

I.- Distribución Normal Multivariada

1.1 Resultados de Álgebra Lineal

Lema 1.1.1. A y B son dos matrices cuadradas con inversa cada una entonces
 $(AB)^{-1} = B^{-1}A^{-1}$

Lema 1.1.2. a) $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$
b) $\text{tr}(AB) = \text{tr}(BA)$ siempre que AB y BA puedan efectuarse

Lema 1.1.3. Si una matriz A es simétrica entonces también lo es su inversa A^{-1} . O sea, si A es simétrica, $(A^{-1})^t = A^{-1}$.

Lema 1.1.4. a) $|AB| = |A| |B|$
b) $|A^{-1}| = \frac{1}{|A|}$

Lema 1.1.5. Las raíces característica de una matriz simétrica son reales.

Lema 1.1.6. Las raíces características de A distintas de cero son igual al rango de A.

Lema 1.1.7. Una matriz simétrica con raíces iguales a cero o a uno es una matriz idempotente.

Lema 1.1.8 Si $A + B = I$ y $AB = 0$, entonces A y B son idempotentes.

Lema 1.1.8. Si A es idempotente y simétrica de rango r , existe una matriz ortogonal P tal que $P^tAP = E_r$ donde E_r es una matriz diagonal con r elementos iguales a uno y el resto a ceros.

Lema 1.1.9. Si A es idempotente y simétrica de rango r entonces $\text{tr}(A) = r$.

Lema 1.1.10 Si A es idempotente y simétrica y P es ortogonal P^tAP es idempotente y simétrica.

Definición 1.1.1. Si $X^tAX > 0$ se dice que X^tAX es definida positiva, para todo $X \neq 0$ y A es positiva.

Definición 1.1.2. Si $X^tAX \geq 0$ se dice que X^tAX es semipositiva, para todo $X \neq 0$ y A es semipositiva.

Definición 1.1.3. Si A es positiva o semipositiva A es no negativa.

Lema 1.1.11. La matriz A es positiva si, y sólo si todos los determinantes de sus matrices angulares son positivos.

Lema 1.1.12. Si P es no singular P^tAP es o no es definida positiva (semi) si A es o no lo es.

Lema 1.1.13. Matrices definidas positivas son no singulares.

Lema 1.1.14. Matrices semipositivas son singulares (el caso contrario no siempre ocurre es decir una matriz singular no es siempre semipositiva).

Lema 1.1.15. Las raíces de una matriz positiva (semi) son todas mayores que cero (mayor o igual a cero).

Lema 1.1.16. AA^t es positiva cuando A es de rango completo por filas y semipositiva en cualquier otro caso.

Lema 1.1.17. A^tA es positiva cuando A tiene rango completo por columna y semipositiva en cualquier otro caso.

Lema 1.1.18. Si A es simétrica de orden n y rango r puede escribirse como $A = LL^t$, L es de $n \times r$ de rango r .

Lema 1.1.19. Si A es simétrica, es positiva si, y sólo si puede escribirse como PP^t con P no singular.

Lema 1.1.20 (Lema de Loynes). Si B es simétrica e idempotente y Q es simétrica y no negativa y si $I - B - Q$ es no negativa entonces $BQ = QB = 0$.

Dermostración.

Sea $Y = BX$ para algún X , entonces

$$Y^tBY = Y^tBBX = Y^tBX = Y^tY$$

y

$$Y^t(I - B - Q)Y = -Y^tQY$$

como $I-B-Q$ es no negativa entonces $Y^t QY = 0$, como Q es simétrica, $Q = LL^t$, $Y^t QY = Y^t LL^t Y = 0$ implica que $L^t Y = 0$ y entonces $LL^t Y = 0$, lo que es lo mismo $QY = QBX = 0$ como es para algún X , $QB = 0$ y así tenemos

$$(QB)^t = B^t Q^t = BQ = 0$$

□

Teorema 1.1.1. Si A_i son matrices de $n \times n$ simétricas de rango k_i , $i = 1, \dots, p$.

Sea $A = \sum_{i=1}^p A_i$, simétrica de rango k

Entonces las condiciones

a) A_i es idempotente para todo i

b) $A_i A_j = 0$ para todo $i \neq j$

c) A es idempotente

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

I) Para alguna de las dos a), b) y c) implica a), b), c) y d).

II) si c) y d) \Rightarrow a) y b)

III) si c) y A_1, A_2, \dots, A_{p-1} idempotentes con A_p no negativa implica A_p idempotente implica a) y por consiguiente b) y d).

Demostración.

I) Si se cumple a) y c) \Rightarrow b)

Podemos observar que $I-A$ es idempotente y por tanto no negativa y $A - A_i - A_j = \sum_{\substack{r=1 \\ i \neq r \neq j}}^p A_r$

por a) es no negativa $I - A + A - A_i - A_j = I - A_i - A_j$, entonces por Loynes $A_i A_j = 0$ para $i \neq j$.

I) Si se cumple b) y a) \Rightarrow c)

$$AA = \left(\sum_{i=1}^p A_i \right)^2 = \sum_{i=1}^p A_i^2 + \sum_{\substack{i=1 \\ i \neq j}}^p \sum_{j=1}^p A_i A_j = \sum_{i=1}^p A_i^2 = A.$$

I) Si se cumple b) y c) \Rightarrow a)

Sea $A_1 v = \lambda v$ para todo $\lambda \neq 0$, despejando $v = \frac{A_1 v}{\lambda}$, entonces $A_i v = \frac{A_i A_1 v}{\lambda}$, $i \neq 1$

Por b) $A_1 v = 0$, luego entonces tenemos que $A_1 v = A_1 v = \lambda v$, como A_1 es idempotente, entonces sus raíces deben ser 0 ó 1, por lo tanto las raíces de A_1 son 0 ó 1 lo que significa que A_1 es idempotente haciendo lo mismo para el resto de las matrices tenemos que a) se verifica.

I) Si se cumple a) y c) \Rightarrow d)

$$k = \text{rango}(A) = \text{tr}(A) = \text{tr}\left(\sum_{i=1}^p A_i\right) = \sum_{i=1}^p \text{tr}(A_i) = \sum_{i=1}^p k_i$$

II) Si c) y d) \Rightarrow a) y b)

Como A es idempotente $A - I$ lo es y tiene rango $n - k$, tenemos que

$$\begin{aligned} (A - I)X = 0 & \text{ tiene } n - k \text{ ecuaciones linealmente independientes} \\ A_1 X = 0 & \text{ tiene } k_1 \text{ ecuaciones linealmente independientes} \\ & \vdots \\ A_p X = 0 & \text{ tiene } k_p \text{ ecuaciones linealmente independientes} \end{aligned}$$

El número máximo de ecuaciones linealmente independientes del sistema anterior es $n - k_1$, el sistema anterior lo podemos reducir a $A_1 X = X$, tiene por lo menos $n - (n - k_1) = k_1$ ecuaciones linealmente independientes. Tenemos que 1 es una raíz característica de A_1 , con multiplicidad de al menos k_1 , pero como $\text{rango}(A_1) = k_1$, así tenemos A_1 que es idempotente, con lo que queda demostrado el punto II.

Ahora III) Si c) y A_1, A_2, \dots, A_{p-1} idempotentes con A_p no negativa implica A_p idempotente que implica a) y por consiguiente b) y d).

Como A es no negativa entonces también lo es $I - A$, como A_1, A_2, \dots, A_{p-1} idempotentes, significa que son no negativas y como también es no negativa entonces

$$A - A_i - A_j = \sum_{\substack{r=1 \\ i \neq r \neq j}}^p A_r \text{ es no negativa}$$

Luego $I - A + A - A_i - A_j = I - A_i - A_j$, es no negativa entonces por Loynes $A_i A_j = 0$ para $i \neq j$. Con lo que b) se cumple y de ahí todas las demás. □

Teorema 1.1.2. Si X es un vector de $p \times 1$ con elementos x_i y sea A un vector de $p \times 1$ con

elementos a_i y sea $Z = X^t A = A^t X$ la derivada de Z con respecto a X es $\frac{\partial Z}{\partial X} = \begin{pmatrix} \frac{\partial Z}{\partial x_1} \\ \vdots \\ \frac{\partial Z}{\partial x_p} \end{pmatrix} = A$

Demostración.

El i -ésimo elemento de $\frac{\partial Z}{\partial X}$ es;

$$\frac{\partial Z}{\partial x_i} = \frac{\partial \left(\sum_{j=1}^p a_j x_j \right)}{\partial x_i}$$

así que el i -ésimo elemento de $\frac{\partial Z}{\partial X}$ es a_i , por lo que $\frac{\partial Z}{\partial X} = A$

□

Teorema 1.1.3. Sea A un vector de $p \times 1$ y B de $q \times 1$ y X una matriz de $p \times q$ cuyo ij -ésimo elemento es x_{ij} . Sea $Z = A^t X B = \sum_{m=1}^q \sum_{n=1}^p a_n x_{nm} b_m$, entonces $\frac{\partial Z}{\partial X} = AB^t$.

Demostración.

El ij -ésimo elemento de $\frac{\partial Z}{\partial X}$ es

$$\frac{\partial Z}{\partial x_{ij}} = \frac{\partial \left(\sum_{m=1}^q \sum_{n=1}^p a_n x_{nm} b_m \right)}{\partial x_{ij}} = a_i b_j$$

por tanto se sigue que $\frac{\partial Z}{\partial X} = AB^t$.

□

Teorema 1.1.4. Sea A una matriz simétrica de $p \times p$ y X de $p \times 1$. Sea $Z = X^t A X$, entonces $\frac{\partial Z}{\partial A} = 2XX^t - D(XX^t)$, donde $D(XX^t)$ es la matriz diagonal cuyos elementos son la diagonal de XX^t .

Demostración.

El ij -ésimo elemento de $\frac{\partial Z}{\partial A}$ es

$$\frac{\partial Z}{\partial a_{ij}} = \frac{\partial \left(\sum_{m=1}^p \sum_{n=1}^p x_n x_m a_{mn} \right)}{\partial a_{ij}}$$

si $i = j$, $\frac{\partial Z}{\partial a_{ij}} = x_i^2$. Si $i \neq j$, $\frac{\partial Z}{\partial a_{ij}} = 2x_i x_j$. Así tenemos que $\frac{\partial Z}{\partial A} = 2XX^t - D(XX^t)$.

□

Teorema 1.1.5. Sea A una matriz simétrica de $p \times p$ y X de $p \times 1$. Sea $Z = X^t A X$, entonces

$$\frac{\partial Z}{\partial X} = 2AX.$$

Demostración.

El ij -ésimo elemento de $\frac{\partial Z}{\partial X}$ es

$$\begin{aligned} \frac{\partial Z}{\partial x_i} &= \frac{\partial \left(\sum_{m=1}^p \sum_{n=1}^p x_n x_m a_{mn} \right)}{\partial x_i} \\ &= \frac{\partial \left(\sum_{m=1}^p x_m^2 a_{mm} + \sum_{\substack{m=1 \\ m \neq n}}^p \sum_{n=1}^p x_n x_m a_{mn} \right)}{\partial x_i} \\ &= 2x_i a_{ii} + 2 \sum_{\substack{n=1 \\ n \neq i}}^p x_n a_{in} = 2 \sum_{n=1}^p x_n a_{in} \end{aligned}$$

luego entonces $\frac{\partial Z}{\partial X} = 2AX$.

□

1.2 Normal Multivariada

1.2.1 Definición y propiedades

La generalización de la función de densidad de una distribución Normal a varias dimensiones juega un papel importante en el análisis de regresión.

Una ventaja de la distribución normal multivariada parte del hecho de que es matemáticamente tratable y se pueden obtener resultados “buenos”. Frecuentemente este no es el caso con otras distribuciones multivariadas.

Antes de definir la distribución Normal Multivariada, necesitamos definir algunos momentos de un vector aleatorio, es decir, un vector cuyos componentes están

distribuidos conjuntamente. La media o esperanza de un vector aleatorio $X = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ de

$m \times 1$ esta definido por ser el vector de esperanzas:

$$E(X) = \begin{pmatrix} E(x_1) \\ \vdots \\ E(x_m) \end{pmatrix}$$

Mas generalmente, si $Z = (z_{ij})$ es una matriz aleatoria de $p \times q$, entonces $E(Z)$, la esperanza de Z , es la matriz cuyo ij -ésimo elemento es $E(z_{ij})$. Aunque no entremos en detalle del desarrollo del siguiente resultado, es simple comprobar que si B , C y D son matrices de constantes de $m \times p$, $p \times n$ y $m \times n$ respectivamente, entonces

$$E(BZC + D) = B E(Z) C + D \quad (1.1)$$

Si X tiene media μ la *matriz de varianza-covarianza* (*varcov*) de X esta definida por la matriz de

$$\Sigma = \text{Cov}(X) = E[(X - \mu)(X - \mu)^t].$$

El elemento ij -ésimo de Σ es

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

la covarianza entre las variables x_i y x_j , y el ii -ésimo elemento es

$$\sigma_{ii} = E[(x_i - \mu_i)^2],$$

la varianza de x_i , la matriz de varianza-covarianza la podemos poner como

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

es decir, es una matriz de orden p y donde σ_{ji} es la covarianza entre la variable j e i , como podemos ver los elementos de la diagonal de Σ son no negativos. Σ es simétrica, es decir, $\Sigma = \Sigma^t$. Además, la clase de matrices de varianza-covarianza coincide con la

clase de matrices no-negativas definida. Recordemos que una matriz simétrica A de $m \times m$ es no negativa definida si

$$\alpha^t A \alpha \geq 0 \quad \text{para toda } \alpha \in \mathbb{R}^m$$

y definida positiva si

$$\alpha^t A \alpha > 0 \quad \text{para toda } \alpha \in \mathbb{R}^m .$$

Teorema 1.2.1. La matriz Σ de $m \times m$ es una matriz de varianza-covarianza , si, y solo si es no-negativa definida.

Demostración.

Supongamos que Σ es la matriz de varianza-covarianza de un vector aleatorio X , donde X tiene media μ , entonces para toda $\alpha \in \mathbb{R}^{m \times 1}$,

$$\begin{aligned} \text{Var} (\alpha^t X) &= E [(\alpha^t X - \alpha^t \mu) (\alpha^t X - \alpha^t \mu)^t] & (1.2) \\ &= E [\alpha^t (X - \mu) (X - \mu)^t \alpha] \\ &= \alpha^t \Sigma \alpha \geq 0 \end{aligned}$$

luego Σ es no negativa definida.

Ahora supongamos que Σ es una matriz no-negativa definida de rango r , digamos ($r \leq m$). Escribimos $\Sigma = C C^t$, donde C es una matriz de $m \times r$ de rango r . Sea Y un vector de $r \times 1$ de variables aleatorias independientes con media 0 y $\text{Cov} (Y) = I$ para $X = CY$. Entonces $E (X) = 0$ y

$$\begin{aligned} \text{Cov} (X) &= E [XX^t] = E[C Y Y^t C^t] \\ &= C E (Y Y^t) C^t \\ &= C C^t = \Sigma , \end{aligned}$$

tal que Σ es una matriz varcov. □

Comúnmente hacemos transformaciones lineales de vectores aleatorios y necesitamos saber como las matrices de varianza-covarianza son transformadas. Supongamos que X es un vector aleatorio de $m \times 1$ con media μ_x y matriz de varianza-covarianza Σ_x y sea $Y = BX + b$, donde B es de $k \times m$ y b es de $k \times 1$. la media de Y es, $\mu_y = B \mu_x + b$, y la matriz de varianza-covarianza de Y es;

$$\begin{aligned} \Sigma_y &= E [(Y - \mu_y) (Y - \mu_y)^t] & (1.3) \\ &= E [(BX + b - (B\mu_x + b)) (BX + b - (B\mu_x + b))^t] \\ &= B E [(X - \mu_x) (X - \mu_x)^t] B^t \end{aligned}$$

$$= B \Sigma B^t$$

Recordemos que la distribución normal univariada, con media μ y varianza σ^2 , tiene función de densidad:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

ahora considerando el exponente:

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu) (\sigma^2)^{-1} (x-\mu)$$

podemos generalizarlo para un vector de las observaciones sobre varias variables, X de $p \times 1$ tal que tendríamos

$$(X - \mu_X)^t \Sigma^{-1} (X - \mu_X)$$

donde el vector μ_X de $p \times 1$ representa el valor esperado del vector aleatorio X y la matriz Σ es la matriz de varianza covarianza.

La constante de la normal univariada $\frac{1}{\sqrt{2\pi\sigma^2}}$ se sustituya por una constante mas general, la constante para la distribución normal multivariada es $(2\pi)^{-p/2} |\Sigma|^{-1/2}$, consecuentemente la *densidad normal p -dimensional* para el vector aleatorio X tiene la forma:

$$f_X(X) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(X-\mu_X)^t \Sigma^{-1} (X-\mu_X)}$$

donde $-\infty < x_i < \infty$, $i = 1, \dots, p$. y Σ es de rango p . La media de X esta definida por $E[X] = \mu$, y la matriz de varianza covarianza por $E[(X - \mu_X)(X - \mu_X)^t] = \Sigma$.

Usualmente la densidad normal p -dimensional se denota por $N_p(\mu, \Sigma)$, la cual es análoga al caso univariado.

Ahora daremos la definición formal de la densidad normal multivariada.

Definición 1.2.1. Si y_1, y_2, \dots, y_p son p variables aleatorias y si Y es el vector $p \times 1$ de estas variables

$$f(y_1, y_2, \dots, y_p) = K e^{-\frac{1}{2}(Y-\mu)^t R (Y-\mu)} \quad -\infty < y_i < \infty, i = 1, \dots, p$$

es la *función de densidad de una normal multivariada* si se cumple

- a) R es una matriz definida positiva, r_{ij} son constantes.

- b) K es una constante positiva.
 c) μ_i es el i -ésimo elemento de $\boldsymbol{\mu}$, μ_i son constantes.

$$K = \frac{|\mathbf{R}|^{1/2}}{(2\pi)^{p/2}}$$

Haciendo $\mathbf{R} = \boldsymbol{\Sigma}^{-1}$, tenemos la misma función deducida anteriormente a partir de la función de densidad normal univariada.

Otra definición alterna es la siguiente.

Definición 1.2.2. El vector aleatorio de $m \times 1$ \mathbf{X} se dice que tiene una distribución normal m -variada si, para cada $\boldsymbol{\alpha} \in \mathbb{R}^{m \times 1}$, la distribución de $\boldsymbol{\alpha}^t \mathbf{X}$ es normal univariada.

A partir de las definiciones ahora estableceremos algunas propiedades de la distribución normal multivariada.

Teorema 1.2.2. Si \mathbf{X} es $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces la función característica de \mathbf{X} es

$$\phi_{\mathbf{X}}(\mathbf{v}) = \exp\left(i\mathbf{v}^t \boldsymbol{\mu} - \frac{1}{2} \mathbf{v}^t \boldsymbol{\Sigma} \mathbf{v}\right) \quad (1.4)$$

$\mathbf{v} \in \mathbb{R}^{m \times 1}$

Demostración.

Aquí

$$\phi_{\mathbf{X}}(\mathbf{v}) = E[\exp(i\mathbf{v}^t \mathbf{X})] = \phi_{\mathbf{v}^t \mathbf{X}}(1),$$

El lado derecho denota la función característica de la variable aleatoria $\mathbf{v}^t \mathbf{X}$ evaluada en 1. Ya que \mathbf{X} es $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces $\mathbf{v}^t \mathbf{X}$ se distribuye como una normal univariada, $\mathbf{v}^t \mathbf{X} \sim N(\mathbf{v}^t \boldsymbol{\mu}, \mathbf{v}^t \boldsymbol{\Sigma} \mathbf{v})$ tal que

$$\phi_{\mathbf{v}^t \mathbf{X}}(1) = \exp\left(i\mathbf{v}^t \boldsymbol{\mu} - \frac{1}{2} \mathbf{v}^t \boldsymbol{\Sigma} \mathbf{v}\right).$$

lo cual completa la demostración. □

Si u_i con $i = 1, 2, \dots, r$, son v.a.i.i.d con distribución normal estándar, si \mathbf{U} es el vector de $r \times 1$ cuyos elementos son dichas variables.

$$\phi_{\mathbf{U}}(\mathbf{v}) = E[\exp(i\mathbf{v}^t \mathbf{U})]$$

$$\begin{aligned}
&= \prod_{j=1}^r E[\exp(iv_j u_j)] && \text{(por} \\
&\text{independencia)} \\
&= \prod_{j=1}^r \exp\left(-\frac{1}{2} v_j^2\right) && \text{(por normalidad)} \\
&= \exp\left(-\frac{1}{2} \mathbf{v}^t \mathbf{v}\right)
\end{aligned}$$

Ahora si ponemos

$$\mathbf{X} = \mathbf{C}\mathbf{U} + \boldsymbol{\mu}$$

donde \mathbf{C} es una matriz de $m \times r$ de rango r tal que $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^t$, y $\boldsymbol{\mu} \in \mathbb{R}^{m \times 1}$. Entonces \mathbf{X} tiene función característica (1.4),

$$\begin{aligned}
E[\exp(i\mathbf{v}^t \mathbf{X})] &= E[\exp(i\mathbf{v}^t \mathbf{C}\mathbf{U})] \exp(i\mathbf{v}^t \boldsymbol{\mu}) \\
&= \phi_{\mathbf{U}}(\mathbf{C}^t \mathbf{v}) \exp(i\mathbf{v}^t \boldsymbol{\mu}) \\
&= \exp(-\frac{1}{2} \mathbf{v}^t \mathbf{C}\mathbf{C}^t \mathbf{v}) \exp(i\mathbf{v}^t \boldsymbol{\mu}) \\
&= \exp(i\mathbf{v}^t \boldsymbol{\mu} - \frac{1}{2} \mathbf{v}^t \boldsymbol{\Sigma} \mathbf{v}).
\end{aligned}$$

Vale comentar que podríamos haber definido la distribución normal multivariada $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ por medio de la transformación lineal sobre variables normal estándar independientes. Tal que dicha representación en la práctica es muy útil.

1.2.2 Formas lineales, Marginales y Condicionales.

Regresando a las propiedades de la distribución normal multivariada, los siguientes resultados muestran que cualquier transformación lineal de un vector normal tiene una distribución normal.

Teorema 1.2.3. Si \mathbf{X} es $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y \mathbf{B} es una matriz de dimensión $k \times m$, \mathbf{b} es un vector de $k \times 1$, entonces

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b} \text{ es } N_k(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t).$$

Demostración.

El hecho de que \mathbf{Y} es normal k -variada es una consecuencia directa de la definición 1.2.2, puesto que todas las funciones lineales de los componentes de \mathbf{Y} son

funciones lineales de los componentes de X y estos son todos normales. La media y la covarianza de la matriz Y están claramente definidas. □

Una propiedad importante de la distribución normal multivariada es que *todas* las distribuciones marginales son normales.

Teorema 1.2.4. Si X es $N_m(\mu, \Sigma)$, entonces la distribución marginal de cualquier subconjunto de $k < m$ componentes de X es normal k -variada.

Demostración.

Esto se sigue directamente de la definición, o del teorema 1.2.3. Por ejemplo, hacemos una partición de X , μ y Σ como

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

donde X_1 y μ_1 son de $k \times 1$ y Σ_{11} es de $k \times k$, poniendo

$$B = [I_k : 0] \quad \text{es de dimensión } k \times m \text{ y } b = \mathbf{0}$$

En el teorema 1.2.3 se muestra inmediatamente que X_1 es $N_m(\mu_1, \Sigma_{11})$. □

Una consecuencia de este teorema es que la distribución marginal de cada componente de X es normal univariada. Lo inverso no es cierto en general; esto es, el hecho de que cada componente de un vector aleatorio no es (marginalmente) normal no implica que el vector tenga una distribución normal multivariada. [Esta es una de las razones por la cual el problema de *pruebas* (de hipótesis) de normalidad multivariada es hasta cierto punto un tanto complicado en la práctica].

Recordemos que la independencia de dos variables aleatorias implican que la covarianza entre ellas, si esta existe, es cero, pero que el caso contrario en general no es cierto. Esta característica, para la distribución normal multivariada, es como la muestra el siguiente resultado.

Teorema 1.2.5. Si X es $N_m(\mu, \Sigma)$ y X , μ y Σ es una partición como la siguiente:

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

donde X_1 y μ_1 son de $k \times 1$ y Σ_{11} es de $k \times k$, entonces los subvectores X_1 y X_2 son independientes si, y solo si $\Sigma_{12} = 0$.

Demostración.

Σ_{12} es la matriz de covarianzas entre los componentes de X_1 y los componentes de X_2 , de tal manera que la independencia de X_1 y X_2 implica que $\Sigma_{12} = 0$.

Sea $\Sigma_{12} = 0$, la función de densidad de X es,

$$\begin{aligned} f(\mathbf{X}) &= \frac{1}{(2\pi)^{p/2} \left| \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right|^{1/2}} e^{-\frac{1}{2} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix}^t \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix}} \\ &= \frac{1}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2} |\Sigma_{22}|^{1/2}} e^{-\frac{1}{2} ((X_1 - \mu_1)^t \Sigma_{11}^{-1} (X_1 - \mu_1) + (X_2 - \mu_2)^t \Sigma_{22}^{-1} (X_2 - \mu_2))} \\ &= f(X_1) f(X_2) \end{aligned}$$

por lo tanto X_1, X_2 son independientes.

□

Este teorema se puede extender al caso donde se hacen particiones de \mathbf{X} en un número de subvectores. Lo importante de esto es que para determinar si dos subvectores de un vector distribuido normalmente son independientes es suficiente comprobar que la matriz de covarianzas entre los dos subvectores es cero.

Teorema 1.2.6. Si los vectores aleatorios X y Y de $m \times 1$ son independientes y $X + Y$ tienen una distribución normal m -variada, entonces X y Y se distribuyen como normal.

Demostración.

Para cada $\alpha \in \mathbb{R}^{m \times 1}$, la distribución de $\alpha^t(X + Y) = \alpha^t X + \alpha^t Y$ es normal (por el teorema 1.2.3, ya que $X + Y$ es normal). Puesto que $\alpha^t X$ y $\alpha^t Y$ son independientes, implica que ambas son normales, y por lo tanto X y Y son normal m -variada.

□

Una propiedad bien conocida de la distribución normal univariada es que las combinaciones lineales de variables normales independientes son normales. La generalización a la situación multivariada, es como se muestra a continuación.

Teorema 1.2.7. Si X_1, \dots, X_N son todas independientes, y X_i es $N_m(\mu_i, \Sigma_i)$ para $i = 1, \dots, N$, entonces para cualquier constante fija $\alpha_1, \dots, \alpha_N$,

$$\sum_{i=1}^N \alpha_i X_i \quad \text{es} \quad N_m \left(\sum_{i=1}^N \alpha_i \mu_i, \sum_{i=1}^N \alpha_i^2 \Sigma_i \right).$$

La demostración de este teorema se sigue de la definición 1.2.2, o por inspección de la función característica de $\sum_{i=1}^N \alpha_i X_i$.

Teorema 1.2.8. Si $X \sim N(\mu, \Sigma)$, $X \in R^{k \times 1}$, con $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, (X_1, X_2 son conjuntamente normales) $X_2 | X_1 \sim N(\mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})$.

Demostración.

Sea $A = \begin{pmatrix} I & 0 \\ -\Sigma_{21} \Sigma_{11}^{-1} & I \end{pmatrix}$, es claro que $|A| = I$, si

$$W = A(X - \mu) = \begin{pmatrix} I & 0 \\ -\Sigma_{21} \Sigma_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix} = \begin{pmatrix} X_1 - \mu_1 \\ -\Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1) + (X_2 - \mu_2) \end{pmatrix}$$

Luego

$$W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \end{pmatrix} \right)$$

$$f_{W_1 W_2}(W_1, W_2) =$$

$$\frac{1}{(2\pi)^{k_1/2} |\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(W_1^t \Sigma_{11}^{-1} W_1)} \frac{1}{(2\pi)^{k_2/2} |\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}|^{1/2}} e^{-\frac{1}{2}(W_2^t (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1} W_2)}$$

$$f_{X_1, X_2}(X_1, X_2) = f_{W_1, W_2}(A(X - \mu)) |I|$$

$$= f_{W_1, W_2}(X_1 - \mu_1, (X_2 - \mu_2) - \Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1))$$

$$= \frac{1}{(2\pi)^{k_1/2} |\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(W_1^t \Sigma_{11}^{-1} W_1)} \frac{1}{(2\pi)^{k_2/2} |\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}|^{1/2}} e^{-\frac{1}{2}(W_2^t (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1} W_2)}$$

$$f_{X_2 | X_1}(X_2 | X_1) =$$

$$\frac{1}{(2\pi)^{k_2/2} |\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}} e^{-\frac{1}{2}\left((X_2 - \mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)\right)^t (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \left((X_2 - \mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)\right)}$$

Como podemos ver $X_2 | X_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.

□

Definición 1.2.3. A $Q = (X - \mu)^t \Sigma^{-1}(X - \mu)$ se le denomina la forma cuadrática asociada a la función de densidad $f_X(X) = K e^{-\frac{1}{2}(X - \mu_X)^t \Sigma^{-1}(X - \mu_X)}$

Teorema 1.2.9. Si $X \sim N(\mu, \Sigma)$, $X \in \mathbb{R}^{k \times 1}$ con forma cuadrática Q , el vector de medias μ es aquel que da la solución al sistema de ecuaciones $\frac{\partial Q}{\partial X} = 0$.

Demostración.

La densidad la podemos poner como $f_X(X) = K e^{-\frac{1}{2}Q}$ el valor de X que maximiza $f_X(X)$ es el valor de X tal que $Q=0$, como $Q=(X - \mu)^t \Sigma^{-1}(X - \mu)$ y es definida positiva, Q sólo puede ser cero en el punto $X - \mu = 0$, es decir el punto en donde $X = \mu$, por lo que podemos decir que μ es el punto que maximiza a $f_X(X)$.

La solución al sistema $\frac{\partial f_X(X)}{\partial X} = 0$ da el punto máximo de $f_X(X)$ y por lo mencionado anteriormente tal punto es μ . Pero como

$$\frac{\partial f_X(X)}{\partial X} = K e^{-\frac{1}{2}Q} \left(-\frac{1}{2} \right) \frac{\partial Q}{\partial X}$$

Se sigue que el vector que satisface $\frac{\partial f_X(X)}{\partial X} = 0$, es el mismo que satisface $\frac{\partial Q}{\partial X} = 0$. Por

tanto el vector μ es la solución al sistema $\frac{\partial Q}{\partial X} = 0$

□

Ejemplo. Sea x_1, x_2 , dos variables distribuidos conjuntamente como una normal multivariada con forma cuadrática

$$Q = x_1^2 + 2x_2^2 - x_1x_2 - 3x_1 - 2x_2 + 4$$

A partir de la forma cuadrática deseamos encontrar el vector de medias y la matriz de varianzas covarianzas. La forma cuadrática la podemos escribir como

$$Q = (X - \mu)' \Sigma^{-1} (X - \mu) = X' \Sigma^{-1} X - \mu' \Sigma^{-1} X - X' \Sigma^{-1} \mu + \mu' \Sigma^{-1} \mu$$

El único término que tiene términos de segundo grado es $X' \Sigma^{-1} X$, así tenemos que

$$X' \Sigma^{-1} X = x_1^2 + 2x_2^2 - x_1x_2 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}' \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\text{así tenemos que } \Sigma^{-1} = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 2 \end{pmatrix} \text{ y } \Sigma = \begin{pmatrix} \frac{8}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{4}{7} \end{pmatrix}$$

ahora sólo nos hace falta obtener la media, por el teorema anterior esta la podemos obtener resolviendo el sistema $\frac{\partial Q}{\partial X} = 0$.

$$\text{Tenemos que } \frac{\partial Q}{\partial x_1} = 2x_1 - x_2 - 3 = 0$$

$$\frac{\partial Q}{\partial x_2} = -x_1 + 4x_2 - 2 = 0$$

La solución al sistema anterior es $x_2 = 1$, $x_1 = 2$ así tenemos que $\mu_1 = 2$ y $\mu_2 = 1$.

1.3 Distribución de formas cuadráticas.

Recordemos que una función de densidad es una mezcla de funciones si:

$$f_X(x) = \sum_{j=1}^n p_j f_j(x)$$

con p_j es tal que $p_j \geq 0$ y $\sum_{j=1}^n p_j = 1$ y $f_j(x)$ son funciones de densidad, $j = 1, \dots, n$.

Definición 1.3.1 La función de densidad

$$f_X(x; \lambda, p) = \sum_{j=0}^{\infty} p_j f_{p+2j}(x)$$

donde $f_{p+2j}(x)$ es la función de densidad de una $\chi^2_{(p+2j)}$ y $p_j = e^{-\lambda} \frac{\lambda^j}{j!}, j = 0, 1, 2, \dots$

a la función de densidad anterior se le denomina χ^2 no central con parámetro de no centralidad λ y p grados de libertad ($\chi^2_{(p, \lambda)}$).

Teorema 1.3.1 El valor esperado y la varianza de la variable definida anteriormente son:

$$\begin{aligned} E(X) &= p+2\lambda \\ V(X) &= 2(p+4\lambda) \end{aligned}$$

Demostración.

$$E(X) = \int_0^{\infty} x \sum_{j=0}^{\infty} p_j f_{p+2j}(x) dx = \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} (p+2j) = p+2\lambda$$

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 \sum_{j=0}^{\infty} p_j f_{p+2j}(x) dx \\ &= \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \left(2(p+2j) + (p+2j)^2 \right) \\ &= \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \left(2p + p^2 + (4+4p)j + 4j^2 \right) \\ &= 2p + p^2 + (4+4p)\lambda + 4(\lambda + \lambda^2) \\ &= 4\lambda^2 + 8\lambda + 4p\lambda + 2p + p^2 \end{aligned}$$

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 \\ &= 4\lambda^2 + 8\lambda + 4p\lambda + 2p + p^2 - (p+2\lambda)^2 \\ &= 2(p+4\lambda) \end{aligned}$$

□

Teorema 1.3.2 La función característica de una $\chi^2_{(p, \lambda)}$ es;

$$\varphi_X(t) = (1 - 2it)^{-\frac{p}{2}} e^{\lambda \left(\frac{1}{1-2it} - 1 \right)}$$

Demostración.

$$\begin{aligned}
\varphi_X(t) &= \int_0^{\infty} e^{itx} \sum_{j=0}^{\infty} p_j f_{p+2j}(x) dx \\
&= \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \int_0^{\infty} e^{itx} f_{p+2j}(x) dx \\
&= \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \left(\frac{1}{1-2it} \right)^{\frac{p+2j}{2}} \\
&= \left(\frac{1}{1-2it} \right)^{\frac{p}{2}} e^{-\lambda} \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\lambda}{1-2it} \right)^j \\
&= (1-2it)^{-\frac{p}{2}} e^{\lambda \left(\frac{1}{1-2it} - 1 \right)}
\end{aligned}$$

□

Corolario 1.3.2.1. Si X se distribuye como una $\chi_{(p,\lambda)}^2$, si $\lambda = 0$ entonces X es una $\chi_{(p)}^2$.

Definición 1.3.2. Si $W \sim \chi_{(p,\lambda)}^2$ y $U \sim \chi_{(q,0)}^2$ son dos variables independientes, la variable

$$V = \frac{W/p}{U/q},$$

tiene función de densidad F no central con parámetro de no centralidad λ y p y q grados de libertad en el numerador y denominador respectivamente ($F_{(p,q,\lambda)}$).

Teorema 1.3.3. Si $X \sim N_p(\mu, I)$ entonces $X^t X \sim \chi_{(p,\lambda)}^2$, con $\lambda = \frac{1}{2} \mu^t \mu$.

Demostración.

$$\begin{aligned}
\varphi_{X^t X}(t) &= E\left(e^{itX^t X}\right) = E\left(\prod_{j=1}^p e^{itx_j}\right) \\
&= \prod_{j=1}^p \int e^{itx_j} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_j - \mu_j)^2} dx_j \\
&= \prod_{j=1}^p \int \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2} \left\{ \frac{[x_j(1-2it) - \mu_j]^2}{1-2it} + \mu_j^2 \left[1 - \frac{1}{1-2it} \right] \right\}} dx_j \\
&= e^{-\frac{1}{2} \sum_{j=1}^p \mu_j^2 \left[1 - \frac{1}{1-2it} \right]} \prod_{j=1}^p \int \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(1-2it) \left\{ \left[x_j - \frac{\mu_j}{1-2it} \right]^2 \right\}} dx_j
\end{aligned}$$

$$\begin{aligned}
&= e^{-\frac{1}{2} \sum_{j=1}^p \mu_j^2 \left[1 - \frac{1}{1-2it} \right]} \prod_{j=1}^p \left(\frac{1}{1-2it} \right)^{\frac{1}{2}} \\
&= (1-2it)^{-\frac{p}{2}} e^{\lambda \left(\frac{1}{1-2it} - 1 \right)}
\end{aligned}$$

Si tomamos como $\lambda = \frac{1}{2} \mu^t \mu$, hemos llegado a la función característica de una $\chi_{(p,\lambda)}^2$, con $\lambda = \frac{1}{2} \mu^t \mu$

□

Corolario 1.3.3.1 Si $X \sim N_p(\mathbf{0}, I)$ entonces $X^t X \sim \chi_{(p)}^2$.

Teorema 1.3.4. Si $X \sim N_n(\mu, \Sigma)$ entonces $X^t A X \sim \chi_{(p,\lambda)}^2$, con $\text{rango}(A) = p$ y $\lambda = \frac{1}{2} \mu^t A \mu$, si, y sólo si, $A \Sigma$ es idempotente.

Demostración.

La función característica de $X^t A X$ es

$$\begin{aligned}
\varphi_{X^t A X}(v) &= \int \cdots \int e^{iv X^t A X} \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2} (X-\mu)^t \Sigma^{-1} (X-\mu)} dX \\
&= \frac{e^{-\frac{1}{2} (\mu^t \Sigma^{-1} \mu)}}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \int \cdots \int e^{-\frac{1}{2} (X^t (I-2ivA\Sigma) \Sigma^{-1} X) + \mu^t \Sigma^{-1} X} dX \\
&= e^{-\frac{1}{2} (\mu^t \Sigma^{-1} \mu)} |\Sigma|^{-\frac{1}{2}} \left| \Sigma (I-2ivA\Sigma)^{-1} \right|^{\frac{1}{2}} e^{\frac{1}{2} \mu^t \Sigma^{-1} \Sigma (I-2ivA\Sigma)^{-1} \Sigma^{-1} \mu} \\
&= \left| (I-2ivA\Sigma) \right|^{-\frac{1}{2}} e^{-\frac{1}{2} \mu^t (I-(I-2ivA\Sigma)^{-1}) \Sigma^{-1} \mu} \\
&= \prod_{j=1}^n (1-2iv\lambda_j)^{-\frac{1}{2}} e^{-\frac{1}{2} \mu^t \left(-\sum_{k=1}^{\infty} (2iv)^k (A\Sigma)^k \right) \Sigma^{-1} \mu}
\end{aligned}$$

donde λ_j son las raíces características de $A\Sigma$, con $A\Sigma$ es idempotente de rango p , p valores de λ_j son uno y $n-p$ son ceros y como $(A\Sigma)^p = A\Sigma$

$$\begin{aligned}
\varphi_{X^t A X}(v) &= \prod_{j=1}^p (1-2iv)^{-\frac{1}{2}} e^{-\frac{1}{2} \mu^t \left[-\sum_{k=1}^{\infty} (2iv)^k \right] A\Sigma \Sigma^{-1} \mu} \\
&= (1-2iv)^{-\frac{1}{2} p} e^{-\frac{1}{2} \mu^t A \mu (1-(1-2iv)^{-1})}
\end{aligned}$$

luego entonces $X^tAX \sim \chi_{(p,\lambda)}^2$, $\lambda = \frac{1}{2}\mu^t A\mu$.

Ahora suponiendo que $X^tAX \sim \chi_{(p,\lambda)}^2$, hay que probar que $A\Sigma$ es idempotente con $\text{rango}(A\Sigma) = p$.

Tenemos que

$$\phi_{X^tAX}(v) = \left| (I - 2ivA\Sigma) \right|^{-\frac{1}{2}} e^{-\frac{1}{2}\mu^t (I - 2ivA\Sigma)^{-1} \Sigma^{-1} \mu} = (1 - 2iv)^{-\frac{1}{2}p} e^{-\frac{1}{2}\mu^t A\mu (1 - 2iv)^{-1}}$$

Lo anterior es para cualquier μ , sin perdida de generalidad tomemos a $\mu = 0$, substituyendo $\mu = \mathbf{0}$ en la igualdad anterior

$$(1 - 2iv)^{-\frac{1}{2}p} = \left| (I - 2ivA\Sigma) \right|^{-\frac{1}{2}}$$

haciendo $u = 2iv$

$$(1 - u)^p = \left| (I - uA\Sigma) \right|$$

sean $\lambda_1, \lambda_2, \dots, \lambda_n$ las raíces características de $A\Sigma$, entonces tenemos que

$$(1 - u)^p = \prod_{j=1}^n (1 - u\lambda_j)$$

para que la igualdad se cumpla debemos tener $n-p$ raíces características iguales a cero, luego entonces, nos queda que

$$(1 - u)^p = \prod_{j=1}^p (1 - u\lambda_j)$$

sacando logaritmo de ambos lados tenemos que los λ_j son iguales a uno, luego entonces $A\Sigma$ es una matriz idempotente con $\text{rango}(A\Sigma) = p$.

□

Corolario 1.3.4.1 Si $X \sim N_n(\mathbf{0}, I)$ entonces $X^tAX \sim \chi_{(k)}^2$, con $\text{rango}(A) = k$, si y sólo si A es idempotente.

Corolario 1.3.4.2 Si $X \sim N_n(\mu, I)$ entonces $X^tAX \sim \chi_{(p,\lambda)}^2$, con $\text{rango}(A) = p$ y $\lambda = \frac{1}{2}\mu^t A\mu$, si, y sólo si, A es idempotente.

Teorema 1.3.5. Si $X \sim N(\mu, \Sigma)$ entonces X^tAX y BX son independientes, si, y sólo si $B\Sigma A = 0$.

Demostración.

Por ser A simétrica $A=LL^t$, donde L es de rango completo .

Por lo que

$$\begin{aligned}\text{Cov}(BX, L^t X) &= E(BX-B\mu)(L^t X - L^t \mu)^t \\ &= E(B(X-\mu)(X-\mu)^t L) \\ &= B\Sigma L\end{aligned}$$

Pero como tenemos que $B\Sigma A = 0$ tenemos que $B\Sigma LL^t=0$, como L es de rango completo $L^t L$ tiene inversa y

$$B\Sigma LL^t=0 \text{ implica } B\Sigma L=0$$

Por lo que BX es independiente de $L^t X$. Entonces BX es independiente de $X^t LL^t X = X^t AX$.

Ahora demostremos que si $X^t AX$ y BX son independientes entonces $B\Sigma A = 0$. Como $X^t AX$ y BX son independientes tenemos que $\text{Cov}(BX, X^t AX) = 0$

$$\begin{aligned}\text{Cov}(BX, X^t AX) &= E(BX-B\mu)(X^t AX - E(X^t AX))^t \\ &= BE(X-\mu)(X^t AX - p - \mu^t A\mu) \\ &= BE(X-\mu)[(X-\mu)^t A(X-\mu) + 2(X-\mu)^t A\mu - p] \\ &= 2B\Sigma A\mu\end{aligned}$$

lo que significa que $B\Sigma A = 0$.

□

Teorema 1.3.6. Si $X \sim N(\mu, \Sigma)$, A y B simétricas, $X^t AX$ y $X^t BX$ son independientes si y sólo, $A\Sigma B=0$ (o equivalentemente $B\Sigma A=0$)
(Considerando que, $A\Sigma$ y $B\Sigma$ son idempotentes y el rango(A)= q_1 y rango(B)= q_2)

Demostración.

Si $X^t AX$ y $X^t BX$ son independientes, tenemos que $X^t AX + X^t BX = X^t (A+B)X$ se distribuye como una $\chi^2_{(p, \lambda)}$, con $p=q_1+q_2$ y $\lambda = \frac{1}{2}\mu^t (A+B)\mu$. Tenemos que $(A+B)\Sigma$ debe ser idempotentes, luego entonces $A\Sigma B=0$.

Ahora demostremos que si $A\Sigma B=0$ entonces $X^t AX$ y $X^t BX$ son independientes,

$$\begin{aligned}((A+B)\Sigma)((A+B)\Sigma) &= A\Sigma A\Sigma + A\Sigma B\Sigma + B\Sigma A\Sigma + B\Sigma B\Sigma \\ &= A\Sigma + A\Sigma B\Sigma + B\Sigma A\Sigma + B\Sigma \\ &= (A+B)\Sigma\end{aligned}$$

Lo que significa que $X^t(A+B)X$ se distribuye como una $\chi^2_{(p,\lambda)}$ con p y λ definidas como arriba, luego entonces $X^t AX$ y $X^t BX$ son independientes. \square

La demostración del siguiente teorema se sigue a partir del teorema 1.1.1 y los resultados de esta sección.

Teorema 1.3.5. (Teorema de Cochran) Sea $Y \sim N(\mu, \sigma^2 I)$ y A_i son matrices de $n \times n$ simétricas de rango $k_i, i = 1, \dots, p$.

y $A = \sum_{i=1}^p A_i$, simétrica de rango k , entonces las $\frac{1}{\sigma^2} Y' A_j Y \sim \chi^2_{(p,\lambda)}$ con $p = r_j$ y $\lambda = \frac{1}{2\sigma^2} \mu' A_j \mu$, serán independientes y $\frac{1}{\sigma^2} Y' A Y \sim \chi^2_{(p,\lambda)}$ con $p = r$ y $\lambda = \frac{1}{2\sigma^2} \mu' A \mu$ si y sólo si

I) Si se cumplen alguna de las dos a), b) y c) implica a), b), c) y d).

II) c) y d) \Rightarrow a) y b)

III) c) y A_1, A_2, \dots, A_{p-1} idempotentes con A_p no negativa implica A_p idempotente y por tanto implica a) y por consiguiente b) y d).

a) A_i es idempotente para todo i

b) $A_i A_j = 0$ para todo $i \neq j$

c) A es idempotente con $k = \sum_{i=1}^p k_i$.

II.- Modelo de Regresión Lineal.

2.1 Estimación de los parámetros.

Para el modelo lineal general con n observaciones y p coeficientes, en forma vectorial tenemos

$$Y = X\beta + e$$

con

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p-11} \\ 1 & x_{12} & x_{22} & \cdots & x_{p-12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{p-1n} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix},$$

Donde Y , X son conocidas y β desconocida, además $e \sim N(0, \sigma^2 I)$ y X de rango completo,

$$\begin{aligned} f(e; 0, \sigma^2 I) &= \frac{1}{(2\pi\sigma^2)^{p/2}} \exp \left\{ -\frac{1}{2} (e)' \left(\frac{1}{\sigma^2} I \right) (e) \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{p/2}} \exp \left\{ -\frac{1}{2} (Y - X\beta)' \left(\frac{1}{\sigma^2} I \right) (Y - X\beta) \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{p/2}} \exp \left\{ -\frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta) \right\} \end{aligned}$$

como los β 's son desconocidos, busquemos sus valores estimando por máxima verosimilitud, la función de verosimilitud es;

$$\begin{aligned} L(\sigma^2, \beta; e) &= \frac{1}{(2\pi\sigma^2)^{p/2}} \exp \left\{ -\frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta) \right\} \\ \ln(L(\sigma^2, \beta; e)) &= -\frac{p}{2} (\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta) \end{aligned}$$

dado que $\beta^t X^t Y = Y^t X \beta$

$$= -\frac{p}{2} (\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} (Y^t Y - 2\beta^t X^t Y + \beta^t X^t X \beta)$$

$$\frac{\partial \text{Ln}(L(\cdot))}{\partial \beta} = -\frac{1}{2\sigma^2} (-2X^t Y + 2X^t X \beta)$$

$$\frac{\partial \text{Ln}(L(\cdot))}{\partial \sigma^2} = -\frac{p}{2\sigma^2} + \frac{1}{2\sigma^4} (Y^t Y - 2Y^t X \beta + \beta^t X^t X \beta)$$

$$= -\frac{p}{2\sigma^2} + \frac{1}{2\sigma^4} (Y - X\beta)^t (Y - X\beta)$$

igualando a cero ambas ecuaciones,

$$-\frac{1}{2\sigma^2} (-2X^t Y + 2X^t X \beta) = 0$$

$$\Rightarrow -X^t Y + X^t X \beta = 0$$

$$\Rightarrow \hat{\beta} = (X^t X)^{-1} X^t Y$$

$$-\frac{p}{2\sigma^2} + \frac{1}{2\sigma^4} (Y - X\beta)^t (Y - X\beta) = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{(Y - X\hat{\beta})^t (Y - X\hat{\beta})}{p}$$

Ahora obtengamos la esperanza de los estimadores.

$$E(\hat{\beta}) = E\left((X^t X)^{-1} X^t Y\right) \text{ haciendo } S = X^t X \text{ y sustituyendo } Y = X\beta + e$$

$$E(\hat{\beta}) = E(S^{-1} X^t (X\beta + e)) = E(S^{-1} X^t X \beta + S^{-1} X^t e) = \beta + S^{-1} X^t E(e) = \beta$$

es decir, $\hat{\beta}$ es un estimador insesgado de β , ahora calculemos la esperanza de $\hat{\sigma}^2$

$$E(\hat{\sigma}^2) = E\left(\frac{(Y - X\hat{\beta})^t (Y - X\hat{\beta})}{p}\right)$$

sustituyendo el valor de $\hat{\beta}$

$$= \frac{1}{p} E\left((Y - X S^{-1} X^t Y)^t (Y - X S^{-1} X^t Y)\right)$$

$$= \frac{1}{p} E\left(Y^t (I - X S^{-1} X^t)^t (I - X S^{-1} X^t) Y\right)$$

Podemos ver que $I - XS^{-1}X^t$ es idempotente

$$= \frac{1}{p} E(Y^t (I - XS^{-1}X^t) Y)$$

Sustituyendo $Y = X\beta + e$

$$\begin{aligned} &= \frac{1}{p} E((X\beta + e)^t (I - XS^{-1}X^t) (X\beta + e)) \\ &= \frac{1}{p} E(\beta^t X^t X\beta - \beta^t X^t XS^{-1}X^t X\beta + \beta^t X^t (I - XS^{-1}X^t) e + e^t (I - XS^{-1}X^t) X\beta + e^t (I - XS^{-1}X^t) e) \end{aligned}$$

Como $E(e) = 0$

$$E(\hat{\sigma}^2) = \frac{1}{p} E(e^t (I - XS^{-1}X^t) e)$$

Por lo demostrado en la tarea 4 (Si X es el vector $p \times 1$ con $X \sim N(0, \sigma^2 I)$ y A es idempotente de rango k entonces $E(X^t AX) = k\sigma^2$.)

$$= \frac{1}{p} \sigma^2 (n - p)$$

Luego entonces para poder tener un estimador insesgado de la varianza debemos considerar a

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

Ahora obtengamos la matriz de varianza covarianza de $\hat{\beta}$

$$\begin{aligned} V(\hat{\beta}) &= E(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))^t \\ &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t \end{aligned}$$

sustituyendo el valor de Y y de $\hat{\beta}$

$$\begin{aligned} &= E(S^{-1}X^t (X\beta + e) - \beta)(S^{-1}X^t (X\beta + e) - \beta)^t \\ &= E(\beta + S^{-1}X^t e - \beta)(\beta + S^{-1}X^t e - \beta)^t \end{aligned}$$

es claro que $S^{-1} = (S^{-1})^t$

$$\begin{aligned} &= E(S^{-1}X^t e e^t X S^{-1}) \\ &= S^{-1}X^t E(e e^t) X S^{-1} \\ &= \sigma^2 S^{-1}X^t X S^{-1} = \sigma^2 S^{-1} \end{aligned}$$

de aquí podemos deducir que $\hat{\beta} \sim N(\beta, \sigma^2 S^{-1})$.

2.2. Mínimos Cuadrados y Proyección Ortogonal.

En esta sección veremos la relación que existe entre el método de estimación de mínimos cuadrados empleado en el análisis de regresión con la mejor aproximación a un vector, empecemos con un ejemplo antes de empezar con el formalismo del tema.

Ejemplo. Un investigador recolecta información mediante la realización de mediciones y_1, y_2, \dots, y_m en los instantes t_1, t_2, \dots, t_m , respectivamente. Por ejemplo, pueden realizarse mediciones sobre el desempleo en distintas fechas durante un período. Supóngase que se grafican los datos $(t_1, y_1), \dots, (t_m, y_m)$ como puntos del plano. A causa de la distribución de tales puntos, él piensa que existe una correlación lineal entre y y t , tal como $y=ct+d$. El investigador está interesado en encontrar los parámetros c y d de tal manera que la recta $y=ct+d$ represente el mejor “ajuste” posible para los datos recopilados. Una estimación del ajuste es calcular el error e que representa la suma de cuadrados de las distancias verticales de los puntos a la recta, esto es,

$$e = \sum_{i=1}^m (y_i - ct_i - d)^2$$

Así, su problema es encontrar las constantes c y d que minimicen a e . Esto lo conduce a considerar el siguiente sistema de ecuaciones :

$$\begin{aligned} t_1c + d &= y_1 \\ t_2c + d &= y_2 \\ &\vdots \\ t_mc + d &= y_m \end{aligned}$$

$$\text{o bien } AX = y, \text{ donde } A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{pmatrix}, X = \begin{pmatrix} c \\ d \end{pmatrix} \text{ y } y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

Por supuesto que parecería irreal suponer que tal sistema tiene una solución puesto que en la práctica, el número de ecuaciones excede con mucho al número de incógnitas, para resolver este tipo de problema veamos los siguientes resultados.

Para x, y en F^n , denotaremos por $\langle x, y \rangle_n$ al producto interno canónico (ordinario) de x, y en F^n .

Lema 2.2.1. Sean A una matriz de $m \times n$ sobre F , x está en F^n , y en F^m . Entonces

$$\langle Ax, y \rangle_m = \langle x, A^*y \rangle_n$$

Demostración.

Sea $Ax = c$ donde $c \in F^m$ y c está definida por $c_i = \sum_{j=1}^n A_{ij}x_j$ con $i = 1, \dots, m$, por tanto;

$$\begin{aligned} \langle Ax, y \rangle &= c_1y_1 + \dots + c_my_m = y_1 \left(\sum_{j=1}^n A_{1j}x_j \right) + \dots + y_m \left(\sum_{j=1}^n A_{mj}x_j \right) \\ &= \sum_{i=1}^m y_i \left(\sum_{j=1}^n A_{ij}x_j \right) = \sum_{i=1}^m \sum_{j=1}^n A_{ij}x_j y_i \end{aligned}$$

Por otro lado $A^*y = d$ donde d está en F^n y d está definida por $d_j = \sum_{i=1}^m A_{ji}^*y_i = \sum_{i=1}^m A_{ij}y_i$ con $j = 1, \dots, n$. El lado derecho de la igualdad queda como:

$$\begin{aligned} \langle x, A^*y \rangle &= d_1x_1 + \dots + d_nx_n = x_1 \left(\sum_{i=1}^m A_{i1}y_i \right) + \dots + x_n \left(\sum_{i=1}^m A_{in}y_i \right) \\ &= \sum_{j=1}^n x_j \left(\sum_{i=1}^m A_{ij}y_i \right) = \sum_{j=1}^n \sum_{i=1}^m A_{ij}y_i x_j \\ &= \sum_{i=1}^m \sum_{j=1}^n A_{ij}x_j y_i = \langle Ax, y \rangle \end{aligned}$$

□

Lema 2.2.2. Sea A una matriz de $m \times n$ sobre F . Entonces el rango(A^*A) = rango(A).

Demostración.

Sólo tenemos que demostrar que, para x en F^n , $A^*Ax = 0$ si y sólo si $Ax = 0$. Claramente si $Ax = 0$ implica que $A^*Ax = 0$. Por ello supóngase que $A^*Ax = 0$.

Entonces $0 = \langle A^*Ax, x \rangle_n = \langle Ax, A^*x \rangle_m = \langle Ax, Ax \rangle_m$ de modo que $Ax = 0$.

□

Corolario 2.2.1. Si A es una matriz de $m \times n$ tal que el $\text{rango}(A) = n$, entonces A^*A es invertible.

Demostración.

Como A es de rango n entonces por el lema anterior A^*A es de rango n pero como A^*A es una matriz de $n \times n$ entonces A^*A es invertible.

□

Teorema 2.2.1. Sea A una matriz de $m \times n$ sobre F , y en F^m . Entonces, existe x_0 en F^n tal que

$$(A^*A)x_0 = A^*y$$

y

$$\|Ax_0 - y\| \leq \|Ax - y\| \text{ para toda } x \in F^n.$$

Además, si $\text{rango}(A) = n$, entonces $x_0 = (A^*A)^{-1}A^*y$.

Demostración.

Considérese a $Ax = y$, definimos a $W = \{Ax : x \in F^n\}$. Haciendo a \mathbf{E} la proyección ortogonal sobre W , escójase a x_0 en F^n tal que $\mathbf{E}(y) = Ax_0$. Entonces por $\|\mathbf{E}(y) - y\| \leq \|u - y\|$ para toda u en W ; esto es, $\|Ax_0 - y\| \leq \|Ax - y\|$ para toda x en F^n .

Observemos que como \mathbf{E} es una proyección ortogonal, $Ax_0 - y = \mathbf{E}(y) - y$ están en W^\perp , entonces $\langle Ax, Ax_0 - y \rangle_m = 0$ para toda x en F^n . Luego por el *Lema 1*, tenemos que $\langle x, A^*(Ax_0 - y) \rangle_n = 0$ para toda x en F^n ; esto es, $A^*(Ax_0 - y) = 0$. Así únicamente tenemos que encontrar una solución para $A^*Ax_0 = A^*y$. Si además suponemos que $\text{rango}(A) = n$, entonces por *Lema 2* tenemos que $x_0 = (A^*A)^{-1}A^*y$.

□

El teorema anterior se mostró que, si $\text{rango}(A) = n$, entonces existe un elemento $x_0 \in F^n$ único tal que Ax_0 , es el punto en W más cercano a y . Por supuesto, si $\text{rango}(A) < n$, existirá un número infinito de estos vectores.

Esta forma de aproximación es mejor conocida como *mínimos cuadrados* es muy útil para resolver sistemas de ecuaciones cuyos problemas envuelvan en que el número de ecuaciones exceda del número de incógnitas, muy utilizado esto en la área de estadística.

Regresando a nuestro ejemplo anterior, pero ahora con un ejemplo numérico. Supongamos que el investigador recopiló los siguientes datos: (1,2), (2,3), (3,5) y (4,7). Entonces

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \quad \text{y} \quad y = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix};$$

por lo tanto

$$A^*A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}$$

y entonces

$$(A^*A)^{-1} = \frac{1}{20} \begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix}$$

Por consiguiente

$$x_0 = \begin{pmatrix} c \\ d \end{pmatrix} = \frac{1}{20} \begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 1.7 \\ 0 \end{pmatrix}.$$

Así, la recta $y=1.7t$ es la recta de mínimos cuadrados. El error \mathbf{E} puede calcularse directamente como $\|Ax_0 - y\|^2 = 0.3$

Pero que pasa si el rango de A es menor a n , este problema se resuelve por medio de *matrices pseudoinvertas*.

2.3 Intervalos de Confianza.

Para el modelo lineal general con n observaciones y p coeficientes a estimar, que en forma vectorial tenemos

$$Y = X\beta + e$$

con

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p-11} \\ 1 & x_{12} & x_{22} & \cdots & x_{p-12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{p-1n} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix},$$

Donde Y , X son conocidas y β desconocida, además $e \sim N(0, \sigma^2 I)$ y X de rango completo,

Recordemos que, los estimadores máximo verosimiles del modelo son;

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

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

en el caso de la varianza utilizaremos el insesgado.

De las ecuaciones anteriores hemos deducido que $\hat{\beta} \sim N(\beta, \sigma^2 S^{-1})$ donde $S = X^t X$, y también tenemos que $(n - p) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p)}^2$, ahora obtengamos algunos intervalos de confianza.

2.3.1 Intervalos de confianza para β_i

La marginal del i -ésimo elemento de $\hat{\beta}$ es, $\hat{\beta}_i \sim N(\beta_i, \sigma^2 c_{ii})$, donde $c_{ii} = (S^{-1})_{ii}$, entonces

$$Z = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 c_{ii}}} \sim N(0, 1)$$

y

$$T = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 c_{ii}}} \sim t_{(n-p)}$$

Si buscamos el valor de k tal que $P\left(-k \leq \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 c_{ii}}} \leq k\right) = 1 - \alpha$, éste es, $k = t_{\frac{\alpha}{2}}(n - p)$

$$P\left(\hat{\beta}_i - t_{1-\frac{\alpha}{2}}(n-p)\sqrt{\hat{\sigma}^2 c_{ii}} \leq \beta_i \leq \hat{\beta}_i + t_{1-\frac{\alpha}{2}}(n-p)\sqrt{\hat{\sigma}^2 c_{ii}}\right) = 1 - \alpha$$

lo que acabamos de obtener, es el intervalo de confianza para β_i con el $(1 - \alpha)100\%$ de confianza.

2.3.2 Intervalo de confianza para σ^2 .

Para obtener el intervalo de confianza de σ^2 usemos la pivotal $(n-p)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p)}^2$,

$$P\left(k_1 \leq (n-p)\frac{\hat{\sigma}^2}{\sigma^2} \leq k_2\right) = 1 - \alpha$$

$$P\left((n-p)\frac{\hat{\sigma}^2}{k_2} \leq \sigma^2 \leq (n-p)\frac{\hat{\sigma}^2}{k_1}\right) = 1 - \alpha$$

$\sigma^2 \in \left(\frac{n-p}{k_2} \hat{\sigma}^2, \frac{n-p}{k_1} \hat{\sigma}^2\right)$ con una confianza del $(1 - \alpha)100\%$, con $k_1 = \chi_{\alpha/2}^2(n-p)$ y

$$k_2 = \chi_{1-\alpha/2}^2(n-p)$$

2.3.3 Intervalo de confianza para $E(y)$ ($E(y | x_1, x_2, \dots, x_{p-1})$).

$$\begin{aligned} E(y) &= x^* \beta \\ &= \sum_{i=0}^{p-1} \beta_i x_i \end{aligned}$$

(queremos el valor esperado de y dado que conocemos $x_0 = 1, x_1, x_2, \dots, x_{p-1}$).

$$x^* = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_{p-1} \end{pmatrix} \text{ es una observación de las variables independientes y } \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}.$$

El estimador de $E(y)$ es $\widehat{E}(y) = \mathbf{x}^{*t} \hat{\boldsymbol{\beta}} = \sum_{i=0}^{p-1} \hat{\beta}_i x_i$.

Sabemos que $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 \mathbf{S}^{-1})$ luego entonces $\mathbf{x}^{*t} \hat{\boldsymbol{\beta}} \sim N(\mathbf{x}^{*t} \boldsymbol{\beta}, \sigma^2 \mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*)$.

Tenemos que

$$Z = \frac{\mathbf{x}^{*t} \hat{\boldsymbol{\beta}} - \mathbf{x}^{*t} \boldsymbol{\beta}}{\sqrt{\sigma^2 \mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*}} \sim N(0,1)$$

y

$$T = \frac{\mathbf{x}^{*t} \hat{\boldsymbol{\beta}} - \mathbf{x}^{*t} \boldsymbol{\beta}}{\sqrt{\hat{\sigma}^2 (\mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*)}} \sim t_{(n-p)}$$

$$P\left(-k \leq \frac{\mathbf{x}^{*t} \hat{\boldsymbol{\beta}} - \mathbf{x}^{*t} \boldsymbol{\beta}}{\sqrt{\hat{\sigma}^2 (\mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*)}} \leq k\right) = 1 - \alpha$$

$$P\left(\mathbf{x}^{*t} \hat{\boldsymbol{\beta}} - t_{1-\frac{\alpha}{2}}(n-p) \sqrt{\hat{\sigma}^2 (\mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*)} \leq \mathbf{x}^{*t} \boldsymbol{\beta} \leq \mathbf{x}^{*t} \hat{\boldsymbol{\beta}} + t_{1-\frac{\alpha}{2}}(n-p) \sqrt{\hat{\sigma}^2 (\mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*)}\right) = 1 - \alpha$$

$\sum_{i=0}^{p-1} \beta_i x_i \in \left(\sum_{i=0}^{p-1} \hat{\beta}_i x_i - t_{1-\frac{\alpha}{2}}(n-p) \sqrt{\hat{\sigma}^2 (\mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*)}, \sum_{i=0}^{p-1} \hat{\beta}_i x_i + t_{1-\frac{\alpha}{2}}(n-p) \sqrt{\hat{\sigma}^2 (\mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*)}\right)$ es el

intervalo de confianza de $\sum_{i=0}^{p-1} \beta_i x_i$ con una confianza del $(1 - \alpha)100\%$.

Nota. No hay que olvidar que $\mathbf{x}_0 = \mathbf{1}$.

2.3.4 Intervalos de confianza para observaciones simultáneas.

Si tenemos $\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_k^*$ (vectores de $p \times 1$), si quisiéramos inferir simultáneamente para $\mathbf{x}_1^{*t} \boldsymbol{\beta}, \mathbf{x}_2^{*t} \boldsymbol{\beta}, \dots, \mathbf{x}_k^{*t} \boldsymbol{\beta}$ buscamos k -intervalos tales que, $\mathbf{x}_1^{*t} \boldsymbol{\beta} \in I_1, \mathbf{x}_2^{*t} \boldsymbol{\beta} \in I_2, \dots, \mathbf{x}_k^{*t} \boldsymbol{\beta} \in I_k$, si a cada uno de estos eventos los designamos por A_j , es decir, cada uno cumple con $P(A_1) = 1 - \alpha_1, P(A_2) = 1 - \alpha_2, \dots, P(A_k) = 1 - \alpha_k$ nosotros buscamos que el evento $A_1 \cap A_2 \cap \dots \cap A_k$ suceda con una probabilidad de por lo menos $1 - \alpha$

$$\begin{aligned}
P(A_1 \cap A_2 \cap \dots \cap A_k) &= 1 - P(A_1 \cap A_2 \cap \dots \cap A_k)^c \\
&= 1 - P(A_1^c \cup A_2^c \cup \dots \cup A_k^c) \\
&\geq 1 - (P(A_1^c) + P(A_2^c) + \dots + P(A_k^c)) \quad \text{si } P(A_k^c) = \frac{\alpha}{k} \\
&\geq 1 - \left(\frac{\alpha}{k} + \frac{\alpha}{k} + \dots + \frac{\alpha}{k}\right) = 1 - \alpha
\end{aligned}$$

Haciendo $P(A_1) = 1 - \frac{\alpha}{k}$, $P(A_2) = 1 - \frac{\alpha}{k}$, ..., $P(A_k) = 1 - \frac{\alpha}{k}$, podemos obtener lo que queremos. De aquí, podemos decir, que para obtener simultáneamente k -intervalos para $x_1^{*t} \beta$, $x_2^{*t} \beta$, ..., $x_k^{*t} \beta$ tenemos que calcular los k -intervalos

$$I_j = \left(x_j^{*t} \hat{\beta} - t_{1-\frac{\alpha}{2k}}(n-p) \sqrt{\hat{\sigma}^2 (x_j^{*t} S^{-1} x_j^*)}, x_j^{*t} \hat{\beta} + t_{1-\frac{\alpha}{2k}}(n-p) \sqrt{\hat{\sigma}^2 (x_j^{*t} S^{-1} x_j^*)} \right)$$

Así los k -intervalos ocurren simultáneamente con una confianza $\geq (1 - \alpha)100\%$, $j = 1, \dots, k$.

Si ahora quisiéramos un intervalo de confianza no sólo para algunos si no **para todo** x^* lo que buscamos es algo como;

$$P \left(\left| \frac{x^{*t} \hat{\beta} - x^{*t} \beta}{\sqrt{\hat{\sigma}^2 (x^{*t} S^{-1} x^*)}} \right| \leq c \right) = 1 - \alpha \quad \text{para todo } x^*$$

buscando es valor de c , tenemos

$$|x^{*t} \hat{\beta} - x^{*t} \beta| \leq c \sqrt{\hat{\sigma}^2 (x^{*t} S^{-1} x^*)} \quad \text{para todo } x^*$$

$$(x^{*t} \hat{\beta} - x^{*t} \beta)^2 \leq c^2 \hat{\sigma}^2 (x^{*t} S^{-1} x^*) \quad \text{para todo } x^*$$

$$0 \leq x^{*t} \left(c^2 \hat{\sigma}^2 S^{-1} - (\hat{\beta} - \beta)(\hat{\beta} - \beta)^t \right) x^* \quad \text{para todo } x^*$$

$$\text{Sea } W = \begin{pmatrix} A & u \\ u^t & cte \end{pmatrix}$$

$$|W| = \begin{cases} |A| |cte - u^t A^{-1} u| \\ |cte| \left| A - u \frac{1}{cte} u^t \right| \end{cases}$$

W es semidefinida positiva si, y sólo si $cte - u^t A^{-1} u \geq 0$ si, y sólo si $A - u \frac{1}{cte} u^t \geq 0$

Haciendo $\Lambda = S^{-1}$, $u = \hat{\beta} - \beta$ y $\text{cte} = c^2 \hat{\sigma}^2$, W es semidefinida positiva si, y sólo si $c^2 \hat{\sigma}^2 - (\hat{\beta} - \beta)^t S (\hat{\beta} - \beta) \geq 0$ si, y sólo si $S^{-1} - (\hat{\beta} - \beta) \frac{1}{c^2 \hat{\sigma}^2} (\hat{\beta} - \beta)^t \geq 0$

Sabemos que $\hat{\beta} \sim N(\beta, \sigma^2 S^{-1})$ luego entonces $\hat{\beta} - \beta \sim N(\mathbf{0}, \sigma^2 S^{-1})$

$$\Rightarrow (\mathbf{X}^t \mathbf{X})^{\frac{1}{2}} (\hat{\beta} - \beta) \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

$$\Rightarrow \frac{1}{\sigma} (\mathbf{X}^t \mathbf{X})^{\frac{1}{2}} (\hat{\beta} - \beta) \sim N(\mathbf{0}, \mathbf{I})$$

$$\Rightarrow \frac{1}{\sigma^2} (\hat{\beta} - \beta)^t (\mathbf{X}^t \mathbf{X}) (\hat{\beta} - \beta) \sim \chi_{(p)}^2$$

Como $(n-p) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p)}^2$, entonces $\frac{(\hat{\beta} - \beta)^t \mathbf{X}^t \mathbf{X} (\hat{\beta} - \beta) / p}{\hat{\sigma}^2} \sim F_{(p, n-p)}$, por tanto c

es el percentil de una $F_{(p, n-p)}$, el intervalo de confianza buscado es

$$x^{*t} \beta \in \left(x^{*t} \hat{\beta} - \sqrt{p F_{1-\alpha}^F(p, n-p)} \sqrt{\hat{\sigma}^2 (x^{*t} S^{-1} x^*)}, \right. \\ \left. x^{*t} \hat{\beta} + \sqrt{p F_{1-\alpha}^F(p, n-p)} \sqrt{\hat{\sigma}^2 (x^{*t} S^{-1} x^*)} \right) \text{ con}$$

una confianza del $(1 - \alpha)100\%$.

2.3.5 Intervalo de confianza para \bar{y}^F

Sean $y_1^F, y_2^F, \dots, y_s^F$ observaciones independientes de una población $N(x^{*t} \beta, \sigma^2)$ (e independientes de y_1, y_2, \dots, y_n) es decir, $y_1^F, y_2^F, \dots, y_s^F$ no están en la muestra original con la que estimamos los coeficientes del modelo, aquí

$$x^* = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_{p-1} \end{pmatrix}$$

es un vector de $p \times 1$. Si definimos a $\bar{y}^F = \frac{\sum_{i=1}^s y_i^F}{s} \sim N(x^{*t} \beta, \sigma^2/s)$, anteriormente vimos

que $x^{*t} \hat{\beta} \sim N(x^{*t} \beta, \sigma^2 x^{*t} S^{-1} x^*)$, luego entonces

$$\begin{aligned} \bar{y}^F - \mathbf{x}^{*t} \hat{\boldsymbol{\beta}} &\sim N\left(0, \sigma^2 \left(\frac{1}{s} + \mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*\right)\right) \\ \Rightarrow \frac{\bar{y}^F - \mathbf{x}^{*t} \hat{\boldsymbol{\beta}}}{\sqrt{\sigma^2 \left(\frac{1}{s} + \mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*\right)}} &\sim N(0,1) \\ \Rightarrow \frac{\bar{y}^F - \mathbf{x}^{*t} \hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{s} + \mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*\right)}} &\sim t_{(n-p)} \end{aligned}$$

luego entonces el intervalo de confianza para \bar{y}^F es;

$$\bar{y}^F \in \left(\mathbf{x}^{*t} \hat{\boldsymbol{\beta}} - t_{1-\frac{\alpha}{2}}(n-p) \sqrt{\hat{\sigma}^2 \left(\frac{1}{s} + \mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*\right)}, \mathbf{x}^{*t} \hat{\boldsymbol{\beta}} + t_{1-\frac{\alpha}{2}}(n-p) \sqrt{\hat{\sigma}^2 \left(\frac{1}{s} + \mathbf{x}^{*t} \mathbf{S}^{-1} \mathbf{x}^*\right)} \right)$$

es el intervalo de confianza de \bar{y}^F con una confianza del $(1 - \alpha)100\%$.

2.4 Pruebas de Hipótesis.

2.4.1 $H_0 : \boldsymbol{\beta} = \mathbf{0}$

Para el modelo lineal general con n observaciones y p coeficientes a estimar, que en forma vectorial tenemos

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

con

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p-11} \\ 1 & x_{12} & x_{22} & \cdots & x_{p-12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{p-1n} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix},$$

Donde \mathbf{Y} , \mathbf{X} son conocidas y $\boldsymbol{\beta}$ desconocida, además $\mathbf{e} \sim N_n(0, \sigma^2 \mathbf{I})$ y \mathbf{X} de rango completo,

Deseamos buscar la prueba para $H_0: \beta=0$.

Usando el cociente de verosimilitud generalizada, lo que buscamos, es lo siguiente, vamos a rechazar H_0 , si,

$$\lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta, \cdot)}{\sup_{\theta \in \Theta} L(\theta, \cdot)} < k, \text{ k una constante positiva}$$

donde $L(\theta, \cdot)$ es la función de verosimilitud y $\theta \in \Theta_0$ son los parámetros desconocidos bajo la hipótesis nula y $\theta \in \Theta$ son todos los parámetros desconocidos, la función de verosimilitud del modelo es;

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

para encontrar el denominador de λ , lo único que se debe de hacer es sustituir los estimadores máximos verosímiles del modelo en la función de verosimilitud, esto es:

$$\sup_{\theta \in \Theta} L(\theta, \cdot) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp^{-\frac{1}{2\hat{\sigma}^2}(\mathbf{Y}-\mathbf{X}\hat{\beta})^t(\mathbf{Y}-\mathbf{X}\hat{\beta})}$$

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

Ahora bien, bajo H_0 tenemos que $\beta = 0$, así tenemos que,

$$L(\sigma^2; \mathbf{e}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp^{-\frac{1}{2\sigma^2} \mathbf{Y}^t\mathbf{Y}}$$

$$\ln(L(\sigma^2; \mathbf{e})) = -\frac{n}{2}(\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} \mathbf{Y}^t\mathbf{Y}$$

$$\frac{\partial \ln(L(\cdot))}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \mathbf{Y}^t\mathbf{Y}$$

igualando a cero,

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \mathbf{Y}^t\mathbf{Y} = 0 \Rightarrow \hat{\sigma}_0^2 = \frac{\mathbf{Y}^t\mathbf{Y}}{n}$$

De aquí tenemos que el numerador de λ es:

$$\text{SupL}(\theta, \cdot) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} \exp^{-\frac{1}{2\hat{\sigma}_0^2} Y^t Y}$$

sabemos que $Y^t Y = \|Y\|^2$ de manera semejante, $(Y - X\hat{\beta})^t (Y - X\hat{\beta}) = \|Y - X\hat{\beta}\|^2$

$$\begin{aligned} \lambda &= \frac{\text{SupL}(\theta, \cdot)_{\theta \in \Theta_0}}{\text{SupL}(\theta, \cdot)_{\theta \in \Theta}} = \frac{\frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} \exp^{-\frac{1}{2\hat{\sigma}_0^2} \|Y\|^2}}{\frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp^{-\frac{1}{2\hat{\sigma}^2} \|Y - X\hat{\beta}\|^2}} \\ &= \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{\frac{n}{2}} \\ &= \left(\frac{\|Y - X\hat{\beta}\|^2 / n}{\|Y\|^2 / n} \right)^{\frac{n}{2}} \\ &= \left(\frac{\|Y - X\hat{\beta}\|^2}{\|Y\|^2 - \|X\hat{\beta}\|^2 + \|X\hat{\beta}\|^2} \right)^{\frac{n}{2}} \end{aligned}$$

como $\|Y - X\hat{\beta}\|^2 = \|Y\|^2 - \|X\hat{\beta}\|^2$, lo anterior nos queda como

$$\lambda = \left(\frac{1}{1 + \frac{p}{n-p} \frac{\|X\hat{\beta}\|^2 / p}{\|Y - X\hat{\beta}\|^2 / (n-p)}} \right)^{\frac{n}{2}}$$

Rechazamos H_0 si $\frac{\frac{\|\mathbf{X}\hat{\beta}\|^2}{p}}{\frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{n-p}} > k_0$, ahora busquemos el valor de k_0 , es fácil verificar

que $\frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{\sigma^2} \sim \chi^2_{(n-p)}$ y $\frac{\|\mathbf{X}\hat{\beta}\|^2}{\sigma^2} \sim \chi^2_{(p)}$ y como ambas formas cuadráticas son independientes

$$F = \frac{\frac{\|\mathbf{X}\hat{\beta}\|^2}{p}}{\frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{n-p}} \sim F_{(p, n-p)}$$

La probabilidad de cometer el error tipo I sería $P(F > k_0) = \alpha$, esto bajo H_0 , de aquí podemos deducir que $k_0 = F_{1-\alpha}(p, n-p)$.

Alo que hemos llegado, es que, rechazamos H_0 si

$$\frac{\frac{\|\mathbf{X}\hat{\beta}\|^2}{p}}{\frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{n-p}} > F_{1-\alpha}(p, n-p)$$

Lo que hemos hecho se resume en la siguiente tabla denominada como la tabla de Análisis de Varianza

Tabla de Análisis de Varianza				
Variación	g, l	Suma de Cuadrados.	Cuadrados medios	Prueba F
Debido a β	p	$\ \mathbf{X}\hat{\beta}\ ^2$	$\frac{\ \mathbf{X}\hat{\beta}\ ^2}{p}$	$\frac{\frac{\ \mathbf{X}\hat{\beta}\ ^2}{p}}{\frac{\ \mathbf{Y} - \mathbf{X}\hat{\beta}\ ^2}{n-p}}$
Error	$n-p$	$\ \mathbf{Y} - \mathbf{X}\hat{\beta}\ ^2$	$\frac{\ \mathbf{Y} - \mathbf{X}\hat{\beta}\ ^2}{n-p}$	
Total	N	$\ \mathbf{Y}\ ^2$	$\frac{\ \mathbf{Y}\ ^2}{n}$	

2.4.2 $H_0 : \beta_1 = 0$

Para el modelo lineal general con n observaciones y p coeficientes a estimar, que en forma vectorial tenemos

$$Y = X\beta + e$$

Si realizamos la siguiente partición sobre β , $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ $\begin{matrix} r \times 1 \\ (p-r) \times 1 \end{matrix}$ es decir β_1 es un vector de $r \times 1$ que contiene r coeficientes de β y β_2 es un vector de $(p-r) \times 1$ que contiene $p-r$ coeficientes de β (diferentes de β_1). Podemos particionar nuestro modelo original de la siguiente forma

$$Y = (X_1 \ X_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + e, \quad X_1 \text{ es de } n \times r \quad \text{y} \quad X_2 \text{ es de } n \times (p-r)$$

Deseamos buscar la prueba para $H_0: \beta_1=0$.

Usando el cociente de verosimilitud generalizada, lo que buscamos, es que,

$$\lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta, \cdot)}{\sup_{\theta \in \Theta} L(\theta, \cdot)} < k, \quad k \text{ una constante positiva}$$

para poder rechazar H_0 .

La función de verosimilitud del modelo es;

$$L(\sigma^2, \beta; e) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp^{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)}$$

el denominador de λ , sería,

$$\sup_{\theta \in \Theta} L(\theta, \cdot) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp^{-\frac{1}{2\hat{\sigma}^2}(Y-X\hat{\beta})'(Y-X\hat{\beta})}$$

donde $\hat{\beta} = (X'X)^{-1} X'Y$ y $\hat{\sigma}^2 = \frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{n}$

bajo H_0 tenemos que $\beta_1 = 0$, así tenemos que, la función de verosimilitud sería,

$$L(\sigma^2, \beta_2; \mathbf{e}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp^{-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}_2\beta_2)^t(\mathbf{Y}-\mathbf{X}_2\beta_2)}$$

Es claro que, el numerador de λ , es:

$$\text{Sup}_{\theta \in \Theta_0} L(\theta, \cdot) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} \exp^{-\frac{1}{2\hat{\sigma}_0^2}(\mathbf{Y}-\mathbf{X}_2\hat{\beta}_2)^t(\mathbf{Y}-\mathbf{X}_2\hat{\beta}_2)}$$

donde $\hat{\beta}_2 = (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{Y}$ y $\hat{\sigma}_0^2 = \frac{(\mathbf{Y}-\mathbf{X}_2\hat{\beta}_2)^t(\mathbf{Y}-\mathbf{X}_2\hat{\beta}_2)}{n}$

sabemos que $(\mathbf{Y}-\mathbf{X}\hat{\beta})^t(\mathbf{Y}-\mathbf{X}\hat{\beta}) = \|\mathbf{Y}-\mathbf{X}\hat{\beta}\|^2$

$$\begin{aligned} \lambda &= \frac{\text{Sup}_{\theta \in \Theta_0} L(\theta, \cdot)}{\text{Sup}_{\theta \in \Theta} L(\theta, \cdot)} \\ &= \frac{\frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} \exp^{-\frac{1}{2\hat{\sigma}_0^2}\|\mathbf{Y}-\mathbf{X}_2\hat{\beta}_2\|^2}}{\frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp^{-\frac{1}{2\hat{\sigma}^2}\|\mathbf{Y}-\mathbf{X}\hat{\beta}\|^2}} \\ &= \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{\frac{n}{2}} \\ &= \left(\frac{\|\mathbf{Y}-\mathbf{X}\hat{\beta}\|^2 / n}{\|\mathbf{Y}-\mathbf{X}_2\hat{\beta}_2\|^2 / n} \right)^{\frac{n}{2}} \\ &= \left(\frac{1}{1 + \frac{\|\mathbf{Y}-\mathbf{X}_2\hat{\beta}_2\|^2 - \|\mathbf{Y}-\mathbf{X}\hat{\beta}\|^2}{\|\mathbf{Y}-\mathbf{X}\hat{\beta}\|^2}} \right)^{\frac{n}{2}} \end{aligned}$$

$$\lambda = \left(\frac{1}{1 + \frac{r}{n-p} \frac{\|\mathbf{X}\hat{\beta}\|^2 - \|\mathbf{X}_2\hat{\beta}_2\|^2}{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}} \right)^{\frac{n}{2}}$$

Rechazamos H_0 si $\frac{\|\mathbf{X}\hat{\beta}\|^2 - \|\mathbf{X}_2\hat{\beta}_2\|^2}{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2} > k_0$, ahora busquemos el valor de k_0 , tenemos que

$$\|\mathbf{X}\hat{\beta}\|^2 = (\mathbf{X}\hat{\beta})^t (\mathbf{X}\hat{\beta}) = \mathbf{Y}^t \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

y de manera semejante

$$\|\mathbf{X}_2\hat{\beta}_2\|^2 = \mathbf{Y}^t \mathbf{X}_2 (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{Y}$$

de las dos anteriores desigualdades

$$\|\mathbf{X}\hat{\beta}\|^2 - \|\mathbf{X}_2\hat{\beta}_2\|^2 = \mathbf{Y}^t \left(\mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t - \mathbf{X}_2 (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \right) \mathbf{Y}$$

por otro lado tenemos que $\mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t = \mathbf{X}_2$, para la matriz de la forma cuadrática anterior, si hacemos $\mathbf{A} = \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t$ y $\mathbf{B} = \mathbf{X}_2 (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t$.

$(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B}) = \mathbf{A}^2 - \mathbf{AB} - \mathbf{BA} + \mathbf{B}^2 = \mathbf{A} - \mathbf{AB} - \mathbf{BA} + \mathbf{B}$, pero $\mathbf{AB} = \mathbf{B}$ lo que hemos probado que $\mathbf{A} - \mathbf{B}$ es idempotente, de aquí entonces podemos decir que,

$\frac{1}{\sigma^2} \mathbf{Y}^t \left(\mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t - \mathbf{X}_2 (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \right) \mathbf{Y} \sim \chi_{(r)}^2$ y $\frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{\sigma^2} \sim \chi_{(n-p)}^2$ y como ambas formas cuadráticas son independientes

$$F = \frac{\|\mathbf{X}\hat{\beta}\|^2 - \|\mathbf{X}_2\hat{\beta}_2\|^2}{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2} \sim F_{(r, n-p)}$$

La probabilidad de cometer el error tipo I sería $P(F > k_0) = \alpha$, esto bajo H_0 , de aquí podemos deducir que $k_0 = F_{1-\alpha}(r, n-p)$

A lo que hemos llegado, es que, rechazamos H_0 si

$$\frac{\frac{\|X\hat{\beta}\|^2 - \|X_2\hat{\beta}_2\|^2}{r}}{\frac{\|Y - X\hat{\beta}\|^2}{n-p}} > F_{1-\alpha}(r, n-p)$$

Lo que hemos hecho, se resume en la siguiente tabla de Análisis de Varianza.

Tabla de Análisis de Varianza				
Variación	g, l	Suma de Cuadrados.	Cuadrados medios	Prueba F
A la Hipótesis	r	$\ X\hat{\beta}\ ^2 - \ X_2\hat{\beta}_2\ ^2$	$\frac{\ X\hat{\beta}\ ^2 - \ X_2\hat{\beta}_2\ ^2}{r}$	$\frac{\ X\hat{\beta}\ ^2 - \ X_2\hat{\beta}_2\ ^2}{\ Y - X\hat{\beta}\ ^2} \cdot \frac{r}{n-p}$
Modelo reducido	$p-r$	$\ X_2\hat{\beta}_2\ ^2$	$\frac{\ X_2\hat{\beta}_2\ ^2}{p-r}$	
Error	$n-p$	$\ Y - X\hat{\beta}\ ^2$	$\frac{\ Y - X\hat{\beta}\ ^2}{n-p}$	
Total	n	$\ Y\ ^2$	$\frac{\ Y\ ^2}{n}$	