2Ej
34

Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS

LA DISTRIBUCION NORMAL MULTIVARIADA Y SU RELACION CON OTRAS DISTRIBUCIONES

T E S I S

Que para obtener el título de

A C T U A R I O

P R E S E N T A

JOSE RODOLFO MENDOZA BLANCO

México. D. F.

1987

LA DISTRIBUCION NORMAL MULTIVARIADA Y SU RELACION CON OTRAS DISTRIBUCIONES

PREFACIO

PREFACIO

La teoría de Análisis Multivariado, teniendo sus principios en la década de los treintas, trata con el Análisis de observaciones en más de una variable; es decir, se orienta al análisis en forma simultánea de varias mediciones o registros obtenidos para cada objeto o individuo que se estudia. De esta manera natural aparece el interés por la posible existencia de una estructura de relación o asociación entre las variables que se registran en un mismo objeto o individuo y surge la necesidad de hacer explícita esta interdependencia en el análisis de este tipo de observaciones.

La distribución Normal Multivariada ha sido un modelo fundamental en el desarrollo de las distintas técnicas del Análisis Multivariado como es el Análisis de Varianza, técnica con que se analizan modelos de Análisis de Regresión, Diseño de Experimentos y Análisis Discriminante, entre otros. Este hecho está principalmente basado en que las matemáticas se tornan demasiado complejas o intratables para otras distribuciones. Otra razón reside en el hecho de que muchos de los procedimientos desarrollados poseen propiedades óptimas bajo la suposición de Normalidad.

El propósito de este trabajo es proporcionar material didáctico que sirva de apoyo a los cursos que se imparten en la Facultad de Ciencias en el área de Estadística, particularmente de Análisis Multivariado, presentando los principales resultados que sobre la distribución Normal Multivariada existen, esencialmente en estimación y pruebas de hipótesis, manteniendo el mayor grado de generalidad posible en la presentación de los teoremas, ya que en gran parte de los libros en el área, los resultados que se presentan no poseen el grado de generalidad suficiente e incluso la demostración de algunos resultados presenta inconvenientes. Se presenta además la rela ción que existe entre la distribución Normal Multivariada y otras distribuciones de uso común en el Análisis Multivariado, como son las χ^2, Wishart, T^2 de Hotelling y Lambda de Wilks, así como algunas propiedades de estas distribuciones.

Para la comprensión del material que aquí se presenta se supone un buen conocimiento del álgebra de matrices y conocimientos básicos de Estadística Matemática. La mayor parte de los resultados sobre álgebra de matrices utilizados en este trabajo se incluyen en el apéndice como referencia. En relación a la notación utilizada en este trabajo, los vectores y matrices se denotan por letras mayúsculas y las dimensiones de éstos se hacen explícitas sólo cuando en el contexto se desea un mayor énfasis en su dimensión. Todos los teoremas aparecen numerados y aquellos que al número anteponen la letra A aparecen en el apéndice. Posiblemente resulte de alguna utilidad al lector revisar primero el apéndice con el objeto de aclarar cuestiones relacionadas con la notación utilizada en este trabajo.

INDICE

I N D I C E

PÁG.

1

LA DISTRIBUCION NORMAL MULTIVARIADA

1.1 INTRODUCCIÓN.

En el análisis de datos multivariados ha existido una tendencia a utilizar la distribución Normal Multivariada en sus diferentes campos de estudio. Existen diversas razones de ello, entre las cuales pueden mencionarse las siguientes:

a) La relativa fácil generalización del caso univariado a la distribución conjunta de p variables, lo cual no sucede con muchas otras distribuciones.

b) En el caso de variables con distribución conjunta Normal, se obtiene correlación cero si y solo si las variables tienen distribución independiente.

c) Distribuciones marginales y condicionales de variables Normales resultan ser Normales, mas aún, combinaciones lineales de variables Normales resultan ser de la misma familia con parámetros apropiados.

d) En diversos análisis se involucran promedios de observaciones por lo cual, aún cuando los datos no poseen una distribución Normal, puede recurrirse a utilizar el teorema del límite central, que demuestra que la media de las observaciones tiene una distribución asintótica Normal multivariada.

e) Los contornos de igual densidad de una distribución Normal multivariada corresponden a elipsoides que geométricamente permiten derivar e interpretar muchas de sus propiedades.

En virtud de estas razones, un gran esfuerzo ha sido encaminado a realizar desarrollos en torno a esta distribución.

1.2 FUNCIÓN DE DENSIDAD.

La normalidad de un vector aleatorio puede caracterizarse mediante la siguiente definición

DEFINICION 1.2.1 *Se dice que un vector aleatorio* $\mathbf{X}'=(x_1,\ldots x_p)$ *tiene distribución Normal multivariada si la función de densidad de* $\mathbf{X}$ *está dada por*

$$f_{\mathbf{X}}(x;\mu,\Sigma) = |2\pi\Sigma|^{-1/2} \exp\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\} \; \forall x \in \mathbb{R}^p \qquad (1.2.1)$$

donde $\mu \in \mathbb{R}^p$ *y* $\Sigma_{p\times p}$ *es definida positiva* $(\Sigma > 0)$.

El vector de medias y la matriz de varianzas y covarianzas de la distribución corresponden a

$$E(X') = \mu' = (\mu_1, \ldots, \mu_p)$$

$$V(X) = E\{(X-\mu)(X-\mu)'\} = \Sigma$$

por conveniencia cuando $\mathbf{X}$ tenga una distribución Normal multivariada de dimensión p se escribirá $\mathbf{X} \sim N_p(\mu,\Sigma)$.

Para demostrar que (1.2.1) es una función de densidad, obsérvese que $f_{\mathbf{X}}(x,\mu,\Sigma) > 0$. Para demostrar que integra 1 consideremos la integral múltiple

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(x,\mu,\Sigma)\, d x_1,\ldots,d x_p \qquad (1.2.2)$$

como $\Sigma > 0$ del teorema A.7b se sigue que $\Sigma^{-1} > 0$ y por el teorema A.14 puede escribirse $\Sigma^{-1} = \Sigma^{-1/2}\Sigma^{-1/2}$ donde $\Sigma^{-1/2}$ es simétrica y definida positiva. La ecuación (1.2.2) puede escribirse como

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |2\pi\Sigma|^{-1/2} \exp\{-\tfrac{1}{2}(\Sigma^{1/2}y)'\,\Sigma^{-1}\,(\Sigma^{1/2}y)\}|\Sigma^{1/2}|\,dy_1\ldots dy_p \quad (1.2.3)$$

$$= |2\pi\Sigma|^{-1/2}\,|\Sigma^{1/2}| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}\sum_{i=1}^{p} y_i^2\}\,dy_1\ldots dy_p$$

$$= (2\pi)^{-p/2} \prod_{i=1}^{p} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}y_i^2\}\,dy_i$$

$$= (2\pi)^{-p/2} \prod_{i=1}^{p} (2\pi)^{1/2}$$

$$= 1$$

por lo que se deduce que (1.2.1) es función de densidad. Para demostrar que el vector de medias y la matriz de covarianzas de X son μ y Σ respectivamente, obsérvese que el integrando de la ecuación (1.2.3) corresponde a la densidad del vector Y y puede escribirse como

$$f_y(y) = (2\pi)^{-p/2} \exp\{-\tfrac{1}{2}\sum_{i=1}^{p} y_i^2\}$$

$$= \prod_{i=1}^{p} (2\pi)^{-1/2} \exp\{-\tfrac{1}{2} y_i^2\} = \prod_{i=1}^{p} h_{Y_i}(y_i)$$

donde $h_{Y_i}(y_i)$ es la densidad de una variable aleatoria Normal en el punto y_i con media cero y varianza 1. La independencia de $Y_1,\ldots,Y_p$ se sigue de que la densidad conjunta puede escribirse como el producto de las densidades de cada variable, por lo que se obtiene que

$$E(Y) = 0$$

$$V(Y) = I$$

como $E(Y) = \Sigma^{-1/2} E(X - \mu) = 0$ se sigue que $E(X) = \mu$.

La matriz de covarianzas de **Y** en términos de la correspondiente a **X** es

$$V(Y) = \Sigma^{-1/2}\ V(X)\ \Sigma^{-1/2} = I$$

de donde se sigue que $V(X) = \Sigma$.

La distribución de variables aleatorias puede ser determinada mediante el uso de la función característica debido a su correspondencia uno a uno con las funciones de distribución o densidad acumulativa. La distribución de variables aleatorias también puede ser caracterizada bajo ciertas condiciones en base a los momentos de la distribución. Las definiciones 1.2.2 y 1.2.3 caracterizan a la función característica y la función generatriz de momentos.

DEFINICION 1.2.2. *Sea* $X_{p\times 1}$ *un vector aleatorio y* $t' = (t_1, \ldots, t_p)$. *La función característica de* X *denotada por* $\phi_X(t)$ *se define como*

$$\phi_X(t) = E\{\exp(it'X)\}$$

Los momentos conjuntos de la distribución pueden obtenerse mediante la relación

$$E(X_1^{K_1}, X_2^{K_2} \ldots X_p^{K_p}) = \frac{1}{i^{K_1+\cdots+K_p}} \left\{ \frac{\partial^{K_1+\cdots+K_p}}{\partial t_1^{K_1} \cdots \partial t_p^{K_p}} \phi_X(t) \right\} \Bigg|_{t=0} \qquad (1.2.4)$$

Esta igualdad puede demostrarse derivando los miembros de la ecuación

$$\phi_X(t) = \int_{-\infty}^{\infty} \exp(i t' x) f_X(x)\, dx \; ; t \in \mathbb{R}^p$$

DEFINICION 1.2.3. *Sea $X_{p\times 1}$ un vector aleatorio y $t' = (t_1, \ldots, t_p)$. La función generatriz de momentos denotada por $M_X(t)$ se define como*

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

Los momentos conjuntos de la distribución pueden obtenerse de la siguiente relación

$$E(X_1^{K_1}\ X_2^{K_2} \ldots X_p^{K_p}) = \left\{ \frac{\partial^{K_1 + \ldots + K_p}}{\partial t_1^{K_1} \ldots \partial t_p^{K_p}} M_X(t) \right\} \Bigg|_{t=0}$$

Puede demostrarse la igualdad anterior obteniendo las derivadas de ambos miembros de la siguiente ecuación.

$$M_X(t) = \int_{-\infty}^{\infty} \exp(t'x) f_X(x)\, dx \; ; \; t \in \mathbb{R}^p$$

TEOREMA 1.2.1. *Sea $X \sim N_p(\mu, \Sigma)$. La función característica de X está dada por*

$$\phi_X(t) = \exp\{i t' \mu - \tfrac{1}{2}\, t' \Sigma t\}$$

DEMOSTRACION. La función característica de X puede obtenerse mediante el uso de la transformación $Y = \Sigma^{-1/2}(X-\mu)$ donde $E(Y)=0$ y $V(Y)=I$. La función característica de Y puede escribirse como

$$\phi_Y(t) = E\{\exp(it'Y)\} = \prod_{i=1}^{p} \phi_{Y_i}(t_i)$$

por ser $Y_1,\ldots,Y_p$ variables independientes. El valor $\phi_{Y_i}(t_i)$ corresponde a

$$E\{\exp(it_iY_i)\} = \int_{-\infty}^{\infty} \exp(it_iy_i)(2\pi)^{-1/2}\exp\{-\tfrac{1}{2}y_i^2\}\,dy_i$$

$$= (2\pi)^{-1/2}\exp\{-\tfrac{1}{2}t_i^2\}\int_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}(y_i - it_i)^2\}\,dy_i$$

$$= \exp\{-\tfrac{1}{2}t_i^2\}$$

por lo que

$$\phi_X(t) = \prod_{i=1}^{p} \phi_{Y_i}(t_i) = \exp\{-\tfrac{1}{2}t't\}$$

de donde se obtiene que

$$\phi_X(t) = E\{\exp(it'X)\} = E\{\exp(it'[\Sigma^{1/2}Y+\mu])\}$$

$$= \exp(it'\mu)E\{\exp(it'\Sigma^{1/2}Y)\}$$

$$= \exp\{it'\mu - \tfrac{1}{2}(t'\Sigma^{1/2})(t'\Sigma^{1/2})'\}$$

$$= \exp\{it'\mu - \tfrac{1}{2}t'\Sigma t\}$$

lo cual demuestra el teorema.

Los teoremas 1.2.2 a 1.2.4 establecen resultados correspondientes a la determinación de la distribución de variables aletorias mediante el uso de la función característica y mediante los momentos de la distribución.

TEOREMA 1.2.2 *(Teorema de Unicidad). Dos vectores aleatorios tienen la misma función característica si y sólo si tienen la misma función de densidad.*

DEMOSTRACION. Ver Cramér(1968, pág.117).

TEOREMA 1.2.3. *Sean X y Y variables aleatorias con momentos r-ésimos definidos por* $\mu_x^r = E(X^r) = \mu_x^r = E(Y^r)$ $r=0,1,2,\ldots$. *Si* μ_x^r *es finito para todo valor r, entonces las variables X y Y tienen la misma distribución si cualesquiera de las propiedades siguientes se satisface*

a) *La serie* $\sum_{r=0}^{\infty} \frac{\mu_x^r}{r!} \delta^r$ *es absolutamente convergente para alguna constante* $\delta > 0$.

b) $\limsup_{r \to \infty} \frac{1}{r} (\mu_x^r)^{1/r} < \infty$.

c) *X es una variable acotada.*

DEMOSTRACION. Para los incisos a) y c) ver Wilks(1962, pp.125 y 126). Para la demostración del inciso b) consultar Feller(1978, pág.572).

COROLARIO 1.2.3. *Si X y Y son vectores aleatorios tales que tienen la misma función generatriz de momentos; es decir* $M_X(t) = M_Y(t)$ *para todo vector t en un rectángulo abierto que incluye el origen, entonces X y Y tienen la misma distribución.*

TEOREMA 1.2.4. *(Teorema de Inversión). Si la función característica $\phi_X(t)$ es absolutamente integrable, entonces X tiene función de densidad de probabilidad dada por*

$$f_X(x) = (2\pi)^{-p} \int_{-\infty}^{\infty} \exp(it'x)\phi_X(t)\, d\, t$$

DEMOSTRACION. Ver Cramér(1968,pág.117).

La independencia entre dos conjuntos de variables puede establecerse mediante el uso de la función característica por medio del siguiente teorema.

TEOREMA 1.2.5. *Sea X_{px1} un vector aleatorio tal que $X'=(X_1',X_2')$. Los vectores X_1 y X_2 son estadísticamente independientes si y sólo si la función característica de X se factoriza como el producto de las funciones características marginales de X_1 y X_2.*

DEMOSTRACION. Supongamos primero que X_1 y X_2 son vectores independientes y calculemos la función característica de X escribiendo $t'=(t_1',t_2')$ obtenemos

$$\phi_X(t) = E\{\exp(it'X)\} = \int_{-\infty}^{\infty} \exp(it'x)\ f_X(x)\, d\, x$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(it_1'x_1 + it_2'x_2) f_{X_1}(x_1) f_{X_2}(x_2) dx_1\, dx_2$$

$$= \phi_{X_1}(x_1)\ \phi_{X_2}(x_2)$$

La demostración inversa resulta de utilidad para caracterizar la independencia entre variables. Sea $X'=(X_1',X_2')$ y $t'=(t_1',t_2')$ tales que

$\phi_X(t) = \phi_{X_1}(t_1)\,\phi_{X_2}(t_2)$. Por el teorema 1.2.4 la densidad $f_X(x)$ puede expresarse como

$$f_X(x) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \exp(it'x)\,\phi_X(t)\,dt$$

$$= \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(it_1'x_1 + it_2'x_2)\,\phi_{X_1}(t_1)\,\phi_{X_2}(t_2)\,dt_1\,dt_2$$

$$= f_{X_1}(x_1)\, f_{X_2}(x_2)$$

de donde se implica la independencia entre X_1 y X_2.

Una caracterización alternativa a la definición 1.2.1 de Normalidad Multivariada se obtiene mediante el siguiente teorema

TEOREMA 1.2.6. *Un vector aleatorio* X *tiene distribución Normal de dimensión* p *si y sólo si* $Y = C'X$ *tiene distribución Normal univariada para todo vector fijo* C *no nulo.*

DEMOSTRACION. La equivalencia entre la definición 1.2.1 y este teorema pude establecerse mediante el uso de funciones características. Primero demostraremos que si **X** es un vector con función de densidad dada por (1.2.1) satisface las condiciones postuladas por este teorema. Por el teorema (1.2.1), la función característica de un vector con función de densidad dada por (1.2.1) corresponde a

$$\phi_X(t) = \exp\{it'\mu - \tfrac{1}{2}\, t'\Sigma t\} \tag{1.2.5}$$

En el caso de una densidad Normal univariada de una variable **W** con media α y varianza β^2, la función característica se obtiene como un caso particular de (1.2.5) y está dada por

$$\phi_W(t) = E\{\exp(i t W)\} = \exp\{i t \alpha - \frac{1}{2} t^2 \beta^2\} \tag{1.2.6}$$

La función característica de Y en términos de la función característica de X corresponde a

$$\phi_Y(k) = E\{\exp(i k Y)\} = E\{\exp(i k C'X)\}$$

$$= E\{\exp(i t'X)\} \quad ; \qquad t = k\,C$$

y utilizando (1.2.5) se obtiene

$$\phi_Y(k) = \exp(i t'\mu - \frac{1}{2} t'\Sigma t)$$

$$= \exp(i k C'\mu - \frac{1}{2} k^2 C'\Sigma C) \tag{1.2.7}$$

de donde se deduce por la ecuación (1.2.6) que (1.2.7) corresponde a la función característica de una variable Normal univariada de media $C'\mu$ y varianza $C'\Sigma C$. Del teorema 1.2.2 se sigue que $Y = C'X$ tiene distribución $N_1(C'\mu, C'\Sigma C)$.

La demostración inversa corresponde a la implicación de la definición 1.2.1 suponiendo las condiciones del teorema 1.2.6. Para esto tomemos la definición de Normalidad implicada por el teorema 1.2.6; es decir, sea X un vector aleatorio de dimensión p tal que $Y = C'X$ tiene una distribución Normal univariada para todo vector fijo C. Sean α y β^2 la media y varianza de Y. Utilizando (1.2.6) se obtiene

$$\phi_Y(k) = E\{\exp(i k Y)\} = \exp(i k\alpha - \frac{1}{2} k^2 \beta^2) \tag{1.2.8}$$

La función $\phi_Y(k)$ puede escribirse en términos de **X** como

$$\phi_Y(k) = E\{\exp(i\,k\,Y)\} = E\{\exp(i\,k\,C'X)\} = \phi_X(t)\,;\; t = k\,C \qquad (1.2.9)$$

de las ecuaciones (1.2.8) y (1.2.9) se obtiene

$$\phi_X(t) = \phi_Y(k) = \exp\{i\,k\,\alpha - \tfrac{1}{2}\cdot k^2\beta^2\} \qquad (1.2.10)$$

Por otra parte, la media y varianza de Y satisfacen las relaciones

$$\alpha = E(Y) = E(C'X) = C'\mu$$

$$\beta^2 = V(Y) = V(C'X) = C'V(X)C = C'\Sigma\,C$$

Substituyendo estas relaciones en (1.2.10) se obtiene

$$\phi_X(t) = \exp\{i\,k\,C'\mu - \tfrac{1}{2}\,(k\,C')\,\Sigma(k\,C)\}$$

$$= \exp\{i\,t'\mu - \tfrac{1}{2}\,t'\,\Sigma\,t\}$$

Por el teorema 1.2.1, esta función característica corresponde a la de una variable con función de densidad dada por (1.2.1) y por el teorema 1.2.2, la función de densidad de X corresponde a la definida en (1.2.1).

Geométricamente, la distribución Normal multivariada tiene densidad constante en elipsoides. Este hecho se deduce igualando a una constante la densidad (1.2.1) de donde se implica que

$$(x-\mu)'\;\Sigma^{-1}\,(x-\mu) = \delta \qquad (1.2.11)$$

donde δ es una constante apropiada. La ecuación (1.2.11) define un elipsoide en $\mathbb{R}^p$. Cuando $\Sigma=I$ la ecuación define una hiperesfera.

Puede obtenerse una interpretación de los elipsoides de **igual** densidad utilizando la descomposición espectral de la matriz Σ como $U\Lambda U'$ (Teorema A.13), donde Λ es una matriz diagonal cuyas entradas corresponden a los valores propios de Λ y U es una matriz ortogonal cuyas columnas corresponden a los vectores propios de Σ. Utilizando la transformación de componentes principales definida por $Y=U'(X-\mu)$, la ecuación (1.2.11) puede escribirse como

$$
\begin{aligned}
(X-\mu)'(U\Lambda U')^{-1}(X-\mu) &= (X-\mu)\,U\Lambda^{-1}U'(X-\mu)\\
&= Y'\Lambda^{-1}Y\\
&= \sum_{i=1}^{p}\frac{Y_i^2}{\lambda_i} = \delta
\end{aligned}
$$

donde los componentes del vector Y representan los ejes del elipsoide. Los valores $(\delta\lambda_i)^{1/2}$ $i=1,\ldots,p$, definen las intersecciones de la elipse con los ejes coordenados.

Si se desea obtener un elipsoide de concentración de 100 α %, el valor de δ puede determinarse de la relación $P((x-\mu)'\Sigma^{-1}(x-\mu)\leq\delta)=\alpha$. Por el corolario 3.2.3, la variable $Z=(x-\mu)'\Sigma^{-1}(x-\mu)\sim\chi^2_p$ y por tanto δ es el cuantil de orden α de la distribución χ^2_p .

EJEMPLO 1.2.1. La figura 1.2.1 muestra un elipsoide de concentración de la distribución Normal bivariada de parámetros

$$\mu=\begin{bmatrix}5\\3\end{bmatrix} \qquad \Sigma=\begin{bmatrix}1.65 & -0.45\\ -0.45 & 2.85\end{bmatrix} \tag{1.2.12}$$

La descomposición en valores singulares de Σ está dada por

$$\Sigma = U \Lambda U' = \begin{bmatrix} 1/\sqrt{10} & -3/\sqrt{10} \\ 3/\sqrt{10} & -1/\sqrt{10} \end{bmatrix} \begin{bmatrix} 3.0 & 0 \\ 0 & 1.5 \end{bmatrix} \begin{bmatrix} 1/\sqrt{10} & 3/\sqrt{10} \\ -3/\sqrt{10} & 1/\sqrt{16} \end{bmatrix}$$

Para este caso, la transformación en componentes principales resulta definida por

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = U'(X-\mu) = \begin{bmatrix} \frac{1}{\sqrt{10}}(X_1+3X_2-14) \\ \frac{1}{\sqrt{10}}(-3X_1+X_2+12) \end{bmatrix}$$

Los ejes coordenados Y_1 y Y_2 pueden obtenerse en función de X_1 y X_2 mediante las ecuaciones $Y_2 = 0$ y $Y_1 = 0$ respectivamente. La gráfica siguiente muestra el elipsoide de concentración del 90%.

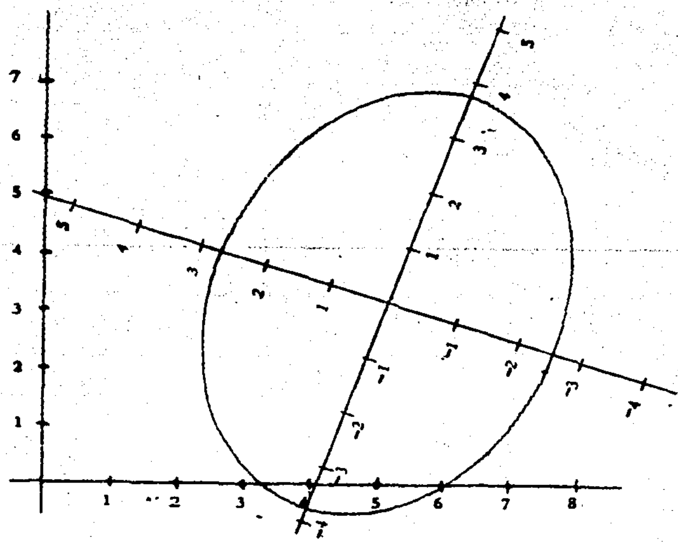

FIGURA 1.2.1 Elipsoide de concentración del 90% de la distribución $N_2(\mu, \Sigma)$ donde μ y Σ se definen en (1.2.12)

1.3 ESTIMACIÓN DE PARÁMETROS .

El propósito de esta sección es determinar los estimadores máximo verosímiles de los parámetros de la distribución $N_p(\mu,\Sigma)$, en base a un conjunto de observaciones vectoriales independientes $X_1,\ldots,X_n$ donde $X_i \sim N_p(\mu,\Sigma)$ $i=1,\ldots,n$. La determinación de los estimadores se hace mediante el método de máxima verosimilitud y se presenta en el siguiente teorema.

TEOREMA 1.3.1. *Sea $X \sim N_p(\mu,\Sigma)$ y $X_1,\ldots,X_n$ una muestra aleatoria de la distribución. Los estimadores máximo verosímiles de μ y Σ están dados por*

$$\hat{\mu}=\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{\Sigma}=S=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'$$

DEMOSTRACION. La función de verosimilitud de los parámetros está dada por la ecuación

$$L(\mu,\Sigma;\underline{X}_n)=|2\pi\Sigma|^{-n/2}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\}$$

donde $\underline{X}_n=\{X_1,\ldots,X_n\}$.

Con objeto de simplificar el álgebra para maximizar la función de verosimilitud puede obtenerse una transformación monótona de L. En este caso la función logaritmo natural resulta apropiada obteniéndose

$$\ell(\mu,\Sigma,\underline{X}_n)=\ln L(\mu,\Sigma,\underline{X}_n)=-\frac{n}{2}\ln|2\pi\Sigma|-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)$$

Utilizando esta expresión, puede escribirse

$$\frac{\partial \ell(\mu,\Sigma,\underline{X}_n)}{\partial \mu} = -\frac{1}{2}\sum_{i=1}^{n}\{-2\ \Sigma^{-1}x_i+2\ \Sigma^{-1}\mu\} = n\,\Sigma^{-1}(\overline{x}-\mu) \qquad (1.3.1)$$

Para obtener las ecuaciones asociadas a la derivada de ℓ respecto a Σ definiendo $V=\Sigma^{-1}$ se sigue que la función ℓ puede escribirse como

$$\ell(\mu,\Sigma,\underline{X}_n) = -\frac{np}{2}\ \ln(2\pi)+\frac{n}{2}\ \ln|V|-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\,V(x_i-\mu)$$

Del teorema A.22 d se obtiene que

$$\frac{\partial\ \ln|V|}{\partial V} = 2\ V^{-1}-\text{Diag}\ V^{-1} = 2\ \Sigma-\text{Diag}\ \Sigma \qquad (1.3.2)$$

Utilizando este resultado se sigue que la derivada de la forma cuadrática está dada por

$$\frac{\partial \sum_{i=1}^{n}(x_i-\mu)'\,V(x_i-\mu)}{\partial V} = \sum_{i=1}^{n}\{2(x_i-\mu)(x_i-\mu)'-\text{Diag}\,(x_i-\mu)(x_i-\mu)'\}$$

$$= 2\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'-\text{Diag}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)' \qquad (1.3.3)$$

de (1.3.2) y (1.3.3) se tiene

$$\frac{\partial \ell(\mu,V,x_n)}{\partial V} = \frac{n}{2}\ (2\,\Sigma-\text{Diag}\ \Sigma)-\frac{1}{2}\{2\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'-\text{Diag}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'\}$$

$$= n\,\Sigma-\frac{n}{2}\ \text{Diag}\ \Sigma-n\,S^{*}+\frac{n}{2}\ \text{Diag}\ S^{*} \qquad (1.3.4)$$

donde

$$S^* = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)(X_i - \mu)'$$

Igualando a cero la ecuación (1.3.1) se obtiene

$$-n\,\Sigma^{-1}(\bar{X} - \hat{\mu}) = 0$$

de donde se sigue que el vector que anula la primera derivada está dado por $\hat{\mu} = \bar{X}$. Substituyendo el valor de μ por $\hat{\mu}$ en (1.3.4) e igualando a cero se produce la ecuación

$$n(\hat{\Sigma} - S) - \frac{n}{2}\,\mathrm{Diag}(\hat{\Sigma} - S) = 0 \qquad (1.3.5)$$

donde

$$S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})'$$

De (1.3.5) se sigue que

$$2(\hat{\Sigma} - S) - \mathrm{Diag}\,(\hat{\Sigma} - S) = 0$$

lo cual ocurre si y solo si $\hat{\Sigma} = S$.

Para demostrar que $\hat{\mu}$ y $\hat{\Sigma}$ maximizan $\ell(\mu, \Sigma\, \underline{X}_n)$ observamos que el valor de μ que maximiza la función ℓ es el que minimiza la expresión

$$\sum_{i=1}^{n} (X_i - \mu)'V(X_i - \mu) = \sum_{i=1}^{n} (X_i - \bar{X})'V(X_i - \bar{X}) + n(\bar{X} - \mu)'V(\bar{X} - \mu) \qquad (1.3.6)$$

Como V es definida positiva $n(\bar{X}-\mu)'V(\bar{X}-\mu)>0$ y es cero si y solo si $\bar{X}-\mu=0$, de donde se sigue que $\hat{\mu}=\bar{X}$ es el valor de μ que minimiza (1.3.6) y por tanto maximiza $\ell(\mu,\Sigma,\underline{X}_n)$ por lo que

$$\ell(\bar{X},\Sigma,\underline{X}_n)>\ell(\mu,\Sigma,\underline{X}_n)\ \forall\mu \tag{1.3.7}$$

Para demostrar que $\hat{\Sigma}=S$ maximiza $\ell(\hat{\mu},\Sigma,\underline{X}_i)$ respecto a Σ obtenemos la diferencia

$$\ell(\hat{\mu},\hat{\Sigma},\underline{X}_n)-\ell(\hat{\mu},\Sigma,\underline{X}_n)=-\frac{n}{2}\ \ell n\ |S|+\frac{n}{2}\ \ell n\ |\Sigma|-\frac{1}{2}\sum_{i=1}^{n}(X_i-\bar{X})\ S^{-1}(X_i-\bar{X})$$

$$+\frac{1}{2}\sum_{i=1}^{n}(X_i-\bar{X})\ \Sigma^{-1}\ (X_i-\bar{X})$$

$$=-\frac{n}{2}\ \ell n\ |\Sigma^{-1}S|-\frac{np}{2}+\frac{n}{2}\ t_r(\Sigma^{-1}S)$$

$$=\frac{np}{2}\ \{\frac{1}{p}\ tr\ \Sigma^{-1}S-1-\ell n\ |\Sigma^{-1}S|^{1/p}\}$$

Como $\Sigma>0$ por el teorema A.7b $\Sigma^{-1}>0$ y del Teorema A.14 se sigue que existe $\Sigma^{-1/2}$. Como se demostrará en (1.3.10), $S>0$ por lo que utilizando el teorema A.7c se obtiene $\Sigma^{-1}S>0$ y del teorema A.15 se sigue que los valores propios positivos de $\Sigma^{-1}S$ son los mismos que los de $W=\Sigma^{-1/2}S\,\Sigma^{-1/2}$. La matriz W es semidefinida positiva ya que

$$Y'WY=(\Sigma^{-1/2}Y)'S\,(\Sigma^{-1/2}Y)>0$$

por ser $S>0$, de donde se sigue que si denotamos por $\lambda_1,\ldots,\lambda_p$ los valores propios no negativos de $\Sigma^{-1}S$, entonces

$$\frac{1}{p}\,\mathrm{tr}\,\Sigma^{-1}S = \frac{1}{p}\sum_{i=1}^{p}\lambda_i = \delta_1 > 0$$

$$|\Sigma^{-1}S|^{1/p} = \left(\prod_{i=1}^{p}\lambda_i\right)^{1/p} = \delta_2 > 0$$

donde δ_1 y δ_2 corresponden a las medias aritmética y geométrica de $\lambda_1,\ldots,\lambda_p$. La función $F(x)$ definida como sigue satisface

$$F(x) = x - 1 - \ln(x) > 0 \quad \forall x > 0 \tag{1.3.8}$$

y tiene su mínimo de cero en $x=1$. Utilizando este hecho y el teorema A.21 se obtiene

$$\begin{aligned} &\ell(\hat{\mu},\hat{\Sigma},\underline{X}_n) - \ell(\hat{\mu},\Sigma,\underline{X}_n) \\ &= \frac{np}{2}(\delta_1 - 1 - \ln\delta_2) \\ &> \frac{np}{2}(\delta_1 - 1 - \ln\delta_1) > 0 \end{aligned} \tag{1.3.9}$$

de (1.3.7) y (1.3.9) se infiere que

$$\ell(\bar{X}, S, \underline{X}_n) > \ell(\mu,\Sigma,\underline{X}_n) \quad \forall \mu,\Sigma$$

por lo que $\bar{X}$ y S maximizan la función de verosimilitud.

Una forma alternativa de expresar la información muestral $X_1,\ldots,X_n$ en un arreglo matricial se presenta mediante la siguiente definición.

DEFINICION 1.3.1. *Se dice que una matriz* X *es una matriz de datos de la distribución* $N_p(\mu,\Sigma)$ *si* X *puede escribirse como*

$$X = \begin{bmatrix} X_1' \\ \cdot \\ \cdot \\ \cdot \\ X_n' \end{bmatrix} = \begin{bmatrix} X_{11} \dots X_{1p} \\ X_{21} \dots X_{2p} \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ X_{n1} \dots X_{np} \end{bmatrix}$$

donde $X_i' = (X_{i1}, X_{i2}, \dots, X_{ip})$, $i=1,\dots,n$ *representa una muestra aleatoria de la distribución* $N_p(\mu,\Sigma)$.

Una forma alternativa de expresar $\bar{X}$ y S se obtiene escribiendo la información muestral $X_1,\dots,X_n$ como una matriz de datos de la distribución $N_p(\mu,\Sigma)$. La media de las observaciones puede escribirse en términos de la matriz de datos X de la manera siguiente

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n} X'1$$

donde 1 es un vector en $\mathbb{R}^n$ con todas sus entradas iguales a 1. La matriz S en términos de X resulta

$$\begin{aligned} S = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})' &= \frac{1}{n}\sum_{i=1}^{n} X_i X_i' - \bar{X}\,\bar{X}' \\ &= \frac{1}{n}\left(X'X - \frac{1}{n} X'1\,1'X\right) \\ &= \frac{1}{n} X'\left(I - \frac{1}{n} 1\,1'\right)X \\ &= \frac{1}{n} X'H\,X \end{aligned}$$

donde $H = I - \frac{1}{n} 1\, 1'$.

Como H es una matriz simétrica e idempotente, se sigue que

$$a'S\,a = \frac{1}{n}\, a'\, X'\, H' H\, X a = \frac{1}{n}\, Y'Y \geq 0 \quad \forall\, a \in \mathbb{R}^p \tag{1.3.10}$$

donde $Y = H X a$, lo cual demuestra que la matriz S es semidefinida positiva. En datos continuos usualmente S resulta definida positiva si el número de observaciones excede la dimensión de éstas; es decir si $n > p+1$. Este hecho se demuestra en el corolario 3.3.9.2.

El estimador máximo verosimil de la media de la distribución resulta insesgado, ya que

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = E(X) = \mu$$

La varianza de la media de la muestra puede obtenerse como

$$V(\bar{X}) = \frac{1}{n^2}\, V(\Sigma X_i) = \frac{1}{n^2}\, [\sum_{i=1}^{n} V(X_i) + \sum_{i \neq 1}^{n} C(X_i, X_j)]$$

$$= \frac{1}{n^2}\, n\, \Sigma = \frac{1}{n}\, \Sigma$$

ya que la covarianza entre X_i y X_j es cero por ser las observaciones independientes.

El estimador $\hat{\Sigma} = S$ no resulta insesgado, ya que

$$E(S) = E\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})(X_i - \mu + \mu - \bar{X})'\}$$

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

$$= \Sigma - \frac{1}{n^2}\ E\{\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)' - \sum_{i\neq j}^{n}(X_i-\mu)(X_j-\mu)\}$$

$$= \Sigma - \frac{1}{n}\ \Sigma = \frac{n-1}{n}\ \Sigma$$

De manera natural se puede considerar el estimador

$$S_u = \frac{n-1}{n}\ S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'$$

que resulta insesgado.

1.4 INDEPENDENCIA DE VARIABLES.

En esta sección se analiza el concepto de independencia entre dos conjuntos de variables bajo el supuesto de Normalidad. El teorema siguiente establece las condiciones necesarias y suficientes.

TEOREMA 1.4.1. *Sea* $\mathbf{X} \sim N_p(\mu, \Sigma)$ y $\mathbf{X}_1$ y $\mathbf{X}_2$ *tales que*

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

donde $\mathbf{X}_i$ y μ_i *tienen dimensión* p_i *y* Σ_{ij} *dimensión* $p_i \times p_j$ *con* $p = p_1 + p_2$. *Los vectores* X_1 *y* X_2 *son independientes si y sólo si* $\Sigma_{12} = 0$

DEMOSTRACION. Si $\Sigma_{12} = 0$, la forma cuadrática del exponente en la densidad de $\mathbf{X}$ puede expresarse como

$$Q = (\mathbf{x} - \mu)\, \Sigma^{-1}\, (\mathbf{x} - \mu)$$

$$= (\mathbf{x}_1' - \mu_1' \;,\; \mathbf{x}_2' - \mu_2') \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \mu_1 \\ \mathbf{x}_2 - \mu_2 \end{bmatrix}$$

$$= (\mathbf{x}_1 - \mu_1)'\, \Sigma_{11}^{-1} (\mathbf{x}_1 - \mu_1) + (\mathbf{x}_2 - \mu_2)'\, \Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2)$$

$$= Q_1 + Q_2$$

donde $Q_1 = (\mathbf{x}_1 - \mu_1)\, \Sigma_{11}^{-1} (\mathbf{x}_1 - \mu_1)$

$Q_2 = (\mathbf{x}_2 - \mu_2)\, \Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2)$

La función de densidad de **X** puede escribirse como

$$f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp(-\tfrac{1}{2} Q)$$

$$= (2\pi)^{-p_1/2} |\boldsymbol{\Sigma}_{11}|^{-1/2} \exp(-\tfrac{1}{2} Q_1) \quad (2\pi)^{-p_2/2} |\boldsymbol{\Sigma}_{22}|^{-1/2} \exp(-\tfrac{1}{2} Q_2)$$

$$= f_{\mathbf{X}_1}(\mathbf{x}_1,\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\, f_{\mathbf{X}_2}(\mathbf{x}_2;\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22})$$

donde por el teorema 1.5.1 $f_{\mathbf{X}_1}$ y $f_{\mathbf{X}_2}$, son las densidades de los vectores $\mathbf{X}_1$ y $\mathbf{X}_2$ respectivamente, por lo que $\mathbf{X}_1$ y $\mathbf{X}_2$ resultan ser independientes.

La implicación de la condición $\boldsymbol{\Sigma}_{12}=0$ por la independencia de $\mathbf{X}_1$ y $\mathbf{X}_2$ puede abordarse utilizando el teorema 1.2.5.

La función característica de **X** está dada por

$$\phi_{\mathbf{X}}(\mathbf{t}) = E\{\exp(i\mathbf{t}'\mathbf{X})\} = \exp(i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2} \mathbf{t}'\boldsymbol{\Sigma}\, \mathbf{t}) \qquad (1.4.1)$$

Las funciones características de $\mathbf{X}_1$ y $\mathbf{X}_2$ están dadas por

$$\phi_{\mathbf{X}_1}(\mathbf{t}_1) = E\{\exp(i\mathbf{t}_1'\mathbf{X}_1)\} = \exp(i\mathbf{t}_1'\boldsymbol{\mu}_1 - \tfrac{1}{2}\mathbf{t}_1' \boldsymbol{\Sigma}_{11} \mathbf{t}_1)$$

$$\phi_{\mathbf{X}_2}(\mathbf{t}_2) = E\{\exp(i\mathbf{t}_2'\mathbf{X}_2)\} = \exp(i\mathbf{t}_2'\boldsymbol{\mu}_2 - \tfrac{1}{2}\mathbf{t}_2' \boldsymbol{\Sigma}_{22} \mathbf{t}_2) \qquad (1.4.2)$$

Si $\mathbf{X}_1$ y $\mathbf{X}_2$ son independientes, la función característica de **X** puede escribirse como

$$\phi_X(t) = \phi_{X_1}(t_1)\ \phi_{X_2}(t_2)$$

y utilizando las ecuaciones (1.4.1) y (1.4.2) esta igualdad se verifica si y solo si

$$i t'\mu - \frac{1}{2} t'\Sigma t = i t_1'\mu_1 + i t_2'\mu_2 - \frac{1}{2} t_1' \Sigma_{11} t_1 - \frac{1}{2} t_2' \Sigma_{22} t_2$$

definiendo $t' = (t_1', t_2)$ la igualdad anterior se verifica si y solo si $t_1'\Sigma_{12} t_2 = 0$ para todo vector t, lo cual ocurre si y sólo si $\Sigma_{12} = 0$.

Es importante mencionar en este punto que no es suficiente que la covarianza entre dos vectores sea cero para que tengandistribución independiente, sino que además se requiere que la distribución conjunta sea Normal multivariada. Esta situación se ilustra mediante el siguiente ejemplo

EJEMPLO 1.4.1. Sea $N_p(x;\mu,\Sigma) = |2\pi\Sigma|^{-1/2} \exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}$ y $X' = (X_1, X_2)$ un vector aleatorio con función de densidad

$$f_X(x;\gamma_1) = \frac{1}{2} N_2(x;0,\Sigma_1) + \frac{1}{2} N_2(x;0,\Sigma_2)$$

$$= \frac{1}{2}|2\pi\Sigma_1|^{-1/2} \exp\{-\frac{1}{2} x'\Sigma_1^{-1} x\} + \frac{1}{2}|2\pi\Sigma_2|^{-1/2} \exp\{-\frac{1}{2} x'\Sigma_2^{-1} x\} \qquad (1.4.3)$$

donde

$$\Sigma_1 = \begin{bmatrix} 1 & \gamma_1 \\ \gamma_1 & 1 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1 & -\gamma_1 \\ -\gamma_1 & 1 \end{bmatrix}$$

y $\gamma_1 \neq 0$. Bajo estas condiciones las siguientes propiedades se satisfacen.

i) $X_1 \sim N_1(0,1)$ y $X_2 \sim N_1(0,1)$

ii) $C(X_1,X_2)=0$

iii) X_1 y X_2 no son independientes

iv) La distribución de X no es Normal bivariada.

Estas afirmaciones se demuestran a continuación

i) La distribución marginal de X_1 está dada por

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_X(x;\gamma_1)\, d\, x_2$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} N_2(x;0,\Sigma_1)\, d\, x_2 + \frac{1}{2} \int_{-\infty}^{\infty} N_2(x;0,\Sigma_2)\, d\, x_2$$

$$= \frac{1}{2} N_1(x_1;0,1) + \frac{1}{2} N_1(x_1;0,1)$$

$$= N_1(x_1;0,1).$$

Análogamente se demuestra que $f_{X_2}(x_2)=N_1(x_2;0,1)$

ii) La covarianza entre las variables X_1 y X_2 puede obtenerse como

$$C(X_1,X_2)= E(X_1 X_2)-E(X_1)E(X_2)$$

$$= \int_{-\infty}^{\infty} X_1X_2 f_X(x;\gamma_1)\, dx$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} x_1x_2 N_2(x;0,\Sigma_1)\,dx+\frac{1}{2}\int_{-\infty}^{\infty} x_1x_2 N_2(x;0,\Sigma_2)\,dx$$

$$= \frac{1}{2}\gamma_1 - \frac{1}{2}\gamma_1$$

$$= 0$$

iii) El producto de las funciones de densidad de X_1 y X_2 puede expresarse como

$$f_{X_1}(x_1)\, f_{X_2}(x_2) = (2\pi)^{-1/2} \exp\left\{-\frac{x_1^2+x_2^2}{2}\right\}$$

y esta función es distinta de $f_X(x,\gamma_1)$ definida en (1.4.3). Este hecho demuestra que aun cuando la covarianza entre X_1 y X_2 es cero, las variables no tienen distribución independiente. Esta implicación se obtiene como resultado de que la distribución conjunta de X_1 y X_2 no es Normal bivariada.

iv) Las condiciones $E(X)=0$, $V(X_1)= V(X_2)=1$ y $C(X_1,X_2)=0$ implican que si X tiene distribución Normal bivariada, de la definición 1.2.1 se sigue que la función de densidad de X está dada por

$$h_X(x)= (2\pi)^{-1} \exp\left\{-\frac{1}{2}x'x\right\}$$

Como $h_X(x)$ es diferente a la función de densidad definida en (1.4.3) se sigue que X no posee distribución Normal bivariada.

1.5. DISTRIBUCIONES MARGINALES.

El teorema siguiente caracteriza la distribución marginal de un subconjunto de variables de un vector **X** con distribución Normal multivariada.

TEOREMA 1.5.1. *Sea* $\mathbf{X} \sim N_p(\mu, \Sigma)$ y $\mathbf{X}_1$ *un subconjunto de* p_1 *variables de* **X**. *La distribución de* $\mathbf{X}_1$ *es Normal multivariada de dimensión* p_1 *con medias, varianzas y covarianzas obtenidas de las respectivas componentes de* μ *y* Σ.

DEMOSTRACION. Sin pérdida de generalidad supóngase que **X** puede escribirse como $\mathbf{X}' = (\mathbf{X}_1', \mathbf{X}_2')$. Si $\mathbf{X}_1$ no corresponde a los primeros p_1 componentes de **X**, puede modificarse el orden de las variables reordenando las correspondientes entradas de μ y Σ .

Sean

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \text{y} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

sea

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \mathbf{M}\,\mathbf{X} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{N} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{N}\mathbf{X}_1 + \mathbf{X}_2 \end{bmatrix} \qquad (1.5.1)$$

Seleccionemos **N** de manera que $\mathbf{Y}_1$ tenga covarianza nula con $\mathbf{Y}_2$; es decir

$$0 = C(Y_1, Y_2) = E\{(X_1 - \mu_1)(N X_1 + X_2 - N\mu_1 - \mu_2)'\}$$

$$= E\{(X_1 - \mu_1)(X_2 - \mu_2)' + (X_1 - \mu_1)(X_1 - \mu_1)' N'\}$$

$$= \Sigma_{12} + \Sigma_{11} N'$$

$$\Rightarrow N = -\Sigma_{21} \Sigma_{11}^{-1}$$

La matriz M dada por

$$M = \begin{bmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_2 \end{bmatrix}$$

define una transformación no singular de X en Y. Utilizando el teorema 2.2.1 se obtiene que Y tiene una distribución Normal multivariada de dimensión p.

El valor esperado de Y corresponde a

$$E(Y) = M\mu = \begin{bmatrix} \mu_1 \\ -\Sigma_{21} \Sigma_{11}^{-1} \mu_1 + \mu_2 \end{bmatrix}$$

La matriz de varianzas y covarianzas de Y_2 está dada por

$$V(Y_2) = E\{[Y_2 - E(Y_2)][Y_2 - E(Y_2)]'\}$$

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

$$= M\Sigma_{11}M' + \Sigma_{22} + M\Sigma_{12} + \Sigma_{21}N'$$

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

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

de donde

$$V(Y) = \begin{bmatrix} \Sigma_{11} & 0 \\ & \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{bmatrix}$$

Como $C(Y_1,Y_2)=0$, por el teorema 1.4.1 Y_1 y Y_2 son independientes de manera que la función de densidad de Y puede escribirse como

$$f_Y(y;M\mu;M\Sigma M') = f_{Y_1}(y_1;\mu_1,\Sigma_{11})\,f_{Y_2}(y_2;\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1,\Sigma_{22.1}) \qquad (1.5.2)$$

donde $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Las funciones f_{Y_1} y f_{Y_2} corresponden a densidades Normales multivariadas de dimensiones p_1 y $p-p_1$ respectivamente. La densidad de Y_1 puede obtenerse de (1.5.2) integrando sobre Y_2 resultando

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_Y(y;M\mu,M\Sigma M')\,dy_2$$

$$= f_{Y_1}(y_1;\mu_1,\Sigma_{11})$$

por lo que la densidad de $Y_1=X_1$ es $N_p(\mu_1,\Sigma_{11})$.

Una forma alternativa de demostrar el teorema 1.5.1 se obtiene observando que $\phi_{x_1}(t_1) = \phi_x(t_1, t_2=0)$. Por el teorema 1.2.1 se sabe que la función característica de **X** está dada por

$$\phi_x(t) = \exp(i t'\mu - \frac{1}{2} t'\Sigma t)$$

Evaluando esta función en $t_2=0$ se obtiene

$$\phi_{x_1}(t_1) = \exp(i t_1'\mu_1 - \frac{1}{2} t_1'\Sigma_{11} t_1)$$

y como esta función es igual a la función característica de una variable con distribución Normal multivariada de media μ_1 y matriz de covarianzas Σ_{11} se sigue por el teorema 1.2.2 el resultado deseado.

Como resultado del teorema 1.5.1 se sigue que si $X'=(X_1,\ldots,X_p)$ tiene distribución Normal de dimensión p cada variable $X_i; i,\ldots,p$ tiene distribución Normal univariada. El caso inverso no siempre es válido; es decir, si $X_1,\ldots,X_p$ tienen distribución Normal univariada no implica que **X** tenga distribución Normal p-variada (ver ejemplo 1.4.1).

1.6. DISTRIBUCIONES CONDICIONALES.

En algunas técnicas estadísticas multivariadas se utilizan procedimientos de selección de variables en los que resulta involucrada la distribución condicional de un conjunto de variables Normales dado otro conjunto de variables con la misma distribución. La distribución condicional se obtiene en el siguiente teorema

TEOREMA 1.6.1. *Sea* $X \sim N_p(\mu, \Sigma)$ *y sean*

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

donde X_i y μ_i *son vectores de dimensión* p_i y Σ_{ij} *es una matriz de* $p_i \times p_j$ *con* $p_1+p_2=p$. *La distribución condicional de* X_2 *dado* X_1 *es Normal de dimensión* p_2 *con media*

$$\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)$$

y matriz de varianzas y covarianzas

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

DEMOSTRACION. Sea Y definido como (1.5.1); es decir,

$$Y = MX = \begin{bmatrix} I_1 & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

La función de densidad de $\mathbf{Y}$ está dada por

$$f_Y(y, M\mu, M\Sigma M') = f_{Y_1}(y_1; \mu_1, \Sigma_{11})\, f_{Y_2}(y_2; \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \Sigma_{22.1})$$

Utilizando el teorema de cambio de variable (Hogg and Craig, 1978, secc.4.5), la densidad de $X = M^{-1}Y$ puede obtenerse como la densidad de $\mathbf{Y}$ valuada en la transformación inversa de $\mathbf{Y}$ definida por MX, multiplicada por el jacobiano de la transformación que es $|M| = 1$, esto es

$$f_X(x; \mu, \Sigma) = (2\pi)^{-p_1/2} |\Sigma_{11}|^{-1/2} \exp\{-\tfrac{1}{2}[x_1 - \mu_1]' \Sigma_{11}^{-1} [x_1 - \mu_1]\}$$

$$\cdot (2\pi)^{-p_2/2} |\Sigma_{22.1}|^{-1/2} \exp\{-\tfrac{1}{2}[x_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]' \Sigma_{22.1}^{-1}$$

$$[x_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]\} \qquad (1.6.1)$$

Naturalmente, esta densidad corresponde a la densidad $N_p(\mu, \Sigma)$. La densidad condicional de $\mathbf{X}_2$ dado $\mathbf{X}_1$ corresponde al cociente de (1.6.1) entre la densidad de $\mathbf{X}_1$; es decir

$$f_{X_2/X_1 = x_1}(x_2) = \frac{f_x(x, \mu, \Sigma)}{f_{x_1}(x_1; \mu_1, \Sigma_{11})} = (2\pi)^{-p_2/2} |\Sigma_{22.1}|^{-1/2} \exp\{-\tfrac{1}{2}[x_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]'$$

$$\Sigma_{22.1}^{-1}[\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)]\}$$

de donde se sigue que la densidad de $X_2/X_1=x_1$ es Normal multivariada de dimensión p_2 con vector de medias

$$E(X_2/X_1=x_1) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1) = \delta(x_1)$$

y matriz de covarianzas

$$V(X_2/X_1=x_1) = E\{[X_2-\delta(x_1)][X_2-\delta(x_1)]'/X_1=x_1\}$$

$$= \Sigma_{22.1}$$

lo que demuestra el teorema.

1.7 TEOREMA DEL LÍMITE CENTRAL.

Aunque gran parte de la teoría desarrollada hasta la fecha sobre Análisis Multivariado, particularmente en Análisis Discriminante, descansa sobre la suposición de Normalidad, no siempre ocurre que esta distribución proporciona una descripción satisfactoria de la variabilidad de la información. En algunos casos mediante el uso de alguna transformación de las variables originales puede resultar que la distribución Normal proporciona una descripción adecuada de la variabilidad inherente al proceso. Sin embargo, la transformación puede ser complicada o difícil de encontrar. Afortunadamente, aún en estos casos las técnicas pueden considerarse relativamente robustas puesto que típicamente basan el análisis en las medias de las observaciones o funciones de éstas. La teoría distribucional para esta situación puede abordarse recurriendo al siguiente teorema.

TEOREMA 1.7.1. *(Teorema del límite central). Si $X_1, X_2, \ldots,$ representa una sucesión infinita de vectores aleatorios independientes e idénticamente distribuídos, provenientes de una población de media μ y matriz de covarianzas Σ entonces*

$$n^{-1/2} \sum_{i=1}^{n} (X_i - \mu) = \sqrt{n}\,(\bar{X} - \mu)$$

converge en distribución a una $N_p(0,\Sigma)$ cuando el valor de n tiende a infinito.

DEMOSTRACION. Ver Anderson (1958 pág.74).

Se dice que un vector aleatorio $X_n' = (X_{n_1}, X_{n_2}, \ldots, X_{n_p})$ converge en distribución a $F_X(X)$ si

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} p(X_{n_1} \leq x_1, \ldots, X_{n_p} \leq x_p)$$

$$= P(X_1 \leq x_1, \ldots, X_p \leq x_p)$$

$$= F_X(x)$$

en los puntos de continuidad de $F_X(X)$.

Un ejemplo de aplicación del teorema del límite central es el siguiente.

EJEMPLO 1.7.1. Sea X un vector de dimensión p con distribución Multinomial con parámetros 1 y $\theta' = (\theta_1, \ldots, \theta_p)$. Sea $X_1, \ldots, X_n$ una muestra aleatoria y $Z = \sum_{i=1}^{n} X_i$. La distribución de Z es multinomial con parámetros n y θ. La esperanza del promedio de la muestra corresponde a

$$E(\overline{X}) = E(X) = \sum_{i=1}^{p} e_i P(X = e_i)$$

$$= \sum_{i=1}^{p} e_i \theta_i = \theta$$

donde e_i es el i-ésimo vector canónico de $\mathbb{R}^p$. La varianza del promedio está dada por

$$V(\overline{X}) = \frac{1}{n} V(X) = \frac{1}{n} E\{(X-\theta)(X-\theta)'\}$$

$$= \frac{1}{n} E(XX') - \frac{1}{n} \theta \theta'$$

Como los valores posibles de X son $e_1, \ldots, e_p$, la matriz XX' es una matriz que toma los valores $e_1e_1',\ldots,e_pe_p'$ con probabilidad $\theta_1,\ldots,\theta_p$ respectivamente, por lo que la varianza de $\mathbf{X}$ resulta

$$V(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\theta_i\; e_i e_i' - \frac{1}{n}\,\theta\,\theta'$$

$$= \frac{1}{n}\left\{\text{diag}\ \theta - \theta\,\theta'\right\}$$

Utilizando el teorema del límite central se obtiene que

$$\overline{X} \sim N_p\left(\theta\,,\ \frac{1}{n}\left\{\text{diag}\ \theta - \theta\theta'\right\}\right)$$

o equivalentemente

$$\sqrt{n}\ (\overline{X} - \theta) \sim N_p(0, \text{diag}\ \theta - \theta\theta')$$

2

DISTRIBUCION DE FORMAS LINEALES

2.1. INTRODUCCIÓN.

El propósito de esta sección es obtener la función de densidad de transformaciones lineales de variables que se distribuyen según una Normal multivariada. Se definen también las condiciones necesarias para la independencia entre dos transformaciones lineales de matrices de datos Normales así como las condiciones necesarias para la independencia entre dos transformaciones lineales de una variable Normal. Por último, se analizan las condiciones para que una transformación lineal de una matriz de datos Normales resulte de nuevo una matriz de datos de la distribución Normal.

2.2. TRANSFORMACIONES LINEALES DE VARIABLES NORMALES.

La distribución de una forma lineal de variables Normales se establece en el siguiente teorema.

TEOREMA 2.2.1. *Sea* $\mathbf{X} \sim \mathbf{N}_p(\mu, \Sigma)$ *y* $\mathbf{Y}_1 = \mathbf{A}_1\mathbf{X}+\mathbf{b}$ *donde* $\mathbf{A}_1$ *es una matriz de constantes de dimensión* $q\mathrm{x}p$ *con* $q<p$ *y* $\mathbf{b} \in \mathbb{R}^q$ *un vector de constantes. La distribución de* $\mathbf{Y}_1$ *es Normal Multivariada de media* $\mathbf{A}_1\mu+\mathbf{b}$ *y matriz de covarianzas* $\mathbf{A}_1 \Sigma \mathbf{A}_1'$.

DEMOSTRACION. Defínase la matriz $\mathbf{A}_2$ de dimensión $p-q\mathrm{x}p$ de manera que

$$\mathbf{A}_{p\mathrm{x}p} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}$$

sea de rango p. Sea $\mathbf{Y}$ definido como

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \mathbf{A}\mathbf{X}$$

La función de densidad de $\mathbf{Y}$ puede obtenerse evaluando la función de densidad de $\mathbf{X}$ en la transformación inversa y multiplicando por el Jacobiano de la transformación. La densidad de $\mathbf{Y}$ está dada por

$$f_Y(y) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\{-\tfrac{1}{2}(A^{-1}y-\mu)'\Sigma^{-1}(A^{-1}y-\mu)\}|A^{-1}|$$

$$= (2\pi)^{-p/2} |A\Sigma A'|^{-1/2} \exp\{-\tfrac{1}{2}(y-A\mu)'(A\Sigma A')^{-1}(y-A\mu)\}$$

que corresponde a la densidad de una variable Normal de dimensión $\mathbf{p}$ de media $\mathbf{A}\mu$ y matriz de covarianza $\mathbf{A}\Sigma\mathbf{A}'$.

La distribución de $\mathbf{Y}_1$ corresponde a una distribución marginal del vector Y y por el teorema 1.5.1 $\mathbf{Y}_1$ tiene una distribución Normal multivariada de dimensión $\mathbf{q}$ de media $\mathbf{A}_1\mu$ y matriz de covarianzas $\mathbf{A}_1\Sigma\mathbf{A}_1'$.

COROLARIO 2.2.1. *Si* $\mathbf{X} \sim N_p(\mu,\Sigma)$ *entonces* $\mathbf{Y} = \Sigma^{-1/2}(\mathbf{X}-\mu) \sim N_p(0,I)$.

DEMOSTRACION. Del teorema 2.2.1 se obtiene que la distribución de $\mathbf{Y}$ es Normal de dimensión $\mathbf{p}$. Los parámetros de la distribución son

$$E(Y) = \Sigma^{-1/2} E(X-\mu) = 0$$

$$V(Y) = \Sigma^{-1/2}\,\Sigma\,\Sigma^{-1/2} = I.$$

La distribución del estimador máximo verosimil de la media de una población $N_p(\mu,\Sigma)$ en base a una muestra aleatoria $X_1,\ldots,X_n$ se caracteriza mediante el siguiente teorema.

TEOREMA 2.2.2. *Si $X_1,\ldots,X_n$ representa una muestra aleatoria de una población $N_p(\mu,\Sigma)$ entonces la distribución de $\bar{X}=\frac{1}{n}X'1$ es $N_p(\mu,\frac{1}{n}\Sigma)$.*

DEMOSTRACION. La función característica de $\bar{X}$ corresponde a

$$\phi_{\bar{X}}(t)= E\{\exp(it'\bar{X})\} = E\{\exp(i\frac{1}{n}\sum_{i=1}^{n} t'X_i)\}$$

Utilizando la independencia de las variables $X_1,\ldots,X_n$ y el teorema 1.2.1, el valor de $\phi_{\bar{X}}(t)$ puede escribirse como

$$\phi_{\bar{X}}(t)=\prod_{i=1}^{n}\phi_{X_i}(\frac{1}{n}t)=\prod_{i=1}^{n}\exp\{i\frac{1}{n}t'\mu-\frac{1}{2n^2}t'\Sigma t\}$$

$$=\exp\{it'\mu-\frac{1}{2}t'(\frac{1}{n}\Sigma)t\}$$

De los teoremas 1.2.1 y 1.2.2 se deduce que $\bar{X}\sim N_p(\mu,\frac{1}{n}\Sigma)$.

2.3. INDEPENDENCIA DE TRANSFORMACIONES LINEALES DE MATRICES DE DATOS NORMALES.

El siguiente teorema establece las condiciones bajo las cuales dos transformaciones lineales de una matriz de datos de una distribución Normal multivariada resultan ser independientes.

TEOREMA 2.3.1. *Si* X_{nxp} *es una matriz de datos de una población* $N_p(\mu,\Sigma)$ *y si* $Y_{qxr} = AXB$ *y* $Z_{sxt} = CXD$ *donde* $qr+st < np$, *entonces los elementos de* Y *son independientes de los elementos de* Z *si y sólo si* $B'\Sigma D=0$ *ó* $AC'=0$.

DEMOSTRACION. Sean Y^a_{qr} y Z^a_{st} vectores columna obtenidos apilando las columnas de Y y Z ; esto es

$$(Y^a)' = (Y'_{(1)},\ldots,Y'_{(r)})$$

$$(Z^a)' = (Z'_{(1)},\ldots,Z'_{(t)})$$

Del teorema A.4h se obtiene que

$$Y^a = (AXB)^a = (B' \otimes A)\; X^a$$

$$Z^a = (CXD)^a = (D' \otimes C)\; X^a \qquad (2.3.1)$$

donde $\otimes$ denota el producto Kronecker.

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

$$f_{\mathbf{X}}(X;\mu,\Sigma)=\prod_{i=1}^{n} f_{X_i}(x_i;\mu,\Sigma)= N_{np}(\mu\otimes 1,\ \Sigma\otimes I) \qquad (2.3.2)$$

Si $qr+st<np$, de (2.3.1), (2.3.2) y el teorema 2.2.1 se obtiene que

$$\mathbf{W}=\begin{bmatrix} Y^a \\ Z^a \end{bmatrix} \sim N_{qr+st}\ (\delta,\ \mathbf{V})$$

donde δ y $\mathbf{V}$ están definidos por

$$\delta= (\mu'B\otimes 1'A' \quad , \quad \mu'D\otimes 1'C')$$

$$\mathbf{V}=\begin{bmatrix} B'\Sigma B\otimes AA' & B'\Sigma D\otimes AC' \\ D'\Sigma B\otimes CA' & D'\Sigma D\otimes CC' \end{bmatrix}$$

Del teorema 1.4.1 se obtiene que Y^a y Z^a son independientes si y solo si su covarianza es cero; es decir $B'\Sigma D\otimes AC'=0$, lo cual ocurre si y solo si $B'\Sigma D=0$ ó $AC'=0$.

COROLARIO 2.3.1. *Si* $\mathbf{X}$ *es una matriz de datos de una distribución* $N_p(\mu,\Sigma)$ *y* $\bar{\mathbf{X}}$ *y* $\mathbf{S}$ *son los estimadores máximo verosímiles de* μ *y* Σ *respectivamente, entonces* $\bar{\mathbf{X}}$ *es independiente de* $\mathbf{S}$.

DEMOSTRACION. Sean $\mathbf{A}=\frac{1}{n}\mathbf{1}'$ y $\mathbf{C}=\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}'$. Por el teorema 2.3.1 $\bar{\mathbf{X}}=\mathbf{A}\,\mathbf{X}$ y $\mathbf{C}\,\mathbf{X}$ tienen distribución independiente ya que $\mathbf{A}\mathbf{C}'=0$ de donde se sigue que $\bar{\mathbf{X}}$ es independiente de $\mathbf{S}=\frac{1}{n}(\mathbf{C}\,\mathbf{X})'(\mathbf{C}\mathbf{X})$.

2.4. INDEPENDENCIA DE TRANSFORMACIONES LINEALES DE VARIABLES NORMALES.

El siguiente teorema caracteriza la independencia entre dos transformaciones lineales de una distribución Normal multivariada.

TEOREMA 2.4.1. *Sea* $X \sim N_p(\mu,\Sigma)$ *y sean* $Y_{r\times 1} = Ax+b$ *y* $Z_{s\times 1} = Cx+d$ *donde* A, C, b *y* d *son respectivamente matrices y vectores de constantes tales que* $r+s \leq p$. *Los vectores* Y *y* Z *son independientes si y sólo si* $A\Sigma C' = 0$.

DEMOSTRACION. Como $r+s \leq p$, del teorema 2.2.1 se obtiene que

$$W = \begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} X + \begin{bmatrix} b \\ d \end{bmatrix} \sim N_{r+s}(\delta, V)$$

donde

$$\delta = \begin{bmatrix} A \\ C \end{bmatrix} \mu + \begin{bmatrix} b \\ d \end{bmatrix}$$

$$V = \begin{bmatrix} A\Sigma A' & A\Sigma C' \\ C\Sigma A' & C\Sigma C' \end{bmatrix}$$

Del teorema 1.4.1 se obtiene que Y es independiente de X si y sólo si $A\Sigma C' = 0$.

2.5. TRANSFORMACIONES LINEALES DE MATRICES DE DATOS NORMALES.

Las condiciones bajo las cuales una transformación lineal de una matriz de datos $\mathbf{X}$ de una distribución Normal multivariada resulta de nuevo una matriz de datos de una distribución Normal se dan en el siguiente teorema.

TEOREMA 2.5.1. *Sea* $\mathbf{X}_{n\times p}$ *una matriz de datos de una distribución* $N_p(\mu,\Sigma)$ *y sea* $Y_{m\times q} = \mathbf{A}\mathbf{X}\mathbf{B}$ *donde* $\mathbf{A}$ *y* $\mathbf{B}$ *son matrices de constantes tales que* $mq < np$. $Y' = (Y_1, \ldots, Y_m)$ *es una matriz de datos de una distribución* $N_q(\delta B'\mu, \eta\, B'\Sigma B)$ *si y sólo si*

a) *(Media Común)* $\mathbf{A}\mathbf{1} = \delta\,\mathbf{1}$ *para algún* $\delta \in \mathbb{R}$ *ó* $B'\mu = 0$

b) *(Independencia)* $\mathbf{A}\mathbf{A}' = \eta\,\mathbf{I}$ *para algún* $\eta \in \mathbb{R}$

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

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

Del teorema A.4h se obtiene que

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

donde $\otimes$ denota el producto Kronecker.

De (2.3.2) se sigue que la distribución de X^a es $N_{np}(\mu \otimes 1, \Sigma \otimes I)$.

Si $mq < np$ del teorema 2.2.1 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 solo 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 solo si

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

donde N es cualquier matriz de $q\times q$. Esta condición equivale a $AA' = \eta I$ para algún número real η.

3

DISTRIBUCION DE FORMAS CUADRATICAS

3.1 INTRODUCCIÓN

En esta sección se determina la distribución de diferentes formas cuadráticas de variables con distribución Normal multivariada. Entre estas distribuciones asociadas a la Normal se encuentran la χ^2, Wishart, T^2 de Hotelling y Λ de Wilks. Para cada una de ellas se analizan diferentes propiedades cuya aplicabilidad se encuentra básicamente en el contexto de las técnicas correspondientes al Análisis de Varianza. Cada una de estas distribuciones se discute por separado en las siguientes secciones.

3.2 LA DISTRIBUCIÓN χ^2.

En este apartado se caracteriza a la distribución de algunas formas cuadráticas de variables Normales multivariadas estlabeciéndose las condiciones para la independencia entre dos formas cuadráticas y entre una forma cuadrática y una lineal.

3.2.1 DEFINICIÓN DE LA DENSIDAD χ^2 .

La siguiente definición proporciona una caracterización de la distribución χ^2.

DEFINICION 3.2.1: *Se dice que una variable aleatoria X tiene distribución χ^2 con n grados de libertad si su función de densidad está dada por* .

$$f_X(x;n)= \{2^{n/2}\Gamma(n/2)\}^{-1}\ x^{n/2\ -1}\ \exp\left(-\frac{x}{2}\right)\ ;\ x>0\ ,n>0$$

donde

$$\Gamma(y) = \int_0^{\infty} u^{y-1} \exp(-u)\, du \quad ; \quad y > 0$$

por conveniencia se denota por $X \sim \chi^2_n$.

TEOREMA 3.2.1. *Si X es una variable aleatoria con distribución* χ^2_n *entonces*

a) $\phi_X(t) = (1-2\,i\,t)^{-n/2}$

b) $M_X(t) = (1-2\,t)^{-n/2}$

DEMOSTRACION.

a) Desarrollando en serie de potencias la función $\exp(itX)$ *se obtiene*

$$\phi_X(t) = E\{\exp(i\,t\,X)\} =$$

$$= \sum_{r=0}^{\infty} \frac{(i\,t)^r}{2^{n/2}\,\Gamma(\frac{n}{2})\, r\,!} \int_0^{\infty} x^{r+\frac{n}{2}-1} \exp(-\frac{x}{2})\, dx$$

$$= \sum_{r=0}^{\infty} \frac{(2\,i\,t)^r\, \Gamma(\frac{n}{2}+r)}{\Gamma(\frac{n}{2})\, r\,!}$$

$$= \sum_{r=0}^{\infty} \frac{(2\,i\,t)^r\, (\frac{n}{2}+r-1)\,(\frac{n}{2}+r-2)\ldots(\frac{n}{2})}{r\,!}$$

$$= \sum_{r=0}^{\infty} (-2\,i\,t)^r\, (-1)^r\, \frac{(-\frac{n}{2})\,(-\frac{n}{2}-1)\ldots(-\frac{n}{2}-r+1)}{r\,!}$$

donde

$$\Gamma(y) = \int_0^\infty u^{y-1}\ \exp(-u)\,du \quad ; \quad y>0$$

por conveniencia se denota por $X \sim \chi^2_n$.

TEOREMA 3.2.1. *Si X es una variable aleatoria con distribución* χ^2_n *entonces*

a) $\phi_X(t) = (1-2\,i\,t)^{-n/2}$

b) $M_X(t) = (1-2\,t)^{-n/2}$

DEMOSTRACION.

a) Desarrollando en serie de potencias la función $\exp(itX)$ *se obtiene*

$$\phi_X(t) = E\{\exp(i\,t\,X)\} =$$

$$= \sum_{r=0}^{\infty} \frac{(i\,t)^r}{2^{n/2}\,\Gamma(\frac{n}{2})\,r\,!} \int_0^\infty x^{r+\frac{n}{2}-1} \exp(-\frac{x}{2})\,dx$$

$$= \sum_{r=0}^{\infty} \frac{(2\,i\,t)^r\,\Gamma(\frac{n}{2}+r)}{\Gamma(\frac{n}{2})\,r\,!}$$

$$= \sum_{r=0}^{\infty} \frac{(2\,i\,t)^r\,(\frac{n}{2}+r-1)\,(\frac{n}{2}+r-2)\ldots.(\frac{n}{2})}{r\,!}$$

$$= \sum_{r=0}^{\infty} (-2\,i\,t)^r\,(-1)^r\,\frac{(-\frac{n}{2})\,(-\frac{n}{2}-1)\ldots.(-\frac{n}{2}-r+1)}{r\,!}$$

$$= \sum_{r=0}^{\infty} \binom{-n/2}{r} (-2\,i\,t)^r$$

$$= (1-2\,i\,t)^{-n/2}$$

por ser la serie binomial para $(1-2\,i\,t)^{-n/2}$.

b) $M_X(t) = E\{\exp(t\,X)\}$

$$= (1-2t)^{-n/2} \int_0^{\infty} \frac{(1-2t)^{n/2}}{2^{n/2}\,\Gamma(\frac{n}{2})}\, x^{n/2-1} \exp\{-\tfrac{x}{2}(1-2t)\}\, dx$$

$$= (1-2\,t)^{-n/2} \quad \text{si } t < 1/2$$

Observando que el integrando resulta una función de densidad Gamma con parámetros n/2 y (1-2t)/2.

3.2.2. FUNCIONES CARACTERÍSTICA Y GENERATRIZ DE MOMENTOS DE FORMAS CUADRÁTICAS DE VARIABLES NORMALES.

TEOREMA 3.2.2. *Si* $X \sim N_p(\mu,\Sigma)$ *y* $A_{p\times p}$ *es una matriz simétrica entonces*

a) $\phi_{X'AX}(t) = |I-2\,i\,t\,A\,\Sigma|^{-1/2} \exp\{i\,t\,\mu'(I-2\,i\,t\,A\,\Sigma)^{-1}A\,\mu\}$

b) $M_{X'AX}(t) = |I-2\,t\,A\,\Sigma|^{-1/2} \exp\{t\,\mu'(I-2\,t\,A\,\Sigma)^{-1}A\,\mu\}$

en una vecindad del cero.

DEMOSTRACION.

a) Como $\Sigma > 0$ y A es simétrica, por el teorema A.19 existe una matriz F de rango completo tal que $F'\Sigma^{-1}F = I$ y $F'AF = D$ donde $D = \text{diag}(d)$ y $d' = (d_{11},\ldots,d_{pp})$. Definiendo $Y = F^{-1}X$ se obtiene del teorema 2.2.1 que $Y \sim N_p(F^{-1}\mu, I)$. Utilizando estas relaciones podemos escribir

$$\begin{aligned}
\phi_{X'AX}(t) &= E\{\exp(it\, X'AX)\} \\
&= E\{\exp(it\, Y'DY)\} \\
&= E\{\exp(it \sum_{j=1}^{p} d_{jj} Y_j^2)\} \\
&= \prod_{j=1}^{p} E\{\exp(it\, d_{jj} Y_j^2)\}
\end{aligned}$$

Definiendo $\theta = F^{-1}\mu$ se obtiene

$$\begin{aligned}
E\{\exp(it\, d_{jj}Y_j^2)\} &= \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\{it\, d_{jj} Y_j^2 - \tfrac{1}{2}(Y_j - \theta_j)^2\}\, dY_j \\
&= \exp\{-\tfrac{1}{2}\theta_j^2 + \tfrac{1}{2}\theta_j^2 (1-2itd_{jj})^{-1}\} \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\{-\tfrac{1}{2}(1-2itd_{jj})[Y_j - \theta_j(1-2itd_{jj})^{-1}]^2\} d \\
&= \exp\{-\tfrac{1}{2}\theta_j^2 + \tfrac{1}{2}\theta_j^2 (1-2it\, d_{jj})^{-1}\}(1-2it\, d_{jj})^{-1/2}
\end{aligned}$$

De esta manera se sigue que

$$\begin{aligned}
\phi_{X'AX}(t) &= [\prod_{j=1}^{p} (1-2itd_{jj})^{-1/2}] \exp\{-\tfrac{1}{2}\sum_{j=1}^{p} [\theta_j^2 - \theta_j^2 (1-2itd_{jj})^{-1}]\} \\
&= |I-2itD|^{-1/2} \exp\{-\tfrac{1}{2}\theta'[I-(I-2itD)^{-1}]\theta\}
\end{aligned}$$

$$=|\Sigma|^{-1/2}\,|F|\,|I-2itD|^{-1/2}\exp\{-\tfrac{1}{2}\mu'[(F^{-1})'F^{-1}-(F^{-1})'(I-2itD)^{-1}F^{-1}]\mu\}$$

$$=|\Sigma|^{-1/2}|\Sigma^{-1}-2itA|^{-1/2}\exp\{-\tfrac{1}{2}\mu'[\Sigma^{-1}-(FF'-2itFDF')^{-1}]\mu\}$$

$$=|I-2itA\Sigma|^{-1/2}\exp\{-\tfrac{1}{2}\mu'[\Sigma^{-1}-(\Sigma-2it\Sigma A\Sigma)^{-1}]\mu\}$$

$$=|I-2itA\Sigma|^{-1/2}\exp\{-\tfrac{1}{2}\mu'[I-(I-2itA\Sigma)^{-1}]\Sigma^{-1}\mu\}$$

$$=|I-2itA\Sigma|^{-1/2}\exp\{-\tfrac{1}{2}\mu'(I-2itA\Sigma)^{-1}[(I-2itA\Sigma)-I]\Sigma^{-1}\mu\}$$

y por tanto, la función característica puede escribirse como

$$\phi_{X'AX}(t)=|I-2itA\Sigma|^{-1/2}\exp\{it\mu'(I-2itA\Sigma)^{-1}A\mu\}$$

b) $M_{X'AX}(t)=E\{\exp(tX'AX)\}$

$$=|2\pi\Sigma|^{-1/2}\exp(-\tfrac{1}{2}\mu'\Sigma^{-1}\mu)\int_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}x'(\Sigma^{-1}-2tA)x+\mu'\Sigma^{-1}x\}\,dx$$

como $\Sigma>0$, del teorema A.7b se obtiene que $\Sigma^{-1}>0$ y definiendo $M=\Sigma^{-1}-2tA$ se obtiene del teorema A.8 que $M>0$ si tomamos $|t|$ lo suficientemente pequeño. Utilizando este hecho y haciendo $\delta=M^{-1}\Sigma^{-1}\mu$, puede escribirse

$$M_{X'AX}(t)=|M\Sigma|^{-1/2}\exp\{-\tfrac{1}{2}(\mu'\Sigma^{-1}\mu-\delta'M\delta)\}\int_{-\infty}^{\infty}|2\pi M^{-1}|^{-1/2}\exp\{-\tfrac{1}{2}(X-\delta)'M(X-\delta)\}\,dx$$

$$=|I-2tA\Sigma|^{-1/2}\exp\{-\tfrac{1}{2}\mu'[I-(I-2tA\Sigma)^{-1}]\Sigma^{-1}\mu\}$$

$$=|I-2tA\Sigma|^{-1/2}\exp\{-\tfrac{1}{2}\mu'(I-2tA\Sigma)^{-1}[(I-2tA\Sigma)-I]\Sigma^{-1}\mu\}$$

$$=|I-2tA\Sigma|^{-1/2}\exp\{t\mu'(I-2tA\Sigma)^{-1}A\mu\}$$

3.2.3. DISTRIBUCIÓN E INDEPENDENCIA DE FORMAS CUADRÁTICAS DE VARIABLES NORMALES.

TEOREMA 3.2.3. *Si* $X \sim N_p(\mu, \Sigma)$ *entonces*

a) $X'AX \sim \chi^2_r$ *si y sólo si* $A\mu=0$ *y* $A\Sigma A=A$
en cuyo caso $r=t_r(A\Sigma)$

b) $X'AX$ *y* $X'BX$ *son independientes si y sólo si* $A\Sigma B=0$

c) $X'AX$ *y* CX *son independientes si y sólo si* $C\Sigma A=0$.

DEMOSTRACION. Las condiciones $A\mu=0$ $A\Sigma A=A$ son suficientes para que $X'AX \sim \chi^2_r$. Para demostrarlo, del teorema 3.2.2a se tiene que la función característica de $X'AX$ está dada por

$$\phi_{X'AX}(t)=|I-2itA\Sigma|^{-1/2}\exp\{it\mu'(I-2itA\Sigma)^{-1}A\mu\}$$

Si $A\mu=0$ se obtiene que

$$\phi_{X'AX}(t)= |I-2itA\Sigma|^{-1/2}=\prod_{j=1}^{p}(1-2it\lambda_j)^{-1/2} \qquad (3.2.1)$$

donde $\lambda_1,\ldots,\lambda_p$ son los valores propios de $A\Sigma$. Si Además $A\Sigma A=A$ se sigue que $A\Sigma A\Sigma=A\Sigma$ de donde se deduce que $A\Sigma$ es idempotente y por tanto exactamente r valores λ_j son iguales a 1 y el resto son iguales a cero con $r=r(A\Sigma)=tr(A\Sigma)$. De (3.2.1) se obtiene que

$$\phi_{X'AX}(t) = (1-2it)^{-r/2}$$

Del teorema 3.2.1a se sigue que la función característica de $X'AX$ es igual a la correspondiente a una variable con distribución χ^2_r y del teorema 1.2.2 se concluye que $X'AX \sim \chi^2_r$.

De manera inversa si $X'AX$ tiene distribución χ^2_s utilizando los teoremas 1.2.2 y 3.2.2a se sigue que

$$\prod_{j=1}^{p}(1-2it\lambda_j)^{-1/2}\exp\{it\mu'W\mu\} = (1-2it)^{-s/2} \qquad (3.2.2)$$

donde $W=(I-2itA\Sigma)^{-1}A$. Derivando ambos miembros de la ecuación (3.2.2) respecto a μ se obtiene

$$\prod_{j=1}^{p}(1-2it\lambda_j)^{-1/2}\exp\{it\mu'W\mu\}(2itW\mu)=0 \quad \forall t\in\mathbb{R}$$

de donde se sigue que $W\mu=0$ lo cual implica la condición $A\mu=0$. Substituyendo este resultado en (3.2.2) se obtiene

$$\prod_{j=1}^{p}(1-2it\lambda_j) = (1-2it)^{S}$$

Los dos polinomios en t son iguales para $\|t\|<\varepsilon$ si y solo si los grados son iguales, es decir $\lambda_{S+1}=\lambda_{S+2}=\ldots=\lambda_p=0$ de donde se sigue que

$$\prod_{j=1}^{S}(1-2it\lambda_j) = (1-2it)^{S} \qquad (3.2.3)$$

Como las raíces de los polinomios deben ser las mismas y el miembro del lado derecho de (3.2.3) tiene una raíz $t=(2i)^{-1}$ de multiplicidad S se obtiene que $\lambda_j=1$ $j=1,\ldots,s$. Del teorema A.9 se obtiene que $A\Sigma$

es idempotente de rango $S=tr(A\Sigma)$, lo cual implica la condición $A\Sigma A = A$.

b) La condición $A\Sigma B=0$ implica independencia entre $X'AX$ y $X'BX$. Para demostrarlo definamos r=rango (A) y S= rango(B). Por el teorema A.16 podemos escribir $A=M\Lambda M'$ y $B= N\theta N'$ donde M y N son matrices de rango completo con r y s columnas. Si $A\Sigma B=0$ se obtiene $M\Lambda M'\Sigma N\theta N'=0$ de donde se sigue que $M'M\Lambda M'\Sigma N\theta N'N$ y como $M'M$, $N'N$, Λ y θ son matrices de rango completo se deduce que $COV(M'X,N'X)=M'\Sigma N=0$. La distribución conjunta de $M'X$ y $N'X$ cuando $r+S<p$ puede deducirse utilizando el teorema 2.2.1, obteniéndose que es Normal multivariada y dado que $C(M'X,N'X)=0$, del teorema 1.4.1 se concluye que $M'X$ y $N'X$ tienen distrbución independiente. De este resultado y observando que $X'AX=(M'X)'\Lambda M'X$ y $X'BX=(N'X)'\theta N'X$ se deduce la independencia entre las dos formas cuadráticas.

De manera inversa si $X'AX$ y $X'BX$ tienen distribución independiente, se deduce que

$$M_{X'(A+B)X}(t)= M_{X'AX}(t)\ M_{X'BX}(t)$$

y esta condición debe verificarse para todo vector μ, en particular para $\mu=0$, obteniéndose del teorema 3.2.2b la condición

$$|I-2t(A+B)\Sigma|=|I-2tA\Sigma|\,|I-2tB\Sigma|$$

y esto ocurre si y sólo si

$$|I-2t(A+B)\Sigma| = |I-2t(A+B)\Sigma + 4t^2 A\Sigma B\Sigma| \qquad (3.2.4)$$

Sea $M=I-2t(A+B)\Sigma$. Por el teorema A.8 la matriz $M\Sigma^{-1}$ es definida positiva y por tanto de rango completo para $|t|$ lo suficientemente pequeño, de donde se sigue por el teorema A.6e que $r(M)=p$. Multiplicando por M^{-1} la ecuación (3.2.4) se obtiene

$$\prod_{i=1}^{p}(1+4t^2\delta_i) = 1 \qquad (3.2.5)$$

donde $\delta_1,\ldots,\delta_p$ son los valores propios de la matriz $A\Sigma B\Sigma M^{-1}$. La igualdad (3.2.5) se verifica para todo valor $|t|<\epsilon$ si y solo si $\delta_i=0$, $i=1,\ldots,p$. Esta condición implica que $A\Sigma B\Sigma M^{-1}=0$ de donde se deduce que $A\Sigma B=0$ puesto que Σ y M^{-1} son matrices de rango completo.

c) Para demostrar que la condición $C\Sigma A=0$ es suficiente para implicar la independencia entre $X'AX$ y CX definamos r=rango(A) y S como el número de renglones de C. Por el teorema A.16 podemos escribir $A=M\Lambda M'$ donde M_{pxr} y Λ_{rxr} son matrices de rango r. La condición $C\Sigma A=0$ implica $C\Sigma M\Lambda M'=0$ de donde se sigue que $C\Sigma M\Lambda M'M=0$ y como $M'M$ es de rango completo se deduce que $C\Sigma M=COV(CX,M'X)=0$. Por el teorema 2.2.1, la distribución conjunta de $M'X$ y CX cuando $r+S<p$ es Normal multivariada y como $COV(CX,M'X)=0$ se sigue del teorema 1.4.1 la independencia entre CX y $M'X$. Observando que $X'AX=(M'X)'\Lambda M'X$ se sigue la independencia entre $X'AX$ y CX.

De manera inversa, si $X'AX$ y CX tienen distribución independiente, se verifica que $COV(CX, X'AX)=0$. La covarianza en términos de A, Σ y C puede calcularse observando que

$$E(X'AX) = E(tr\, AXX') = tr\, A\, E(XX') = tr A\Sigma + \mu' A \mu$$

De esta ecuación se obtiene que la covarianza puede escribirse como

$$COV(CX,\, X'AX) = E\{[CX - C\mu][X'AX - tr\, A\Sigma - \mu' A\mu]\}$$

$$= CE\{[X-\mu][(X-\mu)'A(X-\mu) + 2(X-\mu)'A\mu - tr\, A\Sigma]\}$$

$$= CE\{(X-\mu)(X-\mu)'A(X-\mu)\} + 2C\Sigma A\mu - 0$$

$$= 2\, C\Sigma A\mu \qquad (3.2.6)$$

ya que el tercer momento de $X-\mu$ es cero. Esto puede verificarse definiendo $Y = X-\mu$ y observando que el K-ésimo elemento de $W=(Y'AY)Y$ está dado por

$$W_K = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} Y_i Y_j Y_K \quad ; \quad K=1,\ldots,p$$

Para demostrar que $E(W)=0$ es suficiente demostrar que $E(Y_i Y_j Y_k)=0 \;\forall i,j,k$. Para esto, de (1.2.4) y el teorema 1.2.1 se obtiene que

$$E(Y_i Y_j Y_K) = -\frac{1}{i}\left\{\frac{\partial^3}{\partial t_i\, \partial t_j\, \partial t_K} \exp\left(-\frac{1}{2} t'\Sigma t\right)\right\}$$

Definiendo $\Sigma = (\sigma_{ij})$, se obtiene que las derivadas de $\phi_y(t)$ corresponden a

$$\frac{\partial}{\partial t_K} \phi_Y(t) = -\left(\sum_{\ell=1}^{p} \sigma_{K\ell} t_\ell\right) \exp\left(-\frac{1}{2} t'\Sigma t\right)$$

$$\frac{\partial^2}{\partial t_j \partial t_K} \phi_Y(t) = \left(\sum_{\ell=1}^{p} \sigma_{K\ell} t_\ell\right) \exp\left(-\frac{1}{2} t' \Sigma t\right) \left(\sum_{m=1}^{p} \sigma_{jm} t_m\right)$$

$$- \sigma_{Kj} \exp\left(-\frac{1}{2} t' \Sigma t\right)$$

$$\frac{\partial^3}{\partial t_i \partial t_j \partial t_k} \phi_Y(t) = \left(\sum_{\ell=1}^{p} \sum_{m=1}^{p} \sigma_{K\ell} \sigma_{jm} t_\ell t_m\right) \exp\left(-\frac{1}{2} t'\Sigma t\right) \left(-\sum_{n=1}^{p} \sigma_{in} t_n\right)$$

$$+ \exp\left(-\frac{1}{2} t'\Sigma t\right) \left(\sum_{\ell=1}^{p} \sigma_{K\ell} t_\ell \sigma_{ji} + \sum_{m=1}^{p} \sigma_{jm} t_m \sigma_{Ki}\right)$$

$$+ \sigma_{Kj} \exp\left(-\frac{1}{2} t' \Sigma t\right) \left(\sum_{n=1}^{p} \sigma_{in} t_n\right) \qquad (3.2.7)$$

Valuando en t=0, todos los miembros de la ecuación (3.2.7) son iguales a cero, de donde se sigue que $E(W)=0$. De (3.2.6) se sigue que $C\Sigma A\mu = 0$ y como esta condición se verifica para todo vector μ, se concluye que $C\Sigma A=0$.

COROLARIO.3.2.3. *Sea* $X \sim N_p(\mu,\Sigma)$. *La variable* $Z=(X-\mu)'\Sigma^{-1}(X-\mu)$ *tiene distribución* χ^2_p.

DEMOSTRACION. Del teorema 2.2.1 se tiene que $Y=X-\mu \sim N(0,\Sigma)$. Haciendo $A=\Sigma^{-1}$ y utilizando el teorema 3.2.3a se deduce que $Y'AY = (X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_p$.

3.2.4. OTRAS PROPIEDADES DE LA DISTRIBUCIÓN χ^2.

La siguiente definición y teorema caracterizan dos importantes propiedades de la distribución χ^2.

DEFINICION 3.2.2. *Se dice que una variable aleatoria* X *tiene distribución* F *con parámetros* n *y* m *si su función de densidad está dada por*

$$f_X(x;n,m)=\frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \qquad \frac{(\frac{n}{m})^{n/2}\, x^{n/2-1}}{(1+\frac{n}{m}\, x)^{(n+m)/2}} \quad x>0;n,m=1,2,\ldots$$

para simplificar se denota por $X\sim F_{n,m}$.

TEOREMA 3.2.4. *Si* $X_1,\ldots,X_p$ *son variables aleatorias independientes tales que* $X_j\sim \chi^2_{n_j}$ $j=1,\ldots,p$ *entonces*

a) $Y=\sum_{j=1}^{p} X_j \sim \chi^2_n$; $n=\sum_{j=1}^{p} n_j$

b) $Z= n_2 X_1/n_1 X_2 \sim F_{(n_1,n_2)}$

DEMOSTRACION.

a) La función característica de $\sum_{i=1}^{p} X_i$ puede calcularse como

$$\phi_Y(t)=\prod_{j=1}^{p} \phi_{X_j}(t)$$

$$=\prod_{j=1}^{p}(1-2it)^{-n_j/2}$$

$$=(1-2it)^{-n/2}$$

Del teorema 3.2.1a se deduce que $\phi_Y(t)$ es la función característica de una variable con distribución χ^2_n y por el teorema 1.2.2 se concluye que $Y\sim \chi^2_n$.

b) Sean $Z = n_2 X_1 / n_1 X_2$ y $W = X_2$. La función de densidad conjunta de Z y W está dada por

$$f_{Z\,W}(z, w) = g_{X_1}\left(\frac{n_1}{n_2}\, z\, w\right) h_{X_2}(w)\, \frac{n_1}{n_2}\, w$$

$$= \left\{2^{(n_1+n_2)/2}\, \Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)\right\}^{-1} \left(\frac{n_1}{n_2}\, z\, w\right)^{n_1/2-1} w^{\frac{n_2}{2}-1} \exp\left\{-\frac{1}{2}\left(\frac{n_1}{n_2} z w + w\right)\right\}$$

$$\cdot\ \frac{n_1}{n_2}\, w$$

La densidad de z puede obtenerse como

$$f_Z(z) = \int_0^\infty f_{Z\,W}(z, w)\, dw$$

$$= \int_0^\infty \frac{\left(\frac{n_1}{n_2}\right)^{n_1/2}\, z^{n_1/2-1}}{2^{(n_1+n_2)/2}\, \Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)}\, w^{(n_1+n_2)/2-1} \exp\left\{-\frac{w}{2}\left(1 + \frac{n_1}{n_2} z\right)\right\} dw$$

Efectuando el cambio de variable $x = \frac{w}{2}\left(1 + \frac{n_1}{n_2} z\right)$ se obtiene

$$f_Z(z) = \frac{\left(\frac{n_1}{n_2}\right)^{n_1/2}}{\Gamma\left(\frac{n_2}{2}\right) \Gamma\left(\frac{n_1}{2}\right)}\ \frac{z^{n_1/2-1}}{\left(1 + \frac{n_1}{n_2} z\right)^{(n_1+n_2)/2}} \int_0^\infty x^{(n_1+n_2)/2-1} e^{-x}\, dx$$

$$= \frac{\Gamma\left(\frac{n_1+n_2}{2}\right) \left(\frac{n_1}{n_2}\right)^{n_1/2} z^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right) \left(1 + \frac{n_1}{n_2} z\right)^{(n_1+n_2)/2}} \quad ; \quad z > 0$$

y de la definición 3.2.2 se sigue que la variable Z tiene distribución F con parámetros n_1 y n_2 .

3.3. LA DISTRIBUCIÓN WISHART.

3.3.1. INTRODUCCIÓN.

En esta sección se definen y examinan algunas propiedades de la distribución Wishart, que corresponde a una generalización de la distribución χ^2. Esta distribución juega un papel de particular importancia en el Análisis Multivariado ya que estadísticas como $(n-1)S$ tienen esta distribución. La distribución Wishart se presenta mediante la siguiente definición.

3.3.2. DEFINICIÓN DE LA DENSIDAD WISHART.

DEFINICION 3.3.1. *Se dice que la matriz* $M_{p\times p}$ *tiene distribución Wishart de dimensión* p *con matriz de escala* Σ *y* n *grados de libertad si puede ser escrita como* $M=X'X$ *donde* $X_{n\times p}$ *es una matriz de datos de una distribución* $N_p(0,\Sigma)$. *Para simplificar se emplea la notación* $M\sim W_p(\Sigma,n)$.

Las matrices con distribución Wishart poseen una importante propiedad que se establece con el siguiente teorema.

TEOREMA 3.3.1. *Sea* $M\sim W_p(\Sigma,n)$. *Si* $n\geq p$, *la matriz* M *es positiva definida con probabilidad* 1.

DEMOSTRACION. La matriz M siempre es semidefinida positiva al tener una distribución Wishart ya que podemos escribir $a'Ma=a'X'Xa=Y'Y\geq 0$ donde $Y=Xa$. Para demostrar que M es definida positiva basta demostrar que $a'Ma=0$ ocurre solo cuando $a=0$. Como esta condición se verifica solo cuando M es una matriz de rango completo se sigue que cuando $n\geq p$

$$P(a'Ma>0 \quad \forall a\neq 0) = P(\text{rango } M=p)$$

$$= P(\text{rango } X=p)$$

$$= 1-P(\text{rango } X < p)$$

$= 1-P$(Al menos una columna de X es dependiente de las otras).

$$= 1-P\left(\bigcup_{j=1}^{p}\{\sum_{i=1}^{p} w_i X_{(i)} = 0 \text{ con } w_j = 1\}\right)$$

$$\geq 1-\sum_{j=1}^{p} P\left(\sum_{i=1}^{p} w_i X_{(i)} = 0 \text{ con } w_j = 1\right)$$

$$= 1$$

por tener $Y = \sum_{i=1}^{p} W_i X_{(i)}$ distribución Normal Multivariada de dimensión n.

3.3.3. FUNCIONES CARACTERÍSTICA Y GENERATRIZ DE MOMENTOS DE LA DENSIDAD WISHART.

Al igual que en el caso de variables aleatorias vectoriales la distribución de matrices aleatorias puede determinarse mediante el uso de funciones características debido a su correspondencia uno a uno con las funciones de distribución o densidad acumulativa. La distribución de matrices aleatorias también puede ser caracterizada bajo ciertas condiciones en base a los momentos de la distribución. Las definiciones 3.3.2 y 3.3.3 se refieren respectivamente a la función característica y función generatriz de momentos para el caso de matrices aleatorias.

DEFINICION 3.3.2. *Sea* X_{pxq} *una matriz aleatoria y* θ *una matriz de dimensión* pxq. *La función característica de* X *denotada por* $\phi_X(\theta)$ *se define como*

$$\phi_X(\theta) = E[\exp\{i\,tr(\theta' X)\}]$$

Si p=q *y* X *es simétrica,* θ *se puede restringir al dominio de las matrices simétricas.*

Los momentos conjuntos de la distribución pueden obtenerse mediante la relación

$$E\begin{bmatrix} X_{11}^{k_{11}} & X_{12}^{k_{12}} & \dots X_{qp}^{k_{qp}} \end{bmatrix} = \frac{1}{i^{k_{11}+k_{12}+\dots+k_{qp}}} \left\{ \frac{\partial^{k_{11}+k_{12}+\dots k_{qp}}}{\partial\theta_{11}^{k_{11}}\dots\partial\theta_{qp}^{k_{qp}}}\, \phi_X(\theta) \right\} \Bigg|_{\theta=0} \quad (3.3.1)$$

Esta igualdad puede demostrarse derivando los miembros de la ecuación

$$\phi_X(\theta) = \int_{-\infty}^{\infty} \exp\{i\, \mathrm{tr}\,(\theta' X)\} f_X(X)\, dX$$

DEFINICION 3.3.3. *Sea X_{pxp} una matriz aleatoria y θ una matriz de dimensión qxp. La función generatriz de momentos de X denotada por $M_X(\theta)$ se define como*

$$M_X(\theta) = E[\exp\{\mathrm{tr}(\theta' X)\}]$$

Si p=q y X es simétrica θ se puede restringir al dominio de las matrices simétricas.

Los momentos conjuntos de la distribución pueden obtenerse de la siguiente relación

$$E\begin{bmatrix} X_{11}^{k_{11}} & X_{12}^{k_{12}} & \dots & X_{qp}^{K_{qp}} \end{bmatrix} = \left\{ \frac{\partial^{k_{11}+k_{12}+\dots+k_{qp}}}{\partial\theta_{11}^{k_{11}}\dots\partial\theta_{qp}^{k_{qp}}}\, M_X(\theta) \right\} \Bigg|_{\theta=0} \quad (3.3.2)$$

Puede demostrarse la igualdad anterior obteniendo las derivadas en ambos miembros de la siguiente ecuación

$$M_X(\theta)= \int_{-\infty}^{\infty} \exp\{tr(\theta' X)\} f_X(X)\, dX$$

Comparando las definiciones 3.3.2 y 3.3.3 con las definiciones 1.2.2 y 1.2.3 se observa que comparten la misma idea: A cada variable se asocia un parámetro de tal forma que al tomar las derivadas ya sea de la función característica o función generatriz de momentos respecto a los parámetros apropiados y evaluando posteriormente todos los parámetros en cero se obtienen los momentos conjuntos de la distribución como lo indican las relaciones (3.3.1) y (3.3.2). Puede observarse de esta manera que las definiciones 1.2.2 y 1.2.3 constituyen casos particulares de las definiciones 3.3.2 y 3.3.3 respectivamente. En otras palabras, el hecho de trabajar con vectores ó matrices es solo una forma de escribir el conjunto de variables aleatorias por lo que el teorema de unicidad (Teorema 1.2.2) y el corolario 1.2.3 siguen siendo aplicables a matrices aleatorias estableciéndose de esta manera los dos siguientes teoremas.

TEOREMA 3.3.2. (*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_{pxq} una matriz aleatoria y sea $N'=(M'_{(1)}, M'_{(2)},\ldots, M'_{(q)})$; es decir N es un vector de dimensión pq obtenido apilando las q columnas de la matriz M. La función característica de M puede expresarse como

$$\begin{aligned}\phi_M(\theta) &= E\,[\exp\{i\, tr(\theta' M)\}] \\ &= E\,[\exp\{i\, tr(\delta' N)\}] \\ &= \phi_N(\delta)\end{aligned}$$

donde δ es un vector de dimensión pq obtenido apilando las columnas de θ; esto es $\delta' = (\theta'_{(1)}, \ldots, \theta'_{(q)})$. Del teorema 1.2.2 se obtiene la unicidad de la función de densidad correspondiente a $\phi_N(\delta)$ y por tanto se sigue la unicidad de la densidad asociada a $\phi_M(\theta)$.

TEOREMA 3.3.3. *Si U y V son matrices aleatorias tales que tienen la misma función generatriz de momentos; es decir* $M_u(\theta) = M_v(\theta)$ *para toda matriz* θ *que satisface* $|\theta_{ij}| < \varepsilon_{ij}; i,j,=1,2,\ldots,p$ *entonces U y V tienen la misma distribución.*

DEMOSTRACION. Sean **X** y **Y** vectores columna obtenidos apilando las columnas de las matrices U y V respectivamente, es decir, $X'_{pq} = (U'_{(1)}, \ldots, U'_{(q)})$ y $Y'_{rs} = (V'_{(1)}, \ldots, V'_{(s)})$. La función generatriz de momentos de **U** puede expresarse como

$$\begin{aligned} M_u(\theta) &= E\,[\exp\{tr(\theta' U)\}] \\ &= E\,[\exp\{tr(\alpha' X)\}] \\ &= M_X(\delta) \end{aligned}$$

donde $\delta' = (\theta'_{(1)}, \ldots, \theta'_{(q)})$. Analogamente se demuestra que $M_v(\theta) = M_y(\delta)$ por lo que la condición $M_u(\theta) = M_v(\theta)$ para toda matriz θ que satisface $|\theta_{ij}| < \varepsilon_{ij}$ implica que $M_X(\delta) = M_Y(\delta)$ para todo vector δ en un rectángulo abierto que incluye el origen. Del corolario 1.2.3 se obtiene que los vectores **X** y **Y** tienen la misma distribución de donde se deduce el resultado.

Para efectos de este trabajo, las definiciones de función característica y función generatriz de momentos utilizadas corresponden a las definiciones 3.3.2 y 3.3.3 substituyendo θ por $t\theta$. Este cambio se hace por dos razones. La primera es simplificar el álgebra en la obtención de las funciones generatrices de momentos y la segunda reside en el hecho de que resulta conveniente que la función generatriz de momentos exista para toda matriz θ y para valores de **t** en una vecindad del cero. El cambio de θ por $t\theta$ en la función característica se hace por analogía en la definición de esta función respecto a la utilizada para la función generatriz de momentos. De cualquier forma, valuando en $t=1$ las funciones característica y generatriz de momentos obtenidas en este trabajo, se tienen las correspondientes a las indicadas en las definiciones 3.3.2 y 3.3.3 salvo que la función generatriz de momentos no existe de esta manera para toda matriz θ. Para efectos del cálculo de los momentos conjuntos a partir de la función generatriz de momentos basta obtener las derivadas correspondientes respecto a los parámetros respectivos de la matriz θ, valuando después en $\theta=0$ y posteriormente en $t=1$.

TEOREMA 3.3.4. *Si* $M \sim W_p(\Sigma, n)$ *y* $\theta_{p\times p}$ *es simétrica entonces*

$$a)\quad \phi_M(t\theta) = |\Sigma^{-1}|^{n/2} \big/ |\Sigma^{-1} - 2it\theta|^{n/2}$$

$$b)\quad M_M(t\theta) = |\Sigma^{-1}|^{n/2} \big/ |\Sigma^{-1} - 2t\theta|^{n/2}.$$

DEMOSTRACION.

a) De la definición 3.3.1 podemos escribir $M = \sum_{j=1}^{n} Y_j Y_j'$ donde $Y_1, \ldots, Y_n$ son idependientes con distribución $N_p(0, \Sigma)$. Utilizando esta igualdad se obtiene

$$\phi_M(t\theta) = E\{\exp(it\,\mathrm{tr}\sum_{j=1}^{n} Y_j Y_j' \theta)\}$$

$$= \prod_{j=1}^{n} E\{\exp(it Y_j' \theta Y_j)\}$$

$$= E\{\exp(it Y' \theta Y)\}^n$$

donde $Y \sim N_p(0,\Sigma)$. Utilizando el teorema 3.2.2a se obtiene

$$\phi_M(t\theta) = |I - 2i\, t\, \theta\, \Sigma|^{-n/2}$$

de donde se sigue el resultado.

b) De la definición 3.3.1 podemos escribir $M = \sum_{j=1}^{n} Y_j Y_j'$ donde $Y_1, \ldots, Y_n$ tienen distribución independiente $N_p(0,\Sigma)$. La función generatriz de momentos de M puede calcularse como

$$M_M(t\theta) = E\{\exp(t\, tr \sum_{j=1}^{n} Y_j Y_j' \theta)\}$$

$$E\{\exp(t\, Y'\theta Y)\}^n$$

donde $Y \sim N_p(0,\Sigma)$. Del teorema 3.2.2 b se obtiene

$$M_M(t\theta) = |I - 2\, t\, \theta\, \Sigma|^{-n/2}$$

3.3.4. LA FUNCIÓN DE DENSIDAD WISHART.

El siguiente teorema establece la función de densidad de una distribución Wishart.

TEOREMA 3.3.5. *Si $M \sim W_p(\Sigma, n)$, su función de densidad está dada por*

$$f_M(M) = \frac{|M|^{\frac{1}{2}(n-p-1)} \exp\{-\frac{1}{2}\, tr(\Sigma^{-1} M)\}}{2^{np/2}\, \pi^{p(p-1)/4}\, |\Sigma|^{n/2} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(n+1-i)]} \qquad (3.3.3)$$

para M definida positiva y cero en otro caso. La matriz de parámetros Σ es definida positiva.

DEMOSTRACION. Como **M** y Σ son definidas positivas se obtiene que $|M| > 0$ y $|\Sigma| > 0$ de donde se sigue que $f_M(M) > 0$. Para demostrar que f es función de densidad resta para demostrar que integra 1. Para esto calculemos la integral de la función f sobre el espacio de matrices **M** definidas positivas y simétricas. Definiendo el cambio de variable $W = \frac{1}{2}\Sigma^{-1/2} M \Sigma^{-1/2}$, del teorema A.23 se obtiene que el jacobiano de la transformación es $2^{p(p+1)/2}|\Sigma|^{(p+1)/2}$ de donde se sigue que

$$\int_{M>0} f(M)\, dM$$

$$= \frac{1}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}\int_{W>0}|W|^{\frac{1}{2}(n-p-1)}\exp(-\mathrm{tr}\, W)\, dW$$

Dado que W es simétrica y definida positiva puede escribirse utilizando la descomposición de Cholesky (teorema A.18) como $W = T'T$, donde T es una matriz triangular superior y $t_{ii} > 0$ $i=1,\dots,p$. Del teorema A.24 se obtiene que el jacobiano de la transformación no lineal es $2^p\prod_{i=1}^{p} t_{ii}^{p+1-i}$ por lo que se deduce que

$$\int_{M>0} f(M)\, dM$$

$$= \frac{2^p}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}\int_{T>0}\left[\prod_{i=1}^{p} t_{ii}^{p+1-i}\right]|T'T|^{\frac{1}{2}(n-p-1)}\exp\{-\mathrm{tr}\, T'T\}\, dT$$

$$= \frac{2^p}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}\int_{T>0}\left(\prod_{i=1}^{p} t_{ii}^{n-i}\right)\exp\left(-\sum_{i=1}^{p}\sum_{j=1}^{i} t_{ji}^2\right) dT$$

$$= \frac{2^p}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}\int_{T>0}\left(\prod_{i=1}^{p} t_{ii}^{n-i}\right)\exp\left(-\sum_{i=1}^{p} t_{ii}^2\right) dT\int_{T>0}\exp\left(-\sum_{i=2}^{p}\sum_{j=1}^{i-1} t_{ji}^2\right) dT$$

$$= \frac{2^p}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}\prod_{i=1}^{p}\int_0^\infty t_{ii}^{n-i}\exp(-t_{ii}^2)\,dt_{ii}\left[\prod_{i=1}^{p(p-1)/2}\int_{-\infty}^{\infty}e^{-t^2}\,dt\right]$$

$$= \frac{1}{\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}\prod_{i=1}^{p}\int_0^\infty t_{ii}^{\frac{1}{2}(n-1-i)}\exp(-t_{ii})\,dt_{ii}\left[\prod_{i=1}^{p(p-1)/2}\sqrt{\pi}\right]$$

$$= 1$$

La función generatriz asociada a la densidad f definida en (3.3.3) es

$$M_M(t\theta) = E\{\exp(t\,\mathrm{tr}\,\theta M\}$$

$$= \frac{1}{|\Sigma|^{n/2}}\int \frac{|M|^{\frac{1}{2}(n-p-1)}\exp\{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}-2t\theta)M\}}{2^{np/2}\,\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}\,dM \qquad (3.3.4)$$

donde $dM = dm_{11}dm_{12}\ldots dm_{1p}dm_{22}\ldots dm_{pp}$. Como $\Sigma>0$ del teorema A.7b se obtiene $\Sigma^{-1}>0$ y como θ es simétrica del teorema A.8 se obtiene que la matriz $A=\Sigma^{-1}-2t\theta$ es definida positiva en una vecindad de t y por tanto de rango completo. Multiplicando y dividiendo la ecuación (3.3.4) por $|A|^{-n/2}$ obtenemos

$$M_M(t\theta) = |\Sigma^{-1}-2t\theta|^{-n/2}\,/\,|\Sigma|^{n/2}$$

$$= |\Sigma^{-1}|^{n/2}\,/\,|\Sigma^{-1}-2t\theta|^{n/2}$$

por ser el integrando una función de densidad Wishart con matriz de escala $(\Sigma^{-1}-2t\theta)^{-1}$ y n grados de libertad. Como esta función generatriz de momentos es igual a la obtenida en el teorema 3.3.4b para la distribución $W_p(\Sigma,n)$ se deduce por el teorema 3.3.3. que la función definida en (3.3.1) es la función de densidad M .

La derivación de la función de densidad Wishart puede consultarse en Anderson (1958,pág.157).

3.3.5.TRANSFORMACIONES LINEALES DE VARIABLES WISHART.

TEOREMA 3.3.6. *Si* $M\sim W_p(\Sigma,n)$ *y* B_{pxq} *una matriz de entradas reales entonces* $B'M\ B\sim W_q(B'\Sigma B,n)$.

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

$$B'M\ B= B'X'X\,B=Y'Y$$

donde $Y'=B'X'=(B'X_1,\ldots,B'X_n)$ de donde por ser $X_1,\ldots,X_n$ independientes se obtiene que $B'X_1,\ldots,B'X_n$ son independientes. Por el teorema 2.2.1 $B'X_i\sim N_q(0,B'\Sigma B)$ $i=1,\ldots,n$. Utilizando la definición 3.3.1 se concluye que

$$B'M\,B\sim W_q(B'\Sigma B,n).$$

COROLARIO 3.3.6. *Submatrices de la diagonal de una matriz particionada con distribución* $W_p(\Sigma,n)$ *tienen la misma distribución con matriz de escala correspondiente y los mismos grados de libertad.*

DEMOSTRACION. Sin perder generalidad supóngase que la matriz subdiagonal corresponde a la submatriz de M de las primeras $q\times q$ entradas. Sean B, M y Σ definidas por

$$B_{p\times q} : \begin{bmatrix} I_{q\times q} \\ I_{p-q\times q} \end{bmatrix}, \quad M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

por el teorema 3.3.6 $M_{11} = B'MB \sim W_q(\Sigma_{11}, n)$ donde $\Sigma_{11} = B'\Sigma B$.

TEOREMA 3.3.7. *Si* $M \sim W_p(\Sigma, n)$ *y* α *es un vector de constantes entonces*

$$\frac{\alpha' M \alpha}{\alpha' \Sigma \alpha} \sim \chi^2_{(n)}$$

DEMOSTRACION. Como $M \sim W_p(\Sigma, n)$ podemos escribir $M = X'X$ donde $X' = (X_1, \ldots, X_n)$ es una matriz de datos de una distribución $N_p(0, \Sigma)$, por lo que se tiene que

$$\alpha' M \alpha = \alpha' X' X \alpha = Z'Z = \sum_{i=1}^{n} Z_i^2$$

donde las variables $Z_i = \alpha' X_i$ $i=1,\ldots,n$ son independientes $N_1(0, \sigma_\alpha^2)$ y $\sigma_\alpha^2 = \alpha'\Sigma\alpha > 0$ por ser $\Sigma > 0$. Por el corolario 3.2.3 las variables $Z_i^2/\alpha'\Sigma\alpha$, $i=1,\ldots,n$ tienen distribución independiente χ^2_1 y del teorema 3.2.4a se sigue el resultado.

COROLARIO 3.3.7. *Si* $M \sim W_p(\Sigma,n)$ *entonces* $m_{ii}/\sigma^2_{ii} \sim \chi^2_n$.

DEMOSTRACION. Definiendo $\alpha_i' = (0,0,\ldots,0,1_i,0,\ldots,0,0)$ del teorema 3.3.7 se sigue el resultado.

3.3.6. SUMA DE VARIABLES CON DISTRIBUCIÓN WISHART.

Las matrices con distribución Wishart poseen una importante propiedad aditiva que se establece mediante el siguiente teorema.

TEOREMA 3.3.8. *Si* $M_j \sim W_p(\Sigma,n_j)$ $j=1,\ldots,r$ *son independientes entonces*

$$M = \sum_{j=1}^{r} M_j \sim W_p\ (\Sigma,n)$$

donde $n = \sum_{j=1}^{r} n_j$.

DEMOSTRACION. La función característica de la matriz M puede calcularse utilizando el teorema 3.3.4a obteniéndose

$$\begin{aligned}
\phi_M(t\theta) &= E\{\exp(i\, t\, \mathrm{tr}\, \theta M)\} \\
&= \prod_{j=1}^{r} E\{\exp(i\, t\, \mathrm{tr}\, \theta M_j)\} \\
&= \prod_{j=1}^{r} \frac{|\Sigma^{-1}|^{n_j/2}}{|I-2\, i t\theta|^{n_j/2}} \\
&= \frac{|\Sigma^{-1}|^{n}}{|I-2\, i t\theta|^{n/2}}
\end{aligned}$$

Del teorema 3.3.4a se deduce que esta función característica corresponde a la de una variable con distribución $W_p(\Sigma,n)$ y del teorema 3.3.2 se sigue el resultado.

El siguiente teorema define la distribución de variables de la forma $X'CX$. Su utilidad se encuentra principalmente en el contexto de pruebas de hipótesis en distribuciones Normales Multivariadas, especialmente en relación con la teoría de distribución de algunas estadísticas de prueba.

3.3.7. CONDICIONES PARA QUE UNA FORMA CUADRÁTICA DE MATRICES DE DATOS NORMALES TENGA DISTRIBUCIÓN WISHART.

TEOREMA 3.3.9. *Si* X_{nxp} *es una matriz de datos de una distribución* $N_p(\mu,\Sigma)$ *y* A_{nxn} *es una matriz simétrica, entonces* $X'AX \sim W_p(\Sigma,r)$ *donde* r=rango(A)=tr(A) *si y sólo si las siguientes condiciones se satisfacen*

a) A es idempotente

b) $\mu = 0$ ó $A1=0$

DEMOSTRACION. Consideremos la descomposición espectral de la matriz A definida por

$$A= U \Lambda U'= \sum_{j=1}^{n} \lambda_j U_j U_j'$$

donde $U=(U_1,\ldots,U_n)$. Definiendo $Y_j= X'U_j$ $j=1,\ldots,n$, la matriz Y corresponde en términos de la matriz X a

$$Y'=(Y_1,\ldots,Y_n)=(X'U_1,\ldots,X'U_n)= X'U$$

de donde se obtiene que $Y= U'X$. Por el teorema 2.5.1b, los vectores $Y_1,\ldots,Y_n$ tienen distribución independiente ya que $U'U=I$. La media y matriz de covarianzas de Y_j corresponden a

$$E(Y_j)=M_j=\sum_{k=1}^{n} u_{jk}\,\mu = (U_j'\ 1)\,\mu \qquad ; \quad j=1,\ldots,n$$

$$V(Y_j)=\sum_{k=1}^{n} u_{jk}^2\,\Sigma = \Sigma \qquad ; \quad j=1,\ldots,n$$

Utilizando la descomposición espectral de la matriz $\mathbf{A}$, la matriz $\mathbf{X'AX}$ puede escribirse como

$$\mathbf{X'AX}=\sum_{j=1}^{n} \lambda_j\, \mathbf{X}'U_jU_j'\,\mathbf{X} = \sum_{j=1}^{n} \lambda_j Y_j Y_j'$$

de donde la función característica de $\mathbf{X'AX}$ puede calcularse como

$$\begin{aligned}
\phi_{X'AX}(t\theta) &= E\{\exp(i t \operatorname{tr} X'AX\theta)\} \\
&= E\{\exp(i t \sum_{j=1}^{n} \lambda_j Y_j' \theta Y_j)\} \\
&= \prod_{j=1}^{n} E\{\exp(i t \lambda_j Y_j' \theta Y_j)\} \\
&= \prod_{j=1}^{n} \phi_{Y_j'\theta Y_j}(t\lambda_j)
\end{aligned}$$

Utilizando el teorema 3.2.2a se obtiene que

$$\phi_{X'AX}(t\theta)=\prod_{j=1}^{n}|I-2it\lambda_j\theta\Sigma|^{-1/2}\exp\{itM_j'(I-2it\lambda_j\theta\,\Sigma)^{-1}\theta\,M_j\} \qquad (3.3.5)$$

Para demostrar la suficiencia de las condiciones a) y b) basta observar que si $\mathbf{A}$ es idempotente, exactamente r valores propios λ_j de

A son iguales a 1 y los n-r restantes son iguales a cero donde **r=rango(A)= tr(A).** De esta manera, la ecuación (3.3.5) puede escribirse como

$$\phi_{X'AX}(t\,\theta)= |I-2it\,\theta\,\Sigma|^{-r/2}\exp\{it\sum_{j=1}^{r}M_j'(I-2it\,\theta\,\Sigma)^{-1}\theta\,M_j\}$$

La forma cuadrática del exponente puede expresarse como

$$it\sum_{j=1}^{r}M_j'(I-2it\theta\,\Sigma)^{-1}\theta M_j=it\,\mu'\{\sum_{j=1}^{r}(I-2it\,\theta\,\Sigma)^{-1}\theta\,(1'U_j)(U_j^t 1)\}\,\mu$$

$$=it\mu'(I-2it\,\theta\,\Sigma)^{-1}\,\theta\,\mu\,(1'A1)$$

donde si $\mu=0$ ó $A1=0$, la función característica de **X'AX** es

$$\phi_{X'AX}(t\,\theta)= |I-2it\,\theta\,\Sigma|^{-r/2}$$

Del teorema 3.3.4a se tiene que esta función característica corresponde a la de una variable con distribución $W_p(\Sigma,r)$ y del teorema 3.3.2 se sigue por tanto que $X'AX\sim W_p(\Sigma,r)$.

De manera inversa, si $X'AX\sim W_p(\Sigma,r)$, la función característica de **X'AX** obtenida en (3.3.5) debe ser tal que la siguiente condición se satisface para toda matriz simétrica θ

$$\left[\prod_{j=1}^{n}|I-2\,it\,\lambda_j\,\theta\,\Sigma|^{-1/2}\right]\exp\{i\,t\sum_{j=1}^{n}M_j'(I-2it\lambda_j\theta\Sigma)^{-1}\theta\,M_j\}$$

$$= |I-2\,i\,t\,\theta\,\Sigma|^{-r/2} \tag{3.3.6}$$

Derivando ambos miembros de (3.3.5) respecto a M_j se obtiene la condición

$$(I-2it\,\lambda_j\,\theta\,\Sigma)^{-1}\theta\,M_j=0\ \forall\,\theta \text{ matriz simétrica };j=1,\ldots,n \tag{3.3.7}$$

Premultiplicando cada una de las ecuaciones definidas en (3.3.7) por $i\,t\,M_j'$ y efectuando la suma sobre el índice j se obtiene

$$i\,t\sum_{j=1}^{n} M_j'\ (I-2\,i\,t\lambda_j\,\theta\,\Sigma)^{-1}\ \theta\ M_j=0$$

por lo que la ecuación (3.3.6) puede escribirse como

$$\phi_{X'AX}(t\,\theta)=\prod_{j=1}^{n}|I-2\,i\,t\,\lambda_j\theta\,\Sigma|^{-1/2}=|I-2\,i\,t\,\theta\,\Sigma|^{-r/2}$$

Como los dos polinomios en t son iguales para valores $\|t\|<\varepsilon$, se implica que los grados son iguales, lo cual ocurre si y sólo si $\lambda_{r+1}=\lambda_{r+2}=,\ldots,=\lambda_n=0$, de donde se sigue que

$$\prod_{j=1}^{r}|I-2\,i\,t\,\lambda_j\,\theta\,\Sigma|^{-1/2}=|I-2\,i\,t\,\theta\,\Sigma|^{-r/2}$$

Para que los dos polinomios sean iguales, los coeficientes deben también ser iguales, lo que implica que $\lambda_1=\lambda_2=\ldots=\lambda_r=1$ y por tanto se sigue del teorema A.9 que **A** es idempotente de rango **r**, lo cual impli ca la condición a) del teorema. Para implicar la condición b) del teorema, de las ecuaciones definidas en (3.3.7), haciendo $\theta = I$ y premultiplicando cada ecuación por $(1-2\,i t \lambda_j \theta\, \Sigma)$ se obtienen las ecuaciones

$$M_j = 0 \quad ; \quad j=1,\ldots,n \tag{3.3.8}$$

Premultiplicando por $\lambda_j M_j'$ cada una de las ecuaciones definidas en (3.3.8) y sumando sobre el índice j se tiene que

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

$$\Rightarrow \quad \sum_{j=1}^{n} \mu' 1' (\lambda_j U_j U_j') 1 \mu = 0$$

$$\Rightarrow \quad \mu' 1' A 1 \mu = 0$$

$$\Rightarrow \quad (\mu'\mu)\, 1' A 1 = 0$$

de donde se sigue que

$$\mu'\mu = 0 \quad \text{ó} \quad 1' A 1 = 0$$

Como **A** es simétrica e idempotente se deduce por tanto que

$$\mu'\mu = 0 \quad \text{ó} \quad 1' A' A 1 = 0$$

de donde se concluye que

$$\mu = 0 \quad \text{ó} \quad A 1 = 0$$

lo cual demuestra el teorema.

COROLARIO 3.3.9.1. *La matriz* $nS = X'HX$ *donde* $H = I - \frac{1}{n}\, 1 1'$ *tiene distribución* $W_p(\Sigma, n-1)$.

DEMOSTRACION. Como H es idempotente, su rango puede calcularse como $r(H) = tr(H) = n-1$. La matriz H satisface además la condición $H\,1 = 0$ y del teorema 3.3.9 se sigue el resultado.

COROLARIO 3.3.9.2. *La matriz* $S = \frac{1}{n} X'HX$ *donde* $H = I - \frac{1}{n}\; 1'1$ *es definida positiva con probabilidad* 1 *si* $n > p$.

DEMOSTRACION. Del corolario 3.3.9.1 se obtiene que $nS \sim W_p(\Sigma, n-1)$ y del teorema 3.3.1 se sigue que nS es definida positiva cuando $n > p+1$, lo cual ocurre si y sólo si S es definida positiva cuando $n > p+1$.

3.3.8. INDEPENDENCIA EN FORMAS CUADRÁTICAS Y LINEALES EN MATRICES DE DATOS NORMALES.

Las condiciones para la independencia entre dos formas cuadráticas multivariadas y entre una forma cuadrática multivariada y una forma lineal multivariada se establecen mediante el siguiente teorema.

TEOREMA 3.3.10. *Sea* X_{nxp} *una matriz de datos de una distribución* $N_p(\mu, \Sigma)$ *y sean* $A_{pxp}, B_{pxp}, C_{rxn}, D_{pxs}$ *matrices tales que* A *y* B *son simétricas donde* $t = r(A)$ *y* $u = r(B)$ *y* $t + u < n$. *Las siguientes propiedades se satisfacen*

a) $X'AX$ *y* $X'BX$ *son independientes si y sólo si* $AB = 0$

b) $X'AX$ *y* CXD *son independientes si y sólo si* $CA = 0$

DEMOSTRACION.

a) Primero demostraremos la suficiencia de la condición $AB=0$. Para ello, por el teorema A.16 podemos escribir $A= U\Lambda U'$ y $B=V\theta V'$ donde U_{nxt}, Λ_{txt} son matrices de rango t y V_{nxu} y θ_{uxu} tienen rango u. Utilizando estas descomposiciones podemos escribir

$$X'AX = X'U\Lambda U'X = Y'\Lambda Y$$

$$X'BX = X'V\theta V'X = Z'\theta Z$$

donde $Y= U'X$ y $Z= V'X$.

Si $AB=0$ se sigue que $U\Lambda U'V\theta V'=0$ y esta condición implica $U'U\Lambda U'V\theta V'V=0$. Como $U'U$ y $V'V$ son matrices de rango completo lo mismo que θ y Λ, se sigue que $U'V=0$, de donde por el teorema 2.3.1, Y y Z son independientes y por tanto $X'AX=Y'\Lambda Y$ y $X'BX = Z'\theta Z$ son también independientes.

Para la implicación en sentido inverso, si $X'AX$ y $X'BX$ tienen distribución independiente, entonces para todo vector $\alpha \in \mathbb{R}^p$ distinto de cero $\alpha'X'AX\alpha = Y'AY$ y $\alpha'X'BX\alpha = Y'BY$ donde $Y = X\alpha$ son independientes. Como $Y'=(Y_1,\ldots,Y_n)=(X_1'\alpha,\ldots,X_n'\alpha)$ se tiene que $Y\sim N_n((\alpha'\mu)1, (\alpha'\Sigma\alpha)I)$ y del teorema 3.2.3b se sigue que $(\alpha'\Sigma\alpha)AB=0$ y dado que $\alpha'\Sigma\alpha>0$ para todo vector $\alpha\neq 0$ se concluye que $AB=0$.

b) Para demostrar que $CA=0$ es condición suficiente para la independencia entre $X'AX$ y CXD recurriendo al teorema A.16 podemos escribir a la matriz A como $A= U\Lambda U'$ donde U_{nxt} y Λ_{txt} tienen rango t. Si $CA=0$ se obtiene que $CU\Lambda U'=0$ y por tanto $CU\Lambda U'U = 0$. Como $U'U$ y Λ son de rango t se sigue que $CU=0$.

De esta condición y del teorema 2.3.1 se obtiene que $Y=U'X$ y CXD son independientes y por tanto $X'AX=Y'\Lambda Y$ es independiente de CXD .

La necesidad de la condición $CA=0$ para la independencia entre $X'AX$ y CXD puede demostrarse de la manera siguiente. Del teorema 2.5.1 se sigue que $Y_{nxs}= XD$ es una matriz de datos de una distribución $N_s(D'\mu, D'\Sigma D)$. Por tanto, para todo vector $\alpha\in\mathbb{R}^s$ no nulo, se obtiene que

$$Z=XD\alpha=Y\alpha = \begin{bmatrix} \alpha'Y_1 \\ \vdots \\ \alpha'Y_n \end{bmatrix} \sim N_n[(\alpha'D'\mu)1, (\alpha'D'\Sigma D\alpha) I]$$

Si $X'AX$ y CXD son independientes entonces $D'X'AXD$ y CXD también lo son y por tanto para todo vector α no nulo, $Z'AZ=\alpha'D'X'AXD\alpha$ es independiente de $CZ=CXD\alpha$. Del teorema 3.2.3c se obtiene la condición $(\alpha'D'\Sigma D\alpha)CA=0$ y dado que Σ es definida positiva $\alpha'D'\Sigma D\alpha > 0$ de donde finalmente $CA=0$.

3.3.9. DISTRIBUCIÓN E INDEPENDENCIA DE ALGUNAS FUNCIONES DE MATRICES CON DISTRIBUCIÓN WISHART.

En procedimientos de selección de variables en Análisis Discriminante utilizando técnicas de Análisis de Varianza, resulta involucrada la distribución de ciertas funciones de submatrices de una matriz con distribución Wishart, ya que resultan en estadísticas de prueba de igualdad de medias en distribuciones condicionales. El siguiente teorema resulta de utilidad en este sentido.

TEOREMA 3.3.11. *Sea* $M\sim W_p(\Sigma,n)$ *y sean*

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

donde M_{ij} *y* Σ_{ij} *tienen dimensión* $p_i \times p_j$ *tales que* $p_1+p_2=p$, μ_i *es de dimensión* p_i *y definamos* $M_{22.1}$ *y* $\Sigma_{22.1}$ *como*

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

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

Las siguientes propiedades se satisfacen

a) $M_{22.1} \sim W_{p_2}(\Sigma_{22.1}, n-p_1)$ *y es independiente de* (M_{11}, M_{12})

b) *Si* $\Sigma_{12}=0$ *entonces* $M_{22}-M_{22.1}= M_{21} M_{11} M_{12} \sim W_{p_2}(\Sigma_{22}, P_1)$ *y* $M_{21}M_{11}^{-1}M_{12}$, M_{11} *y* $M_{22.1}$ *son independientes.*

DEMOSTRACION.

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

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \qquad (3.3.9)$$

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

$$\begin{aligned} M_{22.1} &= X_2' X_2 - X_2' X_1 (X_1' X_1)^{-1} X_1' X_2 \\ &= X_2' [I - X_1 (X_1' X_1)^{-1} X_1'] X_2 \\ &= X_2' C X_2 \end{aligned} \tag{3.3.10}$$

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

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

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

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 2.3.1 se deduce que X_1 y $X_{2.1}$ son independientes. Como X es una matriz de datos de una distribución $N_p(0, \Sigma)$, X_1 es una matriz de datos de una distribución $N_{p_1}(\mu_1, \Sigma_{11})$ y utilizando el teorema 2.5.1 se puede comprobar que $X_{2.1}$ es una matriz de datos de una distribución $N_{p_2}(0, \Sigma_{22.1})$. Como $X_{2.1}$ es independiente de X_1 se obtiene que condicionando a X_1 fija, $X_{2.1}$

es una matriz de datos de la distribución $N_{p_2}(0, \Sigma_{22.1})$ y además dada la matriz X_1, la matriz C es una matriz de constantes que satisface $C^2 = C$, $C = C'$ y su rango coincide con $r = \text{rango}(W)\,\text{tr}(W) = n - p_1$ por lo que utilizando el teorema 3.3.9 se deduce que dada X_1 fija,

$$M_{22.1} = X_2' \, C \, X_2 \sim W_{p_2}(\Sigma_{22.1}, \; n-p_1)$$

Como la distribución de $M_{22.1}$ condicionada a X_1 no depende de X_1, se sigue que esta distribución condicional coincide con la correspondiente distribución marginal y por tanto $M_{22.1}$ es independiente de X_1 y en consecuencia también de $M_{11} = X_1' \, X_1$. Para demostrar la independencia entre $M_{22.1}$ y M_{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 2.3.1 se obtiene que dada X_1, las matrices $C\,X_{2.1}$ y $(I-C)X_{2.1}$ tienen distribución independiente de donde se sigue que dada X_1, las matrices $M_{22.1} = (C\,X_{2.1})'(C\,X_{2.1})$ y $(I-C)X_{2.1}$ tienen distribución independiente. Como $M_{22.1}$ es independiente de X_1 , se obtiene que $M_{22.1}$ es independiente de $\{X_1, (I-C)X_{2.1}\}$. Más aún, puesto que $M_{11} = X_1' X_1$ y $M_{12} = X_1'[(I-C)X_{2.1} + X_1 \Sigma_{11}^{-1} \Sigma_{12}] = X_1' \, X_2$ se obtiene que $M_{22.1}$ es independiente de (M_{11}, M_{12}).

b) Utilizando las relaciones (3.3.9), (3.3.10) y la condición $\Sigma_{12} = 0$ se obtiene que

$$\begin{aligned} M_{22} - M_{22.1} &= X_2'(I-C)X_2 \\ &= (X_{2.1} + X_1 \Sigma_{11}^{-1} \Sigma_{12})'(I-C)\,(X_{2.1} + X_1 \Sigma_{11}^{-1} \Sigma_{12}) \\ &= X_{2.1}' \,(I-C)X_{2.1} \\ &= X_{2.1}' \,(X_1 (X_1' \, X_1)^{-1} X_1')X_{2.1} \end{aligned}$$

La matriz $X_1(X_1' X_1)^{-1}X_1'$ es simétrica, idempotente y de rango p_1 y como dada X_1, $X_{2.1}$ es una matriz de datos provenientes de una distribución $N_{p_2}(0,\Sigma_{22})$ cuando $\Sigma_{12}=0$; del teorema 3.3.9 se sigue que dada X_1 fija, $M_{22}-M_{22.1} \sim W_{p_2}(\Sigma_{22},p_1)$. Por el teorema 3.3.10 a dada X_1 fija, las matrices $M_{22.1}$ y $M_{22}-M_{22.1}$ son independientes, ya que $C(I-C)=0$ y como las distribuciones condicionales de $M_{22.1}$ y $M_{22}-M_{22.1}$ no involucran a X_1 se deduce que $M_{22.1}$ y $M_{22}-M_{22.1}$ son independientes y que la distribución de $(M_{22.1}, M_{22}-M_{22.1})$ es independiente de X_1 y por tanto de M_1, de donde las tres matrices $M_{22}-M_{22.1}$, M_{11} y $M_{22.1}$ son independientes.

3.3.10. LA DISTRIBUCIÓN DEL DETERMINANTE DE UNA VARIABLE WISHART COMO PRODUCTO DE VARIABLES χ^2.

El siguiente teorema describe la distribución de $|M|$ cuando M tiene distribución Wishart.

TEOREMA 3.3.12. *Si $M \sim W_p(\Sigma,n)$ y $n>p$ entonces $|M|$ tiene distribución igual a la de $|\Sigma|$ veces el producto de p variables independientes con distribución χ^2 con grados de libertad* $n, n-1, \ldots, n-p+1$.

DEMOSTRACION. La demostración se hará por inducción sobre p. Por el corolario 3.3.7 si $m_{11} \sim W_1(\sigma_{11}^2, n)$ se obtiene

$$m_{11} \sim \sigma_{11}^2 \chi_n^2 \qquad (3.3.11)$$

Supongamos que el teorema es válido para p-1; es decir, sea $M_{11} \sim W_{p-1}(\Sigma_{11},n)$ donde $|M_{11}|$ puede ser expresado como

$$|M_{11}| = |\Sigma_{11}| Y_1, \ldots, Y_{p-1}$$

donde las variables $Y_1, \ldots, Y_{p-1}$ son independientes y $Y_i \sim \chi^2_{n-i+1}$.

Sea $\mathbf{M} \sim W_p(\Sigma, n)$ donde

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \text{ y } \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

por el **teorema** 3.3.11a M_{11} tiene distribución independiente de $M_{22.1} = M_{22} - M_{21} M_{11}^{-1} M_{12} \sim W_1(\Sigma_{22.1}\ , n-p+1)$. Del teorema A.2e y (3.3.11) se obtiene

$$|\mathbf{M}| = |M_{11}|\ |M_{22.1}| = |\Sigma_{11}|Y_1, \ldots, Y_{p-1}|\Sigma_{22.1}|\ Y_p$$

$$= |\Sigma| Y_1, \ldots, Y_p$$

donde $Y_1, \ldots, Y_p$ son independientes y $Y_i \sim \chi^2_{n-i+1}$.

COROLARIO 3.3.12. *(Demostración alternativa del teorema 3.3.1). Si* $\mathbf{M} \sim W_p(\Sigma, n)$ *y* $n > p$ *, la matriz* $\mathbf{M}$ *es definida positiva con probabilidad 1.*

DEMOSTRACION. De la definición 3.3.1 $\mathbf{M}$ puede escribirse como $\mathbf{X'X}$ de donde se sigue que $a\mathbf{M}a = \mathbf{Z'Z} > 0$ con $\mathbf{Z} = \mathbf{X}a$. Para demostrar que $a'\mathbf{M}a = 0$ implica que $a = 0$ basta demostrar que $|\mathbf{M}| \neq 0$. Para esto obtenemos que si $n > p$ $P(|M| > 0) = \prod_{i=1}^{p} P(Y_i > 0 / Y_i \sim \chi^2_{n-i+1}) = 1$ de donde se sigue que $\mathbf{M}$ es definida positiva con probabilidad 1.

3.3.11. DISTRIBUCIÓN DEL COCIENTE DE DETERMINANTES DE VARIABLES WISHART.

TEOREMA 3.3.13. *Si* $M_1 \sim W_p(\Sigma, m)$ *y* $M_2 \sim W_p(\Sigma, n)$ *son independientes, cuando* $m > p$ *y* $n > p$ *entonces*

$$W=\frac{|M_1|}{|M_2|} = |M_2^{-1} \; M_1|$$

tiene distribución proporcional a la de $\prod_{i=1}^{p} z_i$ *donde* $z_1,\ldots,z_p$ *son independientes y tales que* $z_i \sim F_{(m-i+1,n=i+1)}$ $i=1,..,p$.

DEMOSTRACION. Por el teorema 3.3.12 podemos escribir

$$|M_1| = |\Sigma| \prod_{i=1}^{p} X_i \quad ; \; X_i \sim \chi^2_{m-i+1} \qquad i=1,\ldots,p$$
$$|M_2| = |\Sigma| \prod_{i=1}^{p} Y_i \quad ; \; Y_i \sim \chi^2_{n-i+1} \qquad i=1,\ldots,p$$

donde $X_1,\ldots,X_p$ son independientes, $Y_1,\ldots,Y_p$ son independientes y por ser M_1 independiente de M_2 se obtiene que X_i y Y_j son independientes para $i,j=1,..,p$. Por el teorema 3.2.4b la variable

$$z_i=(\frac{n-i+1}{m-i+1})\,\frac{X_i}{Y_i} \sim F_{(m-i+1,\; n-i+1)}$$

de donde se deduce que

$$W=k \prod_{i=1}^{p} z_i \; ; \; k=\prod_{i=1}^{p} \frac{m-i+1}{n-i+1}$$

3.3.12. DISTRIBUCIÓN DE LA INVERSA DE TRANSFORMACIONES LINEALES DE LA MATRIZ INVERSA DE LA DISTRIBUCIÓN WISHART.

TEOREMA 3.3.14. *Si* $M \sim W_p(\Sigma,n)$ *y* $A_{k\times p}$ *es una matriz de rango k entonces*

$$(AM^{-1}A')^{-1} \sim W_k((A\Sigma^{-1}A')^{-1}\;,\; n+k-p)$$

DEMOSTRACION. Sea $C_{pxp}=(B',A')'$ donde $B_{(p-k)xp}$ es cualquier matriz tal que **C** es de rango completo. Por el teorema 3.3.6 se obtiene que

$$W=(C')^{-1}\ M\ C^{-1} \sim W_p((C')^{-1}\ \Sigma\ C^{-1},n)$$

La matriz W^{-1} puede escribirse como

$$W^{-1}=C\,M^{-1}\ C' = \begin{bmatrix} B\ M^{-1}\ B' & B\,M^{-1}\ A' \\ A\ M^{-1}\ B' & A\,M^{-1}\ A' \end{bmatrix} = \begin{bmatrix} W^{11} & W^{12} \\ W^{21} & W^{22} \end{bmatrix}$$

Por el teorema A.3d, la matriz W^{22} puede escribirse en términos de las submatrices de **W** como

$$W^{22}= A\,M^{-1}\ A'= (W_{22}-W_{21}\,W_{11}^{-1}\ W_{12})^{-1}$$

y del teorema 3.3.11a se deduce que

$$(A\,M^{-1}\,A')^{-1}=W_{22}-W_{21}\,W_{11}^{-1}\ W_{12} \sim W_k((A\,\Sigma^{-1}\,A')^{-1},\ n+k-p\ )$$

3.4. LA DISTRIBUCIÓN T^2 DE HOTELLING.

3.4.1. INTRODUCCIÓN.

En esta sección se discute la distribución de funciones de la forma $X'M^{-1}X$ donde X tiene distribución Normal y M tiene distribución Wishart independiente de X. La siguiente definición caracteriza esta distribución.

3.4.2. DEFINICIÓN DE LA DENSIDAD T^2 DE HOTELLING.

DEFINICION 3.4.1. *Si la variable* Z *puede escribirse como* $mX'M^{-1}X$ *donde* X *y* M *tienen distribuciones independientes* $N_p(0,I)$ *y* $W_p(I,m)$ *respectivamente, se dice que* Z *tiene una distribución* T^2 *de Hotelling con parámetros* p *y* m. *Para simplificar se abrevia* $Z \sim T^2(p,m)$.

3.4.3. RELACIÓN ENTRE LA DENSIDAD T^2 DE HOTELLING Y LA DENSIDAD WISHART.

TEOREMA 3.4.1. *Si* $X \sim N_p(\mu,\Sigma)$ *y* $M \sim W_p(\Sigma,m)$ *tienen distribución independiente entonces*

$$m(X-\mu)'M^{-1}(X-\mu) \sim T^2(p,m) .$$

DEMOSTRACION. Por el corolario 2.2.1 $Y = \Sigma^{-1/2}(X-\mu) \sim N_p(0,I)$ y del teorema 3.3.6 se obtiene que $W = \Sigma^{-1/2} M \Sigma^{-1/2} \sim W_p(I,m)$ independiente de Y. De la definición 3.4.1 se sigue que

$$mY'W^{-1}Y = m(X-\mu)'M^{-1}(X-\mu) \sim T^2(p,m) .$$

COROLARIO 3.4.1: *Si* $\overline{X}$ *y* S *son el vector de medias y la matriz de covarianzas obtenidos de una muestra aleatoria de tamaño* n *de una distribución* $N_p(\mu;\Sigma)$ *y* $S_u = \frac{n}{n-1} S$ *entonces*

$$(n-1)(\bar{X}-\mu)' S^{-1}(\bar{X}-\mu) = n(\bar{X}-\mu) S_u^{-1}(\bar{X}-\mu) \sim T^2(p,n-1)$$

DEMOSTRACION. Del teorema 2.2.2 se obtiene que $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$ y del corolario 3.3.3.1 se obtiene que la distribución de nS es $W_p(\Sigma, n-1)$. La independencia entre $\bar{X}$ y S se sigue del corolario 2.3.1. Substituyendo M=nS, m=n-1 y $X-\mu$ por $n^{1/2}(\bar{X}-\mu)$ en el teorema 3.4.1 se obtiene el resultado.

3.4.4. RELACIÓN ENTRE LA DENSIDAD T^2 DE HOTELLING Y LA DENSIDAD F.

La distribución T^2 de Hotelling es proporcional a la distribución F. Su relación la establece el siguiente teorema.

TEOREMA 3.4.2. *Si* $Z \sim T^2(p,m)$, *entonces la distribución de* Z *es igual a la distribución de* $mp/(m-p+1)\,Y$ *donde* Y *tiene distribución* F *de parámetros* p *y* m-p+1.

DEMOSTRACION. Sea Z una variable con distribución T^2 de Hotelling con parámetros p y m . Por la definición 3.4.1 Z puede escribirse como

$$Z = m X' M^{-1} X \tag{3.4.1}$$

donde $X \sim N_p(0,I)$ y $M \sim W_p(I,m)$ independiente de X. Más aún, la variable Z puede escribirse como

$$Z = m\left(\frac{X'M^{-1}X}{X'X}\right) X'X \tag{3.4.2}$$

Por el teorema 3.3.14, haciendo A=X' se deduce que para un vector X fijo,

$$(X'M^{-1}X)^{-1} \sim W_1((X'X)^{-1},\ m-p+1)$$

de donde por el teorema 3.3.7 dado X fijo

$$V = \frac{(X'M^{-1}X)^{-1}}{(X'X)^{-1}} \sim \chi^2_{m-p+1}$$

Como la distribución de V no depende de X se deduce que V es independiente de X y por tanto la distribución marginal de V es precisamente χ^2_{m-p+1}. Utilizando la ecuación (3.4.2), el valor de Z puede escribirse como $Z=m\,V^{-1}X'X$. Dado que $X \sim N_p(0,I)$, por el corolario 3.2.3 se obtiene que $X'X=\sum_{i=1}^{n} X_i^2 \sim \chi^2_p$ y del teorema 3.2.4b se sigue que

$$Y=\left[\frac{m-p+1}{p}\right]\frac{X'X}{V}=\frac{m-p+1}{p}\ X'M^{-1}X \sim F_{p,m-p+1}$$

La variable Z con distribución $T^2(p,m)$ definida en (3.4.1) puede escribirse en términos de Y como $Z=\frac{mp}{m-p+1}\,Y$ de donde se obtiene que en una extensión de la notación $T^2(p,m)=\frac{mp}{m-p+1}\,F_{p,m-p+1}$.

COROLARIO 3.4.2.1. *Si $X \sim N_p(\mu,\Sigma)$ y $M \sim W_p(\Sigma,m)$ tienen distribución independiente entonces*

$$\frac{m-p+1}{p}\,(X-\mu)'M^{-1}(X-\mu) \sim F_{p,m-p+1}\,.$$

DEMOSTRACION. Por el teorema 3.4.1 se tiene que

$$m(X-\mu)'M^{-1}(X-\mu) \sim T^2(p,m)$$

y del teorema 3.4.2 se sigue el resultado.

COROLARIO 3.4.2.2. *Si $\overline{X}$ y S son el vector de medias y la matriz de covarianzas obtenidas de una muestra aleatoria de tamaño* n *de una población* $N_p(\mu,\Sigma)$ *entonces*

$$\frac{n-p}{p}(\overline{X}-\mu)S^{-1}(\overline{X}-\mu) \sim F_{p,n-p}\ .$$

DEMOSTRACION. Del corolario 3.4.1 se deduce que

$$(n-1)(\overline{X}-\mu)'S^{-1}(\overline{X}-\mu) \sim T^2(p,n-1)$$

y del teorema 3.4.2 se sigue el resultado.

3.5. LA DISTRIBUCIÓN Λ DE WILKS.

3.5.1. INTRODUCCIÓN.

La distribución Λ de Wilks representa la extensión multivariada de la distribución Beta. Su aplicación se encuentra principalmente en técnicas de Análisis de Varianza Multivariado, específicamente en el contexto de pruebas de hipótesis de igualdad de medias de distribuciones Normales Multivariadas. En esta sección se analizan algunas de las propiedades de esta distribución y sus relaciones con con las distribuciones Beta, F y χ^2.

3.5.2. DEFINICIÓN DE LA DENSIDAD LAMBDA DE WILKS.

DEFINICION 3.5.1. *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* $X\sim\Lambda(p,m,n)$.

La variable X puede escribirse como $X=\prod_{i=1}^{p}(1+\lambda_i)^{-1}$ donde $\lambda_1,\ldots,\lambda_p$ son los valores propios (no negativos) de $W_1^{-1}W_2$ de donde se sigue que X toma valores entre cero y uno. Si $m>p$, por el teorema 3.3.1 la matriz W_1 es definida positiva con probabilidad 1, por lo que es de rango completo y existe W_1^{-1} con probabilidad 1, justificando de esta manera la representación $X=|I-W_1^{-1}W_2|^{-1}$.

3.5.3. UNA PROPIEDAD DE IGUALDAD EN LA DISTRIBUCIÓN LAMBDA DE WILKS.

TEOREMA 3.5.1. *Las distribuciones* $\Lambda(p,m,n)$ *y* $\Lambda(n,m+n-p,p)$ *son iguales.*

DEMOSTRACION. Sea $Z\sim\Lambda(a,b,c)$ con $c<a$. La variable Z puede escribirse como

$$Z=|I+W_1^{-1}W_2|=\prod_{i=1}^{a}(1+\lambda_i)^{-1}$$

donde W_1 y W_2 son independientes con distribuciones $W_a(I,b)$ y $W_a(I,c)$ respectivamente y $\lambda_1,\ldots,\lambda_a$ son los valores propios de $W_1^{-1}W_2$. Como W_1^{-1} existe con probabilidad 1, el número de valores propios diferentes de cero es igual al rango de W_2 y Z puede escribirse como

$$Z=\prod_{i=1}^{c}(1+\lambda_i)^{-1} \tag{3.5.1}$$

Como $W_2 \sim W_a(I,c)$ podemos escribir $W_2 = X'X$ donde $X_{c\times a}$ tiene rango c con probabilidad 1 y por tanto existe $A=(XX')^{-1/2}X$ con probabilidad 1. La matriz A satisface las condiciones $AA'=I_c$ y $X=XA'A$ de donde se deduce que la matriz

$$W_1^{-1}W_2 = W_1^{-1}X'X = W_1^{-1}A'AX'XA'A$$

tiene los mismos valores propios que la matriz

$$B^{-1}C=(AW_1^{-1}A')(AX'XA')$$

Por el teorema 3.3.14 se obtiene que para una matriz X fija,

$$B=(AW_1^{-1}A')^{-1} \sim W_c(I_c, b+c-a)$$

Como la distribución de B no depende de X se deduce la independencia de las matrices B y X y por tanto la independencia de las matrices B y C. Para demostrar que C tiene distribución Wishart observamos que $C=AX'XA'=XX'$ y como todos los n p elementos de X tienen distribuciones independientes N(0,1) se sigue que $C \sim W_n(I,P)$. De (3.5.1) y puesto que $W_1^{-1}W_2$ tiene los mismos valores propios que

$B^{-1}C$ se sigue que $Z \sim \Lambda(b,b+c-a,a)$ y por tanto $\Lambda(a,b,c)=\Lambda(c,b+c-a,a)$ si $c<a$. Para demostrar el teorema observamos que si $n\leq p$ entonces $\Lambda(p,m,n)=\Lambda(n,m+n-p,p)$; mientras que si $n>p$ se deduce que $\Lambda(n,m+n-p,p)=\Lambda(p,m,n)$ finaliza la prueba.

3.5.4. MOMENTOS DE LA DISTRIBUCIÓN LAMBDA DE WILKS Y ALGUNAS RELACIONES CON LA DISTRIBUCIÓN BETA.

La densidad Λ de Wilks admite diversas representaciones con base en productos de densidades Beta. La siguiente definición y los teoremas 3.5.2 a 3.5.6 establecen las relaciones.

DEFINICION 3.5.2. *Si X es una variable aleatoria con distribución Beta de parámetros α y β, su función de densidad está dada por*

$$f_X(X;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}\,(1-x)^{\beta-1};\ \alpha,\beta>0,\ 0<x<1$$

donde

$$\Gamma(t)=\int_0^\infty u^{t-1}\exp(-u)\,du;\ t>0$$

es la función Gamma evaluada en t. Por conveniencia se utiliza la notación $X\sim B(\alpha,\beta)$.

TEOREMA 3.5.2. *Si $X\sim B(\alpha,\beta)$ el valor esperado de X^h está dado por*

$$E(X^h)=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+h)}{\Gamma(\alpha)\Gamma(\alpha+h+\beta)}\qquad h=1,2,\ldots \tag{3.5.2}$$

DEMOSTRACION. El valor esperado de X^h está dado por

$$E(X^h) = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha+h-1} (1-x)^{\beta-1}\, dx$$

$$= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+h)}{\Gamma(\alpha)\Gamma(\alpha+h+\beta)} \int_0^1 \frac{\Gamma(\alpha+h+\beta)}{\Gamma(\alpha+h)\Gamma(\beta)} x^{\alpha+h-1} (1-x)^{\beta-1}\, dx$$

$$= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+h)}{\Gamma(\alpha)\Gamma(\alpha+h+\beta)}$$

puesto que el integrando resulta la densidad de una variable con distribución $B(\alpha+h,\beta)$. Como X toma valores entre cero y uno, del teorema 1.2.3c se sigue que los momentos definidos en (3.5.2) determinan de manera única la distribución de X.

TEOREMA 3.5.3. Si $X \sim B(m,n)$

$$Z = \frac{m}{n}\left(\frac{1-X}{X}\right) \sim F_{(2n,2m)}$$

DEMOSTRACION. De la definición 3.5.2 y el teorema de cambio de variable se sigue que la densidad de Z está dada por

$$h_Z(z) = f_X\left[\frac{m}{nz+m}\right] \frac{nm}{(nz+m)^2}$$

$$= \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} \left[\frac{m}{nz+m}\right]^{m-1} \left[\frac{nz}{nz+m}\right]^{n-1} \frac{nm}{(nz+m)^2}$$

$$= \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} \frac{(\frac{n}{m})^n z^{n-1}}{(1+\frac{n}{m}z)^{n+m}} \quad z>0\ ;\ m,m=1,2,\ldots$$

de donde se sigue por la definición 3.2.2 el resultado deseado.

TEOREMA 3.5.4. *Sea* $Y \sim \Lambda(p,m,n)$ *y* $k(m,p)$ *definida por*

$$k^{-1}(m,p)= 2^{mp/2}\ \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma[\tfrac{1}{2}(m+1-i)]$$

El valor esperado de Y^h *está dado por*

$$E(Y^h)= \frac{k(m,p)k(m+n+2h,p)}{k(m+2h,p)k(m+n,p)}= \prod_{i=1}^{p} \frac{\Gamma[\frac{1}{2}(m+2h+1-i)]\Gamma[\frac{1}{2}(m+n+1-i)]}{\Gamma[\frac{1}{2}(m+1-i)]\Gamma[\frac{1}{2}(m+n+2h+1-i)]}$$

DEMOSTRACION. Como $Y \sim \Lambda(p,m,n)$, puede escribirse

$$Y= \frac{|W_1|}{|W_1+W_2|}$$

donde $W_1 \sim W_p(I,m)$, $W_2 \sim W_p(I,n)$ y las matrices W_1 y W_2 tienen distribución independiente. El valor esperado de Y^h puede calcularse como

$$E(Y^h)= \int\int |w_1|^h|w_1+w_2|^{-h}\ f_{w_1}(w_1)\, g_{w_2}(w_2)\, dw_1\, dw_2$$

con $f_{w_1}(w_1)$ y $g_{w_2}(w_2)$ las densidades de las variables $W_p(I,m)$ y $W_p(I,n)$ evaluadas en w_1 y w_2 respectivamente. Es inmediato observar que $|w_1|^h\ f_w(w_1)$ es proporcional a la densidad $h_{w_1}(w_1)$ de una variable $W_p(I,m+2h)$ evaluada en w_1. Más especificamente,

$$|w_1|^h f_{w_1}(w_1) = \frac{k(m,p)}{k(m+2h,p)} h_{w_1}(w_1)$$

En consecuencia,

$$E(Y^h) = \frac{k(m,p)}{k(m+2h,p)} \int\int |w_1+w_2|^{-h} h_{w_1}(w_1) g_{w_2}(w_2) dw_1 dw_2$$

Claramente el producto $h_{w_1}(w_1) g_{w_2}(w_2)$ representa la densidad conjunta de dos variables independientes con densidades $W_p(I,m+2h)$ y $W_p(I,n)$ respectivamente, evaluadas en w_1 y w_2. De esta manera si se define $W=W_1+W_2$ se obtiene que

$$E(Y^h) = \frac{k(m,p)}{k(m+2h,p)} E(|W|^{-h})$$

Del teorema 3.3.8 se sigue que $W \sim W_p(\Sigma, n+m+2h)$ y entonces

$$E(Y^h) = \frac{k(m,p)}{k(m+2h,p)} \int |W|^{-h} k_w(w) dw$$

donde $k(w)$ representa la densidad Wishart con matriz de escala I y $n+m+2h$ grados de libertad. Observando que $|w|^{-h} k_w(w)$ es proporcional a la densidad $W_p(I,m+n)$ se sigue que

$$E(Y^h) = \frac{k(m,p)}{k(m+2h,p)} \frac{k(m+n+2h,p)}{k(m+n,p)} \int \ell_w(w) dw$$

donde $\ell_w(w)$ es la densidad $W_p(I,m+n)$ por lo que se deduce que

$$E(Y^h) = \frac{k(m,p)k(m+n+2h,p)}{k(m+2h,p)k(m+n,p)} \quad , \quad h=1,2,... \tag{3.5.3}$$

Como **Y** toma valores entre cero y uno, el teorema 1.2.3c garantiza que los momentos de **Y** definidos en (3.5.3) determinan su distribución de manera única.

TEOREMA 3.5.5. *Sea* $\mathbf{Y} \sim \Lambda(p,m,n)$. *La variable* **Y** *se distribuye como* $\prod_{i=1}^{p} z_i$ *donde* $z_1,...,z_p$ *son variables independientes y* $z_i \sim B([m+1-i]/2, n/2)$ $i=1,2,...,p$.

DEMOSTRACION. Por el teorema 3.5.4. el valor esperado de Y^h puede escribirse como

$$E(Y^h) = \prod_{i=1}^{p} \frac{\Gamma[\frac{1}{2}(m+1-i)+h]\,\Gamma[\frac{1}{2}(m+n+1-i)]}{\Gamma[\frac{1}{2}(m+1-i)]\,\Gamma[\frac{1}{2}(m+n+1-i)+h]}$$

Del teorema 3.5.2 se obtiene que el valor esperado de Z_i^h está dado por

$$E(Z_i^h) = \frac{\Gamma[\frac{1}{2}(m+1-i)+h]\,\Gamma[\frac{1}{2}(m+n+1-i)]}{\Gamma[\frac{1}{2}(m+1-i)]\,\Gamma[\frac{1}{2}(m+n+1-i)+h]}$$

de donde se deduce que si $W = Z_1 Z_2 ... Z_p$

$$E(W^h) = \prod_{i=1}^{p} E(Z_i^h) = E(Y^h)$$

Como Y y W tienen los mismos momentos y son variables acotadas se deduce que tienen la misma distribución.

TEORÉMA 3.5.6. *Sea* $Y \sim \Lambda(p,m,n)$ *y* $Z_1,\ldots,Z_{a+1}$ *variables con distribución independiente donde* $Z_i \sim B(m+1-2i,n)$ $i=1,\ldots,a$ *y* $Z_{a+1} \sim B(\frac{1}{2}[m+1-p],n/2)$. *Definiendo* $U=\prod_{i=1}^{a} Z_i^2$ *y* $V= Z_{a+1}\prod_{i=1}^{a} Z_i^2$ *se obtiene que*

a) *Si* p=2a, *las variables* Y *y* U *tienen la misma distribución.*

b) *Si* p=2a+1, *las variables* Y *y* V *tienen la misma distribución.*

DEMOSTRACION.

a) El valor esperado de U^h puede calcularse utilizando el teorema 3.5.2 como

$$E(U^h)=\prod_{i=1}^{a} E(Z_i^{2h})$$

$$=\prod_{i=1}^{a} \frac{\Gamma(m+n-2i+1)\Gamma(m+2h-2i+1)}{\Gamma(m-2i+1)\Gamma(m+n+2h-2i+1)}$$

y del teorema A.20 se sigue que

$$E(U^h)=\prod_{i=1}^{a} \frac{\Gamma[\frac{1}{2}(m+n-2i)+1]\Gamma[\frac{1}{2}(m+n-2i+1)]\Gamma[\frac{1}{2}(m+2h-2i)+1]\Gamma[\frac{1}{2}(m+2h-2i+1)]}{\Gamma[\frac{1}{2}(m-2i)+1]\Gamma[\frac{1}{2}(m-2i+1)]\Gamma[\frac{1}{2}(m+n+2h-2i)+1]\Gamma[\frac{1}{2}(m+n+2h-2i+1)]}$$

$$=\left[\prod_{i=1,3\ldots}^{p} \frac{\Gamma[\frac{1}{2}(m+n+1-i)]\Gamma[\frac{1}{2}(m+1-i)+h]}{\Gamma[\frac{1}{2}(m+1-i)]\Gamma[\frac{1}{2}(m+n+1-i)+h]}\right]\prod_{i=2,4\ldots}^{p} \frac{\Gamma[\frac{1}{2}(m+n+1-i)]\Gamma[\frac{1}{2}(m+1-i)+h]}{\Gamma[\frac{1}{2}(m+1-i)]\Gamma[\frac{1}{2}(m+n+1-i)+h]}$$

de donde se deduce que

$$E(U^h)=\prod_{i=1}^{p} \frac{\Gamma[\frac{1}{2}(m+n+1-i)]\Gamma[\frac{1}{2}(m+1-i)+h]}{\Gamma[\frac{1}{2}(m+1-i)]\Gamma[\frac{1}{2}(m+n+1-i)+h]} \qquad (3.5.4)$$

del teorema 3.5.5 se obtiene que $E(Y^h) = E(U^h)$ de donde se sigue por el teorema 1.2.3c que Y y U tienen la misma distribución.

b) Si $p=2a+1$, del teorema 3.5.2 se deduce que

$$E(V^h) = E(Z_{a+1}^h) \prod_{i=1}^{a} E(Z_i^{2h})$$

$$= \frac{\Gamma[\frac{1}{2}(m+n+1-p)]\,\Gamma[\frac{1}{2}(m+1-p)+h]}{\Gamma[\frac{1}{2}(m+1-p)]\,\Gamma[\frac{1}{2}(m+n+1-p)+h]}\; E(U^h)$$

De (3.5.4) y el teorema A.20 se tiene que

$$E(V^h) = \prod_{i=1}^{p} \frac{\Gamma[\frac{1}{2}(m+n+1-i)]\,\Gamma[\frac{1}{2}(m+1-i)+h]}{\Gamma[\frac{1}{2}(m+1-i)]\,\Gamma[\frac{1}{2}(m+n+1-i)+h]}$$

de donde se concluye por el teorema 3.5.5 que Y y V tienen los mismos momentos y del teorema 1.2.3c se sigue el resultado.

3.5.5. RELACIONES ENTRE LA DISTRIBUCIÓN LAMBDA DE WILKS Y LA DISTRIBUCIÓN F.

Algunos casos particulares de la distribución Λ de Wilks pueden ser expresados en términos de la distribución F. El teorema siguiente establece la equivalencia entre las distribuciones.

TEOREMA 3.5.7. *Las distribuciones Λ de Wilks y F satisfacen las siguientes relaciones*

a) $$\frac{1-\Lambda(p,m,1)}{\Lambda(p,m,1)} \sim \frac{p}{m-p+1}\, F_{p,m-p+1} \qquad (3.5.5)$$

b) $$\frac{1-\Lambda(1,m,n)}{\Lambda(1,m,n)} \sim \frac{n}{m} F_{n,m} \qquad (3.5.6)$$

c) $$\frac{1-\sqrt{\Lambda(p,m,2)}}{\sqrt{\Lambda(p,m,2)}} \sim \frac{p}{m-p+1} F_{2p,2(m-p+1)} \qquad (3.5.7)$$

d) $$\frac{1-\sqrt{\Lambda(2,m,n)}}{\sqrt{\Lambda(2,m,n)}} \sim \frac{n}{m-1} F_{2n,2(m-1)} \qquad (3.5.8)$$

DEMOSTRACION.

a) Sea $X \sim \Lambda(p,m,1)$. Del teorema 3.5.1 se sigue que $X \sim \Lambda(1,m-p+1,p)$ y por el teorema 3.5.5 la variable X tiene distribución $B\{(m+1-p)/2,\ p/2\}$.

Del teorema 3.5.3 se deduce que

$$Y = \frac{m-p+1}{p}\left\{\frac{1-X}{X}\right\} \sim F_{p,m-p+1}.$$

b) Sea $X \sim \Lambda(1,m,n)$. Del teorema 3.5.1 se sigue que $X \sim \Lambda(n,m+n-1,1)$ y por el inciso a) se obtiene que

$$Y = \frac{m}{n}\left(\frac{1-X}{X}\right) \sim F_{n,m}$$

c) Sea $X \sim \Lambda(p,m,2)$. Del teorema 3.5.1 se obtiene que $X \sim \Lambda(2,m+2-p,p)$ y por el teorema 3.5.6a se sigue que $\sqrt{X} \sim B(m+1-p,p)$. Utilizando el teorema 3.5.3 se concluye que

$$Y = \frac{m+1-p}{p} \left(\frac{1-\sqrt{X}}{\sqrt{X}} \right) \sim F_{[2p,2(m+1-p)]}$$

de donde se sigue el resultado.

d) Del teorema 3.5.6 se deduce que $\sqrt{X} \sim B(m-1,n)$ y utilizando el teorema 3.5.3 se sigue que

$$Y = \frac{m-1}{n} \left(\frac{1-\sqrt{X}}{\sqrt{X}} \right) \sim F_{[2n,2(m-1)]}$$

lo cual demuestra el teorema.

3.5.6. APROXIMACIONES ASINTÓTICAS DE LA DISTRIBUCIÓN LAMBDA DE WILKS.

La distribución $\Lambda(p,m,n)$, en los casos en que $n > 2$ y $p > 2$, si m es grande, puede aproximarse mediante la siguiente relación (Bartlett, 1947).

$$-\left\{m - \frac{1}{2}(p-n+1)\right\} \ln \Lambda(p,m,n) \sim \chi^2_{np} \qquad (3.5.9)$$

cuando $m \to \infty$ Rao. C.R. propone una aproximación alternativa que puede consultarse en Rao(1973, pág.556). La aproximación está definida por

$$R = \left(\frac{\alpha\gamma - 2\beta}{pn} \right) \frac{1 - \Lambda^{1/\gamma}(p,m,n)}{\Lambda^{1/\gamma}(p,m,n)} \sim F_{pn,2\gamma-2\beta} \qquad (3.5.10)$$

donde

$$\alpha = m - \frac{1}{2}(p-n+1)$$

$$\beta = \frac{1}{4}(pn-2)$$

$$\gamma^2 = (p^2n^2-4)/(p^2+n^2-5)$$

4

CONTRASTES DE HIPOTESIS

4.1 PRUEBAS SOBRE UNA POBLACIÓN.

4.1.1. INTRODUCCIÓN.

En esta sección se presenta el desarrollo del contraste de algunas hipótesis relativas a los parámetros de una población Normal Multivariada. Los contrastes de hipótesis que se presentan en esta sección para una población $N_p(\mu,\Sigma)$ corresponden a los siguientes:

H_a: $C\mu = \mu_0$; Σ conocida

H_b: $C\mu = \mu_0$; Σ desconocida

H_c: $\Sigma = \Sigma_0$; μ conocida

H_d: $\Sigma = \Sigma_0$; μ desconocida

H_e: $C\mu = \mu_0$; $\Sigma = \Sigma_0$

H_f: $\Sigma_{12} = 0$; μ desconocida

El contraste de hipótesis tales como Ha y Hc es raro en la práctica, ya que son poco frecuentes las situaciones en que se conoce con absoluta certeza un parámetro de la distribución. Sin embargo, los contrastes han sido desarrollados por que revisten aspectos de interés desde un punto de vista teórico. La hipótesis alternativa elegida en todos los casos corresponde a la negación de la nula correspondiente y la técnica empleada en el contraste de las hipótesis es la del cociente de verosimilitudes generalizado.

En diversas ocasiones resulta dificil determinar la distribución asociada al cociente de verosimilitudes generalizado. En estos casos es conveniente utilizar un resultado asintótico general. Este resultado se establece en el siguiente

TEOREMA 4.1.1. *Sea* X *un vector aleatorio de dimensión* p *cuya distribución depende de un parámetro* θ *contenido en un espacio paramétral* Ω *y* $X_1,\dots,X_n$ *una muestra aleatoria de esta distribución. Sean las hipótesis* $H_0: \theta \in \Omega_0$ vs $H_1: \theta \in \Omega_1$, *donde* Ω_0 *de dimensión* r *es una subregión de* $\Omega \subset \mathbb{R}^q$ *y* $\Omega_1 = \Omega - \Omega_0$. *Si* Λ *representa el cociente de verosimilitudes generalizado asociado al contraste de las hipótesis* H_0 vs H_1, *bajo ciertas condiciones de regularidad para cada* $\theta \in \Omega_0$, $-2 \ln \Lambda$ *tiene distribución asintótica* χ^2 *con* q-r *grados de libertad cuando* n *tiende a infinito.*

DEMOSTRACION. Ver Silvey (1970, pág. 113).

4.1.2. PRUEBAS SOBRE LA MEDIA DE LA DISTRIBUCIÓN NORMAL.

El primer par de hipótesis corresponde a

$$H_a\colon\ C\mu = M_0\ ;\ \Sigma \text{ conocida}$$

vs

$$H_A\colon\ C\mu \neq M_0\ ;\ \Sigma \text{ conocida}$$

donde C_{qxp} es una matriz de constantes de rango q y M_0 es un vector de constantes. La información en este caso consiste en observaciones $X_1,\ldots,X_n$ extraídas en forma independiente de una población $N_p(\mu,\Sigma)$. Definiendo $X'=(X_1,\ldots,X_n)$ como la matriz de datos, se obtiene del teorema 2.5.1 que

$$Y'=(Y_1,\ldots,Y_n)=(CX_1,\ldots,CX_n)$$

es una matriz de datos de una distribución $N_q(M,V)$ donde $M=C\mu$ y $V=C\Sigma C'$. Las hipótesis en términos de Y equivalen a

$$H_a\colon\ M = M_0\ ;\ V \text{ conocida}$$

vs

$$H_A\colon\ M \neq M_0\ ;\ V \text{ conocida}$$

y la función de verosimilitud de M está dada por

$$L(M;Y)=(2\pi)^{-np/2}|V|^{-n/2}\exp\{-\tfrac{1}{2}\sum_{i=1}^{n}(Y_i-M)'V^{-1}(Y_i-M)\} \qquad (4.1.1)$$

Bajo H_a, el estimador máximo verosímil de M es $\hat{M}_a = M_0$. Bajo $H_a \cup H_A$, el estimador de máxima verosimilitud del vector de medias, por el teorema 1.3.1, corresponde a $\hat{M} = \bar{Y}$. El cociente de verosimilitudes generalizado resulta de esta manera

$$\Lambda_a = \frac{\sup_{Ha} L(M,V;X)}{\sup_{H_a \cup H_A} L(M,V;X)} = \exp\left\{ \frac{1}{2} \sum_{i=1}^{n} [(Y_i-\bar{Y})' V^{-1} (Y_i-\bar{Y}) - (Y_i-M_0)' V^{-1} (Y_i-M_0)] \right\}$$

$$= \exp\left\{-\frac{n}{2} [\bar{Y}' V^{-1} \bar{Y} - 2\bar{Y}' V^{-1} M_0 + M_0' V^{-1} M_0]\right\}$$

$$= \exp\left\{-\frac{1}{2} n(\bar{Y}-M_0)' V^{-1} (\bar{Y}-M_0)\right\}$$

La región de rechazo es el conjunto de valores muestrales que satisfacen $\Lambda_a < K$ ó equivalentemente

$$W = n(\bar{Y}-M_0)' V^{-1} (\bar{Y}-M_0) > \delta$$

Bajo H_0, $Y_i \sim N_q(M_0, V)$ y por el teorema 2.2.2 $\bar{Y} \sim N_q(M_0, \frac{1}{n} V)$. Utilizando el corolario 3.2.3 se obtiene que la distribución de la estadística de prueba W es χ_q^2.

El segundo par de hipótesis a contrastar es

H_b: $C\mu = M_0$; Σ desconocida

H_B: $C\mu \neq M_0$; Σ desconocida.

donde $C_{q\times p}$ es una matriz de rango q y M_0 un vector de constantes. Si X es la matriz de información de la distribución $N_p(\mu,\Sigma)$, por el teorema 2.5.1 se obtiene que $Y= XC'$ es una matriz de datos de la distribución $N_q(M,V)$ donde $M=C\mu$ y $V=C\Sigma C'$. Las hipótesis en términos de Y equivalen a

$$H_b: M= M_0 \; ; \; V \text{ desconocida}.$$

$$H_B: M \neq M_0 \; ; \; V \text{ desconocida}$$

y la función de verosimilitud de los parámetros M,V está dada por la relación (4.1.1). Los estimadores máximo verosímiles de los parámetros corresponden en este caso a

$$\hat{M}_b = M_0$$

$$\hat{V}_b = \frac{1}{n} \sum_{i=1}^{n} (Y_i - M_0)(Y_i - M_0)'$$

$\hat{V}_b$ resulta definida positiva con probabilidad 1 si $n> p+1$. Este hecho puede demostrarse definiendo $Z'= \sqrt{n}\;(Y_1-M_0,\ldots,Y_n-M_0)$ donde Z es una matriz de datos de una distribución $N_q(0,n\Sigma)$, por lo que utilizando la definición 3.3.1, bajo H_b $\hat{V}_b= Z'Z \sim W_q(nV,n)$ y del teorema 3.3.1 se sigue que $P(\hat{V}_b$ sea definida positiva$|H_b)=1$. Puede demostrarse que las estimaciones anteriores maximizan el logaritmo de la función (4.1.1) observando que el único posible valor de M bajo Ha es M_0 y por tanto

$$L(M_0,V;\; Y) \geq L(M,V;\; Y) \;\forall\; M \in H_b$$

Para demostrar que $L(M_0,\hat{V}_b;Y) > L(M_0,V;Y)\ \forall\ V$ obtenemos la diferencia de sus logaritmos

$$
\begin{aligned}
A &= \ell(M_0,\hat{V}_b;Y)-\ell(M_0,V;Y) \\
&= -\frac{n}{2}\,\ell n|2\pi\hat{V}_b|+\frac{n}{2}\,\ell n|2\pi V|-\frac{1}{2}\sum_{i=1}^{n}(Y_i-M_0)'\hat{V}_b^{-1}(Y_i-M_0)+\frac{1}{2}\sum_{i=1}^{n}(Y_i-M_0)'V^{-1}(Y_i-M_0) \\
&= -\frac{n}{2}\,\ell n|V^{-1}\hat{V}_b|-\frac{np}{2}+\frac{n}{2}\,\mathrm{tr}\,V^{-1}\hat{V}_b \\
&= \frac{np}{2}\left\{\frac{1}{p}\,\mathrm{tr}\,V^{-1}\hat{V}_b-1-\ell n\,|V^{-1}\hat{V}_b|^{1/p}\right\}
\end{aligned}
$$

Como ya se verificó, la matriz V_b es definida positiva y dado que $V>0$, se obtiene que $V^{-1}>0$ y del teorema A.7c se deduce que $V^{-1}\hat{V}_b \geq 0$. Utilizando el teorema A.15 se sigue que los valores propios positivos de $V^{-1}\hat{V}_b$ son los mismos que los correspondientes a $N=V^{-1/2}\hat{V}_b V^{-1/2}$. Puede verificarse que $N\geq 0$, por lo que los valores propios de $V^{-1}\hat{V}_b$ son no negativos, de donde utilizando el teorema A.21

$$
\begin{aligned}
A &= \frac{np}{2}\left\{\frac{1}{p}\sum_{i=1}^{p}\lambda_i-1-\ell n\Big(\prod_{i=1}^{p}\lambda_i\Big)^{1/p}\right\} \\
&\geq \frac{np}{2}\left\{\frac{1}{p}\sum_{i=1}^{p}\lambda_i-1-\ell n\Big(\frac{1}{p}\sum_{i=1}^{p}\lambda_i\Big)\right\}
\end{aligned}
$$

y de (1.3.8) se obtiene $A\geq 0$. Bajo $H_b \cup H_B$ no existen restricciones impuestas a los estimadores y por el teorema 1.3.1 éstos corresponden a

$$
\begin{aligned}
\hat{M} &= \bar{Y} \\
\hat{V} &= S_y=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})(Y_i-\bar{Y})'
\end{aligned}
$$

De esta manera, el cociente de verosimilitudes generalizado corresponde a

$$\Lambda_b = \left[\frac{|\hat{V}|}{|\hat{V}_b|}\right]^{n/2}$$

El valor de $\hat{V}_b$ puede escribirse como

$$\hat{V}_b = \frac{1}{n}\sum_{i=1}^{n}(Y_i - M_0)(Y_i - M_0)' = S_y + (\bar{Y} - M_0)(\bar{Y} - M_0)'$$

por lo que la región de rechazo corresponde al conjunto de muestras que satisfacen

$$\frac{|\hat{V}|}{|\hat{V}_b|} = \frac{|S_u|}{|S_u + \frac{n}{n-1} S_u W|} = \left|I_p + \frac{n}{n-1} W\right|^{-1} < K$$

donde $W = S_u^{-1}(\bar{Y} - M_0)(\bar{Y} - M_0)'$ y $S_u = \frac{n}{n-1} S_y$. La matriz **W** tiene rango 1 ya que

$$\begin{aligned} r[W] &= r[(\bar{Y} - M_0)(\bar{Y} - M_0)'] \\ &= r[\bar{Y} - M_0] \\ &= 1 \end{aligned}$$

Por el teorema A.13 existe una matriz ortogonal **T** tal que $T'WT = \Lambda$ donde Λ es diagonal de los valores propios de **W** y de rango 1. Utilizando esta propiedad se deduce que

$$|I_p + \frac{n}{n-1} W| = |I_p + \frac{n}{n-1} T' W T|$$

$$= |I_p + \frac{n}{n-1} \Lambda|$$

$$= \prod_{i=1}^{p} (1 + \frac{n}{n-1} \lambda_i)$$

$$= 1 + (\frac{n}{n-1}) \lambda_1$$

$$= 1 + (\frac{n}{n-1}) \sum_{i=1}^{p} \lambda_i$$

$$= 1 + \frac{n}{n-1} \operatorname{tr} W$$

$$= 1 + \frac{n}{n-1} (\bar{Y} - M_0)' S_u^{-1} (\bar{Y} - M_0)$$

La región de rechazo puede escribirse entonces como el conjunto de muestras que satisfacen

$$T = n(\bar{Y} - M_0)' S_u^{-1} (\bar{Y} - M_0) > \delta$$

donde bajo H_b, la estadística T, por el corolario 3.4.1 tiene distribución T^2 de Hotelling con parámetros q y $n-1$. Utilizando el teorema 3.4.2 se puede comprobar que

$$T \sim \frac{(n-1)q}{n-q} F_{(q,n-q)}$$

de donde

$$T' = \frac{(n-q)n}{(n-1)q} (\bar{Y} - M_0)' S_u^{-1} (\bar{Y} - M_0) \sim F_{(q,n-q)}$$

puede utilizarse como estadística de prueba para el contraste.

4.1.3. PRUEBAS SOBRE LA MATRIZ DE COVARIANZAS DE LA DISTRIBUCIÓN NORMAL.

La afirmación sobre la matriz de covarianzas cuando la media de la distribución es conocida corresponde al tercer juego de hipótesis

$$H_c: \Sigma = \Sigma_0 \;;\; \mu \text{ conocida}$$

$$H_c: \Sigma \neq \Sigma_0 \;;\; \mu \text{ conocida}$$

La verosimilitud de Σ está dada por

$$L(\Sigma;X) = (2\pi)^{-np/2}\,|\Sigma|^{-n/2} \exp\{-\tfrac{1}{2}\sum_{i=1}^{n}(X_i-\mu)'\,\Sigma^{-1}(X_i-\mu)\} \qquad (4.1.2)$$

Bajo H_c, el estimador máximo verosímil de Σ es $\hat{\Sigma}_c = \Sigma_0$ por ser el único posible valor de Σ bajo la hipótesis nula. Bajo $H_c \cup H_c$ no existe restricción en la matriz Σ y por tanto

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)'$$

Para demostrar que $L(\mu,\hat{\Sigma};X) > L(\mu,\Sigma;X)$ para cualquier matriz Σ observamos que μ es conocida y la demostración es análoga a la del caso de la hipótesis H_b. El cociente de verosimilitudes generalizado resulta de esta manera

$$\Lambda_c = |\Sigma_0^{-1}\hat{\Sigma}|^{n/2} \exp\{-\tfrac{n}{2}\,\mathrm{tr}(\Sigma_0^{-1}\hat{\Sigma}) + \tfrac{np}{2}\}$$

Por el teorema 4.1.1, la variable W_c definida como

$$W_c = -2\,\ell n\,\Lambda_c = n\,\{tr(\Sigma_0^{-1}\,\hat{\Sigma}) - \ell n|\Sigma_0^{-1}\,\hat{\Sigma}| - p\}$$

$$= np(\delta_1 - \ell n\,\delta_2 - 1)$$

donde δ_1 y δ_2 son la media aritmética y geométrica de los valores propios de $\Sigma^{-1}\,\hat{\Sigma}$ tiene distribución asintótica χ^2 con tantos grados de libertad como parámetros totalmente determinados hay en la hipótesis nula; es decir, $\frac{1}{2}\,p(p+1)$. La región de rechazo de W_c es de la forma $\{W_c > K\}$.

Cuando la media de la distribución es desconocida el juego de hipótesis H_c se traduce en

H_d: $\Sigma = \Sigma_0$; μ desconocida

H_D: $\Sigma \neq \Sigma_0$; μ desconocida.

La verosimilitud de μ y Σ está dada por

$$L(\mu,\Sigma;X) = (2\pi)^{-np/2}\,|\Sigma|^{-n/2}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)'\Sigma^{-1}(X_i-\mu)\} \qquad (4.1.3)$$

Bajo H_d los estimadores de máxima verosimilitud de los parámetros son $\hat{\mu}_d = \bar{X}$ y $\hat{\Sigma}_d = \Sigma_0$. Como Σ_0 es el único posible valor de Σ bajo la hipótesis nula, se obtiene que $L(\mu,\Sigma_0;X) \geq L(\mu,\Sigma;X)$ bajo la hipótesis nula. Para demostrar que $L(\bar{X},\Sigma_0;X) \geq L(\mu,\Sigma_0;X)$ basta observar que la desigualdad es un caso particular de (1.3.7) cuando $\Sigma = \Sigma_0$.

Bajo $H_d \cup H_D$ no existen restricciones en los parámetros y nuevamente por el teorema 1.3.1, los estimadores máximo verosímiles son

$$\hat{\mu} = \overline{X}$$

$$\hat{\Sigma} = S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})'$$

El cociente de verosimilitudes generalizado asociado a este contraste está definido como

$$\Lambda_d = |\Sigma_0^{-1} S|^{n/2} \exp\{-\frac{n}{2} \operatorname{tr} \Sigma_0^{-1} S + \frac{np}{2}\}$$

Por el teorema 4.1.1, la variable W_d definida por

$$W_d = -2\ln \Lambda_d = n\{\operatorname{tr} \Sigma_0^{-1} S - \ln |\Sigma_0^{-1} S| - p\}$$

$$= np\{\delta_1 - \ln \delta_2 - 1\}$$

donde δ_1 y δ_2 son ahora las medias aritmética y geométrica de los valores propios de $\Sigma_0^{-1} S$, tiene distribución asintótica χ^2 con $\frac{1}{2}p(p+1)$ grados de libertad. La región de rechazo de H_d es de la forma $\{W_d > K\}$.

4.1.4. PRUEBA SOBRE LA MEDIA Y MATRIZ DE COVARIANZAS DE LA DISTRIBUCIÓN NORMAL.

Este contraste se establece en el juego de hipótesis

$$H_e:\ \mu=\mu_0\ ,\ \Sigma=\Sigma_0$$

vs

$$H_E:\ \mu\neq\mu_0\ \text{ó}\ \Sigma\neq\Sigma_0$$

Si X_{nxp} es la matriz de datos de la distribución $N_p(\mu,\Sigma)$, la función de verosimilitud de μ y Σ se define como en (4.1.3). Bajo H_e los parámetros quedan determinados de manera única y por tanto los estimadores máximo verosímiles de éstos son precisamente $\hat{\mu}_e=\mu_0$ y $\hat{\Sigma}_e=\Sigma_0$. Bajo $H_e\cup H_E$ no existen restricciones en μ y Σ y por el teorema 1.3.1 los estimadores son $\hat{\mu}=\bar{X}$ y $\hat{\Sigma}=S$. El cociente de verosimilitudes generalizado queda definido por

$$\Lambda_e=|\Sigma_0^{-1}S|^{n/2}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu_0)'\Sigma_0^{-1}(X_i-\mu_0)+\frac{np}{2}\}$$

Utilizando el teorema 4.1.1 se obtiene que la variable

$$W_e=-2\ln\Lambda_e=\sum_{i=1}^{n}(X_i-\mu_0)'\Sigma_0^{-1}(X_i-\mu_0)-n\ln|\Sigma_0^{-1}S|-np$$

tiene distribución asintótica χ^2 con $\frac{1}{2}p(p+3)$ grados de libertad y la región de rechazo de H_e es de la forma $\{W_e>K\}$.

4.1.5. PRUEBA DE INDEPENDENCIA.

La independencia entre dos conjuntos de variables Normales puede ser caracterizada contrastando el siguiente juego de hipótesis

$$H_f:\ \Sigma_{12} = 0 \quad ; \ \mu \text{ desconocida}$$

$$H_F:\ \Sigma_{12} \neq 0 \quad ; \ \mu \text{ desconocida}$$

Sea el vector $X' = (X_1', X_2')$ donde X_1 y X_2 tienen dimensiones p_1 y p_2 respectivamente. Definamos las particiones correspondientes en los parámetros de la distribución como

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} , \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} , \quad \Sigma_{12} = \Sigma_{21}'$$

La ecuación (4.1.1) define la verosimilitud de los parámetros μ y Σ. Por el teorema 1.4.1, la función de verosimilitud de μ y Σ bajo H_f puede escribirse como

$$L(\mu,\Sigma;X) = \prod_{i=1}^{n} f_{X_{1_i}}(x_{1_i}, \mu_1, \Sigma_{11}) \prod_{i=1}^{n} f_{X_{2_i}}(x_{2_i}, \mu_2, \Sigma_{22})$$

$$= L_1(\mu_1, \Sigma_{11}; X)\ L_2(\mu_2, \Sigma_{22}; X)$$

Maximizar la función $L(\mu,\Sigma;X)$ es entonces equivalente a maximizar por separado L_1 y L_2. Aplicando el teorema 1.3.1 los estimadores que se obtienen son

$$\tilde{\mu}_1 = \bar{X}_1 \qquad \hat{\Sigma}_{11} = S_{11}$$

$$\tilde{\mu}_2 = \bar{X}_2 \qquad \hat{\Sigma}_{22} = S_{22}$$

donde S_{11} y S_{22} son las matrices de covarianzas muestrales obtenidas de los componentes de X_1 y X_2 respectivamente. De esta manera bajo H_f ,

$$\tilde{\mu}_f = \bar{X}$$

$$\hat{\Sigma}_F = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} \end{bmatrix}$$

Bajo $H_f \cup H_F$ el espacio parametral no incluye restricciones y por el teorema 1.3.1 los estimadores correspondientes son una vez más $\hat{\mu} = \bar{X}$ y $\hat{\Sigma} = S$. El cociente de verosimilitudes generalizado para este caso resulta

$$\Lambda_f = |\hat{\Sigma}_f^{-1} S|^{n/2} \exp\{- \tfrac{n}{2} \operatorname{tr} \hat{\Sigma}_f^{-1} S + \tfrac{np}{2} \}$$

La traza de $\hat{\Sigma}_f^{-1}$ S es p ya que

$$\hat{\Sigma}_f^{-1} S = \begin{bmatrix} S_{11}^{-1} & 0 \\ 0 & S_{22}^{-1} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} I_1 & S_{11}^{-1} S_{12} \\ S_{22}^{-1} S_{21} & I_2 \end{bmatrix}$$

De esta manera, la región de rechazo puede escribirse como el conjunto de valores muestrales tales que

$$W_f = |\hat{\Sigma}_f^{-1} S| = \frac{|S|}{|S_{11}||S_{22}|} < K$$

Utilizando el teorema A.2e, el valor de W_f puede escribirse como

$$W_f = |S_{22} - S_{21} \; S_{11}^{-1} \; S_{12}| / |S_{22}| \tag{4.1.4}$$

Por el corolario 3.3.9.1 la matriz $nS \sim W_p(\Sigma, n-1)$ y utilizando el teorema 3.3.11a se obtiene que bajo H_f

$$n\,S_{22\cdot1} = n\,S_{22} - n\,S_{21}\,S_{11}^{-1}\;S_{12} \sim W_{p_2}(\Sigma_{22}, n-1-p_1)$$

Del teorema 3.3.11b, bajo la hipótesis nula se deduce que, $n(S_{22} - S_{22\cdot1})$ es independiente de $n\,S_{22\cdot1}$ y además

$$n(S_{22} - S_{22\cdot1}) = nS_{21}\;S_{11}^{-1}S_{12} \sim W_{p_2}(\Sigma_{22}, p_1)$$

Del teorema 3.3.6 y la definición 3.5.1 se tiene que la variable W_f definida en (4.1.4), es tal que cuando $n > p+1$,

$$W_f = \frac{|S_{22} - S_{21}\,S_{11}^{-1}\;\;S_{12}|}{|S_{22}|} = \frac{|n\,\Sigma^{-1/2}\;\;S_{22\cdot1}\;\;\Sigma^{-1/2}|}{|n\;\Sigma^{-1/2}\,S_{22\cdot1}\,\Sigma^{-1/2} + n\Sigma^{-1/2}(S_{22} - S_{22\cdot1})\Sigma^{-1/2}|} \sim \Lambda(p_2, n-1-p_1, p_1)$$

En los casos en que $p_1 = 1,2$ ó $p_2 = 1,2$ la prueba puede realizarse utilizando las relaciones (3.5.5) a (3.5.8). Para otros valores de p_1 y

p_2 puede utilizarse la aproximación de Bartlett definida en la ecuación (3.5.9) que en este caso conduce a

$$-\{n - \frac{1}{2}(p+3)\} \ \ell n |I - S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}| \sim \chi^2_{p_1 p_2}$$

cuando $n \to \infty$.

4.2. PRUEBAS SOBRE VARIAS POBLACIONES.

4.2.1. INTRODUCCIÓN.

Asociados a la comparación de varias poblaciones se presentan los desarrollos correspondientes a las siguientes hipótesis

$$H_g: \mu_1 = \mu_2 = \ldots = \mu_g \quad y \quad \Sigma_1 = \Sigma_2 = \ldots = \Sigma_g$$

$$H_h: \mu_1 = \mu_2 = \ldots = \mu_g \quad \text{dado que } \Sigma_1 = \Sigma_2 = \ldots = \Sigma_g$$

$$H_i: \Sigma_1 = \Sigma_2 = \ldots = \Sigma_g$$

En estos casos, la hipótesis alternativa corresponde a la negación más general de la hipótesis nula correspondiente y el método de contraste empleado es el cociente de verosimilitudes generalizado.

En el contraste de las hipótesis H_g, H_h y H_i que comparan poblaciones, se requiere de información muestral extraída de cada una de las poblaciones a comparar. La información extraída en forma independiente de cada población se denotará por

$$X_i^* = \{X_{i1}, \ldots, X_{in_i}\} \qquad i=1,\ldots,g$$

donde $X_{ij} \sim N_p(\mu_i, \Sigma_i)$ $j=1,\ldots,n_i$; $i=1,\ldots,g$

por conveniencia se define $n = \sum_{i=1}^{g} n_i$.

4.2.2. IGUALDAD DE MEDIAS Y MATRICES DE COVARIANZAS.

El primer juego de hipótesis entre grupos es el siguiente

$$H_g:\ \mu_1=\mu_2=\ldots=\mu_g \quad y \quad \Sigma_1=\Sigma_2=\ldots=\Sigma_g$$

$$H_G:\ \mu_i\neq\mu_j \text{ ó } \Sigma_i\neq\Sigma_j \quad i\neq j \quad i,j=1,\ldots,g$$

Bajo H_g las n observaciones siguen la misma distribución $N_p(\mu,\Sigma)$ y por el teorema 1.3.1 los estimadores correspondientes son

$$\hat{\mu}=\bar{X}=\frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i} X_{ij}$$

$$\hat{\Sigma}=T=\frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})(X_{ij}-\bar{X})'$$

Bajo $H_g\cup H_G$ no existen restricciones impuestas a los parámetros y el máximo de la verosimilitud se obtiene maximizando las funciones de verosimilitud de los parámetros asociados a cada población, por lo que de nuevo utilizando el teorema 1.3.1 los estimadores de máxima verosimilitud están dados por

$$\hat{\mu}_i=\bar{X}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \qquad i=1,\ldots,g$$

$$\hat{\Sigma}_i=S_i=\frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)(X_{ij}-\bar{X}_i)' \qquad i=1,\ldots,g$$

De esta manera, el máximo de la verosimilitud bajo la hipótesis nula resulta ser

$$L(\hat{\mu},\hat{\Sigma}; X_1,\dots,X_g) = (2\pi)^{-np/2}\, |T|^{-n/2} \exp(-np/2)$$

Bajo $H_g \cup H_G$ el máximo de la verosimilitud coincide con

$$L(\hat{\mu}_1,\dots,\hat{\mu}_g,\hat{\Sigma}_1,\dots,\hat{\Sigma}_g; X_1,\dots,X_g) = (2\pi)^{-np/2} \exp(-np/2) \prod_{i=1}^{g} |S_i|^{-n_i/2}$$

Por tanto, el cociente de verosimilitudes generalizado correspondiente es

$$\Lambda_g = \frac{|T|^{-n/2}}{\prod_{i=1}^{g} |S_i|^{-n_i/2}}$$

Del teorema 4.1.1 se obtiene que

$$W_g = -2\,\ell n\, \Lambda_g = n\, \ell n |T| - \sum_{i=1}^{g} n_i \ell n |S_i|$$

tiene una distribución asintótica χ^2. Los grados de libertad pueden obtenerse como el número de restricciones impuestas a los parámetros bajo la hipótesis nula. Respecto a los vectores de medias, el número de restricciones es $p(g-1)$ y como el número de parámetros distintos en cada matriz de covarianzas es $\frac{1}{2}\, p(p+1)$ existen $\frac{1}{2}\,(g-1)p(p+1)$ restricciones adicionales, por tanto el número de grados de libertad de W_g se reduce a

$$p(g-1) + \frac{1}{2}\,(g-1)p(p+1) = \frac{1}{2}\,(g-1)p(p+3)$$

4.2.3. IGUALDAD DE MEDIAS CONDICIONADA A IGUALDAD DE MATRICES DE COVARIANZAS.

El siguiente contraste comparativo entre grupos está dado por las siguientes hipótesis

$$H_h:\ \mu_1 = \mu_2 = \ldots = \mu_g \text{ dado que } \Sigma_1 = \Sigma_2 = \ldots = \Sigma_g$$

$$H_H:\ \mu_i \neq \mu_j \quad i \neq j \quad i,j=1,\ldots,g \text{ dado que } \Sigma_1 = \Sigma_2 = \ldots = \Sigma_g$$

La función de verosimilitud de los parámetros es la siguiente

$$L(\mu_1,\ldots,\mu_g,\Sigma;X_1,\ldots,X_g) = |2\pi\Sigma|^{-n/2}\exp\{-\frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}-\mu_i)'\Sigma^{-1}(X_{ij}-\mu_i)\} \quad (4.2.1)$$

Bajo H_h, la información muestral proviene de una población $N_p(\mu,\Sigma)$ y nuevamente por el teorema 1.3.1 los estimadores correspondientes son

$$\hat{\mu}_h = \overline{X} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i} X_{ij}$$

$$\hat{\Sigma}_h = T = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}-\overline{X})(X_{ij}-\overline{X})'$$

Bajo $H_h \cup H_H$ los estimadores máximo verosímiles de los parámetros están dados por las siguientes ecuaciones

$$\hat{\mu}_i = \overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \qquad i=1,\ldots,g$$

$$\hat{\Sigma} = W = \frac{1}{n}\sum_{i=1}^{g} n_i S_i = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)(X_{ij}-\overline{X}_i)' \qquad (4.2.2)$$

Para demostrar que estos estimadores maximizan la función de verosimilitud observamos que de la ecuación (1.3.7) se obtiene $L_i(\bar{X}_i,\Sigma;X_i) \geq L_i(\mu,\Sigma;X_i)\ \forall\ \mu$ y por tanto se deduce que

$$L(\bar{X}_1,\ldots,\bar{X}_g,\Sigma;X_1,\ldots,X_g) = \prod_{i=1}^{g} L(\bar{X}_i,\Sigma;X_i)$$

$$\geq \prod_{i=1}^{g} L(\mu_i,\Sigma;X_i)$$

$$= L(\mu_1,\ldots,\mu_g,\Sigma;X_1,\ldots,X_g)$$

Para demostrar que $\hat{\Sigma} = W$ maximiza la verosimilitud tomemos la diferencia entre el logaritmo de la verosimilitud evaluada en W y el logaritmo de la verosimilitud evaluada en Σ observando que

$$A = \ell(\bar{X}_1,\ldots,\bar{X}_g;W,X_1,\ldots,X_g) - \ell(\bar{X}_1,\ldots,\bar{X}_g,\Sigma;X_1,\ldots,X_g)$$

$$= -\frac{n}{2}\ \ell n\ |\Sigma^{-1}W| - np/2 + \frac{n}{2}\ \mathrm{tr}\ \Sigma^{-1}W$$

$$= \frac{np}{2}\left\{\frac{1}{p}\ \mathrm{tr}\ \Sigma^{-1}W - 1 - \ell n\ |\Sigma^{-1}W|^{1/p}\right\}$$

La matriz W es semidefinida positiva. Este hecho puede verificarse observando que

$$a'W a = \frac{1}{n}\sum_{i=1}^{g} n_i\ a'\ S_i\ a \geq 0$$

ya que de (1.3.10) se deduce que $S_i \geq 0$. Como $\Sigma > 0$ se obtiene que $\Sigma^{-1} > 0$ y por el teorema A.7c $\Sigma^{-1}W \geq 0$. Utilizando el teorema A.15

se sigue que los valores propios positivos de $\Sigma^{-1}W$ son los mismos que los asociados a $M=\Sigma^{-1/2}W\Sigma^{-1/2}$. Como $M>0$, los valores propios de $\Sigma^{-1}W$ son no negativos de donde por el teorema A.21 se infiere que

$$\Lambda > \frac{np}{2}\left\{\frac{1}{p}\sum_{i=1}^{p}\lambda_i - 1 - \ln\left(\frac{1}{p}\sum_{i=1}^{p}\lambda_i\right)\right\}$$

y por (1.3.8) se obtiene que $\Lambda>0$, lo cual demuestra que los estimadores (4.2.2) maximizan la verosimilitud bajo $H_h \cup H_H$. Puede verificase que el cociente de verosimilitudes resulta de esta manera

$$\Lambda_h = \left\{\frac{|W|}{|T|}\right\}^{n/2}$$

por lo que la región de rechazo equivale al conjunto de muestras tales que $\lambda=\Lambda^{2/n}<K$. La distribución de λ se puede establecer reexpresando la matriz T como

$$T=\frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i+\overline{X}_i-\overline{X})(X_{ij}-\overline{X}_i+\overline{X}_i-\overline{X})'$$

$$=\frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)(X_{ij}-\overline{X}_i)'+\frac{1}{n}\sum_{i=1}^{g}n_i(\overline{X}_i-\overline{X})(\overline{X}_i-\overline{X})'$$

$$=W+B$$

donde la matriz B se define como

$$B=\frac{1}{n}\sum_{i=1}^{g}n_i(\overline{X}_i-\overline{X})(\overline{X}_i-\overline{X})'$$

$\mathbf{W}$ y $\mathbf{B}$ coinciden respectivamente con las matrices usuales de sumas de cuadrados y productos cruzados dentro y entre grupos. λ puede escribirse como

$$\lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = |\mathbf{I} + \mathbf{W}^{-1}\mathbf{B}|^{-1}$$

La información muestral puede expresarse en términos matriciales como $\underline{\mathbf{X}}' = (\mathbf{X}_1', \ldots, \mathbf{X}_g')$, donde $\mathbf{X}_i$ es la matriz de dimensión $n_i \mathrm{x} p$ correspondiente a la muestra proveniente de la i-ésima población. Sea $\mathbf{1}_i$ un vector de dimensión n con la unidad en las entradas correspondientes a la i-ésima muestra y cero en otro lado y sea $\mathbf{I}_i = \mathrm{diag}(\mathbf{1}_i)$. De estas definiciones se obtiene que

$$\mathbf{1} = \sum_{i=1}^{g} \mathbf{1}_i$$

$$\mathbf{I} = \sum_{i=1}^{g} \mathbf{I}_i$$

La matriz $n\mathbf{W}$ es una forma cuadrática de las observaciones por lo que puede expresarse como $n\mathbf{W} = \underline{\mathbf{X}}' \mathbf{M}_1 \underline{\mathbf{X}}$. La forma de la matriz $\mathbf{M}_1$ puede encontrarse observando que la matriz $n\mathbf{W}$ corresponde a la suma de las matrices de covarianzas de los g grupos de observaciones, de manera que si se define

$$\mathbf{H}_i = \mathbf{I}_i - \frac{1}{n_i} \mathbf{1}_i \mathbf{1}_i'$$

se tiene que

$$n_i S_i = \underline{\mathbf{X}}' \mathbf{H}_i \underline{\mathbf{X}}$$

de donde la matriz $\mathbf{M}_1$ está dada por

$$\mathbf{M}_1 = \sum_{i=1}^{g} H_i$$

La matriz de la forma cuadrática asociada a la matriz $n\mathbf{B}$ corresponde a la suma de $\mathbf{g}$ matrices en las que el i-ésimo término corresponde a la matriz $n_i(\overline{X}_i-\overline{X})(\overline{X}_i-\overline{X})'$. Si se define la matriz $\mathbf{C}_i$ como

$$C_i = \frac{1}{n_i} 1_i 1_i' - \frac{1}{ng} 1\, 1'$$

se tiene que

$$\underline{X}' C_i \underline{X} = n_i \overline{X}_i \overline{X}_i' - \frac{n}{g} \overline{X}\, \overline{X}'$$

de donde

$$n B = \sum_{i=1}^{g} n_i (\overline{X}_i-\overline{X})(\overline{X}_i-\overline{X})' = \sum_{i=1}^{g} n \overline{X}_i \overline{X}_i' - n \overline{X}\, \overline{X}' = \underline{X}' \mathbf{M}_2 \underline{X}$$

y $\mathbf{M}_2$ resulta definida como $\mathbf{M}_2 = \sum_{i=1}^{g} C_i$. Las matrices $\mathbf{M}_1$ y $\mathbf{M}_2$ son idempotentes; para verificar el caso de $\mathbf{M}_1$ basta comprobar que

$$H_i H_j = \begin{cases} H_i & i=j \\ 0 & i \neq j \end{cases}$$

de donde se deduce que

$$M_1M_1=\sum_{i=1}^{g} H_i + \sum_{i\neq j} H_i\, H_j = M_1$$

Para demostrar que M_2 es idempotente obsérvese que

$$C_iC_j = \begin{cases} \frac{1}{ng^2}\; 1\, 1' - \frac{1}{ng}(1_i 1' + 1\, 1_i') + \frac{1}{n_i}\, 1_i 1_i' & i=j \\ \frac{1}{ng^2}\; 1\, 1' - \frac{1}{ng}(1_i 1' + 1\; 1_j') & i\neq j \end{cases}$$

y en consecuencia,

$$M_2\, M_2 = \sum_{i=1}^{g} C_i C_i + \sum_{i\neq j} C_i\, C_j$$

$$= \sum_{i=1}^{g} \left\{\frac{1}{ng^2}\; 1\, 1' - \frac{1}{ng}(1_i 1' + 1\, 1_i') + \frac{1}{n_i}\; 1_i 1_i'\right\}$$

$$+ \sum_{i\neq j} \left\{\frac{1}{ng^2}\, 1\, 1' - \frac{1}{ng}(1_i 1' + 1\, 1_j')\right\}$$

$$= \frac{1}{ng}\, 1\, 1' - \frac{2}{ng}\;\; 1\, 1' + \sum_{i=1}^{g} \frac{1}{n_i}\; 1_i 1_i'$$

$$+ \frac{g-1}{ng} 1\, 1' - \frac{1}{ng} \sum_{i=1}^{g} (1-1_i) 1' - \frac{1}{ng} \sum_{i=1}^{g}\; 1(1-1_i)'$$

$$= \sum_{i=1}^{g} \frac{1}{n_i}\; 1_i 1_i' - \frac{1}{ng}\; 1\, 1' + \frac{g-1}{ng}\; 1\, 1' - \frac{2}{ng}\; (g1-1) 1'$$

$$= \sum_{i=1}^{g} \frac{1}{n_i}\; 1_i 1_i' - \frac{1}{ng} 1\, 1' = \sum_{i=1}^{g} C_i = M_2$$

Así, puesto que M_1 y M_2 son matrices idempotentes, su rango coincide con la traza. Para M_1 se tiene

$$r(M_1)=\sum_{i=1}^{g} tr\{I_i-\frac{1}{n_i}1_i1_i'\}$$

$$=\sum_{i=1}^{g}(n_i-1)$$

$$=n-g$$

Por su parte, el rango de M_2 se obtiene como

$$r(M_2)=\sum_{i=1}^{g} tr\{\frac{1}{n_i}1_i1_i'-\frac{1}{ng}1\,1'\}$$

$$=\sum_{i=1}^{g}(1-\frac{1}{g})$$

$$=g-1$$

Las matrices M_1 y M_2 tienen además la propiedad de que sus renglones suman al vector cero. Este hecho puede verificarse como sigue

$$M_1'1=\sum_{i=1}^{g}(I_i-\frac{1}{n_i}1_i1_i')1$$

$$=1-\sum_{i=1}^{g}\frac{1}{n_i}1_i\,n_i$$

$$=0$$

$$M_2'1=\sum_{i=1}^{g}\{\frac{1}{n_i}1_i1_i'-\frac{1}{ng}1\,1'\}1$$

$$=\sum_{i=1}^{g}(1_i-\frac{1}{g}1)$$

$$=0$$

Bajo la hipótesis nula, $\underline{X}$ es una matriz de datos de una distribución $N_p(\mu,\Sigma)$, por lo que utilizando el teorema 3.3.9 se obtiene

$$n\,W = \underline{X}'\,M_1\,\underline{X} \sim W_p(\Sigma,n-g)$$

$$n\,B = \underline{X}'\,M_2\,\underline{X} \sim W_p(\Sigma,g-1)$$

El producto de las matrices M_1 y M_2 puede calcularse como

$$\begin{aligned}
M_1M_2 &= \left(\sum_{i=1}^{g} H_i\right)\left(\sum_{i=1}^{g} C_i\right)\\
&= \left(I-\sum_{i=1}^{g}\frac{1}{n_i}\,1_i1_i'\right)\left(\sum_{j=1}^{g}\frac{1}{n_j}\,1_j1_j' - \frac{1}{n}\,1\,1'\right)\\
&= \sum_{j=1}^{g}\frac{1}{n_j}\,1_j1_j' - \frac{1}{n}\,1\,1' - \sum_{i=1}^{g}\sum_{j=1}^{g}\frac{1}{n_i}\,\frac{1}{n_j}\,1_i1_i'\,1_j1_j'\\
&\quad + \sum_{i=1}^{g}\frac{1}{n_i n}\,1_i1_i'\,1\,1'\\
&= \sum_{j=1}^{g}\frac{1}{n_j}\,1_j1_j' - \frac{1}{n}\,1\,1' - \sum_{i=1}^{g}\frac{1}{n_i}\,1_i1_i' + \frac{1}{n}\,1\,1'\\
&= 0
\end{aligned}$$

Utilizando este resultado y el teorema 3.3.10a se deduce que las matrices $n\,W$ y $n\,B$ tienen distribución independiente. De la definición 3.5.1 se tiene que si $n > p+g$, entonces bajo H_h

$$\lambda = \frac{|n\,W|}{|n\,W+nB|} = |I+W^{-1}B|^{-1} \sim \Lambda(p,n-g,g-1)$$

Si p=1,2 ó g=2,3, diversas funciones de λ tienen distribución F. Estas funciones corresponden a las expresiones (3.5.5) a (3.5.8). Para otros valores de p y g puede utilizarse la aproximación de Bartlett (3.5.9) que en este caso resulta

$$\{n-\tfrac{1}{2}\ (p+g+2)\}\ \ell n\,|I+W^{-1}B| \sim \chi^2_{p(n-g)}$$

cuando $n\to\infty$.

4.2.4. IGUALDAD DE MATRICES DE COVARIANZAS.

El último contraste analizado en esta sección corresponde a la hipótesis de igualdad de matrices de covarianzas. Las hipótesis pueden escribirse como

$$H_i:\ \Sigma_1=\Sigma_2=\ldots=\Sigma_g$$

$$H_I:\ \Sigma_i\neq\Sigma_j\quad i\neq j\quad i,j=1,\ldots,g$$

El cociente de verosimilitudes generalizado para este contraste está dado por

$$\Lambda_i = \frac{\sup\limits_{H_i} L(\mu_1,\ldots,\mu_g,\Sigma;X_1,\ldots,X_g)}{\sup\limits_{H_i\cup H_I} L(\mu_1,\ldots,\mu_g,\Sigma;X_1,\ldots,X_g)}$$

$$= \frac{\sup\limits_{\mu_1,\ldots,\mu_g,\Sigma} L(\mu_1,\ldots,\mu_g,\Sigma;X_1,\ldots,X_g)}{\sup\limits_{\mu\,\Sigma} L(\mu,\Sigma;X_1,\ldots,X_g)}\ \frac{\sup\limits_{\mu,\Sigma} L(\mu,\Sigma;X_1,\ldots,X_g)}{\sup\limits_{\mu_1,\ldots,\mu_g,\Sigma_1,\ldots,\Sigma_g} L(\mu_1,\ldots,\mu_g,\Sigma_1,\ldots,\Sigma_g;X_1,\ldots,X_g)}$$

$$= \Lambda_h^{-1} \quad \Lambda_g$$

$$= \left\{ \frac{|W|}{|T|} \right\}^{-n/2} \quad \frac{|T|^{-n/2}}{\prod_{i=1}^{g} |S_i|^{-n_i/2}}$$

$$= |W|^{-n/2} \Big/ \prod_{i=1}^{g} |S_i|^{-n_i/2}$$

donde W y S_i se definen como

$$W = \frac{1}{n} \sum_{i=1}^{g} n_i S_i$$

$$S_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)(X_{ij} - \overline{X}_i)' \quad i=1,\ldots,g$$

Utilizando el teorema 4.1.1., la distribución de la variable $W_i = -2 \ln \Lambda_i$ puede aproximarse mediante la distribución χ^2. Los grados de libertad pueden calcularse como el número de restricciones impuestas a los parámetros bajo la hipótesis nula. Este número corresponde a $g-1$ restricciones de igualdad en las matrices de covarianzas multiplicado por el número de parámetros libres en cada matriz que es $\frac{1}{2}p(p+1)$, de donde se deduce que

$$W_i = -2 \ln \Lambda_i = n \ln |W| - \sum_{i=1}^{g} n_i \ln |S_i| \sim \chi^2_{\frac{1}{2} p(p+1)(g-1)}$$

cuando $n \to \infty$.

APENDICE

APENDICE

En esta sección se presentan algunas definiciones y 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 incluídas aquí pueden encontrarse en textos como Mardia(1979), Seber(1984), Graybill(1976) ó Rao(1973). La demostración de los teoremas A.22 y A.23 puede encontrarse en Deemer & Olkin(1951).

DEFINICION A.1. *Una matriz* $A_{n\times p}$ *es un arreglo de números en* n *renglones y* p *columnas como sigue*

$$A_{n\times p}=(a_{ij}) = \begin{bmatrix} a_{11} & a_{12} \cdots & a_{1p} \\ a_{21} & a_{22} \cdots & a_{2p} \\ \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{np} \end{bmatrix}$$

Si n=p *se dice que la matriz es cuadrada. Si* $A_{n\times n}$ *es una matriz con* $a_{ii}=1$ $i=1,\ldots,n$ *y* $a_{ij}=0$ $i\neq j$ *se dice que* A *es la matriz identidad 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*

$$A = \begin{bmatrix} A_1' \\ \vdots \\ A_n' \end{bmatrix} = (A_{(1)},\ldots,A_{(p)})$$

$$\mathrm{Diag}\ (A)=\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

DEFINICION A.5. *Una matriz* $A_{nxn}=(a_{ij})$ *es diagonal si* $a_{ij}=0$ *cuando* $i\neq j$.

DEFINICION A.6. *Una matriz* **A** *es simétrica si* $A=A'$.

DEFINICION A.7. *La traza de una matriz se denota por* tr(A) *y se define como* $\mathrm{tr}(A)=\sum_{i=1}^{n} a_{ii}$.

DEFINICION A.8. *Una matriz* A_{nxp} *se denomina particionada cuando se escribe en términos de submatrices; es decir*

$$A_{nxp} = \begin{bmatrix} A_{11\,(n_1xp_1)} & A_{12\,(n_1xp_2)} \\ A_{21\,(n_2xp_1)} & A_{22\,(n_2xp_2)} \end{bmatrix} \qquad (A.1)$$

donde $n=n_1+n_2$ *y* $p=p_1+p_2$.

DEFINICION A.9. *Los vectores* $x_1,\ldots,x_n$ *son linealmente dependientes si existen números* $\delta_1,\ldots,\delta_n$ *no todos cero tales que*

$$\sum_{i=1}^{n} \lambda_i X_i = 0.$$

Si los vectores $X_1, \ldots, X_n$ *no son linealmente dependientes se dice que son linealmente independientes.*

DEFINICION A.10. *El rango de una matriz* $\mathbf{A}_{n\times p}$ *se denota por* $r(\mathbf{A})$ *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.11. *La inversa de una matriz* $\mathbf{A}_{n\times n}$ *de rango completo se denota por* A^{-1} *y satisface las propiedades* $\mathbf{A}\mathbf{A}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$.

DEFINICION A.12. *Se dice que una matriz* $\mathbf{A}_{n\times n}$ *es idempotente si* $\mathbf{A}\mathbf{A}=\mathbf{A}$.

DEFINICION A.13. *La matriz* $\mathbf{A}_{n\times n}$ *es ortogonal si* $\mathbf{A}\mathbf{A}'=\mathbf{I}$.

DEFINICION A.14. *La matriz* $\mathbf{A}_{n\times n}$ *es definida positiva si* $X'\mathbf{A}X > 0 \quad \forall X \neq 0$ *y se denota por* $\mathbf{A} > 0$.

DEFINICION A.15. *La matriz* $\mathbf{A}_{n\times n}$ *es semidefinida positiva si* $X'\mathbf{A}X \geq 0 \quad \forall X \neq 0$ *y se denota por* $\mathbf{A} \geq 0$.

DEFINICION A.16. *El determinante de una matriz cuadrada* $\mathbf{A}$ *de orden* n se denota por $|\mathbf{A}|$ y se define como

$$|\mathbf{A}| = \sum_{t \in P} (-1)^{\delta(t)} a_{1i}\, a_{2j}, \ldots, a_{np}$$

donde P es el conjunto de todas las posibles permutaciones $(i, j, \ldots, p)$ de $(1, 2, \ldots, n)$ y $\delta(t)=1$ si la permutación es impar y $\delta(t)=2$ si la permutación es par. Si $|\mathbf{A}| \neq 0$ se dice que A es no singular.

DEFINICION A.17. *Para una matriz cuadrada* $A_{n\times n}(aij)$, *el cofactor* 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 obtenido eliminando el renglón* i *y la columna* j *de la matriz* A.

DEFINICION A.18. *Sea* $A_{n\times n}$ *cualquier matriz. El polimonio* $P(\lambda)=|A-\lambda I|$ *tiene* n *raíces* $\lambda_1, \ldots, \lambda_n$ *denominados valores propios de la matriz* A. *Los vectores* $\gamma_1, \ldots, \gamma_n$ *que satisfacen* $A\gamma_i = \lambda_i \gamma_i$ $i=1,\ldots,n$ *se denominan vectores propios de la matriz* A.

DEFINICION A.19. *El producto Kronecker de* $A_{n\times p}$ *y* $B_{p\times q}$ *se denota por* $A \otimes B$ *y se define como la matriz de* $np \times nq$ *siguiente*

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \ldots & a_{1p}B \\ a_{21}B & a_{22}B & \ldots & a_{2p}B \\ \vdots & & & \vdots \\ a_{n1}B & a_{n2}B & \ldots & a_{np}B \end{bmatrix}$$

TEOREMA A.1. *Sea* **A** *una matriz cuadrada de orden* **n** *con valores propios* $\lambda_1,\ldots,\lambda_n$ *y* $X_1, X_2, \ldots, X_p$ *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) $tr\left[\sum_{i=1}^{p} X_i' A X_i\right]=tr\left[A\left(\sum_{i=1}^{p} X_i X_i'\right)\right]$

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 contante. 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. *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) $(A W B)^{a} = (B' \otimes A)\ W^{a}$

TEOREMA A.5. *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) $a_i' a_i = a_{(i)}' a_{(i)} = 1 \quad i = 1, \ldots, n$

TEOREMA A.6. *Sean* $A_{nxp}, B_{nxp}, C_{nxn}$ *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) $r(A+B) = r(A) + r(B)$

d) $r(A'A) = r(A\,A') = r(A)$

e) $r(C\,A\,D) = r(A)$

f) *Si* $n=p$, $r(A)=p \iff |A| \neq 0$.

TEOREMA A.7. *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.8. *Sean* $\Sigma > 0$ *y* A *matrices simétricas de orden* p. *La matriz* $W = \Sigma - 2tA$ *es definida positiva en una vecindad de* t.

DEMOSTRACION. Sean los conjuntos P, C y N definidos por

$$P = \{X \in \mathbb{R}^p / X'A\,X > 0\}$$

$$C = \{X \in \mathbb{R}^p / X'A\,X = 0\}$$

$$N = \{X \in \mathbb{R}^p / X'A\,X < 0\}$$

Sea $f(X)$ definida por

$$f(X) = \frac{1}{2}\left[\frac{X'\Sigma\,X}{X'A\,X}\right]$$

para todo vector X tal que $X'A\,X$ es diferente de cero. Si P, C ó N son vacíos no existen restricciones impuestas a los valores de t por el respectivo conjunto para que $X'W\,X > 0$. Para cada conjunto no vacío defínase lo siguiente.

Si $X \in P$ se obtiene que $X'W\,X > 0$ si y sólo si $t < f(X)$. Definiendo $\varepsilon_1 = \min\limits_{X} f(X) > 0$ se obtiene que $X'W\,X > 0$ $\forall X \in P$ si $t < \varepsilon_1$.

Si $X \in C$ se obtiene que $X'W\,X > 0$ $\forall t \in \mathbb{R}$.

Si $X \in N$ se obtiene que $X'W\,X > 0$ si y sólo si $t > f(X)$. Definiendo $\varepsilon_2 = \max\limits_{X} f(X) < 0$ se obtiene que $X'W\,X > 0$ $\forall X \in N$ si $t > \varepsilon_2$.

Haciendo $\varepsilon = \min\{\varepsilon_1, |\varepsilon_2|\} > 0$ se obtiene que $X'W\,X > 0$ $\forall X \in \mathbb{R}^p$ si $|t| < \varepsilon$.

$$P = \{X \in \mathbb{R}^p / X'A\,X > 0\}$$

$$C = \{X \in \mathbb{R}^p / X'A\,X = 0\}$$

$$N = \{X \in \mathbb{R}^p / X'A\,X < 0\}$$

Sea $f(X)$ definida por

$$f(X) = \frac{1}{2}\left[\frac{X'\Sigma\,X}{X'A\,X}\right]$$

para todo vector X tal que $X'A\,X$ es diferente de cero. Si P, C ó N son vacíos no existen restricciones impuestas a los valores de t por el respectivo conjunto para que $X'W\,X > 0$. Para cada conjunto no vacío defínase lo siguiente.

Si $X \in P$ se obtiene que $X'W\,X > 0$ si y sólo si $t < f(X)$. Definiendo $\varepsilon_1 = \min\limits_{X} f(X) > 0$ se obtiene que $X'W\,X > 0$ $\forall X \in P$ si $t < \varepsilon_1$.

Si $X \in C$ se obtiene que $X'W\,X > 0$ $\forall t \in \mathbb{R}$.

Si $X \in N$ se obtiene que $X'W\,X > 0$ si y sólo si $t > f(X)$. Definiendo $\varepsilon_2 = \max\limits_{X} f(X) < 0$ se obtiene que $X'W\,X > 0$ $\forall X \in N$ si $t > \varepsilon_2$.

Haciendo $\varepsilon = \min\{\varepsilon_1, |\varepsilon_2|\} > 0$ se obtiene que $X'W\,X > 0$ $\forall X \in \mathbb{R}^p$ si $|t| < \varepsilon$.

TEOREMA A.9. $A_{n\times n}$ *es idempotente de rango* r *si y sólo si* $\lambda_1=\lambda_2,\ldots,\lambda_r=1$ *y* $\lambda_{r+1}=\lambda_{r+2},\ldots,\lambda_n=0$ *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) $A>0 \Longleftrightarrow \lambda_i>0 \quad i=1,\ldots,n$

b) $A\geq 0 \Longleftrightarrow \lambda_i\geq 0 \quad i=1,\ldots,n$

TEOREMA A.12. *El rango de una matriz* $A_{n\times n}$ *es igual al número de valores propios distintos de cero.*

TEOREMA A.13. *(Descomposición espectral). Cualquier matriz simétrica* $A_{n\times n}$ *puede escribirse como*

$$A = U\Lambda U' = \sum_{i=1}^{n} \lambda_i\, U_{(i)}U'_{(i)}$$

donde $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$, $\lambda_1,\ldots,\lambda_n$ *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 positiva definida* $A^{1/2}$ *tal que* $A = A^{1/2} A^{1/2}$.

TEOREMA A.15. *Si* $A_{nxn} > 0$ *y* $B_{nxn} > 0$, *los valores propios de* $B^{-1}A$, AB^{-1} *y* $B^{-1/2} A B^{-1/2}$ *son los mismos.*

TEOREMA A.16. *Si* A_{nxn} *es una matriz de rango* r, *existe una con columnas ortonormales* U_{nxr} *tal que* $U'AU = \Lambda$, *donde* $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r)$. *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_{mxp} *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. *(Descomposición de Cholesky).* *Si* A_{pxp} *es definida positiva entonces existe una única matriz triangular superior* T *con elementos positivos en la diagonal tal que* $A = T'T$.

TEOREMA A.19. *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.20. *La función Gamma definida por*

$$\Gamma\ (t) = \int_0^{\infty} u^{t-1} \exp(-u)\, du \ ; t > 0$$

satisface la relación

$$\Gamma(y+1) = \left[\frac{2^y}{\sqrt{\pi}}\right] \Gamma\left[\frac{y+2}{2}\right] \Gamma\left[\frac{y+1}{2}\right] = y! \ ; \ y=0,1,2,\ldots.$$

DEMOSTRACION. La función $\Gamma(t)$ satisface las relaciones

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

$$\Gamma(t+1) = t\,\Gamma\,(t) \qquad ; \ t \in \mathbb{R}$$

Utilizando estas relaciones se obtiene que

$$\Gamma(k+\tfrac{1}{2}) = (2k-1)(2k-3)\ldots,5\cdot 3\cdot\frac{\sqrt{\pi}}{2^k} \ ; \ k=0,1,2,\ldots.$$

Si y es un número impar positivo se deduce que

$$\Gamma\left[\frac{y+1}{2}\right] = (y-1)(y-3)\ldots.(4)(2)\ 2^{-(y-1)/2} \qquad ; \ y=1,3,5,\ldots,$$

$$\Gamma\left[\frac{y+2}{2}\right] = y\,(y-2)\ldots.(5)(3)\sqrt{\pi}\ 2^{-(y+1)/2} \qquad ; \ y=1,3,5,\ldots.$$

Si y es un número par no negativo se deduce que

$$\Gamma\left[\frac{y+1}{2}\right] = (y-1)(y-3)\ldots.(5)(3)\sqrt{\pi}\ 2^{-y/2} \qquad ; \ y=0,2,4,\ldots.$$

$$\Gamma\left[\frac{y+2}{2}\right] = y\,(y-2)\ldots.(4)(2)\ 2^{-y/2} \qquad ; \ y=0,2,4,\ldots.$$

de donde se obtiene que si y es un número entero no negativo

$$\Gamma\ (y+1) = \frac{2^y}{\sqrt{\pi}}\ \Gamma\left[\frac{y+2}{2}\right] \Gamma\left[\frac{y+1}{2}\right] = y!$$

TEOREMA A.21. *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

$$f(X) = \left(\prod_{i=1}^{p} X_i\right)^2$$

El máximo de $f(X)$ sujeto a $\| X \| = 1$ puede obtenerse utilizando multiplicadores de Lagrange y se alcanza en los vectores 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

$$Y = \left(\sum_{i=1}^{p} X_i^2\right)^{-1/2} (X_1, \ldots, X_p)$$

el valor de $f(Y)$ es menor o igual a p^{-p} lo cual se traduce en que

$$\left(\prod_{i=1}^{p} X_i^2\right)^{1/p} < p^{-1} \left(\sum_{i=1}^{p} X_i^2\right)$$

o equivalentemente para valores $\lambda_i = X_i^2 > 0$

$$\left(\prod_{i=1}^{p} \lambda_i\right)^{1/p} < \frac{1}{p} \sum_{i=1}^{p} \lambda_i = \bar{\lambda} \;)$$

lo cual demuestra el teorema.

TEOREMA A.22. *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) $\dfrac{\partial\, W'AW}{\partial W} = a$

c) $\dfrac{\partial |X|}{\partial x_{ij}} = \begin{cases} X_{ij} & i = j \\ 2X_{ij} & i \neq j \end{cases}$

donde X_{ij} *es el cofactor* i,j *de* X.

d) $\dfrac{\partial \ln |X|}{\partial X} = 2X^{-1} - \mathit{Diag}\ X^{-1}$

TEOREMA A.23. *Sean* X_{pxp} *y* Y_{pxp} *matrices simétricas. El jacobiano de la transformación* $Y = T(X) = AXA'$ *es* $|A^{-1}|^{p+1}$

TEOREMA A.24. *Sea* X_{pxp} *una matriz simétrica y* T *una matriz triangular superior. El jacobiano de la transformación* $T'T = X$ *es* $2^p \prod_{i=1}^{p} t_{ii}^{p+1-i}$

CONCLUSIONES

CONCLUSIONES

La distribución Normal Multivariada juega un papel fundamental en el desarrollo de los distintos procedimientos del Análisis Multivariado, teniendo un papel de especial importancia en áreas como: Análisis de Varianza Multivariado, Análisis Discriminante y Análisis de Regresión, entre otras. Algunas distribuciones de uso común en Estadística pueden generarse a partir de esta distribución; entre las incluidas en este trabajo se pueden mencionar la χ^2, F, Beta, Wishart, T^2 de Hotelling y Λ de Wilks. De esta manera, el modelo Normal Multivariado constituye parte esencial en la Estadística, por lo que es imprescindible un conocimiento adecuado de los principales resultados que sobre éste existen, pues proporciona las bases para un mejor manejo de las variadas técnicas estadísticas y en especial de las correspondientes al área multivariada.

En relación al propósito de esta tesis, y de acuerdo con lo anterior, en este trabajo se ha tratado de reunir y presentar de manera organizada los principales resultados reportados en la literatura sobre el modelo Normal Multivariado, con énfasis en los resultados referentes a estimación y pruebas de hipótesis.

En mi opinión, el trabajo reune material que puede ser utilizado con dos propósitos: El primero es servir como texto de apoyo o base en cursos del área de Análisis Multivariado a nivel Licenciatura o Posgrado y el segundo, como referencia en algunos de los cursos en las áreas de Probabilidad y Estadística que se imparten en la Facultad de Ciencias, tales como Probabilidad II, Análisis de Regresión y Diseño de Experimentos, entre otros.

Considero que la aportación principal, aparte de haber reunido los resultados más importantes sobre la Distribución Normal Multivariada y presentarlos de una manera consistente, es la de dar demostraciones más generales e incluso diferentes, a las reportadas comúnmente. Por otro lado se encontró que varios resultados, en la literatura revisada, no son satisfactorios ; en estos casos, se incluyeron las hipótesis necesarias para desarrollar una demostración adecuada, y en algunos teoremas se dieron extensiones.

En lo personal, la realización de este trabajo, aparte de haberme brindado una mayor compenetración en el tema, representa para mí una adición a la literatura en el área, disponible actualmente en el país.

BIBLIOGRAFIA

BIBLIOGRAFIA

Anderson T.W.(1958). *An introduction to Multivariate Statistical Analysis*. Wiley, New York.

Arnold S.F.(1981). *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York.

Atchley W.R.& Bryant E.H.(1975). *Multivariate Statistical Methods: Among-Groups Covariation. Benchmark Papers in Systematic and Evolutionary Biology, Vol.1*. Dowden,Hutchinson & Ross,Inc., Pennsylvania.

Bartlett M.S.(1947). *Multivariate Analysis*. Journal of the Royal Statistical Society.B,9,176-197.

Billingsley P.(1979). *Probability and Measure*. Wiley, New York.

Bryant E.H.& Atchley W.R.(1975). *Multivariate Statistical Methods: Within Groups Covariation.Benchmark Papers in Systematic and Evolutionary Biology, Vol.2*. Dowden,Hutchinson & Ross,Inc., Pennsylvania.

Cooley W.W. & Lohnes P.R.(1971). *Multivariate Data Analysis*. Wiley, New York.

Cramér H.(1946). *Mathematical Methods of Statistics*. Princeton University Press.

Cramér,H.(1968). *Métodos Matemáticos de Estadística, Cuarta Edición* Aguilar S.A. de Ediciones, España.

Deemer W.L. & Olkin I.(1951). *The Jacobians of Certain Matrix Transformations Useful in Multivariate Analysis.* Biometrika 38, 345-367.

Dillon W.R. & Goldestein M.(1984). *Multivariate Analysis: Methods and Applications.* Wiley, New York.

Feller W.(1968). *An Introduction to Probability Theory and its Applications.Vol.1.* Wiley, New York.

Feller W.(1971). *An introduction to Probability theory and its Applications,Vol.2.* Wiley, New York.

Feller W.(1978). *Introducción a la Teoría de Probabilidades y sus Aplicaciones, Vol.2, Segunda Edición.* Editorial Limusa S.A., México.

Giri N.C.(1977). *Multivariate Statistical Inference.* Academic Press, New York.

Graybill F.A.(1976). *Theory and Application of the Linear Model.* Duxbury Press.

Harris B.(1966). *Theory of Probability.* Addison-Wesley,Massachusetts.

Hogg R.V. & Craig A.T.(1978). *Introduction to Mathematical Statistics.* Macmillan Publishing Co.,New York.

Johnson N.L. & Kotz S.(1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley,New York.

Kendall M.& Stuart A.(1977). *The Advanced Theory of Statistics, Vol.1.* Charles Griffin & Company. London.

Kendall M.& Stuart A.(1979). *The Advanced Theory of Statistics, Vol.2.* Charles Griffin & Company, London.

Kendall M. & Stuart A.(1976). *The Advanced Theory of Statistics, Vol.3.* Charles Griffin & Company, London.

Lindgren B.W.(1976). *Statistical Theory.* Macmillan Publishing Co., New York.

Mardia K.V., Kent J.T. & Bibby J.M.(1979). *Multivariate Analysis.* Academic Press, London.

Marriot F.H.C.(1974). *The Interpretation of Multiple Observations.* Academic Press, London.

Mood A.M., Graybill F.A.& Boes D.C.(1974). *Introduction to the Theory of Statistics.* McGraw-Hill, Tokyo, Japan.

Morrison D.F.(1976). *Multivariate Statistical Methods.* McGraw-Hill, Tokyo, Japan.

Olkin I.(1952). *Note on the Jacobians of Certain Matrix Transformations Useful in Multivariate Analysis.* Biometrika 40, 43-46.

Rao C.R.(1973). *Linear Statistical Inference & its Applications.* Wiley, New York.

Scheffé H.(1959). *The Analysis of Variance.* Wiley, Toronto Canada.

Searle S.R.(1971). *Linear Models.* Wiley, New York.

Seber G.A.F.(1984). *Multivariate Observations.* Wiley, New York.

Silvey S.D.(1970). *Statistical Inference.* Penguin, Baltimore.

Srivastava M.S. & Khatri C.G.(1979). *An introduction to Multivariate Statistics.* North Holland, New York.

Tatsuoka M.M.(1971). *Multivariate Analysis: Techniques for Educational & Psychological Research.* Wiley, New York.

Wilks S.S.(1962). *Mathematical Statistics.* Wiley, New York.