún $\delta\in\mathbb{R}$. Por otra parte, las variables $Y_{(1)},\ldots,Y_{(m)}$ son independientes si y sólo si

$$B'\Sigma B \otimes AA' = N \otimes I,$$

donde N es cualquier matriz definida positiva de dimensión qxq. Esta condición equivale a $AA'=\eta I$ para algún número real η.

3.6.3 Un Teorema sobre Independencia de Transformaciones Lineales de una Matriz de Datos Normales Multivariados

TEOREMA 3.13. Sea X_{nxp} una matriz aleatoria donde las columnas de $X'=(X_1,\ldots,X_n)$ son independientes tales que

$$X_i \sim N_p(\mu_i,\Sigma) \qquad i=1,\ldots,n.$$

y sean $Y=AXB$ y $Z=CXD$ con A_{qxn}, B_{pxr}, C_{sxn} y D_{pxt} matrices de constantes. La matriz Y tiene distribución independiente de la matriz Z si y sólo si

$$B'\Sigma D=0 \qquad \text{ó} \qquad CA'=0.$$

DEMOSTRACION. Sean θ_1 y θ_2 matrices de dimensiones qxr y sxt respectivamente. La función característica conjunta de Y y Z puede calcularse como

$$\begin{aligned}\varphi_{y,z}(\theta_1,\theta_2) &= E[exp(\imath tr\,\theta_1' Y + \imath tr\,\theta_2' Z)]\\ &= E[exp(\imath tr\,\theta_1' AXB + \imath tr\,\theta_2' CXD)]\\ &= E[exp(\imath tr\,B\theta_1' AX + \imath tr\,D\theta_2' CX)]\\ &= E[exp\{\imath tr\,(B\theta_1' A + D\theta_2' C)X\}]\\ &= \varphi_x(A'\theta_1 B' + C'\theta_2 D')\end{aligned}$$

y del teorema 3.10 se sigue que

$$\varphi_{y,z}(\theta_1,\theta_2) = exp\left\{\imath tr\,\eta'\mu - \tfrac{1}{2}\, tr\,\eta\Sigma\eta'\right\}, \tag{3.3}$$

donde $\eta=A'\theta_1B'+C'\theta_2D'$ y $\mu'=(\mu_1,\ldots,\mu_n)$. La función característica de la matriz Y puede calcularse como función característica marginal de (3.3) y está dada por

$$\begin{aligned}\varphi_y(\theta_1) &= \varphi_{y,z}(\theta_1,0)\\ &= \exp\left\{\imath tr\,B\theta_1' A\mu - \tfrac{1}{2}\, trA'\theta_1 B_1'\Sigma B\theta_1' A\right\}\\ &= \exp\left\{\imath tr\,B\theta_1' A\mu - \tfrac{1}{2}\, tr\theta_1 B'\Sigma B\theta_1' AA'\right\}.\end{aligned}$$

Analogamente se obtiene que

$$\varphi_z(\theta_2)= exp\left\{ i\,tr\, D\theta_2' C\mu - \tfrac{1}{2} tr\, \theta_2 D'\Sigma D\theta_2' CC' \right\}.$$

Del teorema 3.5 se tiene que Y y Z son independientes si y sólo si

$$\varphi_{y,z}(\theta_1,\theta_2)= \varphi_y(\theta_1)\,\varphi_z(\theta_2) \qquad \forall\,\theta_1,\theta_2$$

esto es, si y sólo si

$$tr(\eta\Sigma\eta')= tr(\theta_1 B'\Sigma B\theta_1' AA') + tr(\theta_2 D'\Sigma D\theta_2' CC') \qquad \forall\,\theta_1,\theta_2$$

$$\Leftrightarrow tr\, A'\theta_1 B'\Sigma D\theta_2' C=0 \qquad \forall\,\theta_1,\theta_2$$

$$\Leftrightarrow tr\, CA'\theta_1 B'\Sigma D\theta_2' =0 \qquad \forall\,\theta_1,\theta_2 \tag{3.4}$$

Si CA'=0 ó B'ΣD=0 la ecuación (3.4) se verifica. La implicación inversa puede demostrarse definiendo u= r(CA') y v=r(B'ΣD) y obteniendo del teorema A.17 que

$$CA'= U\Lambda V'$$

y

$$B'\Sigma D= \tilde{U}\tilde{\Lambda}\tilde{V}',$$

donde U_{sxu}, V_{qxu}, $\tilde{U}_{rxv}$, $\tilde{V}_{txv}$ son matrices que satisfacen $U'U=I$, $V'V=I$, $\tilde{U}'\tilde{U}=I$ y $\tilde{V}'\tilde{V}=I$ y las matrices Λ y $\tilde{\Lambda}$ son diagonales con elementos positivos definidos por $\Lambda=Diag(\lambda_1,\dots,\lambda_u)$ y $\tilde{\Lambda}= Diag(\tilde{\lambda}_1,\dots,\tilde{\lambda}_v)$. Definiendo $\theta_1= V1_u1'_v\tilde{U}'$ y $\theta_2= U1_u1'_v\tilde{V}'$ donde 1_u y 1_v son vectores columnas de dimensiónes respectivas u y v con todas sus entradas iguales a la unidad, del teorema A.1.e se sigue que

$$tr\;\; U\Lambda V'V1_u1_v'\;\tilde{U}'\;\tilde{U}\tilde{\Lambda}\tilde{V}'\tilde{V}1_v1_u'U'=0$$

$$\Rightarrow tr\;\; \Lambda 1_u1_v'\;\; \tilde{\Lambda}1_v1_u'=0$$

$$\Rightarrow 1_u'\Lambda 1_u\, 1_v'\tilde{\Lambda}1_v=0$$

$$\Rightarrow \left(\sum_{i=1}^{u}\lambda_i\right)\left(\sum_{i=1}^{v}\tilde{\lambda}_i\right)=0$$

y dado que $\lambda_i>0$ i=1,...,u y $\tilde{\lambda}_i>0$ i=1,...,v se deduce que $\Lambda=0$ ó $\tilde{\Lambda}=0$ de donde se sigue que CA'=0 ó B'ΣD=0 lo cual finaliza la demostración.

3.6.4 La Función Característica de una Transformación Cuadrática General de una Matriz de Datos Normales Multivariados

TEOREMA 3.14. Sea X_{nxp} una matriz aleatoria y μ_{nxp} una matriz de constantes tales que $X'=(X_1,\dots,X_n)$ y $\mu'=(\mu_1,\dots,\mu_n)$ con n>p, donde $X_1,\dots,X_n$ son vectores independientes tales que

$$X_i \sim N_p(\mu_i, \Sigma) \qquad i=1,\dots,n,$$

con Σ_{pxp} una matriz definida positiva. Sean A_{mxn}, B_{pxq}, D_{mxq} y C_{mxq} matrices de constantes con $m<n$, $q\leq p$ tales que $r(A)=m$, $r(B)=q$ y D es una matriz simétrica. La matriz W_{qxq} definida por

$$W = (AXB+C)'\, D\, (AXB+C),$$

tiene función característica dada por

$$\varphi_w(\theta) = \Big[\prod_{j=1}^{n} |I-2\lambda\lambda_j\theta B'\Sigma B|\Big]^{-\frac{1}{2}} exp\Big\{\lambda\, tr \sum_{j=1}^{n} (\mu B+\tilde{C})\,(I-2i\lambda_j\theta B'\Sigma B)^{-1}\theta\,(\mu B+\tilde{C})\,\lambda_j U_j U_j\Big\},$$

donde

$$\tilde{C} = A'(AA')^{-1}C$$

y

$$A'DA = U\Lambda U' \;;\quad U'U=I$$

y las matrices $U=(U_1,\dots,U_n)$ y $\Lambda=diag(\lambda_1,\dots,\lambda_n)$ definen la descomposición espectral de la matriz A'DA.

DEMOSTRACION. La matriz W puede escribirse como

$$
\begin{aligned}
W &= (AXB+C)'D\,(AXB+C)\\
&= [A\{XB+A'(AA')^{-1}C\}]'\, D\, [A\{XB+A'(AA')^{-1}C\}]\\
&= [XB+A'(AA')^{-1}C]'\, A'DA[XB+A'(AA')^{-1}C].
\end{aligned}
$$

Definiendo

$$\tilde{C}_{nxq} = A'(AA')^{-1}C$$

y

$$\tilde{D}_{nxn} = A'DA,$$

se obtiene

$$W = (XB+\tilde{C})'\,\tilde{D}\,(XB+\tilde{C}).$$

Sea la descomposición espectral de la matriz $\tilde{D}$ definida por $\tilde{D}=U\Lambda U'$ donde $U_{nxn}=(U_1,\dots,U_n)$ es una matriz que satisface $U'U=I$ y $\Lambda=diag\,(\lambda_1,\dots,\lambda_n)$. Sea Y_{nxq} una matriz aleatoria definida por

$$Y_{nxq} = U'(XB+\tilde{C}) = \begin{bmatrix} Y_1' \\ \vdots \\ Y_n' \end{bmatrix}.$$

Los vectores $Y_1,\ldots,Y_n$ son independientes ya que $UU'=I$. El vector de medias y matriz de covarianzas de Y_j, $j=1,\ldots n$ están dados por

$$M_j = E(Y_j) = (\mu B+\tilde{C})' U_j$$

y

$$\begin{aligned} V(Y_j) &= V\{(XB)' U_j\} \\ &= V(B'X'U_j) \\ &= V\ (\sum_{k=1}^{n} U_{jk} B'X_k) \\ &= \sum_{k=1}^{n} U_{jk}^2\ B'\Sigma B \\ &= B'\Sigma B. \end{aligned}$$

Utilizando la descomposición espectral de la matriz $\tilde{D}$ se tiene

$$\begin{aligned} (XB+\tilde{C})'\ \tilde{D}\ (XB+\tilde{C}) &= (XB+\tilde{C})' U \Lambda U' (XB+\tilde{C}) \\ &= [U'(XB+\tilde{C})]\ \Lambda\ [U'(XB+\tilde{C})] \\ &= Y' \Lambda Y \\ &= \sum_{j=1}^{n} \lambda_j Y_j Y_j'. \end{aligned}$$

Con este resultado, la función característica de W puede calcularse como

$$\begin{aligned} \varphi_w(\theta) &= E\{exp[\imath tr\, \theta' (XB+\tilde{C})'\ \tilde{D}\ (XB+\tilde{C})]\} \\ &= E\left\{exp\,(\imath tr\, \theta' \sum_{j=1}^{n} \lambda_j Y_j Y_j')\right\} \\ &= E\left\{exp\,(\imath tr \sum_{j=1}^{n} \lambda_j \theta' Y_j Y_j')\right\} \\ &= E\left\{exp\,(\imath \sum_{j=1}^{n} \lambda_j Y_j' \theta Y_j)\right\} \\ &= \prod_{j=1}^{p} E\{exp\,(\imath \lambda_j Y_j' \theta' Y_j)\} \\ &= \prod_{j=1}^{p} \varphi_{y_j' \theta\, y_j}(\lambda_j) \end{aligned}$$

y del corolario 3.8 se sigue que

$$\varphi_w(\theta)=\prod_{j=1}^{n}|I-2i\lambda_j\theta B'\Sigma B|^{-\frac{1}{2}}\, exp\left\{i\lambda_j M_j'(I-2i\lambda_j\theta B'\Sigma B)^{-1}\theta M_j\right\}$$

$$=\left[\prod_{j=1}^{n}|I-2i\lambda_j\theta B'\Sigma B|^{-\frac{1}{2}}\right] exp\left\{i\sum_{j=1}^{n}\lambda_j U'_j(\mu B+\tilde{C})(I-2i\lambda_j\theta B'\Sigma B)^{-1}\theta(\mu B+\tilde{C})'U_j\right\}$$

$$=\left[\prod_{j=1}^{n}|I-2i\lambda_j\theta B'\Sigma B|^{-\frac{1}{2}}\right] exp\left\{i\,tr\sum_{j=1}^{n}(\mu B+\tilde{C})(I-2i\lambda_j\theta B'\Sigma B)^{-1}\theta(\mu B+\tilde{C})'\lambda_j U_j U'_j\right\},$$

lo cual demuestra el teorema.

3.6.5 Las Condiciones que Deben Ser Impuestas para que una Transformación Cuadrática General de una Matriz de Datos Normales Multivariados Posea Distribución Wishart

TEOREMA 3.15. Sea X_{nxp} matriz aleatoria y μ_{nxp} una matriz de constantes con particiones definidas por $X'=(X_1,\ldots,X_n)$ y $\mu'=(\mu_1,\ldots,\mu_n)$ donde $n\geq p$ y $X_1,\ldots,X_n$ variables independientes tales que

$$X_i\sim N_p(\mu_i,\Sigma) \qquad i=1,\ldots,n,$$

con Σ_{pxp} definida positiva. Sean A_{mxn}, B_{pxq}, D_{mxm} y C_{mxq} matrices de constantes con $m\leq n$, $q\leq p$ tales que $r(A)=m$, $r(B)=q$ y D es una matriz simétrica. La matriz W_{qxq} definida por

$$W=(AXB+C)'\,D\,(AXB+C)$$

posee distribución Wishart con matriz de escala $B'\Sigma B$ y d grados de libertad donde $d=r(D)$ si y sólo si las siguientes condiciones se satisfacen

a) $A'DA$ es idempotente.

b) $A'D(A\mu B+C)=0$.

DEMOSTRACION. La matriz W puede escribirse como

$$W=(XB+\tilde{C})'\,\tilde{D}\,(XB+\tilde{C}),$$

donde

$$\tilde{C}=A'(AA')^{-1}C$$

y

$$\tilde{D}=A'DA.$$

Sea la descomposición espectral de la matriz $\tilde{D}$ definida por $\tilde{D} = U \Lambda U'$ donde $U'U = I$. Del teorema 3.14 se obtiene que la función característica de la matriz W está dada por

$$\varphi_w(\theta) \quad \left[\prod_{j=1}^{n} |I - 2\lambda\lambda_j \theta B'\Sigma B\right]^{-\frac{1}{2}} exp\left\{\lambda tr \sum_{j=1}^{n} (\mu B + \tilde{C})(I - 2\lambda\lambda_j \theta B'\Sigma B)^{-1} \theta (\mu B + \tilde{C})' \lambda_j U_j U_j'\right\}, \quad (3.5)$$

donde $\theta_{q\times q}$ es una matriz simétrica.

Primero se demuestra la suficiencia de las condiciones. Para esto, dado que $m \leq n$ se tiene

$$r(\tilde{D}) = r(A'DA) = r(D) = d$$

y dado que $\tilde{D}$ es idempotente, se sigue que d valores propios de $\tilde{D}$ son iguales a uno y los $m-d$ restantes son iguales a cero, por lo que la función característica de W puede escribirse como

$$\begin{aligned}
\varphi_w(\theta) &= |I - 2\lambda\theta B'\Sigma B|^{-d/2} \, exp\{\lambda tr\, (\mu B + \tilde{C})(I - 2\lambda\theta B'\Sigma B)^{-1} \, \theta \, (\mu B + \tilde{C})' \sum_{j=1}^{r} \lambda_j U_j U_j' \} \\
&= |I - 2\lambda\theta \, B'\Sigma B|^{-d/2} \, exp\{\lambda tr\, (\mu B + \tilde{C})(I - 2\lambda\theta B'\Sigma B)^{-1} \theta \, (\mu B + \tilde{C})' \tilde{D}\} \\
&= |I - 2\lambda\theta \, B'\Sigma B|^{-d/2} \, exp\{\lambda tr\, (\mu B + \tilde{C})(I - 2\lambda\theta B'\Sigma B)^{-1} \, \theta \, (\mu B + \tilde{C})' \tilde{D}\tilde{D}\} \\
&= |I - 2\lambda\theta \, B'\Sigma B|^{-d/2} \, exp\{\lambda tr \tilde{D} \, (\mu B + \tilde{C})(I - 2\lambda\theta B'\Sigma B)^{-1} \, \theta \, (\mu B + \tilde{C})' \tilde{D}' \} \\
&= |I - 2\lambda\theta \, B'\Sigma B|^{-d/2} \, exp\{\lambda tr [\tilde{D} \, (\mu B + \tilde{C})] \, (I - 2\lambda\theta B'\Sigma B)^{-1} \, \theta [\tilde{D} \, (\mu B + \tilde{C})]'\}.
\end{aligned}$$

Utilizando la condición b) se deduce que

$$\begin{aligned}
\tilde{D}(\mu B + \tilde{C}) &= A'DA[\mu B + A'(AA')^{-1} C] \\
&= A'D\,[A\mu B + C] \\
&= 0,
\end{aligned}$$

de donde se sigue que

$$\varphi_w(\theta) = |I - 2\lambda\theta B'\Sigma B|^{-d/2}.$$

Del teorema 3.9 se observa que esta función característica coresponde a la de una variable con distribución Wishart de dimensión q con matriz de escala $B'\Sigma B$ y d grados de libertad y por una aplicación del teorema 3.2 se sigue que $W \sim W_q(B'\Sigma B, d)$.

Para la demostración en sentido inverso, se definen los vectores

$$M_j = (\mu B + \tilde{C})' U_j \qquad j=1,\dots,n.$$

Utilizando estas definiciones y la ecuación (3.5) se sigue que la función característica de W puede escribirse como

$$\varphi_W(\theta) = \Big[\prod_{j=1}^{n} |I-2\lambda\lambda_j\theta B'\Sigma B|^{-\frac{1}{2}}\Big]\, exp\Big\{\lambda \sum_{j=1}^{n} \lambda_j M_j'(I-2\lambda\lambda_j\theta B'\Sigma B)^{-1}\theta M_j\Big\}.$$

Utilizando la condición $W \sim W_p(B'\Sigma B, d)$ y la unicidad de la función característica, se sigue que la siguiente condición debe verificarse

$$\Big[\prod_{j=1}^{n} |I-2\lambda\lambda_j\theta B'\Sigma B|^{-\frac{1}{2}}\Big]\, exp\Big\{\lambda \sum_{j=1}^{n} \lambda_j M_j'(I-2\lambda\lambda_j\theta B'\Sigma B)^{-1}\theta M_j\Big\} = |I-2\lambda\theta B'\Sigma B|^{-d/2},$$

$$\forall\, \theta \text{ simétrica.} \qquad (3.6)$$

Elevando al cuadrado ambos miembros de (3.6) y reagrupando términos se obtiene

$$exp\Big\{2\lambda \sum_{j=1}^{n} \lambda_j M_j'(I-2\lambda\lambda_j\theta B'\Sigma B)^{-1}\theta M_j\Big\} = \frac{\prod_{j=1}^{n} |I-2\lambda\lambda_j\theta B'\Sigma B|}{|I-2\lambda\theta B'\Sigma B|^{d}}, \qquad \forall\, \theta \text{ simétrica.}$$

En el caso particular $\theta = t(B'\Sigma B)^{-1}$, la ecuación anterior implica la condición

$$exp\Big\{2\lambda \sum_{j=1}^{n} \lambda_j t(1-2\lambda\lambda_j t)^{-1} M_j'(B'\Sigma B)^{-1} M_j\Big\} = \frac{\prod_{j=1}^{n} (1-2\lambda\lambda_j t)^{q}}{(1-2\lambda\lambda_j t)^{qd}}, \qquad \forall t \in \mathbb{R}. \qquad (3.7)$$

Dado que el argumento de la función exponencial y el miembro del lado derecho son funciones racionales, la ecuación (3.7) puede escribirse como

$$exp\{P(\lambda t)/Q(\lambda t)\} = R(\lambda t)/S(\lambda t),$$

donde P, Q, R y S son polinomios. Utilizando el hecho de que la función exponencial no es racional, se obtiene que $P(\lambda t)/Q(\lambda t)$ y $R(\lambda t)/S(\lambda t)$ son constantes, lo cual implica que

$$2\lambda \sum_{j=1}^{n} \lambda_j t(1-2\lambda\lambda_j t)^{-1} M_j'(B'\Sigma B)^{-1} M_j = c_1 \qquad \forall t \in \mathbb{R}$$

y

$$\prod_{j=1}^{n} (1-2\lambda\lambda_j t)^{q} = c_2 (1-2\lambda\lambda_j t)^{qd} \qquad \forall t \in \mathbb{R}.$$

Valuando en $t=0$ las ecuaciones anteriores se deduce que $c_1=0$ y $c_2=1$, de donde se sigue que

$$2\lambda\sum_{j=1}^{n} \lambda_j t(1-2\lambda\lambda_j t)^{-1} M'_j(B'\Sigma B)^{-1}M_j=0 \qquad \forall t\in\mathbb{R} \tag{3.8}$$

y

$$\prod_{j=1}^{n} (1-2\lambda\lambda_j t)^{q} = (1-2\lambda t)^{qd} \qquad \forall t\in\mathbb{R}. \tag{3.9}$$

La igualdad (3.9) de polinomios en t se verifica si y sólo si los grados son iguales; es decir, n-d valores propios de $\tilde{D}$ son iguales a cero. Seleccionando sin perder generalidad $\lambda_{d+1}=\lambda_{d+2}=.....=\lambda_n=0$, la ecuación (3.9) implica la condición

$$\prod_{j=1}^{n} (1-2\lambda\lambda_j t)^{q} = (1-2\lambda t)^{qd} \qquad \forall t\in\mathbb{R}.$$

Para que la igualdad anterior de polinomios se verifique, sus raíces deben ser iguales, lo cual ocurre si y sólo si $\lambda_j=1$, $j=1,.....d$. Utilizando este resultado y el teorema A.9 se deduce que la matriz $\tilde{D}$ es idempotente y dado que $m\leq n$, se tiene

$$d= tr(\tilde{D})= r(\tilde{D})= r(A'DA)= r(D),$$

lo cual implica la condición a) del teorema.

Utilizando la idempotencia de la matriz $\tilde{D}$, la ecuación (3.8) valuada en $t=1$ implica que

$$\sum_{j=1}^{d} M'_j(B'\Sigma B)^{-1}M_j=0. \tag{3.10}$$

Dado que $B'\Sigma B$ es definida positiva, se sigue que $(B'\Sigma B)^{-1}$ es definida positiva por lo que

$$M'_j(B'\Sigma B)^{-1}M_j \geq 0. \tag{3.11}$$

De (3.10) y (3.11) se tiene que

$$M'_j(B'\Sigma B)^{-1}M_j=0 \qquad j=1,....,d$$

$$\Rightarrow \quad M_j=0 \qquad j=1,....,d$$

y observando que $\lambda_j=0$, $j=d+1,...,n$ se sigue que

$$\lambda_j M'_j M_j =0 \qquad j=1,...,n.$$

Haciendo la suma sobre el índice j se tiene que

$$\sum_{j=1}^{n} \lambda_j M_j' M_j = 0$$

$$\Leftrightarrow \sum_{j=1}^{n} \lambda_j U_j' (\mu B + \tilde{C}) (\mu B + \tilde{C})' U_j = 0$$

$$\Leftrightarrow tr (\mu B + \tilde{C}) (\mu B + \tilde{C})' \sum_{j=1}^{n} \lambda_j U_j U_j' = 0$$

$$\Leftrightarrow tr (\mu B + \tilde{C}) (\mu B + \tilde{C})' \tilde{D} = 0$$

$$\Leftrightarrow tr [\tilde{D} (\mu B + \tilde{C})]' \tilde{D} (\mu B + \tilde{C}) = 0 . \tag{3.12}$$

Definiendo

$$H_{nxq} = \tilde{D} (\mu B + \tilde{C}) = \begin{bmatrix} h_1' \\ \vdots \\ h_n' \end{bmatrix},$$

la ecuación (3.12) puede escribirse como

$$tr\ H' H = 0$$

$$\Leftrightarrow tr \sum_{i=1}^{n} h_i h_i' = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} h_i' h_i = 0$$

$$\Leftrightarrow H = 0$$

$$\Leftrightarrow \tilde{D} (\mu B + \tilde{C}) = 0$$

$$\Leftrightarrow A' D (A \mu B + C) = 0$$

lo cual finaliza la demostración.

COROLARIO 3.15. Sea X_{nxp} una matriz aleatoria donde las columnas de $X' = (X_1, \ldots, X_n)$ son independientes tales que

$$X_i \sim N_p(\delta, \Sigma) \qquad i = 1, \ldots, n$$

y sea D_{nxn} una matriz simétrica de rango r. La variable $W = X'DX$ posee distribución Wishart con matriz de escala Σ y d grados de libertad si y sólo si las siguientes condiciones se satisfacen.

a) D es idempotente.

b) $\delta=0$ ó $D1=0$.

donde 1 es el vector de dimensión n con todas sus entradas iguales a uno.

DEMOSTRACION. Del teorema 3.15, haciendo $A=I_{nxn}$, $B=I_{pxc}$, $C=0$ se obtiene que la matriz $W=X'DX$ tiene distribución Wishart con matriz de escala Σ y r grados de libertad si y sólo si las siguientes condiciones se satisfacen

a) D es idempotente;

b) $D\mu=0$,

donde los n renglores de μ_{nxp} son iguales al vector δ'. Para demostrar el corolario, sólo resta por demostrar que la condición $D\mu=0$ es equivalente a la condición $\delta=0$ ó $D1=0$. Para demostrar esto, observamos que $D\mu=D1\delta'$ de donde se observa que si $\delta=0$ ó $D1=0$ se obtiene $D\mu=0$. De manera inversa se tiene que la condición $D1\delta'$ implica que $(1'D'D1)(\delta'\delta)=0$ lo cual implica que $\delta=0$ ó $D1=0$.

3.6.6 Distribución e Independencia de Algunas Funciones de Matrices con Densidad Wishart

TEOREMA 3.16. Sea $U\sim W_p(\Sigma,m)$ $V\sim W_p(\Sigma,n)$ donde $m\geq p$, tales que U y V tienen distribución independiente. Sean particiones de las matrices U, V, Σ y $T=U+V$ definidas por

$$U=\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, \qquad V=\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

$$T=U+V=\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \quad y \quad \Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

donde U_{11}, V_{11}, T_{11} y Σ_{11} son matrices de dimensión qxq. Definiendo las matrices

$$U_{22.1}=U_{22}-U_{21}U_{11}^{-1}U_{12},$$
$$T_{22.1}=T_{22}-T_{21}T_{11}^{-1}T_{12},$$
$$D_{22.1}=T_{22.1}-U_{22.1},$$

y

$$\Sigma_{22.1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},$$

las siguientes propiedades se satisfacen

a) $U_{22.1} \sim W_{p-q}(\Sigma_{22.1}, m-q)$ y es independiente de (U_{11}, U_{12}),
b) Las variables $U_{22.1}$ y $D_{22.1}$ tienen distribución independiente y
c) $D_{22.1} \sim W_{p-q}(\Sigma_{22.1}, n)$.

DEMOSTRACION.

a) Como $U \sim W_p(\Sigma, m)$, puede escribirse en términos de una matriz X con observaciones independientes de una distribución $N_p(0,\Sigma)$ como sigue

$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} = X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix},$$

donde $X=(X_1,X_2)$. La matriz $U_{22.1}$ a su vez puede expresarse en términos de las submatrices de X'X como

$$\begin{aligned} U_{22.1} &= X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 \\ &= X_2'[I - X_1(X_1'X_1)^{-1}X_1']X_2 \\ &= X_2'CX_2, \end{aligned}$$

donde $C=I-X_1(X_1'X_1)^{-1}X_1'$. El producto $X_1'C$ es cero y por tanto $U_{22.1}=X_{2.1}'CX_{2.1}$ donde $X_{2.1}=X_2-X_1\Sigma_{11}^{-1}\Sigma_{12}$. La matrices X_1 y $X_{2.1}$ en términos de X pueden escribirse como

$$X_1 = XA = (X_1,X_2)\begin{bmatrix} I \\ 0 \end{bmatrix}$$

y

$$X_{2.1} = XB = (X_1,X_2)\begin{bmatrix} -\Sigma_{11}^{-1}\Sigma_{12} \\ I \end{bmatrix}.$$

La matriz $A'\Sigma B$ es la matriz nula, ya que

$$\begin{aligned} A'\Sigma B &= (I,0)\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\begin{bmatrix} -\Sigma_{11}^{-1}\Sigma_{12} \\ I \end{bmatrix} \\ &= (\Sigma_{11},\Sigma_{12})\begin{bmatrix} -\Sigma_{11}^{-1}\Sigma_{12} \\ I \end{bmatrix} \\ &= 0 \end{aligned}$$

y por tanto del teorema 3.13 se deduce que X_1 y $X_{2.1}$ son independientes. Como X es una matriz de datos de una distribución $N(0,\Sigma)$, X_1 es una matriz de datos de una distribución $N_q(\mu_1,\Sigma_{11})$ y utilizando el teorema 3.12 se puede comprobar que $X_{2.1}$ es una matriz de datos de una distribución $N_{p-q}(0,\Sigma_{22.1})$. Como $X_{2.1}$ es independiente de X_1 se obtiene que condicionando a X_1 , $X_{2.1}$ es una matriz de datos de la distribución $N_{p-q}(0,\Sigma_{22.1})$ y además dada X_1, C es una matriz de constantes que satisface $C^2=C$, $C=C'$ y su rango coincide con $r=rango(C)=tr(C)=n-q$ por lo que utilizando el corolario 3.15 se deduce que dada X_1

$$U_{22.1}= X'_2 C X_2 \sim W_{p-q}(\Sigma_{22.1},n-p_1).$$

Como la distribución de $U_{22.1}$ condicionada a X_1 no depende de X_1, se sigue que $U_{22.1}$ es independiente de X_1 y en consecuencia también de $U_{11}=X'_1X_1$. Para demostrar la independencia entre $U_{22.1}$ y U_{12} basta utilizar el hecho de que nuevamente condicionando a X_1 fija las matrices C y I-C son matrices constantes que satisfacen $C(I-C)'=0$, por lo que utilizando el teorema 3.13 se obtiene que dada X_1, las matrices $CX_{2.1}$ y $(I-C)X_{2.1}$ tienen distribución independiente de donde se sigue que dada X_1, las matrices $U_{22.1}=(CX_{2.1})'(CX_{2.1})$ y $(I-C)X_{2.1}$ tienen distribución independiente. Como $U_{22.1}$ es independiente de X_1, se obtiene que $U_{22.1}$ es independiente de $\{X_1,(I-C)X_{2.1}\}$. Más aún, puesto que $U_{11} =X'_1X_1$ y $U_{12}=X'_1[(I-C)X_{2.1}+X_1\ \Sigma_{11}^{-1}\ \Sigma_{12}\ ]=X'_1X_2$ se obtiene que $U_{22.1}$ es independiente de (U_{11},U_{12}).

b) La matriz $D_{22.1}$ puede escribirse como

$$D_{22.1}= U_{22}+V_{22}-(U_{21}+V_{21})(U_{11}+V_{11})^{-1}(U_{12}+V_{12})-[U_{22}- U_{21}U_{11}^{-1}U_{12}]$$

$$= V_{22}+U_{21}U_{11}^{-1}U_{12}-(U_{21}+V_{21})(U_{11}+V_{11})^{-1}(U_{12}+V_{12}).$$

Del inciso a se obtiene que $U_{22.1}$ tiene distribución independiente de U_{11} y U_{12}. Utilizando la independencia entre las matrices U y V y observando que $U_{12}=U'_{21}$ se sigue la independencia entre $U_{22.1}$ y $D_{22.1}$ por ser $D_{22.1}$ función de matrices con distribución independiente de $U_{22.1}$

c) Como U y V tienen distribuciones Wishart independientes, por una aplicación del teorema 3.9 se deduce que la función característica de la matriz T está dada por

$$\varphi_T(\theta)= E\{exp(\iota tr\,\theta'T\}$$

$$= E\{exp(\iota tr\,\theta'U+\iota tr\,\theta'V)\}$$

$$= E\{exp(\iota tr\,\theta'U\} E\{exp(\iota tr\,\theta'V)\}$$

$$= \varphi_U(\theta)\ \varphi_V(\theta)$$

$$= |I-2\lambda\theta\Sigma|^{-(n+m)/2}.$$

Del teorema 3.9 se obtiene que la función característica de T corresponde a la de una variable con densidad $W_p(\Sigma,n+m)$ y dada la unicidad de la función característica se deduce que $T\sim W_p(\Sigma,n+m)$. Por una aplicación del inciso a se sigue que

$$T_{22.1}\sim W_{p-q}(\Sigma_{22.1}, n+m-q). \tag{3.13}$$

Utilizando el inciso b se obtiene

$$\varphi_{T_{22.1}}(\eta)= E\{exp(\lambda tr\eta' T_{22.1})\}$$

$$= E\{exp(\lambda tr\eta' U_{22.1}+\lambda tr\eta' D_{22.1})\}$$

$$= \varphi_{U_{22.1}}(\eta)\,\varphi_{D_{22.1}}(\eta),$$

de donde

$$\varphi_{D_{22.1}}(\eta)= \varphi_{T_{22.1}}(\eta)\ \left(\varphi_{U_{22.1}}(\eta)\right)^{-1}.$$

Del inciso a y (3.13) se sigue que

$$\varphi_{D_{22.1}}(\eta)= |I-2\lambda\eta\Sigma_{22.1}|^{-n/2}.$$

Del teorema 3.9 y dada la unicidad de la función característica se deduce finalmente que $D_{22.1}\sim W_{p-q}(\Sigma_{22.1}, n)$.

3.6.7 Un Resultado sobre la Distribución Λ de Wilks

TEOREMA 3.17. Sea $U\sim W_p(\Sigma,m)$, $V\sim W_p(\Sigma,n)$ donde $m\geq p$, tales que U y V tienen distribución independiente. Sean U, V y T matrices con particiones definidas por

$$U=\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22}\end{pmatrix},\qquad V=\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22}\end{pmatrix}$$

y

$$T= U+V=\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22}\end{pmatrix},$$

donde U_{11}, V_{11} y T_{11} son matrices de dimensión qxq. Sean

$$U_{22.1}= U_{22}-U_{21}U_{11}^{-1}U_{12}$$

y

$$T_{22.1}= T_{22}-T_{21}T_{11}^{-1}T_{12}.$$

La variable

$$\Lambda = \frac{|U_{22.1}|}{|T_{22.1}|}$$

tiene distribución Λ de Wilks con parámetros p-q, m-q y n.

DEMOSTRACION. La variable Λ puede escribirse como

$$\Lambda = \frac{|U_{22.1}|}{|U_{22.1}+D_{22.1}|},$$

donde $D_{22.1} = T_{22.1} - U_{22.1}$. Por una aplicación del teorema 3.16b se tiene que las variables $U_{22.1}$ y $D_{22.1}$ tienen distribución independiente. Del teorema 3.16 incisos a y c se deduce que

$$U_{22.1} \sim W_{p-q}(\Sigma_{22.1}, m-q)$$

y

$$D_{22.1} \sim W_{p-q}(\Sigma_{22.1}, n).$$

Del teorema A.14 se deduce que existe $\Sigma_{22.1}^{-\frac{1}{2}}$ y del teorema 3.11 se sigue que

$$\tilde{U}_{22.1} = \Sigma_{22.1}^{-\frac{1}{2}}\ U_{22.1}\ \Sigma_{22.1}^{-\frac{1}{2}} \sim W_{p-q}(I, m-q)$$

y

$$\tilde{D}_{22.1} = \Sigma_{22.1}^{-\frac{1}{2}}\ D_{22.1}\ \Sigma_{22.1}^{-\frac{1}{2}} \sim W_{p-q}(I, n),$$

donde $\tilde{U}_{22.1}$ y $\tilde{D}_{22.1}$ son tambien independientes. Dado que m-q≥p-q, de la definición 3.3 se deduce que la variable

$$\frac{|\tilde{U}_{22.1}|}{|\tilde{U}_{22.1}+\tilde{D}_{22.1}|}$$

tiene distribución $\Lambda(p-q, m-q, n)$. Observando que

$$\frac{|\tilde{U}_{22.1}|}{|\tilde{U}_{22.1}+\tilde{D}_{22.1}|} = \frac{|\Sigma_{22.1}^{-\frac{1}{2}}\ U_{22.1}\ \Sigma_{22.1}^{-\frac{1}{2}}|}{|\Sigma_{22.1}^{-\frac{1}{2}}(U_{22.1}+D_{22.1})\Sigma_{22.1}^{-\frac{1}{2}}|}$$

$$= \frac{|U_{22.1}|}{|U_{22.1}+D_{22.1}|}$$

$$= \Lambda,$$

se sigue el resultado.

3.7 Teoremas Relacionados con Estimación y Pruebas de Hipótesis en la Densidad Normal Multivariada

En esta sección se presenta un par de teoremas que resultan de utilidad en la estimación de parámetros de la distribución normal multivariada y en la obtención del cociente de verosimilitudes generalizado asociado a la prueba de una hipótesis lineal general bajo el modelo generalizado de Potthoff & Roy.

3.7.1 Un Teorema sobre Estimación de Parámetros en una Generalización de la Densidad Normal Multivariada

TEOREMA 3.18. Sea X una matriz de dimensión nxp y $f(X;\mu,\Sigma)$ definida por

$$f(X;\mu,\Sigma)= |\Sigma|^{-n/2} \quad exp\left\{-\tfrac{1}{2}tr\Sigma^{-1}(X-\mu)'A(X-\mu)\right\},$$

donde A y Σ son matrices definidas postitivas y simétricas. Sean particiones de las matrices A, X, μ y Σ definidas por

$$A_{nxn}=\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22}\end{pmatrix}, \qquad X=\begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22}\end{pmatrix}=(X_1, X_2),$$

$$\mu=\begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22}\end{pmatrix}=(\mu_1,\mu_2) \quad y \quad \Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},$$

donde A_{11} es una matriz cuadrada de orden m, X_{11} y μ_{11} son matrices de dimensión mxq, X_1 y μ_1 son matrices de dimensión nxq y Σ_{11} es una matriz cuadrada de orden q. El supremo de $f(X;\mu,\Sigma)$ sobre Σ y μ_{12} está dado por

$$\sup_{\mu_{12},\Sigma} f(X;\mu,\Sigma)= n^{np/2}\, exp\left\{-\frac{np}{2}\right\} |(X_1-\mu_1)'A(X_1-\mu_1)|^{-n/2}$$

$$|(X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22})-(X_{22}-\mu_{22})'A_{22}(X_{21}-\mu_{21})[(X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21})]^{-1}(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})|^{-n/2}$$

y se alcanza en

$$\hat{\mu}_{12}= X_{12}-(X_{11}-\mu_{11})[(X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21})]^{-1}(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})$$

y

$$\hat{\Sigma}= \frac{1}{n}\begin{bmatrix} (X_1-\mu_1)'A(X_1-\mu_1) & (X_1-\mu_1)'A(X_2-\hat{\mu}_2) \\ (X_2-\hat{\mu}_2)'A(X_1-\mu_1) & (X_2-\hat{\mu}_2)'A(X_2-\hat{\mu}_2)\end{bmatrix},$$

donde $\hat{\mu}_2' = (\hat{\mu}_{12}' \;\; \mu_{22}')$.

DEMOSTRACION. Sea $\Lambda=\Sigma^{-1}$. El doble del logaritmo natural de $f(X;\mu,\Sigma)$ está dado por

$$h(X;\mu,\Lambda)= 2\,ln\, f(X;\mu,\Sigma)= n\,ln|\Lambda|- tr\Lambda(X-\mu)'A(X-\mu).$$

La derivada de $h(X;\mu,\Lambda)$ respecto a Λ puede obtenerse del teorema A.25 incisos d y e y está dada por

$$\frac{\partial h(X;\mu,\Lambda)}{\partial\Lambda} = n\{2\Lambda^{-1}-\mathrm{Diag}\,\Lambda^{-1}\}-\{2(X-\mu)'A(X-\mu)-\mathrm{Diag}(X-\mu)'A(X-\mu)\}. \tag{3.14}$$

La función $h(X;\mu,\Lambda)$ puede expresarse particionando Λ de manera análoga a Σ como

$$h(X;\mu,\Lambda) = n\,ln|\Lambda|-tr\begin{pmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{pmatrix}\begin{bmatrix}(X_1-\mu_1)'\\ (X_2-\mu_2)'\end{bmatrix}A\begin{bmatrix}X_1-\mu_1, & X_2-\mu_2\end{bmatrix}$$

$$= n\,ln|\Lambda|-tr\begin{pmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{pmatrix}\begin{bmatrix}(X_1-\mu_1)'A(X_1-\mu_1) & (X_1-\mu_1)'A(X_2-\mu_2)\\ (X_2-\mu_2)'A(X_1-\mu_1) & (X_2-\mu_2)'A(X_2-\mu_2)\end{bmatrix}$$

$$= n\,ln|\Lambda|-tr\,\Lambda_{11}(X_1-\mu_1)'A(X_1-\mu_1)-tr\Lambda_{22}(X_2-\mu_2)'A(X_2-\mu_2)-2tr\Lambda_{21}(X_1-\mu_1)'A(X_2-\mu_2)$$

$$= n\,ln|\Lambda|-tr\,\Lambda_{11}(X_1-\mu_1)'A(X_1-\mu_1)$$

$$-tr\Lambda_{22}[(X_{12}-\mu_{12})',(X_{22}-\mu_{22})']\begin{pmatrix}A_{11} & 0\\ 0 & A_{22}\end{pmatrix}\begin{pmatrix}X_{12}-\mu_{12}\\ X_{22}-\mu_{22}\end{pmatrix}$$

$$-2tr\Lambda_{21}[(X_{11}-\mu_{11})',(X_{21}-\mu_{21})']\begin{pmatrix}A_{11} & 0\\ 0 & A_{22}\end{pmatrix}\begin{pmatrix}X_{12}-\mu_{12}\\ X_{22}-\mu_{22}\end{pmatrix}$$

$$= n\,ln|\Lambda|-tr\,\Lambda_{11}(X_1-\mu_1)'A(X_1-\mu_1)$$

$$-tr\Lambda_{22}[(X_{12}-\mu_{12})'A_{11}(X_{12}-\mu_{12})+(X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22})]$$

$$-2tr\Lambda_{21}[(X_{11}-\mu_{11})'A_{11}(X_{12}-\mu_{12})+(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})],$$

de manera que la derivada de $h(X;\mu,\Lambda)$ respecto a μ_{12} puede obtenerse utilizando los teoremas A.25 y A.26a y está dada por

$$\frac{\partial h(X;\mu,\Lambda)}{\partial \mu_{12}} = -2A_{11}(X_{11}-\mu_{11})\Lambda_{12} - 2A_{11}(X_{12}-\mu_{12})\Lambda_{22}. \qquad (3.15)$$

Igualando a cero las ecuaciones (3.14) y (3.15) se obtiene

$$\hat{\Lambda}^{-1} = \hat{\Sigma} = \frac{1}{n}(X-\hat{\mu})'A(X-\hat{\mu}) \qquad (3.16)$$

y

$$X_{12}-\hat{\mu}_{12} = (X_{11}-\mu_{11})\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}. \qquad (3.17)$$

Substituyendo estos valores en $f(X;\mu,\Sigma)$ se tiene que

$$f(X;\hat{\mu}_{12},\hat{\Sigma}) = |\hat{\Sigma}|^{-n/2} exp\left\{-\frac{n}{2}\, tr\, I_{pxp}\right\}$$

$$= |\hat{\Sigma}_{11}|^{-n/2}|\hat{\Sigma}_{22}-\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}|^{-n/2} exp\left\{-\frac{np}{2}\right\}. \qquad (3.18)$$

La matriz $(X-\mu)'A(X-\mu)$ puede expresarse como

$$(X-\mu)'A(X-\mu) = \begin{bmatrix} (X_1-\mu_1)'A(X_1-\mu_1) & (X_1-\mu_1)'A(X_2-\mu_2) \\ (X_2-\mu_2)'A(X_1-\mu_1) & (X_2-\mu_2)'A(X_2-\mu_2) \end{bmatrix}$$

$$= \begin{bmatrix} (X_{11}-\mu_{11})'A_{11}(X_{11}-\mu_{11}) & (X_{11}-\mu_{11})'A_{11}(X_{12}-\mu_{12}) \\ (X_{12}-\mu_{12})'A_{11}(X_{11}-\mu_{11}) & (X_{12}-\mu_{12})'A_{11}(X_{12}-\mu_{12}) \end{bmatrix}$$

$$+ \begin{bmatrix} (X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21}) & (X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22}) \\ (X_{22}-\mu_{22})'A_{22}(X_{21}-\mu_{21}) & (X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22}) \end{bmatrix}.$$

Utilizando (3.16) se sigue que

$$n\hat{\Sigma}_{12} = (X_{11}-\mu_{11})'A_{11}(X_{12}-\hat{\mu}_{12}) + (X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})$$

y substituyendo el valor de $X_{12}-\hat{\mu}_{12}$ definido en (3.17) se tiene

$$n\hat{\Sigma}_{12} = (X_{11}-\mu_{11})'A_{11}(X_{11}-\mu_{11})\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12} + (X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})$$

$$= \left[n\hat{\Sigma}_{11}-(X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21})\right]\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12} + (X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})$$

$$= n\hat{\Sigma}_{12}-(X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21})\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12} + (X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22}).$$

de donde se sigue que

$$\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}=[(X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21})]^{-1}(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22}). \tag{3.19}$$

Substituyendo $\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$ en la ecuación (3.17) se tiene

$$\hat{\mu}_{12}= X_{12}-(X_{11}-\mu_{11})[(X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21})]^{-1}(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22}) \tag{3.20}$$

y substituyendo este valor en (3.16) se sigue que

$$\hat{\Sigma}=\frac{1}{n}\begin{bmatrix}(X_1-\mu_1)'A(X_1-\mu_1) & (X_1-\mu_1)'A(X_2-\hat{\mu}_2)\\ (X_2-\hat{\mu}_2)'A(X_1-\mu_1) & (X_2-\hat{\mu}_2)'A(X_2-\hat{\mu}_2)\end{bmatrix}, \tag{3.21}$$

donde $\hat{\mu}_2'=(\hat{\mu}_{12}' \quad \mu_{22}')$.

De (3.16) se tiene que

$$n\hat{\Sigma}_{22}=(X_{12}-\hat{\mu}_{12})'A_{11}(X_{12}-\hat{\mu}_{12})+(X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22})$$

y utilizando (3.17) se deduce que

$$\begin{aligned}n\hat{\Sigma}_{22}&=\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}(X_{11}-\mu_{11})'A_{11}(X_{11}-\mu_{11})\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}+(X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22})\\ &=\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\left[n\hat{\Sigma}_{12}-(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})\right]+(X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22})\\ &=n\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}-\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})+(X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22}).\end{aligned}$$

De la ecuación anterior y la igualdad (3.19) se deduce que

$$\begin{aligned}n\left[\hat{\Sigma}_{22}-\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}\right]&=(X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22})\\ &\quad-(X_{22}-\mu_{22})'A_{22}(X_{21}-\mu_{21})[(X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21})]^{-1}(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})\end{aligned}$$

y substituyendo en (3.18) se deduce finalmente que

$$\begin{aligned}f(X;\hat{\mu}_{12},\hat{\Sigma})&=n^{np/2}\,exp\left\{-\frac{np}{2}\right\}|(X_1-\mu_1)'A(X_1-\mu_1)|^{-n/2}\\ &|(X_{22}-\mu_{22})'A_{22}(X_{22}-\mu_{22})-(X_{22}-\mu_{22})'A_{22}(X_{21}-\mu_{21})[(X_{21}-\mu_{21})'A_{22}(X_{21}-\mu_{21})]^{-1}(X_{21}-\mu_{21})'A_{22}(X_{22}-\mu_{22})|^{-n/2}.\end{aligned} \tag{3.22}$$

Para demostrar que los puntos que anulan la primera derivada $\hat{\mu}_{12}$ y $\hat{\Sigma}$ definidos respectivamente en (3.20) y (3.21) son máximo de $f(X:\mu,\Sigma)$ basta demostrar que

$$\frac{f(X;\hat{\mu}_{12},\hat{\Sigma})}{f(X;\mu_{12},\Sigma)} \geq 1 \qquad \forall \mu_{12} \text{ y } \Sigma>0,$$

o equivalentemente

$$\ln f(X;\hat{\mu}_{12},\hat{\Sigma}) - \ln f(X;\mu_{12},\Sigma) \geq 0 \qquad \forall \mu_{12} \text{ y } \Sigma>0,$$

para lo cual se obtiene

$$\begin{aligned} A &= f(X;\hat{\mu}_{12},\hat{\Sigma}) - \ln f(X;\mu_{12},\Sigma) \\ &= -\tfrac{n}{2}\ln|\hat{\Sigma}| - \tfrac{1}{2}\, tr\,\hat{\Sigma}^{-1}(X-\hat{\mu})'A(X-\hat{\mu}) + \tfrac{n}{2}\ln|\Sigma| + \tfrac{1}{2}\, tr\,\Sigma^{-1}(X-\mu)'A(X-\mu) \\ &= -\tfrac{n}{2}\ln|\Sigma^{-1}\hat{\Sigma}| - \tfrac{np}{2} + \tfrac{n}{2}\, tr(\Sigma^{-1}\hat{\Sigma}) \\ &= \tfrac{np}{2}\left\{\tfrac{1}{p}\, tr(\Sigma^{-1}\hat{\Sigma}) - 1 - \ln|\Sigma^{-1}\hat{\Sigma}|^{1/p}\right\}. \end{aligned}$$

Denotando por $\lambda_1,\ldots,\lambda_p$ los valores propios de $\Sigma^{-1}\hat{\Sigma}$ se sigue que

$$A = \tfrac{np}{2}\{\delta_1 - 1 - \ln\delta_2\},$$

donde

$$\delta_1 = \frac{1}{p}\sum_{i=1}^{p}\lambda_i$$

y

$$\delta_2 = \left(\prod_{i=1}^{p}\lambda_i\right)^{1/p};$$

es decir, δ_1 y δ_2 son respectivamente las medias aritmética y geométrica de $\lambda_1,\ldots,\lambda_p$. Observando que $\hat{\Sigma}\geq 0$, del teorema A.15 se obtiene que los valores propios de $\Sigma^{-1}\hat{\Sigma}$ son los mismos que los de $W=\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2}$ y dado que W es semidefinida positiva se sigue que $\lambda_i\geq 0$ $i=1,\ldots,p$, de donde δ_1 y δ_2 son mayores o iguales a cero. Por una aplicación del teorema A.29 se sigue que

$$A \geq \tfrac{np}{2}\{\delta_1 - 1 - \ln\delta_1\}$$

y dado que la función $f(x)=x-1-\ln x\geq 0$ $\forall x>0$ se deduce finalmente que $A\geq 0$ por lo que los puntos $\hat{\mu}_{12}$ y $\hat{\Sigma}$ definidos en (3.20) y (3.21) maximizan $f(X;\mu_{12},\Sigma)$ y su máximo está dado en (3.22).

3.7.2 Un Teorema Relativo al Cálculo del Cociente de Verosimilitudes Generalizado para Hipótesis sobre la Media de Observaciones Normales Multivariadas

TEOREMA 3.19. Sea Z una matriz aleatoria con función de densidad dada por

$$f(Z)=|2\pi\Sigma|^{-n/2}|A|^{p/2}\, exp\left\{-\tfrac{1}{2}\, tr\Sigma^{-1}(Z-\mu)'A(Z-\mu)\right\},$$

donde $A>0$, $\Sigma>0$, A y Σ simétricas con

$$Z=\begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix}=(Z_1,Z_2), \qquad \mu=\begin{pmatrix} \mu_{11} & \mu_{12} \\ 0 & 0 \end{pmatrix}=(\mu_1,\mu_2)$$

y

$$A=\begin{pmatrix} A_{11\, m\times m} & 0 \\ 0 & A_{22} \end{pmatrix}.$$

El cociente de verosimilitudes generalizado para el constraste

$$H_0:\mu_{11}\in\omega$$
vs.
$$H_1:\mu_{11}\in\Omega$$

dado por

$$\Lambda=\frac{\sup_{H_0} f(Z)}{\sup_{H_0\cup H_1} f(Z)},$$

satisface la igualdad

$$\Lambda=\frac{\sup_{H_0} f(Z_1)}{\sup_{H_0\cup H_1} f(Z_1)}.$$

DEMOSTRACION. El supremo de $f(Z)$ sobre μ_{12} y Σ se obtiene utilizando el teorema 3.18 y está dado por

$$\sup_{\mu_{12},\Sigma} f(Z)=(2\pi)^{-np/2}|A|^{p/2}n^{np/2}exp\left\{-\tfrac{np}{2}\right\}$$
$$|(Z_1-\mu_1)'A(Z_1-\mu_1)|^{-n/2}|Z_{22}'A_{22}Z_{22}-Z_{22}'A_{22}Z_{21}(Z_{21}'A_{22}Z_{21})^{-1}Z_{21}A_{22}Z_{22}|^{-n/2},$$

de manera que

$$\Lambda=\frac{\sup_{\mu_{11}\in\omega}\ \sup_{\mu_{12},\Sigma} f(Z)}{\sup_{\mu_{11}\in\Omega}\ \sup_{\mu_{12},\Sigma} f(Z)}$$

$$= \frac{\sup_{\mu_{11}\varepsilon\omega} |(Z_1-\mu_1)'A(Z_1-\mu_1)|^{-n/2}}{\sup_{\mu_{11}\varepsilon\Omega} |(Z_1-\mu_1)'A(Z_1-\mu_1)|^{-n/2}}.$$

La función de densidad marginal de Z_1 puede calcularse de la manera siguiente:

Sea $W= A^{\frac{1}{2}} Z$ y $\theta=A^{\frac{1}{2}}\mu$. La función de densidad de W está dada por

$$f_w(W)=f_z(A^{-\frac{1}{2}}W)|A|^{-p/2}$$

$$=|2\pi\Sigma|^{-n/2} exp\left\{-\tfrac{1}{2}\, tr\Sigma^{-1}(W-\theta)'(W-\theta)\right\},$$

la cual corresponde a la densidad de una matriz con renglones independientes de distribución Normal. Definiendo $W=(W_{1q},W_{2p-q})$ y $\theta=(\theta_1,\theta_2)$, se deduce que los renglones de W_1 tienen densidad Normal con medias dadas por la relación $E(W_1)=\theta_1$ y matriz de covarianzas Σ_{11}, por lo que la función de densidad de W_1 está dada por

$$f(W)=|2\pi\Sigma_{11}|^{-n/2} exp\left\{-\tfrac{1}{2}\, tr\Sigma^{-1}(W_1-\theta_1)'(W_1-\theta_1)\right\}.$$

Como $Z_1=A^{-\frac{1}{2}}W_1$, la función de densidad de Z_1 resulta

$$f_{z_1}(Z_1)= f_{w_1}(A^{\frac{1}{2}}Z_1)|A|^{-q/2}$$

$$=|2\pi\Sigma_{11}|^{-n/2}|A|^{-q/2} exp\left\{-\tfrac{1}{2}\, tr\Sigma_{11}^{-1}(Z_1-\mu_1)'A(Z_1-\mu_1)\right\}.$$

El supremo de $f_{z_1}(Z_1)$ sobre Σ_{11} puede obtenerse del teorema A.32 y está dado por

$$\sup_{\Sigma_{11}} f(Z_1)=|A|^{-q/2}(2\pi)^{-nq/2}\left|\tfrac{1}{n}(Z_1-\mu_1)'A(Z_1-\mu_1)\right|^{-n/2} exp\left\{-\tfrac{nq}{2}\right\}.$$

Observando que

$$\frac{\sup_{H_0} F(Z_1)}{\sup_{H_0\cup H_1} F(Z_1)} = \frac{\sup_{\mu_{11}\varepsilon\omega} |(Z_1-\mu_1)'A(Z_1-\mu_1)|^{-n/2}}{\sup_{\mu_{11}\varepsilon\Omega} |(Z_1-\mu_1)'A(Z_1-\mu_1)|^{-n/2}}$$

$$=\Lambda,$$

se sigue el resultado.

4

LA PRUEBA DE COCIENTE DE VEROSIMILITUDES GENERALIZADO BAJO EL MODELO GENERAL DE ANALISIS DE VARIANZA MULTIVARIADO

4 La Prueba de Cociente de Verosimilitudes Bajo el Modelo General de Análisis de Varianza Multivariado

En esta sección se discute el concepto de matrices estimables, se determina el estimador máximo verosímil de una matriz estimable, se plantea una reparametrización del modelo general de Análisis de Varianza Multivariado (2.2) formulando una hipótesis lineal general en términos de la reparametrización y finalmente se determina el cociente de verosimilitudes asociado a una hipótesis lineal general sobre parámetros estimables.

4.1 Las Matrices Estimables

El contenido de esta sección está dedicado a presentar una generalización del concepto de funciones lineales estimables (Searle, 1971 y Montgomery, 1976) y demostrar algunas de sus propiedades.

Por conveniencia se describe en esta sección el modelo general de Análisis de Varianza Multivariado descrito en la sección 2.2. El valor esperado de la matriz X, con renglones independientes de distribución Normal con matriz de covarianzas común Σ, está dado por

$$E(X)= D_{n\times g}\mu_{g\times q}E_{q\times p}, \tag{4.1}$$

donde $r(D)=r$, $r(E)=s$, $n>g$, $q<p$ y $\mu_{g\times q}$ es una matriz de parámetros independientes.

4.1.1 Definición y Propiedades

La generalización del concepto de funciones lineales estimables que se discute en esta sección resulta de utilidad en el análisis de modelos de rango incompleto y corresponde a una extensión del concepto de funciones estimables de los parámetros discutido en Searle (1971) y Montgomery (1976). Como se demostrará en la sección 4.1.2, cuando $r<g$ ó $s<q$ la matriz μ no posee un estimador de máxima verosimilitud único dado que el modelo se encuentra sobreparametrizado. De esta manera, resulta atractivo definir condiciones bajo las cuales a una matriz de parámetros pueda asociarsele un estimador máximo verosímil único. En el contexto de pruebas de hipótesis el concepto de matrices estimables resulta de utilidad ya que permite formalizar el planteamiento de la hipótesis lineal general sobre transformaciones lineales estimables de la matriz de parámetros $\mu_{g\times q}$.

En los siguientes resultados y definiciones se hacen referencia el modelo generalizado de Varianza Multivariado (4.1).

DEFINICION 4.1. Una matriz η_{axb} de transformaciones lineales de una matriz μ_{gxq} se define como

$$\eta_{axb} = A_{axg}\, \mu_{gxq}\, B_{qxb},$$

donde A y B son matrices de constantes conocidas.

DEFINICION 4.2. Una matriz de funciones lineales de los parámetros se define como estimable si es igual a alguna transformación lineal del valor esperado de la matriz de observaciones X; es decir, η_{axb} es estimable si existen matrices A y B de constantes conocidas tales que

$$\eta_{axb} = A_{axn} E(X_{nxp})\, B_{pxb} = A_{axn}\, D_{nxg}\, \mu_{gxq}\, E_{qxp}\, B_{pxb}.$$

La definición 4.2 no impone unicidad a las matrices A y B sino simplemente postula que existan. Dado que el operador esperanza es lineal, la condición que se pide para que una matriz η_{axb} sea estimable es que exista una transformación lineal de la matriz de observaciones X cuyo valor esperado sea η. El fin que se persigue al definir la condición de estimabilidad de una matriz η es producir bajo el modelo (4.1) un estimador máximo verosímil $\hat{\eta}$ de η que sea único. Esta propiedad de unicidad se discute en la sección 4.1.3.

Resulta conveniente en este punto señalar algunas propiedades de las matrices estimables.

a) El valor esperado de cualquier observación es estimable. Este hecho puede verificarse particionando a X_{nxp} en términos de sus observaciones como $X'=(X_1,\ldots,X_n)$ y observando que $E(X\ )=E(t_i'X)$ donde t_i es un vector columna con la unidad en la i-ésima entrada y ceros en las posiciones restantes.

b) Transformaciones lineales de matrices estimables son estimables. Para demostrar esta propiedad sea η_{axb} una matriz estimable; es decir, existen matrices M y N tales que $\eta_{axb}=ME(X)N$. Sea $\Upsilon=F\eta G$ una tranformación lineal de η. Para demostrar que Υ es estimable observamos que

$$\begin{aligned}\Upsilon &= F\eta G\\ &= FME(X)NG\\ &= E(AXB)\end{aligned}$$

donde A=FM y B=NG, lo que demuestra la propiedad.

c) Diversas combinaciones lineales de matrices estimables son estimables. Para mostrar este hecho, sean α_{axb} β_{cxb} y γ_{axd} matrices estimables de la forma

$$\alpha_{axb} = E(AXB),$$

$$\beta_{cxb} = E(CXB)$$

y

$$\gamma_{axd} = E(AXD).$$

Las matrices

$$\tilde{\alpha}_{exf} = \tilde{A}_{exa}\,\alpha_{axb}\,\tilde{B}_{bxf} + \tilde{C}_{exc}\,\beta_{cxb}\,\tilde{B}_{bxf}$$

y

$$\tilde{\beta}_{gxh} = \tilde{A}_{gxa}\,\alpha_{axb}\,\tilde{B}_{bxh} + \tilde{A}_{gxa}\,\gamma_{axd}\,\tilde{D}_{dxh}$$

son estimables ya que

$$\tilde{\alpha}_{exf} = E([\tilde{A}A + \tilde{C}C]XB\tilde{B})$$

y

$$\tilde{\beta}_{gxh} = E([\tilde{A}AX[B\tilde{B} + \tilde{C}C]).$$

d) Una matriz es estimable si solo si cada una de sus entradas es estimable. Para demostrar esta afirmación sea η_{axb} una matriz estimable. Por la propiedad b) se deduce que para $i=1,....,a$, $j=1,....,b$, la combinación $\eta_{ij} = a_i'\eta_{axb}b_j$ es estimable donde a_i y b_j son vectores columna con la unicidad en las entradas i y j y cero en otro lado. La implicación inversa puede demostrarse observando que si η_{ij} es estimable para todo valor de i,j entonces

$$\eta_{ij} = a_i'E(X)b_j \qquad i=1,....,a, \qquad j=1,....,b.$$

Definiendo $A'=(a_1,....,a_a)$ y $B=(B_1,....,b_b)$ se obtiene

$$\eta = AE(X)B,$$

lo cual demuestra la afirmación.

e) La forma de una matriz estimable. Para determinar la condición algebraica que satisface una matriz estimable sea η_{axb} una matriz de transformaciones lineales de los parámetros con la propiedad de ser estimable. Esto significa que existen matrices A, B, F y G tales que

$$\eta_{axb} = A_{axg}\,\mu_{gxq}\,B_{qxb} = F_{axn}E(X_{nxp})G_{pxb},$$

lo cual ocurre si y sólo si

$$A\mu B = FD\mu EG. \tag{4.2}$$

Dado que la estimabilidad es un concepto que no depende del particular valor de μ, la ecuación (4.2) debe satisfacerse para toda matriz μ por lo que del teorema A.31 se deduce que

$$A = FD \tag{4.3}$$

y

$$B = EG; \tag{4.4}$$

es decir, los renglones de la matriz A deben pertenecer al espacio generado por los renglones de D y las columnas de B deben pertenecer al espacio generado por las columnas de E.

De manera inversa, si las condiciones (4.3) y (4.4) se satisfacen la deducción de (4.2) es inmediata, por lo que la condición de estimabilidad está dada en forma equivalente por las ecuaciones (4.3) y (4.4); es decir, se dice que $A\mu B$ es estimable si existen matrices F y G tales que $A=FD$ y $B=EG$.

f) Las condiciones de estimabilidad. El determinar si una matriz $\eta_{axb}=A\mu B$ es estimable puede resultar complejo, ya que para A y B conocidas, debe determinarse la existencia de matrices F_{axn} y G_{pxb} que satisfagan las ecuaciones (4.3) y (4.4). Esta tarea puede resultar difícil especialmente cuando X tenga dimensiones grandes. Resulta de esta manera interesante establecer una forma mas simple para determinar la estimabilidad de una matriz. Para ésto, sean M y N inversas generalizadas de D'D y EE' y sean $\tilde{M}=MD'D$ y $\tilde{N}=EE'N$. Si las ecuaciones (4.3) y (4.4) se verifican entonces multiplicando apropiadamente se tiene

$$\begin{aligned} A\tilde{M} &= FD\tilde{M} \\ &= FDMD'D \\ &= FD \\ &= A \end{aligned}$$

y

$$\begin{aligned} \tilde{N}B &= \tilde{N}EG \\ &= EE'NEG \\ &= EG \\ &= B, \end{aligned}$$

de donde se sigue que

$$A\tilde{M} = A \tag{4.5}$$

y

$$\tilde{N}B = B. \tag{4.6}$$

De manera inversa, si las ecuaciones (4.5) y (4.6) se satisfacen, se deduce que

$$A = AMD'D = FD$$

y

$$B = EE'NB = EG$$

donde F=AMD' y G=E'NB; es decir, existen matrices F y G tales que las condiciones (4.3) y (4.4) se verifican. De esta forma puede determinarse con relativa facilidad si una matriz de transformaciones lineales $\eta_{a\times b}=A\mu B$ es estimable. Si las ecuaciones (4.5) y (4.6) se verifican, la matriz η es estimable y si las condiciones (4.5) y (4.6) no se satisfacen entones η no es estimable.

4.1.2 El Estimador Máximo Verosimil de una Matriz Estimable

Resulta atractivo como primer punto de esta sección discutir la utilidad de las matrices estimables. Como se demostrará en la sección 4.1.3, el estimador máximo verosímil de una matriz estimable es único, lo cual no ocurre con matrices que no son estimables; es decir, cuando una matriz η no es estimable existen diferentes matrices $\hat{\eta}$ que maximizan la función de verosimilitud de η. En general, en modelos de regresión y diseños experimentales, la estadística de prueba asociada a hipótesis lineales sobre los parámetros del modelo involucra la estimación de las restricciones impuestas por la hipótesis nula. Esta situación, como se discutirá mas adelante, se presenta en el caso del modelo generalizado de Análisis de Varianza. Cuando las restricciones impuestas a los parámetros bajo la hipótesis nula no pueden ser estimadas de manera única, el valor de la estadística de prueba puede alterarse dependiendo de la particular estimación de las restricciones seleccionada, lo cual produciría una prueba con una característica no deseable. Lo que ocurre en esta situación es que cuando la matriz D ó la matriz E no poseen rango completo, el modelo se encuentra sobreparametrizado y por esta razón no es posible construir un único estimador máximo verosímil para cualesquier parámetro en la matriz μ. Es interesante señalar que la situación anterior no implica que una hipótesis que involucre matrices no estimables no pueda tener una estadística de cociente de verosimilitudes cuyo valor sea único. Como ejemplo puede mencionarse el modelo de diseños experimentales con un criterio de clasificación dado por

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i=1,2 \quad j=1,2,\ldots\ldots,n_i,$$

donde ε_{ij} $i=1,2$ $j=1,2,\ldots\ldots,n_i$ son variables aleatorias independientes con distribución $N(0,\sigma^2)$. Definiendo $C=(1,-1)$ y $\tau'=(\tau_1,\tau_2)$, la hipótesis H_0:$C\tau=0$ puede ser contrastada con un valor único de la estadística de prueba mientras que el vector τ no es estimable; sin embargo, la misma hipótesis puede plantearse equivalentemente como H_0:$C\beta=0$ con $\beta'=(\mu+\tau_1,\mu+\tau_2)$ donde β es estimable.

Es interesante también notar que en el modelo generalizado de Análisis de Varianza Multivariado si D ó E no son de rango completo, la matriz μ no es estimable. Para demostrar esta afirmación de las condiciones de estimabilidad (4.3) y (4.4) deben encontrarse matrices A y B tales que

$$AD = I_g \tag{4.7}$$

y

$$EB = I_g . \tag{4.8}$$

Si r(D)=r<g entonces r(AD)<g para cualquier matriz A y por tanto la condición (4.7) no puede verificarse ya que la matriz identidad es de rango completo. De igual manera, si r(E)=s<q entonces r(EB)<q y por tanto (4.8) no puede verificarse. Naturalmente si D y E son matrices de rango completo, la matriz μ es estimable, ya que en este caso basta tomar $A=(D'D)^{-1}D'$ y $B=E'(EE')^{-1}$.

En el modelo general de Análisis de Varianza (4.1), es posible contrastar hipótesis lineales sobre una matriz con a lo más rxs parámetros independientes. Este es el caso de los modelos de diseños experimentales en los que la matriz E es la identidad, la matriz D es de rango incompleto y solo se pueden probar hipótesis sobre un concepto desarrollado referente a funciones lineales estimables de los parámetros (Searle, 1971 y Montgomery, 1976). El hecho de sólo poder contrastar hipótesis que involucren a lo más a un conjunto de rxs parámetros independientes radica en que en el modelo (4.1) a lo más pueden identificarse rxs parámetros. Este hecho puede verificarse de la siguiente manera.

De los teorema A.23 y A.24 puede escribirse

$$D_{n\times g} = R_{n\times n} \begin{pmatrix} I_r & O_{r\times(g-r)} \\ O_{(n-r)\times r} & O_{(n-r)\times(g-r)} \end{pmatrix} S_{g\times g}$$

y

$$E_{q\times p} = \tilde{S}_{q\times q} \begin{pmatrix} I_s & O_{s\times(p-s)} \\ O_{(q-s)\times s} & O_{(q-s)\times(p-s)} \end{pmatrix} \tilde{R}_{p\times p},$$

donde $R'R=I$, $\tilde{R}'\tilde{R}=I$, $r(S)=g$ y $r(\tilde{S})=q$. Definiendo $Y=R'X\tilde{R}'$ y $\eta_{g\times q}=S\,\mu\,\tilde{S}$ se obtiene que Y es una matriz con renglones independientes de distribución Normal con matriz de covarianzas $\Sigma_y=\tilde{R}\Sigma\tilde{R}'$ y medias dadas por la relación

$$E(Y) = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \eta \begin{pmatrix} I_s & 0 \\ 0 & 0 \end{pmatrix}.$$

Dado que η y μ son funciones uno a uno, constituyen parametrizaciones equivalentes del modelo. Definiendo la partición de η por

$$\eta = \begin{pmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{pmatrix},$$

donde η_{11} es una matriz de rxs parámetros se obtiene que

$$E(Y)=\begin{pmatrix} \eta_{12} & 0 \\ 0 & 0 \end{pmatrix};$$

es decir, en el valor esperado de Y sólo es posible identificar a η_{12}; es decir, a una matriz de rxs parámetros independientes. Dado que μ y η constituyen parametrizaciones equivalentes, se deduce por tanto que el valor esperado de X permite identificar el mismo número de parámetros, es decir rxs; entonces cuando las matrices D ó E no son de rango completo se deduce que el modelo se encuentra sobreparametrizado.

De esta manera, resulta conveniente determinar la forma de una matriz θ_{rxs} de transformaciones lineales de la matriz de parámetros μ_{gxq} que sea estimable, a partir de la cual la estimación de cualquier matriz estimable pueda generarse como una transformación lineal de ésta. Para ésto, utilizando los teoremas A.19 y A.22 las matrices D y E pueden expresarse como

$$D_{nxg}= P_{nxn}\begin{pmatrix} I_{rxr} \\ O_{(n-r)xr} \end{pmatrix} U_{rxg} \tag{4.9}$$

y

$$E_{qxp}= V_{qxs}(O_{sx(p-s)}, I_{sxs})\, Q'_{pxp}, \tag{4.10}$$

donde P'P=I Q'Q=I, r(U)=r y r(V)=s. Utilizando estas descomposiciones, la transformación lineal de la matriz μ se define por

$$\theta_{rxs}= U_{rxg}\, \mu_{gxq}\, V_{qxs} \quad \text{con} \quad r\leq g \text{ y } s\leq q. \tag{4.11}$$

Es importante mencionar que dado que μ es una matriz de parámetros independientes la matriz θ es una matriz de parámetros independientes, ya que r(U)=r y r(V)=s con $r\leq g$ y $s\leq q$. Dado que la elección de las matrices U y V no es única, pueden elegirse tratando de formar en la matriz θ parámetros de interés. Resulta también interesante observar que diferentes reparametrizaciones son equivalentes, ya que son función uno a uno como lo demuestra el siguiente teorema.

TEOREMA 4.1. Sean θ_1 y θ_2 dos reparametrizaciones de la matriz μ definidas por

$$\theta_1= U\mu V$$

y

$$\theta_2= \tilde{U}\mu\tilde{V},$$

donde

$$D_{nxg}= P_{nxn}\begin{pmatrix} I_{rxr} \\ O_{(n-r)xr}\end{pmatrix} U_{rxg} = \tilde{P}_{nxn}\begin{pmatrix} I_{rxr} \\ O_{(n-r)xr}\end{pmatrix} \tilde{U}_{rxg}$$

y

$$E_{qxp}= V_{qxs}(O_{sx(p-s)}, I_{sxs})\, Q'_{pxp} = \tilde{V}_{qxs}(O_{sx(p-s)}, I_{sxs})\, \tilde{Q}'_{pxp}.$$

La matriz θ_1 es una transformación lineal invertible de la matriz θ_2.

DEMOSTRACION. Definiendo las particiones $P=(P_1,P_2)$, $\tilde{P}=(\tilde{P}_1,\tilde{P}_2)$, $Q=(Q_1,Q_2)$ y $\tilde{Q}=(\tilde{Q}_1,\tilde{Q}_2)$ donde P_1 y $\tilde{P}_1$ tienen respectivamente r columnas y las matrices Q_2 y $\tilde{Q}_2$ s columnas se sigue que $D=P_1U$ y $E=VQ_2$. Utilizando estas descomposiciones se deduce que

$$U=A\tilde{U}$$

y

$$V=\tilde{V}B,$$

donde $A= P_1'\tilde{P}_1$ y $B=\tilde{Q}_2'Q_2$. Observando que

$$\begin{aligned}\theta_1 &= U\mu V\\ &= A\tilde{U}\mu\tilde{V}B\\ &= A\theta_2 B\end{aligned}$$

y dado que las matrices A y B son de rango completo, se sigue el resultado.

El teorema anterior resulta de utilidad ya que demuestra que dadas dos reparametrizaciones θ_1 y θ_2 de la matriz μ, es equivalente la estimación de θ_1 y θ_2, ya que a partir de una se obtiene la otra mediante una transformación lineal y por tanto las inferencias pueden obtenerse equivalentemente en términos de cualquier reparametrización.

Es importante observar que la matriz θ_{rxs} definida en (4.11) es estimable, ya que definiendo $P=(P_1,P_2)$ y $Q=(Q_1,Q_2)$ donde P_1 y Q_1 tienen respectivamente r y p-s columnas se obtiene

$$\begin{aligned}\theta_{rxs} &= U_{rxg}\, \mu_{gxq}\, V_{qxs}\\ &= P_1'P_1\, U\mu V\, Q_2'Q_2\end{aligned}$$

$$= P_1' P_{nxn} \begin{pmatrix} I_r \\ O_{(n-r)xr} \end{pmatrix} U \mu V \,(O_{sx(p-s)}, I_s)\, Q_{pxp}\, Q_2$$

$$= P_1' D \mu E Q_2'$$

$$= P_1' E(X) Q_2',$$

de donde se sigue la estimabilidad de la matriz θ.

El teorema siguiente resulta de utilidad tanto en la estimación de matrices estimables como en la formulación de hipótesis sobre matrices estimables.

TEOREMA 4.2. Sea θ_{rxs} definida como en (4.11). El espacio de matrices estimables es el espacio de transformaciones lineales de θ.

DEMOSTRACION. Sean particiones de las matrices P y Q definidas por $P_{nxn}=(P_{1nxr}, P_{2nx(n-r)})$, $Q_{pxp}=(Q_{1px(p-s)}, Q_{2pxs})$. Sea η_{axb} una matriz estimable; es decir, existen matrices A y B tales que

$$\eta_{axb} = A_{axn} E(X)_{nxp} B_{pxb}$$

$$= A_{axn} P_{nxn} \begin{pmatrix} I_r \\ O_{(n-r)xr} \end{pmatrix} U_{rxg}\, \mu_{gxq}\, V_{qxs}\, (O_{sx(p-s)}, I_s) Q'_{pxp}\, B_{pxb}$$

$$= A^*_{axr}\, \theta_{rxs}\, B^*_{sxb},$$

donde

$$A^*_{axr} = A_{axn}\, P_{1nxr}$$

y

$$B^*_{sxb} = Q'_{2sxp}\, B_{pxb},$$

lo cual implica que cualquier matriz estimable es una transformación lineal de θ. De manera inversa cualquier transformación lineal η de la matriz θ es estimable. Para demostrarlo puede utilizarse el hecho de que $P_1'P_1=I_r$ y $Q_2'Q_2=I_s$ y observar que

$$\eta_{axb} = A_{axr}\, \theta_{rxs}\, B_{sxb}$$

$$= A_{axr}\, U_{rxg}\, \mu_{gxq}\, V_{qxs}\, B_{sxb}$$

$$= A_{axr}\, P'_{1rxn} P_{1nxr}\, U_{rxg}\, \mu_{gxq}\, V_{qxs}\, Q'_{2sxp}\, Q_{2pxs}\, B_{sxb}$$

$$= A^*_{axn}\, P_{1nxr} U_{rxg}\, \mu_{gxq}\, V_{qxs}\, Q'_{2_{sxb}}\, B^*_{pxp},$$

donde

$$A^*_{axn} = A_{axr}\, P'_{1rxn}$$

y

$$B^*_{pxb} = Q_{2pxs}\, B_{sxb}.$$

Considerando que $D_{nxg}=P_1U$ y $E_{qxp}=VQ'_2$ se sigue que $\eta_{axb}=A^*E(X)B^*$ lo cual demuestra el teorema.

La importancia del teorema anterior radica en que el proceso de estimación y prueba de hipótesis sobre funciones estimables de la matriz μ_{gxq} pueden expresarse equivalentemente en términos de la matriz θ_{rxs} cuyas dimensiones son a lo más gxq. Esta propiedad resulta interesante dado que, como se discutió, si r<g ó s<q el modelo generalizado de Análisis de Varianza Multivariado está sobreparametrizado; es decir, existen mas parámetros de los que se pueden estimar de manera única, lo cual no ocurre con la matriz θ ya que ésta, como se discutirá en la sección 4.1.3, sí posee un único estimador de máxima verosimilitud por ser una matriz estimable.

Para determinar el estimador máximo verosímil de la matriz θ_{rxs} pueden utilizarse las descomposiciones de las matrices D y E definidas en (4.9) y (4.10) para definir la transformación Y=P'XQ, donde el valor esperado de la matriz Y está dado por

$$\eta = E(Y)$$

$$= E(P'XQ)$$

$$= P'P \begin{pmatrix} 0 & \theta_{rxs} \\ 0 & 0 \end{pmatrix} Q'Q,$$

de donde se sigue que

$$\eta = \begin{pmatrix} 0 & \theta_{rxs} \\ 0 & 0 \end{pmatrix} \tag{4.12}$$

y por el teorema de cambio de variable, la densidad de Y está dada por

$$f_y(Y;\eta,\Lambda) = |2\pi\Lambda|^{-n/2} exp\left\{-\tfrac{1}{2} tr\,\Lambda^{-1}(Y-\eta)'(Y-\eta)\right\},$$

donde $\Lambda=Q'\Sigma Q$. De la densidad anterior se deduce que Y es una matriz aleatoria cuyos renglones son independientes y tienen distribución Normal con medias dadas por la relación (4.12) y matriz de covarianzas común Λ.

Los estimadores máximo verosímiles de las matrices θ y Λ pueden obtenerse a partir de la densidad de Y, ya que es una transformación uno a uno de la matriz X. Definiendo particiones de la matriz Y por

$$Y_{nxp}=(Y_{1nx(p-s)},\, Y_{2nxs}),$$

donde

$$Y_{1nx(p-s)}=\begin{pmatrix} Y_{11rx(p-s)} \\ Y_{21(n-r)x(p-s)} \end{pmatrix} \quad y \quad Y_{2nxs}=\begin{pmatrix} Y_{12rxs} \\ Y_{22(n-r)xs} \end{pmatrix},$$

se obtiene por una aplicación del teorema 3.18 que los estimadores máximo verosímiles de las matrices θ y Λ están dados por

$$\hat{\theta}= Y_{12}-Y_{11}(Y_{21}'\, Y_{21})^{-1}\, Y_{21}'\, Y_{22} \tag{4.13}$$

y

$$\hat{\Lambda}=\frac{1}{n}\begin{pmatrix} Y_1' Y_1 & Y_1'(Y_2-\hat{\delta}) \\ (Y_2-\hat{\delta})' Y_1 & (Y_2-\hat{\delta})'(Y_2-\hat{\delta}) \end{pmatrix}, \tag{4.14}$$

donde

$$\hat{\delta}_{sxn}'=(\hat{\theta}_{sxr}',\, 0_{sx(n-r)}). \tag{4.15}$$

Definiendo $P=(P_1,P_2)$ y $Q=(Q_1,Q_2)$ donde P_1 y Q_1 tienen respectivamente r y p-s columnas se obtiene que

$$\begin{aligned} Y=P'XQ &= \begin{pmatrix} P_1' \\ P_2' \end{pmatrix} X(Q_1,Q_2) \\ &= \begin{pmatrix} P_1' XQ_1 & P_1' XQ_2 \\ P_2'XQ_1 & P_2'XQ_2 \end{pmatrix} \\ &= (P'XQ_1\, ,\, P'XQ_2), \end{aligned}$$

de donde

$$Y_{11}= P_1' XQ_1,$$

$$Y_{12} = P_1' XQ_2,$$

$$Y_{21} = P_2' XQ_1,$$

$$Y_{22} = P_2' XQ_2,$$

$$Y_1 = P' XQ_1$$

y

$$Y_2 = P' XQ_2,$$

por lo que los estimadores máximo verosímiles de θ y Σ pueden escribirse en términos de la información original X como

$$\hat{\theta} = P_1' XQ_2 - P_1' XQ_1(Q_1' X'P_2P_2' XQ_1)^{-1}Q_1'X'P_2P_2'XQ_2$$

y

$$\hat{\Sigma} = Q\hat{\Lambda}Q'.$$

Por una aplicación del Teorema A.30 se deduce que el estimador $\hat{\theta}$ puede escribirse como

$$\hat{\theta} = P_1' XW^{-1}Q_2[Q_2' W^{-1}Q_2]^{-1}, \tag{4.16}$$

donde

$$W = X'P_2P_2'X = X'(I-P_1P_1')X.$$

La estimación de la matriz Σ puede simplificarse, puesto que

$$n\hat{\Sigma} = Q \begin{pmatrix} Q_1'X' \ XQ_1 & Q_1'X' P[P'XQ_2-\hat{\delta}] \\ Q_2'X' P-\hat{\delta}')P'XQ_1 & [Q_2'X'P-\hat{\delta}'][P'XQ_2-\delta] \end{pmatrix} Q'$$

$$= Q_1Q_1'X'XQ_1Q_1' + Q_2(Q_2'X'P-\hat{\delta}')P'XQ_1Q_1'$$

$$+ Q_1Q_1' X'P(P'XQ_2-\hat{\delta})Q_2' + Q_2(Q_2'X'P-\hat{\delta}')(P'XQ_2-\hat{\delta})Q_2'$$

$$= (Q_1Q_1'+Q_2Q_2')X'XQ_1Q_1' - Q_2\hat{\delta}'P'XQ_1Q_1'$$

$$+ (Q_1Q_1'+Q_2Q_2')X'XQ_2Q_2' - Q_1Q_1'X'P\hat{\delta}Q_2'$$

$$- Q_2Q_2'X'P\hat{\delta}Q_2' - Q_2\hat{\delta}'P'XQ_2Q_2' + Q_2\hat{\delta}'\hat{\delta}Q_2'$$

$$= X'X(Q_1Q_1' + Q_2Q_2') - Q_2\hat{\delta}'P'X(Q_1Q_1' + Q_2Q_2')$$

$$- (Q_1Q_1' + Q_2Q_2')X'P\hat{\delta}Q_2' + Q_2\hat{\delta}'\hat{\delta}Q_2'$$

$$= X'X - Q_2\hat{\delta}'P'X - X'P\hat{\delta}Q_2' + Q_2\hat{\delta}'\hat{\delta}Q_2',$$

por lo que la matriz $\hat{\Sigma}$ puede escribirse como

$$\hat{\Sigma} = \frac{1}{n}(X - \hat{M})'(X - \hat{M}), \tag{4.17}$$

donde la matriz $\hat{M}$ es el estimador máximo verosimil del valor esperado de la matriz X; es decir,

$$\hat{M} = \hat{E}(X) = P_1\hat{\theta}Q_2'.$$

El estimador de máxima verosimilitud de una matriz estimable $\eta = AE(X)B$ puede obtenerse en función de $\hat{\theta}$ como

$$\eta_{axb} = A_{axn}E(X)_{nxp}B_{pxb}$$

$$= A_{axn}\, P_{nxn} \begin{pmatrix} I_r \\ O_{(n-r)xr} \end{pmatrix} U_{rxg}\, \mu_{gxq}\, V_{qxs}\, (O_{sx(p-s)}, I_s)\, Q'_{pxp}\, B_{pxb}$$

$$= A^*_{axr}\, \theta_{rxs}\, B^*_{sxb},$$

donde

$$A^*_{axr} = A_{axn}\, P_{1nxr}$$

y

$$B^*_{sxb} = Q'_{2sxp}\, B_{pxb}.$$

Por el principio de invarianza se tiene

$$\hat{\eta} = A^*_{axr}\, \hat{\theta}_{rxs}\, B^*_{sxb}, \tag{4.18}$$

donde $\hat{\theta}$ se define como en (4.16). Es importante mencionar que como se discutió con anterioridad, la elección de las matrices A y B no necesariamente es única; es decir, pueden existir diferentes matrices A y B tales que $\eta = AE(X)B$, por lo cual podría pensarse que la estimación $\hat{\eta}$ definida en (4.18) depende de la particular elección de las matrices A y B. En la sección 4.1.3 se demuestra la unicidad de la matriz $\hat{\eta}$, lo cual implica que el estimador máximo verosímil de η no depende de la particular elección de las matrices A y B bajo la condición $\eta = AE(X)B$.

4.1.3. Propiedades del Estimador Máximo Verosimil de una Matriz Estimable

El valor esperado de $\hat{\theta}$ puede obtenerse de (4.16). Dado que X es una matriz con renglones independientes se deduce del teorema 3.13 que las matrices $P_1'X$ y $P_2'X$ tienen distribución independiente, de donde se sigue la independencia entre $P_1'X$ y $W=X'P_2P_2'X$. Utilizando ésto y las descomposiciones de las matrices D y E definidas en (4.9) y (4.10) se sigue que

$$\begin{aligned}
E(\hat{\theta}) &= E[P_1'XW^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}] \\
&= E[P_1'X]\,E[W^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}] \\
&= P_1'D\mu E\; E[W^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}] \\
&= P_1'P_1U\mu VQ_2'\,E[W^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}] \\
&= \theta E[Q_2'W^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}] \\
&= \theta E[I] \\
&= \theta,
\end{aligned}$$

lo cual demuestra que $\hat{\theta}$ es un estimador insesgado de θ.

El estimador de la matriz estimable $\eta=AE(X)B$ definido en (4.18) dado por

$$\hat{\eta}=AP_1\hat{\theta}Q_2'B \tag{4.19}$$

es un estimador insesgado de η, ya que

$$\begin{aligned}
E(\hat{\eta}) = E(A^*\hat{\theta}B^*) = A^*\theta B^* &= AP_1U\mu VQ_2'B \\
&= AD\mu EB \\
&= AE(X)B \\
&= \eta.
\end{aligned}$$

Resulta interesante determinar si la matriz $\hat{\eta}$ no depende de la particular elección de A y B como podría esperarse, ya que la elección de éstas no es única para una matriz

η determinada. También resulta interesante cuestionarse si al elegir diferentes reparametrizaciones de la matriz μ en θ; es decir, diferentes matrices U y V, puede alterarse la estimación de una matriz estimable. El siguiente teorema responde a éstas preguntas.

TEOREMA 4.3. Sea η=AE(X)B una matriz estimable. El estimador máximo verosímil de η es único.

DEMOSTRACION. De (4.19) se tiene que el estimador máximo verosímil de η puede obtenerse mediante una transformación lineal del estimador máximo verosímil de la matriz θ definida en (4.11). De (4.16), el estimador de θ sólo depende de la información muestral X y de las matrices P_1 y Q_2. Dadas las matrices U y V se define de manera única la matriz θ y también las matrices P_1 y Q_2 dado que

$$P_1=DU'(UU')^{-1}$$

y

$$Q_2=(VV')^{-1}V'E,$$

de donde se sigue la unicidad de la matriz $\hat{\theta}$ definida en (4.16). De esta manera se deduce que cada reparametrización θ tiene asociado un único estimador máximo verosímil $\hat{\theta}$. Otra fuente de no unicidad respecto a $\hat{\theta}$ en la estimación de η es la particular elección de las matrices U y V. Para demostrar primero que la estimación de η no depende de la reparametrización seleccionada, se eligen dos diferentes reparametrizaciones θ_1 y θ_2 de la matriz μ. Para ésto, sean

$$D_{nxg} = P_{nxn}\begin{pmatrix} I_r \\ O_{(n-r)xr} \end{pmatrix} U_{rxg} = \tilde{P}_{nxn}\begin{pmatrix} I_r \\ O_{(n-r)xr} \end{pmatrix} \tilde{U}_{rxg},$$

$$E_{qxp} = V_{qxs}(O_{sx(p-s)}, I_s)\ Q_{pxp} = \tilde{V}_{qxs}(O_{sx(p-s)}, I_s)\ \tilde{Q}'_{pxp},$$

$$\theta_1 = U\mu V$$

y

$$\theta_2 = \tilde{U}\mu\tilde{V}.$$

Considérese también las particiones

$$P= (P_1,P_2),$$

$$Q= (Q_1,Q_2),$$

$$\tilde{P}=(\tilde{P}_1,\tilde{P}_2)$$

y

$$\tilde{Q}=(\tilde{Q}_1,\tilde{Q}_2),$$

donde $P'P=I$, $Q'Q=I$, $r(U)=r(\tilde{U})=r$, $R(V)=r(\tilde{V})=s$, las matrices P_1 y $\tilde{P}_1$ tienen r columnas y las matrices Q_2 y $\tilde{Q}_2$ tienen s columnas. De las igualdades anteriores se sigue que

$$\theta_1=M\theta_2N,$$

donde

$$M=P'_1\tilde{P}_1$$

y

$$N=\tilde{Q}'_2Q_2.$$

Como

$$D=P_1U=\tilde{P}_1\tilde{U}$$

y

$$E=VQ'_2=\tilde{V}\tilde{Q}'_2,$$

entonces

$$P_1=\tilde{P}_1\tilde{U}\,U'(UU')^{-1}$$

y

$$Q_2=\tilde{Q}_2\tilde{V}'V(VV')^{-1},$$

de donde

$$0=P'_2P_1=P'_2\tilde{P}_1\tilde{U}U'(UU')^{-1}$$

y

$$0=Q'_2Q_2=Q'_1Q_2\tilde{V}'V(VV')^{-1},$$

lo cual implica que $P'_2\tilde{P}_1=0$ y $Q'_1\tilde{Q}_2=0$ por ser $\tilde{U}U'$ y $\tilde{V}'V$ matrices de rango completo. Utilizando este resultado y dado que

$$I_n=PP'=P_1P'_1+P_2P'_2$$

y

$$I_p=QQ'=Q_1Q'_1+Q_2Q'_2,$$

se tiene

$$\begin{aligned}P_1\hat{\theta}_1Q'_2&=P_1M\hat{\theta}_2NQ'_2\\&=P_1P'_1\tilde{P}_1\hat{\theta}_2\tilde{Q}'_2Q_2Q'_2\\&=(I-P_2P'_2)\tilde{P}_1\hat{\theta}_2\tilde{Q}'_2(I-Q_1Q'_1)\end{aligned}$$

$$=\tilde{P}_1\hat{\theta}_2\tilde{Q}_2',$$

lo cual demuestra que

$$\hat{\eta}=AP_1\hat{\theta}_1Q_2'B=AP_1\hat{\theta}_2\tilde{Q}_2'B,$$

por lo que la estimación máximo verosímil de η no depende de la particular reparametrización del modelo seleccionada.

De esta manera, la única fuente restante posible de no unicidad de la matriz $\hat{\eta}$ es la particular elección de las matrices A y B. Para demostrar que el estimador de η no depende de la elección de A y B bajo la condición η=AE(X)B considérense dos representaciones diferentes de la matriz η dadas por

$$\eta = AE(X)B = \tilde{A}E(X)\tilde{B},$$

de donde se sigue que

$$\eta = AD\mu EB = \tilde{A}D\mu E\tilde{B} \qquad \forall\mu$$

$$\Leftrightarrow \quad \eta = AP_1U\mu VQ_2'B=\tilde{A}P_1U\mu VQ_2'\tilde{B} \qquad \forall\mu$$

$$\Leftrightarrow \quad \eta = AP_1\theta Q_2'B=\tilde{A}P_1\theta Q_2'\tilde{B} \qquad \forall\theta \tag{4.20}$$

$$\Rightarrow \quad \hat{\eta} = AP_1\hat{\theta} Q_2'B=\tilde{A}P_1\hat{\theta} Q_2'\tilde{B},$$

lo cual demuestra la invarianza de la estimación de η respecto a la elección de las matrices A y B y concluye la demostración del teorema.

De hecho, por una aplicación del teorema A.31 se deduce de (4.20) que

$$AP_1 = \tilde{A}P_1$$

y

$$Q_2'B = Q_2'\tilde{B};$$

es decir, los productos AP_1 y $Q_2'B$ son invariantes a la particular elección de las matrices A y B bajo la condición η=AE(X)B.

El material anterior puede ilustrarse mediante un ejemplo. Considérese el modelo de diseño experimental completamente al azar con un criterio de clasificación con dos niveles definido por

$$Y_{ij}= \delta+\alpha_i+\varepsilon_{ij} \qquad i=1,2 \quad j=1,\ldots,n_i, \tag{4.21}$$

donde ε_{ij} $i=1,2$, $j=1,\ldots,n_i$ son variables aleatorias independientes con distribución $N(0,\sigma^2)$. En términos del modelo generalizado de Análisis de varianza (4.1), el modelo (4.21) puede escribirse como

$$Y_{n\times 1} = D_{n\times 3}\mu_{3\times 1},$$

donde $n=n_1+n_2$ y las matrices Y, D y μ se definen como

$$Y = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{bmatrix}, \qquad D = \begin{bmatrix} 1 & 1 & 0 \\ & \vdots & \\ 1_{n_1} & 1_{n_1} & 0_{n_1} \\ 1_{n_1+1} & 0_{n_1+1} & 1_{n_1+1} \\ & \vdots & \\ 1 & 0 & 1 \end{bmatrix} \quad \text{y} \quad \mu = \begin{bmatrix} \delta \\ \alpha_1 \\ \alpha_2 \end{bmatrix},$$

donde 0_i y 1_i denotan el cero y la unidad en el renglón i-ésimo de la matriz D. Dado que el rango de la matriz D es dos, se deduce que a lo más pueden estimarse dos parámatros independientes en forma única. Supongase que se está interesado en determinar estimaciones de las medias de cada tratamiento y la diferencia de medias de los tratamientos; esto es, interesa estimar el vector η definido por

$$\eta = \begin{pmatrix} \mu+\alpha_1 \\ \mu+\alpha_2 \\ \alpha_1-\alpha_2 \end{pmatrix}.$$

Para demostrar que η es estimable obsérvese que puede representarse como

$$\eta = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \delta \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1_{n_1+1} & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 & -1_{n_1+1} & 0 & \ldots & 0 \end{pmatrix} \begin{bmatrix} \mu+\alpha_1 \\ \vdots \\ (\mu+\alpha_1)_{n_1} \\ (\mu+\alpha_2)_{n_1+1} \\ \vdots \\ \mu+\alpha_2 \end{bmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 1_{n_1+1} & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 & -1_{n_1+1} & 0 & \dots & 0 \end{pmatrix} E(Y),$$

donde 1_i denota la unidad en la columna i-ésima y $(\mu+\alpha_j)_i$ denota el valor $\mu+\alpha_j$ en el i-ésimo renglón. Como se mencionó con anterioridad, la reparametrización de la matriz μ en la matriz θ puede elegirse tratando de formar parámetros de interés; para lo cual a manera de ejemplo se especifican dos diferentes reparametrizaciones definidas por

$$\theta_1 = \begin{bmatrix} \sqrt{n_1}(\mu+\alpha_1) \\ \sqrt{n_2}(\mu+\alpha_2) \end{bmatrix}$$

y

$$\theta_2 = \begin{bmatrix} \sqrt{n_1}(\mu+\alpha_1) \\ \sqrt{n_2}(\alpha_1-\alpha_2) \end{bmatrix}.$$

Asociadas a estas reparametrizaciones se toman dos descomposiciones de la matriz D defindas por $D=P_1U$ y $D=\tilde{P}_1\tilde{U}$ donde

$$P_1 = \begin{bmatrix} n_1^{-\frac{1}{2}} & 0 \\ \vdots & \vdots \\ n_1^{-\frac{1}{2}} & 0 \\ 0 & n_2^{-\frac{1}{2}} \\ \vdots & \vdots \\ 0 & n_2^{-\frac{1}{2}} \end{bmatrix}, \qquad U = \begin{pmatrix} \sqrt{n_1} & \sqrt{n_1} & 0 \\ \sqrt{n_2} & 0 & \sqrt{n_2} \end{pmatrix},$$

$$\tilde{P}_1 = \begin{bmatrix} n_1^{-\frac{1}{2}} & 0 \\ \vdots & \vdots \\ n_1^{-\frac{1}{2}} & 0 \\ n_1^{-\frac{1}{2}} & -n_2^{-\frac{1}{2}} \\ \vdots & \vdots \\ n_1^{-\frac{1}{2}} & -n_2^{-\frac{1}{2}} \end{bmatrix} \quad y \quad \tilde{U} = \begin{pmatrix} \sqrt{n_1} & \sqrt{n_1} & 0 \\ 0 & \sqrt{n_2} & -\sqrt{n_2} \end{pmatrix},$$

con las condiciones $P_1P_1'=\tilde{P}_1\tilde{P}_1'=I_2$ y las matrices U y $\tilde{U}$ de rango completo. Puede verificarse que con las elecciones de las matrices U y $\tilde{U}$ se cumple que $\theta_1=U\mu$ y $\theta_2=\tilde{U}\mu$. Es interestante observar que el estimador de η puede generarse de manera equivalente en función de $\hat{\theta}_1$ y $\hat{\theta}_2$ ya que

$$\theta_1=\begin{pmatrix} 1 & 0 \\ \left(\frac{n_2}{n_1}\right)^{\frac{1}{2}} & -1 \end{pmatrix}\theta_2;$$

es decir, θ_1 y θ_2 son función uno a uno. Si se elige la reparametrización θ_1, la matriz η puede representarse en términos del valor esperado de Y de diferentes maneras, entre las cuales dos posibles son

$$\eta=AE(Y)=\tilde{A}E(Y),$$

donde

$$A=\begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1_{n_1+1} & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 & -1_{n_1+1} & 0 & \ldots & 0 \end{pmatrix},$$

$$\tilde{A}=\begin{pmatrix} 0 & 0 & 0 & \ldots & 1_{n_1} & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & \ldots & 1_{n_1} & 0 & 0 & \ldots & -1 \end{pmatrix}$$

y 1_i representa la unidad en la i-ésima columna de las matrices A y $\tilde{A}$. Puede verificarse fácilmente que

$$AP_1=\tilde{A}P_2=\begin{pmatrix} n_1^{-\frac{1}{2}} & 0 \\ 0 & n_2^{-\frac{1}{2}} \\ n_1^{-\frac{1}{2}} & -n_2^{-\frac{1}{2}} \end{pmatrix},$$

de donde se tiene la propiedad de invarianza de AP_1 ante la elección de la matriz A, bajo la condición $\eta=AE(Y)$. De esta manera, se deduce que el estimador máximo verosímil de la matriz η, definiendo $\hat{\theta}_1'=(\hat{\theta}_{11},\hat{\theta}_{12})$, está dado por

$$\eta=AP_1\hat{\theta}_1$$

$$=\begin{pmatrix} n_1^{-\frac{1}{2}} & 0 \\ 0 & n_2^{-\frac{1}{2}} \\ n_1^{-\frac{1}{2}} & -n_2^{-\frac{1}{2}} \end{pmatrix}\begin{pmatrix} \hat{\theta}_{11} \\ \hat{\theta}_{12} \end{pmatrix}$$

$$= \begin{bmatrix} n_1^{-\frac{1}{2}}\,\hat{\theta}_{11} \\ n_2^{-\frac{1}{2}}\,\hat{\theta}_{12} \\ n_1^{-\frac{1}{2}}\,\hat{\theta}_{11} - n_2^{-\frac{1}{2}}\,\hat{\theta}_{12} \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\eta}_1 \\ \hat{\eta}_2 \\ \hat{\eta}_3 \end{bmatrix},$$

donde

$$\eta_1=\delta+\alpha_1,$$

$$\eta_2=\delta+\alpha_2$$

y

$$\eta_3=\alpha_1-\alpha_2.$$

Cabe mencionar que no se ha escrito, por ejemplo, $\eta_1=\hat{\mu}+\hat{\alpha}_1$ ya que, como puede verificarse, μ y α_1 no son estimables.

Como último punto de esta sección se demuestra que la condición indicada en (2.7), impuesta por Potthoff & Roy para garantizar que la hipótesis H_0:$A_{txg}\mu_{gxp}B_{pxu}=0$ pueda ser contrastada, no es más que una condición de estimabilidad de la matriz de parámetros $\eta_{txu}=A_{txg}\mu_{gxp}B_{pxu}$. Para ésto, considérese el modelo correspondiente definido por

$$E(X_{nxp})=D_{nxg}\mu_{gxp}.$$

De las ecuaciones de estimabilidad (4.3) y (4.4) se deduce que para la matriz η sea estimable deben existir matrices F y G tales que

$$A=FD$$

y

$$B=IG.$$

De las ecuaciones anteriores se tiene que basta tomar G=B para que la segunda igualdad se cumpla, por lo que se deduce que la matriz η es estimable para toda matriz B. De la ecuación (2.7) se tiene

$$A_{txg}=\left(A_{1txr},A_{2tx(g-r)}\right)$$

$$=\left[A_1, A_1(D_1'D_1)^{-1}D_1'D_2\right]$$

$$= A_1(D_1'D_1)^{-1}D_1'[D_1, D_2]$$

$$= A_1(D_1'D_1)^{-1}D_1' D,$$

por lo que tomando $F=A_1 (D_1'D_1)^{-1}D_1'$ se tiene $A=FD$, de donde se sigue η es estimable y por lo tanto la ecuación (2.7) es una condición de estimabilidad. Sin embargo, cabe mencionar que la matriz F que se deriva de la proposición de Potthoff & Roy está determinada, por lo que no resulta clara la posibilidad de que la condición (2.7) pueda no cumplirse, pero sí exista una matriz F, distinta a la propuesta por estos autores, que satisfaga la condición de estimabilidad.

4.2. Formulación de la Hipótesis Lineal General Sobre Matrices Estimables

Dada la generalidad del modelo (4.1), no es posible contrastar cualquier hipótesis lineal respecto a la matriz μ de parámetros desconocidos. El hecho de poder realizar un contraste particular sobre la matriz μ depende, como se discutió en el proceso de estimación, de la relación que guardan las matrices de la hipótesis con las matrices D y E.

Con objeto de plantear la hipótesis lineal general sobre matrices estimables, primero se define una transformación lineal de la matriz μ que involucra a las matrices D y E, sobre la cual es posible realizar cualquier contraste lineal en los parámetros. Con esta reparametrización del modelo es posible desarrollar como caso particular el contraste asociado a una hipótesis lineal general sobre μ cuando D y E sean matrices de rango completo y como se discutirá en su oportunidad, puede contrastarse cualquier hipótesis que involucre matrices estimables. La transformación lineal de la matriz μ está dada por θ_{rxs} definida en (4.11). Cabe señalar, como se discutió en la sección 4.1.2, que la elección de las matrices U y V no es única, por lo que pueden elegirse tratando de formar en la matriz θ parámetros de interés.

En términos de la matriz θ, es posible realizar cualquier contraste de hipótesis sobre transformaciones lineales de la matriz θ. Como se discutió en la sección 4.1.2, toda matriz estimable es una transformación lineal de θ, asi que establecer una forma de contrastar una hipótesis lineal general sobre la matriz θ proporciona la solución al contraste de la hipótesis lineal general sobre cualquier matriz estimable. El contraste lineal general sobre la matriz θ que se analiza en este trabajo está definido por

$$H_0:\ A_{txr}\ \theta_{rxs}\ B_{sxu} = C_{txu} \qquad (4.22)$$

vs

$$H_1:\ A_{txr}\ \theta_{rxs}\ B_{sxu} \neq C_{txu}$$

donde A, B y C son matrices de constantes tales que

$$r(A)=t \quad t\leq r$$

y

$$r(B)=u \quad u\leq s.$$

Resulta conveniente en este punto hacer algunos comentarios respecto a las restricciones $t\leq r$ y $u\leq s$. Si $t>r$ y/o $u\geq s$ la hipótesis nula establecería un mayor número de restricciones lineales que el número de renglones y/o el número de columnas de la matriz θ que el posible, sin agregar restricciones redundantes y/o inconsistentes. Esta situación puede ilustrarse a manera de ejemplo suponiendo que A y B son matrices de rango completo tales que $t>r$ y $u>s$ y obeniendo de los teoremas A.19 y A.20 que las matrices A y B pueden escribirse como

$$A_{t\times r}= R_{t\times t}\begin{pmatrix} I_r \\ O_{(t-r)\times r}\end{pmatrix} S_{r\times r}$$

y

$$B_{s\times u}= \tilde{S}_{s\times s}\,(I_s,\ O_{s\times(u-s)})\,\tilde{R}_{u\times u},$$

donde $R'R=I$, $\tilde{R}'\tilde{R}=I$. $r(S)=r$ y $r(\tilde{S})=s$. La hipótesis H_0 puede escribirse equivalentemente como

$$H_0:\begin{pmatrix} S\theta\tilde{S}_{r\times s} & 0 \\ 0 & 0\end{pmatrix}=\tilde{C},$$

donde $C=R'C\tilde{R}'$. De esta manera se obtiene que algunas submatrices de $\tilde{C}$ son cero en cuyo caso existen restricciones redundantes o existe una inconsistencia en las ecuaciones especificadas en la hipótesis nula.

El hecho de poder realizar el contraste (4.22) para cualesquier matrices A y B está garantizado, dado que θ es una matriz de parámetros que pueden ser estimados de manera única en base a la información proporcionada por la matriz X. En términos de la matriz de parámetros originales μ, el juego de hipótesis (4.22) está equivalentemente dado por

$$H_0:\ A^*_{t\times g}\,\mu_{g\times g}\,B^*_{q\times u}= C_{t\times u}$$

vs.

$$H_1:\ A^*_{t\times g}\,\mu_{g\times g}\,B^*_{q\times u}\neq C_{t\times u},$$

donde

$$A^*= AU,\qquad B^*= VB \tag{4.23}$$

y las matrices U y V se definen en (4.9) y (4.10). La hipótesis anterior no puede ser contrastada para cualesquier par de matrices A^* y B^*, pero sí para cualesquier matrices A^* y B^* de la forma (4.23) donde U y V son cualesquier par de matrices que satisfacen respectivamente (4.9) y (4.10). Este hecho garantiza que los renglones de A^* y las columnas de B^* estén respectivamente, en los espacios generados por los renglones de D y columnas de E; es decir,

$$A^* = FD$$
y
$$B^* = EG,$$
donde

$$F = A(I_r, O_{rx(n-r)})P'$$
y
$$G = Q\begin{pmatrix} Q_{(p-s)xs} \\ I_s \end{pmatrix} B.$$

En otras palabras, la construcción de las matrices A^* y B^* garantizan la estimabilidad de la matriz $\eta = A^*\mu B^*$. De esta manera y como se discutirá en la siguiente sección, es posible entonces, realizar contraste sobre μ en términos de las matrices A y B. Así, dada una reparametrización θ de la matriz μ, deben determinarse matrices A y B que satisfagan la ecuación (4.23) para escribir la hipótesis en términos de la matriz θ y utilizar los resultados presentados en la sección 4.3.

Generalmente cuando se plantea una hipótesis se escribe primero en términos de los parámetros originales del modelo; es decir, el contraste se plantea como

$$H_0: \tilde{A}\mu\tilde{B} = C$$
vs
$$H_1: \tilde{A}\mu\tilde{B} \neq C. \qquad (4.24)$$

Para efectos de realizar la prueba, garantizando que la estimación de $\eta = \tilde{A}\mu\tilde{B}$ sea única, se reqiere que η sea estimable con objeto de producir un único estimador máximo verosímil $\hat{\eta}$ de η y compararlo con la matriz $\tilde{C}$ a través de la estadística de cociente de verosimilitudes generalizado.

Resulta conveniente para realizar la prueba escribir el juego de hipótesis (4.24) es la forma (4.22) con objeto de utilizar los desarrollos presentados en la sección 4.3. En algunas ocasiones el modelo resulta suficientemente simple para determinar matrices A^* y B^* tales que el juego de hipótesis pueda plantearse directamente como

$$H_0: A^*E(X)B^* = C$$
vs
$$H_1: A^*E(X)B^* \neq C,$$

pudiendo escribirse en la forma (4.22) tomando $A=A^{*}P_1$ y $B=Q_2'B^{*}$. En otras ocasiones, puede recurrirse a las ecuaciones de estimabilidad (4.3) y (4.4) y determinar del juego de hipótesis (4.24) matrices F y G tales que

$$\tilde{A}=FD$$

y

$$\tilde{B}=EG,$$

de donde la hipótesis puede expresarse en la forma (4.22) tomando $A=FP_1$ y $B=Q_2'G$. Cuando las dimensiones del modelo son lo suficientemente grandes para determinar con relativa facilidad el contraste deseado en términos de la matriz θ, pueden utilizarse inversas generalizadas. Para ésto, tómese F y G como inversas generalizadas de las matrices DD' y EE' para determinar primero matrices A^{*} y B^{*} que satisfacen las condiciones de estimabilidad; es decir,

$$\tilde{A}=A^{*}D$$

y

$$\tilde{B}=EB^{*}.$$

Si las ecuaciones anteriores se cumplen, entonces

$$\tilde{A}D'=A^{*}DD'$$

y (4.25)

$$E'\tilde{B}=E'EB^{*},$$

de donde una solución está dada por

$$A^{*}=\tilde{A}D'F$$

y (4.26)

$$B^{*}=GE'\tilde{B},$$

ya que substituyendo estas matrices en (4.25) se tiene

$$\tilde{A}D'=\tilde{A}D'FDD'=\tilde{A}D',$$

$$E'\tilde{B}=E'EGE'\tilde{B}=E'\tilde{B}$$

y de esta manera dadas las matrices A^{*} y B^{*} como en (4.26), el contraste (4.24) puede escribirse en la forma (4.22) tomando $A=A^{*}P_1$ y $B=Q_2'B^{*}$.

De cualquier forma, se debe buscar la manera de escribir la hipótesis deseada en términos de alguna reparametrización de la matriz de parámetros originales μ usando una matriz $\theta=U\mu V$ donde U y V son matrices que satisfacen (4.9) y (4.10) respectivamente.

4.3. El Cociente de Verosimilitudes Asocado a la Hipótesis Lineal General Sobre Matrices Estimables.

Con objeto de obtener el cociente de verosimilitudes generalizado asociado al contraste (4.22), se efectúa primero una transformación lineal de la matriz X que facilita la obtención de los supremos de la función de verosimilitud. La transformación se define como sigue

Sean $\widetilde{A}_{rxr}$, $\widetilde{B}_{sxs}$ matrices definidas por

$$\widetilde{A}_{rxr}=\begin{bmatrix} A_{1(r-t)xr} \\ A_{2_{txr}} \end{bmatrix} \quad y \qquad \widetilde{B}_{sxs}=\begin{bmatrix} B_{sxu} & B^{*}_{sx(s-u)} \end{bmatrix}$$

donde

$$A_2=(AA')^{-\frac{1}{2}}A,$$

$$A_1A_2'=0,$$

$$A_1A_1'=I_{r-t},$$

$$B^{*'}B=0$$

y

$$B^{*'}B^{*}=I_{s-u}.$$

Sean las matrices L y N definidas por

$$L_{nxn}=\begin{pmatrix} \widetilde{A}_{rxr} & 0_{rx(n-r)} \\ 0_{(n-r)xr} & I_{n-r} \end{pmatrix} P'_{nxn}$$

y

$$N_{pxp}=Q_{pxp}\begin{pmatrix} I_{p-s} & 0_{(p-s)xs} \\ 0_{sx(p-s)} & \widetilde{B}_{sxs} \end{pmatrix}.$$

De la definición de L y N se observa que son matrices de rango completo; es decir, r(L)=n y r(N)=p. Sea Y una matriz aleatoria definida por

$$Y_{nxp}=LXN.$$

El valor esperado de Y está dado por

$$E(Y)=LE(X)N$$

$$= LD\mu EN$$

$$= \begin{pmatrix} \tilde{A}_{rxr} & 0_{rx(n-r)} \\ 0_{(n-r)xr} & I_{n-r} \end{pmatrix} \begin{pmatrix} U_{rxg} \\ 0_{(n-r)xg} \end{pmatrix} \mu_{gxq} \begin{pmatrix} 0_{qx(p-s)} & V_{qxs} \end{pmatrix} \begin{pmatrix} I_{p-s} & 0_{(p-s)xs} \\ 0_{sx(p-s)} & \tilde{B}_{sxs} \end{pmatrix}$$

$$= \begin{pmatrix} \tilde{A}_{rxr} & 0_{rx(n-r)} \\ 0_{(n-r)xr} & I_{n-r} \end{pmatrix} \begin{pmatrix} 0_{rx(p-s)} & \theta_{rxs} \\ 0_{(n-r)x(p-s)} & 0_{(n-r)xs} \end{pmatrix} \begin{pmatrix} I_{p-s} & 0_{(p-s)xs} \\ 0_{sx(p-s)} & \tilde{B}_{sxs} \end{pmatrix}$$

$$= \begin{pmatrix} 0_{rx(p-s)} & \tilde{A}\theta_{rxs} \\ 0_{(n-r)x(p-s)} & 0_{(n-r)xs} \end{pmatrix} \begin{pmatrix} I_{p-s} & 0_{(p-s)xs} \\ 0_{sx(p-s)} & \tilde{B}_{sxs} \end{pmatrix}$$

$$= \begin{pmatrix} 0_{rx(p-s)} & \tilde{A}\theta\tilde{B}_{rxs} \\ 0_{(n-r)x(p-s)} & 0_{(n-r)xs} \end{pmatrix}.$$

Por su parte, la matriz $\tilde{A}\theta\tilde{B}$ puede escribirse como

$$\tilde{A}\theta\tilde{B}_{rxs} = \begin{pmatrix} A_{1(r-t)xr} \\ A_{2txr} \end{pmatrix} \theta_{rxs} \begin{pmatrix} B_{sx\mu} & B^{*}_{sx(s-u)} \end{pmatrix}$$

$$= \begin{pmatrix} A_1\theta B_{(r-t)xu} & A_1\theta B^{*}_{(r-t)x(s-u)} \\ A_2\theta B_{txu} & A_2\theta B^{*}_{tx(s-u)} \end{pmatrix},$$

donde, por ser $\tilde{A}$ y $\tilde{B}$ matrices cuadradas de rango completo, se deduce que $\tilde{A}\theta\tilde{B}$ es una matriz de parámetros independientes dado que θ es una matriz de parámetros independientes. De esta manera, el valor esperado de la matriz Y puede expresarse como

$$E(Y) = \begin{pmatrix} 0 & \eta_{12} & \eta_{13} \\ 0 & \eta_{22} & \eta_{23} \\ 0 & 0 & 0 \end{pmatrix},$$

donde

$$\eta_{12} = A_1\theta B,$$

$$\eta_{22}= A_2\theta B,$$

$$\eta_{13}= A_1\theta B^*$$

y

$$\eta_{23}= A_2\theta B^*.$$

Definiendo $\tilde{C}=(AA')^{-\frac{1}{2}}C$, el juego de hipótesis (4.22) es equivalente a

$$H_0: \quad \eta_{22}= \tilde{C}$$

vs (4.27)

$$H_1: \quad \eta_{22}\neq \tilde{C}.$$

Definiendo $M=D\mu E$ y particionando la matriz M como

$$M_{nxp}=\begin{bmatrix} m'_1 \\ \vdots \\ m'_n \end{bmatrix},$$

la función de densidad de la matriz X puede escribirse como

$$\begin{aligned} f_x(X;M,\Sigma) &= \prod_{i=1}^{n} \left[|2\pi\Sigma|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}(X_i-m_i)'\Sigma^{-1}(X_i-m_i)\right\}\right] \\ &= |2\pi\Sigma|^{-n/2} exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}(X_i-m_i)'\Sigma^{-1}(X_i-m_i)\right\} \\ &= |2\pi\Sigma|^{-n/2} exp\left\{-\tfrac{1}{2}\, tr\Sigma^{-1}\sum_{i=1}^{n}(X_i-m_i)'(X_i-m_i)\right\} \\ &= |2\pi\Sigma|^{-n/2} exp\left\{-\tfrac{1}{2}\, tr\Sigma^{-1}(X-M)'(X-M)\right\}. \end{aligned}$$

Del teorema A.28, el jacobiano de la transformación inversa de $Y=LXN$ está dado por

$$\left|\frac{\partial X}{\partial Y}\right|= |L|^{-p}\,|N|^{-n},$$

de manera que la función de densidad de la matriz Y puede escribirse como

$$\begin{aligned} f_y(Y) &= f_x(L^{-1}YN^{-1})\,|L|^{-p}\,|N|^{-n} \\ &= |2\pi\Sigma|^{-n/2} exp\left\{-\tfrac{1}{2}\, tr\,\Sigma^{-1}(L^{-1}YN^{-1}-M)'(L^{-1}YN^{-1}-M)\right\}|L|^{-p}\,|N|^{-n} \\ &= |2\pi N'\Sigma N|^{-n/2}\,|L|^{-p} exp\left\{-\tfrac{1}{2}\, tr\,\Sigma^{-1}\left[L^{-1}(Y-LMN)N^{-1}\right]'\left[L^{-1}(Y-LMN)N^{-1}\right]\right\} \\ &= |2\pi N'\Sigma N|^{-n/2}\,|L|^{-p} exp\left\{-\tfrac{1}{2}\, tr(N'\Sigma N)^{-1}(Y-LMN)'(LL')^{-1}(Y-LMN)\right\} \end{aligned}$$

$$= |2\pi\tilde{\Sigma}|^{-n/2}|L|^{-p} exp\{-\tfrac{1}{2} tr\, \tilde{\Sigma}^{-1}(Y-LMN)'(LL')^{-1}(Y-LMN)\},$$

donde $\tilde{\Sigma}$=N'ΣN. Dado que $\tilde{\Sigma}$ es una transformación uno a uno de Σ, se sigue que $\tilde{\Sigma}$ es una matriz de parámetros independientes.

Por su parte, la matriz LL' puede expresarse como

$$LL' = \begin{pmatrix} \tilde{A}_{rxr} & 0_{rx(n-r)} \\ 0_{(n-r)xr} & I_{n-r} \end{pmatrix} P'P \begin{pmatrix} \tilde{A}'_{rxr} & 0_{rx(n-r)} \\ 0_{(n-r)xr} & I_{n-r} \end{pmatrix}$$

$$= \begin{pmatrix} \tilde{A}\tilde{A}'_{rxr} & 0_{rx(n-r)} \\ 0_{(n-r)xr} & I_{n-r} \end{pmatrix}$$

$$= \begin{pmatrix} A_1A'_1 & A_1A'_2 & 0 \\ A_2A'_1 & A_2A'_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} I_{r-t} & 0 & 0 \\ 0 & I_t & 0 \\ 0 & 0 & I_{n-r} \end{pmatrix}$$

$$= I_n,$$

de manera que la función de densidad de Y puede escribirse como

$$f_y(Y)= |2\pi\tilde{\Sigma}|^{-n/2} exp\{-\tfrac{1}{2} tr\, \tilde{\Sigma}^{-1}(Y-LMN)'(Y-LMN)\},$$

de donde se sigue que los renglones de Y tienen distribución Normal independiente con matriz de covarianzas $\tilde{\Sigma}$ y vectores de medias definidos por la relación E(Y)=LMN.

Definiendo particiones de las matrices Y y E(Y) como

$$Y_{nxp}= (Y_{1nx(p-s+u)}\ Y_{2nx(s-u)})$$

y

$$E(Y_{nxp})= (\eta_{1nx(p-s+u)}\ \eta_{2nx(s-u)})$$

se obtiene por el teorema 3.19 que

$$\Lambda = \frac{\sup_{H_0} f_y(Y)}{\sup_{H_0 \cup H_1} f_y(Y)} = \frac{\sup_{H_0} f_{y_1}(Y_1)}{\sup_{H_0 \cup H_1} f_{y_1}(Y_1)}.$$

Dado que los renglones de Y_1 tienen distribución Normal independiente con matriz de covarianzas $\tilde{\Sigma}_{11}$ y vectores de medias dados por la relación $E(Y_1)=\eta_1$, se sigue que la densidad de Y_1 está dada por

$$f_{y_1}(Y_1) = |2\pi\tilde{\Sigma}_{11}|^{-n/2} exp\left\{-\tfrac{1}{2} tr\,\tilde{\Sigma}_{11}^{-1}(Y_1-\eta_1)'(Y_1-\eta_1)\right\},$$

donde

$$\eta_1 = \begin{pmatrix} 0_{(r-t)\times(p-s)} & \eta_{12(r-t)\times u} \\ 0_{t\times(p-s)} & \eta_{22t\times u} \\ 0_{(n-r)\times(p-s)} & 0_{(n-r)\times u} \end{pmatrix}.$$

Definiendo las matrices $\tilde{Y}_1$, $\tilde{Y}_2$, $\tilde{\eta}_1$, $\tilde{\eta}_2$ por

$$\tilde{Y}_{1\,n\times(p-s)} = \begin{pmatrix} Y_{11(r-t)\times(p-s)} \\ Y_{21\,t\times(p-s)} \\ Y_{31(n-r)\times(p-s)} \end{pmatrix}, \qquad \tilde{Y}_2 = \begin{pmatrix} Y_{12(r-t)\times u} \\ Y_{22t\times u} \\ Y_{32(n-r)\times u} \end{pmatrix},$$

$$\tilde{\eta}_1 = \begin{pmatrix} 0_{(r-t)\times(p-s)} \\ 0_{t\times(p-s)} \\ 0_{(n-r)\times(p-s)} \end{pmatrix} \quad y \quad \tilde{\eta}_2 = \begin{pmatrix} \eta_{12(r-t)\times u} \\ \eta_{22t\times u} \\ 0_{(n-r)\times u} \end{pmatrix},$$

se sigue que $Y_1 = (\tilde{Y}_1, \tilde{Y}_2)$ y $\eta_1 = (\tilde{\eta}_1, \tilde{\eta}_2)$ y el cociente de verosimilitudes generalizado puede obtenerse mediante dos aplicaciones consecutivas del teorema 3.18. Bajo la hipótesis nula, $\eta = \tilde{C}$ y el supremo de la verosimilitud bajo H_0 se obtiene maximizándola sobre η_{12} y $\tilde{\Sigma}_{11}$, el cual está dado por

$$\sup_{H_0} f_{y_1}(Y_1) = (2\pi)^{-nd/2}\, n^{nd/2}\, exp\left\{-\tfrac{nd}{2}\right\} |\tilde{Y}_1'\tilde{Y}_1|^{-n/2}$$

$$\cdot |Y_{32}'Y_{32} + (Y_{22}-\tilde{C})'(Y_{22}-\tilde{C}) - [Y_{32}'Y_{31} + (Y_{22}-\tilde{C})'Y_{21}][Y_{21}'Y_{21} + Y_{31}'Y_{31}]^{-1}[Y_{31}'Y_{32} - Y_{21}'(Y_{22}-\tilde{C})]|^{-n/2},$$

donde $d=p-s+u$. Bajo $H_0 \cup H_1$ las matrices η_{12}, η_{22} y Σ_{11} no poseen restricciones y nuevamente por el teorema 3.18 se tiene que

$$\sup_{H_0 \cup H_1} f_{y_1}(Y_1)=(2\pi)^{-nd/2}\, n^{nd/2}\, exp\{-\tfrac{nd}{2}\}\, |\tilde{Y}_1'\tilde{Y}_1|^{-n/2}\, |Y_{32}'Y_{32}-Y_{32}'Y_{31}[Y_{31}'Y_{31}]^{-1}Y_{31}'Y_{32}|^{-n/2},$$

de manera que el cociente de verosimilitudes generalizado asociado al contraste (4.27) está dado por

$$\Lambda=\frac{\sup_{H_0} f_{y_1}(Y_1)}{\sup_{H_0\cup H_1} f_{y_1}(Y_1)}$$

$$=\frac{|Y_{32}'Y_{32}+(Y_{22}-\tilde{C})'(Y_{22}-\tilde{C})-[Y_{32}'Y_{31}+(Y_{22}-\tilde{C})'Y_{21}][Y_{21}'Y_{21}+Y_{31}'Y_{31}]^{-1}[Y_{31}'Y_{32}+Y_{21}'(Y_{22}-\tilde{C})]|^{-n/2}}{|Y_{32}'Y_{32}-Y_{32}'Y_{31}[Y_{31}'Y_{31}]^{-1}Y_{31}'Y_{32}|^{-n/2}}. \tag{4.28}$$

El cociente de verosimilitudes anterior puede escribirse en términos de la información original X, considerando la siguiente partición de la matriz Y

$$Y=\begin{bmatrix} Y_{11} & Y_{12} & Y_{12} \\ Y_{21} & Y_{22} & Y_{23} \\ Y_{31} & Y_{32} & Y_{33} \end{bmatrix},$$

donde Y_{11} es una matriz de dimensión (r-t)x(p-s) y Y_{22} es una matriz de dimensión txu. La matriz L puede escribirse como

$$L=\begin{pmatrix} \tilde{A} & 0 \\ 0 & I \end{pmatrix} P'$$

$$=\begin{pmatrix} A_1 & 0 \\ A_2 & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} P_1' \\ P_2' \end{pmatrix}$$

$$=\begin{pmatrix} A_1P_1' \\ A_2P_1' \\ P_2' \end{pmatrix}.$$

A su vez, la matriz N está dada por

$$N = Q \begin{pmatrix} I & 0 \\ 0 & \tilde{B} \end{pmatrix}$$

$$= (Q_1 \;\; Q_2) \begin{pmatrix} I & 0 & 0 \\ 0 & B & B^* \end{pmatrix}$$

$$= (Q_1 \quad Q_2B \quad Q_2B^*).$$

La matriz Y puede escribirse en términos de X utilizando las particiones anteriores de L y N y está dada por

$$Y = LXN$$

$$= \begin{pmatrix} A_1P_1' \\ A_2P_1' \\ P_2' \end{pmatrix} X\,(Q_1 \quad Q_2B \quad Q_2B^*)$$

$$= \begin{bmatrix} A_1P_1'XQ_1 & A_1P_1'XQ_2B & A_1P_1'XQ_2B^* \\ A_2P_1'XQ_1 & A_2P_1'XQ_2B & A_2P_1'XQ_2B^* \\ P_2'XQ_1 & P_2'XQ_2B & P_2'XQ_2B^* \end{bmatrix},$$

de donde se sigue que

$$Y_{21} = A_2P_1'XQ_1 = (AA')^{-\frac{1}{2}}AP_1'XQ_1, \tag{4.29}$$

$$Y_{22} = A_2P_1'XQ_2B = (AA')^{-\frac{1}{2}}AP_1'XQ_2B, \tag{4.30}$$

$$Y_{31} = P_2'XQ_1 \tag{4.31}$$

y

$$Y_{32} = P_2'XQ_2B. \tag{4.32}$$

Substituyendo estas expresiones en (4.28) se obtiene una manera de cálculo directo del cociente de verosimilitudes generalizado en términos de la información original.

5

LA DISTRIBUCION DE LA ESTADISTICA DE PRUEBA

5 La Distribución de la Estadística de Prueba

Del cociente de verosimilitudes (4.28) se deduce que la variable T definida por

$$T=\frac{|Y'_{32}Y_{32}-Y'_{32}Y_{31}[Y'_{31}Y_{31}]^{-1}Y'_{31}Y_{32}|}{|Y'_{32}Y_{32}+(Y_{22}-\tilde{C})'(Y_{22}-\tilde{C})-[Y'_{32}Y_{31}+(Y_{22}-\tilde{C})'Y_{21}][Y'_{21}Y_{21}+Y'_{31}Y_{31}]^{-1}[Y'_{31}Y_{32}+Y'_{21}(Y_{22}-\tilde{C})]|} \quad (5.1)$$

puede utilizarse para realizar el contraste, siendo la región de rechazo de la hipótesis nula de la forma

$$\mathcal{C}=\{Y_{21},\ Y_{22},\ Y_{31},\ Y_{32}\,/\,T<k\},$$

donde Y_{21}, Y_{22}, Y_{31} y Y_{32} se definen en las ecuaciones (4.29) a (4.32) y k es una constante apropiada. La distribución de la estadística de prueba T, puede determinarse considerando dos submatrices de la matriz Y definidas por

$$Z_{1t\times(p-s+u)}=(Y_{21t\times(p-s)},\ Y_{22t\times u})$$

y

$$Z_{2(n-r)\times(p-s+u)}=(Y_{31(n-r)\times(p-s)},\ Y_{32(n-r)\times u}).$$

Dado que Y es una matriz con renglones independientes de distribución Normal con matriz de covarianzas $\tilde{\Sigma}$ y vectores de medias dados por E(Y)=LMN se deduce que Z_1 y Z_2 son independientes y que los renglones de las matrices Z_1 y Z_2 poseen distribución Normal con matriz de covarianzas $\tilde{\Sigma}_{11}$ y vectores de medias dados por

$$E(Z_1)=(0_{t\times(p-s)}\quad \eta_{22t\times u})$$

y

$$E(Z_2)=(0_{(n-r)\times(p-s)}\quad 0_{(n-r)\times u}).$$

De la definición 3.2 se deduce que

$$U=[Z_1-E(Z_1)]'[Z_1-E(Z_1)]\sim W_{p-s+u}(\tilde{\Sigma}_{11},\ t)$$

y

$$V=Z'_2Z_2\sim W_{p-s+u}(\tilde{\Sigma}_{11},\ n-r),$$

donde las matrices U y V son independientes por ser Z_1 y Z_2 independientes. Por otro lado, las matrices U y V pueden escribirse como

$$U=\begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22}\end{bmatrix}$$

$$=\begin{bmatrix} Y'_{21}Y_{21} & Y'_{21}(Y_{22}-\eta_{22}) \\ (Y_{22}-\eta_{22})'Y_{21} & (Y_{22}-\eta_{22})'(Y_{22}-\eta_{22})\end{bmatrix}$$

y

$$V=\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22}\end{bmatrix}$$

$$=\begin{bmatrix} Y'_{31}Y_{31} & Y'_{31}Y_{32} \\ Y'_{32}Y_{31} & Y'_{32}Y_{32}\end{bmatrix}.$$

Definiendo

$$W=\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22}\end{bmatrix}=\begin{bmatrix} U_{11}+V_{11} & U_{12}+V_{12} \\ U_{21}+V_{21} & U_{22}+V_{22}\end{bmatrix},$$

se obtiene que la estadística de prueba (5.1) puede expresarse como

$$T=\frac{|V_{22.1}|}{|W_{22.1}|}=\frac{|V_{22}-V_{21}\ V_{11}^{-1}\ V_{12}|}{|W_{22}-W_{21}\ W_{11}^{-1}\ W_{12}|}$$

y por una aplicación del teorema 3.17 se deduce finalmente que $T\sim\Lambda(u,\ n-r-p+s,\ t)$.

Los cuantiles de la distribución Λ de Wilks pueden consultarse en Kres (1983). En los casos en que u ó t tomen el valor 1 ó 2, la distribución Λ de Wilks posee algunas relaciones con la distribución F que se resumen a continuación:

$$\frac{1-\Lambda(p,m,1)}{\Lambda(p,m,1)}\sim\frac{p}{m-p+1}F_{p,m-p+1}, \tag{5.2}$$

$$\frac{1-\Lambda(1,m,n)}{\Lambda(1,m,n)}\sim\frac{n}{m}F_{n,m}, \tag{5.3}$$

$$\frac{1-\sqrt{\Lambda(p,m,2)}}{\sqrt{\Lambda(p,m,2)}}\sim\frac{p}{m-p+1}F_{2p,2(m-p+1)} \tag{5.4}$$

y

$$\frac{1-\sqrt{\Lambda(2,m,n)}}{\sqrt{\Lambda(2,m,n)}} \sim \frac{n}{m-1} F_{2n,2(m-1)}. \tag{5.5}$$

La demostración de estas propiedades puede consultarse en Mendoza (1987). Utilizando estas relaciones, las constantes de la región de rechazo pueden determinarse alternativamente, utilizando tablas de la distribución F.

6

LA TABLA DE ANALISIS DE VARIANZA ASOCIADA A LA HIPOTESIS LINEAL GENERAL EN EL MODELO E(X) = D E

6 El Análisis de Varianza Asociado a la Prueba de la Hipótesis Lineal General en el Modelo E(X)=DμE

En este capítulo se expone un resumen sobre la estimación de parámetros y la prueba de una hipótesis lineal general sobre parámetros estimables, bajo el modelo generalizado de análisis de varianza de Potthoff y Roy. Dado que dependiendo del rango de las matrices D y E, el conjunto de parámetros estimables del modelo resulta distinto, se tiene que tanto el proceso de estimación como el de prueba de hipótesis presentan diferencias que resultan conveniente discutir por separado. Cabe señalar que el esquema de análisis presentado en este trabajo, permite recuperar como casos particulares el análisis estadístico corrrespondiente a los modelos en lo que las matrices D o E poseen rango completo. De esta manera, discutiendo por separado estos casos del modelo, se facilita la obtención de los cálculos que resultan necesarios en una aplicación particular. En las siguientes 4 secciones se discute cada una de las situaciones que resultan de considerar la completez del rango de las matrices D y E.

6.1 El Caso Cuando D y E son de Rango Incompleto

Dada que esta situación es la que se discutió en capítulos anteriores, resulta conveniente describir en esta sección el modelo, la hipótesis y los cálculos asociados con la estimación y prueba de hipótesis.

El modelo generalizado de análisis de varianza multivariado de Pottoff y Roy considera una matriz X_{nxp} con renglones independientes de distribución Normal con matriz de covarianzas común Σ y vectores de medias definidos por la relación

$$E(X_{nxp})=D_{nxg}\mu_{gxq}E_{qxp}, \tag{6.1}$$

donde $r(D)=r<g<n$ y $r(E)=s<q<p$. Para plantear la hipótesis lineal general sobre los parámetros estimables de este modelo, se consideran las descomposiciones de las matrices D y E definidas por

$$D_{nxg}=P_{nxn}\begin{pmatrix} I_{rxr} \\ O_{(n-r)xr} \end{pmatrix} U_{rxg}$$

y

$$E_{qxp}=V_{qxs}(O_{sx(p-s)}, I_{sxs})\,Q'_{pxp},$$

donde $P'P=I$, $Q'Q=I$, $r(U)=r$ y $r(V)=s$. Utilizando éstas descomposiciones se define la matriz $\theta_{r\times s}$ por

$$\theta_{r\times s} = U_{r\times g}\, \mu_{g\times q}\, V_{q\times s} \qquad r\leq g,\ s\leq q\,. \tag{6.2}$$

Como se demostró en su oportunidad, cualquier matriz de parámetros estimables se obtiene mediante una transformación lineal de θ. Utilizando la transformación anterior y considerando las particiones $P=(P_1,P_2)$ y $Q=(Q_1,Q_2)$ donde P_1 y Q_1 tienen respectivamente r y p-s columnas, se obtiene que el modelo (6.1) puede escribirse como

$$E(X_{n\times p}) = P_1\theta Q_2'$$

y los estimadores de máxima verosimilitud de las matrices θ y Σ están dados por

$$\hat{\theta}=P_1'XW^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}$$

y

$$\hat{\Sigma}=\tfrac{1}{n}(X-P_1\hat{\theta}Q_2')'(X-P_1\hat{\theta}Q_2'),$$

donde $W=X'(I-P_1P_1')X$. Con estas estimaciones, el estimador de máxima verosimilitud de la matriz estimable $\eta_{a\times b}=F_{a\times n}E(X)G_{p\times b}$ puede calcularse como

$$\hat{\eta}_{a\times b}=FP_1\hat{\theta}Q_2'G\,.$$

El juego de hipótesis lineal general sobre los parámetros estimables del modelo está dado por

$$\begin{aligned} &H_0:\ A_{t\times r}\,\theta_{r\times s}\,B_{s\times u} = C_{t\times u} \\ \text{vs}\quad & \\ &H_1:\ A_{t\times r}\,\theta_{r\times s}\,B_{s\times u} \neq C_{t\times u}, \end{aligned} \tag{6.3}$$

donde A, B y C son matrices de constantes tales que

$$r(A)=t \qquad t\leq r$$

y

$$r(B)=u \qquad u\leq s\,.$$

La estadística de prueba de cociente de verosimilitudes para el juego de hipótesis anterior, puede calcularse definiendo las matrices

$$Y_{21} = (AA')^{-\frac{1}{2}}AP_1'XQ_1,$$

$$Y_{22} = (AA')^{-\frac{1}{2}}AP_1'XQ_2B,$$

$$Y_{31} = P_2' X Q_1 ,$$

$$Y_{32} = P_2' X Q_2 B$$

y está dada por

$$T = \frac{|Y_{32}' Y_{32} - Y_{32}' Y_{31} [Y_{31}' Y_{31}]^{-1} Y_{31}' Y_{32}|}{|Y_{32}' Y_{32} + (Y_{22} - \tilde{C})'(Y_{22} - \tilde{C}) - [Y_{32}' Y_{31} + (Y_{22} - \tilde{C})' Y_{21}][Y_{21}' Y_{21} + Y_{31}' Y_{31}]^{-1}[Y_{31}' Y_{32} + Y_{21}'(Y_{22} - \tilde{C})]|} ,$$

donde $\tilde{C} = (AA')^{-\frac{1}{2}} C$. La región de rechazo de la hipótesis nula es de la forma

$$\mathcal{C} = \{Y_{21}, Y_{22}, Y_{31}, Y_{32} \mid T < k\}$$

y la constante de la región de rechazo puede obtenerse utilizando el hecho de que $T \sim \Lambda(u, n-r-p+s, t)$.

6.2 El Caso Cuando D es de Rango Completo

Cuando la matriz D es de rango completo, no existe necesidad de premultiplicar la matriz μ por la matriz U para obtener parámetros estimables del modelo. Este hecho puede demostrarse considerando las particiones de las matrices D y E definidas por

$$D_{n\times g} = P_{n\times n} \begin{pmatrix} I_{g\times g} \\ O_{(n-g)\times g} \end{pmatrix} U_{g\times g}$$

y

$$E_{q\times p} = V_{q\times s} (O_{s\times(p-s)}, I_{s\times s}) Q'_{p\times p} ,$$

donde $P'P=I$, $Q'Q=I$, $r(U)=g$ y $r(V)=s$. Dado que cualquier transformación lineal de la matriz θ definida en (6.2) es una matriz de parámetros estimables y considerando en este caso que la matriz U^{-1} esta bien definida, en esta sección se considera la matriz estimable $U^{-1}\theta$, que por simplicidad seguirá denotándose por θ, de manera que la matriz de parámetros estimables para esta situación resulta definida por

$$\theta_{g\times s} = \mu_{g\times q} V_{q\times s} \qquad s \leq q . \tag{6.4}$$

De esta manera, el procedimiento de estimación y prueba de hipótesis se obtiene substituyendo la matriz θ definida en la sección 6.1 por $U\theta$. Utilizando este razonamiento y considerando las particiones $P=(P_1, P_2)$ y $Q=(Q_1, Q_2)$ donde P_1 y Q_1 tienen respectivamente r y p-s columnas, se deduce que el estimador de máxima verosimilitud de la matriz θ puede obtenerse como

$$\hat{\theta}=U^{-1}P_1'XW^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}$$

$$=(U'U)^{-1}U'P_1'XW^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}$$

$$=(D'D)^{-1}D'XW^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}.$$

A su vez, la matriz W puede escribirse como

$$W=X'[I-P_1P_1']X$$

$$=X'[I-P_1U(U'U)^{-1}U'P_1]X$$

$$=X'[I-D(D'D)^{-1}D']X\ ,$$

de donde se sigue que los estimadores de θ y Σ están dados por

$$\hat{\theta}=(D'D)^{-1}D'XW^{-1}Q\ [Q'W^{-1}Q\ ]^{-1}$$

y

$$\hat{\Sigma}=\frac{1}{n}(X-D\hat{\theta}Q_2')'(X-D\hat{\theta}Q_2'),$$

con $W=X'[I-D(D'D)^{-1}D']X$. Utilizando estas ecuaciones, el estimador máximo verosímil de una matriz estimable $\eta_{axb}=F_{axn}E(X)G_{pxb}$ se obtiene como

$$\hat{\eta}_{axb}=FD\hat{\theta}Q_2'G.$$

El juego de hipótesis para esta situación corresponde a

$$H_0:\ A_{txg}\,\theta_{gxs}\,B_{sxu}=\ C_{txu}$$

vs $\qquad$ (6.5)

$$H_1:\ A_{txg}\,\theta_{gxs}\,B_{sxu}\neq C_{txu},$$

donde A, B y C son matrices de constantes tales que

$$r(A)=t \qquad t\leq g$$

y

$$r(B)=u \qquad u\leq s.$$

Para el cálculo de la estadística de prueba, se observa que el juego de hipótesis (6.5) se obtiene del (6.3) substituyendo la matriz A por AU^{-1}, por lo que en este caso las matrices involucradas en el cálculo de la estadística de prueba están dadas por

$$Y_{21}=[A(U'U)^{-1}A')]^{-\frac{1}{2}}AU^{-1}P_1'XQ_1=[A(D'D)^{-1}A']^{-\frac{1}{2}}A(D'D)^{-1}D'XQ_1\ ,$$

$$Y_{22}= [A(U'U)^{-1}A')]^{-\frac{1}{2}} A U^{-1}P_1' XQ_2B = [A(D'D)^{-1}A']^{-\frac{1}{2}} A(D'D)^{-1}D' XQ_2B \ ,$$

$$Y_{31}= P_2' XQ_1$$

y

$$Y_{32}= P_2' XQ_2B \ .$$

Utilizando estas matrices y definiendo $\tilde{C}=\{A(D'D)^{-1}A'\}^{-\frac{1}{2}}C$, la estadística de prueba se obtiene como

$$T=\frac{|Y_{32}'Y_{32}-Y_{32}'Y_{31}[Y_{31}'Y_{31}]^{-1}Y_{31}'Y_{32}|}{|Y_{32}'Y_{32}+(Y_{22}-\tilde{C})'(Y_{22}-\tilde{C})-[Y_{32}'Y_{31}+(Y_{22}-\tilde{C})'Y_{21}][Y_{21}'Y_{21}+Y_{31}'Y_{31}]^{-1}[Y_{31}'Y_{32}+Y_{21}'(Y_{22}-\tilde{C})]|} \ ,$$

siendo la región de rechazo de la hipótesis nula de la forma

$$\mathcal{C} = \{Y_{21},\ Y_{22},\ Y_{31},\ Y_{32}\ | T < k\}$$

y la constante de la región de rechazo se obtiene utilizando el hecho de que $T \sim \Lambda(u, n-g-p+s, t)$.

6.3 El Caso Cuando E es de Rango Completo

En esta situación no resulta necesario reparametrizar el modelo postmultiplicando la matriz μ por la matriz V, ya que esta última resulta de rango completo. En este caso, las descomposiciones de las matrices D y E están dadas por

$$D_{n\times g}= P_{n\times n}\begin{pmatrix} I_{r\times r} \\ O_{(n-r)\times r}\end{pmatrix} U_{r\times g}$$

y

$$E_{q\times p}= V_{q\times q}(O_{q\times(p-q)}, I_{q\times q})\, Q'_{p\times p} \ ,$$

donde $P'P=I$, $Q'Q=I$, $r(U)=r$ y $R(V)=q$. Considerando que cualquier transformación lineal de la matriz θ definida en (6.2) resulta estimable y dado que en este caso existe V^{-1}, resulta conveniente considerar la matriz estimable θV^{-1}, que por simplicidad seguirá denotándose de nuevo por θ. De esta manera, la matriz de parámetros estimables para esta situación resulta

$$\theta_{r\times q} = U_{r\times g}\mu_{g\times q} \qquad r \leq g \ .$$

El cálculo de estimadores y el procedimiento de contraste de hipótesis se obtiene substituyendo la matriz θ definida en (6.2) por θV. Considerando las particiones

$P=(P_1,P_2)$ y $Q=(Q_1,Q_2)$ donde P_1 y Q_1 tienen respectivamente r y p-s columnas, se sigue que el estimador máximo verosímil de la matriz θ está dado por

$$\begin{aligned}\hat{\theta} &= P_1'XW^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}V^{-1}\\ &= P_1'XW^{-1}Q_2V'(V')^{-1}[Q_2'W^{-1}Q_2]^{-1}V^{-1}\\ &= P_1'XW^{-1}Q_2V'[VQ_2'W^{-1}Q_2V']^{-1}\\ &= P_1'XW^{-1}E'[EW^{-1}E']^{-1}\ ,\end{aligned}$$

de donde los estimadores de θ y Σ pueden escribirse como

$$\hat{\theta} = P_1'XW^{-1}E'[EW^{-1}E']^{-1}$$

y

$$\hat{\Sigma} = \tfrac{1}{n}(X-P_1\hat{\theta}E)'(X-P_1\hat{\theta}E)\ ,$$

con $W=X'(I-P1P_1')X$. Utilizando estas estimaciones, el estimador máximo verosímil de una matriz estimable $\eta_{axb}=F_{axn}E(X)G_{pxb}$ está dado por

$$\hat{\eta}_{axb}=FP_1\hat{\theta}EG\ .$$

El juego de hipótesis para esta situación resulta definido por

$$\begin{aligned}&H_0:\ A_{txr}\theta_{rxq}B_{qxu}=C_{txu}\\ \text{vs}\quad &\\ &H_1:\ A_{txr}\theta_{rxq}B_{qxu}\neq C_{txu}\ ,\end{aligned} \tag{6.6}$$

donde A, B y C son matrices de constantes tales que

$$\begin{aligned}&r(A)=t \qquad t\leq r\\ \text{y}\quad&\\ &r(B)=u \qquad u\leq q\ .\end{aligned}$$

Dado que el juego de hipótesis (6.6) se obtiene del (6.3) substituyendo la matriz B por $V^{-1}B$, la estadística de prueba asociada al juego de hipótesis (6.6) puede calcularse utilizando las matrices

$$Y_{21}=(AA')^{-\frac{1}{2}}AP_1'XQ_1\ ,$$

$$Y_{22}=(AA')^{-\frac{1}{2}}AP_1'XQ_2V'(V')^{-1}V^{-1}B=(AA')^{-\frac{1}{2}}AP_1'XE'(EE')^{-1}B\ ,$$

$$Y_{31}=P_2'XQ_1\ ,$$

$$Y_{32}=P_2'XQ_2V'(V')^{-1}V^{-1}B=P_2'XE'(EE')^{-1}B$$

y está dada por

$$T=\frac{|Y'_{32}Y_{32}-Y'_{32}Y_{31}[Y'_{31}Y_{31}]^{-1}Y'_{31}Y_{32}|}{|Y'_{32}Y_{32}+(Y_{22}-\tilde{C})'(Y_{22}-\tilde{C})-[Y'_{32}Y_{31}+(Y_{22}-\tilde{C})'Y_{21}][Y'_{21}Y_{21}+Y'_{31}Y_{31}]^{-1}[Y'_{31}Y_{32}+Y'_{21}(Y_{22}-\tilde{C})]|},$$

donde $\tilde{C}=(AA')^{-\frac{1}{2}}C$. La región de rechazo de la hipótesis nula es de la forma

$$\mathcal{C}=\{Y_{21},\ Y_{22},\ Y_{31},\ Y_{32}\,|\,T<k\}$$

y la constante k se obtiene utilizando que $T\sim\Lambda(u,n-r-p+q,t)$.

6.4 El Caso Cuando D y E son de Rango Completo

Este es el caso más simple de analizar, ya que no se requiere transformar la matriz μ dado que ésta es estimable. Para el cálculo de la estadística de prueba se utilizan las descomposiciones siguientes de las matrices D y E

$$D_{n\times g}=P_{n\times n}\begin{pmatrix} I_{g\times g}\\ O_{(n-g)\times g}\end{pmatrix}U_{g\times g}$$

y

$$E_{q\times p}=V_{q\times q}(O_{q\times(p-q)},I_{q\times q})\,Q'_{p\times p}\,,$$

donde $P'P=I$, $Q'Q=I$, $r(U)=g$ y $r(V)=q$. La estimabilidad de la matriz μ puede demostrarse considerando que en este caso las matrices U^{-1} y V^{-1} existen y obteniendo la transformación lineal μ de la matriz θ dada en (6.2) definida por $\mu=U^{-1}\mu V^{-1}$. De esta manera, la matriz de parámetros estimables para esta situación resulta

$$\theta_{g\times q}=\mu_{g\times q}\,.$$

En este caso, la estimación y contraste de hipótesis se obtiene como caso particular de los resultados de la sección 6.2 substituyendo la matriz θ por $U\mu V$. Utilizando de nuevo las particiones de las matrices P y Q definidas por $P=(P_1,P_2)$ y $Q=(Q_1,Q_2)$ donde P_1 y Q_1 tienen respectivamente r y p-s columnas, se tiene que el estimador de máxima verosimilitud de la matriz μ puede obtenerse como

$$\hat{\mu}=U^{-1}P_1'XW^{-1}Q_2[Q_2'W^{-1}Q_2]^{-1}V^{-1}$$

$$=(U'U)^{-1}U'P_1'XW^{-1}Q_2V'(V')^{-1}[Q_2'W^{-1}Q_2]^{-1}V^{-1}$$

$$=(D'D)^{-1}D'XW^{-1}Q_2V'[VQ_2'W^{-1}Q_2V']^{-1}$$

$$=(D'D)^{-1}D'XW^{-1}E'[EW^{-1}E']^{-1},$$

donde $W=X'[I-D(D'D)^{-1}D']X$, por lo que los estimadores máximo verosímiles de las matrices μ y Σ están dados por

$$\hat{\mu}=(D'D)^{-1}D'XW^{-1}E'[EW^{-1}E']^{-1} \tag{6.7}$$

y

$$\hat{\Sigma}=\frac{1}{n}(X-D\hat{\mu}E)'(X-D\hat{\mu}E). \tag{6.8}$$

El juego de hipótesis en este caso puede plantearse directamente en términos de la matriz μ y está dado por

$$H_0:\ A_{t\times g}\,\mu_{g\times q}\,B_{q\times u}=C_{t\times u}$$
vs
$$H_1:\ A_{t\times g}\,\mu_{g\times q}\,B_{q\times u}\neq C_{t\times u}, \tag{6.9}$$

donde A, B y C son matrices de constantes tales que

$$r(A)=t \qquad t\leq g$$
y
$$r(B)=u \qquad u\leq q .$$

Dado que el juego de hipótesis (6.9) se obtiene del (6.3) substituyendo las matrices A y B por AU^{-1} y $V^{-1}B$ respectivamente, las matrices que se utilizan en este caso para el cálculo de la estadística de prueba están dadas por

$$Y_{21}=[A(U'U)^{-1}A')]^{-\frac{1}{2}}AU^{-1}P_1'XQ_1=[A(D'D)^{-1}A']^{-\frac{1}{2}}A(D'D)^{-1}D'XQ_1,$$

$$Y_{22}=[A(U'U)^{-1}A')]^{-\frac{1}{2}}AU^{-1}P_1'XQ_2V'(V')^{-1}V^{-1}B$$

$$=[A(D'D)^{-1}A']^{-\frac{1}{2}}A(D'D)^{-1}D'XE'(EE')^{-1}B,$$

$$Y_{31}=P_2'XQ_1,$$

$$Y_{32}=P_2'XQ_2V'(V')^{-1}V^{-1}B=P_2'XE'(EE')^{-1}B$$

y utilizando las matrices anteriores, la estadística de prueba asociada al contraste (6.9) puede calcularse como

$$T=\frac{|Y_{32}'Y_{32}-Y_{32}'Y_{31}[Y_{31}'Y_{31}]^{-1}Y_{31}'Y_{32}|}{|Y_{32}'Y_{32}+(Y_{22}-\tilde{C})'(Y_{22}-\tilde{C})-[Y_{32}'Y_{31}+(Y_{22}-\tilde{C})'Y_{21}][Y_{21}'Y_{21}+Y_{31}'Y_{31}]^{-1}[Y_{31}'Y_{32}+Y_{21}'(Y_{22}-\tilde{C})]|}.$$

donde $\tilde{C}=\{A(D'D)^{-1}A'\}^{-\frac{1}{2}}C$ y la región de rechazo de la hipótesis nula es de la forma

$$\mathcal{C} = \{Y_{21}, Y_{22}, Y_{31}, Y_{32} \,|\, T < k\}\ .$$

La constante de la región de rechazo se obtiene utilizando que $T \sim \Lambda(u, n-g-p+q\ ,\ t)$.

7

APLICACIONES DE LA HIPOTESIS LINEAL GENERAL

7 Aplicaciones del Modelo Generalizado de Análisis de Varianza Multivariado

Como se indicó en el capítulo 1, existen 3 tipos de modelos lineales que han tenido un amplio desarrollo dada la diversidad de estudios que pueden ser analizados mediante éstos. Estos son los modelos de regresión lineal, modelos de diseño experimental y modelos de crecimiento. Dado que estas 3 clases de modelos resultan ser casos particulares del modelo generalizado de Potthoff y Roy, en este capítulo se exponen los cálculos involucrados en la estimación y pruebas de hipótesis para algunos de estos modelos, utilizando los resultados presentados en el capítulo 6.

7.1 Modelos de Rango Completo

Primero se discute el análisis de algunos modelos en los que las matrices D y E del modelo (6.1) son de rango completo. Entre estos modelos se encuentran los de regresión lineal y los de crecimiento, los cuales se discuten por separado en las siguientes secciones.

7.1.1 El Modelo Univariado de Regresión Lineal

El análisis de este modelo puede derivarse de los resultados presentados en la sección 6.4, substituyendo p=q=1 y las matrices E y B por las matrices identidad de orden 1. El modelo de regresión lineal multiple en el caso univariado considera un vector X_{nx1} cuyas entradas son independientes de distribución normal con varianza común σ^2 y medias definidas por la relación

$$E(X_{nx1})=D_{nxg}\mu_{gx1},$$

donde las g columnas de la matriz D son linealmente independientes y se denominan variables explicativas, las entradas de μ son desconocidas y se denominan coeficientes de regresión. El vector de parámetros estimables para esta situación está dado directamente por μ_{gx1}, cuyo estimador de máxima verosimilitud se obtiene de la ecuación (6.7) y se define por

$$\hat{\mu}=(D'D)^{-1}D'X\ .$$

El estimador de la varianza común σ^2 se obtiene a su vez de la ecuación (6.8) y está dado por

$$\hat{\sigma}^2=\frac{1}{n}(X-D\hat{\mu})'(X-D\hat{\mu})\ .$$

El juego de hipótesis lineal general para esta situación puede definirse como

$$H_0: A_{t\times g}\mu_{g\times 1} = C_{t\times 1}$$

vs

$$H_1: A_{t\times g}\mu_{g\times 1} \neq C_{t\times 1} ,$$

donde A y C son matrices de constantes conocidas tales que $r(A)=t\leq g$. La estadística de prueba para el juego de hipótesis anterior puede obtenerse en base a las matrices

$$Y_{21} = [A(D'D)^{-1}A']^{-\frac{1}{2}} A(D'D)^{-1} D'XQ_1 = [A(D'D)^{-1}A']^{-\frac{1}{2}} A\hat{\mu}Q_1 ,$$

$$Y_{22} = [A(D'D)^{-1}A']^{-\frac{1}{2}} A(D'D)^{-1} D'X = [A(D'D)^{-1}A']^{-\frac{1}{2}} A\hat{\mu} ,$$

$$Y_{31} = P_2'XQ_1$$

y

$$Y_{32} = P_2'X .$$

Definiendo $W=X'[I-D(D'D)^{-1}]X$, el numerador de la estadística de prueba puede calcularse como

$$Y_{32}'Y_{32} - Y_{32}'Y_{31}(Y_{31}'Y_{31})^{-1}Y_{31}'Y_{32}$$

$$= W - WQ_1(Q_1'WQ_1)^{-1}Q_1'W$$

$$= Q_2(Q_2'W^{-1}Q_2)^{-1}Q_2'$$

y dado que en este caso $Q_2 = \pm 1$, se sigue que

$$Y_{32}'Y_{32} - Y_{32}'Y_{31}(Y_{31}'Y_{31})^{-1}Y_{31}'Y_{32} = W .$$

El denominador de la estadística de prueba puede simplificarse observando que

$$(Y_{21}'Y_{21} + Y_{31}'Y_{31})^{-1}$$

$$= \{Q_1'\hat{\mu}'A'[A(D'D)^{-1}A']^{-1}A\hat{\mu}Q_1 + Q_1'WQ_1\}^{-1}$$

$$= \{Q_1'ZQ_1\}^{-1} ,$$

donde $Z=\hat{\mu}'A'[A(D'D)^{-1}A']^{-1}A\hat{\mu}+W$. Utilizando esta expresión se tiene

$$(Y_{21}'Y_{21} + Y_{31}'Y_{31})^{-1} = Q_1'Q_1\{Q_1'ZQ_1\}^{-1}Q_1'Q_1$$

$$=Q_1'\{Z^{-1}-Z^{-1}Q_2(Q_2'Z^{-1}Q_2)^{-1}Q'Z^{-1}\}Q_1$$

$$=Q_1\{0\}Q_1$$

$$=0$$

dado que $Q_2=\pm 1$. De esta manera, se deduce que el denominador de la estadística de prueba puede escribirse como

$$Y_{32}'Y_{32}+(Y_{22}-\tilde{C})'(Y_{22}-\tilde{C})$$

$$=W+(A\hat{\mu}-C)'\{A(D'D)^{-1}A'\}^{-1}(A\hat{\mu}-C)\ ,$$

por lo que la estadística de prueba en este caso está dada por

$$T=\frac{W}{W+(A\hat{\mu}-C)'\{A(D'D)^{-1}A'\}^{-1}(A\hat{\mu}-C)}$$

y tiene distribución $\Lambda(1,n-g,t)$, siendo la región de rechazo de H_0 de la forma

$$C=\{X_{nx1}|T<k\}\ .$$

La constante de la región de rechazo puede determinarse alternativamente utilizando tablas de la distribución F, para lo cual puede utilizarse la ecuación (5.3), obteniéndose que la variable

$$T'=\left(\frac{n-g}{t}\right)\frac{1-T}{T}=\frac{(A\hat{\mu}-C)'\{A(D'D)^{-1}A'\}^{-1}(A\hat{\mu}-C)/t}{W/(n-g)}$$

tiene distribución $F(t,n-g)$ y la región de rechazo de la hipótesis nula es de la forma

$$C=\{X_{nx1}|T'>k\}\ .$$

7.1.2 El Modelo Multivariado de Regresión Lineal

La estimación de parámetros y contraste de hipótesis para este modelo puede obtenerse utilizando los resultados presentados en la sección 6.4, substituyendo q=p y $E=I_p$.

El modelo multivariado de regresión lineal multiple considera una matriz X_{nxp} con renglones independientes de distribución normal con matriz de covarianzas común Σ y vectores de medias definidos por la relación

$$E(X_{nxp})=D_{nxg}\mu_{gxp}\ ,$$

donde $r(D)=g$ y μ_{gxp} es una matriz de coeficientes de regresión desconocidos. La matriz de parámetros estimables en esta situación corresponde a μ_{gxp} y su estimador máximo verosímil puede obtenerse de la ecuación (6.1) y está dado por

$$\hat{\mu}_{gxp}=(D'D)^{-1}D'X$$

y el estimador de máxima verosimilitud de Σ resulta

$$\hat{\Sigma}=\frac{1}{n}(X-D\hat{\mu})'(X-DD\hat{\mu})\ .$$

El contraste lineal general sobre μ puede escribirse como

$$\begin{aligned} &H_0\colon A_{txg}\mu_{gxp}B_{pxu}=C_{txu} \\ &\text{vs} \\ &H_1\colon A_{txg}\mu_{gxp}B_{pxu}\neq C_{txu}\ . \end{aligned} \qquad (7.1)$$

donde A, B y C son matrices de constantes con

$$r(A)=t\leq g$$

y

$$r(B)=u\leq p\ .$$

Las matrices involucradas en el cálculo de la estadística de prueba se definen por

$$Y_{21}=[A(D'D)^{-1}A']^{-\frac{1}{2}}A\hat{\mu}Q_1\ ,$$

$$Y_{22}=[A(D'D)^{-1}A']^{-\frac{1}{2}}A\hat{\mu}B\ ,$$

$$Y_{31}=P_2'XQ_1$$

y

$$Y_{32}=P_2'XB\ .$$

Utilizando estas matrices y definiendo $W=X'[I-D(D'D)^{-1}D']X$, el numerador de la estadística de prueba puede calcularse como

$$|Y_{32}'Y_{32}-Y_{32}'Y_{31}(Y_{31}'Y_{31})^{-1}Y_{31}'Y_{32})|$$

$$=|B'WB-BWQ_1(Q_1'WQ_1)^{-1}Q_1'WB|$$

$$=|B'\{W-WQ_1(Q_1'WQ_1)^{-1}Q_1'W\}B|$$

$$=|B'Q_2(Q_2'W^{-1}Q_2)^{-1}Q_2B|\ .$$

Dado que en este caso puede seleccionarse $Q_2=I_p$, se tiene que

$$|Y_{32}'Y_{32}-Y_{32}'Y_{31}(Y_{31}'Y_{31})^{-1}Y_{31}'Y_{32}|=|B'WB| .$$

El denominador de la estadística de prueba puede simplificarse considerando que

$$[Y_{21}'Y_{21}+Y_{31}'Y_{31}]^{-1}$$

$$=[Q_1'\hat{\mu}'A'\{A(D'D)^{-1}A'\}^{-1}A\hat{\mu}Q_1+Q_1'WQ_1]^{-1}$$

$$=[Q_1'ZQ_1]^{-1} ,$$

con $Z=\hat{\mu}'A'\{A(D'D)^{-1}A'\}^{-1}\hat{A\mu}+W$. De esta manera se deduce que

$$\begin{aligned}[Y_{21}'Y_{21}+Y_{31}'Y_{31}]^{-1} &= Q_1'Q_1[Q_1'ZQ_1]^{-1}Q_1'Q_1\\ &=Q_1'\{Z^{-1}-Z^{-1}Q_2(Q_2'Z^{-1}Q_2)^{-1}Q_2'Z^{-1}\}Q_1\\ &=Q_1'\,0\,Q_1\\ &=0\end{aligned}$$

dado que se seleccionó $Q_2=I_p$. Utilizando este último resultado, se sigue que el denominador de la estadística de prueba resulta

$$Y_{32}'Y_{32}+(Y_{22}-\tilde{C})'(Y_{22}-\tilde{C})=B'WB+(A\hat{\mu}B-C)'[A(D'D)^{-1}A']^{-1}[A\hat{\mu}B-C] ,$$

por lo que la estadística de prueba puede escribirse como

$$T=\frac{|B'WB|}{|B'WB+(A\hat{\mu}B-C)'[A(D'D)^{-1}A']^{-1}(A\hat{\mu}B-C)|}$$

y tiene distribución $\Lambda(u,n-g,t)$, siendo la región de rechazo de la hipótesis nula de la forma

$$\mathcal{C}=\{X_{nxp}|T<k\} .$$

Cabe señalar que en el cálculo de la estadística de prueba, la expresión $(A\hat{\mu}B-C)'[A(D'D)^{-1}A']^{-1}(A\hat{\mu}B-C)$ resulta más sencilla de evaluar que la propuesta en Mardia, Kent & Bibby (1979, pág. 162). El juego de hipótesis (7.1) resulta también ser más general que los discutidos en Seber (1984, cap. 8).

7.1.3 Un Modelo de Crecimiento

Existen diversas situaciones en las que pueden modelarse curvas de crecimiento en base al modelo (6.1), como puede consultarse en Potthoff & Roy (1964). En esta sección, se considera un modelo simple de crecimiento dado que el propósito es sólo ilustar el procedimiento de análisis.

La situación que se ejemplifica en esta sección considera un grupo de n individuos, sujetos a las mismas condiciones, los cuales son observados en p puntos en el tiempo $t_1,\ldots,t_p$. Se supone que las p mediciones realizadas en cada uno de los n individuos siguen una distribución normal con un esquema de asociación modelado por una matriz de covarianza común Σ. Se supone también que la curva de crecimiento asociada a un individuo es un polinomio en el tiempo de grado q-1, de manera que denotado por X_{ij} la j-ésima medición en el i-ésimo individuo se tiene

$$E(X_{ij})=\mu_0+\mu_1 t_j+\mu_2 t_j^2+\cdots\cdots+\mu_{q-1}t_j^{q-1} .$$

El modelo anterior puede escribirse en términos del modelo (6.1), haciendo g=1 y definiendo las matrices D, μ y E por

$$D_{n\times 1}=1_n=\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix},$$

$$\mu_{1\times q}=(\mu_0,\ldots\ldots,\mu_{q-1})$$

y

$$E_{q\times p}=\begin{bmatrix}1 & 1 & \cdots\cdots & 1\\ t_1 & t_2 & \cdots\cdots & t_p\\ t_1^2 & t_2^2 & \cdots\cdots & t_p^2\\ \vdots & \vdots & & \vdots\\ t_1^{q-1} & t_2^{q-1} & \cdots\cdots & t_p^{q-1}\end{bmatrix}$$

El análisis de este modelo puede derivarse utilizando los resultados de la sección 6.4 como sigue:

Sea $F_{n\times n}$ la matriz con todas sus entradas iguales a 1 y sea W=X'[I-F]X. Los estimadores máximo verosimiles μ y Σ están dados por

$$\hat{\mu}=\frac{1}{n}D'XW^{-1}E'[EW^{-1}E']^{-1}=\bar{X}'W^{-1}E'[EW^{-1}E']^{-1}$$

y

$$\hat{\Sigma}=\frac{1}{n}(X-D\hat{\mu}E)'(X-D\hat{\mu}E)\ ,$$

donde $\bar{X}_{p\times 1}$ es el vector de medias de las p mediciones en la matriz X. El juego de hipótesis sobre μ está dado por

$$H_0:\mu_{1\times q}B_{q\times u}=C_{1\times u}$$

vs

$$H_1:\mu_{1\times q}B_{q\times u}\neq C_{1\times u}\ .$$

donde B y C son matrices de constantes tales que $r(B)=u\leq q$. Como caso particular de la hipótesis general anterior, pueden formularse las aseveraciones de que la ordenada al origen del polinomio es igual a un valor determinado o que un polinomio de grado menor que q-1 es suficiente para describir adecuadamente la información. Las matrices asociadas al cálculo de la estadística de prueba para el constraste de hipótesis anterior resultan definidas por

$$Y_{21}=\sqrt{n}\ \bar{X}'Q_1\ ,$$

$$Y_{22}=\sqrt{n}\ \bar{X}'E'(EE')^{-1}B\ ,$$

$$Y_{31}=P_2'XQ_1$$

y

$$Y_{32}=P_2'XE'(EE')^{-1}B$$

Utilizando estas matrices y definiendo $\tilde{C}=\sqrt{n}(AA')^{-\frac{1}{2}}C$, la estadística de prueba puede calcularse como

$$T=\frac{|Y_{32}'Y_{32}-Y_{32}'Y_{31}[Y_{31}'Y_{31}]^{-1}Y_{31}'Y_{32}|}{|Y_{32}'Y_{32}+(Y_{22}-\tilde{C})'(Y_{22}-\tilde{C})-[Y_{32}'Y_{31}+(Y_{22}-\tilde{C})'Y_{21}][Y_{21}'Y_{21}+Y_{31}'Y_{31}]^{-1}[Y_{31}'Y_{32}+Y_{21}'(Y_{22}-\tilde{C})]|}$$

y sigue una distribución $\Lambda(u,n-1-p+q,1)$, siendo la región de rechazo de H_0 de la forma

$$C=\{Y_{21},Y_{22},Y_{31}\text{ y }Y_{32}|T<k\}\ .$$

7.2 Modelos de Rango Incompleto (Diseños Experimentales)

En las siguientes secciones se discuten algunos modelos que resultan de uso común en el análisis de observaciones provenientes de diseños experimentales. Dado que este tipo de modelos es una clase particular del modelo (6.1), en las siguientes secciones se deriva el análisis asociado al proceso de estimación y prueba de hipótesis para algunos modelos de este tipo.

7.2.1 El Modelo Univariado con un Criterio de Clasificación

Para definir este modelo se consideran X_{ij} $i=1,\ldots,a.$ $j=1,\ldots,n_i$ variables aleatorias independientes con distribución normal de medias dadas por la relación

$$E(X_{ij})=\delta+\alpha_i \quad i=1,\ldots,r, \quad j=1,\ldots,n_i \tag{7.2}$$

y varianza común σ^2. El modelo anterior puede ser escrito en forma matricial considerando las matrices

$$X_{n\times 1}=\begin{bmatrix} X_{11} \\ \vdots \\ X_{1n_1} \\ \vdots \\ X_{r1} \\ \vdots \\ X_{rn_r} \end{bmatrix}, \qquad \mu=\begin{bmatrix} \delta \\ \alpha_1 \\ \vdots \\ \alpha_r \end{bmatrix}$$

y

$$D=\begin{array}{c} \left[\begin{matrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \end{matrix}\right] \end{array}
\begin{array}{l} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right] n_1 \text{ renglones} \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right] n_2 \text{ renglones} \\ \vdots \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right] n_r \text{ renglones} \end{array}$$

de manera que el modelo (7.2) puede escribirse alternativamente como

$$E(X_{nx1}) = D_{nx(r+1)}\,\mu_{(r+1)x1}\ .$$

Definiendo p=q=1, g=r+1 y haciendo $E=I_1$, el análisis de este modelo puede derivarse de los resultados presentados en la sección 6.3. Dado que en este caso la matriz μ no es estimable, debe seleccionarse una reparametrización en base a una descomposición de la matriz D de la forma.

$$D_{nx(r+1)} = P_{1nxr}\,U_{rx(r+1)}\ .$$

Para el análisis de este modelo se considera la descomposición de la matriz D definida por las matrices

$$P_1 = \begin{array}{c} \left[\begin{array}{cccccc} n_1^{-1/2} & 0 & 0 & \cdots & 0 & 0 \\ n_1^{-1/2} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & \vdots \\ n_1^{-1/2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & n_2^{-1/2} & 0 & \cdots & 0 & 0 \\ 0 & n_2^{-1/2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & n_2^{-1/2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & n_r^{-1/2} \\ 0 & 0 & 0 & \cdots & 0 & n_r^{-1/2} \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & n_r^{-1/2} \end{array}\right] \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\\1\end{array}}\right\} n_1 \text{ renglones} \\ \left.\vphantom{\begin{array}{c}1\\1\\1\\1\end{array}}\right\} n_2 \text{ renglones} \\ \vdots \\ \left.\vphantom{\begin{array}{c}1\\1\\1\\1\end{array}}\right\} n_r \text{ renglones} \end{array} \quad \text{y} \quad U_{rx(r+1)} \begin{bmatrix} \sqrt{n_1} & \sqrt{n_1} & 0 & 0 \cdots & 0 & 0 \\ \sqrt{n_2} & 0 & \sqrt{n_2} & 0 \cdots & 0 & 0 \\ \vdots & \vdots & & & & \\ \sqrt{n_r} & 0 & \cdots & \cdots & 0 & \sqrt{n_r} \end{bmatrix},$$

de manera que la matriz θ_{rx1} que resulta de la anterior descomposición de D está definida por

$$\theta_{rx1} = U\mu = \begin{bmatrix} \sqrt{n_1}\,(\mu+\alpha_1) \\ \\ \sqrt{n_r}\,(\mu+\alpha_r) \end{bmatrix}\ .$$

El estimador máximo verosímil de θ puede obtenerse como

$$\hat{\theta}=P_1'X=\begin{bmatrix} n_1^{-\frac{1}{2}}\sum_{j=1}^{n_1} & X_{1j} \\ n_2^{-\frac{1}{2}}\sum_{j=1}^{n_2} & X_{2j} \\ \vdots & \\ n_r^{-\frac{1}{2}}\sum_{j=1}^{n_r} & X_{rj} \end{bmatrix} = \begin{bmatrix} n_1^{\frac{1}{2}} & \overline{X}_1 \\ \vdots & \\ \vdots & \\ n_r^{\frac{1}{2}} & \overline{X}_r \end{bmatrix} .$$

A su vez, el estimador de σ^2 puede obtenerse como

$$\hat{\sigma}^2=\frac{1}{n}(X-P_1\hat{\theta})'(X-P_1\hat{\theta})$$

$$=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\frac{1}{n_i}\sum_{k=1}^{n_i}X_{ik})^2$$

$$=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2 .$$

El juego de hipótesis lineal general para esta situación está dado por

$$H_0: A_{t\times r}\,\theta_{r\times 1}=C_{t\times 1}$$

vs

$$H_1: A_{t\times r}\,\theta_{r\times 1}\neq C_{t\times 1} ,$$

donde A y C son matrices de constantes tales que $r(A)=t<r$. Para calcular la estadística de prueba asociada a este constrate se definen las matrices

$$Y_{21}=(AA')^{-\frac{1}{2}}AP_1'XQ_1=(AA')^{-\frac{1}{2}}A\hat{\theta}Q_1 ,$$

$$Y_{22}=(AA')^{-\frac{1}{2}}AP_1'X=(AA')^{-\frac{1}{2}}A\hat{\theta} ,$$

$$Y_{31}=P_2'XQ_1$$

y

$$Y_{32}=P_2'X$$

Definiendo $W=X'(I-P_1P_1')X$, el numerador de la estadística de prueba puede calcularse como

$$Y_{32}'Y_{32}-Y_{32}'Y_{31}(Y_{31}'Y_{31})^{-1}Y_{31}'Y_{32}$$

$$=W-WQ_1(Q_1'WQ_1)^{-1}Q_1'W$$

$$=Q_2(Q_2W^{-1}Q_2)^{-1}Q_2'$$

y dado que en este caso $Q_2=+1$ se tiene que

$$Y_{32}'Y_{32}-Y_{32}'Y_{31}(Y_{31}'Y_{31})^{-1}Y_{31}'Y_{32}=W\,.$$

A su vez, la matriz W puede calcularse observando que

$$P_1P_1'X=\begin{bmatrix}\bar{X}_1\\ \vdots\\ \bar{X}_1\\ \bar{X}_2\\ \vdots\\ \bar{X}_2\\ \vdots\\ \bar{X}_r\\ \vdots\\ \bar{X}_r\end{bmatrix}\begin{array}{l}\left.\begin{array}{l}\\ \\ \\ \end{array}\right\} n_1 \text{ renglones}\\ \left.\begin{array}{l}\\ \\ \\ \end{array}\right\} n_2 \text{ renglones}\\ \vdots\\ \left.\begin{array}{l}\\ \\ \\ \end{array}\right\} n_r \text{ renglones}\end{array}$$

de manera que

$$\begin{aligned}W&=X'(I-P_1P_1')X\\ &=\sum_{i=1}^{r}\sum_{j=1}^{n_i}X_{ij}(X_{ij}-\bar{X}_i)\\ &=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2\\ &=n\hat{\sigma}^2\end{aligned}$$

El denominador de la estadística de prueba puede simplificarse considerando

$$\begin{aligned}(Y_{21}'Y_{21}+Y_{31}'Y_{31})^{-1}&=Q_1'\hat{\theta}'A'(AA')^{-1}A\hat{\theta}Q_1+Q_1'WQ_1\\ &=Q_1'ZQ_1\,,\end{aligned}$$

donde $Z=\hat{\theta}'A'(AA')^{-1}A\hat{\theta}+W$. Observando que

$$(Q_1'ZQ_1)^{-1} = Q_1'Q_1(Q_1'ZQ_1)^{-1}Q_1'Q_1 = Q_1'[Z^{-1}-Z^{-1}Q_2(Q_2'Z^{-1}Q_2)^{-1}Q_2'Z^{-1}]Q_1$$
$$= Q_1'\,0\,Q_1$$
$$=0$$

dado que en este caso $Q_2 = \pm 1$, se sigue que el denominador de la estadística de prueba puede obtenerse como

$$Y_{32}'Y_{32} + (Y_{22}-\tilde{C})'(Y_{22}-\tilde{C}) = W + (A\hat{\theta}-C)'(AA')^{-1}(A\hat{\theta}-C) \ ,$$

de manera que la estadística de prueba está dada por

$$T = \frac{W}{W+(A\hat{\theta}-C)'(AA')^{-1}(A\hat{\theta}-C)} \ ,$$

donde $W = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2$. La región de rechazo de H_0 es de la forma

$$C = \{X_{n\times 1} | T < k\}$$

y la constante de la región de rechazo se determina utilizando que $T \sim \Lambda(1,n-r,t)$. Una prueba equivalente puede derivarse de la ecuación (5.3) utilizado como estadística de prueba la variable.

$$T' = \left(\frac{n-r}{t}\right)\frac{1-T}{T} = \frac{(A\hat{\theta}-C)'(AA')^{-1}(A\hat{\theta}-C)\,/\,t}{W/(n-r)} \ ,$$

la cual sigue una distribución $F(t,n-r)$ y tiene asociada un región critica de la forma

$$C = \left\{X_{n\times 1} | T' > k\right\} \ .$$

7.2.2 El Modelo Multivariado con un Criterio de Clasificación

El modelo multivariado con un criterio de clasificación está definido por las relaciones

$$E(X_{ij}) = \delta + \alpha_i \qquad i=1,\dots,r \qquad j=1,\dots,n_i \ ,$$

donde X_{ij} $i=1,\dots,r$ $j=1,\dots,n_i$ son variables aleatorias con distribución independiente y matriz de covarianzas común Σ. En notación matricial, este modelo puede escribirse como

$$E(X_{nxp}) = D_{nx(r+1)}\mu_{(r+1)xp} \, ,$$

donde

$$X = \begin{bmatrix} X'_{11} \\ \vdots \\ X'_{1n_1} \\ \vdots \\ X'_{r1} \\ \vdots \\ X'_{rn_r} \end{bmatrix} , \qquad \mu = \begin{bmatrix} \delta' \\ \alpha'_1 \\ \vdots \\ \alpha'_r \end{bmatrix}$$

y

$$D_{nx(r+1)} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & & \vdots \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & & \vdots \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & & & & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix}1\\1\\ \vdots\\1\end{matrix}}\right] n_1 \text{ renglones} \\ \left.\vphantom{\begin{matrix}1\\1\\ \vdots\\1\end{matrix}}\right] n_2 \text{ renglones} \\ \vdots \\ \left.\vphantom{\begin{matrix}1\\1\\ \vdots\\1\end{matrix}}\right] n_r \text{ renglones} \end{array}$$

Definiendo $p=q$, $g=r+1$ y haciendo $E=I_p$, el análisis de este modelo se obtiene utilizando los resultados disuctidos en la sección 6.3. Como la matriz μ no es estimable, resulta necesario reparametrizar el modelo para lo cual se considera la siguiente descomposición de la matriz D

$$D_{nx(r+1)} = P_{1_{nxr}} U_{rx(r+1)} \, ,$$

donde

$$P_1 = \begin{array}{c} \left[\begin{array}{cccccc} n_1^{-\frac{1}{2}} & 0 & 0 & \cdots & 0 & 0 \\ n_1^{-\frac{1}{2}} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & \vdots \\ n_1^{-\frac{1}{2}} & 0 & 0 & \cdots & 0 & 0 \\ 0 & n_2^{-\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ 0 & n_2^{-\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & n_2^{-\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & n_r^{-\frac{1}{2}} \\ 0 & 0 & 0 & \cdots & 0 & n_r^{-\frac{1}{2}} \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & n_r^{-\frac{1}{2}} \end{array}\right] \end{array} \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right] n_1 \text{ renglones} \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right] n_2 \text{ renglones} \\ \vdots \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right] n_r \text{ renglones} \end{array}$$

$$U_{a\times(a+1)} \begin{bmatrix} \sqrt{n_1} & \sqrt{n_1} & 0 & 0 & \cdots & 0 & 0 \\ \sqrt{n_2} & 0 & \sqrt{n_2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \\ \sqrt{n_a} & 0 \cdots & 0 \cdots & 0 & \cdots & 0 & \sqrt{n_{r+1}} \end{bmatrix}$$

y se define una reparametrización por la relación

$$\theta_{r\times p} = U_{r\times(r+1)}\,\mu_{(r+1)\times p}$$

$$= \begin{bmatrix} \sqrt{n_1}\,(\delta+\alpha_1)' \\ \vdots \\ \sqrt{n_r}\,(\delta+\alpha_r)' \end{bmatrix} .$$

Considerando la reparametrización anterior, el estimador de máxima verosimilitud de la matriz θ está dado por

$$\hat{\theta} = P_1' X$$

$$= \begin{bmatrix} n_1^{\frac{1}{2}} & \overline{X}_1' \\ \vdots & \\ n_r^{\frac{1}{2}} & \overline{X}_r' \end{bmatrix} .$$

donde

$$\overline{X}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} \qquad i=1,\dots,r\ .$$

El estimador de máxima verosimilitud de Σ se obtiene como

$$\hat{\Sigma}=\frac{1}{n}(X-P_1\hat{\theta})'(X-P_1\hat{\theta})$$

$$=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)'(X_{ij}-\overline{X}_i)\ .$$

En esta situación, el juego de hipótesis lineal general resulta definido por

$$\begin{aligned}&H_0{:}A_{t\times r}\ \hat{\theta}_{r\times p}B_{p\times u}=C_{t\times u}\\ \text{vs}\\ &H_0{:}A_{t\times r}\ \hat{\theta}_{r\times p}B_{p\times u}\neq C_{t\times u}\ ,\end{aligned} \tag{7.3}$$

donde A, B y C son matrices de constantes

$$r(A)=t \qquad t\leq r$$

y

$$r(B)=u \qquad u\leq p\ .$$

La estadística de prueba puede calcularse utilizando las matrices

$$Y_{21}=(AA')^{-\frac{1}{2}}A\hat{\theta}\,Q_1\ ,$$

$$Y_{22}=(AA')^{-\frac{1}{2}}A\hat{\theta}\,B\ ,$$

$$Y_{31}=P_2'XQ_1$$

y

$$Y_{32}=P_2'XB$$

Definiendo $W=X'(I-P_1P_1')X$, el numerador de la estadística de prueba resulta definido en este caso por

$$|Y_{32}'Y_{32}-Y_{32}'Y_{31}(Y_{31}'Y_{31})^{-1}Y_{31}'Y_{32}|$$

$$=|B'WB-B'WQ_1(Q_1'WQ_1)^{-1}Q'WB|$$

$$=|B'\{W-WQ_1(Q_1'WQ_1)^{-1}Q'W\}B|$$

$$=|B'\{Q_2(Q_2'W^{-1}Q_2)^{-1}Q_2'W\}B|$$

$$=|B'WB|$$

dado que en este caso puede seleccionarse $Q_2=I_p$. Una manera alternativa de escribir la matriz W puede obtenerse considerando que

$$P_1P_1'X = \begin{bmatrix} \bar{X}_1' \\ \vdots \\ \bar{X}_1' \\ \bar{X}_2' \\ \vdots \\ \bar{X}_2' \\ \vdots \\ \bar{X}_r' \\ \vdots \\ \bar{X}_r' \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a\\ a\\ a \end{matrix}}\right\} n_1 \text{ renglones} \\ \left.\vphantom{\begin{matrix} a\\ a\\ a \end{matrix}}\right\} n_2 \text{ renglones} \\ \vdots \\ \left.\vphantom{\begin{matrix} a\\ a\\ a \end{matrix}}\right\} n_r \text{ renglones} \end{matrix}$$

por lo que

$$\begin{aligned} E &= X'(I-P_1P_1')X \\ &= \sum_{i=1}^{r} \sum_{j=1}^{n_i} X_{ij}(X_{ij}-\bar{X}_i)' \\ &= \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij}-\bar{X}_i)(X_{ij}-\bar{X}_i)' \\ &= n\hat{\Sigma} \ . \end{aligned}$$

Definiendo $Z=\hat{\theta}'A'(AA')^{-1}A\hat{\theta}+W$ y observando que

$$\begin{aligned} (Y_{21}'Y_{21}+Y_{31}'Y_{31})^{-1} &= Q_1'Q_1(Q_1'ZQ_1)^{-1}Q_1'Q \\ &= Q_1'\{Z^{-1}-Z^{-1}Q_2(Q_2'Z^{-1}Q_2)^{-1}Q_2'Z^{-1}\}Q_1 \\ &= Q_1'0Q_1 \\ &= 0 \end{aligned}$$

dado que puede seleccionarse $Q_2=I_p$, se sigue que el denominador de la estadística de prueba puede escribirse como

$$Y_{32}'Y_{32}-(Y_{22}-\tilde{C})'(Y_{22}'-\tilde{C}) = W+(A\hat{\theta}B-C)'(AA')^{-1}(A\hat{\theta}B-C) \ ,$$

por lo que la estadística de prueba resulta dada por

$$T=\frac{|B'WB|}{|B'WB-(A\hat{\theta}B-C)'(AA')^{-1}(A\hat{\theta}B-C)|}$$

y sigue una distribución $\Lambda(u.n\text{-}r.t)$, siendo la región de rechazo de H_0 de la forma

$$\mathcal{C}=\{X|T<k\} .$$

Una hipótesis de interés que resulta en este modelo es determinar si todas las observaciones provienen de la misma población; es decir, si poseen el mismo valor medio. El juego de hipótesis correspondiente a esta aseveración puede escribirse como

$$\begin{aligned} &H_0: \alpha_1 = \ldots\ldots = \alpha_r \\ \text{vs} \\ &H_1: \text{no } H_0 \ . \end{aligned} \tag{7.4}$$

La hipótesis anterior puede plantearse en términos del juego (7.3) definiendo

$$A_{(r-1)xr}=\begin{bmatrix} n_1^{-\frac{1}{2}}, & -n_2^{-\frac{1}{2}} & , & 0\cdots, & \cdot 0\cdots, & 0\cdots\cdots\cdots & 0 & , & 0 \\ 0 & , & n_2^{-\frac{1}{2}} & ,-n_3^{-\frac{1}{2}} & , & 0\cdots, & 0\cdots\cdots\cdots & 0 & , & 0 \\ 0 & , & 0 & , & n_3^{-\frac{1}{2}} & , -n_4^{-\frac{1}{2}}, & 0\cdots\cdots\cdots & 0 & , & 0 \\ \vdots & & & & & \ddots & & \vdots & & \vdots \\ 0 \\ \vdots \\ 0 & , & 0 & , & 0 & ,\cdot 0\cdots, & 0\cdots\cdots\cdots & n_{r-1}^{-\frac{1}{2}}, & -n_r^{-\frac{1}{2}} \end{bmatrix},$$

$$B_{pxp}=I_p$$

y

$$C_{txp}=0_{txp}$$

Con estas particulares definiciones de A, B y C se tiene

$$AA'=\begin{bmatrix} \frac{1}{n_1}+\frac{1}{n_2}, & -\frac{1}{n_2} & , & 0 & , & 0 & ,\cdot 0\cdots\cdots 0 & , & 0 \\ -\frac{1}{n_2} & , & \frac{1}{n_2}+\frac{1}{n_3} & , & -\frac{1}{n_3} & , & 0 & ,\cdot 0\cdots\cdots 0 & , & 0 \\ 0 & , & -\frac{1}{n_3} & , & \frac{1}{n_3}+\frac{1}{n_3} & , & -\frac{1}{n_4} & ,\cdot 0\cdots\cdots & 0 & , & 0 \\ \vdots & & & & & \ddots & & & \vdots & & \vdots \\ 0 & , & 0 & , & 0 & , & 0 & ,\,.\,0\ldots & -\frac{1}{n_{r-1}} & , & \frac{1}{n_{r-1}}+\frac{1}{n_r} \end{bmatrix},$$

cuya inversa está dada por

$(AA')^{-1}=$

$$\frac{1}{n}\begin{bmatrix} n_1(n-n_1) & , n_1(n-n_1-n_2) & , n_1(n-n_1-n_2-n_3) & ,\cdots, & n_1(n-\sum_{i=1}^{r-2} n_i) & , n_1(n-\sum_{i=1}^{r-1} n_i) \\ n_1(n-n_1-n_2) & , (n_1+n_2)(n-n_1-n_2) & , (n_1+n_2)(n-n_1-n_2-n_3) & ,\cdots, & (n_1+n_2)(n-\sum_{i=1}^{r-2} n_i) & , (n_1+n_2)(n-\sum_{i=1}^{r-1} n_i) \\ n_1(n-\sum_{i=1}^{3} n_i) & , (n_1+n_2)(n-\sum_{i=1}^{3} n_i) & , (n_1+n_2+n_3)(n-\sum_{i=1}^{3} n_i) & ,\cdots, & (\sum_{i=1}^{3} n_i)(n-\sum_{i=1}^{r-2} n_i) & , (\sum_{i=1}^{3} n_i)(n-\sum_{i=1}^{r-1} n_i) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ n_1(n-\sum_{i=1}^{r-2} n_i) & , (n_1+n_2)(n-\sum_{i=1}^{r-2} n_i) & , (n_1+n_2+n_3)(n-\sum_{i=1}^{r-2} n_i) & ,\cdots, & (\sum_{i=1}^{r-2} n_i)(n-\sum_{i=1}^{r-2} n_i) & , (\sum_{i=1}^{r-2} n_i)(n-\sum_{i=1}^{r-1} n_i) \\ n_1(n-\sum_{i=1}^{r-1} n_i) & , (n_1+n_2)(n-\sum_{i=1}^{r-1} n_i) & , (n_1+n_2+n_3)(n-\sum_{i=1}^{r-1} n_i) & ,\cdots, & (\sum_{i=1}^{r-2} n_i)(n-\sum_{i=1}^{r-1} n_i) & , (\sum_{i=1}^{r-1} n_i)(n-\sum_{i=1}^{r-1} n_i) \end{bmatrix}$$

Utilizando las matrices anteriores se deduce que

$$A'(AA')^{-1}A=\begin{bmatrix} \frac{n-n_1}{n} & , -\frac{(n_1 n_2)^{-\frac{1}{2}}}{n} & , -\frac{(n_1 n_3)^{-\frac{1}{2}}}{n} & ,\cdots\cdots, & -\frac{(n_1 n_{r-1})^{-\frac{1}{2}}}{n} & , -\frac{(n_1 n_r)^{-\frac{1}{2}}}{n} \\ -\frac{(n_2 n_1)^{\frac{1}{2}}}{n} & , \frac{n-n_2}{n} & , -\frac{(n_2 n_3)^{-\frac{1}{2}}}{n} & ,\cdots\cdots, & -\frac{(n_2 n_{r-1})^{-\frac{1}{2}}}{n} & , -\frac{(n_2 n_r)^{-\frac{1}{2}}}{n} \\ \vdots & \vdots & & & \vdots & \vdots \\ -\frac{(n_{r-1} n_1)^{-\frac{1}{2}}}{n} & , -\frac{(n_{r-1} n_2)^{-\frac{1}{2}}}{n} & , -\frac{(n_{r-1} n_3)^{-\frac{1}{2}}}{n} & ,\cdots\cdots, & -\frac{n-n_{r-1}}{n} & , -\frac{(n_{r-1} n_r)^{-\frac{1}{2}}}{n} \\ -\frac{(n_r n_1)^{\frac{1}{2}}}{n} & , -\frac{(n_r n_2)^{-\frac{1}{2}}}{n} & , -\frac{(n_r n_3)^{-\frac{1}{2}}}{n} & ,\cdots\cdots, & -\frac{(n_r n_{r-1})^{-\frac{1}{2}}}{n} & , -\frac{n-n_r}{n} \end{bmatrix}$$

y por tanto se tiene que

$(A\hat{\theta}B-C)'(AA')^{-1}(A\hat{\theta}B-C)$

$=\hat{\theta}A'(AA')^{-1}A\hat{\theta}$

$$
\begin{array}{lllll}
=\frac{1}{n}\Big[(n-n_1)n_1\overline{X}_1\overline{X}_1' & -n_1n_2\overline{X}_1\overline{X}_2 & -n_1n_3\overline{X}_1\overline{X}_3' \cdots & -n_1n_{r-1}\overline{X}_1\overline{X}_{r-1}' & -n_1n_r\overline{X}_1\overline{X}_r' \\
\quad -n_2n_1\overline{X}_2\overline{X}_1' & +(n-n_2)n_2\overline{X}_2\overline{X}_2' & -n_2n_3\overline{X}_2\overline{X}_3' \cdots & -n_2n_{r-1}\overline{X}_1\overline{X}_{r-1}' & +n_2n_r\overline{X}_2\overline{X}_r' \\
\quad \vdots & \vdots & \vdots & \vdots & \vdots \\
\quad -n_{r-1}n_1\overline{X}_{r-1}'\overline{X}_1' & -n_{r-1}n_2\overline{X}_{r-1}'\overline{X}_2' & -n_{r-1}n_3\overline{X}_{r-1}'\overline{X}_3' \cdots & -n_{r-1}n_{r-1}\overline{X}_{r-1}\overline{X}_{r-1}' & -n_{r-1}n_r\overline{X}_{r-1}\overline{X}_r' \\
\quad -n_rn_1\overline{X}_r\overline{X}_1' & -n_rn_2\overline{X}_r\overline{X}_2' & -n_rn_3\overline{X}_r\overline{X}_3' \cdots & -n_rn_{r-1}\overline{X}_1'\overline{X}_{r-1}' & +(n-n_r)n_r\overline{X}_r\overline{X}_r\Big]
\end{array}
$$

$$
=\frac{1}{n}\Big[nn_1\overline{X}_1\overline{X}_1' - n_1\sum_{j=1}^{r}n_j\overline{X}_1\overline{X}_j' - nn_2\overline{X}_2\overline{X}_2' - n_2\sum_{j=1}^{r}n_2\overline{X}_2\overline{X}_j' \cdots\cdots
$$

$$
\cdots + nn_{r-1}\overline{X}_{r-1}\overline{X}_{r-1}' - n_{r-1}\sum_{j=1}^{r}\overline{X}_{r-1}\overline{X}_j' + nn_r\overline{X}_r\overline{X}_r' - n_r\sum_{j=1}^{r}n_j\overline{X}_r\overline{X}_j'\Big]
$$

$$
=\Big[\sum_{j=1}^{r}n_i\overline{X}_i\overline{X}_i' - \frac{1}{n}\sum_{i=1}^{r}n_i\sum_{j=1}^{r}n_j\overline{X}_i\overline{X}_j'\Big]
$$

$$
=\Big[\sum_{i=1}^{r}n_i\overline{X}_i\overline{X}_i' - \sum_{i=1}^{r}n_i\overline{X}_i\frac{1}{n}\sum_{j=1}^{r}n_j\overline{X}_j'\Big]
$$

$$
=\Big[\sum_{i=1}^{r}n_i\overline{X}_i\overline{X}_i' - \sum_{i=1}^{r}n_i\overline{X}_i\overline{X}'\Big]
$$

$$
=\Big[\sum_{i=1}^{r}n_i\overline{X}_i(\overline{X}_i-\overline{X}')\Big]
$$

$$
=\Big[\sum_{i=1}^{r}n_i(\overline{X}_i-\overline{X})'(\overline{X}_i-\overline{X})'\Big],
$$

por lo que la estadística de prueba asociada al juego de hipótesis (7.4) puede escribirse como

$$
T=\frac{\left|\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)(X_{ij}-\overline{X}_i)'\right|}{\left|\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)(X_{ij}-\overline{X}_i)'+\sum_{i=1}^{r}n_i(\overline{X}_i-\overline{X})'(\overline{X}_i-\overline{X})'\right|}
$$

y sigue una distribución $\Lambda(p,n-r,r-1)$. Esta forma de la estadística de prueba coincide con la presentada en Mardia, Kent & Bibby (1979, Cap. 12).

8

CONCLUSIONES

8 Conclusiones

El material presentado en este trabajo permite hacer una serie de consideraciones referentes al análisis de modelos lineales que pueden sintetizarse de la siguiente manera:

La proposición de un modelo lineal generalizado de análisis de varianza multivariado realizada por Potthoff & Roy (1964), ha dado origen a una serie de trabajos en los que se discuten diferentes proposiciones y apectos del análisis de este modelo. La importancia del modelo de Potthoff y Roy es clara, considerando el hecho de que permite englobar una serie de modelos que históricamente han sido discutido por separado en la literatura, como es el caso de los modelos de análisis de regresión, los modelos de diseño experimental y algunos modelos de crecimiento entre otros. Sin embargo y hasta la fecha, los procedimientos de análisis estadístico que se han sugerido para este modelo generalizado presentan deficiencias, algunas de las cuales ya han sido señaladas en la literatura y otras que se discuten en este trabajo.

Diferentes ventajas asociadas con determinar un esquema de análisis apropiado del modelo de Potthoff y Roy pueden señalarse. En el análisis de modelos de diseño experimental, no se requiere de la elaboración de un cociente de verosimilitudes para cada modelo y para cada hipótesis particular, determinado en cada caso la distribución correspondiente de la estadística de prueba. Existen proposiciones de reparametrizar los modelos de diseño experimental y analizar éstos utilizando técnicas de análisis de regresión, sin embargo la elección de una particular reparametrización en la mayoria de los casos no resulta clara. La ventaja de analizar los modelos de diseño experimental con técnicas de análisis de regresión reside en el hecho de que las técnicas de análisis elaboradas para éstos, requieren de procedimientos matemáticos más simples e incluso existe la ventaja de que los supuestos del modelo tales como normalidad, independencia y varianza constante de los errores, pueden ser analizados mediante una serie de graficas de residuos contra diferentes variables como son las variables esplicativas, el valor estimado de la variable dependiente, etc. Esta verificación de la suposición del modelo a través de gráficas de residuos no resulta una característica común en los esquemas de análisis de modelos de diseño experimental. En los modelos multivariados de análisis de regresión puede señalarse que existen diferentes proposiciones de lo que es una hipótesis lineal general, que en algunas situaciones corresponden a casos particulares de la hipótesis lineal general propuesta por Potthoff y Roy y en otras coinciden con ésta.

Entre las apotaciones de este trabajo pueden señalarse las siguientes:

-Se elaboran teoremas generales sobre distribución de matrices de formas cuadráticas.

-Se expone una derivación exacta, mediante procedimientos matemáticos simples, del cociente de verosimilitudes asociado a una hipótesis lineal general en el modelo de Potthoff y Roy.

-Se propone una manera de analizar el concepto de estimabilidad lineal en el caso multivariado, que reproduce como caso particular la definición existente para los modelos univariados.

-Se define el concepto de parametrizaciones analizándose algunas de sus propiedades. Mediante esta defınción es posible generar reparametrizaciones de manera simple para los modelos de rango incompleto.

-Se deriva la distribución exacta de la estadística de prueba de cociente de verosimilitudes generalizado asociada a una hipótesis lineal general en el modelo de Potthoff y Roy.

Gran parte del análisis de modelo multivariado de Potthoff y Roy asociado con la definición de reparametrizaciones y los procesos de estimación y contraste de hipótesis requiere del cálculo de descomposiciones y productos de matrices. Histόricamente este tipo de procedimientos no resultaban atractivos dada la complejidad de los cálculos involucrados en una aplicación particular; sin embargo, cabe señalar que con los actuales equipos de cómputo, incluyendo los de uso personal, es posible manejar con relativa facilidad los aspectos numéricos asociados con operaciones matriciales. Resulta oportuno mencionar que una implementación adecuada de los procedimientos de análisis de modelos lineales que se discuten en este trabajo, resultaria en un paquete de cómputo con una gran versatilidad en comparación a la mayoria de los paquetes de cómputo que actualmente circulan en el mercado, cuyas capacidades se limitan por ejemplo al análisis de modelos de diseño experimental con algunos esquemas predeterminados de diseños.

Por último, resulta de interés señalar algunas lineas de investigación que pueden seguirse en el análisis del modelo Potthoff y Roy, entre las cuales se pueden mencionar las siguientes:

-La elaboración de intervalos de confianza con cuantiles obtenidos de manera exacta.

-Elaboración de propuestas de análisis de los supuestos del modelo a través del uso de residuos.

-Determinar la distribución del estimador de la matriz de parámetros estimables θ y la distribución del estimador de la matriz de covarianza común a las observaciones Σ.

-Determinar la forma del estimador de la matriz de parámetros estimables θ bajo la hipótesis nula.

-Determinar el cociente de verosimilitudes para la hipótesis lineal general en función de las estimaciones de la matriz θ bajo la hipótesis nula y sin restricciones.

-Elaboración de rutinas de cómputo que implementen el esquema de análisis propuesto en este trabajo.

APENDICE

Apendice

En esta sección se presentan algunos resultados básicos del álgebra de matrices y tiene el propósito de servir de referencia a los resultados utilizados en este trabajo. Las demostraciones de los teoremas no incluidas aquí pueden consultarse en textos como Mardia et al(1979), Seber(1984), Graybill(1976) ó Rao(1973). La demostración del teorema A. puede encontrarse en Deemer & Olkin (1951).

DEFINICION A.1. Una matriz A_{nxp} es un arreglo de números en n renglones y p columnas como sigue

$$A_{nxp}=(a_{ij})=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

Si n=p se dice que la matriz es cuadrada. Si A_{nxn} es una matriz con $a_{ij}=1$ i=1,.....,n y $a_{ij}=0$ $i\neq j$ se dice que A es la matriz indentidad y se denota por I.

Los renglones de la matriz A se denotan por $A'_1,\ldots,A'_n$ y las columnas por $A_{(1)},\ldots,A_{(p)}$ de manera que A puede escribirse como

$$A=\begin{bmatrix} A' \\ \vdots \\ A'_n \end{bmatrix}=(A_{(1)},\ldots\ldots,A_{(p)})$$

DEFINICION A.2. La matriz transpuesta de una matriz $A_{nxp}=(a_{ij})$ se denota por A' y se define como la matriz

$$A'_{nxp}=(a_{ji})=\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & & & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{np} \end{bmatrix}$$

DEFINICION A.3. Las matrices de una columna se denominan vector columna y se denotan por

$$a=\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} .$$

Los vectores renglón se denotan por el transpuesto del correspondiente vector columna. En este trabajo, la denominación vector corresponde a vector columna.

DEFINICION A.4. El vector de dimensión n con todas sus entradas iguales a la unidad se denota por 1_n. En algunas ocasiones, cuando la dimensión n se sobreentiende, este vector se denota simplemente por 1.

DEFINICION A.5. Si $A=(a_{ij})$ es una matriz de orden nxn, se define el vector columna Diag(A) como el vector

$$\mathrm{Diag}(A)=\begin{bmatrix} a_{11} \\ \vdots \\ a_{nn} \end{bmatrix}$$

Se define también la matriz Diag(a) donde $a'=(a_{11},\ldots,a_{nn})$ como

$$\mathrm{Diag}(A)=\begin{bmatrix} a_{11} & 0 \ldots\ldots & 0 \\ 0 & a_{22} \ldots\ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \ldots\ldots & a_{nn} \end{bmatrix}$$

DEFINICION A.6. Una matriz $A_{nxn}=(a_{ij})$ es diagonal si $a_{ij}=0$ cuando $i\neq j$.

DEFINICION A.7. Una matriz A es simétrica si A=A'.

DEFINICION A.8. La traza de una matriz A_{nxn} se denota por $tr(A)$ y se define como $tr(A)=\sum_{i=1}^{n} a_{ii}$.

DEFINICION A.9. Una matriz A_{nxp} se denomina particionada cuando se escribe en términos de submatrices; es decir

$$A_{nxp}=\begin{bmatrix} A_{11(n_1xp_1)} & A_{12(n_1xp_2)} \\ A_{21(n_2xp_1)} & A_{22(n_2xp_2)} \end{bmatrix} \tag{a.1}$$

donde $n=n_1+n_2$ y $p=p_1+p_2$.

DEFINICION A.10. Los vectores $X_1,....,X_n$ son linealmente dependientes si existen números $\delta_1,....,\delta_n$ no todos cero tales que

$$\sum_{i=1}^{n} \delta_i X_i = 0$$

Si los vectores $X_1,....,X_n$ no son linealmente dependientes se dice que son linealmente independientes.

DEFINICION A.11. El rango de una matriz A_{nxp} se denota por r(A) y se define como el máximo número de renglones linealmente independientes ó el máximo número de columnas linealmente independientes. Si el rango coincide con n ó p se dice que la matriz es de rango completo. Si n=p y r(A)=n se dice que la matriz es no singular.

DEFINICION A.12. La inversa de una matriz A_{nxn} de rango completo se denota por A^{-1} y satisface las propiedades $AA^{-1}=A^{-1}A=I$.

DEFINICION A.13. Una inversa generalizada de una matriz A_{nxp} es cualquier matriz G que satisface AGA=A.

DEFINICION A.14. Se dice que una matriz A_{nxn} es idempotente si AA=A.

DEFINICION A.15. La matriz A_{nxn} es ortogonal si su inversa coincide con su transpuesta; es decir, si AA'=I.

DEFINICION A.16. La matriz A_{nxn} es definida positiva si para todo vector no nulo $X \in \mathbb{R}^n$, $X'AX>0$ y se denota por $A>0$.

DEFINICION A.17. La matriz A_{nxn} es semidefinida positiva si para todo vecto $X \in \mathbb{R}^n$, $X'AX \geq 0$ y se denota por $A \geq 0$.

DEFINICION A.18. El determinante de una matriz cuadrada A de orden n se denota por |A| y se define como

$$|A| = \sum_{t \in P} (-1)^{\delta(t)} a_{1i}\, a_{2j}, \ldots\ldots a_{np}$$

donde P es el conjunto de todas las posibles permutaciones $(i,j,\ldots\ldots,p)$ de $(1,2,\ldots\ldots,n)$ y $\delta((t)=1$ si la permutación es impar y $\delta(t)=2$ si la permutación es par. Si $|A|\neq 0$ se dice que A es no singular.

DEFINICION A.19. Para una matriz cuadrada $A_{n\times n}=(a_{ij})$, el cofactor de a_{ij} se denota por A_{ij} y se define por $(-1)^{i+j}$ veces el menor de a_{ij}, donde el menor de a_{ij} es el determinante de la matriz que resulta al eliminar el renglón i y la columna j de la matriz A.

DEFINICION A.20. Sea $A_{n\times n}$ cualquier matriz. El polinomio $P(\lambda)=|A-\lambda I|$ tiene n raíces $\lambda_1,\ldots\ldots,\lambda_n$ denominados valores priopios de la matriz A. Los vectores $\Upsilon_1,\ldots\ldots,\Upsilon_n$ que satisfacen $A\Upsilon_i=\lambda_i\Upsilon_i$ $i=1,\ldots\ldots,n$ se denominan vectores propios de la matriz A.

DEFINICION A.21. El producto Kronecker de $A_{n\times p}$ y $B_{p\times p}$ se denota por $A\otimes B$ y se define como la siquiente matriz de dimensión $np\times nq$

$$A\otimes B = \begin{bmatrix} a_{11}B & a_{12}B \ldots\ldots & a_{1p}B \\ a_{21}B & a_{22}B \ldots\ldots & a_{2p}B \\ \vdots & & \vdots \\ a_{n1}B & a_{n2}B \ldots\ldots & a_{np}B \end{bmatrix}$$

TEOREMA A.1. Sea A una matriz cuadrada de orden n con valores propios $\lambda_1,\ldots\ldots,\lambda_n$ y sean $X_1,X_2,\ldots\ldots,X_p$ vectores de dimensión n. Las siguientes relaciones se verifican

a) $tr(A)=\sum_{i=1}^{n}\lambda_i$

b) $tr(A\pm B)=tr(A)\pm tr(B)$

c) $tr(\alpha A)=\alpha\, tr(A)$

d) $tr\left[\sum_{i=1}^{p} X'_i A X_i\right] = tr\left[A(\sum_{i=1}^{p} X_i X'_i)\right]$

e) $tr(AB)=tr(BA)$ para matrices $A_{p\times n}$ y $B_{n\times p}$.

TEOREMA A.2. Sean A y B matrices cuadradas de orden n, c una constante y $\lambda_1,\ldots,\lambda_n$ los valores propios de A. La función determinante satisface las siguientes propiedades

a) $|A| = \prod_{i=1}^{n} \lambda_i$

b) $|cA| = c^n |A|$

c) $|A^{-1}| = |A|^{-1}$ si $|A| \neq 0$

d) $|AB| = |A|\,|B|$

e) Definiendo A como en (a.1) se obtiene que

$$|A| = |A_{11}|\,|A_{22} - A_{21} A_{11}^{-1} A_{12}|$$

$$= |A_{22}|\,|A_{11} - A_{12} A_{22}^{-1} A_{21}|$$

TEOREMA A.3. Sean $A_{n\times n}$ y $B_{n\times n}$ matrices cuadradas de rango n y c una constante. Las siguientes propiedades se verifican:

a) $A^{-1} = \frac{1}{|A|} (A_{ij})'$

b) $(cA)^{-1} = c^{-1} A^{-1}$

c) $(AB)^{-1} = B^{-1} A^{-1}$

d) Sea A definida como en (a.1). La matriz A^{-1} en términos de las submatrices de A corresponde a

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$$

donde

$$A^{11} = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}$$

$$A^{12} = -A_{11}^{-1} A_{12} A^{22}$$

$$A^{21} = -A_{22}^{-1} A_{21} A^{11}$$

$$A^{22} = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}$$

TEOREMA A.4. Si G_{pxp} es una inversa generalizada de X'X donde X es una matriz de dimensión nxp entonces

a) G' es una inversa generalizada de X'X

b) XGX'X=X; es decir, GX' es una inversa generalizada de X

c) XGX' es invariante a G.

d) XGX' es simétrica lo sea ó no G.

TEOREMA A.5. Sean A_{nxp}, B_{pxq} y W_{pxq} matrices y sea W^a el vector de dimensión pq definido por

$$W^a = \begin{bmatrix} W_{(1)} \\ \vdots \\ W_{(q)} \end{bmatrix}$$

Las siguientes propiedades se satisfacen:

a) $\alpha(A \otimes B) = (\alpha A) \otimes B = A \otimes (\alpha B)$

b) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$

c) $(A \otimes B)' = A' \otimes B'$

d) $(A \otimes B)(F \otimes G) = (AF) \otimes (BG)$

e) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ si $r(A)=n=p$ y $r(B)=p=q$

f) $(A+B) \otimes C = (A \otimes C) + B \otimes C$

g) $A \otimes (B+C) = A \otimes B + A \otimes C$

h) $(AWB)^a = (B' \otimes A) W^a$

TEOREMA A.6. Si A_{nxn} es ortogonal entonces

a) $A^{-1} = A'$

b) $AA' = A'A = I$

c) $|A| = \pm 1$

d) $a_i' a_j = a_{(i)}' a_{(j)} = 0 \quad i \neq j$

e) $a_i' a_i = a_{(i)}' a_{(i)} = 1 \quad i=1,\ldots\ldots n$

TEOREMA A.7. Sean A_{nxp}, B_{pxq}, C_{nxn} y D_{pxp} matrices tales que r(C)=n y r(D)=p. La función rango satisface las siguientes propiedades:

a) $0 \leq r(A) \leq \min(n,p)$

b) $r(A)=r(A')$

c) $r(AB) = \min\{r(A), r(B)\}$

d) $r(A'A)=r(AA')=r(A)$

e) $r(CAD)=r(A)$

f) Cuando n=p, $r(A)=p \Leftrightarrow |A| \neq 0$

TEOREMA A.8. Si $A_{nxn}>0$ y $B_{nxn}\geq 0$ entonces

a) $|A|>0$

b) $A^{-1}>0$

c) $A^{-1}B \geq 0$

TEOREMA A.9. A_{nxn} es idempotente de rango r si y sólo si $\lambda_1=\lambda_2=\ldots.\lambda_r=1$ y $\lambda_{r+1}=,\ldots.,=\lambda_n=0$ donde $\lambda_1,\lambda_2,\ldots.,\lambda_n$ son los valores propios de la matriz A.

TEOREMA A.10. Todos los valores propios de una matriz simétrica son reales.

TEOREMA A.11. Si A es una matriz cuadrada de orden n y $\lambda_1,\ldots.,\lambda_n$ los valores propios de A entonces

a) $A>0 \Leftrightarrow \lambda_i>0 \quad i=1,\ldots.,n$

b) $A\geq 0 \Leftrightarrow \lambda_i \geq 0 \quad i=1,\ldots.,n$

TEOREMA A.12. El rango de una matriz A_{nxn} es igual al número de valores propios distintos de cero.

TEOREMA A.13. (Descomposición espectral). Cualquier matriz simétrica A_{nxn} puede escribirse como

$$A = U \Lambda U' = \sum_{i=1}^{n} \lambda_i U_{(i)} U'_{(i)}$$

donde $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$, $\lambda_1, \ldots, \lambda_n$ son los valores propios de la matriz A y U es una matriz ortogonal cuyas columnas son los vectores propios estandarizados de A.

TEOREMA A.14. Si $A_{nxn} > 0$, existe una matriz simétrica y definida positiva $A^{\frac{1}{2}}$ tal que $A = A^{\frac{1}{2}} A^{\frac{1}{2}}$.

TEOREMA A.15. Si $A_{nxx} > 0$ y $B_{nxn} > 0$, los valores propios de $B^{-1}A$, AB^{-1} y $B^{-\frac{1}{2}} AB^{-\frac{1}{2}}$ son los mismos.

TEOREMA A.16. Si A_{nxn} es una matriz de rango r, existe una matriz U_{nxr} con columnas ortonormales tal que $U'AU = \Lambda$, donde $\Lambda = diag(\lambda_1, \ldots, \lambda_r)$. Los valores $\lambda_1, \ldots, \lambda_r$ corresponden a los valores propios de A diferentes de cero y las columnas de U son los vectores propios correspondientes.

TEOREMA A.17. (Descomposición en valores singulares). Si A_{nxp} es una matriz de rango r, entonces A puede ser escrita como $A = U \Sigma V'$ donde U_{nxr}, V_{pxr} son matrices con columnas ortonormales; es decir, $U'U = V'V = I$ y Σ es una matriz diagonal con elementos positivos.

TEOREMA A.18. Si A_{pxp} es positiva definida y B_{pxp} es simétrica entonces existe una matriz W_{pxp} de rango completo tal que

i) $W'AW = I$

ii) $W'BW = D$

donde D es una matriz diagonal de los valores propios de BA^{-1}.

TEOREMA A.19. Sea A_{mxn} una matriz de rango r donde $m > n$. La matriz A puede expresarse como

$$A_{mxn} = U_{mxm} \begin{pmatrix} I_r \\ 0_{(m-r)xr} \end{pmatrix} V_{rxn}$$

donde $U'U = I$ y $r(V) = r$.

TEOREMA A.20. Sea A_{mxn} una matriz de rango r donde m<n. La matriz A puede expresarse como

$$A_{mxn} = U_{mxr}(I_r \quad O_{rx(n-r)})\,V_{nxn}$$

donde V'V=I y r(U)=r.

TEOREMA A.21. Sea A_{mxn} una matriz de rango r donde m>n. La matriz A puede expresarse como

$$A_{mxn} = U_{mxm}\begin{pmatrix} 0_{(m-r)xr} \\ I_r \end{pmatrix} V_{rxn}$$

donde U'U=I y r(V)=r.

TEOREMA A.22. Sea A_{mxn} una matriz de rango r donde m<n. La matriz A puede expresarse como

$$A_{mxn} = U_{mxr}(0_{rx(n-r)}\; I_{rxr})\, V'_{nxn}$$

donde V'V=I y r(U)=r

TEOREMA A.23. Sea A_{mxn} una matriz de rango r donde m>n. La matriz A puede expresarse como

$$A_{mxn} = U_{mxm}\begin{pmatrix} I_r & O_{rx(n-r)} \\ O_{(m-r)xr} & O_{(m-r)x(n-r)} \end{pmatrix} V_{nxn}$$

donde U'U=I y r(V)=n.

TEOREMA A.24. Sea A_{mxn} una matriz de rango r donde m<n. La matriz A puede expresarse como

$$A_{mxn} = U_{mxm}\begin{pmatrix} I_r & O_{rx(n-r)} \\ O_{(m-r)xr} & O_{(m-r)x(n-r)} \end{pmatrix} V_{nxn}$$

donde V'V=I y r(U)=m.

TEOREMA A.25. Sea X_{nxn} una matriz simétrica, $W'=(w_1,\ldots,w_p)$ y matrices de constantes A_{nxn} y a_{nx1}. Los siguientes resultados se verifican

a) $\dfrac{\partial a'W}{\partial W} = a$

b) $\frac{\partial W'AW}{\partial W} = 2AW$

c) $\frac{\partial |X|}{\partial x_{ij}} = \begin{cases} X_{ij} & i=j \\ 2X_{ij} & i \neq j \end{cases}$

donde X_{ij} es el cofactor i,j de X.

d) $\frac{\partial \, ln|X|}{\partial X} = 2X^{-1} - \text{Diag}\, X^{-1}$

e) $\frac{\partial \, tr XA}{\partial X} = A + A' - \text{Diag}(A)$

TEOREMA A.26. Sean A_{pxp} , B_{nxn}, X_{nxp} y μ_{nxp} matrices, donde las entradas de la matriz μ son distintas. Definiendo

$$f(\mu) = tr A(X-\mu)' B(X-\mu)$$

se tiene

$$\frac{\partial f(\mu)}{\partial \mu} = -2B(X-\mu)A$$

TEOREMA A.27. Sean A_{pxn}, B_{pxq} y μ_{nxp} matrices, donde las entradas de la matriz μ son distintas. Las siguientes propiedades se satisfacen:

a) $\frac{\partial \, tr(A\mu B)}{\partial \mu} = A'B'$

b) Si L y N son matrices simétricas.

$$\frac{\partial \, tr(\mu' A\mu B)}{\partial \mu} = 2A\mu B$$

c) Si n=p y μ es no singular

$$\frac{\partial \, ln|\mu|}{\partial \mu} = (\mu')^{-1}$$

TEOREMA A.28. Sea X_{nxp} una matriz y sea Y_{nxp} una transformación definida por Y_{nxp}=AXB. El jacobiano de la transformación inversa de Y=AXB está dado por

$$\left|\frac{\partial X}{\partial Y}\right| = |A|^{-p}|B|^{-n}$$

TEOREMA A.29. Sean $\lambda_1,\lambda_2,\ldots,\lambda_p$ valores no negativos. La media geométrica de $\lambda_1\lambda_2,\ldots,\lambda_p$ es menor o igual a la media aritmética de $\lambda_1,\lambda_2,\ldots,\lambda_p$.

DEMOSTRACION. Sea $X=(X_1,\ldots,X_p)$ y sea $f(X)$ definida por

$$f(X)=(\prod_{i=1}^{p} X_i)^2$$

El máximo de $f(X)$ sujeto a $\|X\|=1$ puede obtenerseutilizando multiplicadores de Lagrange y se alcanza en los vectore α de la forma

$$\alpha'=(\pm p^{-1/2},\ldots,\pm p^{-1/2})$$

El valor de $f(\alpha)$ es p^{-p}, por lo que se sigue que para todo vector unitario

$$Y=(\sum_{i=1}^{p} X_i^2)^{-1/2}\ (X_1,\ldots,X_p)$$

el valor de $f(Y)$ es menor o igual a p^{-p} lo cual se traduce en que

$$(\prod_{i=1}^{p} X_i^2)^{1/p}\leq p^{-1}(\sum_{i=1}^{p} X_i^2)$$

o equivalentemente para valores $\lambda_i=X_i^2\geq 0$

$$(\prod_{i=1}^{p}\lambda_i)^{1/p}\leq\frac{1}{p}\sum_{i=1}^{p}\lambda_i=\overline{\lambda})$$

lo cual demuestra el teorema.

TEOREMA A.30. Sea V_{pxp} una matriz definida positiva y sea $Q_{pxp}=(Q_1,Q_2)$ donde las matrices Q_1 y Q_2 son de rango completo con q y p-q columnas respectivamente tales que $Q_1'Q_2=0$. La siguiente ecuación se satisface

$$Q_1(Q_1'VQ_1)^{-1}Q_1'=V^{-1}-V^{-1}Q_2(Q_2'V^{-1}Q_2)^{-1}Q_2'V^{-1}$$

TEOREMA A.31. Sea μ_{gxq}, A_{axg}, B_{qxb}, C_{axg} y D_{qxb} matrices. Si la condición

$$A\mu B=C\mu D$$

se satisface para toda matriz μ, entonces $A=C$ y $B=D$.

TEOREMA A.32. Sean Σ_{pxp} y A_{pxp} matrices simétricas y definidas positivas. Sea la función $f(\Sigma)$ definida por

$$f(\Sigma) = |\Sigma|^{-n}\ exp(-\,tr\ \Sigma^{-1}A)$$

El supremo de $f(\Sigma)$ sobre Σ se alcanza en $\hat{\Sigma} = n^{-1} A$ y está dado por

$$f(\hat{\Sigma}) = |\,n^{-1}A\,|^{-n}\ exp(-\,np)$$

BIBLIOGRAFIA

Bibliografía

Anderson T. W. (1987). An Introduction to Multivariate Analysis. Wiley, New York.

Cramér H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

Deemer W.L. & Olkin I. (1951). The Jacobians of Certain Matrix Transformations Useful in Multivariate Analysis. *Biometrika* 38, 345-367.

Fujikoshi Y. (1974). The Likelihood Ratio Tests for the Dimensionality of Regression Coefficients. *Journal of Multivariate Analysis*, Vol. 4, pp. 327-340.

Geisser S. (1980). Growth Curve Analysis. In: *Handbook of Statistics*, Vol. 1. Edited by P. R. Krishnaiah, pp. 89-115. North-Holland, New York.

Gleser L.J. and Olkin I. (1970). Linear Models in Multivariate Analysis. In: *Essays in Probability and Statistics.* Edited by Bose R.C., Chakravarti I.M., Mahalanobis P.C., Rao C.R. and Smith K.J.C. The University of North Carolina Press.

Graybill F.A. (1976). *Theory and Application of the Linear Model.* Duxburry Press.

Grizzle J. and Allen D. M. (1969). Analysis of Growth and Dose Response Curves. *Biometrics* 25, pp. 357-381.

Heck D.L. (1960). Charts of Some Upper Percentage Points of the Distribution of the Largest Characteristic Root. *Annals of Mathematical Statistics* 31, pp. 625-642.

Hotelling H. (1951). A generalized T Test and Measure of Multivariate Dispersion. *Proceedings of the Second Berkeley Symposium on Mathematical statistics and Probability.* pp. 23-41. Berkeley and los Angeles: University of California Press.

Kabe D. G. (1975). Some Results for the GMANOVA Model. *Comunications in Statistics* 9, pp. 813-820.

Kenward M. G. (1986). The Distribution of a Generalized Least Squares Estimator with Covariance Adjustment. *Journal of Multivariate Analysis*, Vol. 20, pp. 244-250.

Kres H. (1983). *Statistical Tables for Multivariate Analysis.* Springer-Verlag, New York.

Lee J. C. (1982). Clasification of Growth Curves. In: *Handbook of Statistics*, Vol. 2, pp. 121-137. Edited by P. R. Krishnaiah and L. N. Kanal. North-Holland Publishing Company.

Lee Y. H. K. (1974). A Note on Rao's Reduction of Potthoff and Roy's Generalized Linear Model. *Biometrika* 61, pp. 349-351.

Mardia K.V., Kent J.T. and Bibby J.M. (1979). *Multivariate Analysis*. Academic Press, London.

Mendoza J.R. (1987). *La Distribución Normal Multivariada y su Rélación con Otras Distribuciones*. Tesis de Licenciatura en Actuaría, Facultad de Ciencias, UNAM., México.

Mendoza J.R. (1988). *Algunos Teoremas Generales sobre Distribución e Independencia de Formas Cuadráticas y Lineales en Matrices de Datos Normales*. Reporte #11, Laboratorio de Estadística, Facultad de Ciencias, UNAM.

Montgomery D.C. (1976). *Design and Analysis of Experiments*. Wiley, New York

Mudholkar G.S., Davidson M. L. and Subbaiah P. (1974). Extended Linear Hypotheses and Simultaneous Tests in Multivariate Analysis of Variance. *Biometrika* 61, pp. 457-477.

Posten H. O. and Bargmann R. E. (1964). Power of the Likelihood-Ratio Test of the General Linear Hypothesis in Multivariate Analysis. *Biometrika* 51, pp. 467-480.

Potthoff R.F. and roy S.N. (1964). A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems. *Biometrika* 51, pp. 313-326

Rao C. R. (1959). Some Problems Involving Linear Hypothesis in Multivariate Analysis. *Biometrika* 46, pp. 49-58.

Rao C. R. (1965). The Theory of Least Squares When the Parameters are Stochastic and its Application to the Analysis of Growth Curves. *Biometrika* 52, pp. 447-458.

Rao C. R. (1966). Covariance Adjustment and Related Problems in Multivariate Analysis. In: *Multivariate Analysis: Proceedings of an International Symposium*. Edited by P. R. Krishnaiah. Academic Press, New York.

Rao C.R. (1973). *Linear Statistical Inference & its Applications*. Wiley, New York.

Roy S. N. (1953). On a Heuristic Method of Test Construction and its Use in Multivariate Analysis. *Annals of Mathematical Statistics* 24, pp. 220-238.

Searle S.R. (1971). Linear Models. Wiley, New York.

Seber G.A.F. (1984). *Multivariate Observations*. Wiley, New York.

Smith, H., Granadesikan R. and Hughes J. B. (1962). Multivariate Analysis of Variance. (MANOVA). *Biometrics* 18, pp. 22-41.

Srivastava J. N. (1966). Some Generalizations of Multivariate Analysis of Variance. In: *Multivariate Analysis: Poceedings of an International Symposium*. Edited by P. R. Krishnaiah. Academic Press, New York.

Timm N. H. (1980) Multivariate Analysis of Repeated Measurements. In: *Handbook of Statistics*, Vol. 1, pp. 41-87. Edited by P. R. Krishnaiah. North-Holland, New York.

Tubbs J. D., Lewis T. O. and Duran B. S. (1975). A Note on the Analysis of the MANOVA model and its Application to Growth Curves. *Communications in Statistics* 7, pp. 643-653.

Wilks S.S. (1932). Certain Generalizations in the Analysis of Variance. *Biometrika* 24, pp. 471-494.